

UNIVERSIDAD NACIONAL AUTÓNOMA
DE MÉXICO

FACULTAD DE CIENCIAS

Una aplicación del teorema de Perron-Frobenius
a métodos iterativos: El Teorema de Stein-Rosenberg

T E S I S

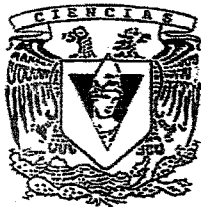
QUE PARA OBTENER EL TÍTULO DE:

MATEMÁTICO

P R E S E N T A :

RICARDO GUZMÁN MELGAREJO

DIRECTORA DE TESIS: DRA. BERTHA MARIA TOMÉ ARREOLA



FACULTAD DE CIENCIAS
UNAM

2009



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

FACULTAD DE CIENCIAS
Secretaría General
División de Estudios Profesionales

Votos Aprobatorios

ACT. MAURICIO AGUILAR GONZÁLEZ
Jefe de la División de Estudios Profesionales
Facultad de Ciencias
Presente

Por este medio hacemos de su conocimiento que hemos revisado el trabajo escrito titulado:

Una aplicación del Teorema de Perron-Frobenius a métodos iterativos: El teorema de Stein-Rosenberg

realizado por **Guzmán Melgarejo Ricardo** con número de cuenta **0-8401387-5** quien ha decidido titularse mediante la opción de tesis en la licenciatura en **Matemáticas**. Dicho trabajo cuenta con nuestro voto aprobatorio.

- Propietario Dr. Juan Morales Rodríguez *Juan Morales R.*
- Propietario Dr. Hugo Alberto Rincón Mejía *Hugo A. Rincón M.*
- Propietario Dra. Bertha María Tomé Arreola *Bertha Tomé*
- Tutora
- Suplente Dra. Diana Avella Alaminos *Diana Avella Alaminos*
- Suplente Dra. Berta Zavala Santana *Berta Zavala S.*

Atentamente,
"POR MI RAZA HABLARÁ EL ESPÍRITU"
Ciudad Universitaria, D. F., a 14 de enero de 2009
EL COORDINADOR DEL COMITÉ ACADÉMICO DE LA LICENCIATURA EN MATEMÁTICAS

M. EN C. FRANCISCO DE JESÚS STRUCK CHÁVEZ

Señor sinodal: antes de firmar este documento, solicite al estudiante que le muestre la versión digital de su trabajo y verifique que la misma incluya todas las observaciones y correcciones que usted hizo sobre el mismo.

Agradecimientos.

Agradezco a mi madre por que mi educación es fruto de su esfuerzo.

A mi esposa por permitirme formar con ella una familia.

A mis hermanas y hermanos por su ayuda y confianza.

A mi directora de tesis, Doctora Bertha Maria Tomé Arreola por su paciencia.

Gracias.

Ricardo Guzmán Melgarejo.

Índice.

Introducción	2
1. Nociones preliminares	3
1.1 Espacios con producto interno.....	3
1.2 Vectores y valores propios. Diagonalizabilidad	4
1.3 Operadores autoadjuntos, unitarios y normales	7
2. Teorema de Jordan	10
2.1 Vectores propios generalizados.....	10
2.2 Existencia de la base de Jordan	14
3. El teorema de Perron-Frobenius	21
3.1 Normas vectoriales	21
3.2 Normas matriciales	24
3.3 El radio espectral y la norma espectral	27
3.4 Matrices convergentes	30
3.5 Matrices irreducibles	33
3.6 El teorema de Perron-Frobenius para matrices irreducibles no negativas	40
4. Dos métodos iterativos clásicos y un teorema de comparación	47
4.1 Un problema modelo	47
4.2 Descripción de los métodos iterativos	51
4.3 Sobre la convergencia y estimación de errores	55
4.4 Velocidades de convergencia	60
4.5 Un teorema de comparación	68
5. Apéndice	71
Bibliografía	73

Introducción.

El álgebra lineal cobra sus mayores bríos cuando es aplicada. En el presente trabajo nos moveremos en la frontera del álgebra lineal avanzada y el álgebra lineal numérica, sin pretender en ningún momento instalarnos en el análisis numérico. El resultado que aquí demostraremos, a saber, el teorema de Stein-Rosenberg, es enunciado en múltiples libros de análisis numérico sin demostración y en algunos especializados en métodos iterativos como un resultado básico, estos hechos son la principal motivación para presentarlo de una manera breve como una aplicación del teorema de Perron-Frobenius. Este último es un teorema cuya demostración no es objeto de estudio de los cursos básicos de álgebra lineal que se imparten en nuestra facultad. Así llega a ser uno de nuestros principales objetivos enunciarlo y demostrarlo.

El capítulo uno contiene las nociones preliminares necesarias para el desarrollo de este trabajo.

En el capítulo dos se da la demostración clásica del teorema de Jordan para poner en contexto un lema que nos ilustra el comportamiento de las potencias de los bloques de Jordan, este último es la base del tratamiento que le damos a la convergencia de matrices.

Precisamente en el capítulo tres desarrollamos los resultados clásicos de la convergencia de matrices, definiendo normas de matrices y centrándonos en la que se conoce como norma espectral. Concluimos dicho capítulo con la demostración del teorema de Perron-Frobenius para matrices irreducibles no negativas, cuya demostración viene a ser consecuencia lógica de una serie de lemas. El teorema básico de Perron establece que si A es una matriz positiva entonces A tiene un valor propio positivo ρ de multiplicidad uno, tal que para todo valor propio λ_i de A se tiene que $\rho > |\lambda_i|$ y ρ tiene asociado un vector propio x positivo.

En el capítulo cuatro describimos los métodos iterativos de Jacobi y Gauss-Seidel para la resolución de sistemas de ecuaciones lineales, que son el objeto del teorema de comparación de Stein-Rosenberg. Se define la velocidad promedio de convergencia, la velocidad asintótica de convergencia para una matriz y se establece el teorema de comparación de Stein-Rosenberg en base al radio espectral de la matriz de iteración de Jacobi y de la matriz de iteración de Gauss-Seidel con la restricción de que estas últimas sean L-matrices o M-matrices según convenga para la simplicidad.

1 Nociones preliminares.

En el desarrollo del presente trabajo \mathbb{k} denota un subcampo de los números complejos \mathbb{C} y $M_n(\mathbb{k})$ el \mathbb{k} -espacio vectorial de las matrices de $n \times n$ con coeficientes en \mathbb{k} .

Con el propósito de hacer este trabajo autocontenido, en este capítulo presentamos una reseña de los resultados básicos del álgebra lineal que en él se utilizan y cuyas pruebas el lector puede encontrar en [5].

1.1 Espacios con producto interno.

Definición 1.1.1. Sea V un espacio vectorial sobre \mathbb{k} . Un *producto interno* en V es una función $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{k}$, tal que para todo $x, y, z \in V$ y todo $\alpha \in \mathbb{k}$ se tiene que:

$$i) \langle x + z, y \rangle = \langle x, y \rangle + \langle z, y \rangle.$$

$$ii) \langle \alpha x, y \rangle = \alpha \langle x, y \rangle.$$

$$iii) \langle x, y \rangle = \overline{\langle y, x \rangle}, \text{ donde la barra indica la conjugación compleja.}$$

$$iv) \langle x, x \rangle > 0 \text{ si } x \neq 0.$$

Ejemplo 1.1.1. Sea $V = \mathbb{k}^n$. Para $x = (x_1, x_2, \dots, x_n)$ y $y = (y_1, y_2, \dots, y_n)$ defínase

$$\langle x, y \rangle = \sum_{i=1}^n x_i \overline{y_i}.$$

Ejemplo 1.1.2. Sea $V = H([0, 2\pi])$ el espacio vectorial de las funciones continuas de valor complejo en $[0, 2\pi]$. Para $f, g \in V$, defínase $\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(t) \overline{g(t)} dt$.

Ejemplo 1.1.3. Sea $V = M_n(\mathbb{k})$ y defínase $\langle A, B \rangle = \text{tr}(B^* A)$ para $A, B \in V$, donde B^* representa la matriz transpuesta conjugada de B y $\text{tr}(A) = \sum_{i=1}^n a_{ii}$ es la traza de A .

Ejemplo 1.1.4. Sea $V = P_3[\mathbb{R}]$ y defínase $\langle p, q \rangle = \sum_{i=1}^3 p(x_i) q(x_i)$, donde $p, q \in P_3[\mathbb{R}]$, $x_1 = -1$, $x_2 = 0$ y $x_3 = 1$.

Definición 1.1.2. Sea V un espacio vectorial con producto interno. Para $x \in V$ definimos la *norma* (o *longitud*) de x mediante $\|x\| = \sqrt{\langle x, x \rangle}$.

Teorema 1.1.1. Sea V un espacio vectorial sobre \mathbb{k} con producto interno. Entonces para toda $x, y \in V$ y $\alpha \in \mathbb{k}$ tenemos:

$$i) \|x\| = 0 \text{ si y sólo si } x = 0; \text{ en cualquier caso } \|x\| \geq 0.$$

$$ii) \|\alpha x\| = |\alpha| \|x\|.$$

$$iii) |\langle x, y \rangle| \leq \|x\| \|y\| \text{ (desigualdad de Cauchy-Schwarz).}$$

$$iv) \|x + y\| \leq \|x\| + \|y\| \text{ (desigualdad del triángulo).}$$

Definición 1.1.3. Sea V un espacio vectorial sobre \mathbb{k} con producto interno. Un vector $x \in V$ es un vector *unitario* si $\|x\| = 1$. Los vectores x y y son *ortogonales* (*perpendiculares*) si $\langle x, y \rangle = 0$. Un subconjunto S de V es *ortogonal* si cualquier par de elementos distintos de S son ortogonales. Finalmente, un subconjunto S de V es *ortonormal* si S es ortogonal y está formado únicamente de vectores unitarios.

Ejemplo 1.1.5. Sean $V = H([0, 2\pi])$ y $S = \{e^{i\alpha x} = \cos \alpha x + i \sin \alpha x : \alpha \in \mathbb{Z}\}$. Claramente, $S \subseteq H$ y dado que $\langle e^{i\alpha x}, e^{i\beta x} \rangle = 0$ si $\alpha \neq \beta$ y $\langle e^{i\alpha x}, e^{i\alpha x} \rangle = 1$, se tiene que S es un subconjunto ortonormal de H .

Ejemplo 1.1.6. El conjunto $S = \{(1, 1, 1, 1), (-1, 4, 4, -1), (4, -1, 2, 0)\}$ es base de un subespacio W de dimensión 3 en \mathbb{R}^4 . Este conjunto no es ortogonal, sin embargo $S' = \{(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, \frac{1}{2}), (-\frac{1}{2}, \frac{1}{2}, \frac{1}{2}, -\frac{1}{2}), (\frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2})\}$ es una base ortonormal de W .

Si V es un espacio vectorial sobre \mathbb{k} con producto interno y S es un subconjunto finito linealmente independiente de V , el proceso de Gram-Schmidt nos permite construir una base ortogonal del subespacio generado por S . Normalizando los vectores de dicha base obtenemos una base ortonormal del subespacio generado por S . Como consecuencia de este hecho obtenemos el siguiente resultado.

Teorema 1.1.2. Si V es un espacio vectorial sobre \mathbb{k} con producto interno de dimensión finita entonces V tiene una base ortonormal.

1.2 Vectores y valores propios. Diagonalizabilidad.

Definición 1.2.1. Sea T un operador lineal sobre un \mathbb{k} -espacio vectorial V de dimensión finita. Se dice que T es *diagonalizable* si existe una base ordenada β para V tal que $[T]_\beta$ es una matriz diagonal.

Una matriz $A \in M_n(\mathbb{k})$ es *diagonalizable* si A es similar a una matriz diagonal.

Claramente, T es diagonalizable si y sólo si $[T]_\beta$ es diagonalizable para cualquier base ordenada β de V .

Definición 1.2.2. Sea T un operador lineal sobre un \mathbb{k} -espacio vectorial V de dimensión finita. Un elemento $0 \neq x \in V$ se llama *vector propio* de T si existe un escalar λ tal que $T(x) = \lambda x$. Al escalar λ se le llama *valor propio* de T correspondiente al vector x .

Análogamente, si $A \in M_n(\mathbb{k})$, un elemento $0 \neq x \in \mathbb{k}^n$ se denomina *vector propio* de A si $Ax = \lambda x$ para algún escalar λ . Al escalar λ se le llama *valor propio* de A correspondiente al vector x .

Se sigue de las definiciones anteriores que un operador lineal T sobre un \mathbb{k} -espacio vectorial V de dimensión finita es diagonalizable si y sólo si existe una base ordenada β de V compuesta por vectores propios de T . En este caso, $[T]_\beta$ es una matriz diagonal cuyos elementos diagonales son los valores propios de T .

Teorema 1.2.1. Sea T un operador lineal sobre un \mathbb{k} -espacio vectorial V de dimensión finita. Entonces:

- 1) Un escalar $\lambda \in \mathbb{k}$ es valor propio de T si y sólo si $\det(T - \lambda I_V) = 0$.
- 2) Si λ es un valor propio de T , un vector $x \in V$ es un vector propio de T correspondiente a λ si y sólo si $x \neq 0$ y $x \in \text{Ker}(T - \lambda I_V)$.

Corolario 1.2.1. Sea $A \in M_n(\mathbb{k})$. Un escalar $\lambda \in \mathbb{k}$ es un valor propio de A si y sólo si $\det(A - \lambda I) = 0$. Si λ es un valor propio de A , un vector $x \in \mathbb{k}^n$ es un vector propio de A correspondiente a λ si y sólo si $x \neq 0$ y $x \in \text{Ker}(A - \lambda I)$.

Sean T un operador lineal sobre un \mathbb{k} -espacio vectorial V de dimensión finita, β una base ordenada de V y $A = [T]_\beta$. Ya que $\det(T - \lambda I_V) = \det(A - \lambda I)$ se tiene que los valores

propios de T son los de A .

Definición 1.2.3. Sea $A \in M_n(\mathbb{k})$. El polinomio $\det(A - tI_n)$ en la incógnita t se denomina *polinomio característico* de A .

Los valores propios de A son entonces las raíces o ceros del polinomio característico de A y como éste es de grado n , A tiene a lo más n valores propios distintos.

Se prueba fácilmente que las matrices similares tienen el mismo polinomio característico por lo que podemos dar la siguiente definición.

Definición 1.2.4. Sea T un operador lineal sobre un \mathbb{k} -espacio vectorial V de dimensión finita con base ordenada β . Definimos al *polinomio característico* $f(t)$ de T como el polinomio característico de $A = [T]_\beta$; esto es,

$$f(t) = \det(A - tI).$$

Al igual que para las matrices, los valores propios de T son las raíces del polinomio característico de T , por lo que T tiene a lo más $\dim V$ valores propios distintos. Es importante señalar que T puede no tener valores propios, como lo muestra el siguiente ejemplo.

Ejemplo 1.2.1. Sea T el operador lineal sobre \mathbb{R}^2 representado en la base canónica por la matriz $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. Entonces el polinomio característico de A (y por lo tanto de T) es $f(x) = \det(A - xI) = x^2 + 1$, cuyas raíces son complejas. Luego T no tiene valores propios reales.

Definición 1.2.5. Sea T un operador lineal sobre un \mathbb{k} -espacio vectorial V y sea λ un valor propio de T . Definamos

$$E_\lambda := \{x \in V : T(x) = \lambda x\} = \text{Ker}(T - \lambda I_V)$$

El conjunto E_λ se denomina el *espacio propio* de T correspondiente al valor propio λ . Análogamente, si $A \in M_n(\mathbb{k})$ y λ es un valor propio de A , el *espacio propio* de A correspondiente a λ es $E_\lambda = \text{Ker}(A - \lambda I)$.

Los siguientes resultados están encaminados a determinar cuándo un operador lineal o una matriz son diagonalizables.

Teorema 1.2.2. Sea T un operador lineal sobre un \mathbb{k} -espacio vectorial V y sean $\lambda_1, \dots, \lambda_k$ valores propios distintos de T . Si x_1, \dots, x_k son vectores propios de T correspondientes a $\lambda_1, \dots, \lambda_k$ respectivamente, entonces $\{x_1, \dots, x_k\}$ es linealmente independiente.

Corolario 1.2.2. Sea T un operador lineal sobre un espacio vectorial V de dimensión n . Si T tiene n valores propios distintos, entonces T es diagonalizable.

Corolario 1.2.3. Sea T un operador lineal sobre un espacio vectorial V y sean $\lambda_1, \dots, \lambda_k$ valores propios distintos de T . Entonces

$$E_{\lambda_i} \cap \sum_{j \neq i} E_{\lambda_j} = \{0\} \quad 1 \leq i \leq k.$$

Esto es, los espacios propios correspondientes a valores propios distintos de T son independientes.

Cuando un operador lineal T sobre un espacio vectorial V de dimensión n es diagonalizable, el polinomio característico de T se descompone en un producto de n factores lineales no necesariamente distintos en $\mathbb{k}[x]$.

Definición 1.2.6. Sea λ un valor propio de un operador lineal T o de una matriz A cuyo polinomio característico es $f(t)$. La *multiplicidad algebraica* de λ es el mayor entero positivo k para el que $(t - \lambda)^k$ es factor de $f(t)$.

Lema 1.2.1. Sea T un operador lineal sobre un espacio vectorial V de dimensión finita. Si λ es un valor propio de T de multiplicidad m , entonces $1 \leq \dim E_\lambda \leq m$.

Finalmente, tenemos

Teorema 1.2.3. Sea T un operador lineal sobre un \mathbb{k} -espacio vectorial V de dimensión n tal que su polinomio característico se descompone en un producto de factores lineales en $\mathbb{k}[x]$ y sean $\lambda_1, \dots, \lambda_k$ los valores propios distintos de T . Entonces son equivalentes:

- T es diagonalizable.
- $V = E_{\lambda_1} \oplus \dots \oplus E_{\lambda_k}$.
- $n = \dim E_{\lambda_1} + \dots + \dim E_{\lambda_k}$.
- Si m_j es la multiplicidad de λ_j , entonces $\dim E_{\lambda_j} = m_j$ para $1 \leq j \leq k$.

Ejemplo 1.2.2. Sea T un operador lineal en \mathbb{R}^3 definido mediante

$$T(x_1, x_2, x_3) = (4x_1 + x_3, 2x_1 + 3x_2 + 2x_3, x_1 + 4x_3).$$

Si β es la base ordenada canónica para \mathbb{R}^3 entonces

$$[T]_\beta = \begin{pmatrix} 4 & 0 & 1 \\ 2 & 3 & 2 \\ 1 & 0 & 4 \end{pmatrix},$$

por lo tanto, el polinomio característico de T es

$$\det([T]_\beta - tI) = \det \begin{pmatrix} 4-t & 0 & 1 \\ 2 & 3-t & 2 \\ 1 & 0 & 4-t \end{pmatrix} = -(t-5)(t-3)^2.$$

Luego, los valores propios de T son $\lambda_1 = 5$ y $\lambda_2 = 3$ con multiplicidades 1 y 2 respectivamente. Dado que

$$E_{\lambda_1} = \text{Ker}(T - \lambda_1 I) = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} -1 & 0 & 1 \\ 2 & -2 & 2 \\ 1 & 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\},$$

E_{λ_1} es el espacio generado por $\begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}$, de donde $\dim E_{\lambda_1} = 1$.

De manera análoga, E_{λ_2} es el espacio generado por $\left\{ \left(\begin{array}{c} 1 \\ 0 \\ -1 \end{array} \right), \left(\begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right) \right\}$ y

$\dim E_{\lambda_2} = 2$.

En este caso la multiplicidad de cada valor propio λ_i es igual a la dimensión del espacio propio correspondiente E_{λ_i} y $\left\{ \left(\begin{array}{c} 1 \\ 2 \\ 1 \end{array} \right), \left(\begin{array}{c} 1 \\ 0 \\ -1 \end{array} \right), \left(\begin{array}{c} 0 \\ 1 \\ 0 \end{array} \right) \right\}$ es una base para \mathbb{R}^3 que consta de vectores propios de T . Por lo tanto, T es diagonalizable.

Ejemplo 1.2.3. Sea T un operador lineal en \mathbb{R}^3 definido mediante

$$T(x_1, x_2, x_3) = (x_2, x_3, x_1 - 3x_2 + 3x_3).$$

Si β es la base ordenada canónica para \mathbb{R}^3 entonces

$$[T]_{\beta} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & -3 & 3 \end{pmatrix},$$

por lo tanto, el polinomio característico de T es

$$\det([T]_{\beta} - tI) = \det \begin{pmatrix} -t & 1 & 0 \\ 0 & -t & 1 \\ 1 & -3 & 3-t \end{pmatrix} = -(t-1)^3.$$

Luego, los valores propios de T son $\lambda_1 = 1$ con multiplicidad 3. Dado que

$$E_{\lambda_1} = \text{Ker}(T - \lambda_1 I) = \left\{ \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in \mathbb{R}^3 : \begin{pmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & -3 & 2 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \right\}$$

E_{λ_1} es el espacio generado por $\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$, de donde $\dim E_{\lambda_1} = 1 \neq 3$. Por lo tanto, T no es diagonalizable.

1.3 Operadores autoadjuntos, unitarios y normales.

Cuando V es \mathbb{k} -un espacio con producto interno de dimensión finita cualquier funcional f sobre V es el producto interno con un único vector fijo en V , esto es, existe un único vector $y \in V$ tal que $f = \langle _, y \rangle$. Este resultado permite probar la existencia del operador adjunto de un operador lineal T sobre V .

Teorema 1.3.1. Sean V un \mathbb{k} -espacio vectorial con producto interno de dimensión finita y T un operador lineal sobre V . Entonces existe un único operador lineal T^* sobre V , llamado el *adjunto* de T , tal que $\langle T(x), y \rangle = \langle x, T^*(y) \rangle$ para cualesquiera $x, y \in V$.

Un resultado útil para obtener adjuntos es la siguiente proposición.

Proposición 1.3.1. Sean V un \mathbb{k} -espacio vectorial con producto interno de dimensión finita, β una base ordenada ortonormal de V , T un operador lineal sobre V y $A = [T]_{\beta}$. Entonces $[T^*]_{\beta} = A^*$, la matriz adjunta o transpuesta conjugada de A .

En lo que sigue, V es un \mathbb{k} -espacio vectorial con producto interno de dimensión finita.

Definición 1.3.1. Un operador lineal T sobre V se llama *autoadjunto* (o *Hermitiano*) si $T = T^*$. Una matriz $A \in M_n(\mathbb{k})$ se llama *autoadjunta* (o *Hermitiana*) si $A = A^*$.

Definición 1.3.2. Un operador lineal T sobre V se llama *unitario* si $UU^* = I_V = U^*U$. Una matriz $A \in M_n(\mathbb{k})$ se llama *unitaria* si $AA^* = I$. Cuando $\mathbb{k} = \mathbb{R}$ tales operadores o matrices se llaman también *ortogonales*.

Definición 1.3.3. Un operador lineal T sobre V se llama *normal* si $TT^* = T^*T$. Una matriz $A \in M_n(\mathbb{k})$ se llama *normal* si $AA^* = A^*A$.

Claramente, todo operador autoadjunto o unitario es normal. Además, si β es una base ordenada ortonormal de V , entonces T es autoadjunto, unitario o normal si y sólo si $[T]_{\beta}$ es autoadjunta, unitaria o normal, respectivamente.

Ejemplo 1.3.1. Sea $0 < \theta < \pi$ y defínase a $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ como la rotación por un ángulo θ . La matriz de T en la base ordenada canónica estándar es $A = \begin{pmatrix} \cos \theta & -\operatorname{sen} \theta \\ \operatorname{sen} \theta & \cos \theta \end{pmatrix}$. Claramente, $AA^* = A^*A$, pero $A \neq A^*$. Esto significa que A es normal y es unitaria pero no es autoadjunta.

Las principales características de estos operadores están contenidas en los siguientes resultados.

Teorema 1.3.2. Sean V un \mathbb{k} -espacio vectorial con producto interno de dimensión finita y T un operador lineal autoadjunto sobre V . Entonces:

- Todo valor propio de T es real.
- Vectores propios asociados a valores propios distintos de T son ortogonales.
- Si $\mathbb{k} = \mathbb{R}$, T tiene un valor propio real.

Teorema 1.3.3. Sean V un \mathbb{k} -espacio vectorial con producto interno de dimensión finita y U un operador lineal sobre V . Las siguientes condiciones son equivalentes:

- U es unitario.
- U es un isomorfismo de espacios con producto interno.
- $\langle U(x), U(y) \rangle = \langle x, y \rangle$ para cualesquier $x, y \in V$.
- $\|U(x)\| = \|x\|$ para todo $x \in V$.
- Para toda base ortonormal β de V , $T(\beta)$ es base ortonormal de V .
- Existe una base ortonormal β de V tal que $T(\beta)$ es base ortonormal de V .

Teorema 1.3.4. Sean V un \mathbb{k} -espacio vectorial con producto interno de dimensión finita y T un operador lineal normal sobre V . Entonces:

- $\|T(x)\| = \|T^*(x)\|$ para todo $x \in V$.
- Si $x \in V$, x es un vector propio de T con valor propio λ si y sólo si x es un vector propio de T^* con valor propio $\bar{\lambda}$.

c) Vectores propios asociados a distintos valores propios de T son ortogonales.

La importancia de estos operadores radica en el siguiente teorema.

Teorema espectral 1.3.5. Sean V un \mathbb{k} -espacio vectorial de dimensión finita con producto interno y T un operador lineal sobre V .

a) Si $\mathbb{k} = \mathbb{R}$, entonces T es autoadjunto si y sólo si V tiene una base ortonormal compuesta por vectores propios de T .

b) Si $\mathbb{k} = \mathbb{C}$, entonces T es normal si y sólo si V tiene una base ortonormal compuesta por vectores propios de T .

