51

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE CIENCIAS

EL TEOREMA DE PERRON-FROBENIUS Y CADENAS DE MARKOV

T E S I S

QUE PARA OBTENER EL TITULO DE

ACTUARIO

P R E S E N T A N:

AYDÉ GUTIÉRREZ ZERMEÑO
GERARDO ERNESTO MORA RÍOS

DIRECTORA DE TESIS:
DRA. BERTHA MARÍA TOMÉ ARREOLA

2002

VNIVERSIDAD NACIONAL
AVTONOMA DE
MEXICO

M. EN C. ELENA DE OTEYZA DE OTEYZA
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo escrito:

El Teorema de Perron-Frobenius y Cadenas de Markov

realizado por Aydé Gutiérrez Zermeño
Gerardo Ernesto Mora Ríos

con número de cuenta 8632657-7
8634590-9 , quién cubrió los créditos de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis Propietario	Dra. Bertha María Tomé Arreola	Beatriz Tomé
Propietario	Dra. María del Carmen Gómez Laveaga	
Propietario	Dra. Edith Corina Sáenz Valadez	Edith C. Sáenz V.
Suplente	Dr. Juan Morales Rodríguez	Juan Morales R.
Suplente	Mat. Hugo Villaseñor Hernández	

Consejo Departamental de Matemáticas

M. en C. José Antonio Flores Díaz

CONSEJO DEPARTAMENTAL DE MATEMATICAS

El Teorema de Perron-Frobenius y Cadenas de Markov

Aydé Gutiérrez Zermeño
Gerardo Ernesto Mora Ríos

Directora de tesis:
Dra. Bertha Tomé Arreola

A la lic. Sagrario del Socorro Zermeño Santana
gracias mamá por tu amor, tu apoyo y tu ejemplo.

AGZ.

A mis padres

GEMR.

Gracias a . . .

Bertha Tomé
nuestra maestra y amiga
por todo su apoyo y su gran paciencia.

Reyna Caballero
Edith Gutiérrez
Rosalía Lemus
Angélica Loza
Rebeca Mora
por su valiosa ayuda.

Índice General

Introducción

En esta tesis presentamos una prueba del teorema de Perron-Frobenius y su aplicación a las cadenas de Markov. Los elegantes resultados obtenidos por Perron en 1907 para matrices cuadradas con entradas positivas están sintetizados en el teorema de Perron. Este teorema afirma que si A es una de tales matrices entonces su radio espectral $\rho(A)$, que es el máximo de los módulos o valores absolutos de los valores propios de A, es positivo y es de hecho un valor propio algebraicamente simple de A; más aún, es el único valor propio de A de módulo máximo. Además el vector propio que genera el espacio propio asociado a $\rho(A)$ puede escogerse con todas sus coordenadas positivas. En consecuencia, hay un único vector propio positivo asociado a $\rho(A)$ cuya suma de coordenadas es 1; a este vector se le llama vector de Perron de A. Finalmente el teorema de Perron asegura que el límite cuando $m \to \infty$ de la sucesión de matrices $\{\rho(A)^{-1}A^m\}$ existe y es una matriz L de rango 1 cuyas columnas son todas iguales al vector de Perron de A. En consecuencia L es una matriz estocástica por columnas y es esta última parte del teorema de Perron la que tiene mayor relevancia en el estudio de las cadenas de Markov en probabilidad.

La generalización del teorema de Perron a cierta clase de matrices con entradas no negativas se debe a Frobenius (1912). Frobenius introdujo la noción de primitividad: una matriz cuadrada es primitiva si es no negativa, irreducible y tiene un único valor propio de módulo máximo. El teorema de Perron-Frobenius asegura entonces que las afirmaciones del teorema de Perron son válidas para las matrices primitivas.

La prueba del teorema de Perron-Frobenius que presentamos aquí no es la más corta posible, pero en cambio, es muy fácil de entender ya que se requieren sólo los conocimientos adquiridos en los primeros 3 ó 4 semestres de la licenciatura en Matemáticas o Actuaría. De hecho, pensando en que sea de alguna utilidad a aquellos alumnos de la carrera de Actuaría que toman los cursos de Probabilidad sin haber tomado el de Algebra Lineal II, decidimos hacer la tesis autocontenida.

Así, en el capítulo I presentamos los conceptos básicos de espacios con producto interno y matrices unitarias necesarios para comprender el teorema de la triangularización unitaria de Schur. Este importante teorema se presenta al final del capítulo II, después de los conceptos básicos de valores y vectores propios y diagonalizabilidad de una matriz cuadrada. En este capítulo también introducimos la noción de radio espectral y algunas de sus propiedades que utilizaremos después.

En el capítulo III, que es la parte central de esta tesis, presentamos la prueba del teorema de Perron-Frobenius. Para ello necesitamos introducir nuevos conceptos como son los de normas vectoriales y matriciales, el del límite de una sucesión de matrices, el de la gráfica dirigida de una matriz y los de matriz irreducible y matriz primitiva. Cabe aclarar que este tema no forma parte de los cursos de Algebra Lineal que se imparten en las licenciaturas de la Facultad.

Finalmente, en el capítulo IV presentamos la aplicación del teorema de Perron-Frobenius a la resolución de algunos de los problemas que surgen en el estudio de las cadenas de Markov.

Capítulo 1

Preliminares

1.1 Productos internos

A lo largo de este trabajo F denotará el campo de los números reales o el campo de los números complejos.

Definición 1.1.1 *Sea V un espacio vectorial sobre F. Una función* $<\cdot,\cdot>: V \times V \to F$ *es un* **producto interno** *si para todo $x, y, z \in V$ y para todo $c \in F$ se cumplen:*

i) $<x,x> \geq 0$; $<x,x>=0$ *si y sólo si* $x=0$

ii) $<cx+y,z> = c<x,z> + <y,z>$

iii) $<x,y>=\overline{<y,x>}$.

Ejemplo 1. Sea $V = F^n$. Para $x = (x_1, x_2, \ldots, x_n)$, $y = (y_1, y_2, \ldots, y_n)$ en V, $<x,y>=\sum_{i=1}^{n} x_i \overline{y_i}$ es un producto interno llamado canónico u ordinario. Cuando $F = \mathbb{R}$ se le conoce como producto punto y cuando $F = \mathbb{C}$ como producto hermitiano canónico.

Ejemplo 2. Sea $V = \mathcal{M}_n$ el espacio de las matrices $n \times n$ sobre F. Para $A, B \in V$, $<A,B>= tr(B^*A)$, donde B^* es la matriz transpuesta conjugada de B, es un producto interno. Este producto generaliza al del ejemplo anterior ya que $tr(B^*A) = \sum_i (B^*A)_{ii} = \sum_i \sum_j B^*_{ij} A_{ji} = \sum_i \sum_j \overline{B_{ji}} A_{ji}$.

Ejemplo 3. Sea $V = \mathcal{M}_{n\times 1}$ el espacio de las matrices (columnas) $n \times 1$ sobre F y sea Q una matriz invertible en $\mathcal{M}_n$, entonces $<X,Y>= Y^*Q^*QX$ es un producto interno en V. Se identifica la matriz 1×1 del segundo miembro con su único elemento. Cuando Q es la matriz unidad, este producto es esencialmente el mismo del ejemplo 1.

Ejemplo 4. Sea V el espacio vectorial de las funciones continuas de valor complejo en el intervalo $[0,1]$, entonces $<f,g>= \int_0^1 f(t)\overline{g(t)}dt$ es un producto interno en V.

Ejemplo 5. Si $< \cdot, \cdot >$ es un producto interno en un espacio vectorial V y $r > 0$, podemos definir otro producto interno en V mediante la regla $(x, y) = r < x, y >$ para cualesquier $x, y \in V$.

Ejemplo 6. Sean V y W espacios vectoriales sobre F y $T : V \to W$ una transformación lineal inyectiva. Si $< \cdot, \cdot >$ es un producto interno en W, podemos definir un producto interno en V mediante la regla $(x, y) = < T(x), T(y) >$ para cualesquier $x, y \in V$. El ejemplo 3 es un caso particular de esta situación.

Las principales propiedades de un producto interno están contenidas en el siguiente teorema.

Teorema 1.1.1 *Sea V un espacio con producto interno $< \cdot, \cdot >$. Para cualesquier $x, y, z \in V$ y cualquier $c \in F$ se satisfacen:*

a) $< x, cy + z > = \bar{c} < x, y > + < x, z >$

b) $< x, y > = 0$ *para todo* $y \in V$ *si y sólo si* $x = 0$

c) $< x, < x, y > \ y > = | < x, y > |^2$

d) $| < x, y > |^2 \leq < x, x > < y, y >$.

Demostración.

a) $< x, cy + z > = \overline{< cy + z, x >} = \overline{c < y, x > + < z, x >} = \overline{c < y, x >} + \overline{< z, x >} = \bar{c} < x, y > + < x, z >$.

b) Supongamos que $< x, y > = 0$ *para todo* $y \in V$*, entonces* $< x, x > = 0$ *de donde* $x = 0$*. Inversamente, si* $x = 0$, $< x, y > = < cx, y > = c < x, y >$ *para todo* $c \in F$*; entonces* $< x, y > = 0$ *para todo* $y \in V$.

c) $< x, < x, y > \ y > = \overline{< x, y >} < x, y > = | < x, y > |^2$.

d) Si $y = 0$ *la afirmación se cumple por b). Si* $y \neq 0$*, definimos* $p(t) = < x + ty, x + ty >$ *para* $t \in \mathbb{R}$*. Entonces*

$$\begin{aligned} p(t) &= < x, x > + t < x, y > + t < y, x > + t^2 < y, y > \\ &= < y, y > t^2 + 2Re < x, y > t + < x, x > \end{aligned}$$

Este es un polinomio de 2º grado con coeficientes reales y por cómo se definió, no puede tener raíces reales simples; por lo tanto su discriminante no puede ser positivo. Entonces

$$(2Re < x, y >)^2 - 4 < x, x > < y, y > \leq 0,$$

de donde,

$$(Re < x, y >)^2 \leq < x, x > < y, y > .$$

Como ésto es cierto para todo $x, y \in V$, en particular,

$$(Re < x, < x,y > \ y >)^2 \leq < x,x >< < x,y > \ y, \ < x,y > \ y >,$$

de donde por c),

$$(| < x,y > |^2)^2 \leq | < x,y > |^2 < x,x >< y,y > .$$

Si $< x,y >= 0$ la afirmación es obvia. Si $< x,y >\neq 0$, la afirmación se obtiene dividiendo entre $| < x,y > |^2$. $\square$

La desigualdad en d) se llama desigualdad de Cauchy-Schwarz. Observemos que en ella se da la igualdad si y sólo si x y y son linealmente dependientes:
Supongamos que x y y son linealmente dependientes, esto es, $x = \alpha y$ para algún $\alpha \in F$, entonces

$$\begin{aligned} | < x,y > |^2 &= | < \alpha y,y > |^2 =< \alpha y,y > \overline{< \alpha y,y >} =< \alpha y,y >< y,\alpha y > \\ &= \overline{\alpha} < \alpha y,y >< y,y >=< \alpha y,\alpha y >< y,y >=< x,x >< y,y > . \end{aligned}$$

Inversamente, supongamos que la igualdad se cumple. Si $< x,y >= 0$ entonces $< x,x >= 0$ o $< y,y >= 0$, es decir, $x = 0$ o $y = 0$, de donde x y y son linealmente dependientes. Si $< x,y >\neq 0$ entonces

$$\begin{aligned} (Re < x, < x,y > y >)^2 &= (Re(\overline{< x,y >} < x,y >))^2 = (Re| < x,y > |^2)^2 \\ &= (| < x,y > |^2)^2 = | < x,y > |^2 < x,x >< y,y > \\ &= < x,x >< < x,y > y, < x,y > y > . \end{aligned}$$

Haciendo $w = < x,y > y$, tenemos que

$$(Re < x,w >)^2 =< x,x >< w,w >,$$

por lo que

$$(2Re < x,w >)^2 - 4 < x,x >< w,w >= 0.$$

Entonces el polinomio $p(t) =< x + tw, x + tw >$ con $t \in \mathbb{R}$ tiene una raíz real doble c, por lo que $x + cw = 0$, y por lo tanto, $x = -c < x,y > y$, esto es, x y y son linealmente dependientes.

1.2 Longitud o norma de un vector

Sabemos que para todo vector $x \in \mathbb{R}^2$ la longitud o norma de x, denotada por $\| x \|$, está determinada por $\| x \|= \sqrt{x_1^2 + x_2^2}$. Entonces nos resulta natural generalizar el concepto de norma de un vector en un espacio con producto interno arbitrario de acuerdo con la siguiente definición.

Definición 1.2.1 *Sea V un espacio con producto interno. Para $x \in V$ definimos la* **norma (o longitud) de** x *mediante* $\| x \| = \sqrt{< x, x >}$.

Como es de esperarse, las propiedades de la longitud en $\mathbb{R}^2$ se satisfacen en general, como se muestra en el siguiente teorema.

Teorema 1.2.1 *Sea V un espacio con producto interno. Entonces para todo $x, y \in V$ y para todo $c \in F$ se cumplen:*

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

ii) $\| cx \| = |c| \cdot \| x \|$

iii) $\| x + y \| \leq \| x \| + \| y \|$ *(desigualdad del triángulo).*

Demostración.
i) y ii) son inmediatas. Probaremos iii).

$$\begin{aligned} \| x + y \|^2 &= < x + y, x + y > = < x, x > + < y, x > + < x, y > + < y, y > \\ &= \| x \|^2 + 2Re < x, y > + \| y \|^2 \leq \| x \|^2 + 2| < x, y > | + \| y \|^2 \\ &\leq \| x \|^2 + 2 \| x \| \cdot \| y \| + \| y \|^2 = (\| x \| + \| y \|)^2. \quad \square \end{aligned}$$

Definición 1.2.2 *Sea V un espacio con producto interno. Un vector $x \in V$ es un* **vector unitario** *si* $\| x \| = 1$.

Para cualquier vector no nulo x en un espacio con producto interno se tiene que $\frac{x}{\|x\|}$ es un vector unitario.

1.3 Ortogonalidad y ortonormalidad de vectores en un espacio con producto interno

Recordemos que para el producto punto en $\mathbb{R}^2$ ó $\mathbb{R}^3$, $< x, y > = \| x \| \cdot \| y \| \cos\theta$ donde θ es el ángulo ($0 \leq \theta \leq \pi$) entre x y y. Esta ecuación implica la desigualdad de Cauchy-Schwarz ya que $|\cos\theta| \leq 1$. Además x y y son perpendiculares si y sólo si $\cos\theta = 0$, es decir, si y sólo si $< x, y > = 0$. Entonces nos resulta natural generalizar la noción de perpendicularidad a un espacio con producto interno cualquiera.

Definición 1.3.1 *Sean V un espacio con producto interno y x, $y \in V$. Los vectores x y y son* **ortogonales** *si* $< x, y > = 0$. *Un* **subconjunto** S **de** V **es ortogonal** *si cualquier par de elementos distintos de S es ortogonal. Finalmente, un* **subconjunto** S **de** V **es ortonormal** *si S es ortogonal y está formado por vectores unitarios.*

Ejemplo 1. Los vectores (x, y), $(-y, x)$ en $\mathbb{R}^2$ son ortogonales con respecto al producto punto.

Ejemplo 2. La base canónica en F^n es un conjunto ortonormal con respecto al producto interno canónico.

Ejemplo 3. La base canónica en $\mathcal{M}_n$ es un conjunto ortonormal con respecto al producto interno dado en el ejemplo 2.

Teorema 1.3.1 *Sea V un espacio con producto interno y sea S un conjunto ortogonal formado por vectores no nulos en V. Entonces S es linealmente independiente.*
Demostración.
Sean $x_1, x_2, \ldots, x_n$ elementos distintos en S y supóngase que para $a_1, a_2, \ldots, a_n \in F$

$$0 = \sum_{i=1}^{n} a_i x_i.$$

Entonces para cualquier j, $1 \leq j \leq n$,

$$0 =< 0, x_j >=< \sum_{i=1}^{n} a_i x_i, x_j >= \sum_{i=1}^{n} a_i < x_i, x_j >= a_j \parallel x_j \parallel^2$$

puesto que $< x_i, x_j >= 0$ para $i \neq j$. Como $x_j \neq 0$, tenemos que $a_j = 0$. Por lo tanto, S es linealmente independiente. □

Las bases ortonormales desempeñan un papel importante en los espacios con producto interno. El siguiente teorema muestra que a partir de un subconjunto finito linealmente independiente de vectores en un espacio con producto interno, podemos encontrar una base ortonormal del subespacio generado por dicho subconjunto. El proceso para encontrar una base ortogonal del subespacio es conocido como proceso de ortogonalización de Gram-Schmidt. Una vez obtenida una base ortogonal, sólo es necesario normalizar los vectores de dicha base.

Teorema 1.3.2 *Sea V un espacio con producto interno y sean $y_1, y_2, \ldots, y_n$ vectores linealmente independientes cualesquiera de V. Entonces se pueden construir vectores ortogonales $x_1, x_2, \ldots, x_n$ en V tales que para cada $k = 1, \ldots, n$ el conjunto*

$$\{x_1, x_2, \ldots, x_k\}$$

sea una base del subespacio generado por $y_1, y_2, \ldots, y_k$.
Demostración.
Primero, sea $x_1 = y_1$. Los otros vectores están entonces dados inductivamente de la siguiente forma:
Supóngase que $x_1, x_2, \ldots, x_m$ ($1 \leq m \leq n$) hayan sido elegidos de modo que para cada $k = 1, \ldots, m$, $\{x_1, x_2, \ldots, x_k\}$ es una base ortogonal para el subespacio de V generado por $y_1, y_2, \ldots, y_k$. Para construir el siguiente vector x_{m+1}, sea

$$x_{m+1} = y_{m+1} - \sum_{k=1}^{m} \frac{< y_{m+1}, x_k >}{\parallel x_k \parallel^2} x_k \tag{1.1}$$

Entonces $x_{m+1} \neq 0$, ya que en caso contrario y_{m+1} es una combinación lineal de $x_1, x_2, \ldots, x_m$ y por lo tanto una combinación lineal de $y_1, y_2, \ldots, y_m$. Además, si $1 \leq j \leq m$, entonces

$$\begin{aligned} < x_{m+1}, x_j > &= < y_{m+1}, x_j > - \sum_{k=1}^{m} \frac{< y_{m+1}, x_k >}{\| x_k \|^2} < x_k, x_j > \\ &= < y_{m+1}, x_j > - < y_{m+1}, x_j >= 0. \end{aligned}$$

Luego, $\{x_1, x_2, \ldots, x_{m+1}\}$ es un conjunto ortogonal que consta de $m+1$ vectores no nulos en el subespacio generado por $y_1, y_2, \ldots, y_{m+1}$. Por el teorema 1.3.1, $\{x_1, x_2, \ldots, x_{m+1}\}$ es una base para este subespacio. Así los vectores $x_1, x_2, \ldots, x_n$ pueden construirse uno tras otro de acuerdo con (1.1). $\square$

1.4 Matrices unitarias

Definición 1.4.1 *Una matriz $U \in \mathcal{M}_n$ se llama* **unitaria** *si $U^*U = I$. En caso de que $U \in \mathcal{M}_n(\mathbb{R})$, se dice que U es* **ortogonal real.**

Ejemplo 1. La matriz $U = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & -\frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{pmatrix}$ es una matriz unitaria, más aún, es ortogonal real.

Ejemplo 2. Para $\theta \in \mathbb{R}$, la matriz $U = \begin{pmatrix} cos\theta & -sen\theta \\ sen\theta & cos\theta \end{pmatrix}$ es también una matriz ortogonal real.

Ejemplo 3. La matriz $U = \begin{pmatrix} i & 0 \\ 0 & i \end{pmatrix}$ es una matriz unitaria.

Teorema 1.4.1 *Sea $U \in \mathcal{M}_n$. Las siguientes afirmaciones son equivalentes:*

1) U es unitaria.

2) U es invertible y $U^* = U^{-1}$.

3) $UU^* = I$.

4) U^* es unitaria.

5) Las columnas de U forman un conjunto ortonormal.

6) Los renglones de U forman un conjunto ortonormal.

7) Para todo $x \in \mathbb{C}^n$, $\| Ux \| = \| x \|$.

