



"El saber de mis hijos
hará mi grandeza".

UNIVERSIDAD DE SONORA
UNIDAD REGIONAL SUR
DIVISIÓN DE CIENCIAS E INGENIERÍA
DEPARTAMENTO DE FÍSICA, MATEMÁTICAS E INGENIERÍA

**ELIMINACIÓN DE LA INESTABILIDAD NUMÉRICA
DEL MÉTODO SIMPLIFICADO DE NEWTON
EL CÁLCULO DE LA P-ÉSIMA RAÍZ PRINCIPAL
DE UNA MATRIZ NO SINGULAR**

**TRABAJO DE TESIS
QUE PARA OBTENER EL TÍTULO DE**

INGENIERO INDUSTRIAL Y DE SISTEMAS

PRESENTAN:

JEAN WOOD EMMANUEL DESIR

JOSEPH WILKY DESIR

NAVOJOA, SONORA

ABRIL DEL 2016

Universidad de Sonora

Repositorio Institucional UNISON



**"El saber de mis hijos
hará mi grandeza"**



Excepto si se señala otra cosa, la licencia del ítem se describe como openAccess



EL SABER DE MIS HIJOS
HARA MI GRANDEZA

UNIVERSIDAD DE SONORA
UNIDAD REGIONAL SUR
DIVISIÓN DE CIENCIAS E INGENIERÍA
Departamento de Física, Matemáticas e Ingeniería

Navojoa, Sonora a 28 de marzo de 2016

L.M. Santos Velarde Palomares
Jefe de Departamento de Física, Matemáticas e Ingeniería
Unidad Regional Sur
PRESENTE:

Por este conducto, hago de su conocimiento que estamos de acuerdo que se realice el examen profesional de:

JEAN WOOD EMMANUEL DESIR
JOSEPH WILKY DESIR

el día viernes 15 de abril de 2016 en la Sala de Juntas del edificio B a las 11:00 horas.

A T E N T A M E N T E

MIEMBROS DEL JURADO

	NOMBRE
PRESIDENTE	C.DR. ALFREDO MENDOZA MEXIA, (22769)
SECRETARIO	Dr. IGNACIO YOCUPICIO VILLEGAS, (24188)
VOCAL	M.A. MA. DEL ROSARIO CASTREJON LEMUS, (26925)
SUPLENTE	M.C. RAFAEL VERDUGO MIRANDA, (30492)

FIRMA

AGRADECIMIENTOS

Emmanuel:

Primeramente, a **Dios** por su gracia, su amor y sus bendiciones, por permitirme estar en México y por haber llegado a una meta más en mi vida en este país.

A mi familia, principalmente a mis padres **Marie Rosette ORCEL** y **Joseph Wesner DESIR** por su apoyo y su amor incondicional y por haberme dado una educación adecuada para ser lo que hasta hoy soy.

A mi hermano, amigo y compañero **Joseph Wilky DESIR** por su apoyo en todo momento.

A la Universidad de Sonora y a mis maestros, en especial al **cDr. Alfredo Mendoza Mexía** director de este trabajo de tesis, por el apoyo a lo largo de mi carrera, las enseñanzas, apoyos y oportunidades brindadas.

Al rector Dr. Heriberto Grijalva Monteverde y a la vicerrectora Dra. Luz Haydee Cruz Morales por sus apoyos económicos.

A la Lic. Bertha Estrella Valenzuela por su apoyo incondicional.

A **México** por brindarme una segunda oportunidad al permitirme llegar, establecerme, trabajar y estudiar.

GRACIAS.

AGRADECIMIENTOS

Wilky:

Primeramente, a **Dios** porque bajo su bendición me permití presentar este trabajo, a cuidarme a todo lo largo de esta carrera, ayudándome a aprender y superar los momentos difíciles, gracias mi Señor.

A mi **familia**, cada uno que me ha dado su ánimo, y especialmente a mi madre **Marie Rosette ORCEL** y mi padre **Joseph Wesner DESIR** por la vida que me han dado, cuidándome desde niño hasta el día de hoy y ser mis amigos.

A mi amigo, compañero y hermano **Jean Wood Emmanuel DESIR** por estar siempre conmigo en esta lucha.

A **mis compañeros** de la Universidad de Sonora, Unidad Regional Sur por su apoyo y haberme aceptado como parte de ellos.

A la **Universidad de Sonora** por haberme aceptado y tener los mejores maestros para orientarme hacia un mundo intelectual especialmente al **cDr. Alfredo Mendoza Mexía** por ser mi guía a lo largo de esta carrera, en las prácticas profesionales y en mi tesis.

Al rector Dr. Heriberto Grijalva Monteverde y a la vicerrectora Dra. Luz Haydee Cruz Morales por sus apoyos económicos.

A la Lic. Bertha Estrella Valenzuela por su apoyo incondicional.

A **México** por brindarme una segunda oportunidad al permitirme llegar, establecerme, trabajar y estudiar.

GRACIAS

DEDICATORIAS

Emmanuel:

A mi madre **Marie Rosette ORCEL**, mi padre **Joseph Wesner DESIR**, mis hermanos, mis primos y mis amigos que han contribuido de una forma u otra manera para seguir adelante.

A cada uno de los maestros de la **Universidad de Sonora, Unidad Regional Sur** de la carrera Ingeniería Industrial y de Sistemas

Wilky:

A mi madre **Marie Rosette ORCEL** por siempre apoyarme en cada meta, ser mi apoyo y mi amiga gracias por ti he logrado alcanzarlas TE AMO MAMÁ.

A mi padre **Joseph Wesner DESIR** por ser mi modelo a seguir en la vida.

A mi hermano **Jean Wood Emmanuel DESIR** por siempre estar a mi lado y apoyarme en los tiempos difíciles.

A cada uno de mis maestros por el tiempo dedicado a mis conocimientos.

ÍNDICE

Agradecimientos	i
Dedicatorias	iii
Índice	iv
Lista de tablas	vii
Lista de gráficas	viii
Introducción	1
Antecedentes	2
Definición del problema	4
Justificación	5
Hipótesis	6
Objetivo principal	6
 CAPÍTULO 1 ÁLGEBRA MATRICIAL	 7
1.1 Definiciones importantes	7
1.2 Operaciones básicas con matrices	9
1.2.1 Suma de matrices	9
1.2.2 Multiplicación por un escalar	10
1.2.3 Resta de matrices	10
1.2.4 Multiplicación de matrices	10
1.2.5 Trasposición de matrices	11
1.2.6 Traza de una matriz	11
1.2.7 Determinantes	12

1.2.8 Matriz inversa	13
1.3 Matrices cuadradas especiales	14
1.3.1 Matriz diagonal	14
1.3.2 Matriz identidad	14
1.3.3 Matriz escalar	15
1.3.4 Matriz triangular inferior	15
1.3.5 Matriz nula	15
1.3.6 Matriz simétrica	15
1.3.7 Matriz idempotente	15
1.3.8 Matriz ortogonal	15
1.4 Valores propios y vectores propios de una matriz	15
1.5 Diagonalización de matrices	17
CAPÍTULO 2 FORMA CANÓNICA DE JORDAN Y EL CÁLCULO DE LA P-ÉSIMA RAÍZ DE UNA MATRIZ	19
2.1 Algunas aplicaciones del cálculo de la p-ésima raíz de una matriz no singular	19
2.2 La forma canónica de Jordan	22
2.3 La función p-ésima raíz de una matriz no singular	27
2.4 Ejemplos de cómo calcular la p-ésima raíz de una matriz no singular utilizando la forma canónica de Jordan.	29
CAPÍTULO 3 LOS MÉTODOS DE NEWTON PARA EL CASO MATRICIAL	38
3.1 Los métodos de Newton	38
3.2 El Método de Newton (MN) para el cálculo de la p-ésima raíz de una matriz	40

3.3 Análisis de convergencia del (MSN) para el cálculo de la p -ésima raíz de una matriz no singular	43
CAPÍTULO 4 MÉTODO SIMPLIFICADO DE NEWTON CON FACTORIZACIONES SUCESIVAS PARA EL CÁLCULO DE LAP-ÉSIMARAÍZ DE UNA MATRIZ NO SINGULAR (MSNFSPR).	52
4.1 Deducción del (MSNFSPR)	52
4.2 Análisis de convergencia del (MSNFSPR)	53
4.3 Análisis de estabilidad numérica del (MSNFSPR)	55
CAPÍTULO 5 EXPERIMENTOS NUMÉRICOS	59
CONCLUSIONES	66
BIBLIOGRAFÍA	67

LISTA DE TABLAS

TABLA	DESCRIPCIÓN	PÁGINA
2.1	Raíces cúbicas de los valores propios de A	31
2.2	Raíces cúbicas de los valores propios de B	36
5.1	Comportamiento de la convergencia para el experimento No.1	60
5.2	Comportamiento de la convergencia para el experimento No.2	61
5.3	Comportamiento de la convergencia para el experimento No.3	63
5.4	Comportamiento de la convergencia para el experimento No.4	64
5.5	Costo computacional por iteración para el cálculo de la de la p-ésima raíz de una matriz no singular	65

LISTA DE GRÁFICAS

GRÁFICA	DESCRIPCIÓN	PÁGINA
5.1	Comportamiento de la convergencia para el experimento No.1	60
5.2	Comportamiento de la convergencia para el experimento No.2	62
5.3	Comportamiento de la convergencia para el experimento No.3	63
5.4	Comportamiento de la convergencia para el experimento No.4	64

INTRODUCCIÓN

Hasta hace poco tiempo, la utilización del Método Simplificado de Newton (MSN) era muy escasa y poco atractiva para el cálculo de la p -ésima raíz de una matriz no singular con $p > 2$ principalmente por dos razones: sus propiedades de convergencia no eran bien comprendidas en presencia de valores propios complejos, además de que sus propiedades de estabilidad numérica eran muy pobres.

En este trabajo de tesis se propone y se demuestra que la variante de ir factorizando la matriz $\mathbf{A}_n = \mathbf{A} \|\mathbf{A}\|_F^{-1} = \mathbf{B}_k \mathbf{C}_k^{p-1} = \mathbf{C}_k^{p-1} \mathbf{B}_k$ en cada k -ésima iteración del (MSN) como el producto de dos matrices \mathbf{B}_k y \mathbf{C}_k^{p-1} que conmuten entre sí, elimina su inestabilidad numérica para el cálculo de la p -ésima raíz de una matriz no singular, convirtiéndolo en un método atractivo por ser robusto, convergente, computacionalmente económico, fácil de implementar y para propósitos prácticos es numéricamente estable como lo demuestran los experimentos numéricos llevados a cabo. El algoritmo tiene un costo computacional del orden de $2n^3(3 + \vartheta \log(p))$, $1 \leq \vartheta \leq 2$ operaciones aritméticas por iteración.

ANTECEDENTES

Los Métodos de Newton para calcular la p -ésima raíz principal de una matriz han sido estudiados durante casi 60 años, y en la actualidad estos métodos ya son bien comprendidos. En lo que respecta al (MSN), en 1958 Laasonen [26] demostró su convergencia, pero también observó sin demostrarlo, que para algunos casos como por ejemplo en el cálculo de la raíz cuadrada presentaba inestabilidad numérica cuando el número de condición de la matriz era mayor que nueve. Desde entonces y hasta la fecha han estado surgiendo variantes del Método de Newton (MN) para el cálculo de la p -ésima raíz de una matriz no singular, que son convergentes y numéricamente estables.

En cuanto a los algoritmos disponibles para el cálculo de la p -ésima raíz de una matriz no singular, se comenta brevemente lo siguiente.

Björck y Hammarling [3] propusieron un método para calcular la raíz cuadrada y raíz cúbica de una matriz utilizando la descomposición de Schur, sin embargo, si la matriz es real, este método puede requerir aritmética compleja, incluso si la raíz que se está buscando es real. Este método fue posteriormente extendido por Higham [11], quien sugirió el uso de la forma real de la descomposición de Schur, y generalizado después por Smith [33] para el cálculo de la p -ésima raíz de una matriz. El método de Smith implementado en MATLAB toolbox [13], es numéricamente estable y requiere $(28 + (p - 1)/3)n^3$ operaciones aritméticas por iteración [14, p.176].

Iannazzo [18] [19] propuso dos nuevas iteraciones para el cálculo de la p -ésima raíz que son convergentes y numéricamente estables. Más recientemente Chun-Hua Guo [8] analizó la manera de como calcular la p -ésima raíz de una matriz, preconditionándola si fuese necesario, utilizando el (MSN) o el Método de Halley, y también explicó cómo la convergencia de estos

métodos se puede mejorar cuando los valores propios de la matriz son conocidos. En este trabajo de tesis se presenta una variante para eliminar la inestabilidad numérica del (MSN) para calcular la p -ésima raíz de una matriz no singular, la variante consiste en lo siguiente: en cada $k + 1$ -ésima iteración se factoriza de manera conveniente la matriz $\mathbf{A}_k = \mathbf{A} \|\mathbf{A}\|_F^{-1}$ expresándola como el producto de dos matrices \mathbf{B}_k y \mathbf{C}_k^{p-1} que conmutan entre sí, es decir $\mathbf{A}_k = \mathbf{B}_k \mathbf{C}_k^{p-1} = \mathbf{C}_k^{p-1} \mathbf{B}_k$. El proceso anterior elimina la inestabilidad numérica del (MSN) convirtiéndolo en un método atractivo para el cálculo de la p -ésima raíz de una matriz no singular, por ser robusto, convergente, computacionalmente económico y para propósitos prácticos es numéricamente estable como será demostrado posteriormente.

DEFINICIÓN DEL PROBLEMA

Algunos métodos actuales para calcular la p -ésima raíz principal de una matriz no singular como por ejemplo el (MN) y el (MSN) tienen ciertas limitantes o características que los hacen poco versátiles y atractivos, como por ejemplo el (MSN) es convergente y numéricamente estable bajo ciertas condiciones muy restrictivas, como por ejemplo la condición de estabilidad numérica para el cálculo de la raíz quinta de una matriz definida positiva es que su número de condición debe ser menor que 4.52, esto representa una muy fuerte restricción lo cual hace prohibitivo su uso. El (MN) aunque es convergente y numéricamente estable para cualquier matriz no singular, en contraparte es computacionalmente muy costoso, ya que para calcular la raíz cuadrada de una matriz no singular de $n \times n$, se debe resolver un sistema lineal de $n^2 \times n^2$, lo cual lo hace poco atractivo para el cálculo de p -ésimas raíces.