Corolario 1.3.1. Sea T un operador lineal sobre un espacio vectorial real V con producto interno de dimensión finita. Entonces T es autoadjunto y ortogonal si y sólo si V tiene una base ortonormal de vectores propios de T con valores propios de valor absoluto 1.

Corolario 1.3.2. Sea T un operador lineal sobre un espacio vectorial complejo V con producto interno de dimensión finita. Entonces T es unitario si y sólo si V tiene una base ortonormal de vectores propios de T con valores propios de módulo 1.

Para dar la versión matricial del teorema espectral necesitamos la siguiente definición.

Definición 1.3.4. Sean $A, B \in M_n(\mathbb{k})$. Se dice que B es unitariamente equivalente a la matriz A si existe una matriz unitaria P tal que $B = P^{-1}AP = P^*AP$.

Corolario 1.3.3. Sea $A \in M_n(\mathbb{R})$. Entonces A es autoadjunta (simétrica) si y sólo si A es ortogonalmente equivalente a una matriz diagonal real.

Corolario 1.3.4. Sea $A \in M_n(\mathbb{C})$. Entonces A es normal si y sólo si A es unitariamente equivalente a una matriz diagonal.

2 El teorema de Jordan.

Como se ha exhibido en el capítulo anterior, no todos los operadores lineales y, por consiguiente, no todas las matrices cuadradas son diagonalizables, de hecho si lo fueran, no habría casi nada que hacer en la teoría de matrices. Sin embargo, muchos problemas, por ejemplo en geometría y ecuaciones diferenciales, se pueden atacar a través de una variación de los planteamientos hechos en el capítulo anterior, aunque no tengamos matrices diagonales de que echar mano. Nos bastará contar con matrices suficientemente sencillas, estructuralmente “casi diagonales”.

2.1 Vectores propios generalizados.

Definición 2.1.1. Sea T un operador lineal sobre un \mathbb{k} -espacio vectorial V . Un subespacio W de V se llama subespacio T -invariante de V si $T(W) \subseteq W$, esto es, $T(x) \in W$ para toda $x \in W$.

Son subespacios T -invariantes $\{0\}$, V , $\text{Im} T$ y $\text{Ker} T$. Si x es un vector propio de T , entonces el subespacio generado por $\{x\}$ es T -invariante; más aún, dado un valor propio λ de T , E_λ es T -invariante. Cuando el subespacio W es T -invariante, T induce un operador lineal T_W en el espacio W . El operador T_W está definido por $T_W(x) = T(x)$ para todo $x \in W$.

Definición 2.1.2. Un bloque de Jordan $J_n(\lambda)$ es una matriz de $n \times n$ de la forma

$$\begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

El único valor propio de $J_n(\lambda)$ es λ . Podemos escribir también $J_n(\lambda) = \lambda I_n + J_n(0)$

Una matriz de Jordan se construye por medio de bloques de Jordan colocados en forma diagonal, de la manera siguiente:

$$\begin{pmatrix} J_{n_1}(\lambda_1) & 0 & 0 & \dots & 0 & 0 \\ 0 & J_{n_2}(\lambda_2) & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & J_{n_{s-1}}(\lambda_{s-1}) & 0 \\ 0 & 0 & 0 & \dots & 0 & J_{n_s}(\lambda_s) \end{pmatrix}$$

Esta matriz también se denota por $J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \dots \oplus J_{n_s}(\lambda_s)$.

Para los fines que se persiguen en el presente trabajo, será suficiente considerar el caso de los operadores T sobre un \mathbb{k} -espacio vectorial V de dimensión finita cuyos polinomios característicos se descomponen en un producto de factores de grado 1. Demostraremos en este caso que existe una base ordenada β de V tal que:

$$A = [T]_{\beta} = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \cdots \oplus J_{n_s}(\lambda_s).$$

La matriz A se denomina *forma canónica de Jordan* de T y la base β se denomina *base canónica de Jordan*.

Consideremos por ejemplo, la matriz de $J \in M_6(\mathbb{R})$

$$J = J_1 \oplus J_2 \oplus J_3 = \begin{pmatrix} a & 1 & 0 & 0 & 0 & 0 \\ 0 & a & 1 & 0 & 0 & 0 \\ 0 & 0 & a & 0 & 0 & 0 \\ 0 & 0 & 0 & b & 0 & 0 \\ 0 & 0 & 0 & 0 & c & 1 \\ 0 & 0 & 0 & 0 & 0 & c \end{pmatrix}$$

J es una forma canónica de Jordan de un operador lineal $T : \mathbb{R}^6 \rightarrow \mathbb{R}^6$, es decir, existe una base $\beta = \{x_1, x_2, \dots, x_6\}$ para \mathbb{R}^6 tal que $[T]_{\beta} = J$. Aquí el polinomio característico de T y, por consiguiente de J es $f(t) = \det(J - tI) = (t - a)^3(t - c)^2(t - b)$.

Por otra parte, de los vectores x_1, x_2, \dots, x_6 solamente x_1, x_4 y x_5 son vectores propios de T , esto significa que x_2, x_3 y x_6 no lo son. Sin embargo, como $[T]_{\beta} = J$, $T(x_2) = x_1 + ax_2$ y $T(x_3) = x_2 + ax_3$. Entonces

$$(T - aI)^2(x_2) = (T - aI)(T(x_2) - ax_2) = (T - aI)(x_1) = 0$$

y similarmente,

$$(T - aI)^3(x_3) = (T - aI)^2(T(x_3) - ax_3) = (T - aI)^2(x_2) = 0$$

Así, aún cuando $(T - aI)(x_2) \neq 0$ y $(T - aI)(x_3) \neq 0$, $(T - aI)^p(x_2) = 0 = (T - aI)^p(x_3)$ para $p \geq 3$. Si llamamos $\beta_1 = \{x_1, x_2, x_3\}$, $\beta_2 = \{x_4\}$, $\beta_3 = \{x_5, x_6\}$, entonces el subespacio W_i generado por β_i , $1 \leq i \leq 3$, es un subespacio T -invariante y, $[T_{W_i}]_{\beta_i} = J_i$, para $1 \leq i \leq 3$.

Definición 2.1.3. Sea T un operador lineal en un \mathbb{k} -espacio vectorial V de dimensión finita. Un elemento $x \neq 0 \in V$ se llama *vector propio generalizado* de T si existe un escalar λ tal que $(T - \lambda I)^p(x) = 0$ para algún entero positivo p . Diremos que x es un vector propio generalizado correspondiente a λ .

Definición 2.1.4. Sea T un operador lineal en un \mathbb{k} -espacio vectorial V de dimensión finita y sea x un vector propio generalizado de T correspondiente al valor propio λ . Si p es el entero positivo más pequeño tal que $(T - \lambda I)^p(x) = 0$, entonces el conjunto ordenado

$$\{(T - \lambda I)^{p-1}(x), (T - \lambda I)^{p-2}(x), \dots, (T - \lambda I)(x), x\}$$

se llama *ciclo de vectores propios generalizados* de T que corresponden a λ . Los elementos $(T - \lambda I)^{p-1}(x)$ y x se llaman *vector inicial* y *vector terminal* del ciclo respectivamente y p es la *longitud* del ciclo.

Teorema 2.1.1. Sea T un operador lineal en un \mathbb{k} -espacio vectorial V de dimensión finita y sea γ un ciclo de vectores propios generalizados de T que corresponden al valor propio λ . Entonces:

a) El vector inicial de γ es un vector propio de T correspondiente al valor propio λ y ningún otro elemento de γ es vector propio de T .

b) γ es linealmente independiente.

c) Sea β una base ordenada de V . Entonces β es una base canónica de Jordan para T si y sólo si β es unión ajena de ciclos de vectores propios generalizados de T .

Demostración

Para a), sea $\gamma = \{(T - \lambda I)^{p-1}(x), (T - \lambda I)^{p-2}(x), \dots, (T - \lambda I)(x), x\}$ un ciclo de vectores propios generalizados de longitud p . Consideremos el vector inicial $y = (T - \lambda I)^{p-1}(x)$, aplicándole $(T - \lambda I)$ obtenemos

$$(T - \lambda I)(y) = (T - \lambda I)(T - \lambda I)^{p-1}(x) = (T - \lambda I)^p(x) = 0,$$

lo cual implica que $(T - \lambda I)^{p-1}(x)$ es vector propio de T correspondiente a λ . Por otra parte, como $(T - \lambda I)(T - \lambda I)^{p-j}(x) \neq 0$ para $2 \leq j \leq p$, tenemos que ningún otro vector del ciclo es vector propio de T correspondiente a λ .

La demostración para b) se hará por inducción sobre la longitud del ciclo γ . Si γ tiene longitud 1, entonces $\gamma = \{x\}$ es linealmente independiente ya que x es vector propio y, por lo tanto, no nulo.

Supongamos que los ciclos de longitud $k - 1$ son linealmente independientes para algún entero $k - 1 \geq 1$.

Supongamos que $\gamma = \{(T - \lambda I)^{k-1}(x), (T - \lambda I)^{k-2}(x), \dots, (T - \lambda I)(x), x\}$ es un ciclo de vectores propios generalizados que corresponden al valor propio λ y que

$$\sum_{i=1}^k \alpha_{k-i} (T - \lambda I)^{k-i}(x) = 0,$$

donde $\alpha_i \in \mathbb{k} \quad \forall i = 1, 2, \dots, k$. Aplicando $(T - \lambda I)$ a esta ecuación se tiene

$$\sum_{i=1}^k \alpha_{k-i} (T - \lambda I)^{k-i+1}(x) = \sum_{i=2}^k \alpha_{k-i} (T - \lambda I)^{k-i+1}(x) = 0,$$

donde esta suma es una combinación lineal del ciclo $\{(T - \lambda I)^{k-1}(x), (T - \lambda I)^{k-2}(x), \dots, (T - \lambda I)(x)\}$ de longitud $k - 1$. Por lo tanto, $\alpha_{k-i} = 0$ para $i = 2, 3, \dots, k$, por lo que

$$\sum_{i=1}^k \alpha_{k-i} (T - \lambda I)^{k-i}(x) = 0$$

se reduce a $\alpha_{k-1}(T - \lambda I)^{k-1}(x) = 0$. Pero como $(T - \lambda I)^{k-1}(x) \neq 0$, se tiene que $\alpha_{k-1} = 0$, de manera que $\alpha_{k-i} = 0 \quad \forall i = 1, 2, \dots, k$. Por lo tanto, γ es linealmente independiente.

Demostremos finalmente c). En efecto, supongamos que β es una base canónica de Jordan para T , es decir,

$$[T]_{\beta} = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \cdots \oplus J_{n_s}(\lambda_s),$$

donde $n_1 + n_2 + \cdots + n_s = \dim V$. Para $1 \leq k \leq s$, sea $\beta_k = \{x_1, x_2, \dots, x_{n_k}\}$ el subconjunto de β que da lugar al bloque de Jordan $J_{n_k}(\lambda_k)$. Entonces $T(x_1) = \lambda_k x_1, T(x_2) = x_1 + \lambda_k x_2, \dots, T(x_{n_k}) = x_{n_k-1} + \lambda_k x_{n_k}$, esto es, $(T - \lambda_k I)(x_1) = 0, x_1 = (T - \lambda_k I)(x_2), \dots, x_{n_k-1} = (T - \lambda_k I)(x_{n_k})$. Por lo tanto, $\beta_k = \{(T - \lambda_k I)^{n_k-1}(x_{n_k}), (T - \lambda_k I)^{n_k-2}(x_{n_k}), \dots, x_{n_k}\}$ es un ciclo de vectores propios generalizados de T correspondientes a λ_k .

Claramente, $\beta = \bigcup_{k=1}^s \beta_k$ es la unión ajena de ciclos de vectores propios generalizados de T .

Inversamente, si $\beta = \bigcup_{k=1}^s \beta_k$, donde

$$\beta_k = \{(T - \lambda_k I)^{n_k-1}(x), (T - \lambda_k I)^{n_k-2}(x), \dots, (T - \lambda_k I)(x), x\},$$

es un ciclo de longitud n_k de vectores propios generalizados de T correspondientes a λ_k , entonces el subespacio W_k generado por β_k es T -invariante y $[T_{W_k}]_{\beta_k} = J_{n_k}(\lambda_k)$, lo cual implica que

$$[T]_{\beta} = J_{n_1}(\lambda_1) \oplus J_{n_2}(\lambda_2) \oplus \cdots \oplus J_{n_s}(\lambda_s)$$

quedando demostrado el teorema. \square

Definición 2.1.5. Sea λ un valor propio de un operador lineal T en un \mathbb{k} -espacio vectorial V de dimensión finita. El *espacio propio generalizado* de T correspondiente a λ y denotado por K_{λ} es el conjunto

$$K_{\lambda} = \{x \in V : (T - \lambda I)^p(x) = 0 \text{ para algún entero positivo } p\}.$$

Luego, K_{λ} consta del vector nulo y de todos los vectores propios generalizados de T correspondientes a λ .

Teorema 2.1.2. Sea λ un valor propio de un operador lineal T en un \mathbb{k} -espacio vectorial V de dimensión n . Entonces K_{λ} es un subespacio de V que contiene a E_{λ} .

Demostración

Es claro que $0 \in K_{\lambda}$. Supongamos que $x, y \in K_{\lambda}$, entonces existen enteros positivos p y q tales que $(T - \lambda I)^p(x) = 0$ y $(T - \lambda I)^q(y) = 0$. De aquí

$$\begin{aligned} (T - \lambda I)^{p+q}(x + y) &= (T - \lambda I)^{p+q}(x) + (T - \lambda I)^{p+q}(y) \\ &= (T - \lambda I)^q(T - \lambda I)^p(x) + (T - \lambda I)^p(T - \lambda I)^q(y) \\ &= (T - \lambda I)^q(0) + (T - \lambda I)^p(0) \\ &= 0 + 0 = 0, \end{aligned}$$

y entonces $x + y \in K_{\lambda}$. Finalmente, para cualquier escalar α ,

$$(T - \lambda I)^p(\alpha x) = \alpha(T - \lambda I)^p(x) = \alpha 0 = 0,$$

es decir, $\alpha x \in K_\lambda$. Por lo tanto K_λ es subespacio de V .

Claramente, $E_\lambda = \text{Ker}(T - \lambda I) \subseteq K_\lambda$. \square

Teorema 2.1.3. Sean $\lambda_1, \lambda_2, \dots, \lambda_k$ los distintos valores propios de un operador lineal T en un \mathbb{k} -espacio vectorial V de dimensión n . Entonces

$$K_{\lambda_i} \cap \sum_{j \neq i} K_{\lambda_j} = \{0\} \quad \text{para } i = 1, 2, \dots, k.$$

Demostración

Sin pérdida de generalidad, sea $i = 1$. Supongamos que

$$\sum_{j=2}^k x_j = x_1,$$

con $x_j \in K_{\lambda_j}$ para $1 \leq j \leq k$.

Sea p_j ($1 \leq j \leq k$) el entero positivo más pequeño tal que $(T - \lambda_j I)^{p_j}(x_j) = 0$. Supongamos que $x_1 \neq 0$, entonces $(T - \lambda_1 I)^{p_1-1}(x_1)$ es un vector propio correspondiente al valor propio λ_1 . Aplicando $(T - \lambda_1 I)^{p_1-1}(T - \lambda_2 I)^{p_2} \dots (T - \lambda_k I)^{p_k}$ a ambos lados de la ecuación anterior y en virtud de que para cualquier polinomio $g(x) \in \mathbb{k}(x)$, $g(T)(T - \lambda_1 I)^{p_1-1}(x_1) = g(\lambda_1)(T - \lambda_1 I)^{p_1-1}(x_1)$ se tiene que

$$\begin{aligned} 0 &= (T - \lambda_1 I)^{p_1-1}(T - \lambda_2 I)^{p_2} \dots (T - \lambda_k I)^{p_k}(x_1) \\ &= (T - \lambda_2 I)^{p_2} \dots (T - \lambda_k I)^{p_k}(T - \lambda_1 I)^{p_1-1}(x_1) \\ &= (\lambda_1 - \lambda_2)^{p_2} \dots (\lambda_1 - \lambda_k)^{p_k}(T - \lambda_1 I)^{p_1-1}(x_1) \end{aligned}$$

Por lo tanto, dado que $\lambda_1, \lambda_2, \dots, \lambda_k$ son distintos, se tiene que $(T - \lambda_1 I)^{p_1-1}(x_1) = 0$, contradiciendo el hecho de que $(T - \lambda_1 I)^{p_1-1}(x_1)$ es un vector propio. Así $x_1 = 0$, de donde se deduce que $K_{\lambda_i} \cap \sum_{j \neq i} K_{\lambda_j} = \{0\}$ para $i = 1, 2, \dots, k$. \square

2.2 Existencia de la base de Jordan.

Lema 2.2.1 Sea T un operador lineal en un \mathbb{k} -espacio vectorial V de dimensión finita. Sea S_i ($1 \leq i \leq k$) un ciclo de vectores propios generalizados de T correspondientes al valor propio λ y sean p_i y y_i la longitud y el vector inicial de S_i , respectivamente. Si $\{y_1, y_2, \dots, y_k\}$ es un conjunto linealmente independiente que contiene k elementos, entonces

$$\bigcup_{i=1}^k S_i$$

es un conjunto linealmente independiente que contiene $\sum_{i=1}^k p_i$ elementos.

Demostración

Supongamos sin pérdida de generalidad que $p_1 \geq p_2 \geq \dots \geq p_k$. La demostración se hará por inducción sobre p_1 .

Si $p_1 = 1$, entonces $p_1 = p_2 = \dots = p_k = 1$. Por lo tanto, cada S_i contiene un solo elemento, de manera que

$$\bigcup_{i=1}^k S_i = \{y_1, y_2, \dots, y_k\}$$

es un conjunto linealmente independiente que por hipótesis contiene $\sum_{i=1}^k p_i = k$ elementos.

Supongamos ahora, que el teorema es cierto para $p_1 < n$ y sea S_i ($1 \leq i \leq k$) un ciclo de vectores propios generalizados de T correspondientes al valor propio λ con longitud p_i y vector inicial y_i . Supongamos que $n = p_1 \geq p_2 \geq \dots \geq p_k$ y sea r ($1 \leq r \leq k$) el subíndice más grande tal que $p_r > 1$. Sea

$$S = \bigcup_{i=1}^k S_i,$$

y sea S'_i ($1 \leq i \leq r$) el ciclo obtenido al suprimir el vector terminal x_i de S_i . Entonces S'_i es un ciclo de vectores propios generalizados que corresponde al valor propio λ con longitud $p_i - 1$ y vector inicial y_i . Como $\{y_1, \dots, y_r\}$ es linealmente independiente, se tiene por hipótesis de inducción que

$$S' = \bigcup_{i=1}^r S'_i$$

es un conjunto linealmente independiente que contiene $\sum_{i=1}^r (p_i - 1)$ elementos. Es claro ahora que

$$\bigcup_{i=1}^k S_i = S' \cup \{x_1, \dots, x_k\}$$

es una unión ajena. Entonces $\bigcup_{i=1}^k S_i$ contiene $\sum_{i=1}^r (p_i - 1) + k = \sum_{i=1}^k (p_i - 1) + k = \sum_{i=1}^k p_i$ elementos.

Finalmente, demostraremos que $S = \bigcup_{i=1}^k S_i$ es un conjunto linealmente independiente. Supongamos que para algunos escalares α_z

$$\sum_{z \in S} \alpha_z z = 0 \tag{1}$$

Por el teorema 2.1.1 a), y_i es un vector propio de T correspondiente al valor propio λ , luego $(T - \lambda I)(y_i) = 0$ para $1 \leq i \leq k$. Por lo tanto, al aplicar $(T - \lambda I)$ a ambos lados de la ecuación (1) se tiene

$$0 = \sum_{z \in S} \alpha_z (T - \lambda I)(z) = \sum_{z \in Z} \alpha_z (T - \lambda I)(z), \quad (2)$$

donde $Z = \{v \in S : v \neq y_i \text{ para } 1 \leq i \leq k\}$. Pero la suma final en la ecuación (2) es una combinación lineal de elementos de S' y como S' es linealmente independiente, se tiene que $\alpha_z = 0$ si $z \in Z$. Luego la ecuación (1) se reduce a una combinación lineal de $\{y_1, \dots, y_k\}$, el cual, por hipótesis, es linealmente independiente. Por lo tanto, todos los coeficientes α_z de la ecuación (1) son iguales a cero, demostrando que S es linealmente independiente. \square

Teorema 2.2.1. (Jordan) Sea T un operador lineal en un \mathbb{k} -espacio vectorial V de dimensión n , tal que el polinomio característico de T se descompone en un producto de factores de grado 1. Entonces existe una base canónica de Jordan para T ; esto es, existe una base ordenada β para V que es unión ajena de ciclos de vectores propios generalizados de T .

Demostración

La demostración se hará por inducción sobre n .

Es claro que el resultado es verdadero para $n = 1$, ya que toda matriz de 1×1 es una forma canónica de Jordan. Supongamos que la conclusión es cierta para espacios vectoriales de dimensión menor que n y que $\dim V = n$. Entonces tenemos:

Caso 1. $\dim \text{Im } T < n$. Puesto que $\text{Im } T$ es un subespacio T -invariante de V podemos definir a $T_1 : \text{Im } T \rightarrow \text{Im } T$ como la restricción de T a $\text{Im } T$. Ya que el polinomio característico de T_1 divide al de T , la suposición del Caso 1 nos permite aplicar la hipótesis de inducción a T_1 para concluir que existe una base canónica de Jordan γ para T_1 que contiene r elementos ($r < n$). Por lo tanto, en virtud del teorema 2.1.1 c), γ es una unión ajena de ciclos de vectores propios generalizados de T_1 (y por lo tanto de T).

Sean S_1, S_2, \dots, S_k todos los ciclos en γ que corresponden al valor propio cero, y sean y_i y w_i respectivamente los vectores inicial y terminal de S_i . Como $w_i \in \gamma \subseteq \text{Im } T$, existe $x_i \in V$ tal que $T(x_i) = w_i$. Ahora definamos $Y = \{y_1, y_2, \dots, y_k\}$, $X = \{x_1, x_2, \dots, x_k\}$ y

$$\gamma_o = \bigcup_{i=1}^k S_i.$$

Supongamos que γ_o contiene p elementos ($p \leq r$). Por el teorema 2.1.1 a) cada y_i es un vector propio que corresponde al valor propio cero, por lo tanto $Y \subseteq \text{Ker } T$ y como Y es linealmente independiente por ser un subconjunto de γ , puede ser extendido a una base $Y \cup Z$ para $\text{Ker } T$. Observemos que Z debe contener $n - r - k$ elementos puesto que $\dim \text{Ker } T = n - \dim \text{Im } T$.

Sea $S'_i = S_i \cup \{x_i\}$, entonces S'_i es un ciclo de vectores propios generalizados de T que corresponden al valor propio cero y que tiene como vector inicial a y_i . Además, si $z \in Z$,

entonces $\{z\}$ es un ciclo correspondiente al valor propio cero cuya longitud es 1. Por lo tanto, el lema 2.2.1 implica que $\left(\bigcup_{i=1}^k S'_i\right) \cup Z = (\gamma_o \cup X) \cup Z$ es un conjunto linealmente independiente que contiene $p+k+(n-r-k) = n-(r-p)$ elementos, puesto que $Y \cup Z$ es un conjunto linealmente independiente de vectores iniciales para los ciclos que corresponden al valor propio cero.

Ahora, probaremos que $\beta = \gamma \cup X \cup Z$ es la base que se desea encontrar. En primera instancia, observemos que si $\gamma_o = \gamma$ (de manera que $p = r$), entonces $\beta = \gamma_o \cup X \cup Z$ es un conjunto linealmente independiente que contiene $n - (r - p) = n$ elementos. Entonces β es una base para V .

Por otra parte, si $\gamma_o \neq \gamma$, entonces $\gamma_1 = \{v \in \gamma : v \notin \gamma_o\}$ es una unión no vacía de ciclos ajenos de vectores propios generalizados que corresponden a valores propios no nulos $\lambda_2, \dots, \lambda_m$ de T . Supongamos que

$$0 = \sum_{v \in \beta} \alpha_v v = \sum_{v \in \gamma_o \cup X \cup Z} \alpha_v v + \sum_{v \in \gamma_1} \alpha_v v.$$

Entonces

$$\sum_{v \in \gamma_o \cup X \cup Z} (-\alpha_v) v = \sum_{v \in \gamma_1} \alpha_v v.$$

Pero el lado izquierdo de la igualdad anterior es un elemento de K_{λ_1} , donde $\lambda_1 = 0$ y el lado derecho es un elemento de $K_{\lambda_2} + \dots + K_{\lambda_m}$. Luego, por el teorema 2.1.3, ambos lados de la igualdad anterior son iguales a cero. Así, como $\gamma_o \cup X \cup Z$ y γ_1 son conjuntos linealmente independientes, se tiene que $\alpha_v = 0 \quad \forall v \in \beta$. Por lo tanto, β es linealmente independiente, y como $\beta = \gamma \cup X \cup Z$ contiene $r+k+(n-r-k) = n$ elementos, β es una base para V . Finalmente, es claro que β es una unión ajena de ciclos de vectores propios generalizados de T , en virtud del teorema 2.1.1 c), β es una base canónica de Jordan para T .