Demostración.
Dado que una matriz que tiene inversa izquierda o derecha es invertible y su inversa es única, 1), 2) *y* 3) *son equivalentes. La equivalencia de estas afirmaciones con* 4) *se sigue de que* $(U^*)^* = U$.
Considerando la multiplicación de matrices y denotando por U^i *la i-ésima columna de* U $(1 \leq i \leq n)$, *el enunciado* $U^*U = I$ *significa que*

$$(U^i)^*U^j = (U^*U)_{ij} = \begin{cases} 0 & si\ j \neq i \\ 1 & si\ j = i \end{cases}$$

Es decir, las columnas de U *forman un conjunto ortonormal respecto al producto interno canónico y por lo tanto* 1) *y* 5) *son equivalentes. Similarmente* 4) *y* 6) *son equivalentes.*
Si $y := Ux$ *y* U *es unitaria entonces* $\| y \|^2 = y^*y = x^*U^*Ux = x^*Ix = x^*x = \| x \|^2$, *esto es,* 1) *implica* 7). *Para el recíproco, consideremos primero el caso* $n = 2$. *Suponiendo* 7) *y tomando* $x = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, *encontramos que* $1 = x^*x = y^*y = x^*U^*Ux = (U^*U)_{11}$. *Similarmente, tomando* $x = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, *concluimos que* $(U^*U)_{22}$ *es también* 1 *y dado que* $(U^*U)^* = U^*U$ *(*U^*U *es hermítica o autoadjunta),* U^*U *debe tener la forma* $\begin{pmatrix} 1 & a \\ \overline{a} & 1 \end{pmatrix}$ *donde* a *es el producto interno de la segunda columna de* U *y la primera columna de* U *y* $\overline{a}$ *es el producto interno de la primera columna de* U *y la segunda columna de* U. *Tomando* $x = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ *obtenemos que* $2 = x^*x = y^*y = x^*U^*Ux = 2 + (a + \overline{a})$ *y tomando* $x = \begin{pmatrix} 1 \\ i \end{pmatrix}$ *obtenemos que* $2 = 2 + i(a - \overline{a})$, *de donde* $a + \overline{a} = 2Rea = 0 = 2iIma = a - \overline{a}$, *por lo que* $a = 0$ *y entonces* $U^*U = I$, *esto es,* U *es unitaria.*
Consideremos ahora el caso $n > 2$. *Sean* $A = U^*U$ *y* $x \in \mathbb{C}^n$ *tal que todas sus componentes distintas de* i *y* j $(i \leq j)$ *son cero. Entonces*

$$\begin{aligned} x^*Ax &= a_{ii}\overline{x_i}x_i + a_{ji}x_i\overline{x_j} + a_{ij}\overline{x_i}x_j + a_{jj}x_i\overline{x_j} \\ &= \begin{pmatrix} \overline{x_i} & \overline{x_j} \end{pmatrix} \begin{pmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{pmatrix} \begin{pmatrix} x_i \\ x_j \end{pmatrix} \end{aligned}$$

lo cual se reduce a la condición x^*U^*Ux *para todo* $x \in \mathbb{C}^2$ *y* $U^*U = \begin{pmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{pmatrix}$.
Por el caso anterior, $\begin{pmatrix} a_{ii} & a_{ij} \\ a_{ji} & a_{jj} \end{pmatrix} = I \in \mathcal{M}_2$. *Dado que* i *y* j *son arbitrarios, concluimos que cada submatriz principal de* 2×2 *de* A *es* $I \in \mathcal{M}_2$. *La única matriz que satisface esta propiedad es* $A = I \in \mathcal{M}_n$ *y dado que el caso* $n = 1$ *es obvio, concluimos que* 7) *implica* 1), *lo cual completa la prueba.* □

Observamos que la inversa de una matriz unitaria y el producto de matrices unitarias son también matrices unitarias. En efecto, $A \in \mathcal{M}_n$ es unitaria si y sólo si $A^{-1} = A^*$.

Así, si A es unitaria entonces $(A^{-1})^{-1} = (A^*)^{-1} = (A^{-1})^*$. Si A y B son unitarias entonces $(AB)^{-1} = B^{-1}A^{-1} = B^*A^* = (AB)^*$.

Definición 1.4.2 *Una matriz $B \in \mathcal{M}_n$ es* **unitariamente equivalente** *a $A \in \mathcal{M}_n$ si existe una matriz unitaria $U \in \mathcal{M}_n$ tal que $B = U^*AU$. Si U es real (y por tanto ortogonal), entonces B es* **ortogonalmente equivalente** *a A.*

Ejemplo. Las matrices $A = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ y $B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & i & 0 \\ 0 & 0 & -i \end{pmatrix}$ son unitariamente equivalentes. En efecto, la matriz $U = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{i}{\sqrt{2}} \\ 0 & \frac{i}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ 1 & 0 & 0 \end{pmatrix}$ es unitaria con adjunta $U^* = \begin{pmatrix} 0 & 0 & 1 \\ \frac{1}{\sqrt{2}} & \frac{-i}{\sqrt{2}} & 0 \\ \frac{-i}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{pmatrix}$ y un simple cálculo muestra que $B = U^*AU$.

Capítulo 2

Diagonalización

2.1 Introducción

Sabemos que todo operador lineal T sobre un espacio vectorial V de dimensión n se puede representar por medio de una matriz $n \times n$. Más aún, si β es una base ordenada de V y $A = [T]_\beta$, entonces el conjunto de todas las matrices que representan a T con respecto a alguna base ordenada de V es el conjunto de todas las matrices similares a A, a saber, el conjunto

$$\{S^{-1}AS \mid S \in \mathcal{M}_n \quad \text{es invertible}\}.$$

Inversamente, si $A \in \mathcal{M}_n$, entonces A induce un operador lineal T_A sobre cualquier espacio vectorial V de dimensión n, en particular, A induce un operador lineal L_A sobre F^n definido por $L_A(x) = Ax$ para todo vector columna $x \in F^n$ y A es la matriz que representa a L_A con respecto a la base canónica de F^n. Así pues, es lo mismo trabajar con operadores que con matrices.

En este capítulo intentamos representar al operador T por una matriz especialmente sencilla. Como los cálculos que involucran a las matrices diagonales son sencillos (por ejemplo, la nulidad, el rango y el determinante de una matriz diagonal son muy fáciles de calcular), lo que buscamos es una base ordenada $\gamma = \{v_1, \ldots, v_n\}$ de V tal que $[T]_\gamma$ sea diagonal, digamos

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

Observamos que $[T]_\gamma = D$ si y sólo si $T(v_k) = \lambda_k v_k$ para $1 \leq k \leq n$. Si β y A son como en el primer párrafo, esto equivale a encontrar una matriz invertible $S \in \mathcal{M}_n$ tal que $S^{-1}AS = D$. De hecho, S debe ser una matriz de cambio de base, a saber, la que transforma coordenadas en γ en coordenadas en β, esto es, $S = [1_V]_\gamma^\beta$. Observamos que las columnas de S forman una base de F^n y que $S^{-1}AS = D$ si y sólo si

$AS = SD$ si y sólo si para $1 \leq j \leq n$ la j-ésima columna de AS es la j-ésima columna de SD, esto es, $AS^j = (AS)^j = (SD)^j = \lambda_j S^j$.

En lo sucesivo trabajaremos con matrices en lugar de operadores.

2.2 Valores y vectores propios

Definición 2.2.1 *Una matriz $A \in \mathcal{M}_n$ es* **diagonalizable** *si es similar a una matriz diagonal.*

Como veremos mas adelante, no toda matriz A es diagonalizable. Luego, estamos interesados en encontrar condiciones necesarias y suficientes para que una matriz A sea diagonalizable.

Definición 2.2.2 *Si $A \in \mathcal{M}_n$, un* **valor propio de A** *es un escalar $\lambda \in F$ para el que existe $0 \neq x \in F^n$ con $Ax = \lambda x$. Si λ es un valor propio de A, entonces cualquier $0 \neq x \in F^n$ tal que $Ax = \lambda x$ se llama un* **vector propio de A asociado al valor propio λ.**

Los valores y vectores propios se llaman también valores y vectores característicos.

Teorema 2.2.1 *Sea $A \in \mathcal{M}_n$. Entonces A es diagonalizable si y sólo si existe una base $\beta = \{x_1, \ldots, x_n\}$ de F^n formada por vectores propios de A. De hecho, si A es similar a la matriz diagonal D, entonces los elementos de la diagonal principal de D son los valores propios de A.*
Demostración.
Si A es diagonalizable, existe una matriz invertible S tal que $D = S^{-1}AS$ es una matriz diagonal. En el último párrafo de la sección anterior vimos que las columnas de S son una base para F^n formada por vectores propios de A con valores propios los elementos de la diagonal principal de D.
Inversamente, si $\beta = \{x_1, \ldots, x_n\}$ es una base de F^n tal que $Ax_i = \lambda_i x_i$ para $1 \leq i \leq n$, sean $S = [x_1, \ldots, x_n]$ la matriz cuyas columnas son $x_1, \ldots, x_n$ y

$$D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}.$$

Entonces S es invertible y $S^{-1}AS = S^{-1}[Ax_1, \ldots, Ax_n] = S^{-1}[\lambda_1 x_1, \ldots, \lambda_n x_n] = S^{-1}SD = D$. Luego A es diagonalizable. $\square$

Definición 2.2.3 *Si $A \in \mathcal{M}_n$ y λ es un valor propio de A, el* **espacio propio de A asociado a λ** *es el conjunto $E_\lambda = \{x \in F^n \mid Ax = \lambda x\}$*

Observamos que si $A \in \mathcal{M}_n$ y λ es un valor propio de A, entonces

$$E_\lambda = \{x \in F^n \mid Ax - \lambda x = 0\} = \{x \in F^n \mid (A - \lambda I)x = 0\} = N(A - \lambda I)$$

por lo que E_λ es efectivamente un subespacio de F^n.

El siguiente resultado proporciona un método para calcular los valores propios de una matriz.

Teorema 2.2.2 *Sean $A \in \mathcal{M}_n$ y $\lambda \in F$. Las siguientes afirmaciones son equivalentes:*

1) λ es valor propio de A.

2) $A - \lambda I$ no es invertible.

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

Demostración.
Por la observación anterior, λ es un valor propio de A si y sólo si $N(A - \lambda I) \neq \{0\}$. Sabemos que esta condición es equivalente a que la matriz $A - \lambda I$ no sea invertible y a que su determinante sea 0. □

Ya que λ es un valor propio de A si y sólo si $\det(A - \lambda I) = 0$, o bien, si y sólo si $\det(\lambda I - A) = 0$, podemos construir la matriz $xI - A$ con elementos polinómicos y considerar el polinomio

$$det(xI - A) = \begin{vmatrix} x - a_{11} & -a_{12} & \cdots & -a_{1n} \\ -a_{21} & x - a_{22} & \cdots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \cdots & x - a_{nn} \end{vmatrix} = x^n + \text{términos de menor grado}$$

Los valores propios de A son los escalares $\lambda \in F$ tales que $\det(\lambda I - A) = 0$, esto es, son las raíces o ceros del polinomio $\det(xI - A)$ en F. Esto muestra que A tiene a lo más n valores propios distintos. Es importante señalar que A puede carecer de valores propios.

Definición 2.2.4 *Si $A \in \mathcal{M}_n$, el polinomio $f_A(x) = det(xI - A)$ se llama* **el polinomio característico de A.**

Lema 2.2.1 *Las matrices similares tienen el mismo polinomio característico.*
Demostración.
Si $B = P^{-1}AP$, entonces

$$\begin{aligned} det(xI - B) &= det(xI - P^{-1}AP) \\ &= det(P^{-1}(xI - A)P) \\ &= detP^{-1} \cdot det(xI - A) \cdot detP \\ &= det(xI - A). \quad \square \end{aligned}$$

Ejemplo 1. Sea $A = \begin{pmatrix} 3 & -5 \\ 1 & -1 \end{pmatrix}$. El polinomio característico de A es $f_A(x) = det(xI - A) = \begin{vmatrix} x-3 & 5 \\ -1 & x+1 \end{vmatrix} = (x-3)(x+1)+5 = x^2 - 2x + 2$ cuyo discriminante es negativo. Luego A no tiene valores propios reales.

Ejemplo 2. Sea $A = \begin{pmatrix} 4 & 6 & 6 \\ 1 & 3 & 2 \\ -1 & -5 & -2 \end{pmatrix}$. El polinomio característico de A es $f_A(x)$ $= \begin{vmatrix} x-4 & -6 & -6 \\ -1 & x-3 & -2 \\ 1 & 5 & x+2 \end{vmatrix} = x^3 - 5x^2 + 8x - 4 = (x-1)(x-2)^2$, con lo que los valores propios de A son 1 y 2.

Ahora, $A - I = \begin{pmatrix} 3 & 6 & 6 \\ 1 & 2 & 2 \\ -1 & -5 & -3 \end{pmatrix}$ tiene rango 2 y nulidad 1, por lo que $E_1 = N(A - I)$ es unidimensional. El vector $\begin{pmatrix} 4 \\ 1 \\ -3 \end{pmatrix}$ es un vector propio asociado a 1, por lo que E_1 está generado por este vector.

Por otro lado, $A - 2I = \begin{pmatrix} 2 & 6 & 6 \\ 1 & 1 & 2 \\ -1 & -5 & -4 \end{pmatrix}$ tiene también rango 2 y nulidad 1, por lo que $E_2 = N(A - 2I)$ es también unidimensional. El vector $\begin{pmatrix} 3 \\ 1 \\ -2 \end{pmatrix}$ es un vector propio asociado a 2, por lo que E_2 está generado por este vector. Observamos que $\begin{pmatrix} 4 \\ 1 \\ -3 \end{pmatrix}$ y $\begin{pmatrix} 3 \\ 1 \\ 2 \end{pmatrix}$ son linealmente independientes. Sin embargo, A no es diagonalizable pues no existe una base de $\mathbb{R}^3$ formada por vectores propios de A.

Ejemplo 3. Sea $A = \begin{pmatrix} 7 & -2 & -4 \\ 3 & 0 & -2 \\ 6 & -2 & -3 \end{pmatrix}$. El polinomio característico de A es $f_A(x)$ $= \begin{vmatrix} x-7 & 2 & 4 \\ -3 & x & 2 \\ -6 & 2 & x+3 \end{vmatrix} = (x-1)^2(x-2)$, con lo que los valores propios de A son 1 y 2. En este caso, $A - I = \begin{pmatrix} 6 & -2 & -4 \\ 3 & -1 & -2 \\ 6 & -2 & -4 \end{pmatrix}$ tiene rango 1 y nulidad 2, por lo que E_1 es bidimensional. Los vectores $\begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}$ y $\begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}$ son vectores propios linealmente independientes asociados a 1, por lo que E_1 está generado por ellos.

Por otro lado, $A-2I = \begin{pmatrix} 5 & -2 & -4 \\ 3 & -2 & -2 \\ 6 & -2 & -5 \end{pmatrix}$ tiene rango 2 y nulidad 1, por lo que E_2 es unidimensional. El vector $\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$ es un vector propio asociado a 2, por lo que E_2 está generado por este vector.

Es fácil ver que $\beta = \left\{ \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix}, \begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix} \right\}$ es una base de $\mathbb{R}^3$ formada por vectores propios de A y por lo tanto A es diagonalizable.

La matriz diagonal similar a A es $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$ y una matriz invertible P tal que $P^{-1}AP = D$ es $P = \begin{pmatrix} 1 & 2 & 2 \\ 3 & 0 & 1 \\ 0 & 3 & 2 \end{pmatrix}$. Observemos que en la diagonal principal de D aparecen los valores propios de A contando su multiplicidad como raices de $f_A(x)$ y que las columnas de P son la base de vectores propios de A.

Definición 2.2.5 *El* **espectro** *de una matriz $A \in \mathcal{M}_n$ es el conjunto de todos los valores propios de A en $\mathbb{C}$.*

Definición 2.2.6 *El* **radio espectral** $\rho(A)$ *de una matriz $A \in \mathcal{M}_n$ es $\rho(A) = max\{|\lambda| : \lambda$ es un valor propio de $A\}$.*

Ejemplo 1. Sea $A = \begin{pmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{pmatrix}$. El polinomio característico de A es $f_A(x) = (x-2)(x+3)(x-1)$, por lo que el espectro de A es el conjunto $\{1, 2, -3\}$ y el radio espectral de A es $\rho(A) = 3$.

Ejemplo 2. Sea $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$. El polinomio característico de A es $f_A(x) = x^2+1 = (x-i)(x+i)$, por lo que el espectro de A es el conjunto $\{i, -i\}$ y el radio espectral de A es $\rho(A) = 1$.

Observamos que si λ es cualquier valor propio de A, entonces $|\lambda| \leq \rho(A)$; más aún, hay al menos un valor propio λ para el cual $|\lambda| = \rho(A)$.

2.3 Diagonalizabilidad

En la sección anterior vimos que no toda matriz es diagonalizable. En esta sección damos un criterio para determinar si una matriz puede ser diagonalizada. Comenzamos con el siguiente resultado.

Teorema 2.3.1 *Sean $A \in \mathcal{M}_n$ y $\lambda_1, \lambda_2, \ldots, \lambda_k$ valores propios distintos de A. Si $x_1, x_2, \ldots, x_k$ son vectores propios de A tales que $Ax_i = \lambda_i x_i$, $1 \leq i \leq k$, entonces $\{x_1, x_2, \ldots, x_k\}$ es linealmente independiente.*
Demostración.
Procedemos por inducción sobre k. Supongamos que $k = 1$. Entonces $x_1 \neq 0$ pues x_1 es un vector propio de A. Luego $\{x_1\}$ es linealmente independiente.
Supongamos válido el teorema para $k-1$ vectores propios, donde $k-1 \geq 1$, y que tenemos k vectores propios $x_1, x_2, \ldots, x_k$ correspondientes a distintos valores propios $\lambda_1, \lambda_2, \ldots, \lambda_k$. Queremos demostrar que $\{x_1, x_2, \ldots, x_k\}$ es linealmente independiente. Supongamos que se tienen $a_1, a_2, \ldots, a_k \in F$ tales que

$$a_1x_1 + a_2x_2 + \ldots + a_kx_k = 0 \tag{2.1}$$

Multiplicando por A ambos lados de la ecuación (2.1) obtenemos

$$a_1Ax_1 + a_2Ax_2 + \ldots + a_kAx_k = a_1\lambda_1x_1 + a_2\lambda_2x_2 + \ldots + a_k\lambda_kx_k = 0 \tag{2.2}$$

Multiplicando ambos lados de la ecuación (2.1) por λ_k obtenemos

$$a_1\lambda_kx_1 + a_2\lambda_kx_2 + \ldots + a_k\lambda_kx_k = 0 \tag{2.3}$$

Restando la ecuación (2.3) de la (2.2) tenemos que

$$a_1(\lambda_1 - \lambda_k)x_1 + a_2(\lambda_2 - \lambda_k)x_2 + \ldots + a_{k-1}(\lambda_{k-1} - \lambda_k)x_{k-1} = 0$$

Por la hipótesis de inducción, $a_1(\lambda_1 - \lambda_k) = a_2(\lambda_2 - \lambda_k) = \ldots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0$ y ya que $\lambda_1, \lambda_2, \ldots, \lambda_k$ son distintos, $a_1 = a_2 = \ldots = a_{k-1} = 0$. Así la ecuación (2.1) se reduce a $a_kx_k = 0$. Como $x_k \neq 0$, $a_k = 0$; por lo tanto, $a_1 = a_2 = \ldots = a_k = 0$. □

Corolario 2.3.1 *Sea $A \in \mathcal{M}_n$. Si A tiene n valores propios distintos entonces A es diagonalizable.*
Demostración.
Supongamos que A tiene n valores propios distintos $\lambda_1, \lambda_2, \ldots, \lambda_n$ y sean $x_1, x_2, \ldots, x_n$ los vectores propios de A tales que $Ax_i = \lambda_i x_i$ para $1 \leq i \leq n$. Por el teorema anterior, $\{x_1, x_2, \ldots, x_n\}$ es linealmente independiente y por lo tanto es una base de F^n formada por vectores propios de A. Entonces A es diagonalizable. □

Ejemplo. Sea $A = \begin{pmatrix} 2 & -1 \\ -4 & 2 \end{pmatrix}$. El polinomio característico de A es $f_A(x) = x(x-4)$ y por lo tanto los valores propios de A son 0 y 4. Por el corolario anterior, A es diagonalizable. En efecto, el vector columna $\begin{pmatrix} 1 \\ 2 \end{pmatrix}$ genera al espacio propio de A asociado a 0, el vector columna $\begin{pmatrix} 1 \\ -2 \end{pmatrix}$ genera al espacio propio de A asociado a 4 y $\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ -2 \end{pmatrix} \right\}$ es la base que diagonaliza a A.

La condición suficiente para la diagonalizabilidad dada en el corolario anterior no es necesaria, como lo muestra el ejemplo 3 de la sección anterior.

Supongamos que $A \in \mathcal{M}_n$ es diagonalizable y que $\lambda_1, \lambda_2, \ldots, \lambda_k$ con $k < n$ son los valores propios distintos de A. Entonces existe una matriz invertible S (cuyas columnas son una base de F^n formada por vectores propios de A) tal que

$$\mathbf{S^{-1}AS = D} = \begin{pmatrix} \lambda_1 I_1 & & \\ & \ddots & \\ & & \lambda_k I_k \end{pmatrix}$$

donde I_j es la matriz identidad de $m_j \times m_j$. Observamos que:

1) Por el lema 2.2.1, $f_A(x) = f_D(x) = det(xI - D) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k}$.

2) Para $1 \leq j \leq k$, $dimE_{\lambda_j} = dimN(A - \lambda_j I) = dimN(D - \lambda_j I) = m_j$, ya que las matrices similares tienen la misma nulidad y el mismo rango y la matriz $D - \lambda_j I$ tiene m_j ceros en la diagonal principal.

Esta relación entre la dimensión de cada espacio propio y la multiplicidad del valor propio correspondiente como raíz del polinomio característico nos proporcionará un criterio para determinar cuándo una matriz dada es diagonalizable.

Definición 2.3.1 *Sea λ un valor propio de una matriz $A \in \mathcal{M}_n$ cuyo polinomio característico es $f(x)$. La* **multiplicidad algebraica de** *λ es el mayor entero positivo m para el que $(x - \lambda)^m$ es un factor de $f(x)$.*

Ejemplo. Para la matriz $A = \begin{pmatrix} 4 & 6 & 6 \\ 1 & 3 & 2 \\ -1 & -5 & -2 \end{pmatrix}$ vimos en la sección anterior que su polinomio característico es $f_A(x) = (x-1)(x-2)^2$. En este caso los valores propios $\lambda_1 = 1$ y $\lambda_2 = 2$ tienen multiplicidades algebraicas 1 y 2 respectivamente.

Definición 2.3.2 *Sea λ un valor propio de una matriz $A \in \mathcal{M}_n$. La* **multiplicidad geométrica de** *λ es la dimensión del espacio propio E_λ asociado a λ.*

Ejemplo. Para la matriz $A = \begin{pmatrix} 4 & 6 & 6 \\ 1 & 3 & 2 \\ -1 & -5 & -2 \end{pmatrix}$ vimos en la sección anterior que los espacios propios E_1 y E_2 asociados a los valores propios $\lambda_1 = 1$ y $\lambda_2 = 2$ son unidimensionales, por lo que ambos valores propios tienen multiplicidad geométrica 1.