Resumiendo, el problema es que el (MSN) es numéricamente inestable para el cálculo de p -ésimas raíces, y el (MN) aunque es numéricamente estable, tiene el inconveniente de que es computacionalmente muy costoso.

JUSTIFICACIÓN

Algunos de los métodos actuales para el cálculo de la p -ésima raíz principal de una matriz no singular, presentan ciertas limitaciones, por ejemplo el Método de Newton (MN) a pesar de ser convergente y numéricamente estable presenta un alto costo computacional, ya que al resolver una raíz cuadrada de una matriz de tamaño $n \times n$ la convierte en una de $n^2 \times n^2$, lo cual lo vuelve poco atractivo, simplificando al (MN) surgen dos iteraciones que representan el (MSN) ambas presentan muy buenas propiedades de convergencia, pero su estabilidad numérica es muy pobre.

Al revisar la literatura que trata sobre el cálculo de la p -ésima raíz principal de una matriz no singular, nos podemos dar cuenta de la rica variedad de métodos que existen para su cálculo, las cuestiones fundamentales relacionadas con el análisis de su convergencia y estabilidad numérica siguen hoy en día vigentes. Sigue habiendo mucho trabajo por hacer, ya que en la actualidad lo anteriormente mencionado son temas de estudio e investigación de gran interés.

Por todo lo anterior en este trabajo de tesis se presenta como propuesta un nuevo método para el cálculo de la p -ésima raíz de una matriz, el cual es atractivo por ser computacionalmente sencillo de implementar, robusto, económico, convergente y numéricamente estable para cualquier matriz no singular.

OBJETIVO PRINCIPAL

Este trabajo de tesis, tiene como objetivo principal realizar una revisión bibliográfica exhaustiva sobre algunos métodos para el cálculo de la p-ésima raíz principal de una matriz no singular para analizar sus propiedades de convergencia y estabilidad numérica y, en base a lo anterior, proponer una modificación original al (MSN) que eliminará su inestabilidad numérica convirtiéndolo en un método atractivo para el cálculo de p-ésimas raíces, ya que será convergente, numéricamente estable, computacionalmente económico y sencillo de implementar.

HIPOTESIS

La inestabilidad numérica observada del (MSN) para el cálculo de la p-ésima raíz principal de una matriz no singular, se debe a que en el proceso iterativo llega un momento en que los errores se empiezan a distribuir de una manera muy irregular dentro de la matriz (se van cargando hacia una determinada zona de la matriz), causando con ello que sus valores propios se alteren, provocando con ello que se pierda la conmutatividad $X_k H_k = H_k X_k$ causando la inestabilidad numérica observada. Lo anterior puede ser corregido expresando a la matriz en cuestión como el producto de dos matrices que conmutan entre sí, es decir: $A_n = B_k C_k^{p-1} = C_k^{p-1} B_k$; $A_n = A \|A\|_F^{-1}$; con lo anterior se eliminará su inestabilidad numérica.

CAPÍTULO 1 ÁLGEBRA MATRICIAL

El cálculo de la p -ésima raíz de una matriz requiere cierta familiaridad con el álgebra matricial. La teoría de matrices simplifica la descripción, desarrollo y aplicación de los métodos para calcular p -ésimas raíces. En este capítulo, se resumen algunos conceptos fundamentales del álgebra matricial que se usarán a lo largo de este trabajo de tesis [1] [7].

1.1 Definiciones importantes.

Una matriz A de orden $m \times n$ es un conjunto de elementos a_{ij} ($i = 1, \dots, m; j = 1, \dots, n$) dispuestos en m filas y n columnas, por ejemplo, la matriz A se representa como:

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

Las matrices se representan con letras mayúsculas en negrita. El elemento de la fila i -ésima y de la columna j -ésima se representa por una letra minúscula con un par de subíndices a_{ij} .

Por lo tanto, un modo abreviado de escribir una matriz es $A = [a_{ij}]$ para $i = 1, \dots, m; j = 1, \dots, n$. El orden o dimensión de la matriz $m \times n$ nos indica el número de filas y de columnas.

La matriz A se denomina cuadrada cuando $m = n$ y rectangular si $m \neq n$.

Los elementos de una matriz pueden ser numéricos o de cualquier clase. En este trabajo de tesis sólo se consideran matrices de números complejos, es decir $A \in C^{n \times n}$, no perdiendo de vista que el conjunto de los números reales es un subconjunto de los complejos. Por ejemplo, la matriz

A

$$A = \begin{pmatrix} 6 & 5 & 7 & 4 \\ 5 & 4 & 2 & 5 \\ 1 & 1 & 11 & 1 \end{pmatrix}$$

es una matriz rectangular de orden $m \times n$; el elemento de la fila 3 y columna 3 es 11. Por ejemplo, la traspuesta de la matriz A es:

$$A' = \begin{pmatrix} 6 & 5 & 1 \\ 5 & 4 & 1 \\ 7 & 2 & 11 \\ 4 & 5 & 1 \end{pmatrix}$$

Un vector columna es una matriz de orden $m \times 1$, es decir, una matriz que sólo tiene una columna

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{pmatrix}$$

Un vector columna se denota por una letra minúscula en negrita y se escribe de forma abreviada como $\mathbf{a} = [a_i]$. Cada elemento del vector tiene un subíndice que indica la posición en la columna. Un vector fila es una matriz de orden $1 \times m$, es decir, una matriz que sólo tiene una fila

$$\mathbf{a}' = (a_1, a_2, \dots, a_m)$$

La traspuesta del vector columna $\mathbf{a} = (a_1, a_2, \dots, a_m)'$ es un vector fila $\mathbf{a}' = (a_1, a_2, \dots, a_m)$

Observe que la notación $(a_1, a_2, \dots, a_m)'$ indica la traspuesta de un vector fila (que es un vector columna) y se usa para escribir un vector columna en una línea de texto.

Sean $\mathbf{a} = (a_1, a_2, \dots, a_m)'$ y $\mathbf{b} = (b_1, b_2, \dots, b_m)'$ dos vectores columna del mismo orden $m \times 1$, su producto escalar se define como $\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = a_1b_1 + a_2b_2 + \dots + a_mb_m = \sum_{i=1}^m a_ib_i$ que es la suma de los productos de cada elemento de \mathbf{a} por el correspondiente elemento de \mathbf{b} .

La norma de un vector \mathbf{X} se define como

$$\|\mathbf{X}\| = \sqrt{\mathbf{X}'\mathbf{X}}$$

siendo el vector normalizado $\frac{\mathbf{x}}{\|\mathbf{x}\|}$ un vector unitario

Dos vectores $\mathbf{a} = (a_1, a_2, \dots, a_m)'$ y $\mathbf{b} = (b_1, b_2, \dots, b_m)'$ son ortogonales si y sólo si, si su producto escalar es cero

$$\mathbf{a}'\mathbf{b} = \mathbf{b}'\mathbf{a} = a_1b_1 + a_2b_2 + \dots + a_mb_m = \sum_{i=1}^m a_ib_i = 0$$

1.2 Operaciones básicas con matrices.

Dos matrices $\mathbf{A} = [a_{ij}]$ y $\mathbf{B} = [b_{ij}]$ del mismo orden $m \times n$ son iguales si $a_{ij} = b_{ij}$ para toda $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, n$.

1.2.1 Suma de matrices, la suma de dos matrices $\mathbf{A} = [a_{ij}]$ y $\mathbf{B} = [b_{ij}]$ del mismo orden $m \times n$ es una matriz $\mathbf{C} = [c_{ij}]$ de orden $m \times n$ tal que $c_{ij} = a_{ij} + b_{ij}$ para toda $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, n$.

La suma de matrices cumple las siguientes propiedades

- Commutativa: $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- Asociativa: $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
- Existencia de elemento neutro nulo $\mathbf{0} = [0]$: $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$
- Existencia de matriz opuesta $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$

Por ejemplo: La suma de las matrices \mathbf{A} y \mathbf{B} es

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \underbrace{\begin{pmatrix} 6 & 5 & 7 & 4 \\ 5 & 4 & 2 & 5 \\ 1 & 1 & 11 & 1 \end{pmatrix}}_{\mathbf{A}} + \underbrace{\begin{pmatrix} 7 & 11 & 2 & 9 \\ 5 & 8 & 8 & 1 \\ 6 & 10 & 8 & 10 \end{pmatrix}}_{\mathbf{B}} = \underbrace{\begin{pmatrix} 13 & 16 & 9 & 13 \\ 10 & 12 & 10 & 6 \\ 7 & 11 & 19 & 11 \end{pmatrix}}_{\mathbf{C}}$$

1.2.2 Multiplicación por un escalar, el producto de una matriz $A = [a_{ij}]$ por un escalar λ es una matriz $B = [b_{ij}] = [\lambda a_{ij}]$, esto es, se multiplican todos los elementos de la matriz por el escalar

Ejemplo. La multiplicación de la matriz A por 2 es

$$E = 2A = \begin{pmatrix} 20 & 10 & 14 & 8 \\ 10 & 8 & 4 & 10 \\ 2 & 2 & 22 & 2 \end{pmatrix}$$

1.2.3 Resta de matrices, la resta de dos matrices $A = [a_{ij}]$ y $B = [b_{ij}]$ del mismo orden $m \times n$ es una matriz $C = [c_{ij}]$ de orden $m \times n$ tal que $c_{ij} = a_{ij} - b_{ij}$ para toda $i = 1, 2, \dots, m$ y $j = 1, 2, \dots, n$. La operación resta puede definirse también a partir de la suma de matrices y la multiplicación de una matriz por un escalar.

1.2.4 Multiplicación de matrices, Sean $A = [a_{ij}]$ y $B = [b_{ij}]$ dos matrices de ordenes $m \times n$ y $n \times p$, respectivamente (el número de columnas de A es igual al número de filas de B). El producto de A y B , AB es una matriz $C = [c_{ij}]$ de orden $m \times p$ tal que:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj} \text{ para toda } i = 1, 2, \dots, m \text{ y } j = 1, 2, \dots, p.$$

Observe que el elemento c_{ij} es el producto escalar de la fila i -ésima de A por la columna j -ésima de B .

La multiplicación de matrices cumple las propiedades:

- a) Asociativa: $(AB)C = A(BC)$
- b) Distributiva: $A(B + C) = AB + AC$

Observación. La multiplicación de matrices generalmente no siempre cumple con la propiedad conmutativa: $AB \neq BA$

$$F = A \times B' = \begin{pmatrix} 6 & 5 & 7 & 4 \\ 5 & 4 & 2 & 5 \\ 1 & 1 & 11 & 1 \end{pmatrix} \begin{pmatrix} 7 & 5 \\ 11 & 8 \\ 2 & 8 \\ 9 & 1 \end{pmatrix} = \begin{pmatrix} 147 & 130 & 182 \\ 128 & 78 & 136 \\ 49 & 104 & 114 \end{pmatrix}$$

1.2.5 Trasposición de matrices, la trasposición de matrices, ya definida, cumple las propiedades:

- a) Reflexiva: $(A')' = A$
- b) La traspuesta de la suma de dos matrices es la suma de las matrices traspuestas $(A + B)' = A' + B'$
- c) $(AB)' = B'A'$ la traspuesta del producto de dos matrices es el producto de las traspuestas en orden invertido $(AB)' = B'A'$. Esta propiedad puede extenderse al producto de tres o más matrices: $(ABC)' = (A(BC))' = (BC)'A' = C'B'A'$.

1.2.6 Traza de una matriz, la traza de una matriz cuadrada $A = [a_{ij}]$ de orden $m \times m$ es la suma de los elementos de la diagonal principal

$$tr(A) = a_{11} + a_{22} + \dots + a_{mm} = \sum_i^m a_{jj}$$

Es claro que se cumplen las siguientes propiedades:

- a) $tr(A) = tr(A')$
- b) $tr(A + B) = tr(A) + tr(B)$
- c) $tr(AB) = tr(BA)$

Ejemplo. La traza de la matriz F es

$$\text{tr}(F) = \begin{pmatrix} 130 & 182 \\ 128 & 78 & 136 \\ 147 & 102 & 114 \\ 49 & & \end{pmatrix} = 147 + 78 + 114 = 339$$

1.2.7 Determinantes.

El determinante es un escalar o lo que es lo mismo, una matriz de tamaño 1×1 . El determinante de una matriz $A = [a_{ij}]$ de orden 2×2 es

$$|A| = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

que es el producto de los elementos de la diagonal principal menos el producto de los elementos situados fuera de la diagonal. El determinante de una matriz $A = [a_{ij}]$ de orden 3×3 es

$$|A| = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{13}a_{22}a_{31} - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32}$$

El determinante de una matriz cuadrada A puede calcularse por expansión de sus menores

$$|A| = \sum_{j=1}^m a_{ij}(-1)^{i+j} |M_{ij}|$$

Algunas propiedades de los determinantes son las siguientes

1. $|AB| = |BA| = |A||B|$ si A y B son matrices cuadradas del mismo orden.
2. $|A^t| = |A|$
3. $|\lambda A| = \lambda^m |A|$
4. $|A^{-1}| = |A|^{-1}$

Definición. Se dice que una matriz cuadrada A es singular si su determinante es cero, $|A| = 0$, y no singular si su determinante es distinto de cero, $|A| \neq 0$.

Ejemplo. Sea la matriz \mathbf{G}

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

El determinante de \mathbf{G} desarrollado a partir de la tercera columna es

$$|G| = \begin{vmatrix} 1 & 1 & 3 \\ 1 & 1 & 0 \\ 3 & 1 & 2 \end{vmatrix} = 3 \begin{vmatrix} 1 & 1 \\ 3 & 1 \end{vmatrix} - 0 \begin{vmatrix} 1 & 1 \\ 3 & 1 \end{vmatrix} + 2 \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = -6 + 0 + 0 = -6$$

1.2.8 Matriz inversa.

Sea \mathbf{A} una matriz cuadrada de orden $m \times m$. Si existe una matriz \mathbf{B} tal que $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, entonces \mathbf{B} se denota como \mathbf{A}^{-1} y se denomina matriz inversa.

La inversa de un matriz \mathbf{A} se calcula del siguiente modo

$$\mathbf{A}^{-1} = \frac{1}{|\mathbf{A}|} \text{adj}(\mathbf{A})$$

donde $\text{adj}(\mathbf{A})$ es la matriz adjunta o traspuesta de la matriz de cofactores de \mathbf{A} . Vemos que la condición necesaria y suficiente para que una matriz tenga inversa es que su determinante sea distinto de cero.