Caso 2. $\dim \text{Im } T = n$. Como el polinomio característico de T se descompone como un producto de factores de grado 1, T tiene un valor propio λ . Utilizando el caso 1 para el operador no invertible $T - \lambda I$ en V , se obtiene una base ordenada β para V , tal que $[T - \lambda I]_{\beta} = J$ es una forma canónica de Jordan para $T - \lambda I$. Por lo tanto, $[T]_{\beta} = J + \lambda I_n$ es una forma canónica de Jordan para T . \square

Teorema 2.2.2. Sea T un operador lineal en un \mathbb{k} -espacio vectorial V de dimensión n tal que el polinomio característico de T se descompone en un producto de factores de grado 1. Supongamos que $\lambda_1, \lambda_2, \dots, \lambda_k$ son los distintos valores propios de T y que la multiplicidad de λ_i es m_i ($1 \leq i \leq k$) y sea β una base canónica de Jordan para T . Definamos $\beta_i = \beta \cap K_{\lambda_i}$ ($1 \leq i \leq k$). Entonces

a) $V = K_{\lambda_1} \oplus K_{\lambda_2} \oplus \dots \oplus K_{\lambda_k}$.

b) β_i es una base para K_{λ_i} .

c) K_{λ_i} ($1 \leq i \leq k$) es un subespacio T -invariante de V .

d) $\dim K_{\lambda_i} = m_i$, para $1 \leq i \leq k$.

e) $K_{\lambda_i} = \text{Ker}(T - \lambda_i I)^{m_i}$, para $1 \leq i \leq k$.

f) T es diagonalizable si y sólo si $E_{\lambda_i} = K_{\lambda_i}$ para todo $i = 1, \dots, k$.

Demostración

a) y b) Para $1 \leq i \leq k$, sea W_i el subespacio generado por β_i . Entonces $W_i \subseteq K_{\lambda_i}$ y por lo tanto $\dim W_i \leq \dim K_{\lambda_i}$. Por el teorema 2.1.3, $K_{\lambda_1}, \dots, K_{\lambda_k}$ son independientes y consecuentemente, β es la unión ajena de β_1, \dots, β_k , por lo que $V = W_1 \oplus \dots \oplus W_k \subseteq K_{\lambda_1} \oplus \dots \oplus K_{\lambda_k} \subseteq V$. Luego $V = K_{\lambda_1} \oplus \dots \oplus K_{\lambda_k}$. Además $W_i = K_{\lambda_i}$ para todo $i = 1, \dots, k$ pues de lo contrario se tendría que $\dim W_j < \dim K_{\lambda_j}$ para algún $j \in \{1, \dots, k\}$ y consecuentemente, $\dim V = \sum_{i=1}^k \dim W_i < \sum_{i=1}^k \dim K_{\lambda_i} = \dim V$, lo cual es una contradicción.

c) Por el teorema 2.1.2, K_{λ_i} es subespacio de V , y por el inciso b), β_i es una base para K_{λ_i} formada por ciclos de vectores propios generalizados que corresponden a λ_i . Pero la imagen bajo T de cualquier vector en un ciclo es una combinación lineal de vectores en dicho ciclo y en consecuencia es un elemento de K_{λ_i} . Así, $T(\beta_i) \subseteq K_{\lambda_i}$ y por lo tanto, K_{λ_i} es T -invariante.

d) Definamos a T_i ($1 \leq i \leq k$) como la restricción de T a K_{λ_i} . Entonces, de acuerdo con el inciso b), $A_i = [T_i]_{\beta_i}$ es una forma canónica de Jordan para T_i y $[T]_{\beta} = A_1 \oplus A_2 \oplus \dots \oplus A_k$.

Si $n_i = \dim K_{\lambda_i}$, entonces el polinomio característico de T_i es $\det(A_i - tI_{n_i}) = (\lambda_i - t)^{n_i}$ puesto que $A_i - tI_{n_i}$ es una matriz triangular superior que tiene a $\lambda_i - t$ en cada posición diagonal. Si $f(t)$ es el polinomio característico de T , entonces

$$\begin{aligned} f(t) &= \det(A_1 - tI_{n_1}) \det(A_2 - tI_{n_2}) \cdots \det(A_k - tI_{n_k}) \\ &= (\lambda_1 - t)^{n_1} (\lambda_2 - t)^{n_2} \cdots (\lambda_k - t)^{n_k}, \end{aligned}$$

por lo que la multiplicidad de λ_i es n_i ; esto es, $m_i = n_i = \dim K_{\lambda_i}$.

e) Es claro que $\text{Ker}(T - \lambda_i I)^{m_i} \subseteq K_{\lambda_i}$. Sea $x \in K_{\lambda_i}$, entonces el ciclo S con un vector terminal x es un subconjunto de K_{λ_i} linealmente independiente según el teorema 2.1.1 b). Como $\dim K_{\lambda_i} = m_i$, se tiene que la longitud de S no puede exceder a m_i , es decir, $(T - \lambda_i I)^p(x) = 0$ para algún entero positivo $p \leq m_i$. Por lo tanto, $x \in \text{Ker}(T - \lambda_i I)^{m_i}$, de donde $K_{\lambda_i} \subseteq \text{Ker}(T - \lambda_i I)^{m_i}$.

f) En efecto, si $E_{\lambda_i} = K_{\lambda_i}$ para $1 \leq i \leq k$, entonces, por el inciso a)

$$E_{\lambda_1} \oplus E_{\lambda_2} \oplus \dots \oplus E_{\lambda_k} = K_{\lambda_1} \oplus K_{\lambda_2} \oplus \dots \oplus K_{\lambda_k} = V$$

y así, por el teorema 1.2.3 b), T es diagonalizable.

Inversamente, si T es diagonalizable entonces $\dim E_{\lambda_i} = m_i$ por el teorema 1.2.3 d). Pero como E_{λ_i} es subespacio de K_{λ_i} y $\dim K_{\lambda_i} = m_i$ por el inciso d), se tiene que $E_{\lambda_i} = K_{\lambda_i}$ para $1 \leq i \leq k$. \square

Definición 2.2.1. Sea $A \in M_n(\mathbb{k})$ tal que el polinomio característico de A se descompone en un producto de factores de grado 1. Entonces la *forma canónica de Jordan* de A se define como la forma canónica de Jordan del operador lineal L_A de \mathbb{k}^n en \mathbb{k}^n dado por la multiplicación por A .

Ejemplo 2.2.1.

Sea D el operador derivación sobre el espacio vectorial V generado por e^x, xe^x, x^2e^x y e^{2x} . La matriz que representa al operador D en esta base es

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

cuyo polinomio característico es $f(x) = (x - 1)^3(x - 2)$. Luego los valores propios de A y D son $\lambda_1 = 1$ y $\lambda_2 = 2$. Sabemos que $V = K_{\lambda_1} \oplus K_{\lambda_2}$, donde $K_{\lambda_1} = Ker(A - I)^3$ y $K_{\lambda_2} = Ker(A - 2I) = E_{\lambda_2}$, tienen dimensiones 3 y 1 respectivamente. Ya que

$$A - I = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

es de rango 3, $E_{\lambda_1} = Ker(A - I)$ es de dimensión 1, por lo que hay sólo un ciclo (de longitud 3) de vectores propios generalizados correspondientes a λ_1 . Un simple cálculo muestra que $K_{\lambda_1} = \langle e^x, xe^x, x^2e^x \rangle$ y $K_{\lambda_2} = \langle e^{2x} \rangle$, así que $\beta_1 = \{2e^x, 2xe^x, x^2e^x\}$ es el ciclo de vectores propios generalizados correspondientes a λ_1 que es base de K_{λ_1} . Si $\beta_2 = \{e^{2x}\}$ entonces $\beta = \beta_1 \cup \beta_2$ es una base de Jordan para D y la forma canónica de Jordan de D es

$$J = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

Observación:

Puede haber muchas bases canónicas de Jordan para el mismo operador T . Sin embargo, sujetos a la convención de que podemos ordenar los ciclos en orden de longitud decreciente, la forma canónica de Jordan de un operador lineal es única hasta el ordenamiento de sus valores propios.

Terminamos este capítulo con el siguiente lema que nos será de mucha utilidad en los siguientes capítulos.

Lema 2.2.2 Sea $J \in M_p(\mathbb{C})$ el siguiente bloque de Jordan

$$J = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 & \vdots \\ 0 & 0 & \dots & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix} = \lambda I + E.$$

Entonces

a) $E^p = O$ y para $1 \leq r \leq p$,

$$E_{ij}^r = \begin{cases} 1 & \text{si } j = i + r \\ 0 & \text{en otro caso} \end{cases}$$

b) Para cualquier entero $m \geq p$

$$J^m = \begin{pmatrix} \lambda^m & C_m^1 \lambda^{m-1} & C_m^2 \lambda^{m-2} & \dots & \dots & C_m^{m-(p-1)} \lambda^{m-(p-1)} \\ 0 & \lambda^m & C_m^1 \lambda^{m-1} & \dots & \dots & C_m^{m-(p-2)} \lambda^{m-(p-2)} \\ 0 & 0 & \lambda^m & C_m^1 \lambda^{m-1} & \dots & C_m^{m-(p-3)} \lambda^{m-(p-3)} \\ \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \lambda^m & C_m^1 \lambda^{m-1} \\ 0 & 0 & 0 & 0 & 0 & \lambda^m \end{pmatrix}$$

donde $C_m^r = \frac{m(m-1)\dots(m-r+1)}{r!}$.

Demostración

Un cálculo directo (o bien inducción sobre r) prueba a).

Por la expansión binomial y a) se tiene, para $m \geq p$, que

$$\begin{aligned} J^m &= (\lambda I + E)^m = \lambda^m I + C_m^1 \lambda^{m-1} E + C_m^2 \lambda^{m-2} E^2 + \dots + C_m^{p-1} \lambda^{m-p+1} E^{p-1} = \\ &= \begin{pmatrix} \lambda^m & C_m^1 \lambda^{m-1} & C_m^2 \lambda^{m-2} & \dots & \dots & C_m^{m-(p-1)} \lambda^{m-(p-1)} \\ 0 & \lambda^m & C_m^1 \lambda^{m-1} & \dots & \dots & C_m^{m-(p-2)} \lambda^{m-(p-2)} \\ 0 & 0 & \lambda^m & C_m^1 \lambda^{m-1} & \dots & C_m^{m-(p-3)} \lambda^{m-(p-3)} \\ \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \lambda^m & C_m^1 \lambda^{m-1} \\ 0 & 0 & 0 & 0 & 0 & \lambda^m \end{pmatrix}. \quad \square \end{aligned}$$

3 El teorema de Perron-Frobenius.

En 1907 el matemático alemán Oskar Perron demostró el teorema para el caso de matrices cuyas entradas son todas positivas. Más tarde el teorema fue generalizado por el matemático alemán Georg Frobenius a cierta clase de matrices con entradas no negativas. La versión que demostraremos aquí es para matrices irreducibles con entradas no negativas.

3.1 Normas Vectoriales.

En este capítulo y el siguiente trabajaremos con el espacio $M_{n \times 1}(\mathbb{C})$ al que denotaremos por $V_n(\mathbb{C})$.

Definición 3.1.1. Una función $\|\cdot\| : V_n(\mathbb{C}) \rightarrow \mathbb{R}$ es una *norma vectorial* si para todo $x, y \in V_n(\mathbb{C})$ y para todo $\alpha \in \mathbb{C}$ se satisfacen:

i) $\|x\| \geq 0$; $\|x\| = 0$ si y sólo si $x = 0$

ii) $\|\alpha x\| = |\alpha| \|x\|$

iii) $\|x + y\| \leq \|x\| + \|y\|$.

Para la norma de la diferencia de 2 vectores se cumple una desigualdad importante que se obtiene a partir de la definición 3.1.1 iii), a saber,

$$\|x - y\| \geq \left| \|x\| - \|y\| \right| \quad \forall x, y \in V_n(\mathbb{C}).$$

Ejemplo 3.1.1.

a) La norma ℓ_1 está dada por

$$\|x\|_1 = \sum_{i=1}^n |x_i|$$

b) La norma ℓ_2 (o norma euclidiana) está dada por

$$\|x\|_2 = \left(\sum_{i=1}^n |x_i|^2 \right)^{\frac{1}{2}}$$

c) La norma ℓ_∞ está dada por

$$\|x\|_\infty = \max \{|x_1|, \dots, |x_n|\}$$

d) La norma ℓ_p está dada por

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

Definición 3.1.2. Sea $\{x^{(k)}\}_{k=1}^\infty$ una sucesión infinita de vectores en $V_n(\mathbb{C})$, decimos que esta sucesión *converge* al vector $x \in V_n(\mathbb{C})$ si

$$\lim_{m \rightarrow \infty} x_j^{(m)} = x_j \quad \text{para } 1 \leq j \leq n,$$

donde $x_j^{(m)}$ y x_j son la j -ésima entrada del vector $x^{(m)}$ y el vector x respectivamente.

Analógamente, por la convergencia de una serie infinita $\sum_{m=0}^{\infty} y^{(m)}$ de vectores en $V_n(\mathbb{C})$

al vector $y \in V_n(\mathbb{C})$, entenderemos que

$$\lim_{N \rightarrow \infty} \sum_{m=0}^N y_j^{(m)} = y_j \quad \text{para } 1 \leq j \leq n.$$

Se sigue de la definición 3.1.2 que la condición necesaria y suficiente para la convergencia de la sucesión de vectores $\{x^{(k)}\}$ al vector x es la convergencia de la sucesión de escalares $\{x_i^{(k)}\}$ al escalar x_i para $i = 1, \dots, n$. Ya que

$$\|x^{(k)} - x\|_{\infty} = \max_{1 \leq i \leq n} |x_i^{(k)} - x_i|$$

esta condición es equivalente a la convergencia a cero de la sucesión de normas $\|x^{(k)} - x\|_{\infty}$.

Podemos afirmar aún mas:

Teorema 3.1.1. Sea $\|\cdot\|$ una norma en $V_n(\mathbb{C})$. Si $\{x^{(k)}\}_{k=1}^{\infty}$ es una sucesión de vectores en $V_n(\mathbb{C})$ y $x \in V_n(\mathbb{C})$, entonces la sucesión $\{x^{(k)}\}$ converge a x si y sólo si la sucesión de normas $\|x^{(k)} - x\|$ converge a 0.

La prueba de este teorema se sigue de los siguientes resultados:

Recordemos que una sucesión de funciones $f : V_n(\mathbb{C}) \rightarrow \mathbb{R}$ es *uniformemente continua* en $A \subset V_n(\mathbb{C})$ si para todo $\varepsilon > 0$ existe $\delta(\varepsilon) > 0$ tal que si $x, y \in A$ y $\|x - y\|_2 < \delta$ entonces $|f(x) - f(y)| \leq \varepsilon$.

Lema 3.1.1. Cualquier norma $\|\cdot\|$ en $V_n(\mathbb{C})$ es uniformemente continua.

Demostración

Sean $\beta = \{e_1, e_2, \dots, e_n\}$ la base canónica, $x = \sum_{i=1}^n x_i e_i$ y $y = \sum_{i=1}^n y_i e_i$, donde $x_i, y_i \in \mathbb{C}$ para $1 \leq i \leq n$. Entonces

$$\| \|x\| - \|y\| \| \leq \|x - y\| = \left\| \sum_{i=1}^n (x_i - y_i) e_i \right\| \leq \sum_{i=1}^n |x_i - y_i| \|e_i\| \leq n \max_{1 \leq i \leq n} |x_i - y_i|.$$

Luego

$$\| \|x\| - \|y\| \| \leq n \max_{1 \leq i \leq n} |x_i - y_i| \leq n \|(x_1, x_2, \dots, x_n) - (y_1, y_2, \dots, y_n)\|_2.$$

Entonces, para que $\| \|x\| - \|y\| \| < \varepsilon$ basta con elegir $\delta = \frac{\varepsilon}{n}$. \square

Teorema 3.1.2. Sea $\beta = \{v_1, v_2, \dots, v_n\}$ una base ordenada de $V_n(\mathbb{C})$. Si $f_1, f_2 : V_n(\mathbb{C}) \rightarrow \mathbb{R}$ son funciones que satisfacen:

a) $f_i(v) \geq 0$ para todo $v \in V$; $f_i(v) = 0$ si y sólo si $v = 0$,

b) $f_i(\alpha v) = |\alpha| f_i(v)$ para todo $\alpha \in \mathbb{R}$ y todo $v \in V_n(\mathbb{C})$,

c) $f_i(v(x))$ es continua en \mathbb{R}^n , donde $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ y $v(x) = \sum_{k=1}^n x_k v_k$,

entonces existen $C_m, C_M \in \mathbb{R}^+$ tales que para cualquier $v \in V_n(\mathbb{C})$,

$$C_m f_1(v) \leq f_2(v) \leq C_M f_1(v).$$

Demostración

Sea $S = \{x \in \mathbb{R}^n : \|x\|_2 = 1\}$ la esfera unitaria y sea $h : S \rightarrow \mathbb{R}$ la función dada por

$$h(x) = \frac{f_2(v(x))}{f_1(v(x))},$$

donde $v(x) = x_1 v_1 + \cdots + x_n v_n$. Ya que $v(x) = 0$ si y sólo si $x = 0$, por a), h está bien definida en S ; de hecho, $h(x) > 0$ para todo $x \in S$. Por c), h es continua en S que es un compacto en \mathbb{R}^n . Luego, por el teorema del valor máximo y mínimo ([8] pag. 65), h alcanza su máximo C_M y su mínimo C_m en S y ambos son positivos. De aquí, para todo $x \in S$,

$$C_m f_1(v(x)) \leq f_2(v(x)) \leq C_M f_1(v(x)).$$

Probaremos que las desigualdades anteriores valen para cualquier $x \in \mathbb{R}^n$. Por a) valen para $x = 0$. Si $x \neq 0$, entonces $\frac{x}{\|x\|_2} \in S$ y $v(\frac{x}{\|x\|_2}) = \frac{1}{\|x\|_2} v(x)$; luego, por b), valen también para $x \neq 0$. Finalmente, como β es base, cualquier vector $v \in V_n(\mathbb{C})$ es de la forma $v = v(x)$ para algún $x \in \mathbb{R}^n$. Por lo tanto, para cualquier $v \in V_n(\mathbb{C})$,

$$C_m f_1(v) \leq f_2(v) \leq C_M f_1(v). \quad \square$$

Corolario 3.1.1. Sean $\|\cdot\|_\alpha$ y $\|\cdot\|_\beta$ normas en $V_n(\mathbb{C})$. Entonces existen constantes positivas C_m y C_M tales que $\forall x \in V_n(\mathbb{C})$

$$C_m \|x\|_\alpha \leq \|x\|_\beta \leq C_M \|x\|_\alpha.$$

Corolario 3.1.2. Sean $\|\cdot\|_\alpha$ y $\|\cdot\|_\beta$ normas en $V_n(\mathbb{C})$. Si $\{x^{(k)}\}_{k=1}^\infty$ es una sucesión de vectores en $V_n(\mathbb{C})$ entonces $\|x^{(k)}\|_\alpha$ converge a 0 si y sólo si $\|x^{(k)}\|_\beta$ converge a 0.

Demostración

Por el corolario 3.1.1 existen $C_m, C_M \in \mathbb{R}^+$ tales que para todo $k \in \mathbb{N}$

$$C_m \|x^{(k)}\|_\alpha \leq \|x^{(k)}\|_\beta \leq C_M \|x^{(k)}\|_\alpha.$$

Por lo tanto, $\|x^{(k)}\|_\alpha$ converge a 0 si y sólo si $\|x^{(k)}\|_\beta$ converge a 0. \square

3.2 Normas matriciales.

Sea $A \in M_n(\mathbb{C})$. Ya que $M_n(\mathbb{C})$ es un espacio vectorial de dimensión n^2 , podemos tomar cualquier norma vectorial en $V_{n^2}(\mathbb{C})$ para definir la norma de A , y ésta deberá satisfacer las condiciones de la definición 3.1.1. Como el producto de 2 matrices de orden n es también una matriz de orden n , habrá que añadir una cuarta condición para la norma de matrices que considere este hecho.

Es posible seleccionar un número ilimitado de formas distintas para definir la norma de la matriz A y que ésta cumpla las condiciones anteriores. Sin embargo, en la mayoría de los casos en que se utiliza la teoría de las aproximaciones para analizar un problema, se considerarán al mismo tiempo matrices y vectores, como ocurre con la ecuación matricial $Ax = b$. Por lo tanto, es conveniente que al definir las normas de matrices éstas estén relacionadas de manera lógica con las normas de vectores.

Dados una norma vectorial $\|\cdot\|$ en $V_n(\mathbb{C})$, $0 \neq x \in V_n(\mathbb{C})$ y $A \in M_n(\mathbb{C})$, la proporción $\frac{\|Ax\|}{\|x\|}$ muestra qué tanto la multiplicación por A magnifica la norma del vector x . Podemos medir la ‘magnitud’ de A como el ‘alargamiento’ relativo. –

Definición 3.2.1. Sean $A \in M_n(\mathbb{C})$ y $\|\cdot\|$ una norma vectorial en $V_n(\mathbb{C})$ entonces

$$\|A\| := \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$$

es llamada la norma de la matriz A *inducida* por $\|\cdot\|$.

Para que esta definición tenga sentido, debemos mostrar que el máximo existe. Observemos que el valor $\frac{\|Ax\|}{\|x\|}$ no varía si x es reemplazado por αx , donde α es un escalar. Así podemos escribir

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

Pero el conjunto de todos los vectores en $V_n(\mathbb{C})$ con $\|x\| = 1$ es compacto y $\|Ax\|$ es una función continua de x definida en este conjunto, entonces por el teorema del valor máximo y mínimo ([8] pag 65) existe un vector $y \in V_n(\mathbb{C})$ con $\|y\| = 1$ tal que

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

Las principales propiedades de la norma inducida están contenidas en el siguiente teorema.

Teorema 3.2.1. Si $A, B \in M_n(\mathbb{C})$, entonces se satisfacen:

- a) $\|A\| \geq 0$; $\|A\| = 0$ si y sólo si $A = O$.
- b) $\|\alpha A\| = |\alpha| \|A\|$ para todo escalar $\alpha \in \mathbb{C}$.
- c) $\|A + B\| \leq \|A\| + \|B\|$.
- d) $\|AB\| \leq \|A\| \|B\|$.

Demostración

a) Por el párrafo anterior, para cada $A \in M_n(\mathbb{C})$ podemos hallar $u \in V_n(\mathbb{C})$ unitario tal que $\|A\| = \|Au\|$. Además, de la definición 3.2.1 se deduce que para cualquier vector $x \in V_n(\mathbb{C})$ se cumple la condición $\|Ax\| \leq \|A\| \|x\|$. De estas dos observaciones se sigue a).

b) Sea u el vector unitario tal que $\|\alpha A\| = \|(\alpha A)u\|$. Entonces

$$\|\alpha A\| = \|(\alpha A)u\| = \|\alpha(Au)\| = |\alpha| \|Au\| \leq |\alpha| \|A\| \|u\| = |\alpha| \|A\|.$$

Sea ahora v el vector unitario tal que $\|A\| = \|Av\|$. Entonces

$$|\alpha| \|A\| = |\alpha| \|Av\| = \|\alpha(Av)\| = \|(\alpha A)v\| \leq \|\alpha A\| \|v\| = \|\alpha A\|.$$

De estas desigualdades se sigue b).

c) Sea u el vector unitario tal que $\|A + B\| = \|(A + B)u\|$. Entonces

$$\|A + B\| = \|(A + B)u\| = \|Au + Bu\| \leq \|Au\| + \|Bu\| \leq \|A\| \|u\| + \|B\| \|u\|.$$

Ya que $\|Au\| \leq \|A\| \|u\|$, $\|Bu\| \leq \|B\| \|u\|$ y $\|u\| = 1$

$$\|A + B\| \leq \|A\| + \|B\|.$$

d) Finalmente, sea u el vector unitario tal que $\|AB\| = \|(AB)u\|$. Entonces

$$\|AB\| = \|(AB)u\| = \|A(Bu)\| \leq \|A\| \|Bu\| \leq \|A\| \|B\| \|u\| = \|A\| \|B\|. \quad \square$$

Una función $\|\cdot\| : M_n(\mathbb{C}) \rightarrow \mathbb{R}$ que satisface las propiedades a), b), c) y d) del teorema anterior se llama *norma matricial*. Luego, las normas inducidas por normas vectoriales en $V_n(\mathbb{C})$ son normas matriciales. Claramente, toda norma matricial es una norma vectorial.

Concluimos esta sección con la descripción de las normas inducidas por $\|\cdot\|_\infty$ y $\|\cdot\|_1$ en $V_n(\mathbb{C})$.

Teorema 3.2.2. Sea $A \in M_n(\mathbb{C})$. Entonces $\|A\|_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$.

Demostración

Sea $0 \neq x \in V_n(\mathbb{C})$. Consideremos la componente k del vector Ax , que se obtiene multiplicando el renglón k de la matriz A por el vector x . Entonces de la definición de $\|\cdot\|_\infty$ se sigue que

$$|(Ax)_k| = \left| \sum_{j=1}^n a_{kj} x_j \right| \leq \sum_{j=1}^n |a_{kj}| |x_j| \leq \|x\|_\infty \sum_{j=1}^n |a_{kj}|.$$

Pero $\sum_{j=1}^n |a_{kj}| \leq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$, por lo que

$$|(Ax)_k| \leq \|x\|_\infty \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

La desigualdad anterior se cumple para cualquier componente k del vector Ax , por lo tanto se cumplirá para la componente que posee el máximo valor, así que

$$\|Ax\|_\infty = \max_k |(Ax)_k| \leq \|x\|_\infty \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|.$$

De la definición 3.2.1 tenemos que

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

Supongamos ahora que el máximo en la desigualdad anterior se obtiene para el renglón k , es decir,

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

y consideremos el vector y con componentes

$$y_j = \begin{cases} \frac{|a_{kj}|}{a_{kj}} & \text{si } a_{kj} \neq 0 \\ 1 & \text{si } a_{kj} = 0 \end{cases}$$

La componente k del vector Ay será entonces

$$(Ay)_k = \sum_{j=1}^n a_{kj} y_j = \sum_{j=1}^n |a_{kj}| = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$$

De la definición de $\|\cdot\|_\infty$ se tiene que

$$\|Ay\|_\infty \geq |(Ay)_k| = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|,$$

y como $\|y\|_\infty = 1$ se obtiene

$$\|A\|_\infty = \max_{\|x\|=1} \|Ax\|_\infty \geq \|Ay\|_\infty \geq \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|,$$

con lo que queda demostrado el teorema. \square

Teorema 3.2.3. Sea $A \in M_n(\mathbb{C})$. Entonces $\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|$.