Teorema 2.3.2 *Sean $A \in \mathcal{M}_n$ y λ un valor propio de A de multiplicidad algebraica m, entonces $1 \leq dim E_\lambda \leq m$.*

Demostración.

Sea $\{x_1, x_2, \ldots, x_p\}$ una base ordenada de E_λ y sea $\beta = \{x_1, x_2, \ldots, x_p, x_{p+1}, \ldots, x_n\}$ una extensión a una base ordenada de F^n. Consideremos el operador L_A definido sobre F^n por $L_A(x) = Ax$ y sea $B = [L_A]_\beta$. Ya que $x_1, x_2, \ldots, x_p$ son vectores propios de A correspondientes a λ

$$B = \begin{pmatrix} B_1 & B_2 \\ 0 & B_3 \end{pmatrix},$$

donde $B_1 = \lambda I_p$ y 0 es la matriz cero.

Como A y B son similares, por el lema 2.2.1 tenemos que

$$\begin{aligned} f_A(x) &= det(xI - A) \\ &= det(xI - B) \\ &= det \begin{pmatrix} xI_p - B_1 & -B_2 \\ 0 & xI_{n-p} - B_3 \end{pmatrix} \\ &= det(xI_p - B_1) \cdot det(xI_{n-p} - B_3). \end{aligned}$$

Sea $g(x) = det(xI_{n-p} - B_3)$ el polinomio característico de B_3. Claramente, $det(xI_p - B_1) = (x - \lambda)^p$. Por lo tanto $f_A(x) = (x - \lambda)^p g(x)$, de modo que la multiplicidad algebraica m de λ es al menos p. Pero $dim E_\lambda = p$, por lo que $1 \leq dim E_\lambda \leq m$. □

Lema 2.3.1 *Sean $A \in \mathcal{M}_n$, $\lambda_1, \lambda_2, \ldots, \lambda_k$ los valores propios distintos de A y $W = E_{\lambda_1} + E_{\lambda_2} + \cdots + E_{\lambda_k}$. Entonces $dim\ W = \sum_{i=1}^{k} dim E_{\lambda_i}$.*

Demostración.

Para $1 \leq i \leq k$, sea β_i una base ordenada de E_{λ_i} y sea $\beta = (\beta_1, \beta_2, \ldots, \beta_k)$ la sucesión ordenada de las bases. Claramente, β genera el subespacio $W = E_{\lambda_1} + E_{\lambda_2} + \ldots + E_{\lambda_k}$. Falta ver que β es linealmente independiente. Cualquier relación lineal entre los vectores en β tiene la forma $w_1 + w_2 + \ldots + w_k = 0$ con $w_i \in E_{\lambda_i}$ para $1 \leq i \leq k$. Si algún $w_i \neq 0$, se tiene una relación de dependencia lineal entre vectores propios de A asociados a valores propios distintos, lo que contradice al teorema 2.3.1. Luego $w_i = 0$ para cada $i = 1, \ldots, k$. Ya que w_i es una combinación lineal de vectores en la base β_i, se ve que sólo se tiene la relación lineal trivial entre los vectores en β.
□

Teorema 2.3.3 *Sean $A \in \mathcal{M}_n$ y $\lambda_1, \lambda_2, \ldots, \lambda_k$ los valores propios distintos de A en F. Las siguientes afirmaciones son equivalentes:*

i) A es diagonalizable.

ii) El polinomio característico de A es $f_A(x) = (x-\lambda_1)^{d_1}\cdots(x-\lambda_k)^{d_k}$ *y* $dim E_{\lambda_i} = d_i$ *para* $i = 1, \ldots, k$.

iii) $dim E_{\lambda_1} + \ldots + dim E_{\lambda_k} = n$.

Demostración.
En la discusión anterior a la definición 2.3.1 observamos que i) implica ii). Si el polinomio característico $f_A(x)$ *es producto de factores lineales, como en ii), entonces* $d_1 + \ldots + d_k = n$. *En efecto, la suma de los* d_i *es el grado del polinomio característico y ese grado es* n. *Por tanto ii) implica iii).*
Supongamos que tenemos iii). Por el lema 2.3.1 tenemos que $F^n = E_{\lambda_1} + \ldots + E_{\lambda_k}$, *es decir, los vectores propios de* A *generan* F^n. □

Observemos que la equivalencia entre i) y ii) proporciona un criterio para la diagonalizabilidad. En efecto, $A \in \mathcal{M}_n$ es diagonalizable si y sólo si $f_A(x)$ se descompone en un producto de factores lineales y para cada valor propio λ de A el número de factores de $f_A(x)$ de la forma $(x - \lambda)$ es $n-$ rango$(A - \lambda I)$. Según el teorema 2.3.2 esto siempre ocurre para las raíces simples de $f_A(x)$, por lo que sólo debe verificarse para las raíces múltiples.

Ejemplo. Para la matriz $A = \begin{pmatrix} 7 & -2 & -4 \\ 3 & 0 & -2 \\ 6 & -2 & -3 \end{pmatrix}$ vimos en la sección anterior que su polinomio característico es $f_A(x) = (x-1)^2(x-2)$ y que los espacios propios E_1 y E_2 correspondientes a los valores propios 1 y 2 tienen dimensión 2 y 1 respectivamente, por lo que A es diagonalizable. En efecto, como E_1 está generado por $\left\{ \begin{pmatrix} 1 \\ 3 \\ 0 \end{pmatrix}, \begin{pmatrix} 2 \\ 0 \\ 3 \end{pmatrix} \right\}$ y E_2 por $\begin{pmatrix} 2 \\ 1 \\ 2 \end{pmatrix}$, juntamos las bases de E_1 y E_2 para formar la matriz invertible $P = \begin{pmatrix} 1 & 2 & 2 \\ 3 & 0 & 1 \\ 0 & 3 & 2 \end{pmatrix}$ que es la matriz de cambio de base para la cual $P^{-1}AP$ es diagonal.

2.4 El teorema de triangularización unitaria de Schur

Concluimos este capítulo con un resultado de fundamental importancia por su utilidad, a saber, que cualquier matriz $A \in \mathcal{M}_n$ es unitariamente equivalente a una matriz triangular superior cuyas entradas en la diagonal principal son los valores propios de A en un orden prescrito.

Teorema 2.4.1 (Schur). *Dada* $A \in \mathcal{M}_n$ *con valores propios* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *en cualquier orden prescrito, existe una matriz unitaria* $U \in \mathcal{M}_n$ *tal que* $U^*AU =$

$T = (t_{ij})$ es triangular superior con entradas en la diagonal principal $t_{ii} = \lambda_i$, $i = 1, \ldots, n$. Mas aún, si $A \in \mathcal{M}_n(\mathbb{R})$ y $\lambda_1, \lambda_2, \ldots, \lambda_n \in \mathbb{R}$, entonces U se puede escoger real y ortogonal.

Demostración.

La prueba es algorítmica y consiste en una sucesión de reducciones.

Sea $u_1 \in \mathbb{C}^n$ un vector propio de A asociado a λ_1 y de norma 1. Extendemos u_1 a una base ortonormal $\{u_1, w_2, \ldots, w_n\}$ de $\mathbb{C}^n$. Sea U_1 la matriz unitaria cuyas columnas son, de izquierda a derecha, $u_1, w_2, \ldots, w_n$. Ya que la primera columna de AU_1 es $\lambda_1 u_1$ y los renglones de U_1^* son $\overline{u_1}, \overline{w_2}, \ldots, \overline{w_n}$, donde la barra indica la conjugación compleja, un simple cálculo muestra que la matriz $U_1^* AU_1$ tiene la forma

$$\mathbf{U_1^* A U_1} = \begin{pmatrix} \lambda_1 & * \\ 0 & A_1 \end{pmatrix}$$

Por el lema 2.2.1, la matriz $A_1 \in \mathcal{M}_{n-1}$ tiene valores propios $\lambda_2, \ldots, \lambda_n$. Sea $u_2 \in \mathbb{C}^{n-1}$ un vector propio de A_1 asociado a λ_2 y de norma 1. Repitiendo el proceso anterior, obtenemos una matriz unitaria $U_2 \in \mathcal{M}_{n-1}$ tal que

$$\mathbf{U_2^* A_1 U_2} = \begin{pmatrix} \lambda_2 & * \\ 0 & A_2 \end{pmatrix}$$

Sea $\mathbf{V_2} = \begin{pmatrix} 1 & 0 \\ 0 & U_2 \end{pmatrix}$. Entonces V_2 es unitaria y por lo tanto también lo es U_1V_2. Además $V_2^* U_1^* AU_1 V_2$ tiene la forma

$$\mathbf{V_2^* U_1^* A U_1 V_2} = \begin{pmatrix} \lambda_1 & * \\ 0 & U_2^* A_1 U_2 \end{pmatrix} = \begin{pmatrix} \lambda_1 & * & * \\ 0 & \lambda_2 & \\ & 0 & A_2 \end{pmatrix}$$

Continuando este proceso de reducción, obtenemos matrices unitarias $U_i \in \mathcal{M}_{n-i+1}$ para $i = 1, \ldots, n-1$ y matrices unitarias $V_i \in \mathcal{M}_n$ para $i = 2, \ldots, n-1$. La matriz $U = U_1 V_2 V_3 \cdots V_{n-1}$ es unitaria y $U^* AU$ tiene la forma deseada.

Si todos los valores propios de $A \in \mathcal{M}_n(\mathbb{R})$ son reales, entonces los vectores propios correspondientes se pueden escoger con coordenadas reales y por lo tanto U es ortogonal real. □

Como consecuencia del teorema de Schur obtenemos el siguiente resultado que nos será útil en el siguiente capítulo.

Teorema 2.4.2 *Si $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los valores propios distintos de $A \in \mathcal{M}_n$ entonces para cualquier $m \in \mathbb{N}$, $\lambda_1^m, \lambda_2^m, \ldots, \lambda_k^m$ son todos los valores propios (no necesariamente distintos) de A^m.*

Demostración.

Sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los valores propios distintos de A con multiplicidades algebraicas

$n_1, n_2, \ldots, n_k$ respectivamente. Por el teorema de Schur existe $U \in \mathcal{M}_n$ unitaria tal que

$$\mathbf{U^*AU} = \mathbf{T} = \begin{pmatrix} \lambda_1 & & * \\ & \ddots & \\ 0 & & \lambda_k \end{pmatrix}$$

es triangular superior y cada λ_i aparece repetido n_i veces en su diagonal principal. Ya que U es unitaria, $U^* = U^{-1}$ y por lo tanto

$$\mathbf{U^*A^mU} = \mathbf{T^m} = \begin{pmatrix} \lambda_1^m & & * \\ & \ddots & \\ 0 & & \lambda_k^m \end{pmatrix}$$

donde cada λ_i^m aparece repetido n_i veces en la diagonal principal. Por el lema 2.2.1, A^m y T^m tienen los mismos valores propios. Luego $\lambda_1^m, \lambda_2^m, \ldots, \lambda_k^m$ son todos los valores propios de A^m. $\square$

Corolario 2.4.1 *Sean $A \in \mathcal{M}_n$ y $m \in \mathbb{N}$, entonces $\rho(A^m) = \rho(A)^m$.*
Demostración.
Sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los valores propios distintos de A. De acuerdo con la definición de radio espectral, existe $1 \leq j \leq k$ tal que $|\lambda_j| = \rho(A)$ y para todo $i \neq j$ $(1 \leq i \leq k)$, $|\lambda_i| \leq \rho(A)$. Por lo tanto, $|\lambda_j^m| = |\lambda_j|^m = \rho(A)^m$ y para todo $i \neq j$, $|\lambda_i^m| = |\lambda_i|^m \leq \rho(A)^m$. Como $\lambda_1^m, \lambda_2^m, \ldots, \lambda_k^m$ son todos los valores propios de A^m se sigue que $\rho(A^m) = \rho(A)^m$. $\square$

Capítulo 3

El Teorema de Perron-Frobenius

3.1 Normas vectoriales

Definición 3.1.1 *Sea V un espacio vectorial sobre F. Una función* $\| \cdot \|: V \to \mathbb{R}$ *es una* **norma vectorial** *si para todo $x, y \in V$ y para todo $c \in F$ se cumplen:*

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

ii) $\| cx \| = |c| \, \| x \|$

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

Una norma vectorial es entonces una generalización de la familiar longitud Euclidiana en el plano o en el espacio y nos permite hablar de "tamaño" y "cercanía" en un espacio vectorial arbitrario.

Lema 3.1.1 *Si* $\| \cdot \|$ *es una norma vectorial en V, entonces para todo $x, y \in V$,*

$$\Big| \| x \| - \| y \| \Big| \leq \| x - y \| .$$

Demostración.
Por iii), $\| x \| = \| x - y + y \| \leq \| x - y \| + \| y \|$, *de donde* $\| x \| - \| y \| \leq \| x - y \|$. *Análogamente,* $\| y \| - \| x \| \leq \| x - y \|$. *Por lo tanto,* $\Big| \| x \| - \| y \| \Big| \leq \| x - y \|$.
□

A continuación damos algunos ejemplos de normas vectoriales. En el capítulo 1 vimos el primero de ellos.

Ejemplo 1. Si $< \cdot , \cdot >$ es un producto interno en V, entonces $\| x \| = < x , x >^{\frac{1}{2}}$ es una norma vectorial, llamada la norma vectorial inducida por el producto interno.

En los ejemplos 2 a 5, $V = \mathbb{C}^n$ y $x \in V$.

Ejemplo 2. La norma ℓ_1 está dada por

$$\| x \|_1 = |x_1| + \ldots + |x_n|$$

Ejemplo 3. La norma ℓ_2 (o norma euclidiana) está dada por

$$\| x \|_2 = \left(|x_1|^2 + \ldots + |x_n|^2\right)^{\frac{1}{2}}$$

Observamos que esta norma es la inducida por el producto interno canónico.

Ejemplo 4. La norma ℓ_p está dada por

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

Ejemplo 5. La norma ℓ_∞ está dada por

$$\| x \|_\infty = \max\left\{|x_1|, \ldots, |x_n|\right\}$$

Ejemplo 6. Si V es un espacio vectorial de dimensión finita sobre F y $\beta = \{v_1, v_2, \ldots, v_n\}$ es una base ordenada de V, sabemos que la función $[\,\cdot\,]_\beta : V \to \mathbb{C}^n$ dada por

$$[v]_\beta = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

para $v = \sum_{i=1}^{n} x_i v_i$, es un isomorfismo de V en $\mathbb{C}^n$. Entonces, si $\| \cdot \|$ es cualquier norma en $\mathbb{C}^n$, $\| v \|_\beta = \| [v]_\beta \| = \| (x_1, x_2, \ldots, x_n)^t \|$ es una norma en V.

En los ejemplos 7 a 10, V es el espacio vectorial de todas las funciones continuas del intervalo [a,b] en F, es decir, $V = C[a,b] = \{\, f : [a,b] \to F \mid f \text{ es continua}\}$.

Ejemplo 7. La norma L_1 está dada por

$$\| f \|_1 = \int_a^b |f(t)|\, dt$$

Ejemplo 8. La norma L_2 está dada por

$$\| f \|_2 = \left(\int_a^b |f(t)|^2\, dt\right)^{\frac{1}{2}}$$

Ejemplo 9. La norma L_p está dada por

$$\| f \|_p = \left(\int_a^b |f(t)|^p \, \mathrm{d}t \right)^{\frac{1}{p}}, \quad p \geq 1$$

Ejemplo 10. La norma L_∞ está dada por

$$\| f \|_\infty = \max \left\{ |f(x)| \mid x \in [a, b] \right\}$$

Si V es un espacio con producto interno, una norma $\| \cdot \|$ definida sobre V está inducida por el producto interno si y sólo si satisface la ley del paralelogramo, es decir, que para cualesquier $x, y \in V$,

$$\frac{1}{2}(\| x + y \|^2 + \| x - y \|^2) = \| x \|^2 + \| y \|^2.$$

Vemos que la norma ℓ_∞ del ejemplo 5 anterior no está inducida por un producto interno, pues para $n = 2$ y para $x = (1, 0)$, $y = (0, 1)$ no se satisface la ley del paralelogramo.

Además de medir el "tamaño" de un vector y determinar si dos vectores están "cerca" uno del otro, las normas se pueden usar para medir la convergencia de una sucesión de vectores.

Definición 3.1.2 *Sea V un espacio vectorial sobre F y sea $\| \cdot \|$ una norma vectorial en V. Decimos que la sucesión $\{v_{(k)}\}$ de vectores en V* **converge** *a un vector $v \in V$* **con respecto** *a la norma $\| \cdot \|$ si y sólo si $\| v_{(k)} - v \| \to 0$ cuando $k \to \infty$ y escribimos*

$$\lim_{k \to \infty} v_{(k)} = v \quad \textit{con respecto a} \quad \| \cdot \|$$

ó

$$v_{(k)} \xrightarrow[\|\cdot\|]{} v$$

Observamos que el límite de una sucesión $\{v_{(k)}\}$ de vectores en V con respecto a una norma $\| \cdot \|$ dada está bien definido. En efecto, si

$$v_{(k)} \xrightarrow[\|\cdot\|]{} v \quad \text{y} \quad v_{(k)} \xrightarrow[\|\cdot\|]{} u,$$

entonces

$$\| v_{(k)} - v \| \to 0 \quad \text{y} \quad \| v_{(k)} - u \| \to 0 \quad \text{cuando} \quad k \to \infty$$

y por lo tanto

$$\| u - v \| = \| u - v_{(k)} + v_{(k)} - v \| \leq \| u - v_{(k)} \| + \| v_{(k)} - v \| \to 0 \quad \text{cuando} \quad k \to \infty.$$

Luego, $v = u$.

Nos preguntamos ahora si es posible que una sucesión dada de vectores converja con respecto a una norma y no converja con respecto a otra. Como veremos enseguida, esto no puede ocurrir en un espacio vectorial de dimensión finita.

Recordemos que una función $f : F^n \to \mathbb{R}$ es **uniformemente continua** en $A \subset F^n$ si para todo $\epsilon > 0$ existe $\delta > 0$ tal que si $x, u \in A$ y $\| x - u \|_2 \leq \delta$ entonces $|f(x) - f(u)| \leq \epsilon$.

Lema 3.1.2 *Sea $\| \cdot \|$ una norma en un espacio vectorial V sobre F y sean $v_1, v_2, \ldots, v_m \in V$ vectores dados. La función $g : F^m \to \mathbb{R}$ definida por*

$$g(x_1, x_2, \ldots, x_m) = \| x_1 v_1 + x_2 v_2 + \ldots + x_m v_m \|$$

es una función uniformemente continua.
Demostración.
Sean $x = \sum_{i=1}^m x_i v_i$ y $y = \sum_{i=1}^m y_i v_i$, donde $x_i, y_i \in F$ para $1 \leq i \leq m$. Entonces

$$\begin{aligned} |g(x_1, x_2, \ldots, x_m) - g(y_1, y_2, \ldots, y_m)| &= \Big| \| x \| - \| y \| \Big| \\ &\leq \| x - y \| = \left\| \sum_{i=1}^m (x_i - y_i) v_i \right\| \\ &\leq \sum_{i=1}^m |x_i - y_i| \, \| v_i \| \\ &\leq m \max_{1 \leq i \leq m} \left(|x_i - y_i| \, \| v_i \| \right) \\ &\leq m \max_{1 \leq i \leq m} \| v_i \| \cdot \max_{1 \leq i \leq m} |x_i - y_i|. \end{aligned}$$

Sea $c = m \max_{1 \leq i \leq m} \| v_i \|$. Como algún $v_i \neq 0$ (en caso contrario, $g = 0$ y no hay nada que probar), $c > 0$. Luego

$$\begin{aligned} |g(x_1, x_2, \ldots, x_m) - g(y_1, y_2, \ldots, y_m)| &\leq c \max_{1 \leq i \leq m} |x_i - y_i| \\ &\leq c \, \| (x_1, x_2, \ldots, x_m) - (y_1, y_2, \ldots, y_m) \|_2 . \end{aligned}$$

Entonces, para que $|g(x_1, x_2, \ldots, x_m) - g(y_1, y_2, \ldots, y_m)| < \epsilon$ basta con elegir $\delta = \frac{\epsilon}{c}$. □

Corolario 3.1.1 *Cualquier norma $\| \cdot \|$ en F^m es uniformemente continua.*
Demostración.
Basta con tomar $v_i = e_i$ en el lema 3.1.2 y observar que $\| x \| = g(x_1, x_2, \ldots, x_m)$ para todo $x = (x_1, x_2, \ldots, x_m) \in F^m$. □

Teorema 3.1.1 *Sean V un espacio vectorial de dimensión finita sobre F y $\beta = \{v_1, v_2, \ldots, v_n\}$ una base ordenada de V. Si $f_1, f_2 : V \to F$ 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* $\alpha \in F$ *y todo* $v \in V$

c) $f_i\Big(v(x)\Big)$ *es continua en* F^n, *donde* $x = (x_1, x_2, \ldots, x_n) \in F^n$ *y* $v(x) = x_1v_1 + x_2v_2 + \ldots + x_nv_n$,

entonces existen $C_m, C_M \in \mathbb{R}^+$ *tales que para cualquier* $v \in V$,

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

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

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

donde $v(x) = x_1v_1 + \ldots + x_nv_n$. *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* F^n. *Luego, por el teorema del valor máximo y mínimo (ver [1] pag. 180),* h *alcanza su máximo* C_M *y su mínimo* C_m *en* S *y ambos son positivos. Por lo tanto, para todo* $x \in S$,

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

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

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

Corolario 3.1.2 *Sean* $\| \cdot \|_\alpha$ *y* $\| \cdot \|_\beta$ *normas en un espacio vectorial* V *de dimensión finita sobre* F. *Entonces existen constantes positivas* C_m *y* C_M *tales que para todo* $v \in V$,

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

Corolario 3.1.3 *Si* $\| \cdot \|_\alpha$ *y* $\| \cdot \|_\beta$ *son normas en un espacio vectorial* V *de dimensión finita sobre* F *y si* $\{v_{(k)}\}$ *es una sucesión dada de vectores, entonces*

$$v_{(k)} \xrightarrow[\|\cdot\|_\alpha]{} v \quad \text{si y sólo si} \quad v_{(k)} \xrightarrow[\|\cdot\|_\beta]{} v.$$

Demostración.
Por el corolario 3.1.2 existen C_m *y* C_M *en* $\mathbb{R}^+$ *tales que para todo* $k \in \mathbb{N}$

$$C_m \| v_{(k)} - v \|_\alpha \leq \| v_{(k)} - v \|_\beta \leq C_M \| v_{(k)} - v \|_\alpha.$$

Por lo tanto, $\| v_{(k)} - v \|_\alpha \longrightarrow 0$ *si y sólo si* $\| v_{(k)} - v \|_\beta \longrightarrow 0$. $\square$

Definición 3.1.3 *Dos normas vectoriales son* **equivalentes** *si cada vez que la sucesión $\{v_{(k)}\}$ converge a un vector v con respecto a una norma, también converge a v con respecto a la otra norma.*

El corolario 3.1.3 muestra entonces que en un espacio vectorial de dimensión finita sobre F todas las normas son equivalentes. En particular, todas las normas en $\mathbb{R}^n$ ó $\mathbb{C}^n$ son equivalentes a la norma ℓ_∞, donde $\| x \|_\infty = \max\left\{|x_1|, \dots, |x_n|\right\}$. Así,

$$x_{(k)} \xrightarrow[\|\cdot\|]{} x \quad \text{si y sólo si} \quad x_{(k)} \xrightarrow[\|\cdot\|_\infty]{} x$$

si y sólo si $(x_{(k)})_i \to x_i$ cuando $k \to \infty$. Es decir, la sucesión de vectores $\{x_{(k)}\}$ en $\mathbb{R}^n$ ó $\mathbb{C}^n$ converge a x con respecto a cualquier norma si y sólo si la sucesión de las i-ésimas componentes de $\{x_{(k)}\}$ converge a x_i para todo $i = 1, \dots, n$.