Algunas propiedades de la matriz inversa son las siguientes:

1. La matriz inversa es única.
2. $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$, la inversa de la inversa es la matriz original.
3. $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$, la inversa del producto es el producto invertido de las inversas.
4. $(\mathbf{A}')^{-1} = (\mathbf{A}^{-1})'$, la inversa de la traspuesta es la traspuesta de la inversa, esto es, el operador transposición y el operador inversión son intercambiables.

Una matriz cuadrada A se denomina ortogonal si $AA' = I$, esto es, si $A' = A^{-1}$

Ejemplo. La inversa de la matriz G es

$$G = \begin{pmatrix} 1 & 1 & 3 \\ 1 & 1 & 0 \\ 3 & 1 & 2 \end{pmatrix}^{-1} = \frac{1}{16} \begin{pmatrix} 2 & 1 & -3 \\ -2 & -7 & 3 \\ -2 & 2 & 0 \end{pmatrix}$$

1.3 Matrices cuadradas especiales.

1.3.1 Matriz diagonal: es una matriz cuadrada $A = [a_{ij}]$ de orden $m \times m$ cuyos elementos situados fuera de la diagonal principal son iguales a cero, $a_{ij} = 0 \quad \forall i \neq j$,

$$A = \begin{pmatrix} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ a_{11} & \vdots & \diagdown & \vdots \\ 0 & 0 & \dots & a_{mm} \end{pmatrix}$$

Escribimos una matriz diagonal como $A = \text{diag}(a_{11}, a_{22}, \dots, a_{mm})$

1.3.2 Matriz identidad: es una matriz cuadrada cuyos elementos de la diagonal principal son todos iguales a uno, y ceros en las demás posiciones y se denota por I .

$$I = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \diagdown & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

1.3.3 Matriz escalar: es una matriz diagonal cuyos elementos de la diagonal principal son todos iguales a $\lambda \in R$. Veremos que una matriz escalar es el producto de un número λ por una matriz identidad, λI .

1.3.4 Matriz triangular inferior: es una matriz cuadrada cuyos elementos por encima de la diagonal principal son todos nulos, $a_{ij} = 0 \quad \forall i < j$.

$$A = \begin{pmatrix} 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ a_{i1} & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{pmatrix}$$

1.3.5 Matriz nula, es una matriz (cuadrada o rectangular) cuyos elementos son todos iguales a cero, se denota por $[0]$.

1.3.6 Matriz simétrica, es una matriz de orden m $A = [a_{ij}]$ cuyos elementos satisfacen la condición $a_{ij} = a_{ji}$. Una matriz simétrica es igual a su traspuesta $A = A'$.

1.3.7 Matriz idempotente, es una matriz cuadrada que cumple $A^2 = A$.

1.3.8 Matriz ortogonal, es una matriz cuadrada que cumple $AA' = I$.

1.4 Valores propios y vectores propios de una matriz.

Definición. Sea A una matriz cuadrada de orden m . La ecuación característica de A es

$$|A - \lambda I| = 0$$

que es una ecuación polinomial en λ de grado m

$$\lambda^m + \alpha_1 \lambda^{m-1} + \dots + \alpha_{m-1} \lambda + \alpha_m = 0$$

Ejemplo. La ecuación característica de la matriz

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}, \text{ es}$$

$$|A - \lambda I| = \begin{vmatrix} 1 - \lambda & 2 \\ 2 & 1 - \lambda \end{vmatrix} = \lambda^2 - 2\lambda - 3 = 0$$

Definición. Las raíces $\lambda_1, \dots, \lambda_m$ de la ecuación característica $|A - \lambda I| = 0$ se denominan autovalores, valores propios, raíces características o raíces latentes de la matriz A .

Los valores propios de una matriz simétrica pertenecen al cuerpo de los números reales.

Ejemplo. Los autovalores de la matriz

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

son las raíces $\lambda_1 = -1$ y $\lambda_2 = 3$ de la ecuación características $|A - \lambda I| = \lambda^2 - 2\lambda - 3 = 0$

Definición. Se llama autovector, vector propio, vector característico o vector latente de la matriz cuadrada A a todo vector \vec{x} de orden $m \times 1$, distinto del vector nulo, que cumple

$$A\vec{x} = \lambda\vec{x}$$

Observación. Si \vec{x} es un autovector de A y $c \in R$, entonces $c\vec{x}$ también es un autovector de A .

Ejemplo. El autovector asociado al autovalor $\lambda_1 = -1$ cumple

$$\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} = -1 \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix}$$

De aquí

$$\vec{x}_1 = \begin{pmatrix} -x_{12} \\ x_{12} \end{pmatrix}$$

y el vector propio normalizado es

$$\begin{pmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix}$$

Proposición. Los vectores propios \vec{x}_i y \vec{x}_j asociados a los valores propios λ_i y λ_j distintos son ortogonales.

Proposición. Se cumplen las siguientes relaciones

1. $tr(A) = \sum_{i=1}^n \lambda_i$
2. $|A| = \prod_{i=1}^n \lambda_i$

1.5 Diagonalización de matrices.

Definición. Una matriz cuadrada $A = [a_{ij}]$ de orden m es diagonalizable cuando existe una matriz $P = [p_{ij}]$ de orden m no singular tal que

$$P^{-1}AP = D$$

donde $D = [d_{ij}]$ es una matriz diagonal de orden m .

Proposición. Si una matriz cuadrada A de orden m es diagonalizable, entonces los elementos de la diagonal principal de D son los autovalores $\lambda_1, \dots, \lambda_m$ de A , y las columnas P , son los correspondientes vectores propios de A $\vec{p}_1, \dots, \vec{p}_m$

Proposición. Una matriz cuadrada A con autovalores distintos es siempre diagonalizable.

Proposición. Si la matriz A es simétrica, $A = A'$, entonces $P^{-1} = P'$ y $A = PDP'$.

Definición. La descomposición espectral de una matriz simétrica de A de orden m es

$$A = \sum_{i=1}^m \lambda_i P_i P_i'$$

Definición. La raíz cuadrada de una matriz definida positiva A de orden m es

$$A^{1/2} = PD^{1/2}P' = \sum_{i=1}^m \sqrt{\lambda_i} P_i P_i'$$

en donde $D = \lambda_1^{1/2}, \dots, \lambda_m^{1/2}$

Definición. La descomposición de Cholesky de una matriz definida positiva A de orden m es

$$A = T'T$$

en donde T es una matriz triangular superior

Los elementos de la primera columna de la matriz T pueden obtenerse mediante las relaciones

$$t_{11} = \sqrt{a_{11}}$$

$$t_{ij} = \frac{a_{1j}}{t_{11}}, \quad j = 2, \dots, m$$

y los elementos de las siguientes columnas

$$t_{ii} = \sqrt{a_{ii} - \sum_{k=1}^{i-1} t_{ki}^2} \quad i = 2, \dots, m$$

$$t_{ij} = \frac{a_{ij} - \sum_{k=1}^{i-1} t_{ki} t_{kj}}{t_{ii}} \quad i = 2, \dots, m; j = i + 1, \dots, m$$

CAPITULO 2 FORMA CANÓNICA DE JORDAN Y EL CÁLCULO DE LA P-ÉSIMA RAÍZ DE UNA MATRIZ

2.1 Algunas aplicaciones del cálculo de la p-ésima raíz de una matriz no singular.

Algunas de las aplicaciones donde se hace necesario el cálculo de la p-ésima raíz de una matriz no singular son las siguientes.

Por ejemplo, la matriz logaritmo natural y la matriz exponencial juegan un papel importante en los modelos de Markov que se utilizan en una variedad de diferentes temas. Considere un proceso de Markov homogéneo en tiempo continuo en el que los individuos se mueven entre "n" estados. La matriz de transición de probabilidades $P(t) \in R^{n \times n}$ tiene (i, j) entradas que son iguales a la probabilidad de que un individuo saliendo del estado i para un tiempo cero esté en el estado j en un tiempo t . La suma de los elementos que pertenecen a cualquier fila de la matriz P es igual a la unidad, por lo que P es una matriz estocástica. Asociada con el proceso de Markov está la matriz de intensidad de transición $Q \in R^{n \times n}$, que está relacionada con P por medio de la siguiente expresión [14].

$$P(t) = e^{Qt}$$

los elementos de Q satisfacen

$$q_{ij} \geq 0, i \neq j, \sum_{j=1}^n q_{ij} = 0, \quad i = 1:n$$

para cualquier Q , e^{Qt} es no negativo para todo $t \geq 0$ y también la suma de sus filas es la unidad, por lo tanto e^{Qt} también es una matriz estocástica. Consideremos ahora un proceso de Markov en tiempo discreto con la matriz de transición de probabilidades P en la que las probabilidades de transición son independientes del tiempo. Podemos preguntarnos si existe alguna $P = e^Q$ para alguna matriz de intensidad Q , es decir, si el proceso se puede considerar como una

manifestación discreta de un proceso subyacente de Markov homogéneo en el tiempo. Si existe tal Q se llama *generador* y P se dice que es integrable. Actualmente no se conocen las condiciones necesarias y suficientes para la existencia de un generador.

Los investigadores en sociología [32], estadística [22], y finanzas [20] han investigado todo lo relacionado a la integración de este problema. Algunas características interesantes son las siguientes.

- Si P tiene, valores propios distintos reales y positivos, entonces el único logaritmo real y por lo tanto, el único generador candidato, es el logaritmo principal.
- Es posible que P pueda tener uno o más valores propios negativos reales, de modo que el logaritmo principal no está definido, sin embargo, todavía puede existir un generador. Por ejemplo, considere la matriz [32].

$$P = \frac{1}{3} \begin{bmatrix} +2x & 1-x & 1-x \\ 1-x & 1+2x & 1-x \\ 1-x & 1-x & 1+2x \end{bmatrix} \quad x = -e^{-2\sqrt{3}\pi} \approx -1.9 \times 10^{-5}. \quad P \text{ es diagonalizable,}$$

con valores propios $(1, x, yx)$. Cada logaritmo primario es complejo, ya que no puede tener valores propios complejos conjugados. Sin embargo, el logaritmo no primario

$$Q = 2\sqrt{3}\pi \begin{bmatrix} -2/3 & 1/2 & 1/6 \\ 1/6 & -2/3 & 1/2 \\ 1/2 & 1/6 & -2/3 \end{bmatrix}$$

es un generador.

- puede existir más de un generador.

Supongamos que una matriz de transición $P \equiv P(1)$ tiene un generador $Q = \log(P)$. Entonces Q puede ser utilizado para construir $P(t)$ en otros tiempos, por medio de $P(t) = \exp(Qt)$. Por ejemplo, si P es la matriz de transición para el período de tiempo de un año entonces la matriz

de transición para un mes es $P\left(\frac{1}{12}\right) = e^{10^{\mathbb{E}}(P)/12}$. Sin embargo, es más directo y eficiente calcular $P\left(\frac{1}{12}\right)$ como $P^{1/12}$ (aquí está el uso de la p-ésima raíz de una matriz no singular), evitando así el cálculo de un generador. De hecho, el escalamiento inverso estándar y el método de la cuadratura del logaritmo principal de una matriz requieren el cálculo de una raíz de la matriz. Del mismo modo, la matriz de transición para una semana puede ser calculada directamente como $P^{1/52}$ (de nuevo aquí se presenta la necesidad del cálculo de la p-ésima raíz de una matriz no singular). Este uso de las p-ésimas raíces de una matriz es sugerido por Waugh y Abel [35], mencionado por parte de Israel, Rosenthal y Wei [20], e investigado en detalle por Kreinin y Sidelnikova [25]. Los últimos autores, quienes motivados por los modelos de riesgo de crédito de cara a los problemas que pudieran generar el hecho que la raíz principal y el logaritmo principal de P pudieran tener valores propios negativos, muestran cómo regularizar de manera óptima estas matrices para lograr que cumplan con las propiedades requeridas.

Otra aplicación donde el cálculo de la p-ésima raíz de una matriz no singular juega un papel importante es en el cálculo de la *función sector-p matricial*, introducida por Shieh, Tsay y Wang [13] que es una generalización de la *Función signo matricial*. Para una p dada, la función sector matricial se puede definir a través de la forma canónica de Jordan, de una manera análoga como para la función signo matricial, pero ahora mapeando cada valor propio de la p-ésima raíz que esté más cercano a la unidad.

La función sector-p matricial se puede definir como $sect_p(A) = A(A^p)^{-1/p}$ (de nuevo aquí se presenta la necesidad del cálculo de la p-ésima raíz de una matriz no singular). La *función sector-p matricial* ha despertado el interés en la literatura de la teoría de control, ya que puede ser usada para determinar el número de valores propios en un sector específico y así obtener los

correspondientes subespacios invariantes [23]. Sin embargo, en la actualidad se carece de un buen método numérico para el cálculo de la función *sector-p matricial*.

2.2 LA FORMA CANÓNICA DE JORDAN

Sea una matriz $A \in \mathbb{C}^{n \times n}$ con p valores propios distintos $\lambda_1, \lambda_2, \dots, \lambda_p$ de multiplicidades m_1, m_2, \dots, m_p de modo que

$$m_1 + m_2 + \dots + m_p = n.$$

Entonces existe una matriz P no singular de tal manera que

$$A = PJP^{-1} \quad (2.1)$$

donde

$$J = \text{diag}(J_1, J_2, \dots, J_p), \quad J_i = J_i(\lambda_i) = \begin{pmatrix} \lambda_i & & & 1 \\ & \ddots & & \\ & & \lambda_i & \\ \lambda_i & & & \lambda_i \end{pmatrix} \in \mathbb{C}^{m_i \times m_i} \quad (2.2)$$

además, también se cumple que:

$$f(A) = Pf(J)P^{-1} \in \mathbb{C}^{n \times n}$$

los $J_i(\lambda_i)$ son los llamados bloques de Jordan y están dispuestos en diagonal, la forma (2.2) es llamada la forma canónica de Jordan de A .

El bloque de Jordan $J_i(\lambda_i)$ de tamaño $k \times k$ y valor propio λ_i es la matriz cuadrada $k \times k$ que tiene el número λ_i en la diagonal principal; el número 1 en la supradiagonal principal y el número 0 en las posiciones restantes.

Por ejemplo, el bloque de Jordan de tamaño 3×3 y valor propio 5 es

$$\begin{bmatrix} 5 & 1 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}$$

el bloque de Jordan de tamaño 2×2 y valor propio -4 es

$$\begin{bmatrix} -4 & 1 \\ 0 & -4 \end{bmatrix}$$

el bloque de Jordan de tamaño 1×1 y valor propio 8 es $\{8\}$.