Demostración

Sea $0 \neq x \in V_n(\mathbb{C})$. Entonces

$$\|Ax\|_1 = \sum_{i=1}^n |(Ax)_i| = \sum_{i=1}^n \left| \sum_{j=1}^n a_{ij}x_j \right| \leq \sum_{i=1}^n \sum_{j=1}^n |a_{ij}| |x_j| = \sum_{j=1}^n |x_j| \left(\sum_{i=1}^n |a_{ij}| \right).$$

Pero $\max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| \geq \sum_{i=1}^n |a_{ij}|$, por lo que

$$\|Ax\|_1 \leq \sum_{j=1}^n |x_j| \left(\max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| \right) = \|x\|_1 \left(\max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| \right).$$

De la definición 3.2.1 se sigue que

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

Supongamos que el máximo en la desigualdad anterior se obtiene para la columna k , es decir,

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

y consideremos el vector y con componentes

$$y_j = \begin{cases} 0 & \text{si } j \neq k \\ 1 & \text{si } j = k \end{cases}$$

Es evidente que $\|y\|_1 = 1$ y que $Ay = (a_{1k}, a_{2k}, \dots, a_{nk})$. Entonces

$$\|A\|_1 = \max_{\|x\|=1} \|Ax\|_1 \geq \|Ay\|_1 = \sum_{i=1}^n |a_{ik}| = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}|.$$

Por lo tanto, queda demostrado el teorema. \square

Dejamos la descripción de la norma inducida en $M_n(\mathbb{C})$ por la norma euclidiana para la siguiente sección.

3.3 El radio espectral y la norma espectral.

Definición 3.3.1. Sea $A \in M_n(\mathbb{C})$, con valores propios λ_i , $1 \leq i \leq n$. Entonces

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|$$

es llamado el *radio espectral* de la matriz A .

Ejemplo 3.3.1. Sea $A \in M_3(\mathbb{R})$ tal que $A = \begin{pmatrix} 1 & 2 & -4 \\ 2 & -2 & -2 \\ -4 & -2 & 1 \end{pmatrix}$ que tiene como

polinomio característico $\det(A - \lambda I) = -\lambda^3 + 27\lambda + 54 = -(\lambda + 3)^2(\lambda - 6)$. Luego $\rho(A) = 6$

Lema 3.3.1. Si $\lambda_1, \lambda_2, \dots, \lambda_k$ son los valores propios distintos de $A \in M_n(\mathbb{C})$ entonces para cualquier $m \in \mathbb{N}$, $\lambda_1^m, \lambda_2^m, \dots, \lambda_k^m$ son todos los valores propios (no necesariamente distintos) de A^m .

Demostración

Sean $\lambda_1, \lambda_2, \dots, \lambda_k$ los valores propios distintos de A con multiplicidades n_1, n_2, \dots, n_k respectivamente y sea J una forma canónica de Jordan de A . Entonces cada λ_i aparece repetido n_i veces en la diagonal principal de J . Sea $Q \in M_n(\mathbb{C})$ la matriz invertible tal que $A = Q^{-1}JQ$. Entonces para cualquier $m \in \mathbb{N}$, $A^m = Q^{-1}J^mQ$ y, por el lema 2.2.2, cada λ_i^m aparece repetido n_i veces en la diagonal principal de J^m . Luego $\lambda_1^m, \lambda_2^m, \dots, \lambda_k^m$ son todos los valores propios de A^m . \square

Corolario 3.3.1. Sean $A \in M_n(\mathbb{C})$ y $m \in \mathbb{N}$. Entonces $\rho(A^m) = \rho(A)^m$.

Demostración

Sean $\lambda_1, \lambda_2, \dots, \lambda_k$ los valores propios distintos de A y sea $j \in \{1, \dots, k\}$ tal que $|\lambda_j| = \rho(A)$. Luego para todo $i \neq j$ con $1 \leq i \leq k$, $|\lambda_i| \leq \rho(A)$. Por lo tanto $|\lambda_j^m| = |\lambda_j|^m = \rho(A)^m$ y para todo $i \neq j$ con $1 \leq i \leq k$, $|\lambda_i^m| = |\lambda_i|^m \leq \rho(A)^m$. Por el lema anterior, $\rho(A^m) = \rho(A)^m$. \square

Definición 3.3.2. La norma inducida en $M_n(\mathbb{C})$ por la norma euclidiana es llamada *norma espectral*.

Teorema 3.3.1. Sea $A \in M_n(\mathbb{C})$, entonces $\|A\|_2 = (\rho(A^*A))^{\frac{1}{2}}$.

Demostración

La matriz A^*A es Hermitiana y sus valores propios son reales. Probaremos ahora que son no negativos. Sean λ un valor propio de A^*A y x un vector propio asociado a λ . Entonces

$$\|Ax\|_2^2 = \langle Ax, Ax \rangle = \langle x, A^*Ax \rangle = \langle x, \lambda x \rangle = \lambda \langle x, x \rangle = \lambda \|x\|_2^2,$$

de donde $\lambda \geq 0$. Por el teorema espectral existe una base ortonormal $\{x_1, \dots, x_n\}$ de $V_n(\mathbb{C})$ formada por vectores propios de A^*A . Supongamos que $A^*Ax_i = \lambda_i x_i$ para $1 \leq i \leq n$, donde $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0$; es decir que $\lambda_1 = \rho(A^*A)$.

Si $x = \sum_{i=1}^n c_i x_i$ es un vector no nulo, entonces

$$\begin{aligned} \|Ax\|_2^2 &= \langle Ax, Ax \rangle = \langle x, A^*Ax \rangle = \left\langle \sum_{i=1}^n c_i x_i, \sum_{j=1}^n c_j \lambda_j x_j \right\rangle = \\ &= \sum_{i=1}^n |c_i|^2 \lambda_i \leq \lambda_1 \sum_{i=1}^n |c_i|^2 = \lambda_1 \|x\|_2^2, \end{aligned}$$

de donde $\|Ax\|_2 \leq \sqrt{\lambda_1} \|x\|_2$. Observemos que para $x = x_1$ se da la igualdad, por lo que

$$\|Ax\|_2 = \sqrt{\lambda_1} = [\rho(A^*A)]^{\frac{1}{2}}. \quad \square$$

Corolario 3.3.2. Sea $A \in M_n(\mathbb{C})$ una matriz Hermitiana, entonces $\|A\|_2 = \rho(A)$.

Demostración

Si A es Hermitiana, entonces

$$\|A\|_2^2 = \rho(A^*A) = \rho(A^2) = [\rho(A)]^2,$$

de donde $\|A\| = \rho(A)$. \square

A continuación probaremos que las normas inducidas proporcionan cotas superiores para el espectro de una matriz. De hecho no es difícil ver que lo mismo ocurre para cualquier norma matricial.

Teorema 3.3.2. Sea $\|\cdot\|$ cualquier norma inducida en $M_n(\mathbb{C})$. Si $A \in M_n(\mathbb{C})$ entonces $\|A\| \geq \rho(A)$.

Demostración

Si λ es cualquier valor propio de A y x es un vector propio asociado a λ , entonces

$$|\lambda| \|x\| = \|\lambda x\| = \|Ax\| \leq \|A\| \|x\|,$$

de donde se concluye que $\|A\| \geq |\lambda|$ para todo λ valor propio de A . \square

Corolario 3.3.3. Si $A \in M_n(\mathbb{C})$ y $\nu = \|A\|_\infty = \max_{1 \leq i < n} \sum_{j=1}^n |a_{ij}|$, entonces $\rho(A) \leq \nu$.

Corolario 3.3.4. Si $A \in M_n(\mathbb{C})$ y $\nu' = \|A\|_1 = \max_{1 \leq j < n} \sum_{i=1}^n |a_{ij}|$, entonces $\rho(A) \leq \nu'$.

Corolario 3.3.5. Si $A \in M_n(\mathbb{C})$ entonces $\rho(A) \leq \min \{\nu, \nu'\}$.

Ejemplo 3.3.2.

Sea $A \in M_2(\mathbb{R})$ tal que

$$A = \begin{pmatrix} 1 & 2 \\ 4 & 3 \end{pmatrix},$$

entonces $\rho(A) \leq \min \{5, 7\}$; de hecho los valores propios de A son $\lambda_1 = 5$ y $\lambda_2 = -1$.

Corolario 3.3.6 Sea $A \in M_n(\mathbb{C})$ y x_1, x_2, \dots, x_n n números reales positivos cualesquiera. Si

$$\nu = \max_{1 \leq i \leq n} \left\{ \frac{\sum_{j=1}^n |a_{ij}| x_j}{x_i} \right\}; \quad \nu' = \max_{1 \leq j \leq n} \left\{ x_j \sum_{i=1}^n \frac{|a_{ij}|}{x_i} \right\},$$

entonces $\rho(A) \leq \min \{\nu, \nu'\}$.

Demostración

Sea $D = \text{diag}(x_1, x_2, \dots, x_n)$ una matriz diagonal con entradas x_1, x_2, \dots, x_n . El resultado se obtiene aplicando los corolarios 3.3.3 y 3.3.4 a la matriz $D^{-1}AD$, cuyo radio espectral coincide necesariamente con el de A . \square

Para ilustrar este importante resultado consideremos la matriz $A = \frac{1}{4} \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$

y $x_1 = 1 = x_4$, $x_2 = 2 = x_3$, entonces $\nu = \nu' = \frac{3}{4}$, lo cual implica que $\rho(A) \leq \frac{3}{4}$. Más aún, si $x_1 = x_2 = x_3 = x_4$, entonces $\nu = \nu' = \frac{1}{2}$, que es el valor exacto de $\rho(A)$. Esto finalmente muestra que la igualdad es posible en los corolarios 3.3.3, 3.3.4, 3.3.5 y 3.3.6

En un intento por mejorar lo que dice el corolario 3.3.3, supongamos que las sumas de los módulos de las entradas de cada renglón de la matriz A no fueran todas iguales a ν . ¿Podríamos esperar que $\rho(A) < \nu$?

La respuesta es no, el contraejemplo está dado por la matriz

$$B = \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}.$$

Para esta matriz $\nu = 3$ y $\rho(B) = 3$.

3.4 Matrices convergentes.

Definición 3.4.1. Sea $\{A^{(m)}\}_{m=1}^{\infty}$ una sucesión infinita de matrices en $M_n(\mathbb{C})$, decimos que esta sucesión *converge* a la matriz $A \in M_n(\mathbb{C})$ si

$$\lim_{m \rightarrow \infty} a_{ij}^{(m)} = a_{ij} \quad 1 \leq i, j \leq n,$$

donde $a_{ij}^{(m)}$ y a_{ij} son la ij -ésima entrada de la matriz $A^{(m)}$ y la matriz A respectivamente.

Análogamente, por la convergencia de una serie infinita $\sum_{m=0}^{\infty} B^{(m)}$ de matrices $B^{(m)} = (b_{ij}^{(m)})$ en $M_n(\mathbb{C})$ a la matriz $B = (b_{ij}) \in M_n(\mathbb{C})$, entenderemos que

$$\lim_{N \rightarrow \infty} \sum_{m=0}^N b_{ij}^{(m)} = b_{ij} \quad 1 \leq i, j \leq n.$$

Como en $V_n(\mathbb{C})$, la sucesión de matrices $\{A^{(m)}\}_{m=1}^{\infty}$ converge a la matriz A si y sólo si para cualquier norma matricial

$$\|A^{(m)} - A\| \rightarrow 0, \text{ cuando } m \rightarrow \infty,$$

y análogamente la serie de matrices $\sum_{m=0}^{\infty} B^{(m)}$ converge a la matriz B si y sólo si

$$\left\| \sum_{m=0}^{\infty} B^{(m)} - B \right\| \rightarrow 0, \text{ cuando } m \rightarrow \infty$$

Cuando $\lim_{m \rightarrow \infty} A^{(m)} = A$ se tiene que $\lim_{m \rightarrow \infty} \|A^{(m)}\| = \|A\|$ por la continuidad de la norma $\|\cdot\|$.

Por otra parte, si $\|\cdot\|$ es una norma inducida y $x \in V_n(\mathbb{C})$ se tiene que

$$\|A^{(m)}x - Ax\| = \|(A^{(m)} - A)x\| \leq \|A^{(m)} - A\| \|x\|.$$

Por consiguiente, si $\|A^{(m)} - A\| \rightarrow 0$, cuando $m \rightarrow \infty$, también $\|A^{(m)}x - Ax\| \rightarrow 0$ cuando $m \rightarrow \infty$. Inversamente, si $\lim_{m \rightarrow \infty} A^{(m)}x = Ax$ para todo x considerando $u \in V_n(\mathbb{C})$ tal que $\|u\| = 1$; $\|A^{(m)} - A\| = \|(A^{(m)} - A)u\| \rightarrow 0$ cuando $m \rightarrow \infty$. De aquí se deduce $\lim_{m \rightarrow \infty} A^{(m)} = A$ si y sólo si $\lim_{m \rightarrow \infty} A^{(m)}x = Ax$ para todo $x \in V_n(\mathbb{C})$.

Los siguientes resultados nos serán de utilidad.

Proposición 3.4.1. Sea $\{A^{(m)}\}_{m=1}^{\infty}$ una sucesión de matrices en $M_n(\mathbb{C})$ tales que

$$\lim_{m \rightarrow \infty} A^{(m)} = L \in M_n(\mathbb{C}).$$

Entonces para cualquier $B \in M_n(\mathbb{C})$ y $D \in M_n(\mathbb{C})$,

$$\lim_{m \rightarrow \infty} BA^{(m)} = BL \text{ y } \lim_{m \rightarrow \infty} A^{(m)}D = LD.$$

Demostración

Para cualquier i ($1 \leq i \leq n$) y j ($1 \leq j \leq n$), se tiene que

$$\lim_{m \rightarrow \infty} [(BA^{(m)})_{ij}] = \lim_{m \rightarrow \infty} \left[\sum_{k=1}^n B_{ik} A_{kj}^{(m)} \right] = \sum_{k=1}^n B_{ik} \left\{ \lim_{m \rightarrow \infty} [A_{kj}^{(m)}] \right\} = \sum_{k=1}^n B_{ik} L_{kj} = (BL)_{ij}.$$

Por lo tanto, $\lim_{m \rightarrow \infty} (BA)^{(m)} = BL$.

La demostración de la segunda afirmación es análoga. \square

Corolario 3.4.1. Sea $A \in M_n(\mathbb{C})$ y sea $\lim_{m \rightarrow \infty} A^m = L$. Entonces para cualquier matriz invertible $Q \in M_n(\mathbb{C})$, $\lim_{m \rightarrow \infty} (QAQ^{-1})^m = QLQ^{-1}$.

Demostración

Puesto que

$$(QAQ^{-1})^m = (QAQ^{-1})(QAQ^{-1}) \dots (QAQ^{-1}) = QA^m Q^{-1},$$

tenemos, de acuerdo con la proposición 3.4.1, que

$$\lim_{m \rightarrow \infty} [(QAQ^{-1})^m] = \lim_{m \rightarrow \infty} (QA^m Q^{-1}) = Q \left(\lim_{m \rightarrow \infty} A^m \right) Q^{-1} = QLQ^{-1}. \quad \square$$

Definición 3.4.2. Sea $A \in M_n(\mathbb{C})$, se dice que A es *convergente a cero* si la sucesión de matrices $\{A, A^2, A^3, \dots\}$ converge a la matriz cero, y que es *divergente* en cualquier otro caso.

Teorema 3.4.1. Sea $A \in M_n(\mathbb{C})$, entonces A es convergente si y sólo si $\rho(A) < 1$.

Demostración

En efecto, sea $A \in M_n(\mathbb{C})$, entonces existe una matriz invertible Q que reduce a A a su forma canónica de Jordan, es decir,

$$QAQ^{-1} = J = \begin{pmatrix} J_1 & & 0 \\ & J_2 & \\ & & \ddots \\ 0 & & & J_r \end{pmatrix}$$

donde cada una de las matrices J_l para $1 \leq l \leq r$ tiene la forma

$$J_l = \begin{pmatrix} \lambda_l & 1 & & 0 \\ & \lambda_l & 1 & \\ & & \ddots & 1 \\ 0 & & & \lambda_l \end{pmatrix}$$

Ahora bien, como cada una de las submatrices J_l es triangular superior, también J lo es y el conjunto $\{\lambda_l\}_{l=1}^r$ incluye todos los distintos valores propios de las matrices A y J , puesto que son similares. Por cálculo directo tenemos para $m \geq 1$ que

$$J^m = \begin{pmatrix} J_1^m & & 0 \\ & J_2^m & \\ & & \ddots \\ 0 & & & J_r^m \end{pmatrix}$$

y por el lema 2.2.2 tenemos para $m \geq n_l$ que

$$J_l^m = \begin{pmatrix} \lambda^m & C_m^1 \lambda^{m-1} & C_m^2 \lambda^{m-2} & \dots & \dots & C_m^{m-(p-1)} \lambda^{m-(n_l-1)} \\ 0 & \lambda^m & C_m^1 \lambda^{m-1} & \dots & \dots & C_m^{m-(p-2)} \lambda^{m-(n_l-2)} \\ 0 & 0 & \lambda^m & C_m^1 \lambda^{m-1} & \dots & C_m^{m-(p-3)} \lambda^{m-(n_l-3)} \\ \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \lambda^m & C_m^1 \lambda^{m-1} \\ 0 & 0 & 0 & 0 & 0 & \lambda^m \end{pmatrix}$$

donde n_l es el orden del bloque de Jordan J_l

Luego, si A es convergente, entonces $A^m \rightarrow O$ cuando $m \rightarrow \infty$, y como $J^m = QA^mQ^{-1}$, por el corolario 3.4.1 se sigue que $J^m \rightarrow O$ cuando $m \rightarrow \infty$. Consecuentemente, cada $(J_l)^m \rightarrow O$ cuando $m \rightarrow \infty$. De esta manera, las entradas diagonales λ_l de J_l deben satisfacer que $|\lambda_l| < 1$ para $1 \leq l \leq r$. Claramente $\rho(A) = \rho(J) = \max_{1 \leq l \leq r} |\lambda_l| < 1$.

Inversamente, si $\rho(A) = \rho(J) < 1$, entonces $|\lambda_l| < 1$ para $1 \leq l \leq r$. Haciendo uso directo de este hecho y de que

$$\lim_{m \rightarrow \infty} (J_l)^m = 0,$$

tenemos que cada J_l es convergente, lo cual implica que J también es convergente. Por lo tanto, como $A^m = Q^{-1}J^mQ$, tenemos que A es convergente. \square

Recordemos que una serie infinita de matrices $\sum_{m=0}^{\infty} A^{(m)}$ converge a A si la sucesión de sumas parciales $S^{(1)} = A^{(1)}, S^{(2)} = A^{(1)} + A^{(2)}$, etc., converge a A .

Corolario 3.4.2. Sea $A \in M_n(\mathbb{C})$, entonces la serie $\sum_{m=0}^{\infty} A^m$ converge si y sólo si $\rho(A) < 1$. En este caso, la matriz $I - A$ es no singular y su inversa es $(I - A)^{-1} = I + A + A^2 + \dots = \sum_{m=0}^{\infty} A^m$.

Demostración

Sea $S^{(n)} = I + A + A^2 + \dots + A^n$, entonces $AS^{(n)} = A + A^2 + \dots + A^{n+1}$ de donde

$$(I - A)S^{(n)} = S^{(n)} - AS^{(n)} = I - A^{n+1}.$$

Si $\rho(A) < 1$ entonces $I - A$ es no singular. Pues si $I - A$ fuese singular entonces 1 sería valor propio de A , lo que contradice que $\rho(A) < 1$. Por otra parte tenemos que

$$(I - A)^{-1}(I - A)S^{(n)} = (I - A)^{-1}(I - A^{n+1}),$$

luego

$$S^{(n)} - (I - A)^{-1} = -(I - A)^{-1}A^{n+1},$$

así $\|S^{(n)} - (I - A)^{-1}\| \leq \|(I - A)^{-1}\| \|A^{n+1}\|$, de aquí y del hecho que $\|A^{n+1}\| \rightarrow 0$ se deduce que $S^{(n)}$ converge a $(I - A)^{-1}$ cuando $n \rightarrow \infty$.

Inversamente, si $\sum_{m=0}^{\infty} A^m$ converge entonces claramente $A^m \rightarrow 0$ y por el teorema anterior $\rho(A) < 1$. \square

3.5 Matrices irreducibles

Definición 3.5.1. Una matriz $P \in M_n(\mathbb{C})$ se llama una *matriz de permutación* si existe una permutación π del conjunto $\{1, 2, \dots, n\}$ tal que

$$p_{ij} = \begin{cases} 1 & \text{si } j = \pi(i) \\ 0 & \text{si } j \neq \pi(i) \end{cases}$$

Observemos que cada renglón y cada columna de una matriz de permutación tiene exactamente una entrada no cero. La multiplicación por la izquierda (derecha) de una matriz A por P permutará sus renglones (columnas) de acuerdo a la permutación π .

Definición 3.5.2. Una matriz $A \in M_n(\mathbb{C})$ se llama *reducible* si existe un subconjunto no vacío $S \subset N = \{1, 2, \dots, n\}$ con $S \neq N$, tal que $a_{ij} = 0$ para todos los pares de índices (i, j) donde $i \in S$ y $j \in N - S$. Una matriz es *irreducible* si no es reducible.

Si A es una matriz compleja de 1×1 , entonces A es irreducible si su única entrada es distinta de cero y reducible en otro caso.

Ejemplo 3.5.1.

Sea A una matriz de la forma

$$\begin{pmatrix} x & x & 0 & 0 & x \\ x & x & 0 & 0 & x \\ x & x & x & x & x \\ x & x & x & x & x \\ x & x & 0 & 0 & x \end{pmatrix}$$

Sea $S = \{1, 2, 5\}$ y $N - S = \{3, 4\}$, entonces $a_{ij} = 0$ siempre que $i \in S$ y $j \in N - S$. Por lo tanto, A es reducible. Notemos que si P es la matriz de permutación

$$P = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} = P^t$$

entonces

$$PA = \begin{pmatrix} x & x & x & x & x \\ x & x & x & x & x \\ x & x & 0 & 0 & x \\ x & x & 0 & 0 & x \\ x & x & 0 & 0 & x \end{pmatrix}$$

y

$$PAP^t = \begin{pmatrix} x & x & x & x & x \\ x & x & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & x & x & x \\ 0 & 0 & x & x & x \end{pmatrix}$$

Teorema 3.5.1. Una matriz $A \in M_n(\mathbb{C})$ es reducible si y sólo si existe una matriz de permutación $P \in M_n(\mathbb{C})$ tal que PAP^t tiene la forma

$$PAP^t = \begin{pmatrix} A_{11} & A_{12} \\ 0 & A_{22} \end{pmatrix}$$

Demstración

Si A es reducible, existe un subconjunto propio no vacío $S \subset N = \{1, 2, \dots, n\}$ tal que $a_{ij} = 0$ siempre que $i \in S$ y $j \in N - S$. Podemos suponer que $S = \{k_{r+1}, \dots, k_n\}$ y $N - S = \{k_1, \dots, k_r\}$.

Sea $P = [e_{k_1}, \dots, e_{k_r}, e_{k_{r+1}}, \dots, e_{k_n}]$, donde $e_{k_i} = (0, \dots, 1, \dots, 0)^t$ es el k_i -ésimo vector coordenado y donde $k_i \neq k_j$ si $i \neq j$, esto es, la permutación π de $\{1, 2, \dots, n\}$ que define a P está dada por $\pi(i) = k_i$.

Entonces $(PAP^t)_{ij} = \sum_{k=1}^n p_{ik} (AP^t)_{kj} = (AP^t)_{k_{ij}} = \sum_{l=1}^n a_{k_i k_l} p_{l j} = a_{k_i k_j}$, que es 0 si $k_i \in S$ y $k_j \in N - S$ esto es, si $r + 1 \leq i \leq n$ y $1 \leq j \leq r$. Por lo tanto,

$$PAP^t = \begin{pmatrix} A_{11} & A_{12} \\ O & A_{22} \end{pmatrix}$$

Inversamente, si $B = PAP^t = \begin{pmatrix} A_{11} & A_{12} \\ O & A_{22} \end{pmatrix}$ para alguna matriz de permutación $P \in M_n(\mathbb{C})$, podemos suponer que la permutación π de $\{1, 2, \dots, n\}$ que define a P está dada por $\pi(i) = k_i$.

Como arriba $a_{k_i k_j} = b_{ij}$ que es 0 si $r + 1 \leq i \leq n$ y $1 \leq j \leq r$, esto es, si $k_i \in S = \{k_{r+1}, \dots, k_n\}$ y $k_j \in N - S = \{k_1, \dots, k_r\}$. Luego $A = P^t B P$ es reducible. \square

Ejemplo 3.5.2.