En un espacio vectorial de dimensión infinita, dos normas distintas pueden no ser equivalentes, como lo muestra el siguiente ejemplo.

Ejemplo. Considérese la sucesión $\{f_{(k)}\}$ de funciones en $C[0,1] = \{f : [0,1] \to F | f$ es continua$\}$ definida por

$$\begin{array}{ll} f_k(x) = 0, & 0 \le x \le \frac{1}{k} \\ f_k(x) = 2(k^{\frac{3}{2}}x - k^{\frac{1}{2}}), & \frac{1}{k} \le x \le \frac{3}{2k} \\ f_k(x) = 2(-k^{\frac{3}{2}}x + 2k^{\frac{1}{2}}), & \frac{3}{2k} \le x \le \frac{2}{k} \\ f_k(x) = 0, & \frac{2}{k} \le x \le 1 \\ \text{para } k = 2, 3, \dots & \end{array}$$

Para la norma L_1 se tiene que

$$\| f_{(k)} \|_1 = \int_0^1 |f_{(k)}(t)|dt = \int_{\frac{1}{k}}^{\frac{3}{2k}} |2(k^{\frac{3}{2}}t - k^{\frac{1}{2}})|dt + \int_{\frac{3}{2k}}^{\frac{2}{k}} |2(-k^{\frac{3}{2}}t + 2k^{\frac{1}{2}})|dt =$$

$$2\int_{\frac{1}{k}}^{\frac{3}{2k}} (k^{\frac{3}{2}}t - k^{\frac{1}{2}})dt + 2\int_{\frac{3}{2k}}^{\frac{2}{k}} (-k^{\frac{3}{2}}t + 2k^{\frac{1}{2}})dt = \frac{1}{2k^{\frac{1}{2}}} \to 0 \quad si \quad k \to \infty.$$

Para la norma L_2 se tiene que

$$\begin{aligned} \left(\| f_{(k)} \|_2\right)^2 &= \int_0^1 |f_{(k)}(t)|^2 dt \\ &= \int_{\frac{1}{k}}^{\frac{3}{2k}} 4(k^3t^2 - 2k^2t + k)dt + \int_{\frac{3}{2k}}^{\frac{2}{k}} 4(k^3t^2 - 4k^2t + 4k)dt \\ &= \frac{1}{3}. \end{aligned}$$

Por lo tanto, $\| f_{(k)} \|_2 = \frac{1}{\sqrt{3}}$ para todo k. Finalmente, para la norma L_∞ se tiene que

$$\| f_{(k)} \|_\infty = \max\left\{|f_{(k)}(x)| \,:\, x \in [0,1]\right\} = \left|f_{(k)}\left(\frac{3}{2k}\right)\right| = f_{(k)}\left(\frac{3}{2k}\right) = k^{\frac{1}{2}} \to \infty$$

si $k \to \infty$.

3.2 Normas matriciales y radio espectral

Dado que $\mathcal{M}_n$ es un espacio vectorial de dimensión n^2, podemos usar cualquier norma vectorial sobre $\mathbb{C}^{n^2}$ para medir el "tamaño" de una matriz. Sin embargo, $\mathcal{M}_n$ además de ser un espacio vectorial, tiene definida una operación natural, la multiplicación de matrices. La definición de una norma matricial permite no sólo medir el "tamaño" de una matriz, sino relacionar el "tamaño" de AB con los "tamaños" de A y de B.

Definición 3.2.1 *Una función* $|||\cdot||| : \mathcal{M}_n \to \mathbb{R}$ *es una* **norma matricial** *si para toda* $A, B \in \mathcal{M}_n$ *y para todo* $c \in F$ *se cumplen:*

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

ii) $|||cA||| = |c|\ |||A|||$

iii) $|||A + B||| \leq |||A||| + |||B|||$

iv) $|||AB||| \leq |||A|||\ |||B|||$.

Claramente, toda norma matricial es una norma vectorial. Por otro lado, sólo algunas de las normas vectoriales introducidas en la sección anterior son normas matriciales al definirlas sobre $\mathcal{M}_n$, como veremos en los siguientes ejemplos.

Ejemplo 1. La norma ℓ_1 definida para $A \in \mathcal{M}_n$ por

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

es una norma matricial. En efecto,

i) $||A||_1 = \sum_{i,j=1}^n |a_{ij}| \geq 0$; $\quad ||A||_1 = 0$ si y sólo si $A = 0$.

ii) $||cA||_1 = \sum_{i,j=1}^n |ca_{ij}| = \sum_{i,j=1}^n |c|\, |a_{ij}| = |c| \sum_{i,j=1}^n |a_{ij}| = |c|\, ||A||_1$.

iii) $||A+B||_1 = \sum_{i,j=1}^n |a_{ij} + b_{ij}| \leq \sum_{i,j=1}^n (|a_{ij}| + |b_{ij}|) = \sum_{i,j=1}^n |a_{ij}| + \sum_{i,j=1}^n |b_{ij}| = ||A||_1 + ||B||_1$.

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

Ejemplo 2. La norma ℓ_2 definida para $A \in \mathcal{M}_n$ por

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

es una norma matricial.

Ejemplo 3. La norma matricial **máxima suma de columnas** se define para $A \in \mathcal{M}_n$ por

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

Ejemplo 4. La norma matricial **máxima suma de renglones** se define para $A \in \mathcal{M}_n$ por

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

Ejemplo 5. Una norma vectorial que no es una norma matricial es la norma ℓ_∞ definida para $A \in \mathcal{M}_n$ por

$$\|A\|_\infty = \max_{i,j} |a_{ij}|.$$

Si consideramos la matriz $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \in \mathcal{M}_2$ entonces $A^2 = 2A$, $\| A \|_\infty = 1$, $\| A^2 \|_\infty = \| 2A \|_\infty = 2\| A \|_\infty = 2$ y por lo tanto no se cumple que $\| A^2 \|_\infty \leq \| A \|_\infty^2$.

Probamos enseguida que una norma matricial se puede transformar en otra por medio de una similitud dada.

Teorema 3.2.1 *Si $|||\cdot|||$ es una norma matricial sobre $\mathcal{M}_n$ y si $S \in \mathcal{M}_n$ es una matriz invertible, entonces $|||\cdot|||_S$ definida para $A \in \mathcal{M}_n$ por*

$$|||A|||_S = |||S^{-1}AS|||$$

es una norma matricial.
Demostración.
Las propiedades i), ii) y iii) se verifican fácilmente para $|||\cdot|||$. Para la propiedad iv) tenemos que

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

Las normas matriciales proporcionan cotas para el espectro de una matriz.

Teorema 3.2.2 *Sea $|||\cdot|||$ cualquier norma matricial. Si $A \in \mathcal{M}_n$, entonces*

$$\rho(A) \leq |||A|||.$$

Demostración.
Por la definición de radio espectral (definición 2.2.6) existe un valor propio λ de A tal que $|\lambda| = \rho(A)$. Sea x un vector propio de A asociado a λ, entonces $Ax = \lambda x$ y $x \neq 0$. Sea $B \in \mathcal{M}_n$ tal que todas sus columnas son iguales al vector propio x, entonces $AB = \lambda B$. Si $|||\cdot|||$ es cualquier norma matricial, entonces

$$|\lambda|\;|||B||| = |||\lambda B||| = |||AB||| \leq |||A|||\;|||B|||.$$

Por lo tanto,

$$|\lambda| = \rho(A) \leq |||A|||. \quad \square$$

Lema 3.2.1 *Sean $A \in \mathcal{M}_n$ y $\epsilon > 0$, entonces existe una norma matricial $|||\cdot|||$ tal que $\rho(A) \leq |||A||| \leq \rho(A) + \epsilon$.*
Demostración.
Por el teorema de Schur, existe $U \in \mathcal{M}_n$ unitaria tal que $\Delta = U^*AU$ es una matriz triangular superior cuyas entradas en la diagonal principal son los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$ de A (en $\mathbb{C}$). Supongamos que

$$\Delta = \begin{pmatrix} \lambda_1 & d_{1\,2} & \ldots & d_{1\,n} \\ 0 & \lambda_2 & \ldots & d_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

y para $t > 0$, sea

$$\mathbf{D_t} = \begin{pmatrix} t & & & 0 \\ & t^2 & & \\ & & \ddots & \\ 0 & & & t^n \end{pmatrix}$$

Entonces

$$\mathbf{D_t}\Delta\mathbf{D_t^{-1}} = \begin{pmatrix} t & & & 0 \\ & t^2 & & \\ & & \ddots & \\ 0 & & & t^n \end{pmatrix} \begin{pmatrix} \lambda_1 & d_{1\,2} & \ldots & d_{1\,n} \\ 0 & \lambda_2 & \ldots & d_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix} \begin{pmatrix} t^{-1} & & & 0 \\ & t^{-2} & & \\ & & \ddots & \\ 0 & & & t^{-n} \end{pmatrix}$$

$$= \begin{pmatrix} t\lambda_1 & td_{1\,2} & \dots & td_{1\,n} \\ 0 & t^2\lambda_2 & \dots & t^2 d_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & t^n\lambda_n \end{pmatrix} \begin{pmatrix} t^{-1} & & & 0 \\ & t^{-2} & & \\ & & \ddots & \\ 0 & & & t^{-n} \end{pmatrix}$$

$$= \begin{pmatrix} \lambda_1 & t^{-1}d_{1\,2} & \dots & t^{1-n}d_{1\,n} \\ 0 & \lambda_2 & \dots & t^{2-n} d_{2\,n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

Haciendo $t > 0$ suficientemente grande, podemos acotar la suma de todos los módulos de las entradas fuera de la diagonal de $D_t\Delta D_t^{-1}$ y hacer dicha suma menor que ϵ. Entonces $|||D_t\Delta D_t^{-1}|||_1 \leq \rho(A) + \epsilon$ para t suficientemente grande.
Definimos $|||\cdot||| : \mathcal{M}_n \to \mathbb{R}$ por

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

para toda $B \in \mathcal{M}_n$. Ya que U es unitaria,

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

Luego, por el teorema 3.2.1, $|||\cdot|||$ es una norma matricial. Esta norma satisface $|||A||| \leq \rho(A) + \epsilon$ para t suficientemente grande. □

Observamos que los resultados anteriores muestran que $\rho(A) = \inf\{|||A||| : |||\cdot|||$ es una norma matricial$\}$.

Recordemos que dada $A \in \mathcal{M}_n$, se define $A^m = A^{m-1}A$ para todo $m \in \mathbb{N}$. En el siguiente capítulo estaremos interesados en determinar el "límite" (si existe) de la sucesión de matrices $\{A^k : k \in \mathbb{N}\}$.

Definición 3.2.2 *Sean $L, A_1, A_2, A_3, \dots$ matrices en $\mathcal{M}_{n\times p}$. Se dice que la sucesión $A_1, A_2, A_3, \dots$ converge a la matriz L, denominada el* **límite de la sucesión**, *si para cada par de índices i y j, con $1 \leq i \leq n$ y $1 \leq j \leq p$, la sucesión de números complejos $(A_1)_{ij}, (A_2)_{ij}, \dots$ converge a L_{ij}. En este caso escribimos $\lim_{m\to\infty} A_m = L$.*

Recordemos que el límite de la sucesión de números complejos $\{z_m : m \in \mathbb{N}\}$ se define como

$$\lim_{m\to\infty} z_m = (\lim_{m\to\infty} Re(z_m)) + i(\lim_{m\to\infty} Im(z_m)).$$

Ejemplos:

$$\text{Si } \mathbf{A_m} = \begin{pmatrix} (-\frac{2}{3})^m & 5 \\ i - \frac{2}{m} & (\frac{i}{6})^m \end{pmatrix} \text{ entonces } \lim_{m\to\infty} \mathbf{A_m} = \begin{pmatrix} 0 & 5 \\ i & 0 \end{pmatrix}$$

$$\text{Si } \mathbf{B_m} = \begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} \text{ entonces } \lim_{m\to\infty} \mathbf{B_m} \text{ no existe.}$$

Teorema 3.2.3 *Sean* $L, A_1, A_2, \ldots$ *matrices en* $\mathcal{M}_{n\times p}$ *tales que* $\lim_{m\to\infty} A_m = L$. *Entonces para cualesquier* $B \in \mathcal{M}_{r\times n}$ *y* $C \in \mathcal{M}_{p\times s}$ *se tiene que*

$$\lim_{m\to\infty} BA_m = BL \quad y \quad \lim_{m\to\infty} A_m C = LC.$$

Demostración.
Para cualquier par de índices i *y* j *con* $1 \le i \le r$ *y* $1 \le j \le p$,

$$\begin{aligned} \lim_{m\to\infty}[(BA_m)_{ij}] &= \lim_{m\to\infty}\Big[\sum_{k=1}^{n} B_{ik}(A_m)_{kj}\Big] \\ &= \sum_{k=1}^{n} B_{ki}\Big\{\lim_{m\to\infty}[(A_m)_{kj}]\Big\} \\ &= \sum_{k=1}^{n} B_{ik}L_{kj} \\ &= (BL)_{ij}. \end{aligned}$$

Por lo tanto, $\lim_{m\to\infty} BA_m = BL$. *Análogamente,* $\lim_{m\to\infty} A_m C = LC$. □

Corolario 3.2.1 *Sea* $A \in \mathcal{M}_n$ *y sea* $\lim_{m\to\infty} A^m = L$. *Entonces para cualquier matriz invertible* $Q \in \mathcal{M}_n$ *se tiene que* $\lim_{m\to\infty}(QAQ^{-1})^m = QLQ^{-1}$.
Demostración.
Dado que $(QAQ^{-1})^m = (QAQ^{-1})(QAQ^{-1})\ldots(QAQ^{-1}) = QA^mQ^{-1}$, *usando el teorema anterior, obtenemos que*

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

En esta sección estamos interesados en caracterizar las matrices $A \in \mathcal{M}_n$ tales que $\lim_{k\to\infty} A^k = 0$.

Lema 3.2.2 *Sea* $A \in \mathcal{M}_n$. *Si* $|||A||| < 1$ *para alguna norma matricial* $|||\cdot|||$, *entonces*

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

Demostración.
Si $|||A||| < 1$, *ya que* $|||A^k||| \le (|||A|||)^k$, *entonces* $|||A^k||| \to 0$ *cuando* $k \to \infty$, *es decir,* $A^k \to 0$ *con respecto a la norma* $|||\cdot|||$. *Dado que todas las normas vectoriales sobre* $\mathcal{M}_n$ *son equivalentes (corolario 3.1.3), debe ocurrir que* $A^k \to 0$ *con respecto a la norma* $\|\cdot\|_\infty$, *esto es, todas las entradas de* A^k *tienden a cero cuando* $k \to \infty$. □

Definición 3.2.3 *Se dice que $A \in \mathcal{M}_n$ es* **convergente** *si* $\lim_{k\to\infty} A^k = 0$.

Teorema 3.2.4 *Sea $A \in \mathcal{M}_n$. Entonces A es convergente si y sólo si $\rho(A) < 1$.*
Demostración.
Si A es convergente, para cualquier vector propio x de A con valor propio λ, $A^k x = \lambda^k x \to 0$ cuando $k \to \infty$. Luego $|\lambda| < 1$ para todo valor propio λ de A y por lo tanto $\rho(A) < 1$.
Si $\rho(A) < 1$, por el lema 3.2.1, existe una norma matricial $|||\cdot|||$ tal que $|||A||| < 1$, de donde A es convergente. □

Corolario 3.2.2 *Sea $|||\cdot|||$ una norma matricial. Entonces para toda $A \in \mathcal{M}_n$,*

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

Demostración.
Sea λ un valor propio de A tal que $\rho(A) = |\lambda|$, entonces λ^k es valor propio de A^k y por lo tanto $\rho(A)^k = |\lambda|^k = |\lambda^k| \leq \rho(A^k) \leq |||A^k|||$. Luego, para todo $k \in \mathbb{N}$,

$$\rho(A) \leq |||A^k|||^{\frac{1}{k}}.$$

Sean $\epsilon > 0$ y $\widetilde{A} = (\rho(A)+\epsilon)^{-1}A$. Si λ es un valor propio de $\widetilde{A}$, entonces existe un vector no nulo x tal que $\lambda x = \widetilde{A}x = (\rho(A)+\epsilon)^{-1}Ax$, de donde $Ax = (\rho(A)+\epsilon)\lambda x$. Luego $(\rho(A)+\epsilon)\lambda$ es un valor propio de A y por lo tanto

$$(\rho(A)+\epsilon)|\lambda| = |(\rho(A)+\epsilon)\lambda| \leq \rho(A).$$

Entonces $|\lambda| \leq (\rho(A)+\epsilon)^{-1}\rho(A) < 1$, por lo que $\rho(\widetilde{A}) < 1$ y $\widetilde{A}$ es convergente.
Ya que $lim_{k\to\infty}|||\widetilde{A}^k||| = 0$, existe $N \in \mathbb{N}$ tal que $(\rho(A)+\epsilon)^{-k}|||A^k||| = |||\widetilde{A}^k||| < 1$ para todo $k \geq N$. Luego, para todo $k \geq N$,

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

Por lo tanto, $lim_{k\to\infty}|||A^k|||^{\frac{1}{k}} = \rho(A)$. □

3.3 Matrices no negativas y matrices positivas

Definición 3.3.1 *Sean $A, B \in \mathcal{M}_{n\times r}(\mathbb{R})$. Se dice que:*

i) B *es* **no negativa**, *o bien $B \geq 0$, si $b_{ij} \geq 0$ para todo $i = 1,\ldots,n$ y todo $j = 1,\ldots,r$.*

ii) B *es* **positiva**, *o bien $B > 0$, si $b_{ij} > 0$ para todo $i = 1,\ldots,n$ y todo $j = 1,\ldots,r$.*

iii) $\mathbf{A} \geq \mathbf{B}$ *si $A - B \geq 0$.*

iv) $\mathbf{A} > \mathbf{B}$ *si* $A - B > 0$.

Observamos que $\geq$ es un orden parcial en $\mathcal{M}_{n\times r}$ y que es compatible con las operaciones usuales de suma y multiplicación en $\mathcal{M}_n$. En particular, probamos los siguientes resultados.

Lema 3.3.1 *Si* $0 \leq A \leq B$ *y* $0 \leq C \leq D$, *entonces* $0 \leq AC \leq BD$.
Demostración.
Dado que $0 \leq a_{ik} \leq b_{ik}$ *y* $0 \leq c_{kj}$ *para todo* $i, j, k = 1, \ldots, n$, *se tiene que*

$$0 \leq \sum_{k=1}^{n} a_{ik}c_{kj} \leq \sum_{k=1}^{n} b_{ik}c_{kj}$$

para todo $i, j = 1, \ldots, n$, *de donde,* $0 \leq AC \leq BC$. *Análogamente,* $BC \leq BD$. □

Corolario 3.3.1 *Si* $0 \leq A \leq B$, *entonces* $0 \leq A^m \leq B^m$ *para todo* $m \in \mathbb{N}$.
Demostración.
Por inducción sobre m. □

Lema 3.3.2 *Si* $A > 0$, $x \geq 0$ *y* $x \neq 0$, *entonces* $Ax > 0$.
Demostración.
Sabemos que existe $k \in \{1, \ldots, n\}$ *tal que* $x_k > 0$. *Entonces para todo* $i = 1, \ldots, n$,

$$(Ax)_i = \sum_{j=1}^{n} a_{ij}x_j = a_{ik}x_k + \sum_{j=1}^{k-1} a_{ij}x_j + \sum_{j=k+1}^{n} a_{ij}x_j > 0$$

pues $a_{ik} > 0$. □

Definición 3.3.2 *Sea* $A \in \mathcal{M}_{n\times r}$. *La* **matriz de los valores absolutos de las componentes de A** *es* $|A| = [|a_{ij}|]$.