Cuando un bloque de Jordan de tamaño $k \times k$ presenta una multiplicidad geométrica 1 y una multiplicidad algebraica k , λ_1 será su único valor propio.

Un suprabloque de Jordan de tamaño $m \times m$ y valor propio λ_1 es una matriz cuadrada $m \times m$ formada por:

- Uno o más bloques de Jordan de diversos o iguales tamaños, todo con el mismo valor propio λ_1 , ubicados diagonalmente (es decir, tales que la diagonal principal de cada bloque es parte de la diagonal principal del suprabloque),
- Ceros en los restantes términos del suprabloque.

Por ejemplo, un suprabloque de Jordan de tamaño 6×6 de valor propio 7 , formado por tres bloques de Jordan de tamaños 2×2 , 3×3 y 1×1 respectivamente es:

$$\begin{bmatrix} \boxed{7} & \boxed{1} & 0 & 0 & 0 & 0 \\ 0 & \boxed{7} & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{7} & \boxed{1} & 0 & 0 \\ 0 & 0 & 0 & \boxed{7} & \boxed{1} & 0 \\ 0 & 0 & 0 & 0 & \boxed{7} & 0 \\ 0 & 0 & 0 & 0 & 0 & \boxed{7} \end{bmatrix}$$

Una matriz cuadrada $n \times n$ es una forma canónica de Jordan si está formada por:

- Uno o más supra bloques de Jordan, ubicados diagonalmente.
- Ceros en los restantes términos de la matriz.

Por ejemplo, una forma canónica de Jordan con un suprabloque 2×2 formado por dos bloques de valor propio 7, un suprabloque 3×3 formado por dos bloques de valor propio cero y un supra bloque 1×1 de valor propio 3, es:

$$\begin{bmatrix} \boxed{7} & 0 & 0 & 0 & 0 & 0 \\ 0 & \boxed{7} & 0 & 0 & 0 & 0 \\ 0 & 0 & \boxed{0} & 0 & 0 & 0 \\ 0 & 0 & 0 & \boxed{0} & 1 & 0 \\ 0 & 0 & 0 & \boxed{0} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \boxed{3} \end{bmatrix}$$

Véase que en este ejemplo la matriz tiene valores propios $\lambda_1 = 7$ con multiplicidad algebraica 2 y geométrica 2, $\lambda_2 = 0$ con multiplicidad algebraica 3 y geométrica 2, y $\lambda_3 = 3$ con multiplicada algebraica 1 y geométrica 1.

Una forma canónica de Jordan tiene como valores propios a los números de la diagonal, con multiplicidad algebraica igual a la cantidad de veces en que esta repetido en la diagonal (o sea igual al tamaño del suprabloque respectivo), y multiplicidad geométrica igual a la cantidad de bloques que forman el respectivo suprabloque.

En particular la forma de Jordan es diagonal (es decir, todos sus términos fuera de la diagonal principal son nulos) si y sólo si todos los bloques de Jordan tienen tamaño 1×1 , y esto ocurre si y sólo si la cantidad de bloques que forman cada suprabloque es igual al tamaño del suprabloque, o sea, la multiplicidad geométrica es igual a la algebraica para todo valor propio.

El teorema fundamental de la forma canónica de Jordan, dice esencialmente que *“toda matriz cuadrada, mediante un cambio de base, se puede llevar a una forma canónica de Jordan”*.

Sea A una matriz cuadrada $n \times n$, entonces existe una matriz J que es una forma canónica de Jordan, y una matriz P invertible (o sea no singular), ambas $n \times n$ con términos complejos, tales que:

$$A = PJP^{-1}$$

Como se mencionó anteriormente, la matriz J es llamada *forma canónica de Jordan de A*. La matriz P se llama *matriz de cambio de base de Jordan*.

La forma canónica de Jordan J y la matriz P de cambio de base de Jordan para la matriz dada A no son únicas. En efecto, permutando los suprabloques de J , o los bloques dentro de cada suprabloque, se obtiene otra forma canónica de Jordan para A .

El siguiente teorema permite en muchos casos, calcular la forma canónica de Jordan de A , mediante el cálculo de los valores propios y los vectores propios de A y de la dimensión de sus subespacios propios.

Sea A una matriz $n \times n$ y sean $\lambda_1, \dots, \lambda_k$ sus valores propios diferentes, con multiplicidades algebraicas m_1, \dots, m_k y multiplicidades geométricas r_1, \dots, r_k respectivamente.

Entonces la forma canónica de Jordan de A tiene k suprabloques, cada uno de valor propio λ_i , tamaño $m_i \times m_i$, y formado por r_i bloques.

Por ejemplo, sea la matriz A :

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

La forma canónica de Jordan de A será:

El polinomio característico de A $\det(A - \lambda I) = -(\lambda - 4)(\lambda + 1)^3 = 0 \Rightarrow \lambda_1 = 4$ y $\lambda_2 = -1$ con multiplicidades algebraicas respectivas $m_1 = 1$ y $m_2 = 3$. Entonces J va a estar formada

por dos suprabloques: uno de valor propio 4 y tamaño 1×1 , y otro de valor propio -1 y tamaño 3×3 . Para determinar completamente la matriz J necesitamos saber cuántos bloques forman el suprabloque de tamaño 3×3 , o sea la multiplicidad geométrica r_2 del valor propio -1. Para ello hallemos el subespacio propio.

Subespacio propio del valor propio -1 = $n - \text{rango}(A - (-1 * I)) = 4 - 2 = 2$.

Luego entonces, el suprabloque 3×3 de valor propio -1, esta formado por dos bloques. Entonces estos dos bloques tendrán tamaños 1×1 y 2×2 , ya que 1 y 2 son los únicos dos números naturales mayores o iguales que 1 que suman 3. Entonces:

$$J = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

Las siguientes propiedades de la forma canónica de Jordan pueden deducirse en forma sencilla de los resultados ya vistos.

1. La forma canónica de Jordanes diagonal si y solo si las multiplicidades geométricas son iguales a las algebraicas para todo valor propio de A , esto sucede si y solo si A es diagonalizable.
2. Si A es diagonalizable entonces su forma canónica de Jordanes la matriz que tiene en la diagonal los valores propios de A repetidos tantas veces como sus multiplicidades algebraicas, y ceros en todos los demás términos.
3. Si la matriz A es 2×2 y tiene un solo valor propio doble λ_1 con multiplicidad geométrica igual a 1, entonces su forma canónica de Jordanes $J = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix}$.

Existen procedimientos, que veremos en la siguiente sección, para encontrar una matriz P de cambio de base Jordan: es decir una matriz invertible P tal que $A = PJP^{-1}$.

2.3 La función p-ésima raíz de una matriz no singular

Dada una matriz $A \in C^{n \times n}$, se dice que la matriz $X \in C^{n \times n}$ es una raíz p-ésima de la matriz A si cumple que

$$X^p = A$$

Como se comentó anteriormente entre las aplicaciones que requieren el cálculo de las raíces p-ésimas de una matriz se encuentra el cálculo de la *función sector-p matricial* definida por,

$$\text{sect}_p(A) = A(A^p)^{-1/p}$$

Otra posible aplicación aparece en el escalado de una matriz para el cálculo del logaritmo de una matriz. En este caso, se determina un entero positivo p de manera que

$$\text{Log}(A) = p \log A^{1/p}$$

con $\|A^{1/p}\|$ menor que un cierto valor.

Al igual que ocurre en el caso particular de raíces cuadradas, la raíz p-ésima de una matriz puede no existir o tener una o más soluciones. Si A es no singular, entonces admite al menos una raíz p-ésima; en caso contrario, la matriz A admite raíces p-ésimas dependiendo de la estructura de los divisores elementales de A correspondientes a los valores propios nulos.

Definición

Sea $A \in C^{n \times n}$ una matriz no singular con valores propios λ_i , $i = 1, 2, \dots, n$ de manera que $\arg(\lambda_i) \neq \pi$, $i = 1, 2, \dots, n$. La p-ésima raíz principal de A , denotada por $A^{1/p} \in C^{n \times n}$, se define como la matriz que satisface las dos condiciones siguientes [33]:

1. $(A^{1/p})^p = A$.
2. Los argumentos de los valores propios z_i de la matriz $A^{1/p}$ deben cumplir que $\{-\frac{\pi}{p} \geq \arg(z_i) \geq \frac{\pi}{p}\}$.

El siguiente teorema [33] permite caracterizar las p -ésimas raíces de una matriz cuadrada compleja a partir de su descomposición de Jordán.

Si $\lambda_i \in \mathbb{C}$ es distinto de cero, entonces el bloque de Jordán

$$J_{\lambda_i} = \begin{bmatrix} \lambda_i & 1 & \dots & 0 & 0 \\ 0 & \lambda_i & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \lambda_i & 1 \\ 0 & 0 & \dots & 0 & \lambda_i \end{bmatrix} \in \mathbb{C}^{n_i \times n_i}$$

tiene precisamente p p -ésimas raíces que son funciones de J_{λ_i} , definidas como

$$f_j(J_{\lambda_i}) = \begin{bmatrix} f_j(\lambda_i) & \frac{f_j^{(1)}(\lambda_i)}{1!} & \dots & \frac{f_j^{(n_i-2)}(\lambda_i)}{(n_i-2)!} & \frac{f_j^{(n_i-1)}(\lambda_i)}{(n_i-1)!} \\ 0 & f_j(\lambda_i) & \dots & \frac{f_j^{(n_i-3)}(\lambda_i)}{(n_i-3)!} & \frac{f_j^{(n_i-2)}(\lambda_i)}{(n_i-2)!} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & f_j \lambda_i & \frac{f_j^{(1)}(\lambda_i)}{1!} \\ 0 & 0 & \dots & 0 & f_j \lambda_i \end{bmatrix} \in \mathbb{C}^{n_i \times n_i} \quad j = 1, 2, \dots, p$$

donde $f_j(\lambda) = \sqrt[p]{\lambda}$ denota una de las p ramas de la función p -ésima raíz en el entorno de λ_i .

En [9] se establece lo siguiente: Sea $A \in \mathbb{C}^{n \times n}$ una matriz no singular con la descomposición de Jordán

$$A = X \text{diag} (J_{\lambda_1}, J_{\lambda_2}, \dots, J_{\lambda_r}) X^{-1}$$

donde $J_{\lambda_s} \in \mathbb{C}^{n_s \times n_s}$, $\sum_{s=1}^r n_s = n$, y s ($s \leq r$) el número de valores propios distintos de A se verifican entonces las siguientes propiedades:

1. A tiene precisamente p^s raíces de índice p , dadas por

2. $F_j = X \text{diag}(f_{j_1}(J_{\lambda_1}), f_{j_2}(J_{\lambda_2}), \dots, f_{j_r}(J_{\lambda_r})) X^{-1}$, $j = 1, 2, \dots, p^s$, correspondientes a todas las posibles elecciones de j_1, j_2, \dots, j_r , $j_i \in \{1, 2, \dots, p\}$, $i = 1, 2, \dots, r$, sujetas a la restricción de que $j_k = j_l$ si $\lambda_k = \lambda_l$.

2.4 Ejemplos de cómo calcular la p-ésima raíz de una matriz no singular utilizando la forma canónica de Jordan.

Ejemplo 1. Dada la matriz A calcule 2 raíces cúbicas

$$A = \begin{pmatrix} 7 & 10 \\ 15 & 22 \end{pmatrix}$$

Se tendrá que calcular la representación canónica de Jordan de la matriz A

$$A = PJP^{-1}$$

Primeramente, se calculará la matriz J cuya diagonal principal estará formada por los valores propios de A , y después se calcularán los vectores propios asociados a cada uno de los valores propios encontrados.

Cálculo de los valores propios de A .

$$|A - \lambda I| = 0 \quad (2.3)$$

desarrollando el determinante de (2.3) se obtiene el polinomio característico que es una función de λ

$$\begin{vmatrix} 7 - \lambda & 10 \\ 15 & 22 - \lambda \end{vmatrix} = (7 - \lambda)(22 - \lambda) - 150 = 0 \quad (2.4)$$

resolviendo (2.4) se tiene que $\lambda_1 = 0.1386$ y $\lambda_2 = 28.8614$ por lo tanto la matriz J tendrá la siguiente forma

$$J = \begin{bmatrix} 0.1386 & 0 \\ 0 & 28.8614 \end{bmatrix} \quad (2.5)$$

Cálculo del vector propio \vec{U} asociado a $\lambda_1 = 0.1386$

$$\begin{bmatrix} 7 - \lambda_1 & 10 \\ 15 & 22 - \lambda_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (2.6)$$

sustituyendo valores en (2.6) se tiene

$$\begin{bmatrix} 6.8614 & 10 \\ 15 & 21.8614 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (2.7)$$

resolviendo el sistema homogéneo (2.7)

$$6.9861u_1 + 10u_2 = 0 \quad (2.8)$$

$$15u_1 + 21.8614u_2 = 0 \quad (2.9)$$

de (2.9) se expresa a u_1 en función de u_2 quedando $u_1 = -1.4574u_2$ luego de (2.8) tenemos que $u_2 = u_2$, por lo tanto, el vector propio buscado es;

$$\vec{U} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} = \begin{bmatrix} -1.4574 u_2 \\ u_2 \end{bmatrix} \quad (2.10)$$

por lo tanto, la primera columna de la matriz P es el vector propio $\begin{bmatrix} -1.4574 \\ 1 \end{bmatrix}$

Cálculo del vector propio \vec{V} asociado a $\lambda_2 = 28.8614$

$$\begin{bmatrix} 7 - \lambda_2 & 10 \\ 15 & 22 - \lambda_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (2.11)$$

sustituyendo valores en (2.11) se tiene

$$\begin{bmatrix} -21.8614 & 10 \\ 15 & -6.1386 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad (2.12)$$

resolviendo el sistema homogéneo (2.12)

$$-21.8614v_1 + 10v_2 = 0 \quad (2.13)$$

$$15v_1 - 6.1386v_2 = 0 \quad (2.14)$$

de (2.14) se expresa a v_1 en función de v_2 quedando $v_1 = 0.4574v_2$ luego de (2.13) tenemos que $v_2 = v_2$, por lo tanto, el vector propio buscado es;

$$\vec{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0.4574v_2 \\ v_2 \end{bmatrix} \quad (2.15)$$