Si

$$A = \begin{pmatrix} a_{11} & \dots & \dots & a_{1r} & \dots & \dots & a_{1n} \\ \vdots & & & & & & \\ \vdots & & & & & & \\ 0 & \dots & 0 & a_{rr} & 0 & \dots & 0 \\ \vdots & & & & & & \\ \vdots & & & & & & \\ a_{n1} & \dots & \dots & a_{nr} & \dots & \dots & a_{nn} \end{pmatrix}$$

entonces A es reducible. En este caso, tenemos que $S = \{r\}$ y $N - S = \{1, \dots, r - 1, r + 1, \dots, n\}$

En efecto, con la matriz de permutación $P = I_{rn}$ la cual permuta los renglones r -ésimo y n -ésimo, tenemos que

$$PAP^t = \begin{pmatrix} a_{11} & \dots & a_{1n} & \dots & a_{1n-1} & a_{1r} \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ a_{n1} & \dots & a_{nn} & \dots & a_{nn-1} & a_{nr} \\ \vdots & & & & & \vdots \\ \vdots & & & & & \vdots \\ 0 & \dots & 0 & \dots & 0 & a_{rr} \end{pmatrix}$$

Observemos que si A es reducible con la forma dada por el teorema 3.5.1, entonces $Ax = y$ puede ser escrita como

$$\begin{pmatrix} A_{11} & A_{12} \\ O & A_{22} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$$

o bien,

$$\begin{aligned} A_{11}x_1 + A_{12}x_2 &= y_1 \\ A_{22}x_2 &= y_2 \end{aligned}$$

Luego, si A_{22} es no-singular, podemos resolver primero para x_2 y entonces sustituir las soluciones en el primer conjunto de ecuaciones para calcular x_1 .

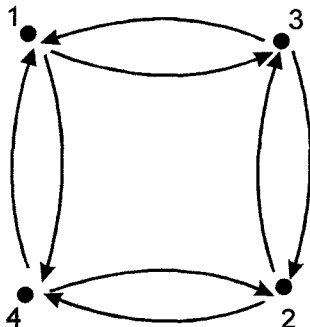
Frecuentemente es difícil determinar cuándo una matriz dada es reducible o no lo es. En este sentido, la interpretación geométrica del concepto de irreducibilidad por medio de la teoría de gráficas es bastante útil. Para esto consideremos algunas nociones elementales de la teoría de gráficas.

Sea $A \in M_n(\mathbb{C})$ y consideremos P_1, P_2, \dots, P_n n puntos arbitrarios en el plano, a los cuales llamaremos *vértices*. Para cada entrada $a_{ij} \neq 0$ de la matriz A , conectamos el vértice P_i al vértice P_j por medio de la trayectoria dirigida $\overrightarrow{P_i P_j}$ de P_i a P_j . De esta manera, a cada matriz A de $n \times n$ se le puede asociar una *gráfica dirigida finita* $G(A)$.

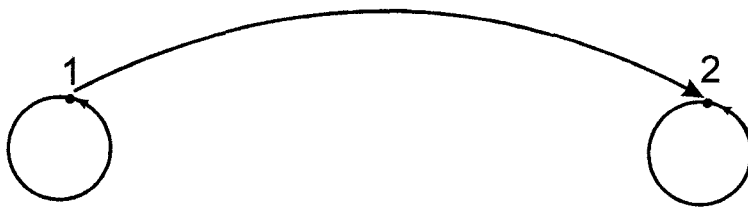
Ejemplo 3.5.3. Sea

$$A = \frac{1}{4} \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}.$$

La matriz A tiene asociada la gráfica dirigida siguiente



Analógamente la matriz $B = \begin{pmatrix} 1 & 1 \\ 0 & 3 \end{pmatrix}$ tiene asociada la gráfica dirigida



Notemos que la gráfica dirigida de una matriz A es independiente de los valores de las entradas no cero. Así, cuando hablemos de gráficas de matrices, es suficiente considerar la llamada "matriz Booleana" asociada con A , la cual tiene una entrada 1 en la posición i -ésima si y sólo si $a_{ij} \neq 0$.

Además, es fácil ver que para cualquier reordenamiento simétrico $B = PAP^t$ de A , donde P es una matriz de permutación, las gráficas dirigidas de A y B difieren sólo con respecto a la etiquetación de los puntos.

Definición 3.5.3. Una gráfica dirigida es *fuertemente conexa* si para cualquier par de vértices distintos P_i y P_j , existe una trayectoria dirigida

$$\overrightarrow{P_i P_{i_1}}, \overrightarrow{P_{i_1} P_{i_2}}, \dots, \overrightarrow{P_{i_{r-1}} P_{i_r} P_j}$$

que conecta a P_i con P_j .

Observemos que en el ejemplo anterior $G(A)$ es fuertemente conexa, mientras que para $G(B)$ no existe una trayectoria de P_2 a P_1 y por lo tanto no es fuertemente conexa.

Teorema 3.5.2. Sea $A \in M_n(\mathbb{C})$, entonces A es irreducible si y sólo si la gráfica dirigida de A es fuertemente conexa.

Demostración (Richard Varga, 1962)

En efecto, sea A una matriz irreducible y sea i_0 un índice arbitrario. Entonces existe al menos un índice j para el cual $a_{i_0 j} \neq 0$, porque de otra manera A sería reducible (como se muestra en el ejemplo 3.5.2).

Sea T el conjunto de todos los índices j tales que existe una trayectoria dirigida de P_{i_0} a P_j , y sea S el conjunto restante, esto es, $S = N - T$, donde $N = \{1, 2, \dots, n\}$. Entonces T y S son ajenos y T es no vacío. Asumamos que S es no vacío y tomemos $j \in T$ y $k \in S$. Entonces $a_{jk} = 0$, ya que de no ser así existiría una trayectoria dirigida de P_{i_0} a P_k vía P_j , esto es,

$$\overrightarrow{P_{i_0} P_{i_1}}, \overrightarrow{P_{i_1} P_{i_2}}, \dots, \overrightarrow{P_{i_{r-1}} P_j}, \overrightarrow{P_j P_k}$$

lo cual significa que $k \in T$. Ahora, notemos que $a_{jk} = 0$ para cualquier $j \in T$, $k \in S$ y $S \cap T = \emptyset$, implica (por definición) que A es reducible. Esta contradicción implica que S debe ser vacío; de aquí, la gráfica de la matriz es fuertemente conexa.

Inversamente, sea $A \in M_n(\mathbb{C})$ y supongamos que su gráfica asociada es fuertemente conexa. Si A es reducible, entonces existen conjuntos S y T ajenos no vacíos tales que $a_{jk} = 0$ siempre que $j \in S$ y $k \in T$.

Dado que la gráfica es fuertemente conexa, existe una trayectoria dirigida

$$\overrightarrow{P_{i_o} P_{i_1}}, \dots, \overrightarrow{P_{i_{r-1}} P_{i_r}},$$

donde $i_o = j \in S$, $i_r = k \in T$.

En la sucesión i_1, \dots, i_{r-1} hay un índice final i_s tal que $i_{s-1} \in S$ e $i_s \in T$ (posiblemente $i_s = k$). Pero se tiene que $a_{i_{s-1}i_s} \neq 0$, lo que contradice la definición de S y T . Por lo tanto A es irreducible. \square

Es claro, por el teorema anterior, que todas las entradas fuera de la diagonal de cualquier renglón o columna no pueden ser cero si la matriz es irreducible.

Un resultado clásico es el siguiente

Teorema 3.5.3. (Gerschgorin) Sea $A \in M_n(\mathbb{C})$ y sea

$$\Lambda_i = \sum_{j=1, j \neq i}^n |a_{ij}|, \quad 1 \leq i \leq n.$$

Entonces, todos los valores propios λ de A caen en la unión de los discos

$$|z - a_{ii}| \leq \Lambda_i, \quad 1 \leq i \leq n.$$

Demostración

Sean λ cualquier valor propio de la matriz A , x un vector propio de A correspondiente a λ y supongamos que $|x_m| = \max_{1 \leq j \leq n} |x_j|$. Como $(\lambda I - A)x = 0$ entonces $(\lambda - a_{mm})x_m + \sum_{j \neq m} (-a_{mj})x_j = 0$, de donde tenemos que

$$|\lambda - a_{mm}| |x_m| = \left| \sum_{j \neq m} a_{mj} x_j \right| \leq \sum_{j \neq m} |a_{mj}| |x_j| \leq \sum_{j \neq m} |a_{mj}| |x_m|.$$

Luego, dividiendo ambos lados de esta desigualdad entre $|x_m| > 0$, obtenemos que $|\lambda - a_{mm}| \leq \sum_{j \neq m} |a_{mj}|$.

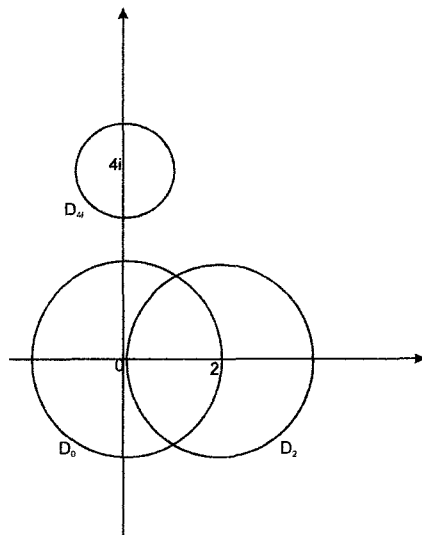
Así, el valor propio λ de A cae en el m -ésimo disco de Gerschgorin, $|\lambda - a_{mm}| \leq \Lambda_m$. Pero como λ fue tomado arbitrariamente, se sigue que los valores propios de la matriz A caen en la unión de los discos $|z - a_{ii}| \leq \Lambda_i$ para $1 \leq i \leq n$. \square

Ejemplo 3.5.4.

Sea $A \in M_3(\mathbb{C})$ tal que

$$A = \begin{pmatrix} 0 & 1 & -1 \\ 1.3 & 2 & -0.7 \\ 0.5 & 0.5i & 4i \end{pmatrix}$$

Los discos de Gerschgorin son los que se muestran en la figura siguiente.



Con el concepto de irreducibilidad podemos afinar el teorema anterior como sigue.

Teorema 3.5.4. Sea $A \in M_n(\mathbb{C})$ una matriz irreducible. Supongamos que λ es un valor propio de A en la frontera de la unión de los discos $|z - a_{ii}| \leq \Lambda_i$, entonces todos los n círculos $|z - a_{ii}| \leq \Lambda_i$ pasan por el punto λ .

Demostración

Sea λ un valor propio de A y x un vector propio de A asociado a λ . Multiplicando a x por un escalar apropiado podemos suponer que $1 = |x_r| \geq |x_i|$, donde $i = 1, \dots, n$. Entonces, como en la demostración del teorema 3.5.3

$$|\lambda - a_{rr}| \leq \sum_{j \neq r} |a_{rj}| |x_j| \leq \Lambda_r,$$

pero como λ está en la frontera de la unión de todos los discos $|z - a_{ii}| \leq \Lambda_i$, necesariamente tenemos que $|\lambda - a_{rr}| = \Lambda_r$. Por lo tanto, tenemos que

$$\sum_{j \neq r} |a_{rj}| |x_j| = \sum_{j \neq r} |a_{rj}| = \Lambda_r.$$

Luego, $|x_j| = 1$ para todo j tal que $a_{rj} \neq 0$. Como A es irreducible, existe al menos un índice $r_1 \neq r$ tal que $a_{rr_1} \neq 0$, así $|x_{r_1}| = 1$. Repitiendo el mismo argumento, se tiene que

$$|\lambda - a_{r_1 r_1}| = \sum_{j \neq r_1} |a_{r_1 j}| |x_j| = \Lambda_{r_1},$$

donde $|x_j| = 1$ para todo j tal que $a_{r_1 j} \neq 0$, y existe al menos un índice $r_2 \neq r_1$ tal que $a_{r_1 r_2} \neq 0$, y así sucesivamente. Dado que A es irreducible, se sigue del teorema 3.5.2, que existe una trayectoria dirigida

$$\overrightarrow{P_r P_{r_1}}, \overrightarrow{P_{r_1} P_{r_2}}, \dots, \overrightarrow{P_{r_m} P_j}$$

para cada $j = 1, \dots, n$ y cualquier r . Así, $a_{rr_1} \neq 0$, $a_{r_1 r_2} \neq 0, \dots, a_{r_m j} \neq 0$ y $|\lambda - a_{jj}| = \Lambda_j$ para $j = 1, \dots, n$; de aquí se tiene que λ está en la intersección de todos los n círculos. \square

En virtud de este teorema, obtenemos un refinamiento del corolario 3.3.6.

Corolario 3.5.1. Sean $A \in M_n(\mathbb{C})$ una matriz irreducible y x_1, x_2, \dots, x_n n números reales positivos arbitrarios. Si

$$\frac{\sum_{j=1}^n |a_{ij}| x_j}{x_i} \leq \nu \quad \text{para } 1 \leq i \leq n,$$

con la desigualdad estricta satisfaciéndose para al menos un subíndice i , entonces $\rho(A) < \nu$.

Análogamente, si

$$x_j \sum_{i=1}^n \frac{|a_{ij}|}{x_i} \leq \nu' \quad \text{para } 1 \leq j \leq n,$$

con la desigualdad estricta satisfaciéndose para al menos un subíndice j , entonces $\rho(A) < \nu'$.

Demostración

Sin pérdida de generalidad, supongamos que $x_1 = x_2 = \dots = x_n = 1$. La demostración completa se sigue de la demostración del corolario 3.3.6 y del teorema 3.5.3. Si

$$\sum_{j=1}^n |a_{ij}| \leq \nu \quad \text{para } 1 \leq i \leq n,$$

con la desigualdad estricta satisfaciéndose para al menos un subíndice i , entonces todos los círculos

$$|z - a_{ii}| = \Lambda_i = \sum_{j=1, j \neq i}^n |a_{ij}|$$

no pasan por un punto común σ , donde $|\sigma| = \nu$. Además, dado que $|z| = |z - a_{ii} + a_{ii}| \leq |z - a_{ii}| + |a_{ii}| \leq \sum_{j=1}^n |a_{ij}| \leq \nu$, todos los discos $|z - a_{ii}| \leq \Lambda_i$ son subconjuntos del disco $|z| \leq \nu$. Por lo tanto, por el teorema 3.5.4, ningún punto del círculo $|z| = \nu$ es valor propio de A , lo cual implica que $\rho(A) < \nu$. \square

3.6 El teorema de Perron-Frobenius para matrices irreducibles no negativas

En esta sección probamos el teorema de Perron-Frobenius para matrices irreducibles no negativas y vemos qué tanto se puede generalizar al caso de matrices reducibles no negativas.

Definición 3.6.1. Sean $A = (a_{ij})$ y $B = (b_{ij})$ dos matrices de $m \times n$ con coeficientes reales. Entonces $A \geq B$ (respectivamente, $A > B$) si $a_{ij} \geq b_{ij}$ (respectivamente, $a_{ij} > b_{ij}$) para todo i, j tales que $1 \leq i \leq m, 1 \leq j \leq n$.

Si $A \geq O$ (respectivamente, $A > O$), decimos que A es una matriz *no-negativa* (respectivamente, *positiva*).

Precisemos que $A > B$ si $A \geq B$ y $A \neq B$, es decir, si $a_{ij} \geq b_{ij}$ para todo i, j y $a_{kl} > b_{kl}$ para algún k, l .

Finalmente, si $B \in M_{m \times n}(\mathbb{C})$, entonces $|B|$ denota a la matriz con entradas $|b_{ij}|$.

Observemos que como los vectores columna son matrices de $n \times 1$, entonces tienen sentido los términos *vector no-negativo* y *vector positivo*.

Los hechos simples contenidos en el siguiente lema se siguen inmediatamente de las definiciones anteriores.

Lema 3.6.1. Sean $A \in M_n(\mathbb{C})$ y $x \in V_n(\mathbb{C})$. Entonces se cumplen:

- $|Ax| \leq |A||x|$.
- Si $A > O$, $x \geq 0$ y $x \neq 0$, entonces $Ax > 0$.
- Si $A \geq O$, $x > 0$ y $Ax = 0$, entonces $A = O$.

Lema 3.6.2. Si $A \geq O$ es una matriz irreducible de $n \times n$, entonces $(I + A)^{n-1} > O$.

Demostración

Basta probar que para cualquier vector no nulo $x > 0$ se tiene que $(I + A)^{n-1}x > 0$.

En efecto, definiendo la sucesión de vectores no-negativos $x_{k+1} = (I + A)x_k$ para $0 \leq k \leq n - 2$, donde $x_0 = x$, la demostración se sigue mostrando que x_{k+1} tiene menos componentes cero que x_k para cada $0 \leq k \leq n - 2$, siempre que x_k tenga al menos una componente cero.

Ya que $x_{k+1} = x_k + Ax_k$, es claro que el número de componentes 0 de x_{k+1} es menor o igual que el número de componentes cero de x_k .

Por otra parte, si x_{k+1} y x_k tienen exactamente el mismo número de componentes cero, entonces para una matriz de permutación P de $n \times n$ adecuada podemos escribir

$$Px_{k+1} = \begin{pmatrix} \alpha \\ 0 \end{pmatrix}, \quad Px_k = \begin{pmatrix} \beta \\ 0 \end{pmatrix}, \quad \alpha > 0, \beta > 0, \quad (1)$$

donde ambos vectores α y β tienen m componentes positivas con $1 \leq m < n$.

Considerando la matriz PAP^t , y sujetándola a la partición inducida por (1) tenemos que

$$\begin{pmatrix} \alpha \\ 0 \end{pmatrix} = \begin{pmatrix} \beta \\ 0 \end{pmatrix} + \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \beta \\ 0 \end{pmatrix}$$

donde A_{11} y A_{22} son cuadradas y A_{11} es de $m \times m$. Esto implica que $A_{21}\beta = 0$, pero como $A_{21} \geq O$ y $\beta > 0$, ésto sólo puede ocurrir cuando $A_{21} = O$.

Esto último contradice la hipótesis de que A es irreducible, por lo tanto x_{k+1} tiene menos componentes cero que x_k y como x_0 tiene a lo más $(n - 1)$ componentes cero, entonces x_k tiene a lo más $(n - k - 1)$ componentes cero. Luego, $x_{n-1} = (I + A)^{n-1}x_0$ es un vector positivo, con lo cual tenemos la conclusión deseada. \square

Si $A = (a_{ij}) \geq O$ es una matriz irreducible de $n \times n$ y $x \geq 0$ es cualquier vector distinto de cero, entonces podemos definir

$$r_x = \min \left\{ \frac{\sum_{j=1}^n a_{ij} x_j}{x_i} \right\} \quad (3.1)$$

donde el mínimo es tomado sobre todos los subíndices i para los cuales $x_i > 0$. Claramente r_x es un número real no-negativo y es el supremo de todos los números $\rho \geq 0$ para los cuales $Ax \geq \rho x$, esto es

$$r_x = \sup \{ \rho \geq 0 \mid Ax \geq \rho x \} \quad (3.2)$$

Consideremos ahora el número real no negativo r definido por

$$r = \sup_{x \geq 0} \{ r_x \}, \quad (3.3)$$

Como $r_{\alpha x}$ tiene el mismo valor que r_x para cualquier escalar $\alpha > 0$, en la definición de r necesitamos considerar sólo el conjunto P de vectores $x \geq 0$ con $\|x\| = 1$. Como no es

claro que r_x sea continua en P ya que el factor $\frac{1}{x_i}$ en la expresión $\frac{\sum_{j=1}^n a_{ij} x_j}{x_i}$ puede tender a infinito, podemos definir el conjunto Q de todos los vectores $y = (I + A)^{n-1} x$, donde $x \in P$. Por el lema 3.6.2 Q consiste sólo de vectores positivos. Además todas las componentes y_i de un vector $y \in Q$ están alejadas de 0. De hecho si β es la menor componente de la matriz positiva $(I + A)^{n-1}$, entonces $y_i \geq \beta > 0$ para todo $i = 1, \dots, n$. Multiplicando ambos lados de la desigualdad $Ax \geq r_x x$ por $(I + A)^{n-1}$ tenemos que $Ay \geq r_x y$ y concluimos por (3.2) que $r_y \geq r_x$. Por lo tanto el número real no negativo r puede ser definido equivalentemente como

$$r = \sup_{y \in Q} \{ r_y \}$$

Como P es un conjunto compacto en $V_n(\mathbb{C})$, Q también lo es, y como r_y es una función continua sobre Q , entonces necesariamente existe un vector positivo $z \in Q$ para el cual $Az \geq rz$ y no existe ningún vector $w \geq 0$ para el cual $Aw > rw$. Llamaremos a todos los vectores no-negativos y no nulos z que satisfacen esta desigualdad *vectores extremales* de la matriz A .

Lema 3.6.2. Si $A \geq O$ es una matriz irreducible de $n \times n$, el número real r es positivo. Además, cada vector extremal z es un vector propio positivo de la matriz A con valor asociado r , esto es, $Az = rz$ con $z > 0$.

Demostración

Si x es un vector cuyas componentes son todas uno, entonces ya que la matriz A es irreducible, ningún renglón de A puede ser cero, y consecuentemente ninguna componente de Ax puede anularse. Así, $r_x > 0$, probando esto que $r > 0$.

Para la segunda parte del lema, sea z un vector extremal tal que $Az - rz = \eta$. Si $\eta \neq 0$, entonces alguna componente de η es positiva; multiplicando ambos lados por la matriz $(I + A)^{n-1}$, tenemos que $Aw - rw > 0$, donde $w = (I + A)^{n-1}z > 0$. Sin embargo, por (3.1) esto implicaría que $r_w > r$, lo cual contradice a la definición de r . Por lo tanto, $Az = rz$. Ahora, un cálculo directo muestra que $w = (1 + r)^{n-1}z$ y como $w > 0$ y $r > 0$ se tiene que $z > 0$. \square

Lema 3.6.3. Sea $A = (a_{ij}) \geq O$ una matriz irreducible de $n \times n$, y sea $B = (b_{ij})$ una matriz compleja de $n \times n$ tal que $|B| \leq A$. Si β es cualquier valor propio de B entonces $|\beta| \leq r$, donde r es la constante positiva definida en (3.2). Más aún, $|\beta| = r$ si y sólo si $|B| = A$, y B tiene la forma

$$B = e^{i\theta} D A D^{-1}$$

donde D es una matriz diagonal cuyas entradas tienen módulo uno.

Demostración

Si $\beta y = B y$ con $y \neq 0$, entonces por el lema 3.6.1 a) se sigue que

$$|\beta| |y| \leq |B| |y| \leq A |y|,$$

lo cual implica que $|\beta| \leq r_{|y|} \leq r$.

Por otro lado, si $|\beta| = r$, entonces $|y|$ es un vector extremal de A . Luego, por el lema 3.6.2 $|y|$ es un vector propio positivo de A con valor propio positivo asociado r , de aquí que

$$r |y| = |B| |y| = A |y|,$$

lo que implica, por el lema 3.6.1 c), que $|B| = A$.

Para el vector y , donde $|y| > 0$, definimos $D = \text{diag}(\frac{y_1}{|y_1|}, \dots, \frac{y_n}{|y_n|})$. Es claro que las entradas de la diagonal de D tienen módulo uno y $y = D |y|$. Poniendo $\beta = r e^{i\theta}$ y sustituyendo en $B y = \beta y$ se tiene que

$$B D |y| = r e^{i\theta} D |y|,$$

de donde $e^{-i\theta} D^{-1} B D |y| = r |y| = |B| |y| = A |y|$.

Sea $C = e^{-i\theta} D^{-1} B D$, entonces $|C| = |B| = A$ y, por el lema 3.6.1 a), se tiene que

$$A |y| = C |y| = |C| |y| \leq |C| |y| = |B| |y| = A |y|.$$

Luego para cada $i = 1, \dots, n$, los números complejos $c_{ij} |y_j|$ con $(1 \leq j \leq n)$ tienen el mismo argumento y podemos escribir $c_{ij} = |c_{ij}| e^{i\theta}$ ($1 \leq j \leq n$), donde θ es independiente de j .

Entonces $C |y| = r |y|$ implica

$$e^{i\theta} \sum_{j=1}^n |c_{ij}| |y_j| = r |y_i|.$$

Como A es irreducible ningún renglón de A , y en consecuencia de $|C|$, pueden ser cero. Luego $e^{i\theta}$ es un número real positivo, de donde $\theta = 0$. Se sigue que $C = |C| = A$ y por lo tanto $B = e^{i\theta} DAD^{-1}$.

Inversamente, es obvio que si B es como en el enunciado del lema, entonces $|B| = A$ y B tiene un valor propio β tal que $|\beta| = r$, lo cual concluye la demostración. \square

Poniendo $B = A$ en el lema 3.6.3. obtenemos el siguiente resultado.

Corolario 3.6.1. Si $A \geq O$ es una matriz irreducible de $n \times n$, entonces el valor propio positivo r del lema 3.6.2 es igual al radio espectral $\rho(A)$ de la matriz A .

Teorema 3.6.4. (Perron-Frobenius) Sea $A \geq O$ una matriz irreducible de $n \times n$, entonces

- a) A tiene un valor propio real positivo igual a su radio espectral,
- b) a $\rho(A)$ le corresponde un vector propio $x > 0$,
- c) $\rho(A)$ se incrementa cuando cualquier entrada de A se incrementa,
- d) $\rho(A)$ es un valor propio simple de A .

Demostración

Sea A como en la hipótesis; a) y b) se siguen inmediatamente del lema 3.6.2 y el corolario 3.6.1.

Para probar la parte c), supongamos que incrementamos alguna entrada de la matriz A , obteniendo así una nueva matriz irreducible \tilde{A} , donde $\tilde{A} \geq A$ y $\tilde{A} \neq A$. Aplicando el lema 3.6.3 concluimos que $\rho(A) < \rho(\tilde{A})$.