Ejemplo. Sea la matriz $\mathbf{A} = \begin{pmatrix} 1+i & 0 \\ 3-4i & -2 \end{pmatrix}$, entonces $|\mathbf{A}| = \begin{pmatrix} \sqrt{2} & 0 \\ 5 & 2 \end{pmatrix}$.

Observamos que $|A|$ es una matriz no negativa para cualquier matriz $A \in \mathcal{M}_{n\times r}$.

Lema 3.3.3 *Si* $A, B \in \mathcal{M}_n$, *entonces* $|AB| \leq |A|\,|B|$.
Demostración.
Para cualquier par de índices $i, j = 1, \ldots, n$ *se tiene que*

$$|AB|_{ij} = \Big|\sum_{k=1}^{n} a_{ik}b_{kj}\Big| \leq \sum_{k=1}^{n} |a_{ik}|\,|b_{kj}| = (|A|\,|B|)_{ij}. \quad \square$$

Corolario 3.3.2 *Si* $A \in \mathcal{M}_n$, *entonces* $|A^m| \leq |A|^m$ *para todo* $m \in \mathbb{N}$.
Demostración.
Por inducción sobre m. □

Lema 3.3.4 *Si $|A| \leq |B|$, entonces $\| A \|_2 \leq \| B \|_2$.*
Demostración.
Dado que $|a_{ij}| \leq |b_{ij}|$ para todo $i, j = 1, \ldots, n$, se tiene que $\sum_{i,j} |a_{ij}|^2 \leq \sum_{i,j} |b_{ij}|^2$, de donde, $\| A \|_2 \leq \| B \|_2$. □

Lema 3.3.5 *Si $A \in \mathcal{M}_n$, entonces $\| A \|_2 = \| \, |A| \, \|_2$.* □

Como consecuencia de estos sencillos resultados obtenemos:

Teorema 3.3.1 *Sean $A, B \in \mathcal{M}_n$. Si $|A| \leq B$, entonces $\rho(A) \leq \rho(|A|) \leq \rho(B)$.*
Demostración.
Por los corolarios 3.3.2 y 3.3.1, $|A^m| \leq |A|^m \leq B^m$ para todo $m \in \mathbb{N}$. Luego, por los lemas 3.3.5 y 3.3.4,

$$\| A^m \|_2 = \| \, |A^m| \, \|_2 \leq \| \, |A|^m \, \|_2 \leq \| B^m \|_2$$

para todo $m \in \mathbb{N}$, de donde

$$\| A^m \|_2^{\frac{1}{m}} \leq \| \, |A|^m \, \|_2^{\frac{1}{m}} \leq \| B^m \|_2^{\frac{1}{m}}$$

para todo $m \in \mathbb{N}$. Entonces, haciendo $m \to \infty$, por el corolario 3.2.2, obtenemos que $\rho(A) \leq \rho(|A|) \leq \rho(B)$. □

Corolario 3.3.3 *Sean $A, B \in \mathcal{M}_n$. Si $0 \leq A \leq B$, entonces $\rho(A) \leq \rho(B)$.* □

Lema 3.3.6 *Sea $A \in \mathcal{M}_n$ tal que $A \geq 0$. Si la suma de renglones de A es constante, entonces $\rho(A) = |||A|||_\infty$.*
Demostración.
Por el teorema 3.2.2, $\rho(A) \leq |||A|||_\infty = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}|$. Si la suma de renglones de A es constante, entonces $x = [1, 1, \ldots, 1]^t$ es un vector propio de A con valor propio $|||A|||_\infty$. Luego $|||A|||_\infty \leq \rho(A)$ y por lo tanto $\rho(A) = |||A|||_\infty$. □

Teorema 3.3.2 *Sea $A \in \mathcal{M}_n$ tal que $A \geq 0$. Entonces*

$$\min_{1 \leq i \leq n} \sum_{j=1}^n a_{ij} \leq \rho(A) \leq \max_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}.$$

Demostración.
Sea $\alpha = \min_{1 \leq i \leq n} \sum_{j=1}^n a_{ij}$ y construyamos una matriz B tal que $0 \leq B \leq A$ y $\sum_{j=1}^n b_{ij} = \alpha$ para todo $i = 1, \ldots, n$ de la siguiente manera. Si $\alpha = 0$, entonces $B = 0$.
Si $\alpha > 0$, entonces hacemos $b_{ij} = \alpha a_{ij} \left(\sum_{k=1}^n a_{ik} \right)^{-1}$. Se sigue de la definición de α que $0 \leq b_{ij} \leq a_{ij}$ para todo $i, j = 1, \ldots, n$ y un simple cálculo muestra que

$\sum_{j=1}^{n} b_{ij} = \alpha$ *para todo* $i = 1, \dots, n$.
Aplicando el lema 3.3.6 a B *y el corolario 3.3.3 obtenemos*

$$\begin{aligned}\min_{1\leq i\leq n} \sum_{j=1}^{n} a_{ij} &= \alpha = \max_{1\leq i\leq n} \sum_{j=1}^{n} b_{ij} = \max_{1\leq i\leq n} \sum_{j=1}^{n} |b_{ij}| = |||B|||_\infty \\ &= \rho(B) \leq \rho(A).\end{aligned}$$

Finalmente, por el teorema 3.2.2,

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

Por lo tanto,

$$\min_{1\leq i\leq n} \sum_{j=1}^{n} a_{ij} \leq \rho(A) \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} a_{ij}. \quad \square$$

Corolario 3.3.4 *Sea* $A \in \mathcal{M}_n$ *tal que* $A \geq 0$. *Si* $\sum_{j=1}^{n} a_{ij} > 0$ *para todo* $i = 1, \dots, n$, *entonces* $\rho(A) > 0$. *En particular,* $\rho(A) > 0$ *si* $A > 0$.
Demostración.
Si $\sum_{j=1}^{n} a_{ij} > 0$ *para todo* $i = 1, \dots, n$, *entonces* $\rho(A) \geq \min_{1\leq i\leq n} \sum_{j=1}^{n} a_{ij} > 0$. $\square$

Podemos generalizar el teorema anterior introduciendo algunos parámetros libres para obtener el siguiente resultado.

Teorema 3.3.3 *Sea* $A \in \mathcal{M}_n$ *tal que* $A \geq 0$. *Entonces para cualquier vector positivo* $x \in \mathbb{C}^n$ *se tiene que*

$$\min_{1\leq i\leq n} \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j \leq \rho(A) \leq \max_{1\leq i\leq n} \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j.$$

Demostración.
Se sigue del lema 2.2.1 y de la definición 2.2.6 que $\rho(S^{-1}AS) = \rho(A)$ *cuando* $S \in \mathcal{M}_n$ *es invertible. Sea* $S = diag(x_1, x_2, \dots, x_n)$, *la matriz diagonal que tiene a* $x_1, x_2, \dots, x_n$ *en su diagonal principal y supongamos que* $x_i > 0$ *para todo* i. *Entonces* $S^{-1}AS = [a_{ij}x_j x_i^{-1}] \geq 0$. *Por el teorema 3.3.2,*

$$\min_{1\leq i\leq n} \sum_{j=1}^{n} a_{ij} x_j x_i^{-1} \leq \rho(S^{-1}AS) = \rho(A) \leq \max_{1\leq i\leq n} \sum_{j=1}^{n} a_{ij} x_j x_i^{-1}.$$

Por lo tanto,

$$\min_{1\leq i\leq n} \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j \leq \rho(A) \leq \max_{1\leq i\leq n} \frac{1}{x_i} \sum_{j=1}^{n} a_{ij} x_j. \quad \square$$

Corolario 3.3.5 *Sean $A \in \mathcal{M}_n$ y $x \in \mathbb{R}^n$ tales que $A \geq 0$ y $x > 0$. Si $\alpha, \beta \geq 0$ satisfacen $\alpha x \leq Ax \leq \beta x$, entonces $\alpha \leq \rho(A) \leq \beta$. Si $\alpha x < Ax$, entonces $\alpha < \rho(A)$; si $Ax < \beta x$, entonces $\rho(A) < \beta$.*
Demostración.
Si $\alpha x \leq Ax$, entonces para todo $i = 1, \ldots, n$, $\alpha x_i \leq \sum_{j=1}^{n} a_{ij}x_j$. Por lo tanto,

$$\alpha \leq \min_{1 \leq i \leq n} \frac{1}{x_i} \sum_{j=1}^{n} a_{ij}x_j \leq \rho(A).$$

Si $\alpha x < Ax$, el razonamiento anterior muestra que existe $\alpha' > \alpha$ tal que $\alpha' x \leq Ax$. En este caso, $\rho(A) \geq \alpha' > \alpha$. Se procede análogamente para las cotas superiores. $\square$

Corolario 3.3.6 *Sea $A \in \mathcal{M}_n$ tal que A es no negativa. Si A tiene un vector propio positivo, entonces su correspondiente valor propio es $\rho(A)$; es decir, si $Ax = \lambda x$, $x > 0$ y $A \geq 0$, entonces $\lambda = \rho(A)$.*
Demostración.
Si $A \geq 0$, $x > 0$ y $Ax = \lambda x$, entonces $\lambda \geq 0$ y $\lambda x \leq Ax \leq \lambda x$. Pero entonces $\lambda \leq \rho(A) \leq \lambda$ por el corolario anterior. $\square$

3.4 El teorema de Perron para matrices positivas

En esta sección presentamos los elegantes resultados obtenidos por Perron en 1907 para matrices positivas. Comenzamos con un resultado técnico del cual se deduce el primer resultado importante para matrices positivas.

Lema 3.4.1 *Sea $A \in \mathcal{M}_n$ tal que $A > 0$. Si $Ax = \lambda x$, $x \neq 0$ y $|\lambda| = \rho(A)$, entonces $A|x| = \rho(A)|x|$ y $|x| > 0$.*
Demostración.
Sea $y = A|x| - \rho(A)|x|$. Dado que $\rho(A)|x| = |\lambda||x| = |\lambda x| = |Ax| \leq |A||x| = A|x|$, se tiene que $y \geq 0$. Ahora, como $|x| \geq 0$ y $|x| \neq 0$, el lema 3.3.2 asegura que $A|x| > 0$. Por otro lado, ya que $A > 0$, el corolario 3.3.4 garantiza que $\rho(A) > 0$. Así, si $y = 0$ entonces $A|x| = \rho(A)|x|$ y $|x| = \rho(A)^{-1}A|x| > 0$. Si $y \neq 0$, de nuevo por el lema 3.3.2, $Ay > 0$. Sea $z = A|x| > 0$, entonces $0 < Ay = A(A|x| - \rho(A)|x|) = Az - \rho(A)z$, de donde $\rho(A)z < Az$. Usando el corolario 3.3.5 obtenemos la contradicción $\rho(A) < \rho(A)$. Luego $y = 0$. $\square$

Teorema 3.4.1 *Sea $A \in \mathcal{M}_n$ tal que A es positiva. Entonces $\rho(A) > 0$ y es un valor propio de A. Más aún, existe un vector positivo x tal que $Ax = \rho(A)x$.*
Demostración.
Existe un valor propio λ de A tal que $|\lambda| = \rho(A) > 0$ y un vector propio $x \neq 0$ asociado a λ. Por el lema anterior, el vector requerido es $|x|$. $\square$

Lema 3.4.2 *Sea $A \in \mathcal{M}_n$ tal que $A > 0$. Si $Ax = \lambda x$, $x \neq 0$ y $|\lambda| = \rho(A)$, entonces existe $\theta \in \mathbb{R}$ tal que $e^{-i\theta}x = |x| > 0$.*
Demostración.
De la hipótesis y el lema anterior sabemos que $|Ax| = |\lambda x| = \rho(A)|x| = A|x|$ y $|x| > 0$. Entonces, para cada $k = 1, \dots, n$,

$$\Big|\sum_{p=1}^{n} a_{kp}x_p\Big| \leq \sum_{p=1}^{n} a_{kp}|x_p| = \sum_{p=1}^{n} |a_{kp}||x_p| = \sum_{p=1}^{n} |a_{kp}x_p|.$$

Como se da la igualdad en la desigualdad del triángulo, los complejos no nulos $a_{kp}x_p$ ($p = 1, \dots, n$) deben estar en el mismo rayo en el plano complejo. Si denotamos su argumento común por θ, entonces $e^{-i\theta}a_{kp}x_p = a_{kp}|x_p| > 0$ para todo $p = 1, \dots, n$. Como $A > 0$, $e^{-i\theta}x_p = |x_p| > 0$ para todo $p = 1, \dots, n$, de dondè $e^{-i\theta}x = |x| > 0$. $\square$

Teorema 3.4.2 *Sea $A \in \mathcal{M}_n$ tal que A es positiva. Entonces $|\lambda| < \rho(A)$ para todo valor propio $\lambda \neq \rho(A)$.*
Demostración.
Tenemos que $|\lambda| \leq \rho(A)$ para todo valor propio λ de A. Supongamos que $|\lambda| = \rho(A)$, $Ax = \lambda x$ y $x \neq 0$. Sea $w = |x| > 0$. Por el lema 3.4.2, $w = e^{-i\theta}x$ para algún $\theta \in \mathbb{R}$. Luego $Aw = Ae^{-i\theta}x = e^{-i\theta}Ax = e^{-i\theta}\lambda x = \lambda e^{-i\theta}x = \lambda w$. Pero entonces $\lambda = \rho(A)$ por el corolario 3.3.6. $\square$

El teorema anterior nos dice que si A es una matriz positiva, entonces $\rho(A)$ es el único valor propio de A de módulo o valor absoluto estrictamente mayor. El siguiente resultado muestra que $\rho(A)$ es un valor propio de multiplicidad geométrica 1. De hecho, veremos más adelante que su multiplicidad algebraica también es 1.

Teorema 3.4.3 *Sea $A \in \mathcal{M}_n$ tal que $A > 0$. Si w y z son vectores distintos de cero tales que $Aw = \rho(A)w$ y $Az = \rho(A)z$, entonces existe $\alpha \in \mathbb{C}$ tal que $w = \alpha z$.*
Demostración.
Por el lema 3.4.2, existen θ_1, $\theta_2 \in \mathbb{R}$ tales que $e^{-i\theta_1}z > 0$ y $e^{-i\theta_2}w > 0$. Sean $p = e^{-i\theta_1}z$, $q = e^{-i\theta_2}w$, $\beta = \min_{1\leq i\leq n} q_i p_i^{-1}$ y $r = q - \beta p$. Dado que $p_i\beta \leq q_i$ para todo $i = 1, \dots, n$, se tiene que $r \geq 0$ y al menos una coordenada de r es 0, así que $r \not> 0$. Por otro lado, por el lema 3.4.1, $Ar = Aq - \beta Ap = \rho(A)q - \beta\rho(A)p = \rho(A)r$; así que si $r \neq 0$, entonces $r = \rho(A)^{-1}Ar > 0$ por el lema 3.3.2. Luego $r = 0$, $q = \beta p$ y $w = \beta e^{-i(\theta_1 - \theta_2)}z$. $\square$

Corolario 3.4.1 *Sea $A \in \mathcal{M}_n$ tal que $A > 0$. Entonces existe un único vector x tal que $Ax = \rho(A)x$, $x > 0$ y $\sum_{i=1}^{n} x_i = 1$.*
Demostración.
Por el lema 3.4.1, existe un vector $w > 0$ tal que $Aw = \rho(A)w$. Como el espacio propio asociado a $\rho(A)$ tiene dimensión 1, el vector $x = (\sum_{k=1}^{n} w_k)^{-1}w$ es el vector requerido. $\square$

Definición 3.4.1 *El* **vector de Perron de una matriz positiva A** *es el único vector propio asociado a $\rho(A)$ que es positivo y cuya suma de componentes es uno. A $\rho(A)$ se le llama la* **raíz de Perron de A.**

En el siguiente capítulo estaremos interesados en el comportamiento de las potencias A^m de una matriz A cuando $m \to \infty$. El siguiente lema es esencial para obtener los resultados necesarios sobre límites de matrices no negativas. Observamos que todas las hipótesis se satisfacen si $A > 0$ y $\lambda = \rho(A)$.

Lema 3.4.3 *Sean $A \in \mathcal{M}_n$, $\lambda \in \mathbb{C}$ y x, y vectores tales que $Ax = \lambda x$, $A^t y = \lambda y$ y $x^t y = 1$. Si definimos $L := xy^t$, entonces:*

a) $Lx = x$

b) $L^m = L$ *para todo* $m \in \mathbb{N}$

c) $A^m L = LA^m = \lambda^m L$ *para todo* $m \in \mathbb{N}$

d) $L(A - \lambda L) = 0$

e) $(A - \lambda L)^m = A^m - \lambda^m L$ *para todo* $m \in \mathbb{N}$

f) todo valor propio distinto de cero de $A - \lambda L$ es valor propio de A.

Si además suponemos que $\lambda \neq 0$ es valor propio de A de multiplicidad geométrica 1, entonces:

g) λ no es valor propio de $A - \lambda L$; es decir, $\lambda I - (A - \lambda L)$ es invertible.

Finalmente, si suponemos que λ es el único valor propio de A con $|\lambda| = \rho(A) > 0$, y si ordenamos los valores propios de A de modo que $|\lambda_1| \leq |\lambda_2| \leq \ldots \leq |\lambda_{n-1}| < |\lambda_n| = |\lambda| = \rho(A)$, entonces:

h) $\rho(A - \lambda L) \leq |\lambda_{n-1}| < \rho(A)$

i) $(\lambda^{-1}A)^m = L + (\lambda^{-1}A - L)^m \longrightarrow L$ *si* $m \longrightarrow \infty$.

Demostración.
Las afirmaciones b), c) y e) se prueban por inducción. (Dejamos las pruebas al lector interesado).

a) Por hipótesis, $Lx = (xy^t)x = x(y^t x) = x(x^t y)^t = x(1)^t = x(1) = x$.

d) De b) y c) se sigue que $L(A - \lambda L) = LA - \lambda L^2 = \lambda L - \lambda L = 0$.

f) Si $\mu \neq 0$ es un valor propio de $A - \lambda L$ y $(A - \lambda L)w = \mu w$ para algún vector $w \neq 0$, entonces por d), $0 = L(A - \lambda L)w = L\mu w = \mu Lw$. Luego $Lw = 0$ y por lo tanto $Aw = (A - \lambda L)w = \mu w$.

g) Supongamos que para algún vector $w \neq 0$, $(A - \lambda L)w = \lambda w$. Por f), w es un vector propio de A. Luego, por hipótesis, $w = \alpha x$ para algún $0 \neq \alpha \in \mathbb{C}$. Pero entonces $\lambda w = (A - \lambda L)w = (A - \lambda L)\alpha x = \alpha\lambda x - \lambda\alpha x = 0$, lo cual es imposible ya que $\lambda \neq 0$, $w \neq 0$.

h) Usando f), g) y la hipótesis, concluimos que $\rho(A - \lambda L) \leq |\lambda_{n-1}| < |\lambda| = \rho(A)$.

i) Ya que $\lambda^{-1}A - L = \lambda^{-1}(A - \lambda L)$, por h),

$$\rho(\lambda^{-1}A - L) = \rho(\lambda^{-1}(A - \lambda L)) = |\lambda^{-1}|.$$

$$\rho(A - \lambda L) = \rho(A)^{-1}\rho(A - \lambda L) < 1.$$

Entonces, usando e) y el teorema 3.2.4, obtenemos que $(\lambda^{-1}A - L)^m = \lambda^{-m}(A^m - \lambda^m L) = (\lambda^{-1}A)^m - L \longrightarrow 0$ cuando $m \longrightarrow \infty$. □

Teorema 3.4.4 *Sea $A \in \mathcal{M}_n$ tal que $A > 0$. Si $Ax = \rho(A)x$, $A^t y = \rho(A)y$, $x > 0$, $y > 0$, $x^t y = 1$ y $L = xy^t$, entonces*

$$\lim_{m \to \infty} (\rho(A)^{-1}A)^m = L$$

Demostración.
Las hipótesis del lema 3.4.3 se cumplen si $\lambda = \rho(A)$, x es el vector de Perron de A y $y = (x^t z)^{-1} z$, donde z es el vector de Perron de A^t. El resultado se sigue de i). □

Corolario 3.4.2 *Sea $A \in \mathcal{M}_n$. Si $A > 0$, entonces $L = \lim_{m \to \infty}(\rho(A)^{-1}A)^m$ es una matriz positiva de rango 1.*
Demostración.
Sean x y y los vectores dados en el teorema anterior. Si

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad y^t = (y_1, y_2, \cdots, y_n)$$

entonces

$$\mathbf{L} = xy^t = \begin{pmatrix} x_1y_1 & x_1y_2 & \cdots & x_1y_n \\ x_2y_1 & x_2y_2 & \cdots & x_2y_n \\ \vdots & \vdots & \ddots & \vdots \\ x_ny_1 & x_ny_2 & \cdots & x_ny_n \end{pmatrix}.$$

Por lo tanto, $L > 0$ y rango $L = 1$. □

Teorema 3.4.5 *Sea $A \in \mathcal{M}_n$ tal que $A > 0$. Entonces $\rho(A)$ es un valor propio de multiplicidad algebraica 1, esto es, $\rho(A)$ es una raíz simple del polinomio característico $p_A(t)$.*

Demostración.
Por el teorema de triangularización de Schur (teorema 2.4.1), podemos escribir $A = U\Delta U^*$, *donde* U *es unitaria,* Δ *es triangular superior con entradas* $\rho, \rho, \ldots, \rho, \lambda_{k+1}, \ldots, \lambda_n$ *en la diagonal principal,* $\rho = \rho(A)$ *es un valor propio de multiplicidad algebraica* $k \geq 1$ *y los valores propios* λ_i *son tales que* $|\lambda_i| < \rho(A)$ *para todo* $i = k+1, \ldots, n$. *Entonces, usando el corolario 3.2.1 y el hecho de que* $|\lambda_i \rho^{-1}| < 1$ *para* $k+1 \leq i \leq n$, *obtenemos*

$$\mathbf{L} = \lim_{m\to\infty} (\rho(A)^{-1}A)^m = \mathbf{U} \lim_{m\to\infty} \begin{pmatrix} 1 & & & & & * \\ & \ddots & & & & \\ & & 1 & & & \\ & & & \lambda_{k+1}\rho^{-1} & & \\ & 0 & & & \ddots & \\ & & & & & \lambda_n\rho^{-1} \end{pmatrix}^m \mathbf{U}^*$$

$$= \mathbf{U} \begin{pmatrix} 1 & & & & & * \\ & \ddots & & & & \\ & & 1 & & & \\ & & & 0 & & \\ & 0 & & & \ddots & \\ & & & & & 0 \end{pmatrix} \mathbf{U}^*$$

donde la entrada igual a 1 en la diagonal principal está repetida k *veces y la entrada igual a cero está repetida* $(n-k)$ *veces. La matriz triangular superior en la última expresión tiene rango* $\geq k$ *y como por el corolario anterior rango* $L = 1$, *entonces* $k = 1$. *Por lo tanto* $\rho(A)$ *es una raíz simple de* $p_A(t)$. □

Resumiendo, obtenemos:

Teorema 3.4.6 (Teorema de Perron) *Si* $A \in \mathcal{M}_n$ *y* $A > 0$, *entonces:*

a) $\rho(A) > 0$

b) $\rho(A)$ *es un valor propio de* A

c) existe un vector $x \in \mathbb{C}^n$ *tal que* $x > 0$ *y* $Ax = \rho(A)x$

d) $\rho(A)$ *es valor propio de* A *algebraicamente (y por lo tanto geométricamente) simple*

e) $|\lambda| < \rho(A)$ *para todo valor propio* $\lambda \neq \rho(A)$, *esto es,* $\rho(A)$ *es el único valor propio de módulo máximo*

f) $[\rho(A)^{-1}A]^m \to L$ *si* $m \to \infty$, *donde* $L = xy^t$, $Ax = \rho(A)x$, $A^t y = \rho(A)y$, $x > 0$, $y > 0$ *y* $x^t y = 1$.