Por lo tanto, la segunda columna de la matriz P es el vector propio $\begin{bmatrix} 0.4574 \\ 1 \end{bmatrix}$

$$P = \begin{bmatrix} -1.4574 & 0.4574 \\ 1 & 1 \end{bmatrix}, \quad y \quad P^{-1} = \begin{bmatrix} -0.5222 & 0.2389 \\ 0.5222 & 0.7611 \end{bmatrix}$$

por lo tanto, la representación canónica de Jordan de la matriz A es

$$A = \begin{bmatrix} -1.4574 & 0.4574 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.1386 & 0 \\ 0 & 28.8614 \end{bmatrix} \begin{bmatrix} -0.5222 & 0.2389 \\ 0.5222 & 0.7611 \end{bmatrix}$$

Por lo tanto, se requiere que:

$$A^{\frac{1}{3}} = \begin{bmatrix} -1.4574 & 0.4574 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.1386 & 0 \\ 0 & 28.8614 \end{bmatrix}^{1/3} \begin{bmatrix} -0.5222 & 0.2389 \\ 0.5222 & 0.7611 \end{bmatrix}$$

Como A tiene $s = 2$ valores propios diferentes entonces tendrá P^s raíces cúbicas, en este caso como $p = 3$ y $s = 2$ entonces tiene 9 raíces cúbicas, de las cuales sólo calcularemos 2 utilizando para ello la fórmula de Moivre (2.16), que para un número complejo ($z = a + bi$) se define como

$$z^{\frac{1}{p}} = |z|^{\frac{1}{p}} \left(\cos \left[\frac{\theta + 2k\pi}{p} \right] + i \operatorname{sen} \left[\frac{\theta + 2k\pi}{p} \right] \right), \quad k = 0, 1, 2, \dots, \quad (2.16)$$

Calculo de las raíces cúbicas de $\lambda_1 = 0.1386 + 0i$

$$\theta = \tan^{-1} \left(\frac{0}{0.1386} \right) = 0^\circ$$

Para $k = 0$ se tiene que la primera raíz cúbica de $\lambda_1^{1/3} = 0.5175$

Para $k = 1$ se tiene que la segunda raíz cúbica de $\lambda_1^{1/3} = -0.2587 + 0.4481i$

Para $k = 2$ se tiene que la tercera raíz cúbica de $\lambda_1^{1/3} = -0.2587 - 0.4481i$

Calculo de las raíces cúbicas de $\lambda_2 = 28.8614 + 0i$

$$\theta = \tan^{-1}\left(\frac{0}{28.8614}\right) = 0^\circ$$

Para $k = 0$ se tiene que la primera raíz cúbica de $\lambda_2^{1/3} = 3.0674$

Para $k = 1$ se tiene que la segunda raíz cúbica de $\lambda_2^{1/3} = -1.5337 + 2.6565i$

Para $k = 2$ se tiene que la tercera raíz cúbica de $\lambda_2^{1/3} = -1.5337 - 2.6565i$

Tabla 2.1 Raíces cúbicas de los valores propios de A

$\lambda_1^{1/3}$	$\lambda_2^{1/3}$
0.5175	3.0674
$-0.2587 + 0.4481i$	$-1.5337 + 2.6565i$
$-0.2587 - 0.4481i$	$-1.5337 - 2.6565i$

Para la primera raíz cúbica de **A** tomaremos a **J** con diagonal principal formada con la primera raíz cúbica de λ_1 y la primera raíz cúbica de λ_2 , por lo tanto, tendremos:

$$A^{1/3} = PJ^{1/3}P^{-1} = \begin{bmatrix} -1.4574 & 0.4574 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5175 & 0 \\ 0 & 3.0674 \end{bmatrix} \begin{bmatrix} -0.5222 & 0.2389 \\ 0.5222 & 0.7611 \end{bmatrix}$$

$$A^{1/3} = \begin{bmatrix} 1.12643 & 0.88770 \\ 1.33161 & 2.45821 \end{bmatrix}$$

Para la segunda raíz cúbica de **A** tomaremos a **J** con diagonal principal formada con la primera raíz cúbica de λ_1 y la segunda raíz cúbica de λ_2 , por lo tanto, tendremos:

$$A^{1/3} = \begin{bmatrix} -1.4574 & 0.4574 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 0.5175 & 0 \\ 0 & -1.5337 + 2.6565i \end{bmatrix} \begin{bmatrix} -0.5222 & 0.2389 \\ 0.5222 & 0.7611 \end{bmatrix}$$

$$A^{1/3} = \begin{bmatrix} 0.02744 + 0.63452i & -0.71407 + 0.92479i \\ -1.07108 + 1.38722i & -1.04369 + 2.02186i \end{bmatrix}$$

Ejemplo 2. Dada la matriz B , calcule 2 raíces cúbicas de B

$$B = \begin{bmatrix} -1 & 1 & 2 \\ -3 & 3 & 5 \\ -4 & 3 & 8 \end{bmatrix}$$

Cálculo de los valores propios de B .

$$|B - \lambda I| = 0 \quad (2.17)$$

Desarrollando el determinante de (2.17) se obtiene el polinomio característico que es una función de λ

$$\begin{vmatrix} -1-\lambda & 1 & 2 \\ -3 & 3-\lambda & 5 \\ -4 & 3 & 8-\lambda \end{vmatrix} = 1 - 9\lambda + 10\lambda^2 - \lambda^3 = 0 \quad (2.18)$$

Resolviendo (2.18) se tiene que $\lambda_1 = 0.129505$, $\lambda_2 = 0.856652$ y $\lambda_3 = 9.01384$ por lo tanto la matriz J tendrá la siguiente forma

$$J = \begin{bmatrix} 0.129505 & 0 & 0 \\ 0 & 0.856652 & 0 \\ 0 & 0 & 9.01384 \end{bmatrix} \quad (2.19)$$

Cálculo del vector propio \vec{U} asociado a $\lambda_1 = 0.129505$

$$\begin{bmatrix} -1 - \lambda_1 & 1 & 2 \\ -3 & 3 - \lambda_1 & 5 \\ -4 & 3 & 8 - \lambda_1 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (2.20)$$

Sustituyendo valores en (2.20) surge el sistema homogéneo formado por las siguientes tres ecuaciones.

$$-1.129505u_1 + u_2 + 2u_3 = 0 \quad (2.21)$$

$$-3u_1 + 2.87049u_2 + 5u_3 = 0 \quad (2.22)$$

$$-4u_1 + 3u_2 + 7.87049u_3 = 0 \quad (2.23)$$

Multiplicando (2.22) por -1.0451, sumando con (2.23) y expresando u_1 en función de u_3 se tiene

$u_1 = \frac{2.814901}{6465} u_3 = 3.05892u_3$. Seguidamente multiplicando (2.21) por -2.65603, sumando con

(2.22) y expresando u_2 en función de u_3 se tiene $u_2 = \frac{0.31206}{0.21446} u_3 = 1.45510u_3$ y la única opción

restante para u_3 es $u_3 = u_3$, por lo tanto el vector propio buscado es

$$\vec{U} = \begin{bmatrix} 3.05892u_3 \\ 1.45510u_3 \\ u_3 \end{bmatrix} = u_3 \begin{bmatrix} 3.05892 \\ 1.45510 \\ 1 \end{bmatrix}$$

la primera columna de la matriz P es $\begin{bmatrix} .05892 \\ 1.45510 \\ 1 \end{bmatrix}$

Cálculo del vector propio \vec{V} asociado a $\lambda_2=0.856652$

$$\begin{bmatrix} -1 - \lambda_2 & 1 & 2 \\ -3 & 3 - \lambda_2 & 5 \\ -4 & 3 & 8 - \lambda_2 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (2.24)$$

Sustituyendo valores en (2.24) surge el sistema homogéneo formado por las siguientes tres ecuaciones.

$$-1.856652v_1 + v_2 + 2v_3 = 0 \quad (2.25)$$

$$-3v_1 + 2.143348v_2 + 5v_3 = 0 \quad (2.26)$$

$$-4v_1 + 3v_2 + 7.143348v_3 = 0 \quad (2.27)$$

Multiplicando (2.26) por -1.0451, sumando con (2.23) y expresando v_1 en función de v_3 se tiene

$v_1 = \frac{-0.144951v_3}{0.199038} = -0.728258v_3$. Seguidamente multiplicando (2.25) por -1.615812,

sumando con (2.22) y expresando v_2 en función de v_3 se tiene $v_2 = \frac{1.768377}{0.527536} v_3 =$

$-3.352144v_3$ y la única opción restante para v_3 es $v_3 = v_3$, por lo tanto el vector propio

buscado es

$$\vec{v} = \begin{bmatrix} -0.728258 \\ -3.352144 \\ v_3 \end{bmatrix} = v_3 \begin{bmatrix} -0.728258 \\ -3.352144 \\ 1 \end{bmatrix}$$

la segunda columna de la matriz P es $\begin{bmatrix} 3.352144 \\ -0.728258 \\ 1 \end{bmatrix}$

Cálculo del vector propio \vec{w} asociado a $\lambda_3 = 9.01384$

$$\begin{bmatrix} -\lambda_3 & 1 & 2 \\ -3 & 3 - \lambda_3 & 5 \\ -1 & -4 & 8 - \lambda_3 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad (2.28)$$

Sustituyendo valores en (2.28) surge el sistema homogéneo formado por las siguientes tres ecuaciones.

$$-10.01384w_1 + w_2 + 2w_3 = 0 \quad (2.29)$$

$$-3w_1 - 6.01384w_2 + 5w_3 = 0 \quad (2.30)$$

$$-4w_1 + 3w_2 - 1.01384w_3 = 0 \quad (2.31)$$

Multiplicando (2.30) por 0.498849, sumando con (2.31) y expresando w_1 en función de w_3 se

tiene $w_1 = \frac{-1.480407}{-5.496548} w_3 = 0.269334w_3$. Seguidamente multiplicando (2.29) por -0.299585,

sumando con (2.30) y expresando w_2 en función de w_3 se tiene $w_2 = \frac{-4.400829}{-6.313425} w_3 =$

$0.697059w_3$ y la única opción restante para w_3 es $w_3 = w_3$, por lo tanto el vector propio

buscado es

$$\vec{w} = \begin{bmatrix} 0.269334w_3 \\ 0.697059w_3 \\ w_3 \end{bmatrix} = w_3 \begin{bmatrix} 0.269334 \\ 0.697059 \\ 1 \end{bmatrix}$$

la tercera columna de la matriz P es $\begin{bmatrix} .269334 \\ 0.697059 \\ 1 \end{bmatrix}$

Por lo tanto, la representación canónica de Jordan de la matriz B es

$$B = \begin{bmatrix} 3.05892 & -0.728258 & 0.269334 \\ 1.45510 & -3.352144 & 0.697059 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.129505 & 0 & 0 \\ 0 & 0.856652 & 0 \\ 0 & 0 & 9.01384 \end{bmatrix} \begin{bmatrix} 0.38419 & -0.09465 & -0.03749 \\ 0.07192 & -0.26468 & 0.16512 \\ -0.45612 & 0.35933 & 0.87237 \end{bmatrix}$$

Utilizando la fórmula de Moivre (2.16) calcularemos las raíces cúbicas de los valores propios de la matriz B las cuales se muestran en la tabla 2.2

Tabla 2.2 Raíces cúbicas de los valores propios de B

$\lambda_1^{1/3}$	$\lambda_2^{1/3}$	$\lambda_3^{1/3}$
0.50594	0.94973	2.081150
-0.25297+0.43815i	-0.47487+0.82249i	-1.04057+1.80233i
-0.2579-0.43815i	-0.47487-0.82249i	-1.04057-1.80233i

De estas nueve raíces sólo seleccionaremos 2 combinaciones de 3 raíces para la matriz J .

Para la primera raíz cúbica de B seleccionamos el primer renglón de la tabla 2.2

$$B^{1/3} = \begin{bmatrix} 3.05892 & -0.728258 & 0.269334 \\ 1.45510 & -3.352144 & 0.697059 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.50594 & 0 & 0 \\ 0 & 0.94973 & 0 \\ 0 & 0 & 2.081150 \end{bmatrix} \begin{bmatrix} 0.38419 & -0.09465 & -0.03749 \\ 0.07192 & -0.26468 & 0.16512 \\ -0.45612 & 0.35933 & 0.87237 \end{bmatrix}$$

$$B^{1/3} = \begin{bmatrix} 0.28918 & 0.23799 & 0.31674 \\ -0.60783 & 1.29425 & 0.71222 \\ -0.68657 & 0.44857 & 1.95339 \end{bmatrix}$$

Para la segunda raíz cúbica de B seleccionamos los dos primeros elementos del renglón 1 y el último elemento del renglón 2 de la tabla 2.2.

$$B^{1/3} = \begin{bmatrix} 3.05892 & -0.728258 & 0.269334 \\ 1.45510 & -3.352144 & 0.697059 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0.50594 & 0 & 0 \\ 0 & 0.94973 & 0 \\ 0 & 0 & -1.0406 + 1.80231i \end{bmatrix} \begin{bmatrix} 0.38419 & -0.09465 & -0.03749 \\ 0.07192 & -0.26468 & 0.16512 \\ -0.45612 & 0.35933 & 0.87237 \end{bmatrix}$$

$$B^{1/3} = \begin{bmatrix} 0.6727 - 0.2214i & -0.0641 + 0.1744i & 0.4167 + 0.4235i \\ 0.3847 - 0.5730i & 0.5123 + 0.4514i & 1.1861 + 1.0959i \\ 0.7373 - 0.8221i & -0.6732 + 0.6476i & 0.7699 + 1.5723i \end{bmatrix}$$

Como puede verse, el cálculo de la p -ésima raíz de una matriz cuadrada no singular utilizando la forma canónica de Jordan no es una tarea sencilla, además la forma canónica de Jordan es

muy sensible a la propagación de pequeños errores y su cálculo implica un costo computacional considerable, por lo tanto, este método no es atractivo para el cálculo de p -ésimas raíces de matrices cuadradas no singulares, sólo es muy útil para su caracterización. Debido a lo anteriormente expuesto, lo más adecuado para el cálculo de p -ésimas raíces de matrices no singulares, son los métodos iterativos que se estudiarán en los siguientes capítulos de este trabajo de tesis.

CAPÍTULO 3 LOS MÉTODOS DE NEWTON PARA EL CASO MATRICIAL

3.1 Los métodos de Newton

Los Métodos de Newton para calcular la p -ésima raíz principal de una matriz se han estudiados durante casi 60 años y actualmente estos métodos ya son bien comprendidos. En lo que respecta al Método Simplificado de Newton (MSN), en 1958 Laasonen [26] demostró su convergencia, pero también observó sin demostrarlo, que para algunos casos como en el cálculo de la raíz cuadrada presentaba inestabilidad numérica cuando el número de condición de la matriz era mayor que nueve. Desde entonces y hasta la fecha han estado surgiendo variantes del Método de Newton (MN) para el cálculo de la p -ésima raíz principal de una matriz no singular, que son convergentes y numéricamente estables. Con respecto a la existencia y caracterización de la p -ésima raíz de una matriz, esto se encuentra bien documentado en [33], [18] y [2].