Para demostrar que $\rho(A)$ es un valor propio simple de A , es decir, $\rho(A)$ es un cero de multiplicidad uno del polinomio característico $f(t) = \det(A - tI)$, mostraremos que no hay un vector propio generalizado de orden 2 asociado con $\rho(A)$.

En efecto, sean $x_1 > 0$ y $y > 0$ vectores propios de A y A^t , respectivamente, asociados con el valor propio $\rho(A)$. De esta manera, se tiene que $(A - \rho(A)I)x_1 = 0$ y $(A^t - \rho(A)I)y = 0$. Supongamos que hay un vector propio generalizado $x_2 \neq 0$ para el cual $(A - \rho(A)I)x_2 = x_1$. Entonces como $y^t(A - \rho(A)I) = 0$, tenemos que $y^t x_1 = 0$. Pero esto último contradice la positividad de x_1 y y .

Por lo tanto, $\rho(A)$ debe tener multiplicidad algebraica uno. \square

Lema 3.6.4. Sean $A \in M_n(\mathbb{C})$ y $\lambda_1, \lambda_2, \dots, \lambda_s$ sus diferentes valores propios tales que λ_j tiene multiplicidad m_j , donde $\sum_{j=1}^s m_j = n$. Entonces para todo $\varepsilon > 0$ lo suficientemente pequeño existe $\delta = \delta(\varepsilon) > 0$ tal que si $|b_{ij} - a_{ij}| \leq \delta$ para $i, j = 1, \dots, n$ entonces la matriz $B = (b_{ij})$ tiene exactamente m_j valores propios en el círculo $|\lambda - \lambda_j| < \varepsilon$ para cada $j = 1, \dots, s$.

Demostración

En esta demostración se hace uso de la teoría de funciones de una variable compleja. Sea B una matriz compleja de $n \times n$ y definamos

$$\phi(\lambda, B) = \det(B - \lambda I).$$

Para $1 \leq i < j \leq s$, sea

$$\varepsilon_o = \frac{1}{2} \min |\lambda_i - \lambda_j|.$$

Consideremos cualquier número positivo $\varepsilon < \varepsilon_o$, entonces los círculos

$$C_j : |\lambda - \lambda_j| = \varepsilon \text{ para } j = 1, \dots, s$$

no se intersecan. Definamos para $1 \leq j \leq s$,

$$\rho_j = \min_{\lambda \in C_j} |\phi(\lambda, A)|.$$

Todos los ρ_j son positivos puesto que las raíces de $\phi(\lambda, A) = 0$ son los centros de los círculos, además el mínimo está definido porque $\phi(\lambda, A)$ depende de manera continua de λ . Dado que el determinante $\phi(\lambda, B)$ depende de manera continua de todas las $1+n^2$ variables $\lambda, b_{11}, \dots, b_{nn}$ existe $\delta > 0$ tal que $\phi(\lambda, B) \neq 0$ para λ en cualquier C_j , si $|b_{ij} - a_{ij}| \leq \delta$ para todo $i, j = 1, \dots, n$. Ahora, consideremos la integral

$$n_j(B) = \frac{1}{2\pi i} \int_{C_j} \frac{\phi'(\lambda, B)}{\phi(\lambda, B)} d\lambda.$$

Esta es la integral que cuenta el número n_j de raíces λ de la ecuación $\phi(\lambda, B) = 0$ que caen dentro del círculo C_j ([8] pag. 403). Por ejemplo, tenemos que $n_j(A) = m_j$ para $j = 1, \dots, s$. Por otro lado, ya que el integrando es una función que depende de manera continua de $\lambda, b_{11}, \dots, b_{nn}$ en los conjuntos cerrados

$$|\lambda - \lambda_j| = \varepsilon, \quad |b_{ij} - a_{ij}| \leq \delta \quad i, j = 1, \dots, n,$$

la integral $n_j(B)$ es una función continua de los b_{ij} que satisfacen que $|b_{ij} - a_{ij}| \leq \delta$. Luego, como la integral no puede brincar continuamente de un entero a otro, debemos tener

$$n_j(B) = n_j(A) = m_j \quad j = 1, \dots, s,$$

para todas las matrices $B = (b_{ij})$ con $|b_{ij} - a_{ij}| \leq \delta$ ($i, j = 1, \dots, n$). \square

Corolario 3.6.2. Sea $A \geq O$ una matriz de $n \times n$, entonces:

- A tiene un valor propio real no negativo igual a su radio espectral.
- A $\rho(A)$ le corresponde un vector propio $x \geq 0$.

Demostración

En efecto, sea A una matriz reducible y en virtud del lema 3.6.4, definamos la matriz $D(t) = (d_{ij}(t)) \in M_n(\mathbb{R})$ de la siguiente manera:

$$d_{ij}(t) = \begin{cases} a_{ij} & \text{si } a_{ij} > 0 \\ t & \text{si } a_{ij} = 0 \end{cases}$$

Entonces $D(t) > 0$ para $t > 0$ y $D(0) = A$. Sea $\rho(t)$ el valor propio maximal de $D(t)$ para $t > 0$. Entonces como todos los valores propios de $D(t)$ son funciones continuas de t y como $\rho(t)$ es igual al radio espectral de $D(t)$ para $t > 0$, tenemos que $\lim_{\substack{m \rightarrow \infty \\ t \rightarrow 0^+}} \rho(t) = r$ es un

valor propio de $D(0) = A$. Además, este valor propio es igual al radio espectral de A .

Claramente, hay un vector propio $x(t)$ que puede ser asociado con $\rho(t)$ que depende continuamente de t para $t \geq 0$. Por el caso irreducible, $x(t) > 0$ para $t > 0$ de donde se tiene que $x(t) \geq 0$ en $t = 0$. \square

Observemos que en general la estructura de una matriz A no depende continuamente de sus elementos. Por esta razón, la conclusión *d*) del teorema anterior no siempre se da en el caso en que A es reducible.

Ejemplo 3.6.1.

Las siguientes matrices no son irreducibles

i) Para $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ el polinomio característico es $p_A(x) = x^2$, por lo que $\rho(A) = 0$ no es valor propio simple. Además el espacio propio correspondiente está generado por $(1, 0)$, por lo que ningún vector propio asociado a $\rho(A)$ es positivo.

ii) Para $A = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}$ el polinomio característico es $p_A(x) = x(x - 1)$, por lo que $\rho(A) = 1$ es un valor propio simple. Sin embargo, el espacio propio asociado a $\rho(A)$ está generado por $(0, 1)$, por lo que ningún vector propio asociado a $\rho(A)$ es positivo.

iii) Para $A = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}$ el polinomio característico es $p_A(x) = x(x - 1)$, por lo que $\rho(A) = 1$ es un valor propio simple. En este caso el espacio propio asociado a $\rho(A)$ está generado por $(1, 1)$, por lo que si hay un vector propio positivo correspondiente a $\rho(A)$. Observemos en este caso que el vector $(2, 1)$ es un vector extremal que no es vector propio asociado a $\rho(A)$.

4 Dos métodos iterativos clásicos y un teorema de comparación.

En este capítulo aplicaremos los resultados del capítulo anterior al estudio de métodos iterativos para resolver sistemas de ecuaciones lineales de la forma

$$Ax = b \quad (1)$$

donde $A = (a_{ij}) \in M_n(\mathbb{R})$ y b es un vector columna de orden n .

Es claro que estamos interesados en casos donde n es grande, quizás en el rango entre 10^4 y 10^6 , y donde A es “dispersa”, esto es, donde A tiene una cantidad de entradas iguales a cero suficientemente grande como para que a la hora de implementar un algoritmo computacional esto redunde en ahorro de tiempo. Supondremos además que A es no singular, de tal manera que el sistema (1) tiene solución única $x = A^{-1}b$.

4.1 Un problema modelo

Los métodos iterativos comienzan con una aproximación inicial $x^{(0)}$ a la solución y pasan por un procedimiento fijo (algoritmo) para obtener una mejor aproximación $x^{(1)}$, y así sucesivamente, para generar una sucesión $x^{(0)}, x^{(1)}, x^{(2)}, \dots$; si el método es apropiadamente elegido, ésta convergerá a la solución exacta x de (1). Como x es desconocida, un criterio para terminar la iteración puede ser cuando

$$\frac{|x_i^{(k+1)} - x_i^{(k)}|}{|x_i^{(k+1)}|} < \varepsilon, \quad i = 1, \dots, n$$

donde ε es un número pequeño preestablecido, dependiendo de la precisión de la computadora que esté siendo utilizada. Esencialmente el criterio de terminación de la iteración es que elegimos $x_i^{(k+1)}$ como la aproximación a x_i cuando la diferencia relativa entre las aproximaciones sucesivas $x_i^{(k)}$ y $x_i^{(k+1)}$ se hacen suficientemente pequeñas para $i = 1, \dots, n$.

Estos algoritmos iterativos tienen la ventaja de que A no es alterada durante el cómputo, de aquí que el problema de errores de redondeo es mucho menos serio que en los métodos directos, donde la matriz es descompuesta en el proceso de iteración. Además, la cantidad de almacenamiento de datos que se requiere para una matriz de coeficientes dispersa es proporcional a n y, en el caso de la *eliminación gaussiana*, dado que el proceso tiende a llenar los ceros de A , la cantidad de almacenamiento de datos requerida es proporcional a n^2 . Así, es claro que esto representa un problema cuando n es grande. Los métodos iterativos que describiremos aquí no requerirán ningún almacenamiento extra y por lo tanto son más prácticos. Una desventaja que ofrecen es que después de resolver $Ax = b_1$, debemos volver a comenzar desde el principio para resolver $Ax = b_2$.

Es importante hacer notar que los sistemas lineales con matrices dispersas pueden generarse de manera muy natural en la práctica, por ejemplo, cuando nos planteamos

modelar un problema continuo de manera discreta. El problema continuo no puede ser resuelto de manera exacta en una computadora. Así, éste tendrá que ser aproximado por un modelo discreto; naturalmente, mientras más precisión en el modelo discreto se desee, se generará más costo. Como un simple pero muy típico problema continuo, elijamos la ecuación diferencial siguiente:

$$-\frac{d^2u}{dx^2} = g(x), \quad 0 \leq x \leq 1 \quad (2)$$

Esta es una ecuación diferencial ordinaria de segundo orden con término no homogéneo $g(x)$. En virtud de que cualquier combinación de la forma $ax + b$ puede ser sumada a cualquier solución $u(x)$ de (2) y producir otra solución, impongamos condiciones en la frontera, es decir,

$$u(0) = \alpha, \quad u(1) = \beta \quad (3)$$

con el objetivo de determinar las constantes a y b . De esta manera (2) y (3) constituyen un problema con dos valores en la frontera y con una solución única $u = u(x)$. Este planteamiento puede ser utilizado para describir la temperatura en una barra de longitud uno con temperaturas fijas α en $x = 0$, β en $x = 1$ y con una fuente de distribución de calor $g(x)$.

Como lo que buscamos es un modelo discreto para este problema continuo, no podemos aceptar más que una cantidad finita de información acerca de la función g . Pongamos valores distribuidos de manera uniforme en el intervalo $[0, 1]$ a través de los puntos de la partición $x_0 = 0, x_1 = h, x_2 = 2h, \dots, x_n = nh, x_{n+1} = 1$ donde $h = \frac{1}{n+1}$.

Calculemos las aproximaciones u_1, \dots, u_n a la solución exacta u en esos puntos, esto es, u_j será la aproximación a $u(jh)$ para $j = 1, \dots, n$. En la frontera tendremos $u_0 = u(0) = \alpha$ y $u_{n+1} = u(1) = \beta$ exactamente. Con el objeto de discretizar la ecuación diferencial ordinaria (2) utilizaremos la aproximación siguiente:

$$\frac{d^2u(x)}{dx^2} \approx \frac{u(x+h) - 2u(x) + u(x-h)}{h^2} \quad (4)$$

El lado derecho de esta expresión aproxima a $\frac{d^2u(x)}{dx^2}$ cuando $h \rightarrow 0$, pero siempre para una h finita. Siendo así, para la partición $x_j = jh$ de $(0, 1)$, la ecuación diferencial (2) es reemplazada por su discreta análoga (4) y después de multiplicar por h^2 tenemos

$$-u_{j+1} + 2u_j - u_{j-1} = h^2 g(jh), \quad j = 1, \dots, n. \quad (5)$$

Así obtenemos exactamente n ecuaciones lineales con incógnitas u_1, \dots, u_n . Notemos que la primera y última ecuación incluyen las expresiones $u_0 = \alpha$ y $u_{n+1} = \beta$, respectivamente, y que pueden ser cambiadas al lado derecho de la ecuación (5). La estructura de las ecuaciones (5) para $h = \frac{1}{5}$ o $n = 4$ está dada por

$$\begin{aligned}
2u_1 - u_2 &= h^2g(h) + \alpha \\
-u_1 + 2u_2 - u_3 &= h^2g(2h) \\
-u_2 + 2u_3 - u_4 &= h^2g(3h) \\
-u_3 + 2u_4 &= h^2g(4h) + \beta
\end{aligned} \tag{6}$$

Ahora, simplemente resolvamos el sistema (6), es decir, la primera ecuación para u_1 , la segunda para u_2 , la tercera para u_3 , etc., de donde obtenemos

$$\begin{aligned}
u_1 &= \frac{1}{2}u_2 + \frac{1}{2}[h^2g(h) + \alpha] \\
u_2 &= \frac{1}{2}u_1 + \frac{1}{2}u_3 + \frac{1}{2}h^2g(2h) \\
u_3 &= \frac{1}{2}u_2 + \frac{1}{2}u_4 + \frac{1}{2}h^2g(3h) \\
u_4 &= \frac{1}{2}u_3 + \frac{1}{2}[h^2g(4h) + \beta]
\end{aligned}$$

Entonces, comenzando con un vector inicial de aproximación $u^o = (u_1^o, u_2^o, u_3^o, u_4^o)^t$ a la solución u , en cualquier etapa, dado $u^i = (u_1^i, u_2^i, u_3^i, u_4^i)^t$, podemos determinar u^{i+1} haciendo

$$\begin{aligned}
u_1^{i+1} &= \frac{1}{2}u_2^i + \frac{1}{2}[h^2g(h) + \alpha] \\
u_2^{i+1} &= \frac{1}{2}u_1^i + \frac{1}{2}u_3^i + \frac{1}{2}h^2g(2h) \\
u_3^{i+1} &= \frac{1}{2}u_2^i + \frac{1}{2}u_4^i + \frac{1}{2}h^2g(3h) \\
u_4^{i+1} &= \frac{1}{2}u_3^i + \frac{1}{2}[h^2g(4h) + \beta]
\end{aligned}$$

Se puede mostrar que el método converge para este ejemplo. Sin embargo, podría necesitarse un número grande de iteraciones si n es grande, es decir, si el ancho de la partición h es pequeño. Este número de iteraciones puede reducirse en aproximadamente 50%, usando los valores de u_j^{i+1} tan pronto estén disponibles en lugar de u_j^i . Así por ejemplo, la segunda ecuación del sistema inmediato anterior sería

$$u_2^{i+1} = \frac{1}{2}u_1^{i+1} + \frac{1}{2}u_3^i + \frac{1}{2}[h^2g(2h)].$$

Muchas de las herramientas utilizadas para establecer la convergencia y las velocidades de convergencia de estos métodos iterativos básicos se basan generalmente en la teoría de matrices no-negativas. Finalmente, en forma matricial $Au = b$, el sistema lineal (6) toma la forma

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} = \begin{bmatrix} h^2g(h) + \alpha \\ h^2g(2h) \\ h^2g(3h) \\ h^2g(4h) + \beta \end{bmatrix}$$

Y en general, para $n \geq 2$, la matriz de coeficientes A asociada a este sistema está dada por

$$A = \begin{bmatrix} 2 & -1 & 0 & \dots & 0 \\ -1 & 2 & -1 & \dots & 0 \\ 0 & -1 & \ddots & \dots & \vdots \\ \vdots & \vdots & \vdots & \ddots & -1 \\ 0 & \dots & \dots & -1 & 2 \end{bmatrix}$$

Notemos que A es simétrica, irreducible (su gráfica dirigida es fuertemente conexa), no-singular y $A^{-1} \geq O$. Esta información es muy importante como lo veremos más adelante.

Algunas de estas propiedades de A se tienen presentes generalmente para sistemas de ecuaciones lineales originados por ecuaciones diferenciales ordinarias y parciales, resultando de manera natural métodos numéricos para la aproximación a sus soluciones.

Definición 4.1.1. Sea $A \in M_n(\mathbb{R})$. A es una L -matriz si $a_{ii} > 0$ para $1 \leq i \leq n$ y $a_{ij} \leq 0$ para $i \neq j$, $i, j = 1, 2, \dots, n$.

Definición 4.1.2. Sea $A \in M_n(\mathbb{R})$. A es una M -matriz si $a_{ij} \leq 0$ para $i \neq j$, $i, j = 1, 2, \dots, n$, A es no singular y $A^{-1} \geq O$.

Teorema 4.1.1. Sean $A \in M_n(\mathbb{R})$ una L -matriz, $D = \text{diag}(a_{11}, \dots, a_{nn})$, $C = D - A$ y $B = D^{-1}C$. Entonces A es una M -matriz si y sólo si $\rho(B) < 1$.

Demostración

Si $\rho(B) < 1$ entonces, por el corolario 3.4.2, $I - B$ es no singular y la serie $I + B + B^2 + \dots$ converge a $(I - B)^{-1}$. Como $D \geq O$ y $C \geq O$ tenemos que $B \geq O$ y $(I - B)^{-1} \geq O$. Dado que D e $I - B$ son no singulares, se sigue que $A = D(I - B)$ es no singular y

$$A^{-1} = (I - B)^{-1}D^{-1} \geq O.$$

Así, A es una M -matriz.

Inversamente, si A es una M -matriz, definamos la matriz

$$\widehat{A} = D^{-\frac{1}{2}}AD^{-\frac{1}{2}} = I - D^{-\frac{1}{2}}CD^{-\frac{1}{2}},$$

donde $D^{-\frac{1}{2}} = \text{diag}(a_{11}^{-\frac{1}{2}}, \dots, a_{nn}^{-\frac{1}{2}})$. Evidentemente, \widehat{A} es una M -matriz. Más aún, B es similar a \widetilde{B} , donde

$$\widetilde{B} = D^{\frac{1}{2}}BD^{-\frac{1}{2}}.$$

Por cálculo directo podemos verificar que

$$(I - \widetilde{B})(I + \widetilde{B} + \widetilde{B}^2 + \dots + \widetilde{B}^m) = I - \widetilde{B}^{m+1},$$

y luego

$$\widehat{A}^{-1} = (I - \widetilde{B})^{-1} = (I + \widetilde{B} + \widetilde{B}^2 + \dots + \widetilde{B}^m) + (I - \widetilde{B})^{-1}\widetilde{B}^{m+1}.$$

Dado que $(I - \widetilde{B})^{-1} \geq O$ y $\widetilde{B} \geq O$, se sigue que cada elemento de la matriz $I + \widetilde{B} + \widetilde{B}^2 + \dots + \widetilde{B}^m$ es una función no decreciente de m , la cual es acotada por el correspondiente elemento de $(I - \widetilde{B})^{-1}$. Luego, la serie $I + \widetilde{B} + \widetilde{B}^2 + \dots + \widetilde{B}^m$ debe converger. Por el teorema 3.4.1 esto sucede sólo si $\rho(\widetilde{B}) = \rho(B) < 1$. \square

4.2 Descripción de los métodos iterativos.

Estamos ahora en posición de describir los métodos iterativos que son objeto de análisis en este trabajo. Cada método de iteración tiene una región de aplicación limitada. A veces un método puede resultar divergente al aplicarlo a determinados sistemas; en otros casos, la convergencia puede resultar tan lenta que en la práctica se vuelve inútil para la obtención de una aproximación satisfactoria. De esto se deduce que en el método de iteración resulte esencial su convergencia y la velocidad de la misma. Para que el proceso de iteración converja lo más rápidamente posible es necesario preparar previamente el sistema (matriz) sustituyéndolo por uno equivalente. Esta preparación jugará un papel fundamental.

El proceso de iteración más simple para la resolución del sistema (1) es el de *aproximaciones sucesivas*. Consideremos el sistema de ecuaciones lineales (1), sumando y restando x_i en el miembro de la izquierda de la i -ésima ecuación para $i = 1, 2, \dots, n$, el sistema (1) se transforma en el sistema equivalente

$$\begin{array}{rcccccc} x_1 = & (1 - a_{11})x_1 - & a_{12}x_2 - & \dots & -a_{1n}x_n + & b_1 \\ x_2 = & -a_{21}x_1 + (1 - a_{22})x_2 - & \dots & -a_{2n}x_n + & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_n = & -a_{n1}x_1 - a_{n2}x_2 - & \dots & +(1 - a_{nn})x_n + & b_n \end{array}$$

cuya expresión matricial es $x = (I - A)x + b$.

Si ahora designamos $H = I - A$ y $c = b$, obtenemos $x = Hx + c$. El proceso de aproximaciones sucesivas se establece por medio de la fórmula

$$x^{(m+1)} = Hx^{(m)} + c \quad \text{para } m = 0, 1, 2, \dots$$

comenzando con cierta aproximación $x^{(0)}$ que, en general, se puede seleccionar de manera arbitraria.

La fórmula anterior corresponde al sistema de ecuaciones

$$x_i^{(m+1)} = \sum_{j=1}^n h_{ij} x_j^{(m)} + c_i \quad \text{para } i = 1, 2, \dots, n ; m = 0, 1, 2, \dots$$

En el proceso de aproximaciones sucesivas, $H = I - A$ se llama la *matriz de iteración*. En una sección posterior se verá que si la matriz de coeficientes del sistema (1) está muy próxima a la matriz identidad I , se obtendrá una rápida convergencia del método. Cuando esta condición no se cumpla es recomendable preparar previamente el sistema hasta lograrla. Esto nos lleva a plantear los métodos iterativos objeto de nuestro estudio.

Sea $A \in M_n(\mathbb{R})$ y supongamos que deseamos resolver el sistema (1), donde suponemos que la matriz de coeficientes $A = (a_{ij})$ es no singular y $a_{ii} \neq 0$ para $i = 1, \dots, n$. Escribamos A en la forma $A = M - N$, donde M es una matriz no singular que sea fácilmente invertible. Entonces el sistema (1) puede reescribirse en la forma

$$Mx = Nx + b \quad \text{o} \quad x = M^{-1}Nx + M^{-1}b.$$

Si hacemos $H = M^{-1}N = I - M^{-1}A$ y $c = M^{-1}b$ entonces $x = Hx + c$.

Para resolver la ecuación anterior, comenzamos con una conjetura inicial $x^{(0)}$, que puede ser cualquier vector en \mathbb{R}^n . Luego hacemos

$$\begin{aligned} x^{(1)} &= Hx^{(0)} + c \\ x^{(2)} &= Hx^{(1)} + c \end{aligned}$$

y, en general,

$$x^{(i+1)} = Hx^{(i)} + c, \quad i = 0, 1, \dots$$

La matriz H es la *matriz de iteración* asociada al método iterativo y se tiene que $x = Hx + c$ si y sólo si $Ax = b$.

A cada método iterativo podemos asociarle los *vectores de error* $e^{(i)}$ definidos por

$$e^{(i)} = x^{(i)} - x, \quad i \geq 0.$$

La elección más simple de M (excluyendo aproximaciones sucesivas) consiste en hacer que M sea una matriz diagonal cuyos elementos diagonales sean los elementos diagonales de A . El esquema de iteración con esta elección de M se conoce como *iteración de Jacobi*.

ITERACIÓN DE JACOBI

Sean

$$M = D = \text{diag}(a_{11}, \dots, a_{nn})$$

y

$$N = - \begin{bmatrix} 0 & a_{12} & \dots & a_{1n} \\ a_{21} & 0 & & a_{2n} \\ \vdots & & \ddots & \\ a_{n1} & a_{n2} & \dots & 0 \end{bmatrix}$$

Hagamos $H = D^{-1}N$ y $c = D^{-1}b$, esto es,

$$H = \begin{bmatrix} 0 & -\frac{a_{12}}{a_{11}} & \dots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & \dots & -\frac{a_{2n}}{a_{22}} \\ \vdots & & \ddots & \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & \dots & 0 \end{bmatrix} \quad \text{y} \quad c = \begin{bmatrix} \frac{b_1}{a_{11}} \\ \frac{b_2}{a_{22}} \\ \vdots \\ \frac{b_n}{a_{nn}} \end{bmatrix}$$

Asumiendo que el i -ésimo vector de aproximación $x^{(i)}$ a $x = A^{-1}b$ ha sido calculado, entonces el $i + 1$ -ésimo vector de aproximación se calcula por

$$x_j^{(i+1)} = \frac{1}{a_{jj}} \left(-\sum_{k=1, k \neq j}^n a_{jk} x_k^{(i)} + b_j \right), \quad j = 1, \dots, n. \quad (\Delta)$$

Observemos que los vectores $x^{(i)}$ y $x^{(i+1)}$ en términos computacionales deben almacenarse por separado.

Una alternativa a la iteración de Jacobi consiste en tomar a M como la parte triangular inferior de A , esto con la esperanza de que como M es una mejor aproximación a A , también M^{-1} lo sea a A^{-1} . El esquema de iteración con esta elección de M recibe el nombre de *iteración de Gauss-Seidel*. Uno de los objetivos de este trabajo es mostrar que esta iteración converge más rápido que la *iteración de Jacobi*.