Ejemplo. Consideremos la matriz positiva

$$A = \begin{pmatrix} 2 & 1 \\ 36 & 2 \end{pmatrix}.$$

En este caso $p_A(t) = (t-8)(t+4)$; entonces los valores propios de A son $\lambda_1 = 8$ y $\lambda_2 = -4$, por lo que $\rho(A) = 8 > 0$, $\rho(A)$ es un valor propio algebraicamente simple de A y es el único valor propio de A de módulo máximo.
Vemos que los vectores positivos

$$x = \begin{pmatrix} \frac{1}{7} \\ \\ \frac{6}{7} \end{pmatrix} \quad y = \begin{pmatrix} \frac{7}{2} \\ \\ \frac{7}{12} \end{pmatrix}$$

satisfacen $Ax = \rho(A)x$, $A^t y = \rho(A)y$, $x^t y = 1$ y que

$$L = xy^t = \begin{pmatrix} \frac{1}{2} & \frac{1}{12} \\ \\ 3 & \frac{1}{2} \end{pmatrix}.$$

3.5 Matrices irreducibles

Estamos interesados en generalizar los resultados de la sección anterior a las matrices no negativas que son con las que uno se topa usualmente en la práctica. Sin embargo, sin mas hipótesis adicionales, no podemos ir mas allá del siguiente resultado.

Teorema 3.5.1 *Sea $A \in \mathcal{M}_n$. Si $A \geq 0$, entonces $\rho(A)$ es un valor propio de A y existe $x \geq 0$, $x \neq 0$ tal que $Ax = \rho(A)x$.*
Demostración.
Para cada $\epsilon > 0$, definimos $A(\epsilon) = [a_{ij} + \epsilon] > 0$. Denotamos por $x(\epsilon)$ al vector de Perron de $A(\epsilon)$, entonces $x(\epsilon) > 0$ y

$$\sum_{i=1}^{n} x(\epsilon)_i = 1.$$

Dado que el conjunto de vectores $\{x(\epsilon) \mid \epsilon > 0\}$ está contenido en el compacto $\{x | x \in \mathbb{C}^n, \| x \|_1 \leq 1\}$, existe una sucesión monótona decreciente $\epsilon_1, \epsilon_2, \ldots$ con $\lim_{k\to\infty} \epsilon_k = 0$ para la cual existe $\lim_{k\to\infty} x(\epsilon_k) = x$. Como $x(\epsilon_k) > 0$ para todo $k \in \mathbb{N}$, entonces $x = \lim_{k\to\infty} x(\epsilon_k) \geq 0$; pero

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} \Big(\lim_{k\to\infty} x(\epsilon_k)_i \Big) = \lim_{k\to\infty} \Big(\sum_{i=1}^{n} x(\epsilon_k)_i \Big) = \lim_{k\to\infty} 1 = 1,$$

luego $x \neq 0$. Por el teorema 3.3.1, $\rho\Big(A(\epsilon_k)\Big) \geq \rho\Big(A(\epsilon_{k+1})\Big) \geq \ldots \geq \rho(A)$ para todo $k \in \mathbb{N}$, así que $\Big\{\rho\Big(A(\epsilon_k)\Big)\Big\}_{k\in\mathbb{N}}$ es una sucesión de números reales monótona *decreciente. Por lo tanto existe* $\lim_{k\to\infty} \rho\Big(A(\epsilon_k)\Big) = \rho$ *y* $\rho \geq \rho(A)$. *Ahora,*

$$Ax = \lim_{k\to\infty} A(\epsilon_k)x(\epsilon_k) = \lim_{k\to\infty} \rho\Big(A(\epsilon_k)\Big)x(\epsilon_k) = \lim_{k\to\infty} \rho\Big(A(\epsilon_k)\Big) \lim_{k\to\infty} x(\epsilon_k) = \rho x.$$

Entonces ρ *es un valor propio de* A *asociado a* x, *por lo que* $\rho \leq \rho(A)$. *Luego* $\rho = \rho(A)$. □

Para una matriz no negativa arbitraria A, el valor propio $\rho(A)$ se llama la raíz de Perron de A. Sin embargo, no existe la noción de vector de Perron, ya que el vector propio positivo asociado a $\rho(A)$ cuya suma de coordenadas es 1 puede no estar únicamente determinado, como lo muestra el primero de los ejemplos sencillos que damos a continuación (comparar con el teorema 3.4.6).

Ejemplo 1. Sea A la matriz no negativa $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. Entonces $\rho(A) = 1$ es un valor propio de A de multiplicidad geométrica y algebraica 2 y A tiene a cada vector positivo como vector propio asociado a $\rho(A)$.

Ejemplo 2. Sea A la matriz no negativa $\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$. En este caso $\rho(A) = 0$ es un valor propio de multiplicidad geométrica 1 y multiplicidad algebraica 2 y no existe ningún vector propio positivo asociado a $\rho(A)$.

Ejemplo 3. Sea A la matriz no negativa $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. En este caso los valores propios de A son 1 y -1. Luego $\rho(A) = 1$, pero $|-1| = \rho(A)$. Además, $\lim_{m\to\infty} A^m$ no existe.

Definición 3.5.1 *Una matriz* $P \in \mathcal{M}_n$ *es una* **matriz de permutación** *si exactamente una entrada en cada renglón y cada columna es igual a 1 y todas las demás entradas son cero.*

Ejemplo. La matriz $\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$ es una matriz de permutación.

Claramente, una matriz de permutación P es aquella que se obtiene de la matriz identidad por medio de una sucesión finita de intercambios de renglones o columnas. Luego, una matriz de permutación P es unitaria y por lo tanto ortogonal, esto es, $P^t = P^{-1}$. Más aún, las matrices de permutación más sencillas son las matrices elementales E que se obtienen de la matriz identidad I por un solo intercambio de renglones o columnas. De hecho, observamos que intercambiar los renglones i y j de I equivale a intercambiar las columnas i y j. Luego, la transformación de similitud $A \to E^tAE = E^{-1}AE = EAE$ tiene el efecto de intercambiar los renglones i y j de A, así como las columnas i y j (ver [4] pag. 20 ó [2] pag. 142). Como cualquier matriz de permutación P es un producto de este tipo de matrices elementales, la transformación de similitud $A \to P^tAP = P^{-1}AP$ intercambia o permuta los renglones y las columnas de A en la misma manera.

Definición 3.5.2 *Una matriz $A \in \mathcal{M}_n$ es* **reducible** *si, o bien,*

a) $n = 1$ *y* $A = 0$, *o bien,*

b) $n \geq 2$ *y existen una matriz de permutación* $P \in \mathcal{M}_n$ *y algún entero* r *con* $1 \leq r \leq n-1$ *tales que* $P^tAP = \begin{pmatrix} B & C \\ 0 & D \end{pmatrix}$, *donde* $B \in \mathcal{M}_r$, $D \in \mathcal{M}_{n-r}$, $C \in \mathcal{M}_{r,n-r}$ *y* $0 \in \mathcal{M}_{n-r,r}$ *es la matriz cero.*

Ejemplo 1. La matriz $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$ es claramente reducible.

Ejemplo 2. La matriz $\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ es reducible ya que para la matriz de permutación $\mathbf{P} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$ se tiene que $\mathbf{P^tAP} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$.

Observamos que si $|A| > 0$, entonces A no es reducible (pues no hay forma de obtener una submatriz de A cuyas entradas sean todas cero) y si A es reducible, entonces debe tener al menos $r(n-r)$ ceros, para alguna $r = 1, \ldots, n$.

Definición 3.5.3 *Una matriz $A \in \mathcal{M}_n$ es* **irreducible** *si no es reducible.*

Ejemplo 1. La matriz $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ es claramente irreducible.

Ejemplo 2. La matriz $\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ es irreducible pues no tiene suficientes ceros.

Deseamos caracterizar la irreducibilidad de una matriz en términos de su gráfica dirigida. Para ello introducimos las siguientes definiciones.

Definición 3.5.4 *Una* **gráfica dirigida** Γ *consta de un conjunto finito no vacío* $V = \{P_1, P_2, \ldots, P_n\}$, *cuyos elementos son los vértices de* Γ, *y un subconjunto* $F \subset V \times V$, *cuyos elementos son las flechas o aristas dirigidas de* Γ. *Si* $P_i, P_j \in V$ *y* $(P_i, P_j) \in F$, *entonces hay una flecha* $P_i \xrightarrow{\alpha} P_j$. *El vértice* P_i *se llama el vértice inicial de* α *y el vértice* P_j, *el vértice final de* α.

Definición 3.5.5 *Sean P_i, P_j vértices de una gráfica dirigida Γ. Un* **camino dirigido de longitud m** *de P_i a P_j en Γ es una sucesión de m flechas $\alpha_1, \alpha_2, \ldots, \alpha_m$ que satisface:*

i) el vértice inicial de α_1 es P_i

ii) para $1 \leq k \leq m-1$, el vértice final de α_k es el vértice inicial de α_{k+1}

iii) el vértice final de α_m es P_j.

Definición 3.5.6 *Una gráfica dirigida Γ es* **fuertemente conexa** *si para cada par de vértices distintos P_i, P_j en Γ hay un camino dirigido de longitud finita de P_i a P_j en Γ.*

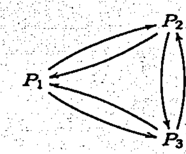
Definición 3.5.7 *La* **gráfica dirigida** de $A \in \mathcal{M}_n$, *denotada por $\Gamma(A)$, es la gráfica dirigida con vértices $\{P_1, P_2, \ldots, P_n\}$ tal que hay una flecha en $\Gamma(A)$ de P_i a P_j si y sólo si $a_{ij} \neq 0$.*

Ejemplo 1. Si A $= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ entonces

$\Gamma(A) =$ 

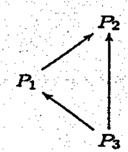
es fuertemente conexa.

Ejemplo 2. Si A $= \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ entonces

$\Gamma(A) =$ 

es fuertemente conexa.

Ejemplo 3. Si A $= \begin{pmatrix} 0 & 2 & 0 \\ 0 & 0 & 0 \\ 3 & 4 & 0 \end{pmatrix}$ entonces

$\Gamma(A) =$ 

no es fuertemente conexa pues, por ejemplo, no hay forma de ir del vértice P_2 al vértice P_3.

Dado que queremos extender el teorema de Perron a cierta clase de matrices no negativas, en lo sucesivo nos restringiremos a este tipo de matrices.

Teorema 3.5.2 *Sea $A \in \mathcal{M}_n$ no negativa y sean P_i, P_j vértices en $\Gamma(A)$. Hay un camino dirigido de longitud m de P_i a P_j en $\Gamma(A)$ si y sólo si $(A^m)_{ij} \neq 0$.*
Demostración.
Haremos la prueba por inducción sobre m. El enunciado es válido para $m = 1$. Supongamos que el enunciado ha sido probado para $m = q$. Entonces

$$(A^{q+1})_{ij} = \sum_{k=1}^{n} (A^q)_{ik} A_{kj} \neq 0$$

si y sólo si para al menos un valor de k, $(A^q)_{ik}$ y A_{kj} son ambos distintos de cero. Esto es equivalente a tener un camino de P_i a P_k de longitud q y uno de P_k a P_j de longitud 1, y esto pasa si y sólo si hay un camino de P_i a P_j de longitud $q + 1$. □

Corolario 3.5.1 *Sea $A \in \mathcal{M}_n$ no negativa. Entonces $A^m > 0$ si y sólo si para cada par de vértices P_i, P_j en $\Gamma(A)$ hay un camino dirigido de longitud m de P_i a P_j en $\Gamma(A)$.* □

Corolario 3.5.2 *Sea $A \in \mathcal{M}_n$ no negativa. Entonces $\Gamma(A)$ es fuertemente conexa si y sólo si $(I + A)^{n-1} > 0$.*
Demostración.
La matriz $(I + A)^{n-1} = I + (n-1)A + \binom{n-1}{2}A^2 + \ldots + \binom{n-1}{n-1}A^{n-1} > 0$ si y sólo si para cada par (i, j) de índices con $i \neq j$ al menos uno de los términos $A, A^2, \ldots, A^{n-1}$ tiene una entrada positiva (i, j). Pero por el teorema 3.5.2 esto pasa si y sólo si existe un camino dirigido en $\Gamma(A)$ de P_i a P_j. Esto es equivalente a que $\Gamma(A)$ sea fuertemente conexa. □

Teorema 3.5.3 *Una matriz no negativa $A \in \mathcal{M}_n$ es irreducible si y sólo si*

$$(I + A)^{n-1} > 0.$$

Demostración.
Probaremos que A es reducible si y sólo si $(I + A)^{n-1}$ tiene al menos una entrada igual a cero. Supongamos que A es reducible y que para alguna matriz de permutación P tenemos que

$$A = P^t \begin{pmatrix} B & C \\ 0 & D \end{pmatrix} P = P^t \widetilde{A} P,$$

donde $B, C, 0$ y D son las matrices de bloque dadas en la definición 3.5.2. Observemos que $\widetilde{A}^2, \widetilde{A}^3, \ldots, \widetilde{A}^{n-1}$ tienen todas el mismo bloque de ceros de $(n-r) \times r$

en la esquina inferior izquierda como $\widetilde{A}$. Entonces

$$\begin{aligned}(I+A)^{n-1} &= (I+P^t\widetilde{A}P)^{n-1} = (P^t[I+\widetilde{A}]P)^{n-1} = P^t(I+\widetilde{A})^{n-1}P \\ &= P^t\Big[I+(n-1)\widetilde{A}+\binom{n-1}{2}\widetilde{A}^2+\ldots+\binom{n-1}{n-1}\widetilde{A}^{n-1}\Big]P\end{aligned}$$

y todos los términos entre los corchetes tienen un bloque de ceros de $(n-r)\times r$ en la esquina inferior izquierda. Luego $(I+A)^{n-1}$ es reducible y no puede tener todas las entradas distintas de cero.
Inversamente, supongamos que para algún par de índices (p,q) con $p\neq q$ la entrada (p,q) de $(I+A)^{n-1}$ es cero. Entonces sabemos que no hay un camino dirigido en $\Gamma(A)$ de P_p a P_q. Partimos el conjunto de vértices $\{P_1,\ldots,P_n\}$ de $\Gamma(A)$ en dos subconjuntos, a saber, sea $S_1=\{P_i \mid P_i=P_q$ o hay un camino en $\Gamma(A)$ de P_i a $P_q\}$ y sea S_2 tal que S_2 contiene todos los vértices de $\Gamma(A)$ que no están en S_1. Entonces $S_1\cup S_2=\{P_1,\ldots,P_n\}$ y $P_q\in S_1\neq\emptyset$, luego $S_2\neq\{P_1,\ldots,P_n\}$. Si hubiera un camino de algún vértice P_i de S_2 a algún vértice P_j de S_1, por la definición de S_1, habría un camino de P_i a P_q y entonces P_i estaría en S_1. Luego no puede haber caminos de ningún vértice de S_2 a ningún vértice de S_1. Ahora, renumerando los vértices de modo que $S_1=\{\widetilde{P}_1,\ldots,\widetilde{P}_r\}$ y $S_2=\{\widetilde{P}_{r+1},\ldots,\widetilde{P}_n\}$ y tomando P como la matriz de permutación correspondiente a esta renumeración obtenemos

$$\widetilde{A}=P^tAP=\begin{pmatrix} B & C \\ 0 & D\end{pmatrix},\quad B\in\mathcal{M}_r,\quad 0\in\mathcal{M}_{n-r\times r}.$$

Por lo tanto A es reducible. $\square$

Resumiendo obtenemos:

Teorema 3.5.4 *Para una matriz no negativa $A\in\mathcal{M}_n$ son equivalentes las siguientes afirmaciones:*

a) *A es irreducible.*

b) $(I+A)^{n-1}>0$.

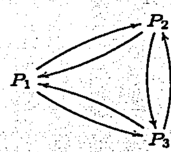
c) *$\Gamma(A)$ es fuertemente conexa.*

Ejemplo 1. La matriz $\mathbf{A}=\begin{pmatrix}0&1&1\\1&0&1\\1&1&0\end{pmatrix}$ es irreducible pues

$$(\mathbf{I}+\mathbf{A})^2=\begin{pmatrix}3&3&3\\3&3&3\\3&3&3\end{pmatrix}>0$$

y

$$\Gamma(A) =$$



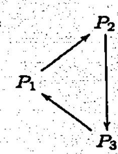
es fuertemente conexa.

Ejemplo 2. La matriz $\mathbf{B} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ es irreducible pues

$$(\mathbf{I} + \mathbf{B})^2 = \begin{pmatrix} 1 & 2 & 1 \\ 1 & 1 & 2 \\ 2 & 1 & 1 \end{pmatrix} > 0$$

y

$$\Gamma(B) =$$



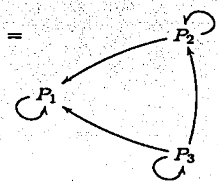
es fuertemente conexa.

Ejemplo 3. La matriz $\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$ es reducible pues

$$(\mathbf{I} + \mathbf{C})^2 = \begin{pmatrix} 4 & 0 & 0 \\ 4 & 4 & 0 \\ 5 & 4 & 4 \end{pmatrix} \not> 0$$

y

$$\Gamma(C) =$$



no es fuertemente conexa.

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

Comenzamos esta sección con los siguientes dos resultados sencillos.

Lema 3.6.1 *Sean $A \in \mathcal{M}_n$ y $\lambda_1, \lambda_2, \ldots, \lambda_n$ los valores propios de A (incluyendo multiplicidades). Entonces $\lambda_1+1, \lambda_2+1, \ldots, \lambda_n+1$ son los valores propios de $I+A$ y $\rho(I+A) \leq 1+\rho(A)$. Si $A \geq 0$ entonces $\rho(I+A) = 1+\rho(A)$.*
Demostración.
Si λ es un valor propio de A de multiplicidad algebraica k, entonces λ es una raíz de multiplicidad k del polinomio característico $p_A(t) = det(tI-A)$. Pero entonces $\lambda+1$ es una raíz de multiplicidad k de $p_{A+I}(s) = det[sI-(A+I)]$ ya que $det(tI-A) = det(tI-A+I-I) = det[(t+1)I-(A+I)]$. Luego λ_1+1, λ_2+1, ..., λ_n+1 son los valores propios de $I+A$ y por lo tanto

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

Ahora, si $A \geq 0$, por el teorema 3.5.1, existe $x \geq 0$, $x \neq 0$ tal que $Ax = \rho(A)x$; así $(I+A)x = x+Ax = x+\rho(A)x = (1+\rho(A))x$. Por lo tanto $\rho(I+A) = 1+\rho(A)$, si $A \geq 0$. □

Lema 3.6.2 *Sea $A \in \mathcal{M}_n$ tal que $A \geq 0$ y $A^k > 0$ para alguna $k \geq 1$, entonces $\rho(A)$ es un valor propio de A algebraicamente simple.*
Demostración.
Sabemos que si $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los valores propios de A, entonces $\lambda_1^k, \lambda_2^k, \ldots, \lambda_n^k$ son los valores propios de A^k (teorema 2.4.2) y que $\rho(A)$ es un valor propio de A (teorema 3.5.1). Así que si $\rho(A)$ es un valor propio múltiple de A, entonces $\rho(A)^k = \rho(A^k)$ (corolario 2.4.1) es un valor propio múltiple de A^k, lo cual no es posible pues $\rho(A^k)$ es un valor propio simple de A^k según el teorema 3.4.5. □

Ahora podemos ver qué tanto del teorema de Perron se generaliza al caso de las matrices no negativas irreducibles. La generalización se debe a Frobenius.

Teorema 3.6.1 (Teorema de Perron-Frobenius) *Sea $A \in \mathcal{M}_n$. Si A es irreducible y $A \geq 0$, entonces:*

a) $\rho(A) > 0$

b) $\rho(A)$ *es un valor propio de* A

c) existe un vector $x \in \mathbb{C}^n$ tal que $x > 0$ y $Ax = \rho(A)x$

d) $\rho(A)$ es valor propio de A algebraicamente (y por lo tanto geométricamente) simple.