En cuanto a los algoritmos disponibles para el cálculo de la p -ésima raíz principal de una matriz no singular, se comenta brevemente lo siguiente.

Björck y Hammarling [3] propusieron un método para calcular la raíz cuadrada y la raíz cúbica de una matriz utilizando la descomposición de Schur, sin embargo, si la matriz es real este método puede requerir aritmética compleja, incluso si la raíz que se está buscando es real. Este método fue posteriormente extendido por Higham [11], quien sugirió el uso de la forma real de la descomposición de Schur, y generalizado después por Smith [33] para el cálculo de la p -ésima raíz de una matriz. El método de Smith implementado en MATLAB toolbox [13], es numéricamente estable y requiere $(28 + (p - 1)/3)n^3$ operaciones aritméticas por iteración [14, p. 176].

Darío A. Bini y otros [2] presentaron resultados teóricos novedosos sobre el cálculo de la p -ésima raíz principal de una matriz. En particular, mostraron que la p -ésima raíz está relacionada con la función signo matricial y con la factorización de Wiener-Hopf, y que puede ser expresada como una integral sobre un círculo unitario. Los resultados se utilizan en el diseño y análisis de nuevos algoritmos para el cálculo de la p -ésima raíz. También analizaron las propiedades de estabilidad y convergencia del (MSN) para el cálculo de la p -ésima raíz inversa de una matriz. Iannazzo [18] [19] propuso dos nuevas iteraciones para el cálculo de la p -ésima raíz que son convergentes y numéricamente estables. Más recientemente Chun-Hua Guo [8] analizó la manera de como calcular la p -ésima raíz de una matriz, preconditionándola si fuese necesario, utilizando el (MSN) o el Método de Halley, y también explicó cómo la convergencia de estos métodos se puede mejorar cuando los valores propios de la matriz son conocidos.

En este trabajo de tesis se presenta una variante para eliminar la inestabilidad numérica del (MSN) para el cálculo de la p -ésima raíz principal de una matriz no singular, la variante consiste en lo siguiente: en cada k -ésima iteración se factoriza la matriz $\mathbf{A}_k = \mathbf{A} \|\mathbf{A}\|_F^{-1}$ expresándola como el producto de dos matrices \mathbf{B}_k y \mathbf{C}_k^{p-1} que conmutan entre sí, es decir $\mathbf{A}_k = \mathbf{B}_k \mathbf{C}_k^{p-1} = \mathbf{C}_k^{p-1} \mathbf{B}_k$. El proceso anterior elimina la inestabilidad numérica del (MSN) convirtiéndolo en un método atractivo para el cálculo de la p -ésima raíz de una matriz no singular, por ser robusto, convergente, computacionalmente económico y para propósitos prácticos es numéricamente estable como será demostrado posteriormente. De aquí en adelante en todos los métodos se trabajará con la matriz \mathbf{A}_k en lugar de la matriz \mathbf{A} .

3.2 El Método de Newton (MN) para el cálculo de la p -ésima raíz de una matriz.

Para dar solución a una función general $F: C^{n \times n} \rightarrow C^{n \times n}$ que tiene la siguiente forma

$$F(X) = 0 \quad (3.1)$$

el (MN) se define como [31, p. 133], [6, p. 86]

$$X_{k+1} = X_k - F'(X_k)^{-1}F(X_k), \quad k = 0, 1, 2, \dots, \quad (3.2)$$

donde X_0 es un valor inicial conocido y F' es la derivada de Fréchet de F [27, p. 225].

El (MN) ha sido utilizado para calcular la raíz cuadrada de matrices por Higham [9] [10] [12], Iannazzo [17], Meini [29], Long [28] y Mendoza y Gómez [30]. El problema más general, que es el de calcular la p -ésima raíz de una matriz no singular es discutido por Smith [33], el cual tiene la siguiente forma

$$F(X) \equiv X^p - An = 0 \quad (3.3)$$

considerando un desarrollo de Taylor a primer orden para F en el entorno de X

$$\begin{aligned} F(X + H) &= F(X) + F'(X)H + O(H^2) \\ F'(X)H &= F(X + H) - F(X) - O(H^2) \end{aligned} \quad (3.4)$$

calculando $F(X + H)$ a partir de (3.3) se tiene

$$\begin{aligned} F(X + H) &= (X + H)^p - An \\ &= F(X) + X^{p-1}H + X^{p-2}HX + \dots + XHX^{p-2} + HX^{p-1} + O(H^2) \end{aligned} \quad (3.5)$$

combinando (3.4) y (3.5)

$$F'(X)H = X^{p-1}H + X^{p-2}HX + \dots + XHX^{p-2} + HX^{p-1} \quad (3.6)$$

de (3.2) se tiene

$$F'(X_k)(X_{k+1} - X_k) = -F(X_k)$$

pero $(X_{k+1} - X_k) = H_k$, entonces

$$F'(X)H = -F(X) \quad (3.7)$$

sustituyendo (3.7) en (3.6) se llega al (MN) para el cálculo de la p-ésima raíz principal de una matriz no singular.

$$\left. \begin{aligned} & \text{dado } X_0 = I \\ & X_k^{p-1}H_k + X_k^{p-2}H_kX_k + \dots + X_kH_kX_k^{p-2} + H_kX_k^{p-1} = An - X_k^p \\ & X_{k+1} = X_k + H_k \\ & A^{\frac{1}{p}} = X(\|A\|_F)^{\frac{1}{p}} \end{aligned} \right\}, k = 0, 1, 2, \dots \quad (3.8)$$

El teorema de convergencia local estándar para el (MN) [31] dice que, siempre que $\|X - X_0\|$ sea suficientemente pequeño y la transformación lineal $F'(X)$ sea no singular, la iteración (3.8) converge cuadráticamente a una p-ésima raíz principal de An , de donde fácilmente puede obtenerse la p-ésima raíz de A como se muestra en (3.8).

Para $p > 2$ el (MN) nos obliga a resolver (3.8) para H_k , esto se puede llevar a cabo con la ayuda del operador vec que para An se define como $vec(An)$ (que significa el apilamiento de las columnas de An una debajo de otra iniciando de izquierda a derecha), junto con el producto de Kronecker $An \otimes B = (a_{ij}B)$.

Aplicando el operador vec a (3.8) y utilizando la propiedad $vec(AnXB) = (B^* \otimes An)vec(X)$ [5, Problema 6.4, p. 358] y [15, cap. 4], se obtiene

$$\left((I \otimes X^{p-1}) + (X^* \otimes X^{p-2}) + \dots + ((X^{p-1})^* \otimes I) \right) vec(H) = vec(An - X^p), \quad (3.9)$$

el sistema lineal (3.9) puede ser resuelto utilizando cualquier método estándar, siempre que la matriz de coeficientes sea no singular. Sin embargo (3.9) es un sistema lineal de $n^2 \times n^2$, lo que implica un alto costo computacional para su solución que de acuerdo a [14, p. 72, 336], es del orden de $2n^{6+\vartheta} \log_2 p$, $1 \leq \vartheta \leq 2$ operaciones aritmeticas por iteración, lo cual hace poco atractivo al (MN) para el cálculo de la p-ésima raíz principal de una matriz no singular.

Una suposición razonable para reducir el costo computacional de (3.8) es que el producto $\mathbf{X}_k \mathbf{H}_k$ conmuta [10, teorema 1], [33] es decir

$$\mathbf{X}_k \mathbf{H}_k = \mathbf{H}_k \mathbf{X}_k, \quad k = 0, 1, 2, \dots, \quad (3.10)$$

entonces (3.8) se puede expresar como

$$\left. \begin{array}{l} \text{dado } \mathbf{X}_0 = \mathbf{I} \\ p\mathbf{X}_k^{p-1} \mathbf{H}_k = p\mathbf{H}_k \mathbf{X}_k^{p-1} = \mathbf{A}\mathbf{n} - \mathbf{X}_k^p \\ \mathbf{X}_{k+1} = \mathbf{X}_k + \mathbf{H}_k \\ \mathbf{A}^{\frac{1}{p}} = \mathbf{X}(\|\mathbf{A}\|_F)^{\frac{1}{p}} \end{array} \right\}, \quad k = 0, 1, 2, \dots, \quad (3.11)$$

a partir de (3.10) y (3.11), se obtienen las dos iteraciones del Método Simplificado de Newton (MSN), que son las siguientes

$$\left. \begin{array}{l} \text{dado } \mathbf{Y}_0 = \mathbf{I} \\ \mathbf{Y}_{k+1} = \frac{1}{p} \left((p-1)\mathbf{Y}_k + \mathbf{Y}_k^{1-p} \mathbf{A}\mathbf{n} \right) \\ \mathbf{A}^{\frac{1}{p}} = \mathbf{Y}(\|\mathbf{A}\|_F)^{\frac{1}{p}} \end{array} \right\}, \quad k = 0, 1, 2, \dots, \quad (3.12)$$

$$\left. \begin{array}{l} \text{dado } \mathbf{Z}_0 = \mathbf{I} \\ \mathbf{Z}_{k+1} = \frac{1}{p} \left((p-1)\mathbf{Z}_k + \mathbf{A}\mathbf{n} \mathbf{Z}_k^{1-p} \right) \\ \mathbf{A}^{\frac{1}{p}} = \mathbf{Z}(\|\mathbf{A}\|_F)^{\frac{1}{p}} \end{array} \right\}, \quad k = 0, 1, 2, \dots, \quad (3.13)$$

siempre y cuando \mathbf{Y}_k y \mathbf{Z}_k sean no singular. Las iteraciones (3.12) y (3.13) implican que la secuencia \mathbf{Y}_k^{1-p} conmuta con $\mathbf{A}\mathbf{n}$ para toda k [4]. El costo computacional de (3.12) y (3.13) es del orden de $2n^3(2 + \vartheta \log_2 p)$, $1 \leq \vartheta \leq 2$ operaciones aritméticas por iteración.

Las propiedades de convergencia y estabilidad de (3.12) y (3.13), son temas importantes que desempeñan un papel fundamental en el diseño de un algoritmo para el cálculo de la p -ésima raíz principal de una matriz. Por ejemplo, para iniciar el proceso iterativo del (MSN), Hoskins y Walton [16], y Smith [33] toman como valor inicial \mathbf{Y}_0 igual a la matriz \mathbf{A} .

Desafortunadamente, como se discute en [33], esta decisión conduce a una región de convergencia que no es lo suficientemente buena como para diseñar un algoritmo simple de convergencia global.

Iannazzo [18, iteración 3.6] ha presentado una modificación adecuada al (MSN), que garantiza su convergencia y estabilidad numérica. Iannazzo demuestra que cuando la solución inicial es $Y_0 = I$, la convergencia se produce cuando los valores propios de A pertenecen al conjunto $D = \{\lambda_i \in \mathbb{C}: \text{Re}(\lambda_i) > 0, |\lambda_i| \leq 1\}$, esta restricción puede ser relajada escalando la matriz por medio de una norma adecuada. La iteración que se obtiene, tiene una convergencia cuadrática y un costo computacional del orden de $2n^3(3 + \vartheta \log(p))$, $1 \leq \vartheta \leq 2$ operaciones aritméticas por iteración. Más recientemente Iannazzo [19, iteración 5.10] presenta una iteración del tipo racional de rápida convergencia y que es numéricamente estable para calcular la p -ésima raíz de una matriz no singular que tiene un costo computacional del orden de $2n^3(5 + \vartheta \log(p))$, $1 \leq \vartheta \leq 2$ operaciones aritméticas por iteración.

3.3 Análisis de convergencia del (MSN) para el cálculo de la p -ésima raíz de una matriz no singular.

Se presenta la demostración *del teorema 1*, donde se concluye que el (MSN) es convergente bajo ciertas condiciones que son fáciles de cumplir, la demostración es diferente a las contempladas en [18] y [33].

Se tendrá en cuenta la relación que existe entre la iteración (3.8) y las iteraciones simplificadas (3.12) y (3.13). Se recalca que las iteraciones del (MN) estarán bien definidas, si y sólo si, para toda k , la ecuación (3.8) tiene solución única, es decir, que la derivada de Fréchet $F'(X_k)$ sea no singular. Para el análisis se trabajará con la iteración (3.12).

Teorema 1. Sea $An \in C^{n \times n}$ una matriz no singular y diagonalizable con valores propios λ_i , $\{\lambda_i \in C, 0 < |\lambda_i| \leq 1, \text{Re}(\lambda_i) > 0, i = 1:n\}$. Si $Y_0 = I$ y $\|Y - Y_0\|$ es suficientemente pequeño con Y_k^{1-p} no singular, entonces el proceso iterativo (3.12) converge a $Y = (\|A\|_F)^{\frac{-1}{p}} A^{\frac{1}{p}}$ cuando $k \rightarrow \infty$, es decir $A^{\frac{1}{p}} = Y(\|A\|_F)^{\frac{1}{p}}$ es la p -ésima raíz principal de A .

Demostración:

por las características de An , existe una matriz no singular Q tal que [24]

$$Q^{-1}An Q = \tau = \text{diag}(\lambda_1, \dots, \lambda_n), i = 1:n \quad (3.14)$$

$$Q^{-1}Y_k Q = D_k = \text{diag}(d_1^{(k)}, \dots, d_n^{(k)}), i = 1:n, k = 0,1,2, \dots, \quad (3.15)$$

donde $\lambda_1, \dots, \lambda_n$ son los valores propios de An y $d_1^{(k)}, \dots, d_n^{(k)}$ son los valores propios de Y_k .

a partir de (3.14 y (3.15), se diagonaliza (3.12) resultando

$$D_{k+1} = \frac{1}{p} \left((p-1)D_k + D_k^{1-p} \tau \right), \quad (3.16)$$

la ecuación (3.16) representa esencialmente n iteraciones escalares desacopladas del (MSN)

para el cálculo de las p -ésimas raíces $(d_i^{(k)})^{\frac{1}{p}}$, es decir

$$d_i^{(k+1)} = \frac{1}{p} \left((p-1)d_i^{(k)} + \frac{\lambda_i}{(d_i^{(k)})^{p-1}} \right), p=3,4, \dots, i=1:n, k=0,1,2, \dots \quad (3.17)$$

En 1879 Cayley [4] señaló que el estudio de la convergencia de la iteración con la forma de (3.17) era muy difícil para $p > 2$. De hecho, el conjunto de valores iniciales para que la iteración (3.17) converja a una raíz específica es un complicado conjunto llamado conjunto de Julia de la iteración [21]. Se analizará la convergencia a partir de la expresión (3.18) que se obtiene manipulando algebraicamente (3.17).

$$G(d_i^{(k)}, p, \lambda_i) = \frac{(d_i^{(k)})^p - \lambda_i}{(d_i^{(k)})^p + \lambda_i}, \quad (3.18)$$

se analizó (3.18) para $p \geq 3$, $k = 0, 1, 2, \dots$, y $d_i^{(0)} = 1$ y se observó el mismo comportamiento para cada uno de los desarrollos algebraicos. A continuación, se reproduce el comportamiento de (3.18) sólo para $p = 3$.