ITERACIÓN DE GAUSS-SEIDEL

Sean

$$D = \text{diag}(a_{11}, \dots, a_{nn})$$

$$L = - \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ a_{21} & 0 & \dots & 0 & 0 \\ \vdots & & \ddots & & \\ a_{(n-1)1} & a_{(n-1)2} & \dots & 0 & 0 \\ a_{n1} & a_{n2} & \dots & a_{n(n-1)} & 0 \end{bmatrix}$$

y

$$U = - \begin{bmatrix} 0 & a_{12} & \dots & a_{1(n-1)} & a_{1n} \\ 0 & 0 & \dots & a_{2(n-1)} & a_{2n} \\ \vdots & \vdots & \ddots & & \\ 0 & 0 & \dots & 0 & a_{(n-1)n} \\ 0 & 0 & \dots & 0 & 0 \end{bmatrix}$$

Hagamos $M = D - L$ y $N = U$. Sea $x = x^{(0)}$ un vector distinto de cero en \mathbb{R}^n . Entonces, se tiene que

$$\begin{aligned} Mx^{(i+1)} &= Nx^{(i)} + b \\ (D - L)x^{(i+1)} &= Ux^{(i)} + b \\ Dx^{(i+1)} &= Lx^{(i+1)} + Ux^{(i)} + b \end{aligned}$$

Podemos resolver esta última ecuación para obtener $x^{(i+1)}$, una coordenada a la vez. La primera coordenada de $x^{(i+1)}$ está dada por

$$x_1^{(i+1)} = \frac{1}{a_{11}} \left(-\sum_{k=2}^n a_{1k}x_k^{(i)} + b_1 \right).$$

La segunda coordenada de $x^{(i+1)}$ se puede resolver en términos de la primera coordenada y las últimas $n - 2$ coordenadas de $x^{(i)}$ como sigue

$$x_2^{(i+1)} = \frac{1}{a_{22}} \left(-a_{21}x_1^{(i+1)} - \sum_{k=3}^n a_{2k}x_k^{(i)} + b_2 \right).$$

En general, se tiene

$$x_j^{(i+1)} = \frac{1}{a_{jj}} \left(-\sum_{k=1}^{j-1} a_{jk}x_k^{(i+1)} - \sum_{k=j+1}^n a_{jk}x_k^{(i)} + b_j \right). \quad (\blacktriangle)$$

Es interesante comparar (\triangle) y (\blacktriangle) . La diferencia entre las iteraciones de Jacobi y de Gauss-Seidel es que en el segundo caso se utilizan las coordenadas de $x^{(i+1)}$ tan pronto como se calculan y no en la siguiente iteración. El programa de la iteración de Gauss-Seidel es realmente más sencillo que el de la iteración de Jacobi (ver apéndice). Los vectores $x^{(i)}$ y $x^{(i+1)}$ se almacenan en el mismo vector x . Cuando se calcula una coordenada de $x^{(i+1)}$, ésta reemplaza la coordenada correspondiente de $x^{(i)}$.

Finalmente, en forma matricial tenemos a la iteración de Jacobi como

$$x^{(i+1)} = D^{-1}(L + U)x^{(i)} + D^{-1}b, \quad i = 0, 1, \dots$$

y a la iteración de Gauss-Seidel como

$$x^{(i+1)} = (D - L)^{-1}Ux^{(i)} + (D - L)^{-1}b, \quad i = 0, 1, \dots$$

o equivalentemente

$$x^{(i+1)} = L'x^{(i+1)} + U'x^{(i)} + c, \quad i = 0, 1, \dots$$

donde $L' = D^{-1}L$ y $U' = D^{-1}U$. Dado que L' es estrictamente triangular inferior, $\det(I - L') = 1$ e $I - L'$ es no singular, entonces

$$x^{(i+1)} = (I - L')^{-1}U'x^{(i)} + (I - L')^{-1}c, \quad i = 0, 1, \dots$$

Entonces, para el método de Jacobi $H = M^{-1}N = D^{-1}(L + U) = L' + U'$ y para el método de Gauss-Seidel $H = M^{-1}N = (D - L)^{-1}U = (I - L')^{-1}U'$.

Definición 4.2.1 Sea $A \in M_n(\mathbb{R})$ no singular con $a_{ii} \neq 0$ para $1 \leq i \leq n$. Definimos la matriz de Jacobi como $B = L' + U'$ y la matriz de Gauss-Seidel como $\mathcal{L} = (I - L')^{-1}U'$.

4.3 Sobre la convergencia y estimación de errores.

Recordemos que nuestro principal objetivo es mostrar que una aplicación importante del teorema de Perron-Frobenius se da en la demostración de un teorema de comparación de la convergencia de los métodos iterativos descritos anteriormente. En primera instancia establezcamos una condición suficiente para la convergencia individual de dichos métodos.

Supongamos entonces que cualquiera de los tres métodos iterativos descritos en la sección anterior (aproximaciones sucesivas, Jacobi, Gauss-Seidel) converge, es decir,

$$x^{(m)} \text{ converge a } x^* \text{ cuando } m \rightarrow \infty.$$

Pasando al límite en la ecuación $x^{(m+1)} = Hx^{(m)} + c$, obtenemos $x^* = Hx^* + c$, lo que indica que x^* satisface al sistema considerado. Por consiguiente, si el proceso iterativo converge, lo hará a la solución del sistema.

Consideremos ahora la ecuación $x = Hx + c$. En cualquiera de los procesos iterativos anteriores obtenemos la iteración $x^{(i)} = Hx^{(i-1)} + c$ para $i = 1, 2, \dots$. Entonces

$$\begin{aligned} x^{(1)} - x &= H(x^{(0)} - x) \\ x^{(2)} - x &= H^2(x^{(0)} - x) \\ &\vdots \\ x^{(i)} - x &= H^i(x^{(0)} - x) \end{aligned}$$

de donde, por el teorema 3.2.1 d),

$$\|x^{(i)} - x\| = \|H^i(x^{(0)} - x)\| \leq \|H^i\| \|x^{(0)} - x\| \leq \|H\|^i \|x^{(0)} - x\|.$$

Luego, si $\|H\| < 1$ tenemos que $\|x^{(i)} - x\| \rightarrow 0$ cuando $i \rightarrow \infty$, esto es, $x^{(i)} \rightarrow x$.

Retomando la discusión de los métodos iterativos de Jacobi y Gauss-Seidel, vemos que no sólo se necesita que la matriz $M = A + N$ sea fácilmente invertible sino que M^{-1} sea una aproximación lo suficientemente buena a A^{-1} como para que

$$\|I - M^{-1}A\| = \|M^{-1}N\| = \|H\| < 1.$$

En virtud del teorema 3.4.1 esta última condición implica que todos los valores propios de H tiene módulo menor que 1.

Teorema 4.3.1. Sea $A = M - N \in M_n(\mathbb{C})$ con A y M matrices no singulares. Entonces para $H = M^{-1}N$ y $c = M^{-1}b$, el método iterativo converge a la solución de $Ax = b$ para cada conjetura inicial $x^{(0)}$ si y sólo si $\rho(H) < 1$.

Demostración

Demostraremos primero el teorema en el caso en que H tiene n vectores propios linealmente independientes. En efecto, si x_1, \dots, x_n son n vectores propios linealmente independientes de H con valores propios $\lambda_1, \dots, \lambda_n$ respectivamente, podemos escribir $x^{(0)} - x = \alpha_1 x_1 + \dots + \alpha_n x_n$ para algunos escalares $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ y por lo tanto

$$\begin{aligned} x^{(i)} - x &= H^i(\alpha_1 x_1 + \dots + \alpha_n x_n) \\ &= \alpha_1 \lambda_1^i x_1 + \dots + \alpha_n \lambda_n^i x_n \end{aligned}$$

de aquí se deduce que $\|x^{(i)} - x\| \rightarrow 0$ si y sólo si $|\lambda_k| < 1$ para $k = 1, \dots, n$. Por lo tanto $x^{(i)} \rightarrow x$ si y sólo si $\rho(H) < 1$.

En el caso general, como en el párrafo anterior al teorema, tenemos la ecuación

$$x^{(i)} - x = H^i(x^{(0)} - x)$$

Tomando el límite cuando $i \rightarrow \infty$, obtenemos que $x^{(i)} \rightarrow x$ si y sólo si

$$\lim_{i \rightarrow \infty} H^i = O,$$

que se da si y sólo si $\rho(H) < 1$ según el teorema 3.4.1. \square

Definición 4.3.1. Sea $A \in M_n(\mathbb{R})$. Se dice que A es *diagonalmente dominante* si

$$|a_{ii}| > \sum_{j=1, j \neq i}^n |a_{ij}| \quad \text{para } i = 1, \dots, n$$

Teorema 4.3.2. Si A es diagonalmente dominante, entonces la iteración de Jacobi y la iteración de Gauss-Seidel convergen a una solución de $Ax = b$.

Demostración

Si A es diagonalmente dominante, la matriz H de la iteración de Jacobi tendrá la propiedad

$$\sum_{j=1}^n |h_{ij}| = \sum_{j=1, j \neq i}^n \frac{|a_{ij}|}{|a_{ii}|} < 1 \quad \text{para } i = 1, \dots, n.$$

Por lo tanto

$$\|H\|_{\infty} = \max_{1 \leq i \leq n} \sum_{j=1}^n |h_{ij}| < 1.$$

Luego, la iteración de Jacobi converge.

Consideremos ahora el caso de la iteración de Gauss-Seidel. Para $j = 1, \dots, n$, sean

$$\alpha_j = \sum_{i=1}^{j-1} |a_{ji}|, \quad \beta_j = \sum_{i=j+1}^n |a_{ji}| \quad \text{y} \quad M_j = \frac{\beta_j}{(|a_{jj}| - \alpha_j)}$$

Como A es diagonalmente dominante, se deduce que $|a_{jj}| > \alpha_j + \beta_j$ y en consecuencia, $M_j < 1$ para $j = 1, \dots, n$. Por lo tanto,

$$M = \max_{1 \leq j \leq n} M_j < 1.$$

Demostraremos que

$$\|H\|_{\infty} = \max_{x \neq 0} \frac{\|Hx\|_{\infty}}{\|x\|_{\infty}} \leq M < 1.$$

Sea x un vector distinto de cero en \mathbb{R}^n y sea $y = Hx$. Elijamos k de modo que

$$\|y\|_{\infty} = \max_{1 \leq i \leq n} |y_i| = |y_k|.$$

Se sigue de la definición de H que $y = Hx = (D - L)^{-1}Ux$, y de aquí

$$y = D^{-1}(Ly + Ux).$$

Comparando la k -ésima coordenada de cada lado, se observa que

$$y_k = \frac{1}{a_{kk}} \left(- \sum_{i=1}^{k-1} a_{ki} y_i - \sum_{i=k+1}^n a_{ki} x_i \right)$$

y por lo tanto

$$\|y\|_{\infty} = |y_k| \leq \frac{1}{|a_{kk}|} (\alpha_k \|y\|_{\infty} + \beta_k \|x\|_{\infty}).$$

De la desigualdad anterior se sigue que

$$\frac{\|Hx\|_{\infty}}{\|x\|_{\infty}} = \frac{\|y\|_{\infty}}{\|x\|_{\infty}} \leq M_k \leq M.$$

Finalmente, se tiene que

$$\|H\|_{\infty} = \max_{x \neq 0} \frac{\|Hx\|_{\infty}}{\|x\|_{\infty}} \leq M < 1$$

y en consecuencia la iteración convergerá a la solución de $Ax = b$. \square

Veamos ahora cómo estimar el error que se comete en el método iterativo. Consideremos cualesquiera dos aproximaciones sucesivas $x^{(m)} = Hx^{(m-1)} + c$ y $x^{(m+1)} = Hx^{(m)} + c$. Restando y tomando normas obtenemos

$$\|x^{(m+1)} - x^{(m)}\| \leq \|H\| \|x^{(m)} - x^{(m-1)}\| \quad \text{para } m = 1, 2, \dots \quad (*)$$

Si $\|H\| < 1$, el método converge y, denotando por x^* a la solución del sistema podemos escribir

$$x^* = x^{(m)} + (x^{(m+1)} - x^{(m)}) + (x^{(m+2)} - x^{(m+1)}) + \dots + (x^{(m+p)} - x^{(m+p-1)}) + \dots,$$

de donde,

$$\|x^* - x^{(m)}\| \leq \|x^{(m+1)} - x^{(m)}\| + \|x^{(m+2)} - x^{(m+1)}\| + \dots + \|x^{(m+p)} - x^{(m+p-1)}\| + \dots$$

y usando (*) recursivamente obtenemos

$$\|x^* - x^{(m)}\| \leq \|H\| \|x^{(m)} - x^{(m-1)}\| (1 + \|H\| + \dots + \|H\|^{p-1} + \dots)$$

Por ser $\|H\| < 1$, $(1 + \|H\| + \dots + \|H\|^{p-1} + \dots)$ converge a $(1 - \|H\|)^{-1}$. Luego

$$\|x^* - x^{(m)}\| \leq \frac{\|H\| \|x^{(m)} - x^{(m-1)}\|}{1 - \|H\|} \quad (**)$$

La desigualdad anterior nos permite estimar el error que se comete en la aproximación $x^{(m)}$ si conocemos su valor y el de su precedente. Por ejemplo, si $\frac{\|H\|}{1 - \|H\|} \leq 1$, entonces

$$\|x^* - x^{(m)}\| \leq \|x^{(m)} - x^{(m-1)}\|.$$

Podemos obtener otra expresión que nos permite estimar el error que se comete en la n -ésima aproximación antes de que la hallamos calculado con sólo conocer la aproximación inicial $x^{(0)}$ y la siguiente $x^{(1)}$. En efecto, de (*) obtenemos recursivamente que

$$\|x^{(m)} - x^{(m-1)}\| \leq \|H^{m-1}\| \|x^{(1)} - x^{(0)}\| \leq \|H\|^{m-1} \|x^{(1)} - x^{(0)}\|,$$

y sustituyendo en (**) obtenemos

$$\|x^* - x^{(m)}\| \leq \frac{\|H\|^m \|x^{(1)} - x^{(0)}\|}{1 - \|H\|},$$

que es una expresión que depende de las 2 primeras aproximaciones.

Con frecuencia es importante comparar la exactitud de 2 aproximaciones sucesivas, es decir, los valores de $x^* - x^{(m)}$ y $x^* - x^{(m+1)}$. Este estimado de la velocidad de convergencia del proceso de aproximaciones sucesivas se obtiene de la siguiente manera. De las igualdades $x^* = Hx^* + c$ y $x^{(m+1)} = Hx^{(m)} + c$ se obtiene, restándolas y tomando normas, que

$$\|x^* - x^{(m+1)}\| \leq \|H\| \|x^* - x^{(m)}\|.$$

Si como antes, para los errores usamos la notación $e^{(m+1)} = x^* - x^{(m+1)}$, podemos escribir

$$\|e^{(m+1)}\| \leq \|H\| \|e^{(m)}\|.$$

De la desigualdad anterior se deduce que el error de cualquier aproximación es proporcional al de la precedente y, cuanto menor sea $\|H\|$, tanto más rápida será la convergencia. Para concluir esta sección, ilustremos los métodos iterativos de Jacobi y Gauss-Seidel.

Ejemplo 4.3.1.

Sea $A \in M_4(\mathbb{R})$ definida como

$$A = \begin{bmatrix} 1 & 0 & -\frac{1}{4} & -\frac{1}{4} \\ 0 & 1 & -\frac{1}{4} & -\frac{1}{4} \\ -\frac{1}{4} & -\frac{1}{4} & 1 & 0 \\ -\frac{1}{4} & -\frac{1}{4} & 0 & 1 \end{bmatrix}.$$

Específicamente, queremos resolver el sistema $Ax = k$, donde

$$k = \frac{1}{2} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

Claramente, la única solución a dicho sistema es

$$x_1 = x_2 = x_3 = x_4 = 1.$$

Si nuestro vector inicial de aproximación es $x^{(0)} = 0$, entonces el vector inicial $e^{(0)}$ tiene todas sus componentes iguales a -1 . Para este caso, la matriz H de la iteración de Jacobi y la de la iteración de Gauss-Seidel son respectivamente

$$H = M^{-1}N = D^{-1}(L + U) = \frac{1}{4} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{bmatrix}$$

$$H = M^{-1}N = (D - L)^{-1}U = \frac{1}{4} \begin{bmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

Si $\alpha^{(m)}$ y $\beta^{(m)}$ denotan los errores para los métodos iterativos de Jacobi y Gauss-Seidel respectivamente, entonces se tiene que

$$\alpha^{(m)} = H^m e^{(0)} = -\frac{1}{2^m} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \beta^{(m)} = H^m e^{(0)} = -\frac{1}{4^m} \begin{bmatrix} 2 \\ 2 \\ 1 \\ 1 \end{bmatrix} \quad m \geq 1$$

Así, como

$$\|\alpha^{(m)}\| = \frac{1}{2^{m-1}} > \|\beta^{(m)}\| = \frac{\sqrt{10}}{4^m} \quad \forall m \geq 1,$$

podemos concluir que $\|\beta^{(m)}\|$ tiende más rápidamente a cero que $\|\alpha^{(m)}\|$ para esta elección particular de $e^{(0)}$. Esto último nos permite sospechar que en algunos casos la iteración de Gauss-Seidel converge más rápido que la iteración de Jacobi.

4.4 Velocidades de convergencia.

En esta sección nos avocaremos a una formulación precisa de lo que significa que un método iterativo sea más rápido que otro.

Comencemos analizando la velocidad de convergencia de un método iterativo convergente. Por el teorema 4.3.1, si $\rho(H) < 1$ entonces para cualquier conjetura inicial $x^{(0)}$, la sucesión $x^{(0)}, x^{(1)}, \dots$ de aproximaciones sucesivas generada por el método iterativo converge a la solución única \hat{x} de $(I - H)x = c$. Poniendo $e^{(n)} = \hat{x} - x^{(n)}$, buscamos determinar la velocidad a la cual $\|e^{(n)}\|$ converge a cero cuando $n \rightarrow \infty$.

Primero, definamos una cantidad conocida como *número de condición* para una matriz A . Sea $A \in M_n(\mathbb{C})$ una matriz no singular y consideremos el sistema (1). Si x^* es la solución exacta al sistema, x' es una aproximación a la solución, $e = x^* - x'$ y $\|\cdot\|$ es una norma en $V_n(\mathbb{C})$ entonces $\|e\|$ es una medida del *error absoluto* y $\frac{\|e\|}{\|x^*\|}$ es una medida del *error relativo*. En general, no tenemos manera de determinar los valores exactos de $\|e\|$ y $\frac{\|e\|}{\|x^*\|}$. Una forma posible de comprobar la exactitud de x' consiste en volverlo a colocar en el sistema (1) y ver qué tanto se acerca $b' = Ax'$ a b ; el vector $r = b - b' = b - Ax'$

se llama *residuo* y se puede calcular fácilmente. La cantidad $\frac{\|b - Ax'\|}{\|b\|} = \frac{\|r\|}{\|b\|}$ se llama *residuo relativo* y proporciona una estimación precisa del error relativo.

Dado que $r = b - Ax' = Ax - Ax' = Ae$, se tiene que $e = A^{-1}r$ y por el teorema 3.2.1 se deduce que $\|e\| \leq \|A^{-1}\| \|r\|$ y $\|r\| \leq \|A\| \|e\|$. Por lo tanto,

$$\frac{\|r\|}{\|A\|} \leq \|e\| \leq \|A^{-1}\| \|r\|.$$

Además $\frac{\|b\|}{\|A\|} \leq \|x^*\| \leq \|A^{-1}\| \|b\|$, luego $\frac{1}{\|A\| \|A^{-1}\|} \frac{\|r\|}{\|b\|} \leq \frac{\|e\|}{\|x^*\|} \leq \|A\| \|A^{-1}\| \frac{\|r\|}{\|b\|}$.

Esta última desigualdad relaciona el tamaño del error relativo con el residuo relativo. Si el número $\|A\| \|A^{-1}\|$ está próximo a 1, el error relativo y el residuo relativo serán similares. Si el número es grande, el error relativo podría ser muchas veces más grande que el residuo relativo.

Ejemplo 4.4.1.

Sea $A = \begin{pmatrix} 3 & 3 \\ 4 & 5 \end{pmatrix}$. Entonces $A^{-1} = \frac{1}{3} \begin{pmatrix} 5 & -3 \\ -4 & 3 \end{pmatrix}$, $\|A\|_{\infty} = 9$ y $\|A^{-1}\|_{\infty} = \frac{8}{3}$. Por lo tanto, $\|A\|_{\infty} \|A^{-1}\|_{\infty} = 24$. Así, el error relativo en la solución calculada a un sistema $Ax = b$ podría ser hasta 24 veces mayor que el residuo relativo.

Definición 4.4.1. Sea $A \in M_n(\mathbb{C})$. El *número de condición* de A se define como $\nu(A) = \|A\| \|A^{-1}\|$, donde $\|\cdot\|$ es la norma espectral.

Evidentemente, tenemos que $\nu(A) \geq 1$ ya que $I = AA^{-1}$ y $1 = \|I\| \leq \|A\| \|A^{-1}\|$. Por otro lado, si A es unitaria entonces por el teorema 3.3.1 $\|A\| = [\rho(A^{-1}A)]^{\frac{1}{2}} = [\rho(I)]^{\frac{1}{2}} = 1$. Análogamente, $\|A^{-1}\| = 1$ y por lo tanto $\nu(A) = 1$ si A es unitaria.

Supongamos ahora que tenemos un método iterativo general

$$x^{(m+1)} = Hx^{(m)} + c, \quad m \geq 0,$$

donde $H \in M_n(\mathbb{C})$. Si $(I - H)$ es no singular entonces existe una única solución de $(I - H)x = c$, y si como antes los vectores de error $e^{(m)}$ para este método iterativo general están definidos como $e^{(m)} = x^{(m)} - x$, entonces $e^{(m)} = H^m e^{(0)}$ para $m \geq 0$. Si $e^{(0)}$ es un vector no nulo entonces

$$\frac{\|e^{(m)}\|}{\|e^{(0)}\|} \leq \|H^m\| \quad \text{para } m \geq 0,$$

por lo que $\|H^m\|$ nos da una estimación de cota superior para el cociente $\frac{\|e^{(m)}\|}{\|e^{(0)}\|}$ de las normas euclidianas de $e^{(m)}$ y $e^{(0)}$ para m iteraciones. Recalquemos aquí el hecho de que consideramos tales razones para toda $m \geq 0$, como $e^{(0)}$ es un error inicial arbitrario no nulo entonces $\|H^m\|$ sirve como una base de comparación para los diferentes métodos iterativos.

Por otra parte, puesto que por el teorema 3.3.2 $\rho(H) \leq \|H\|$, tenemos que para $\varepsilon \geq 0$, $\rho(H) + \varepsilon = \|H\|$, de donde

$$\|e^{(m)}\| \leq \|H^m\| \|e^{(0)}\| \leq \|H\|^m \|e^{(0)}\| = [\rho(H) + \varepsilon]^m \|e^{(0)}\|.$$

Entonces $[\rho(H) + \varepsilon]^m$ es el factor de reducción de la magnitud del error, que se puede tomar aproximadamente como $[\rho(H)]^m$. Por ejemplo, supongamos que se desea que la magnitud del error sea menor que 10^{-s} . El número de iteraciones que se necesitará realizar en este caso será igual al menor valor de m para el cual se obtiene que $[\rho(H)]^m \leq 10^{-s}$. Tomando logaritmos en esta última expresión y exigiendo que $0 < \rho(H) < 1$, tendremos que $\log[\rho(H)]^m \leq \log 10^{-s}$, de donde se desprende que $m \log \rho(H) \leq -s$ y de aquí,

$$m \geq \frac{s}{-\log \rho(H)}.$$

Por consiguiente, el número de iteraciones m que se requiere para reducir el error inicial en un factor 10^{-s} es inversamente proporcional a $-\log \rho(H)$.

Definición 4.4.2. Sea $A = M - N \in M_n(\mathbb{C})$ con A y M matrices no singulares y $H = M^{-1}N$ la matriz de iteración. Dada una norma matricial $\|\cdot\|$, $\|H^m\|$ es llamado *el factor de convergencia* para m pasos y $R_m = \frac{1}{m} \ln \|H^m\|$ es llamado *el factor de convergencia promedio* (por paso para m pasos) para esta norma.

Definición 4.4.3. Sea $A \in M_n(\mathbb{C})$. Si para algún entero positivo m , $\|A^m\| < 1$, entonces

$$R(A^m) := -\ln[(\|A^m\|)^{\frac{1}{m}}] = -\frac{\ln \|A^m\|}{m}$$

es la *velocidad promedio de convergencia* para m iteraciones de la matriz A .

En este sentido, si $A, B \in M_n(\mathbb{C})$ y $R(A^m) < R(B^m)$, entonces B es iterativamente más rápido para m iteraciones que A . Puede resultar que para un vector particular y , $\|B^m y\| < \|A^m y\|$ y sin embargo, $\|A^m\| < \|B^m\|$ de donde, por la definición 4.4.3, A es iterativamente más rápido para m iteraciones que B .

En términos de cálculos efectivos, la importancia de la velocidad promedio de convergencia $R(H^m)$ de la matriz de iteración H es la siguiente. En virtud de la definición 4.4.2

la cantidad $\sigma := \left(\frac{\|e^{(m)}\|}{\|e^{(0)}\|}\right)^{\frac{1}{m}}$ es el *factor de reducción promedio* por iteración para las normas de error sucesivo. Si $\|H^m\| < 1$, entonces por la definición 4.4.3

$$\sigma := \left(\frac{\|e^{(m)}\|}{\|e^{(0)}\|}\right)^{\frac{1}{m}} \leq (\|H^m\|)^{\frac{1}{m}} = e^{-R(H^m)}$$

donde e es la base del logaritmo natural. Definiendo $N_m := (R(H^m))^{-1}$ y teniendo en cuenta que $\sigma \leq e^{-R(H^m)}$, se sigue que

$$\sigma^{N_m} \leq \frac{1}{e},$$

así que N_m es una estimación bruta del número de iteraciones requeridas para reducir la norma del vector inicial de error por un factor e .