Demostración.

a) Si A es irreducible entonces A no puede tener un renglón de ceros, de donde $\sum_{j=1}^{n} a_{ij} > 0$ para todo $i = 1, \ldots, n$. Por lo tanto, por el corolario 3.3.4, $\rho(A) > 0$.

b) Si $A \geq 0$ entonces $\rho(A)$ es un valor propio de A por el teorema 3.5.1.

c) El mismo teorema garantiza que existe $x \geq 0$, $x \neq 0$ tal que $Ax = \rho(A)x$. Entonces $(I + A)x = (1 + \rho(A))x$, de donde $(I + A)^{n-1}x = (1 + \rho(A))^{n-1}x$. Ahora, como A es irreducible, $(I + A)^{n-1} > 0$ por el teorema 3.5.3. Luego, por el lema 3.3.2, tenemos que $x = (1 + \rho(A))^{1-n}(I + A)^{n-1}x > 0$.

d) Si $\rho(A)$ es un valor propio múltiple de A, entonces $1 + \rho(A) = \rho(I + A)$ es un valor propio múltiple de $(I + A)$ según el lema 3.6.1. Pero $(I + A) \geq 0$ e $(I + A)^{n-1} > 0$ por el teorema 3.5.3, entonces $1 + \rho(A)$ debe ser un valor propio simple de $(I + A)$ por el lema 3.6.2, y por lo tanto $\rho(A)$ es un valor propio de A algebraica y geométricamente simple. □

Definición 3.6.1 El vector de Perron de una matriz irreducible no negativa A *es el único vector propio positivo asociado a $\rho(A)$ cuya suma de componentes es 1.*

Cuando $A \in \mathcal{M}_n$ es positiva, el teorema de Perron afirma que $\rho(A)$ es el único valor propio de A de módulo máximo. Cuando A es no negativa, puede haber más de un valor propio de módulo máximo (ver ejemplo 2, pag. 51), pero en este caso A debe tener una forma especial y estos valores propios deben aparecer en un patrón muy regular, como lo muestran los siguientes resultados, cuyas pruebas omitiremos.

Teorema 3.6.2 *Sea $A \in \mathcal{M}_n$ irreducible y no negativa. Supongamos que el conjunto $S = \{\lambda_n = \rho(A), \lambda_{n-1}, \ldots, \lambda_{n-k+1}\}$ de valores propios de módulo máximo $\rho(A)$ tiene exactamente k elementos distintos. Entonces cada valor propio $\lambda_i \in S$ tiene multiplicidad algebraica 1 y*

$$S = \{e^{\frac{2\pi i p}{k}}\rho(A) : p = 0, 1, \ldots, k-1\},$$

esto es, los valores propios de módulo máximo son precísamente las k-ésimas raíces de la unidad por $\rho(A)$. Más aún, si λ es cualquier valor propio de A, entonces $e^{\frac{2\pi i p}{k}}\lambda$ es un valor propio para todo $p = 0, 1, \ldots, k-1$. □

Según este teorema, si A es irreducible no negativa y tiene $k > 1$ valores propios de módulo máximo, entonces cada valor propio no nulo de A cae sobre una circunferencia con centro en 0 en $\mathbb{C}$ que pasa por exactamente k valores propios de A, todos igualmente espaciados alrededor de la circunferencia. En particular, k debe ser un divisor del número de valores propios no nulos de A.

Teorema 3.6.3 *Si A es una matriz irreducible no negativa y A tiene $k > 1$ valores propios de módulo máximo, entonces existe una matriz de permutación P tal que*

$$\mathbf{PAP^t} = \begin{pmatrix} 0 & A_{1,2} & 0 & 0 \\ \vdots & 0 & \ddots & 0 \\ 0 & 0 & \ddots & A_{k-1,k} \\ A_{k,1} & 0 & \cdots & 0 \end{pmatrix}$$

donde los k bloques de ceros de la diagonal principal son matrices cuadradas y los bloques $A_{i,j}$ mostrados no son necesariamente cero. En particular, todas las entradas a_{ii} de la diagonal principal de A son cero.

Como consecuencia de este teorema obtenemos el siguiente resultado.

Corolario 3.6.1 *Si $A \in \mathcal{M}_n$ es irreducible no negativa y alguna entrada en la diagonal principal de A es no nula, entonces $\rho(A)$ es el único valor propio de A de módulo máximo.* $\square$

Concluimos esta sección con algunos ejemplos. El primero muestra que la afirmación recíproca del corolario anterior es falsa y el tercero muestra que para una matriz no negativa reducible los valores propios de módulo máximo difieren de los del teorema 3.6.2.

Ejemplo 1. Consideremos la matriz irreducible no negativa $\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$ cuyo polinomio característico se factoriza como $p_A(x) = (x+1)^2(x-2)$. Luego $\rho(A) = 2$ es el único valor propio de A de módulo máximo. Más aún, el vector de Perron de A es $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ y $\lim_{m\to\infty}(\rho(A)^{-1}A)^m = \frac{1}{3}\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}$.

Ejemplo 2. Consideremos la matriz irreducible no negativa $\mathbf{B} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$ cuyo polinomio característico es $p_B(x) = x^3 - 1$. Luego $\rho(B) = 1$ y las tres raíces cúbicas de la unidad, 1, $-\frac{1}{2} + \frac{\sqrt{3}}{2}i$ y $-\frac{1}{2} - \frac{\sqrt{3}}{2}i$, son los tres valores propios de B de módulo máximo. El vector de Perron de B es $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ y $\lim_{m\to\infty}(\rho(B)^{-1}B)^m$ no existe.

Ejemplo 3. Sean C la matriz irreducible no negativa $\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, B la matriz del ejemplo anterior y $\mathbf{A} = \begin{pmatrix} C & 0 \\ 0 & B \end{pmatrix}$. El polinomio característico de A es $p_A(x) = p_C(x)p_B(x) = (x^2-1)(x^3-1)$. Luego los valores propios de A son ± 1, $-\frac{1}{2} + \frac{\sqrt{3}}{2}i$

y $-\frac{1}{2}-\frac{\sqrt{3}}{2}i$, todos ellos de módulo máximo. En este caso, $\rho(A)=1$ es de multiplicidad algebraica 2, no hay vector de Perron y $\lim_{m\to\infty}(\rho(A)^{-1}A)^m$ no existe.

3.7 Matrices primitivas

Como veremos en el siguiente capítulo, el resultado del teorema de Perron que se aplica mas frecuentemente es el inciso f) o teorema 3.4.4. De hecho, la única hipótesis que falta para poder aplicar el lema 3.4.3 al caso de las matrices irreducibles no negativas es la condición de que el radio espectral sea el único valor propio de módulo máximo. Ya que $A=\begin{pmatrix}0&1\\1&0\end{pmatrix}$ es un ejemplo de una matriz irreducible no negativa con dos valores propios de módulo máximo y para la cual $\lim_{m\to\infty}A^m$ no existe, vemos que necesitamos restringir aún más la clase de matrices con la que estamos trabajando. Desde luego, el procedimiento mas económico es suponer exactamente lo que necesitamos.

Definición 3.7.1 *Una matriz no negativa $A\in\mathcal{M}_n$ es una* **matriz primitiva** *si es irreducible y tiene sólo un valor propio de módulo máximo.*

El siguiente resultado acerca del límite se obtiene ahora directamente del lema 3.4.3 con la misma prueba que para el teorema 3.4.4.

Teorema 3.7.1 *Sea $A\in\mathcal{M}_n$ no negativa y primitiva. Entonces*

$$\lim_{m\to\infty}(\rho(A)^{-1}A)^m=L>0$$

donde $L=xy^t$, $Ax=\rho(A)x$, $A^ty=\rho(A)y$, $x>0$, $y>0$ y $x^ty=1$. □

Ejemplo. La matriz irreducible no negativa $A=\begin{pmatrix}0&1&1\\1&0&1\\1&1&0\end{pmatrix}$ considerada en el ejemplo 1 de la sección anterior es primitiva. El $\lim_{m\to\infty}(\rho(A)^{-1}A)^m$ se calculó usando el teorema 3.7.1.

Finalmente, hemos generalizado todo el teorema de Perron para matrices positivas a la clase de matrices primitivas no negativas. Sin embargo, queda aún el problema de determinar cuándo una matriz dada es primitiva sin recurrir al cálculo explícito de sus valores propios. La siguiente caracterización de primitividad permite dar varios criterios.

Teorema 3.7.2 *Sea $A\in\mathcal{M}_n$ no negativa. Entonces A es primitiva si y sólo si $A^m>0$ para algún $m\geq 1$.*
Demostración.
Si $A\geq 0$ y $A^m>0$, entonces de cada vértice P_i de la gráfica dirigida $\Gamma(A)$ de A

a cada vértice P_j hay un camino dirigido de longitud m, de acuerdo con el corolario 3.5.1. Luego, $\Gamma(A)$ es fuertemente conexa y por el teorema 3.5.4, A es irreducible. Por el lema 3.6.2, $\rho(A)$ es un valor propio de A algebraicamente simple. Sean $\lambda_1, \lambda_2, \ldots, \lambda_n = \rho(A)$ los valores propios de A. Como en la demostración del lema 3.6.2, tenemos que $\lambda_1^m, \lambda_2^m, \ldots, \lambda_n^m = \rho(A)^m = \rho(A^m)$ son los valores propios de A^m. Aplicando el teorema de Perron a A^m, tenemos que $\rho(A^m)$ es un valor propio de A^m algebraicamente simple y es el único valor propio de A^m de módulo máximo, de tal forma que los valores propios $\lambda_1^m, \lambda_2^m, \ldots, \lambda_{n-1}^m$ de A^m satisfacen $|\lambda_i^m| < \rho(A^m)$ para $i = 1, \ldots, n-1$. Por lo tanto, $|\lambda_i| < \rho(A)$ para $i = 1, \ldots, n-1$, de donde $\rho(A)$ es el único valor propio de A de módulo máximo y consecuentemente, A es primitiva.
Inversamente, si A es primitiva, entonces $\lim_{m\to\infty}(\rho(A)^{-1}A)^m = L > 0$ según el teorema 3.7.1, por lo que para algún $m \geq 1$ se debe tener que $(\rho(A)^{-1}A)^m > 0$. $\square$

Si A es una matriz primitiva, el mínimo entero positivo k tal que $A^k > 0$ se llama el índice de primitividad de A y usualmente se denota por $\gamma(A)$. Los siguientes resultados, cuyas pruebas omitiremos, proporcionan cotas superiores para $\gamma(A)$.

Teorema 3.7.3 *Sea $A \in \mathcal{M}_n$ no negativa. Entonces A es primitiva si y sólo si $A^{n^2-2n+2} > 0$.* $\square$

Teorema 3.7.4 *Sea $A \in \mathcal{M}_n$ no negativa tal que todas sus entradas en la diagonal principal son positivas. Entonces A es primitiva si y sólo si $A^{n-1} > 0$.* $\square$

Teorema 3.7.5 *Sea $A \in \mathcal{M}_n$ no negativa con d entradas positivas en la diagonal principal, donde $1 \leq d \leq n$. Entonces A es primitiva si y sólo si $A^{2n-d-1} > 0$.* $\square$

Ejemplo 1. Sea $\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$, entonces $2 = \gamma(A) = n^2 - 2n + 2 = 2n - d - 1$.

Ejemplo 2. Sea $\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$, entonces $3 = \gamma(A) < 4 = 2n - d - 1 < 5 = n^2 - 2n + 2$.

Ejemplo 3. Sea $\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, entonces $2 = \gamma(A) < 5 = n^2 - 2n + 2$.

Ejemplo 4. Sea $\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$, entonces $5 = \gamma(A) = n^2 - 2n + 2$, lo que muestra que la cota $\gamma(A) \leq n^2 - 2n + 2$ es la mejor cota superior posible para matrices cuya diagonal principal es nula.

Comentarios finales: Si se desea comprobar si una matriz no negativa $A \in \mathcal{M}_n$ es irreducible o primitiva y n es bastante grande, es recomendable usar el teorema 3.5.3 o el teorema 3.7.3, respectivamente. En cualquier caso, el número de multiplicaciones a efectuar se reduce considerablemente si la matriz en cuestión ($I + A$ ó A, respectivamente) se eleva al cuadrado repetidamente, hasta que la potencia resultante exceda el valor crítico ($n - 1$ ó $n^2 - 2n + 2$, respectivamente). Esto se debe al hecho de que si $A \geq 0$ y $A^k > 0$ para algún $k \in \mathbb{N}$, entonces $A^m > 0$ para todo $m \geq k$.

Capítulo 4

Cadenas de Markov

4.1 Procesos estocásticos y cadenas de Markov

Un proceso estocástico tiene como finalidad predecir el estado de un objeto que esté restringido a estar exactamente en uno de ciertos posibles estados en un instante dado cualquiera, pero que cambia de estado de alguna manera aleatoria.
Desde un punto de vista no matemático, un proceso estocástico es cualquier proceso probabilístico, es decir, cualquier proceso que se desarrolla en el tiempo y es controlado por leyes de probabilidad.

Definición 4.1.1 *Un* **proceso estocástico** *está compuesto por un espacio de estados E, un conjunto de índices T y las relaciones de dependencia de un conjunto de variables aleatorias X_t ($t \in T$).*
El **espacio de estados E** *es el conjunto de valores que toma cada variable aleatoria X_t.*
El **conjunto de índices T** *se asocia con el tiempo y puede ser $T = (0, 1, \dots)$ en cuyo caso decimos que el proceso es de tiempo discreto, o bien $T = [0, \infty)$ en donde decimos que el proceso es de tiempo continuo.*

Normalmente, la probabilidad de que un objeto se encuentre en un estado particular en un instante dado depende de factores como:

1. el estado en cuestión
2. el instante en cuestión
3. los estados anteriores en los que estuvo el objeto
4. los estados en que se encuentran otros objetos.

Sin embargo, si la probabilidad de que un objeto que está en un cierto estado cambie a otro estado distinto depende únicamente de los dos estados (y no de otros factores), entonces el proceso estocástico se llama proceso de Markov. Esto se expresa formalmente en la siguiente definición.

Definición 4.1.2 *Un* **proceso de Markov** *es un proceso estocástico con la propiedad de que, dado el valor de X_t, los valores de X_s para $s > t$ son independientes de los valores de X_u para $u < t$.*

Definición 4.1.3 *Una* **cadena de Markov** *es un proceso de Markov en el que el número de estados posibles es finito.*

En términos formales, la propiedad de Markov se expresa por

$$Pr\{X_{n+1} = j | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}, X_n = i\} = Pr\{X_{n+1} = j | X_n = i\}$$

para todo $n \in T$ y todo $i_k \in E$.
Frecuentemente el espacio de estados de una cadena de Markov se etiqueta con los enteros no negativos $\{0, 1, 2, \ldots\}$ y se dice que X_n se encuentra en el estado i si $X_n = i$.

Definición 4.1.4 *La* **probabilidad de transición de un paso** *es la probabilidad de que X_{n+1} esté en el estado i dado que X_n está en el estado j y se denota por $P_{ij}^{n,n+1}$, esto es,*

$$P_{ij}^{n,n+1} = Pr\{X_{n+1} = i | X_n = j\}.$$

En general, las probabilidades de transición son funciones no sólo de los estados inicial y final, sino también del tiempo de transición, es decir, de n. Cuando estas probabilidades son independientes de la variable de tiempo n, decimos que la cadena de Markov tiene **probabilidades de transición estacionarias.**
Dado que la gran mayoría de las cadenas de Markov tienen probabilidades de transición estacionarias, limitamos nuestra discusión a este caso y entonces $P_{ij}^{n,n+1} = P_{ij}$ es la probabilidad de que el estado cambie de j a i en un paso. Podemos arreglar estos números en una matriz

$$\mathbf{P} = \begin{pmatrix} P_{00} & P_{01} & P_{02} & \cdots \\ P_{10} & P_{11} & P_{12} & \cdots \\ P_{20} & P_{21} & P_{22} & \cdots \\ \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

y decimos que P es la **matriz de Markov ó matriz de probabilidades de transición de la cadena de Markov.** La $(j+1)$-ésima columna de P es la distribución de probabilidad de los valores de X_{n+1} dado que $X_n = j$. Como el número de estados es finito, entonces P es una matriz cuadrada finita cuyo número de renglones es igual al número r de estados. Claramente, las cantidades P_{ij} satisfacen las condiciones:

a) $P_{ij} \geq 0$ para todo $i, j = 0, 1, 2, \ldots$

b) $\sum_{i=0}^{r} P_{ij} = 1$ para todo $i = 0, 1, 2, \ldots$

Un proceso de Markov está completamente definido una vez que su matriz de probabilidades de transición P y su estado inicial X_0 son especificados.
El análisis de una cadena de Markov consiste esencialmente en el cálculo de las probabilidades de las posibles realizaciones del proceso y para ello es primordial el cálculo de las matrices de probabilidades de transición de n pasos $P^n = (P^n)_{ij}$, donde $(P^n)_{ij}$ denota la probabilidad de que el proceso vaya del estado j al estado i en n transiciones. Formalmente $(P^n)_{ij} = Pr\{X_{m+n} = i | X_m = j\}$. Como estamos tratando solamente con procesos que tienen probabilidades de transición estacionarias, entonces $(P^n)_{ij}$ no depende de m.
La propiedad de Markov nos permite expresar $(P^n)_{ij}$ en términos de P_{ij} como

$$(P^n)_{ij} = \sum_{k=0}^{r} P_{ik}(P^{n-1})_{kj},$$

donde

$$(P^0)_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Vemos que el cálculo de $(P^n)_{ij}$ se puede hacer con la multiplicación de matrices, es decir, $P^n = P \cdot P^{n-1}$. En otras palabras, las probabilidades de transición $(P^n)_{ij}$ son las entradas de la matriz P^n, la n-ésima potencia de P.

Ejemplo 1. El modelo de la urna de Ehrenfest. Este modelo matemático clásico describe la difusión de partículas a través de una membrana. Supongamos que se tienen dos contenedores A con k bolas y B con $2a - k$ bolas. Una bola es seleccionada aleatoriamente (todas las selecciones son igualmente probables) del total de las $2a$ bolas y movida al otro contenedor. Cada selección genera una transición del proceso. Claramente, las bolas fluctúan entre los dos contenedores con un movimiento promedio de la urna con mayor concentración a la otra.
Sea Y_n el número de bolas en la urna A en el estado n y definamos $X_n = Y_n - a$. Entonces $\{X_n\}$ es una cadena de Markov sobre los estados $i = -a, -a + 1, \ldots, -1, 0, 1, \ldots, a$. Calculemos las probabilidades de transición P_{ij}. Vemos que j puede tomar los valores $i + 1$, $i - 1$.
Si $j = i + 1$, entonces $X_n = i$, $X_{n+1} = i + 1$, $Y_n = X_n + a = i + a$, $Y_{n+1} = X_{n+1} + a = i + 1 + a$. P_{ij} es la probabilidad de cambiar de $i + a$ bolas en el contenedor A a tener $i + 1 + a$, esto es, $P_{ij} = \frac{2a-(i+a)}{2a} = \frac{a-i}{2a}$.
Si $j = i - 1$, entonces $X_n = i$, $X_{n+1} = i - 1$, $Y_n = i + a$, $Y_{n+1} = i - 1 + a$. P_{ij} es la probabilidad de cambiar de $i + a$ bolas en el contenedor A a tener $i - 1 + a$. En este caso, $P_{ij} = \frac{i+a}{2a}$.
Resumiendo

$$P_{ij} = \begin{cases} \frac{a-i}{2a} & \text{si } j = i + 1, \\ \frac{a+i}{2a} & \text{si } j = i - 1, \\ 0 & \text{en cualquier otro caso.} \end{cases}$$

Ejemplo 2. Una cadena de Markov para una cola discreta. Los consumidores llegan por un servicio y se forman en una fila de espera. Durante cada periodo de tiempo, un solo consumidor es atendido cuando hay alguien en la fila. Si no hay consumidores en la fila, entonces no se realiza ningún servicio durante este periodo. (Por ejemplo, un sitio de taxis en el cual un taxi llega en intervalos fijos de tiempo para dar el servicio. Si no hay nadie esperando, entonces el taxi parte inmediatamente.) Suponemos que el número de consumidores que llegan durante el n-ésimo periodo es una variable aleatoria ξ_n cuya distribución es independiente del periodo y está dada por $Pr\{k$ consumidores lleguen en un periodo de servicio$\} = Pr\{\xi_n = k\} = a_k$ para $k = 0, 1, \ldots$, donde $a_k \geq 0$ y $\sum_{k=0}^{\infty} a_k = 1$. También asumimos que $\xi_1, \xi_2, \ldots$ son variables aleatorias independientes. El estado del sistema al principio de cada periodo se define como el número de consumidores en la fila esperando el servicio. Si el presente estado es i, entonces después del lapso de un periodo el estado es

$$j = \begin{cases} i - 1 + \xi & \text{si } i \geq 1, \\ \xi & \text{si } i = 0, \end{cases}$$

donde ξ es el número de nuevos consumidores que han llegado en este periodo mientras un solo consumidor fue atendido. En términos de variables aleatorias, podemos expresar el proceso como

$$X_{n+1} = (X_n - 1)^+ + \xi_n,$$

donde $Y^+ = max\{Y, 0\}$. Así, la matriz de probabilidades de transición se puede calcular fácilmente y obtenemos

$$\mathbf{P} = \begin{pmatrix} a_0 & a_0 & 0 & 0 & \cdots \\ a_1 & a_1 & a_0 & 0 & \cdots \\ a_2 & a_2 & a_1 & a_0 & \cdots \\ a_3 & a_3 & a_2 & a_1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

4.2 Aplicación del teorema de Perron-Frobenius a las cadenas de Markov

Lema 4.2.1 *Si λ es un valor propio de una matriz de probabilidades de transición A, entonces $|\lambda| \leq 1$.*

Demostración.