Sustituyendo en (3.18) el valor inicial $d_i^{(0)} = 1$ se obtiene como resultado

$$G(d_i^{(k)}, p, \lambda_i) = (1 - \lambda_i) \frac{1}{1 + \lambda_i}. \quad (3.19)$$

Para calcular $d_i^{(1)}$ se sustituye en (3.17) $d_i^{(0)} = 1$ y se obtiene $d_i^{(1)} = \frac{1}{3}(2 + \lambda_i)$, sustituyendo $d_i^{(1)}$ en (3.18) se obtiene

$$G(d_i^{(k)}, p, \lambda_i) = (1 - \lambda_i)^2 \frac{(8 + \lambda_i)}{8 + 39\lambda_i + 6\lambda_i^2 + \lambda_i^3}. \quad (3.20)$$

Para calcular $d_i^{(2)}$ se sustituye en (3.17) $d_i^{(1)} = \frac{1}{3}(2 + \lambda_i)$ y se obtiene

$$d_i^{(2)} = \frac{1}{3} \left(\frac{9\lambda_i}{(2+\lambda_i)^2} + \frac{2}{3}(2 + \lambda_i) \right), \text{ sustituyendo } d_i^{(2)} \text{ en (3.18) se obtiene}$$

$$G(d_i^{(k)}, p, \lambda_i) = (1 - \lambda_i)^4 \frac{4096 + 8896\lambda_i + 5104\lambda_i^2 + 1403\lambda_i^3 + 176\lambda_i^4 + 8\lambda_i^5}{4096 + 85824\lambda_i + 274032\lambda_i^2 + 367899\lambda_i^3 + 226980\lambda_i^4 + 83682\lambda_i^5 + 18012\lambda_i^6 + 2205\lambda_i^7 + 144\lambda_i^8 + 8\lambda_i^9} \quad (3.21)$$

del comportamiento observado de las ecuaciones de la (3.19) a la (3.21) para $p = 3$ y $k = 0, 1, 2$ se deduce que $(1 - \lambda_i)^2$ aparece como factor del cociente de dos números complejos que son función de λ_i . El número complejo que aparece como numerador está representado por el polinomio $S(\lambda_i)$ el cual es de grado $(p^k - 2^k)$ con todos sus términos enteros positivos. Y el

número complejo que aparece como denominador está representado por el polinomio $T(\lambda_i)$ que es de grado p^k con todos sus términos enteros positivos.

Es obvio que el grado de $S(\lambda_i)$ siempre será menor que el grado de $T(\lambda_i)$ para toda p y k , además todas las potencias de $S(\lambda_i)$ aparecen también en $T(\lambda_i)$ observándose que los coeficientes de las potencias de λ_i pertenecientes a $S(\lambda_i)$ siempre son menores que los coeficientes de las mismas potencias de λ_i pertenecientes a $T(\lambda_i)$. Por lo tanto, el modulo del cociente $S(\lambda_i)/T(\lambda_i)$ siempre será menor que la unidad para toda $p = 3, 4, \dots$, $i = 1: n$, $k = 0, 1, 2, \dots$

Por lo tanto, el comportamiento general de (3.18) es

$$G(d_i^{(k)}, p, \lambda_i) = (1 - \lambda_i)^{2k} \frac{S(\lambda_i)}{T(\lambda_i)}, \quad p = 3, 4, \dots, \quad i = 1: n, \quad k = 0, 1, 2, \dots$$

Como An es una matriz normalizada entonces

$$\{\lambda_i \in \mathcal{C}, 0 < |\lambda_i| \leq 1, \operatorname{Re}(\lambda_i) > 0, \quad i = 1: n\}$$

por lo tanto, se tiene la certeza de que

$$0 \leq |1 - \lambda_i| < 1, \quad \text{para toda } i \quad (3.22)$$

además

$$\left| \frac{S(\lambda_i)}{T(\lambda_i)} \right| < 1, \quad \text{para toda } i, p, k, \lambda_i \quad (3.23)$$

Por lo tanto

$$\left| \frac{(d_i^{(k)})^p - \lambda_i}{(d_i^{(k)})^p + \lambda_i} \right| = \left| (1 - \lambda_i)^{2k} \frac{S(\lambda_i)}{T(\lambda_i)} \right|, \quad p = 3, 4, \dots, \quad i = 1: n, \quad k = 0, 1, 2, \dots \quad (3.24)$$

Considerando (3.22) y (3.23) y tomando límites en ambos lados de (3.24)

$$\frac{\lim_{k \rightarrow \infty} \left| (d_i^{(k)})^p - \lambda_i \right|}{\lim_{k \rightarrow \infty} \left| (d_i^{(k)})^p + \lambda_i \right|} = \lim_{k \rightarrow \infty} |(1 - \lambda_i)^{2k}| \lim_{k \rightarrow \infty} \left| \frac{S(\lambda_i)}{T(\lambda_i)} \right| = 0, \quad \lim_{k \rightarrow \infty} \left| (d_i^{(k)})^p + \lambda_i \right| \neq 0 \quad (3.25)$$

simplificando (3.25)

$$\lim_{k \rightarrow \infty} \left| (d_i^{(k)})^p - \lambda_i \right| = 0. \quad (3.26)$$

La ecuación (3.26) es válida si y sólo si $(d_i^{(k)})^p \rightarrow \lambda_i$ cuando $k \rightarrow \infty$, es decir

$$d_i^{(k)} = (\lambda_i)^{\frac{1}{p}} + \epsilon_i^{(k)}, \quad \epsilon_i^{(k)} \rightarrow 0 \text{ cuando } k \rightarrow \infty. \quad (3.27)$$

Utilizando la notación de (3.14) y (3.15) y de acuerdo con (3.27)

$$D_k = \tau^{\frac{1}{p}}, \text{ cuando } k \rightarrow \infty$$

$$Y_k = Q D_k Q^{-1} = Q (\tau)^{\frac{1}{p}} Q^{-1} = (An)^{\frac{1}{p}}, \text{ cuando } k \rightarrow \infty$$

$$Y_k = Y = (\|A\|_F)^{-\frac{1}{p}} A^{\frac{1}{p}}, \text{ cuando } k \rightarrow \infty$$

finalmente

$$A^{\frac{1}{p}} = Y (\|A\|_F)^{\frac{1}{p}}$$

que es lo que se quería demostrar. Nuevo número ecn = no +13

3.4. Análisis de estabilidad numérica del (MSN) para el cálculo de la p-ésima raíz de una matriz no singular.

Se sabe que (3.8) converge cuadráticamente si el valor de inicio X_0 de la iteración está lo suficientemente cerca de una solución y, bajo supuestos razonables, cualquier error que surja debido a la aritmética de punto flotante se amortigua en las iteraciones subsecuentes [34]. Pero, ¿cómo afecta a las perturbaciones y al comportamiento del (MSN) la consideración de que $X_k H_k = H_k X_k$?. Examinaremos la iteración (3.12) bajo el supuesto que la iteración converge

en aritmética exacta (por ejemplo, a la p -ésima raíz de una matriz \mathbf{An} positiva definida) siendo \mathbf{An} no singular y diagonalizable. Se considera que $\widehat{\mathbf{Y}}_k$ es igual a la k -ésima iteración perturbada con la matriz de perturbaciones Δ_k .

$$\widehat{\mathbf{Y}}_k = \mathbf{Y}_k + \Delta_k$$

y que $\widehat{\mathbf{Y}}_{k+1}$ se calcula utilizando (3.12) en función de $\widehat{\mathbf{Y}}_k$

$$\begin{aligned} \widehat{\mathbf{Y}}_{k+1} &= \frac{1}{p} \left((p-1)\widehat{\mathbf{Y}}_k + \widehat{\mathbf{Y}}_k^{1-p} \mathbf{An} \right) \\ &= \frac{1}{p} \left((p-1)(\mathbf{Y}_k + \Delta_k) + (\mathbf{Y}_k + \Delta_k)^{1-p} \mathbf{An} \right) \end{aligned} \quad (3.28)$$

$$\Delta_{k+1} = \widehat{\mathbf{Y}}_{k+1} - \mathbf{Y}_{k+1}$$

Desarrollando $(\mathbf{Y}_k + \Delta_k)^{1-p}$ en series de Taylor a primer orden para $p = 1, 2, 3, \dots$ se llega a la siguiente expresión:

$$(\mathbf{Y}_k + \Delta_k)^{1-p} = \mathbf{Y}_k^{1-p} - \sum_{r=1}^{p-1} \mathbf{Y}_k^{r-p} \Delta_k \mathbf{Y}_k^{-r} + O(\|\Delta_k\|^2) \quad (3.29)$$

finalmente (3.28) toma la siguiente forma

$$\widehat{\mathbf{Y}}_{k+1} = \frac{1}{p} \left((p-1)(\mathbf{Y}_k + \Delta_k) + \left[\mathbf{Y}_k^{1-p} - \sum_{r=1}^{p-1} \mathbf{Y}_k^{r-p} \Delta_k \mathbf{Y}_k^{-r} \right] \mathbf{An} \right) + O(\|\Delta_k\|^2). \quad (3.30)$$

Restando (3.12) de (3.30) resulta

$$\Delta_{k+1} = \frac{1}{p} \left((p-1)\Delta_k - \left[\sum_{r=1}^{p-1} \mathbf{Y}_k^{r-p} \Delta_k \mathbf{Y}_k^{-r} \right] \mathbf{An} \right) + O(\|\Delta_k\|^2). \quad (3.31)$$

Utilizando la notación de (3.14) y (3.15) para diagonalizar (4.31)

$$\hat{\Delta}_k = \mathbf{Q}^{-1} \Delta_k \mathbf{Q}$$

$$\hat{\Delta}_{k+1} = \frac{1}{p} \left((p-1) \hat{\Delta}_k - \left[\sum_{r=1}^{p-1} \mathbf{D}_k^{r-p} \hat{\Delta}_k \mathbf{D}_k^{-r} \right] \boldsymbol{\tau} \right) + O(\|\Delta_k\|^2). \quad (3.32)$$

De (3.14) y (3.15) tenemos que $\mathbf{D}_k = \text{diag}(d_i^{(k)})$, $\boldsymbol{\tau} = \text{diag}(\lambda_i)$ y $\hat{\Delta}_k = \hat{\delta}_{ij}^{(k)}$ entonces (3.32) puede expresarse como

$$\begin{aligned} \hat{\delta}_{ij}^{(k+1)} &= \frac{1}{p} \left((p-1) \hat{\delta}_{ij}^{(k)} - \left[\sum_{r=1}^{p-1} \frac{\hat{\delta}_{ij}^{(k)}}{(d_i^{(k)})^{p-r} (d_j^{(k)})^r} \right] \lambda_i \right) + O(\|\Delta_k\|^2) \\ &= \xi_{ij}^{(k)} \hat{\delta}_{ij}^{(k)} + O(\|\Delta_k\|^2), \quad i, j = 1:n, \end{aligned} \quad (3.33)$$

donde

$$\xi_{ij}^{(k)} = \frac{1}{p} \left((p-1) - \left[\sum_{r=1}^{p-1} \frac{1}{(d_i^{(k)})^{p-r} (d_j^{(k)})^r} \right] \lambda_i \right).$$

Pero como se demostró anteriormente en la sección 3.3; $\mathbf{D}_k \rightarrow \tau^{\frac{1}{p}}$ cuando $k \rightarrow \infty$ entonces podemos escribir

$$d_i^{(k)} = \lambda_i^{\frac{1}{p}} + \epsilon_i^{(k)}, \quad \epsilon_i^{(k)} \rightarrow 0 \text{ cuando } k \rightarrow \infty$$

entonces

$$\xi_{ij}^{(k)} = \frac{1}{p} \left((p-1) - \left[\sum_{r=1}^{p-1} \frac{1}{(\lambda_i)^{\frac{p-r}{p}} (\lambda_j)^{\frac{r}{p}}} \right] \lambda_j \right) + O(|\epsilon^{(k)}|)$$

$$\xi_{ij}^{(k)} = \frac{1}{p} \left((p-1) - \sum_{r=1}^{p-1} \left(\frac{\lambda_i}{\lambda_j} \right)^{r/p} \right) + O|\epsilon^{(k)}|, \quad \epsilon^{(k)} = \max_i |\epsilon^{(k)}|.$$

Para la estabilidad numérica de la iteración se requiere que el modulo del factor de amplificación del error $\xi_{ij}^{(k)}$ no exceda de 1. Por lo tanto, es necesario que

$$\frac{1}{p} \left| (p-1) - \sum_{r=1}^{p-1} \left(\frac{\lambda_i}{\lambda_j} \right)^{r/p} \right| \leq 1, \quad i, j = 1:n.$$

Esta es una restricción muy severa para la matriz \mathbf{An} y hace que la iteración del (MSN) sea de poca utilidad práctica para el cálculo de p-ésima raíz principal de una matriz no singular. Por ejemplo, si \mathbf{An} es Hermitiana positiva definida, entonces para el caso de la raíz cuadrada ($p = 2$) esto es equivalente a que el número de condición de \mathbf{An} sea menor o igual a nueve $k_2(\mathbf{An}) \leq 9$, donde $k_2(\mathbf{An}) = \|\mathbf{An}\|_2 \|\mathbf{An}^{-1}\|_2$. Este resultado fue notado primero por Laasonen [26] y demostrado posteriormente por Higham [10]. Para la raíz cúbica de una matriz \mathbf{An} Hermitiana definida positiva la condición para la estabilidad es

$$k_2(\mathbf{An})^{\frac{1}{3}} + k_2(\mathbf{An})^{\frac{2}{3}} \leq 5$$

resolviendo esta ecuación cuadrática, nos encontramos con que la condición para la estabilidad es

$$k_2(\mathbf{An}) \leq 5.74$$

para la raíz quinta de una matriz positiva definida se necesita que

$$k_2(\mathbf{An})^{\frac{1}{5}} + k_2(\mathbf{An})^{\frac{2}{5}} + k_2(\mathbf{An})^{\frac{3}{5}} + k_2(\mathbf{An})^{\frac{4}{5}} \leq 9$$

resolviendo la ecuación anterior, la condición para la estabilidad es

$$k_2(\mathbf{An}) \leq 4.52.$$

Es evidente que mientras busquemos las raíces de orden superior, la condición para la estabilidad numérica será aún más restrictiva.