Hagamos hincapié en el hecho de que $\|H^m\|$ da una medida de la cantidad por la cual la norma del error es reducida después de m iteraciones. Usualmente, se itera hasta que $\|e^{(m)}\|$ es reducida a una fracción δ de $\|e^{(0)}\|$, esta reducción se puede lograr eligiendo m tal que

$$\|e^{(m)}\| \leq \delta \|e^{(0)}\|.$$

Podemos satisfacer esta desigualdad si elegimos m lo suficientemente grande para que $\|H^m\| \leq \delta$. Así, para toda m lo suficientemente grande tal que $\|H^m\| < 1$, la última desigualdad es equivalente a que

$$m \geq \frac{-\ln \delta}{\left(-\frac{1}{m} \ln \|H^m\|\right)}.$$

Así, el mínimo número de iteraciones es inversamente proporcional a $R(H^m)$.

Retomemos ahora en la discusión de la comparación de métodos iterativos el radio espectral de la matriz de iteración H .

Si A_1 y A_2 son ambas Hermitianas, entonces por el corolario 3.3.1

$$\|A_i^m\| = (\rho(A_i))^m \quad \text{para } i = 1, 2 \text{ y } m \geq 1,$$

y así, si $\rho(A_1) < \rho(A_2) < 1$, entonces $\|A_1^m\| < \|A_2^m\|$ para todo $m \geq 1$.

Desafortunadamente, aunque sabemos que $\|A^m\| \rightarrow 0$ cuando $m \rightarrow \infty$ para cualquier matriz convergente, la norma espectral $\|A^m\|$ puede comportarse erráticamente. En términos de la definición 4.4.3, es enteramente posible para una matriz A ser iterativamente más rápida que una matriz B para m_1 iteraciones, pero iterativamente más lenta para $m_2 \neq m_1$ iteraciones. Para ilustrar esto, consideremos el siguiente ejemplo.

Ejemplo 4.4.2.

Sean $A, B \in M_2(\mathbb{R})$ dos matrices convergentes tales que

$$A = \begin{bmatrix} \alpha & 4 \\ 0 & \alpha \end{bmatrix}, \quad B = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix}, \quad 0 < \alpha < \beta < 1.$$

Se puede verificar rápidamente que

$$A^m = \begin{bmatrix} \alpha^m & 4m\alpha^{m-1} \\ 0 & \alpha^m \end{bmatrix} \quad \text{y} \quad B^m = \begin{bmatrix} \alpha^m & 0 \\ 0 & \beta^m \end{bmatrix}.$$

Recordando que $\|A^m\| = \rho[(A^m)^t A^m]^{\frac{1}{2}}$,

$$\|A^m\| = \left\{ \alpha^{2m} + 8m^2 \alpha^{2m-2} \left(1 + \sqrt{1 + \frac{\alpha^2}{4m^2}}\right) \right\}^{\frac{1}{2}}$$

y

$$\|B^m\| = \beta^m.$$

Así, para α suficientemente cerca de 1, podemos ver que las normas $\|A^m\|$ están inicialmente incrementándose para $m \geq 1$, y que $\|A^m\| > \|B^m\|$ para valores pequeños de m . Por otro lado, es claro que $\|A^m\| < \|B^m\|$ cuando $m \rightarrow \infty$.

Para potencias grandes de la matriz A , hay disponible información más precisa acerca de $\|A^m\|$.

Convengamos que $g(m) \sim h(m)$ cuando $m \rightarrow \infty$, significa que $\frac{g(m)}{h(m)} \rightarrow 1$ cuando $m \rightarrow \infty$. Además, por $g(m) = o\left(\frac{1}{m}\right)$ cuando $m \rightarrow \infty$ se entenderá que $|mg(m)| \leq \sigma$ para toda m suficientemente grande, donde σ es una constante positiva.

Lema 4.4.1. Sea $J_{\lambda,p} \in M_p(\mathbb{C})$ la siguiente matriz de Jordan

$$J_{\lambda,p} = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ 0 & 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \lambda \end{pmatrix} = \lambda I + E$$

donde $\lambda \neq 0$. Entonces

$$\|J_{\lambda,p}^m\| \sim C_m^{p-1} [\rho(J)]^{m-(p-1)} \text{ cuando } m \rightarrow \infty.$$

Demostración

Por el lema 2.2.2

$$J_{\lambda,p}^m = \begin{pmatrix} \lambda^m & C_m^1 \lambda^{m-1} & C_m^2 \lambda^{m-2} & \dots & \dots & C_m^{m-(p-1)} \lambda^{m-(p-1)} \\ 0 & \lambda^m & C_m^1 \lambda^{m-1} & \dots & \dots & C_m^{m-(p-2)} \lambda^{m-(p-2)} \\ 0 & 0 & \lambda^m & C_m^1 \lambda^{m-1} & \dots & C_m^{m-(p-3)} \lambda^{m-(p-3)} \\ \vdots & \vdots & 0 & \ddots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \lambda^m & C_m^1 \lambda^{m-1} \\ 0 & 0 & 0 & 0 & 0 & \lambda^m \end{pmatrix}$$

para $m \geq p-1$.

En virtud de que $|\lambda| = \rho(J) > 0$ definimos

$$K_m = \frac{J_{\lambda,p}^m}{C_m^{p-1} \lambda^{m-(p-1)}},$$

entonces $K_m = (k_{ij}^{(m)}) \in M_p(\mathbb{C})$ es tal que para todo $1 \leq i, j \leq p$, $k_{ij}^{(m)} = 0(\frac{1}{m})$ cuando $m \rightarrow \infty$, con la sola excepción de que la entrada $k_{1p}^{(m)}$ es 1. Formando $K_m^* K_m = (l_{ij}^{(m)})$, es claro que todas las entradas de este producto son $0(\frac{1}{m})$, con la excepción de la entrada $l_{pp}^{(m)} = 1 + 0(\frac{1}{m})$. Por el lema 3.6.4 los valores propios de una matriz son funciones continuas de las entradas de una matriz, de aquí se sigue que $\|K_m\| \rightarrow 1$ cuando $m \rightarrow \infty$. \square

Teorema 4.4.1. Sea $A \in M_n(\mathbb{C})$ tal que $\rho(A) > 0$ entonces

$$\|A^m\| \sim \Lambda C_m^{p-1} [\rho(A)]^{m-(p-1)} \quad \text{cuando } m \rightarrow \infty,$$

donde p es el orden del bloque de Jordan más grande en la forma canónica de Jordan de A asociado con un valor propio de módulo igual a $\rho(A)$ y Λ es una constante positiva.

Demostración

Sea $S \in M_n(\mathbb{C})$ una matriz no singular tal que $A = SJS^{-1}$, donde J es una forma canónica de Jordan de A . Podemos suponer que la matriz de Jordan J es tal que sus submatrices diagonales J_l , $1 \leq l \leq r$, están arregladas en orden no decreciente de sus radios espectrales y la submatriz diagonal final J_r es tal que $|\lambda| = \rho(J_r) = \rho(A)$ y el orden de J_r no es excedido por el orden de cualquier otra submatriz J_l con $\rho(J_l) = \rho(A)$. Si el orden de J_r es p , entonces la matriz

$$\frac{J^m}{C_m^{p-1} [\rho(A)]^{m-(p-1)}}$$

tiende a la matriz $M \in M_n(\mathbb{C})$ para la cual todas sus entradas son cero excepto para ciertas entradas de módulo uno tales como la situada en el $(n - p + 1)$ -ésimo renglón y la n -ésima columna. Precisamente, el número total de tales entradas de módulo 1 de M es igual al número de las submatrices diagonales J_l para las cuales $\rho(J_l) = \rho(A)$ y el orden de J_l es igual al orden de J_r . Así

$$\left\| \frac{A^m}{C_m^{p-1} [\rho(A)]^{m-(p-1)}} \right\| \rightarrow \|SMS^{-1}\| \equiv \Lambda \quad \text{cuando } m \rightarrow \infty,$$

lo cual completa la demostración. \square

La constante $\Lambda = \|SMS^{-1}\|$ puede ser estimada de la siguiente manera:

Dado que $\nu(S) \geq 1$, $A^m = SJ^mS^{-1}$ y $J^m = S^{-1}A^mS$, entonces se sigue de la definición de $\nu(S)$ y el teorema 3.2.1 d) que

$$\frac{\|J^m\|}{\nu(S)} \leq \|A^m\| \leq \nu(S) \|J^m\|,$$

y así

$$\frac{1}{\nu(S)} \leq \Lambda \leq \nu(S).$$

Para ilustrar este último resultado, consideremos el siguiente ejemplo.

Ejemplo 4.4.3.

Sea $A = \begin{pmatrix} \alpha & 4 \\ 0 & \alpha \end{pmatrix} \in M_2(\mathbb{R})$. En este caso $p = 2$, $\alpha = \rho(A)$ y se puede verificar que

$S = \begin{pmatrix} 4 & 0 \\ 0 & 1 \end{pmatrix}$, $J = \begin{pmatrix} \alpha & 1 \\ 0 & \alpha \end{pmatrix}$ son tales que $A = SJS^{-1}$. Pero como $SM S^{-1} = \begin{pmatrix} 0 & 4 \\ 0 & 0 \end{pmatrix}$

donde $M = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, entonces la norma espectral de esta matriz es 4 y así $\Lambda = 4$. Como $\|S\| = 4$, $\|S^{-1}\| = 1$ entonces $\nu(S) = 4$.

Regresando al tema de velocidades de convergencia, vemos por el lema 4.4.1 que si una matriz A es convergente, entonces la norma espectral de todas las potencias lo suficientemente grandes de la matriz son monótonas decrecientes. En efecto, tomando en consideración la afirmación del teorema 4.4.1 podemos naturalmente probar el siguiente resultado.

Teorema 4.4.2. Sea $A \in M_n(\mathbb{C})$ una matriz convergente. Para toda m suficientemente grande, la velocidad promedio de convergencia para m iteraciones $R(A^m)$ satisface

$$\lim_{m \rightarrow \infty} R(A^m) = -\ln \rho(A)$$

Demostración

Por el resultado del teorema 4.4.1 tomando logaritmo en base e , obtenemos

$$R(A^m) = -\frac{\ln \|A^m\|}{m} \sim -\frac{\ln \Lambda}{m} - \frac{\ln C_m^{p-1}}{m} - C_m^{m-(p-1)} \ln \rho(A).$$

Dado que $\frac{\ln \Lambda}{m} \rightarrow 0$, $\frac{\ln C_m^{p-1}}{m} \rightarrow 0$ y $C_m^{m-p+1} \rightarrow 1$ cuando $m \rightarrow \infty$, se obtiene la conclusión del teorema. \square

Definición 4.4.4. Sea $A \in M_n(\mathbb{C})$. Entonces $R_\infty(A) := \lim_{m \rightarrow \infty} R(A^m) = -\ln \rho(A)$ es la *velocidad asintótica promedio de convergencia* o *velocidad asintótica de convergencia*.

En virtud del corolario 3.3.1 y el teorema 3.3.2, $\|A^m\| \geq [\rho(A)]^m$ para todo $m \geq 1$. Así, tenemos inmediatamente.

Corolario 4.4.1. Si $A \in M_n(\mathbb{C})$ es una matriz convergente arbitraria entonces

$$R_\infty(A) \geq R(A^m)$$

para cualquier entero positivo m para el cual $\|A^m\| < 1$.

Debemos decir que la velocidad asintótica de convergencia $R_\infty(A)$ para una matriz convergente A es simplemente la medida de rapidez de convergencia de una matriz de uso más común. Sin embargo, su uso indiscriminado o con poco cuidado puede darnos información engañosa.

Ejemplo 4.4.4.

Sea $A = \begin{pmatrix} \alpha & 4 \\ 0 & \alpha \end{pmatrix} \in M_2(\mathbb{R})$ con $0 < \alpha < 1$. Si $\alpha = 0.99$ entonces A es convergente y $R_\infty(A) = 0.01005$, de donde su recíproco $N_\infty = 99.50$. Esto indicaría que aproximadamente sólo 100 iteraciones serían suficientes para reducir la norma euclidiana de cualquier vector inicial de error por un factor de e . Sin embargo, como hemos visto para este ejemplo particular, $\|A^m\|$ inicialmente se incrementa y $\|A^m\| \geq 1$ para todo $m < 805$. Con el objeto de tener que $\|A^m\| \leq \frac{1}{e}$ es necesario que $m \geq 918$.

4.5 Un Teorema de Comparación.

Con la definición de velocidad asintótica de convergencia, compararemos ahora las velocidades asintóticas de convergencia, o equivalentemente, el radio espectral de la matriz de Jacobi B y la matriz de Gauss-Seidel \mathcal{L} asociadas a los respectivos métodos iterativos. Esta comparación fundamental basada en el Teorema de Perron-Frobenius es debida a Stein-Rosenberg(1948).

Teorema 4.5.1. (Stein-Rosenberg).

Sean A una L-matriz, $B = L + U$ la matriz de Jacobi y $\mathcal{L} = (I - L)^{-1}U$ la matriz de Gauss-Seidel. Entonces:

a) $\rho(B) < 1$ si y sólo si $\rho(\mathcal{L}) < 1$.

b) $\rho(B) < 1$ (y $\rho(\mathcal{L}) < 1$) si y sólo si A es una M-matriz; si $\rho(B) < 1$ entonces $\rho(\mathcal{L}) \leq \rho(B)$.

c) Si $\rho(B) \geq 1$ y $\rho(\mathcal{L}) \geq 1$ entonces $\rho(\mathcal{L}) \geq \rho(B) \geq 1$.

Demostración

Sea A una L-matriz. Dado que L es una matriz estrictamente triangular inferior, $L^n = O$ y entonces tenemos que

$$(I - L)^{-1} = I + L + L^2 + \dots + L^{n-1} \geq O,$$

de donde $\mathcal{L} = (I - L)^{-1}U \geq O$.

Ahora, sea $\bar{\lambda} = \rho(\mathcal{L})$ y $\bar{\mu} = \rho(B)$. Por el teorema de Perron-Frobenius $\bar{\lambda}$ es un valor propio de \mathcal{L} , luego $\mathcal{L}w = \bar{\lambda}w$ para algún $w \neq 0$. De aquí se sigue que

$$(\bar{\lambda}L + U)w = \bar{\lambda}w,$$

por lo que $\bar{\lambda}$ es un valor propio de $\bar{\lambda}L + U$ y por lo tanto tenemos que

$$\bar{\lambda} \leq \rho(\bar{\lambda}L + U).$$

Si $\bar{\lambda} \leq 1$ entonces, por el lema 3.6.3., $\rho(\bar{\lambda}L + U) \leq \rho(L + U) = \bar{\mu}$, esto es, $\bar{\lambda} \leq \bar{\mu}$.

Por otro lado, si $\bar{\lambda} \geq 1$ entonces $\bar{\lambda} \leq \rho(\bar{\lambda}L + U) \leq \rho(\bar{\lambda}L + \bar{\lambda}U) = \bar{\lambda}\bar{\mu}$, de donde $\bar{\mu} \geq 1$.

Así, hemos mostrado que:

(i) Si $\bar{\lambda} \leq 1$ entonces $\bar{\lambda} \leq \bar{\mu}$.

(ii) Si $\bar{\lambda} \geq 1$ entonces $\bar{\mu} \geq 1$.

Esto último implica que:

(iii) Si $\bar{\mu} < 1$ entonces $\bar{\lambda} < 1$.

Dado que $B \geq O$, por el teorema de Perron-Frobenius $\bar{\mu}$ es un valor propio de B , es decir, para algún $v \neq 0$ se tiene que $Bv = \bar{\mu}v$. Entonces, tomando $Q = (I - \alpha L)^{-1}U$ con $\alpha = \frac{1}{\bar{\mu}}$, tenemos que $Qv = \bar{\mu}v$. Por lo tanto, $\bar{\mu} \leq \rho(Q)$.

Pero si $\bar{\mu} \geq 1$, entonces como $\alpha \leq 1$, tenemos que

$$(I - \alpha L)^{-1} = I + \alpha L + \alpha^2 L^2 + \dots + \alpha^{n-1} L^{n-1} \leq I + L + L^2 + \dots + L^{n-1} = (I - L)^{-1}.$$

Se sigue que $Q \leq \mathcal{L}$, es decir,

$$\bar{\mu} \leq \rho(Q) \leq \rho(\mathcal{L}) = \bar{\lambda}.$$

Así, hemos mostrado que:

(iv) Si $\bar{\mu} \geq 1$ entonces $\bar{\lambda} \geq \bar{\mu} \geq 1$.

Esto implica que:

(v) Si $\bar{\lambda} < 1$ entonces $\bar{\mu} < 1$.

Por (iii) y (v), tenemos *a*). Por (i) y el teorema 4.1.1., tenemos *b*). Por (iv), tenemos *c*). Luego la demostración del teorema está completa. \square

Así, la matriz de Jacobi B y la matriz de Gauss-Seidel \mathcal{L} , son ambas convergentes o ambas divergentes. Como una consecuencia inmediata tenemos el siguiente resultado.

Corolario 4.5.1. Si la matriz de Jacobi B es tal que $0 < \rho(B) < 1$, entonces $R_\infty(\mathcal{L}) > R_\infty(B)$.

Así, el resultado de Stein-Rosenberg nos da nuestro primer teorema de comparación para los diferentes métodos iterativos. Interpretado de una manera más práctica, el método iterativo de Gauss-Seidel no sólo es computacionalmente más conveniente por cuestión de almacenamiento de datos que el método iterativo de Jacobi, sino que además es asintóticamente más rápido cuando la matriz de Jacobi B es no-negativa con $0 < \rho(B) < 1$.

Regresando al ejemplo 4.3.1, la matriz de Jacobi convergente

$$B = \frac{1}{4} \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \end{pmatrix}$$

ciertamente satisface la hipótesis del corolario 4.5.1, y así la velocidad asintótica de convergencia de la matriz de iteración de Gauss-Seidel es más grande que la de la matriz de iteración de Jacobi como es sugerido anteriormente.

Sería de interés comparar ahora la velocidad promedio de convergencia del método iterativo de Jacobi para m iteraciones con la del método iterativo de Gauss-Seidel, para cada $m \geq 0$, en el caso de que la matriz de Jacobi B es no-negativa y $0 < \rho(B) < 1$. Si examinamos otra vez el ejemplo de la sección anterior, vemos que la matriz B es real y simétrica, así que por el corolario 3.3.1

$$\|B^m\| = \rho(B)^m = \frac{1}{2^m}.$$

Por otro lado, por cálculo directo

$$\mathcal{L}^m = \frac{1}{2(4^m)} \begin{pmatrix} 0 & 0 & 2 & 2 \\ 0 & 0 & 2 & 2 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \text{ para } m \geq 1,$$

por lo que

$$(\mathcal{L}^m)^* (\mathcal{L}^m) = \frac{1}{4^{2m+1}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 10 & 10 \\ 0 & 0 & 10 & 10 \end{pmatrix} \text{ para } m \geq 1.$$

Luego, $\|\mathcal{L}^m\| = \frac{\sqrt{5}}{4^m}$ para $m \geq 1$ y vemos que $\|\mathcal{L}^m\| \geq \|B^m\|$ para $m = 1$. De hecho si elegimos nuestro vector de error inicial como

$$e^{(0)} = \sqrt{2} \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \end{pmatrix},$$

entonces $\|e^{(0)}\| = 2$, $\|Be^{(0)}\| = 1$ y $\|\mathcal{L}e^{(0)}\| = \frac{\sqrt{5}}{2} > 1 = \|Be^{(0)}\|$, de donde se deriva la sorprendente conclusión que para exactamente una iteración, el método de Jacobi es iterativamente más rápido que el método iterativo de Gauss-Seidel. Este simple contraejemplo nos muestra que no podemos extender la desigualdad del corolario 4.5.1 del Teorema de Stein-Rosenberg de velocidades asintóticas de convergencia a velocidades promedio de convergencia para todo $m \geq 1$. Enfatizamos que en el caso no-negativo hemos restringido nuestra comparación de los métodos iterativos de Jacobi y Gauss-Seidel a matrices irreducibles no-negativas en virtud de que el propósito era exhibir a la teoría de Perron-Frobenius como base de la comparación. Desafortunadamente, para el caso en que la matriz de Jacobi no es no-negativa, no es posible probar que las desigualdades relativas al radio espectral de las matrices inducidas por los dos métodos iterativos en cuestión se dan en el caso general. Específicamente, se pueden dar contraejemplos donde un método iterativo es convergente mientras que el otro es divergente, así que en el caso general ninguno de los dos métodos puede ser universalmente favorecido.

Ejemplo 4.5.1.

Sea $A = \begin{pmatrix} 1 & 2 & -2 \\ 1 & 1 & 1 \\ 2 & 2 & 1 \end{pmatrix}$ entonces $B = \begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix}$, aquí $\rho(B) < 1$ mientras

que $\rho(\mathcal{L}) > 1$.

Ejemplo 4.5.2.

Sea $A = \begin{pmatrix} 2 & -1 & 1 \\ 2 & 2 & 2 \\ -1 & -1 & 2 \end{pmatrix}$ entonces $B = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 0 & -2 \\ 1 & 1 & 0 \end{pmatrix}$, aquí $\rho(B) > 1$ mientras

que $\rho(\mathcal{L}) < 1$.

5 Apéndice.

Programas en MATLAB de los algoritmos numéricos subyacentes a los métodos iterativos de Jacobi y Gauss-Seidel.

Programa 1 (Método iterativo de Jacobi). Resolución de un sistema lineal $Ax = b$ mediante la generación de una sucesión $\{x^{(k)}\}$ que converge a la solución, a partir de un punto inicial $x^{(0)}$. Recordemos que una condición suficiente para que el método sea aplicable es que A sea estrictamente diagonal dominante.

```
function X=jacobi(A,B,P,delta,max1)
%Datos
%      A es una matriz invertible de orden  $n \times n$ .
%      B es una matriz de orden  $n \times 1$ .
%      P es una matriz de orden  $n \times 1$ ; el punto inicial  $x^{(0)}$ .
%      delta es la tolerancia para P.
%      max1 es el número máximo de iteraciones.
%Resultados
%      X es una matriz de orden  $n \times 1$  :
%      la aproximación a la solución de  $AX = B$ 
%      generada por el método iterativo de Jacobi.
N=length(B);
for k=1:max1
for j=1:N
    X(j)=(B(j)-A(j,[1:j-1,j+1:N])*P([1:j-1,j+1:N]))/A(j,j);
end
err=abs(norm(X'-P));
relerr=err/(norm(X)+eps);
P=X';
    if(err<delta)\((relerr<delta)
        break
    end
end
X=X';
```

Programa 2 (Método iterativo de Gauss-Seidel). Resolución de un sistema lineal $Ax = b$ mediante la generación de una sucesión $\{x^{(k)}\}$ que converge a la solución, a partir de un punto inicial $x^{(0)}$. Recordemos que una condición suficiente para que el método sea aplicable es que A sea estrictamente diagonal dominante.

```
function X=gseid(A,B,P,delta,max1)
%Datos
%      A es una matriz invertible de orden  $n \times n$ .
%      B es una matriz de orden  $n \times 1$ .
%      P es una matriz de orden  $n \times 1$ ; el punto inicial  $x^{(0)}$ .
```

```

%      delta es la tolerancia para P.
%      max1 es el número máximo de iteraciones.
%Resultados
%      X es una matriz de orden  $n \times 1$  :
%      la aproximación a la solución de  $AX=B$ 
%      generada por el método iterativo de Gauss-Seidel.
N=length(B);
for k=1:max1
for j=1:N
    if j==1
        X(1)=(B(1)-A(1,2:N)*P(2:N))/A(1,1);
    elseif j==N
        X(N)=(B(N)-A(N,1:N-1)*(X(1:N-1)))'/A(N,N);
    else
        % X contiene la aproximación k-ésima
        % y P la (k-1)-ésima
        X(j)=B(j)-A(j,1:j-1)*X(1:j-1)-A(j,j+1:N)*P(j+1:N))/A(j,j);
    end
end
err=abs(norm(X'-P));
relerr=err/(norm(X)+eps);
P=X';
    if(err<delta)\((relerr<delta)
        break
    end
end
X=X';

```

Bibliografía.

- [1] Axelsson, Owe. (2004); “Iterative Solution Methods”, Ed. Cambridge University Press, USA.
- [2] Bartle, Robert G. (1987); “Introducción al Análisis Matemático”. Ed. Limusa, México.
- [3] Berman, Abraham y Plemmons, Robert J. (1979); “Nonnegative Matrices in the Mathematical Sciences”, Ed. Academic Press, USA.
- [4] Franklin, Joel N. (1968); “Matrix Theory”, Ed. Prentice Hall, USA.
- [5] Friedberg, Stephen e Insel, Arnold (1982); “Álgebra Lineal”, Ed. Publicaciones Cultural, México.
- [6] Horn, Roger A. y Johnson, Charles R. (1999); “Matrix Analysis”, Ed. Cambridge University Press, USA.
- [7] Leon, Steven J (1998); “Álgebra Lineal con Aplicaciones”, Ed. C.E.C.S.A., México.
- [8] Marsden, Jerrold E. y Hoffman, Michael J. (1996); “Análisis Básico de Variable Compleja”. Ed. Trillas, México.
- [9] Mathews, John H. y Fink, Kurtis D. (2000); “Métodos Numéricos con Matlab”, Ed. Prentice Hall, Madrid.
- [10] Varga, Richard S. (1962); “Matrix Iterative Analysis”, Ed. Prentice Hall, USA.
- [11] Young, David M. (1971); “Iterative Solution of Large Linear Systems”, Ed. Academic Press, USA.