Sea λ un valor propio de A. Por el teorema 3.2.2, $\rho(A) \leq |||A|||_1 = 1$ y como $|\lambda| \leq \rho(A)$ entonces $|\lambda| \leq 1$. □

Teorema 4.2.1 *Toda matriz de probabilidades de transición tiene a 1 como valor propio.*

Demostración.

Sea $A \in \mathcal{M}_n$ una matriz de probabilidades de transición y sea $u = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \in \mathbb{R}^n$ *el vector columna en donde cada coordenada es* 1*. Entonces* $A^t u = u$*, de donde* 1 *es un valor propio de* A^t *y como* A *y* A^t *tienen los mismos valores propios entonces* 1 *es valor propio de* A. $\square$

Este teorema muestra que se ha alcanzado la cota superior en el lema anterior; es decir, si A es una matriz de probabilidades de transición, entonces $\rho(A) = 1$.

Definición 4.2.1 *Si alguna potencia de una matriz de probabilidades de transición contiene únicamente elementos positivos, entonces la matriz se llama* **matriz regular de transición.**

El teorema 3.7.2 nos dice que si A es una matriz regular de transición, entonces A es primitiva.

Definición 4.2.2 *Un vector* $v \in \mathbb{R}^n$ *es un* **vector de probabilidad** *si* $v \geq 0$ *y* $\sum_{i=1}^{n} v_i = 1$.

Lema 4.2.2 *El producto de dos matrices de probabilidades de transición en* $\mathcal{M}_n$ *es una matriz de probabilidades de transición en* $\mathcal{M}_n$*. En particular, cualquier potencia de una matriz de probabilidades de transición es una matriz de probabilidades de transición.*

Demostración.

Sean A, B *matrices de probabilidades de transición en* $\mathcal{M}_n$*. Entonces* $(AB)_{ij} = \sum_{k=1}^{n} A_{ik}B_{kj} \geq 0$ *dado que* $A_{ik}, B_{kj} \geq 0$ *para* $1 \leq k \leq n$*. Además* $\sum_{i=1}^{n}(AB)_{ij} = \sum_{i=1}^{n}\left(\sum_{k=1}^{n} A_{ik}B_{kj}\right) = \sum_{k=1}^{n}\left(\sum_{i=1}^{n} A_{ik}B_{kj}\right) = \sum_{k=1}^{n}\left(B_{kj}\sum_{i=1}^{n} A_{ik}\right) = \sum_{k=1}^{n} B_{kj} = 1$*. Por lo tanto* AB *es una matriz de probabilidades de transición.* $\square$

Lema 4.2.3 *El producto de una matriz de probabilidades de transición por un vector de probabilidad es un vector de probabilidad.*

Demostración.

Sean A *una matriz de probabilidades de transición y* v *un vector de probabilidad. Entonces* $(Av)_i = \sum_{k=1}^{n} A_{ik}v_k \geq 0$ *dado que* $A_{ik}, v_k \geq 0$ *para* $1 \leq k \leq n$*. Además* $\sum_{i=1}^{n}\left(\sum_{k=1}^{n} A_{ik}v_k\right) = \sum_{k=1}^{n}\left(\sum_{i=1}^{n} A_{ik}v_k\right) = \sum_{k=1}^{n}\left(v_k\sum_{i=1}^{n} A_{ik}\right) = \sum_{k=1}^{n} v_k = 1$*, por lo que* Av *es un vector de probabilidad.* $\square$

Teorema 4.2.2 *Sea* $A \in \mathcal{M}_n$ *una matriz regular de transición. Entonces:*

a) la multiplicidad de 1 *como valor propio de* A *es* 1

b) existe $\lim_{m\to\infty} A^m$

c) *$L = \lim_{m\to\infty} A^m$ es una matriz de probabilidades de transición*

d) *$AL = LA = L$*

e) *las columnas de L son idénticas, de hecho, cada columna de L es igual al único vector de probabilidad v que es también un vector propio correspondiente al valor propio 1 de A*

f) *para cualquier vector de probabilidad x, $\lim_{m\to\infty}(A^m x) = v$.*

Demostración.

a) *Por el teorema 4.2.1, $\rho(A) = 1$ es valor propio de A. Como además A es no negativa y pimitiva (y por tanto irreducible), aplicando el inciso d) del teorema de Perron-Frobenius (teorema 3.6.1), tenemos que 1 es un valor propio de multiplicidad 1.*

b) *Como A es primitiva y no negativa, el resultado se sigue del teorema 3.7.1.*

c) *Según el lema 4.2.2, A^m es una matriz de probabilidades de transición, así que cada elemento de A^m es no negativo para todo $m \in \mathbb{N}$. Por tanto, para $1 \le i, j \le n$,*

$$L_{ij} = \lim_{m\to\infty} (A^m)_{ij} \ge 0.$$

Además, para $1 \le j \le n$,

$$\sum_{i=1}^{n} L_{ij} = \sum_{i=1}^{n} \left(\lim_{m\to\infty} (A^m)_{ij} \right) = \lim_{m\to\infty} \left(\sum_{i=1}^{n} (A^m)_{ij} \right) = \lim_{m\to\infty} (1) = 1.$$

Luego L es una matriz de probabilidades de transición.

d) *De acuerdo con el teorema 3.2.3,*

$$AL = A\Big(\lim_{m\to\infty} A^m \Big) = \lim_{m\to\infty} (AA^m) = \lim_{m\to\infty} A^{m+1} = L.$$

Análogamente, $LA = L$.

e) *Como $AL = L$, de acuerdo con el inciso d), cada una de las columnas de L es un vector propio de A correspondiente al valor propio 1. Además, de acuerdo con el inciso c), cada columna de L es un vector de probabilidad. Se sigue entonces de a) que cada columna de L es igual al único vector de probabilidad v correspondiente al valor propio 1 de A.*

f) *Sea x cualquier vector de probabilidad y defínase $y := Lx$. Entonces, por el lema 4.2.3, y es un vector de probabilidad y por el inciso d), $Ay = ALx = Lx = y$. Por lo tanto y es también un vector propio que corresponde al valor propio 1 de A. Entonces, de acuerdo con el inciso e), $y = v$.* □

Definición 4.2.3 *El vector v del inciso e) del teorema anterior se denomina* **vector de probabilidad fija (o vector estacionario)** *de la matriz regular de transición A.*

Ejemplo 1. En 1995 una investigación municipal del uso de la tierra mostró que el 10% de la tierra del municipio era urbana, el 50% no estaba utilizada y el 40% estaba destinada a usos agrícolas. Cinco años después una investigación de actualización reveló que el 70% de la superficie urbana había permanecido urbana, 10% se había convertido en no utilizada y el 20% se había transformado en superficie agrícola. De la misma manera, 20% de la superficie no utilizada se había convertido en urbana, 60% había permanecido no utilizada y 20% se había convertido en superficie agrícola. Finalmente, la investigación del 2000 mostró que el 20% de la superficie agrícola se había convertido en no utilizada, mientras que el 80% permaneció agrícola. Suponiendo que las tendencias indicadas por la investigación del 2000 continuén, calcularemos los porcentajes de las superficies urbanas, no utilizadas y agrícolas en el municipio en el 2005 y los porcentajes eventuales correspondientes.

Podemos describir la situación anterior como una cadena de Markov de 3 estados, donde los estados son los usos de la tierra, haciendo que los estados uno, dos y tres sean respectivamente el uso urbano, el no utilizado y el agrícola. Vemos que el vector de probabilidad que indica la probabilidad inicial de estar en cada uno de los estados es $P = \begin{pmatrix} 0.1 \\ 0.5 \\ 0.4 \end{pmatrix}$ y la matriz de transición es

$$\mathbf{A} = \begin{pmatrix} 0.7 & 0.2 & 0 \\ 0.1 & 0.6 & 0.2 \\ 0.2 & 0.2 & 0.8 \end{pmatrix}.$$

Los porcentajes de las superficies urbanas, no utilizadas y agrícolas en el municipio en el 2005 son las coordenadas del vector AP, donde $AP = \begin{pmatrix} 0.17 \\ 0.39 \\ 0.44 \end{pmatrix}$ y los porcentajes eventuales de las superficies correspondientes son las coordenadas de $\lim_{m\to\infty} A^m P$. La matriz A es regular pues

$$\mathbf{A}^2 = \begin{pmatrix} 0.51 & 0.26 & 0.04 \\ 0.17 & 0.42 & 0.28 \\ 0.32 & 0.32 & 0.68 \end{pmatrix} > 0.$$

Entonces, según el teorema anterior, $\lim_{m\to\infty} A^m$ es una matriz L en la que cada columna es igual al único vector de probabilidad fija para A. Este vector v es tal que $(A-I)v = 0$. Si $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ vemos que $(A-I)v = 0$ nos proporciona el siguiente sistema de ecuaciones lineales

$$\begin{aligned}-0.3v_1 + 0.2v_2 &= 0\\ 0.1v_1 - 0.4v_2 + 0.2v_3 &= 0\\ 0.2v_1 + 0.2v_2 - 0.2v_3 &= 0.\end{aligned}$$

Resolviendo el sistema, encontramos que $\left\{ \begin{pmatrix} 2 \\ 3 \\ 5 \end{pmatrix} \right\}$ es una base para el espacio de soluciones de este sistema, de donde el vector v es $\begin{pmatrix} 0.2 \\ 0.3 \\ 0.5 \end{pmatrix}$. Entonces a largo plazo esperamos que el 20% de la tierra utilizada sea urbana, el 30% no esté utilizada y el 50% se destine a usos agrícolas.

Ejemplo 2. En 1990 la industria automotriz determinó que el 40% de los americanos poseedores de autos conducía autos grandes, 20% conducía autos de tamaño mediano y el 40% conducía autos pequeños. Una segunda investigación en el 2000 mostró que el 70% de los dueños de autos grandes de 1990 aún poseía autos grandes en el 2000, pero el 30% había cambiado a autos de tamaño mediano. De aquellos que poseían automóviles de tamaño mediano en 1990, 10% había cambiado a autos grandes, el 70% seguía teniendo autos medianos y el 20% había cambiado a autos pequeños en el 2000. Finalmente, de los dueños de autos pequeños de 1990, el 10% poseía autos medianos y el 90% poseía autos pequeños en en el 2000. Suponiendo que esta tendencia continúe, determinaremos el porcentaje de americanos que poseerán autos de cada uno de los tamaños en el 2010 y los porcentajes eventuales correspondientes.
Podemos describir la situación anterior como una cadena de Markov de 3 estados, donde los estados son los tamaños de los autos, haciendo que los estados uno, dos y tres sean respectivamente el tamaño grande, mediano y pequeño. Vemos que el vector de probabilidad que indica la probabilidad inicial de estar en cada uno de los estados es $P = \begin{pmatrix} 0.4 \\ 0.2 \\ 0.4 \end{pmatrix}$ y la matriz de transición es

$$\mathbf{A} = \begin{pmatrix} 0.7 & 0.1 & 0 \\ 0.3 & 0.7 & 0.1 \\ 0 & 0.2 & 0.9 \end{pmatrix}.$$

Los porcentajes de americanos que tendrán autos grandes, medianos y pequeños en 1985 son las coordenadas del vector AP, donde $AP = \begin{pmatrix} 0.3 \\ 0.3 \\ 0.4 \end{pmatrix}$ y los porcentajes eventuales de los tamaños correspondientes son las coordenadas de

$\lim_{m\to\infty} A^m P$. La matriz A es regular pues

$$\mathbf{A}^2 = \begin{pmatrix} 0.52 & 0.14 & 0.01 \\ 0.42 & 0.54 & 0.16 \\ 0.06 & 0.32 & 0.83 \end{pmatrix} > 0,$$

por lo que $\lim_{m\to\infty} A^m$ es una matriz L en la que cada columna es igual al vector de probabilidad fija para A. Este vector v es tal que $(A-I)v = 0$. Si $v = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ vemos que $(A-I)v = 0$ nos proporciona el siguiente sistema de ecuaciones lineales

$$\begin{aligned} -0.3v_1 + 0.1v_2 &= 0 \\ 0.3v_1 - 0.3v_2 + 0.1v_3 &= 0 \\ 0.2v_2 - 0.1v_3 &= 0. \end{aligned}$$

Resolviendo el sistema, encontramos que el vector buscado es $v = \begin{pmatrix} 0.1 \\ 0.3 \\ 0.6 \end{pmatrix}$.

Entonces a largo plazo esperamos que el 10% de los americanos poseerán autos grandes, 30% conducirán autos medianos y 60% tendrán autos pequeños.

Ejemplo 3. Una unidad de traumatología de un hospital ha determinado que al momento de llegar al hospital el 30% de sus pacientes es ambulatorio y el 70% debe guardar cama. Un mes después de su llegada, el 60% de los pacientes ambulatorios se ha recuperado, 20% permanece ambulatorio y 20% ahora debe guardar cama. Después del mismo lapso, 10% de los pacientes encamados se ha recuperado, 20% ahora es ambulatorio, 50% permanece encamado y el 20% ha muerto. Determinaremos el porcentaje de pacientes que se recuperan, son ambulatorios, están encamados y han muerto un mes después de su llegada, así como el porcentaje eventual de pacientes de cada tipo.

Podemos describir la situación anterior como una cadena de Markov de 4 estados, donde los estados son la situación de los pacientes haciendo que los estados uno, dos, tres y cuatro sean respectivamente el estar sano o recuperado, el ser ambulatorio, el guardar cama y el estar muerto. Vemos que el vector de probabilidad que indica la probabilidad inicial de estar en cada uno de los estados es $P = \begin{pmatrix} 0 \\ 0.3 \\ 0.7 \\ 0 \end{pmatrix}$ y la matriz de transición es

$$\mathbf{A} = \begin{pmatrix} 1 & 0.6 & 0.1 & 0 \\ 0 & 0.2 & 0.2 & 0 \\ 0 & 0.2 & 0.5 & 0 \\ 0 & 0 & 0.2 & 1 \end{pmatrix}.$$

Los porcentajes de los estados de los pacientes un mes después de su llegada son las coordenadas del vector AP, donde $AP = \begin{pmatrix} 0.25 \\ 0.2 \\ 0.41 \\ 0.14 \end{pmatrix}$ y los porcentajes eventuales de los estados correspondientes son las coordenadas de $\lim_{m\to\infty} A^m P$. Sin embargo, no podemos aplicar el teorema 4.2.2 porque A no es regular ya que la primera columna de A^k es $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}$ para todo $k \in \mathbb{N}$. Sin embargo, A es diagonalizable, pues si

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & 2 & -11 \\ 0 & 0 & -1 & 18 \\ 0 & 0 & -2 & -9 \\ 0 & 1 & 1 & 2 \end{pmatrix}$$

entonces Q es invertible con inversa

$$\mathbf{Q}^{-1} = \begin{pmatrix} 1 & \frac{8}{9} & \frac{5}{9} & 0 \\ 0 & \frac{1}{9} & \frac{4}{9} & 1 \\ 0 & -\frac{1}{5} & -\frac{2}{5} & 0 \\ 0 & \frac{2}{45} & -\frac{1}{45} & 0 \end{pmatrix}$$

y así

$$\mathbf{D} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0.6 & 0 \\ 0 & 0 & 0 & 0.1 \end{pmatrix}.$$

Por lo tanto

$$\lim_{m\to\infty} A^m = \lim_{m\to\infty} (QDQ^{-1})^m = Q(\lim_{m\to\infty} D^m)Q^{-1} =$$

$$\begin{aligned}
&= \begin{pmatrix} 1 & 0 & 2 & -11 \\ 0 & 0 & -1 & 18 \\ 0 & 0 & -2 & -9 \\ 0 & 1 & 1 & 2 \end{pmatrix} \lim_{m\to\infty} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & (0.6)^m & 0 \\ 0 & 0 & 0 & (0.1)^m \end{pmatrix} \begin{pmatrix} 1 & \frac{8}{9} & \frac{5}{9} & 0 \\ 0 & \frac{1}{9} & \frac{4}{9} & 1 \\ 0 & -\frac{1}{5} & -\frac{2}{5} & 0 \\ 0 & \frac{2}{45} & -\frac{1}{45} & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & 2 & -11 \\ 0 & 0 & -1 & 18 \\ 0 & 0 & -2 & -9 \\ 0 & 1 & 1 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & \frac{8}{9} & \frac{5}{9} & 0 \\ 0 & \frac{1}{9} & \frac{4}{9} & 1 \\ 0 & -\frac{1}{5} & -\frac{2}{5} & 0 \\ 0 & \frac{2}{45} & -\frac{1}{45} & 0 \end{pmatrix} \\
&= \begin{pmatrix} 1 & \frac{8}{9} & \frac{5}{9} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{9} & \frac{4}{9} & 1 \end{pmatrix},
\end{aligned}$$

de donde

$$\lim_{m\to\infty} A^m P = \begin{pmatrix} 1 & \frac{8}{9} & \frac{5}{9} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{1}{9} & \frac{4}{9} & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 0.3 \\ 0.7 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{59}{90} \\ 0 \\ 0 \\ \frac{31}{90} \end{pmatrix}.$$

Entonces a largo plazo esperamos que el $\frac{59}{90}$% de los pacientes que llegaron al hospital inicialmente se recuperen y el $\frac{31}{90}$% restante muera.

Ejemplo 4. Un jugador principia un juego de azar colocando una ficha en el casillero 2 (marcado salida), de acuerdo con la siguiente figura

Gana	Salida		Pierde
1	2	3	4

Se lanza un dado y la ficha se mueve un cuadro a la izquierda si se obtiene 1 ó 2 y un cuadro a la derecha si se obtiene 3,4,5 ó 6. Este proceso continúa hasta que la ficha llega al cuadro 1 (en cuyo caso el jugador gana el juego) o al cuadro 4 (en cuyo caso el jugador pierde el juego). ¿Cuál es la probabilidad de ganar este juego?

Podemos describir el juego anterior como una cadena de Markov de 4 estados, donde los estados son los lugares donde se encuentra la ficha (1,2,3,4). La matriz de probabilidades de transición es

$$\mathbf{A} = \begin{pmatrix} 1 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 \\ 0 & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & 1 \end{pmatrix}$$

Como en cualquier juego, una vez que se gana o se pierde, se termina el juego; traduciendo esto en términos de cadenas de Markov, esto quiere decir que una vez que se llega a los estados 1 ó 4 se permanece indefinidamente en dicho estado respectivamente.

El vector de probabilidad que indica la probabilidad inicial de estar en cada uno de los estados es $P = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$ y la respuesta a la probabilidad de ganar este juego la proporciona la primera coordenada de $\lim_{m\to\infty} A^m P$. Sin embargo, como en el ejemplo anterior, no podemos aplicar el teorema 4.2.2, pero A es diagonalizable, pues si

$$\mathbf{Q} = \begin{pmatrix} 1 & 0 & \frac{1}{4-6\sqrt{2}} & \frac{-1}{4+6\sqrt{2}} \\ 0 & 0 & \frac{1}{2\sqrt{2}} & \frac{1}{2\sqrt{2}} \\ 0 & 0 & \frac{1}{2} & -\frac{1}{2} \\ 0 & 1 & \frac{1}{-3+\sqrt{2}} & \frac{1}{3+\sqrt{2}} \end{pmatrix}$$

entonces Q es invertible con inversa

$$\mathbf{Q}^{-1} = \begin{pmatrix} 1 & \frac{3}{7} & \frac{1}{7} & 0 \\ 0 & \frac{4}{7} & \frac{6}{7} & 1 \\ 0 & \sqrt{2} & 1 & 0 \\ 0 & \sqrt{2} & -1 & 0 \end{pmatrix}$$

y así

$$\mathbf{D} = \mathbf{Q}^{-1}\mathbf{A}\mathbf{Q} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{\sqrt{2}}{3} & 0 \\ 0 & 0 & 0 & -\frac{\sqrt{2}}{3} \end{pmatrix}.$$

Por lo tanto

$$\begin{aligned} \lim_{m\to\infty} A^m &= \lim_{m\to\infty} (QDQ^{-1})^m = \lim_{m\to\infty} (QD^mQ^{-1}) \\ &= Q(\lim_{m\to\infty} D^m)Q^{-1} = \begin{pmatrix} 1 & \frac{3}{7} & \frac{1}{7} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{4}{7} & \frac{6}{7} & 1 \end{pmatrix} \end{aligned}$$

de donde

$$\lim_{m\to\infty} A^m P = \begin{pmatrix} 1 & \frac{3}{7} & \frac{1}{7} & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \frac{4}{7} & \frac{6}{7} & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{3}{7} \\ 0 \\ 0 \\ \frac{4}{7} \end{pmatrix}.$$

Entonces la probabilidad de ganar el juego es $\frac{3}{7}$.

Bibliografía

[1] BARTLE, Robert G.
Introducción al análisis matemático.
Primera Edición, Editorial Limusa, S. A. de C. V., México, 1989.

[2] FRIEDBERG, Stephen H., INSEL, Arnold J. y SPENCE, Lawrence E.
Algebra lineal.
Primera Edición, Publicaciones Cultural, S. A., México, 1982.

[3] GROSSMAN, Stanley I.
Aplicaciones de álgebra lineal.
Cuarta Edición, McGraw-Hill, México, 1992.

[4] HOFFMAN, Kenneth y KUNZE, Ray.
Álgebra lineal.
Prentice-Hall Hispanoamericana, México, 1988.

[5] HORN, Roger A. y JOHNSON, Charles R.
Matrix analysis.
Cambridge University Press, Estados Unidos, 1999.

[6] ROSS, Sheldon M.
Stochastic Processes.
John Wiley & Sons, Estados Unidos, 1983.

[7] TAYLOR, Howard M. y KARLIN, Samuel.
An Intoduction to Stochastic Modeling.
Tercera Edición, Academic Press, Estados Unidos, 1998.