El análisis muestra que dependiendo de los valores propios de An , una pequeña perturbación Δ_k en Y_k puede causar perturbaciones que aumentan la norma en las iteraciones, causando que la secuencia \hat{Y}_k diverja de la verdadera secuencia Y_k . La pérdida de estabilidad del (MSN) se debe a la propagación inestable de los errores de redondeo, lo que resulta en una pérdida de conmutatividad en las iteraciones. Por lo tanto, al simplificar el (MN) (3.8), para obtener las iteraciones (3.11), (3.12) y (3.13), por lo general se pierde la estabilidad numérica.

CAPÍTULO 4 MÉTODO SIMPLIFICADO DE NEWTON CON FACTORIZACIONES SUCESIVAS PARA EL CÁLCULO DE LA P-ÉSIMA RAÍZ DE UNA MATRIZ NO SINGULAR (MSNFSPR)

4.1 Deducción del (MSNFSPR)

Analizando (3.12) para el caso donde $p = 2$ y considerando que Y_k conmuta con An para toda k [4, teorema 1], [17], [18] la iteración (3.12) puede expresarse como

$$\left. \begin{aligned} & \text{dado } Y_0 = I \\ Y_{k+1} &= \frac{1}{2} (Y_k + \sqrt{An} Y_k^{-1} \sqrt{An}) \end{aligned} \right\}, \quad k = 0, 1, 2, \dots \quad (4.1)$$

$$A^{\frac{1}{2}} = Y(\|A\|_F)^{\frac{1}{2}}$$

En (4.1) An es factorizada en dos matrices que conmutan con respecto a su producto, por lo tanto, conmutan con Y_k . Utilizando el mismo procedimiento tratado en [10, sección 3] puede probarse que (4.1) es convergente y numéricamente estable.

Para el caso general del cálculo de la p -ésima raíz de An se utilizará un argumento similar al caso $p = 2$, que es el siguiente: en cada k -ésima iteración factorizar An como el producto de p matrices de las cuales $(p - 1)$ serán idénticas. Una de las $(p - 1)$ matrices estará representada por la matriz C_k y la matriz desigual estará representada por B_k , de tal manera que en cada k -ésima iteración se cumpla que:

$$An = B_k C_k^{p-1} = C_k^{p-1} B_k \quad (4.2)$$

en otras palabras B_k y C_k^{p-1} conmutan, con esto se garantiza que Y_k^{1-p} conmute con B_k y C_k^{p-1} , lo cual implica que

$$Y_{k+1} = \frac{1}{p} \left((p-1)Y_k + Y_k^{1-p} B_k C_k^{p-1} \right) \quad (4.3)$$

considerando la conmutatividad

$$Y_{k+1} = \frac{1}{p} \left((p-1)Y_k + B_k Y_k^{1-p} C_k^{p-1} \right) \quad (4.4)$$

manipulando algebraicamente (4.4)

$$Y_{k+1} = \frac{1}{p} \left((p-1)Y_k + B_k (Y_k^{-1} C_k)^{p-1} \right) \quad (4.5)$$

Ahora el problema se reduce a encontrar en cada k -ésima iteración las matrices B_k y C_k que cumplan con (4.2), lo anterior se resuelve fácilmente de la siguiente manera: Dado $B_0 = An$, $C_0 = Y_0 = I$ y habiéndose calculado Y_{k+1} a partir de (4.5), se hace $C_k = Y_k$, por consiguiente $B_{k+1} = pY_{k+1} - (p-1)Y_k$, luego se hace $C_{k+1} = C_k$. Con lo anterior se garantiza que en cada k -ésima iteración se cumple con (4.2). De esta manera se obtiene el Método Simplificado de Newton con Factorizaciones Sucesivas para el Cálculo de la p -ésima raíz de la matriz A no singular (MSNFSPR). El algoritmo tiene un costo computacional del orden de $2n^3(3 + \vartheta \log(p))$, $1 \leq \vartheta \leq 2$ operaciones aritméticas por iteración, y es el siguiente.

$$\left. \begin{array}{l} B_0 = An, C_0 = I, Y_0 = I \\ \text{for } k = 0 : w \\ \quad Y_{k+1} = \frac{1}{p} \left((p-1)Y_k + B_k (Y_k^{-1} C_k)^{p-1} \right) \\ \quad C_k = Y_k \\ \quad B_{k+1} = pY_{k+1} - (p-1)Y_k \\ \quad C_{k+1} = C_k \\ \text{end} \\ Y = Y_{k+1} = Y_k \\ A^{\frac{1}{p}} = Y (\|A\|_F)^{\frac{1}{p}} \end{array} \right\} \quad (4.6)$$

4.2 Análisis de convergencia del (MSNFSPR)

Teorema 2. Sea $B_k C_k^{p-1} = C_k^{p-1} B_k = An \in C^{n \times n}$, $k=0, 1, 2, \dots$, una matriz no singular y diagonalizable con valores propios λ_i , $\{\lambda_i \in C, 1 \geq |\lambda_i| > 0, \text{Re}(\lambda_i) > 0, i = 1:n\}$, $B_0 = An$, $C_0 = Y_0 = I$, si $\|Y - Y_0\|$ es suficientemente pequeño y Y_k^{-1} es no singular, entonces el

proceso iterativo (4.6) converge a $Y = Y_{k+1} = B_k = C_k = Y_k = (\|A\|_F)^{-\frac{1}{p}} A^{\frac{1}{p}}$ cuando $k \rightarrow \infty$, $A^{\frac{1}{p}} = Y(\|A\|_F)^{\frac{1}{p}}$ es la p -ésima raíz principal de A .

Demostración:

Por las características de la matriz An , existe una matriz Z no singular tal que [24]

$$\tau = Z^{-1} A n Z = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (4.7)$$

$$D_k = Z^{-1} Y_k Z = \text{diag}(d_1^{(k)}, \dots, d_n^{(k)}) \quad (4.8)$$

$$F_k = Z^{-1} B_k Z = \text{diag}(f_1^{(k)}, \dots, f_n^{(k)}) \quad (4.9)$$

$$G_k = Z^{-1} C_k Z = \text{diag}(g_1^{(k)}, \dots, g_n^{(k)}) \quad (4.10)$$

$f_1^{(k)}, \dots, f_n^{(k)}$ y $g_1^{(k)}, \dots, g_n^{(k)}$ son los valores propios de B_k y C_k respectivamente. Además de (4.2) se tiene

$$\tau = F_k G_k^{p-1} = G_k^{p-1} F_k, \quad \text{para toda } k \quad (4.11)$$

desacoplando (4.11)

$$\begin{aligned} \left(\frac{(g_i^{(k)})^p}{g_i^{(k)}} \right) f_i^{(k)} &= \lambda_i \\ \left(\frac{(g_i^{(k)})^p}{g_i^{(k)}} \right) &= \frac{\lambda_i}{f_i^{(k)}} \end{aligned} \quad (4.12)$$

utilizando desde (4.7) hasta (4.11) se diagonaliza (4.6)

$$D_{k+1} = \frac{1}{p} \left((p-1) D_k + F_k (D_k^{-1} G_k)^{p-1} \right) \quad (4.13)$$

considerando la propiedad de conmutatividad y manipulando algebraicamente

$$D_{k+1} = \frac{1}{p} \left((p-1) D_k + D_k^{1-p} F_k G_k^{p-1} \right) \quad (4.14)$$

de(4.11)

$$D_{k+1} = \frac{1}{p} \left((p-1)D_k + D_k^{1-p} \tau \right) \quad (4.15)$$

la iteración (4.15) tiene la misma forma que (3.16), por lo tanto, se concluye que (4.6) es convergente, lo cual implica que:

$$Y_{k+1} = Y_k = An^{\frac{1}{p}} \text{ cuando } k \rightarrow \infty$$

Además, tenemos que

$$C_{k+1} = C_k = Y_k = An^{\frac{1}{p}} \text{ cuando } k \rightarrow \infty$$

$$B_{k+1} = B_k = pY_{k+1} - (p-1)Y_k = pAn^{\frac{1}{p}} - (p-1)An^{\frac{1}{p}} = An^{\frac{1}{p}} \text{ cuando } k \rightarrow \infty$$

Finalmente

$$B_k = C_k = Y_k = Y = An^{\frac{1}{p}} = (\|A\|_F)^{-\frac{1}{p}} A^{\frac{1}{p}} \text{ cuando } k \rightarrow \infty \quad (4.16)$$

de(4.16)se diagonaliza B_k , C_k , y Y_k

$$f_i^{(k)} = \lambda_i^{\frac{1}{p}} + \epsilon_i^{(k)}, \quad g_i^{(k)} = \lambda_i^{\frac{1}{p}} + \epsilon_i^{(k)}, \quad d_i^{(k)} = \lambda_i^{\frac{1}{p}} + \epsilon_i^{(k)} \quad (4.17)$$

donde $\epsilon_i^{(k)} \rightarrow 0$ cuando $k \rightarrow \infty$

finalmente

$$A^{\frac{1}{p}} = (\|A\|_F)^{\frac{1}{p}} Y$$

4.3 Análisis de estabilidad numérica del (MSNFSPR)

Se analiza la estabilidad numérica de (4.6) bajo el supuesto que converge en aritmética exacta siendo An diagonalizable y no singular (por ejemplo, el cálculo de la p -ésima raíz de una

matriz positiva definida). Consideremos que \hat{Y}_k es la k -ésima iteración Y_k perturbada con la matriz de perturbaciones Δ_k .

$$\hat{Y}_k = Y_k + \Delta_k$$

\hat{Y}_{k+1} se calcula utilizando (5.6) en función de \hat{Y}_k

$$\begin{aligned} \hat{Y}_{k+1} &= \frac{1}{p} \left((p-1) \hat{Y}_k + B_k (\hat{Y}_k^{-1} C_k)^{p-1} \right) \\ &= \frac{1}{p} \left((p-1)(Y_k + \Delta_k) + B_k ((Y_k + \Delta_k)^{-1} C_k)^{p-1} \right) \end{aligned} \quad (4.18)$$

considerando la conmutatividad entre Y_k , C_k y B_k

$$\hat{Y}_{k+1} = \frac{1}{p} \left((p-1)(Y_k + \Delta_k) + B_k (Y_k + \Delta_k)^{1-p} C_k^{p-1} \right) \quad (4.19)$$

$$\Delta_{k+1} = \hat{Y}_{k+1} - Y_{k+1}$$

desarrollando $(Y_k + \Delta_k)^{1-p}$ en series de Taylor a primer orden para $p = 1, 2, 3, \dots$ se llega a la siguiente expresión:

$$(Y_k + \Delta_k)^{1-p} = Y_k^{1-p} - \sum_{r=1}^{p-1} Y_k^{r-p} \Delta_k Y_k^{-r}. \quad (4.20)$$

Finalmente (4.19) toma la siguiente forma

$$\hat{Y}_{k+1} = \frac{1}{p} \left((p-1)(Y_k + \Delta_k) + B_k \left[Y_k^{1-p} - \sum_{r=1}^{p-1} Y_k^{r-p} \Delta_k Y_k^{-r} \right] C_k^{p-1} \right) + O(\|\Delta_k\|^2). \quad (4.21)$$

restando (4.6) de (4.21) resulta

$$\Delta_{k+1} = \frac{1}{p} \left((p-1)\Delta_k - B_k \left[\sum_{r=1}^{p-1} Y_k^{r-p} \Delta_k Y_k^{-r} \right] C_k^{p-1} \right) + O(\|\Delta_k\|^2). \quad (4.22)$$

Utilizando la notación desde (4.7) hasta (4.12) para diagonalizar (4.22)

$$\hat{\Delta}_k = \mathbf{Q}^{-1} \Delta_k \mathbf{Q} = \hat{\delta}_{ij}^{(k)}$$

$$\hat{\Delta}_{k+1} = \frac{1}{p} \left((p-1) \hat{\Delta}_k - \mathbf{F}_k \left[\sum_{r=1}^{p-1} \mathbf{D}_k^{r-p} \hat{\Delta}_k \mathbf{D}_k^{-r} \right] \mathbf{G}_k^{p-1} \right) + O(\|\Delta_k\|^2). \quad (4.23)$$

Desacoplando (4.23)

$$\begin{aligned} \delta_{ij}^{(k+1)} &= \frac{1}{p} \left((p-1) \delta_{ij}^{(k)} - f_i^{(k)} \left[\sum_{r=1}^{p-1} \frac{\delta_{ij}^{(k)}}{(d_i^{(k)})^{p-r} (d_j^{(k)})^r} \right] \frac{(g_j^{(k)})^p}{g_j^{(k)}} \right) + O(\|\Delta_k\|^2) \quad (4.24) \\ &= \xi_{ij}^{(k)} \delta_{ij}^{(k)} + O(\|\Delta_k\|^2), \quad i, j = 1:n, \end{aligned}$$

donde

$$\xi_{ij}^{(k)} = \frac{1}{p} \left((p-1) - f_i^{(k)} \left[\sum_{r=1}^{p-1} \frac{\delta_{ij}^{(k)}}{(d_i^{(k)})^{p-r} (d_j^{(k)})^r} \right] \frac{(g_j^{(k)})^p}{g_j^{(k)}} \right) + O(\|\Delta_k\|^2)$$

de acuerdo con (4.17) resulta

$$\begin{aligned} \xi_{ij}^{(k)} &= \frac{1}{p} \left((p-1) - \lambda_i^{\frac{1}{p}} \left[\sum_{r=1}^{p-1} \frac{1}{(\lambda_i)^{\frac{p-r}{p}} (\lambda_j)^{\frac{r}{p}}} \right] \frac{\lambda_j}{\lambda_j^{\frac{1}{p}}} \right) + O|\epsilon^{(k)}| \\ &= \frac{1}{p} \left((p-1) - \sum_{r=1}^{p-1} \left(\frac{\lambda_i}{\lambda_j} \right)^{1-p+r} \right) + O|\epsilon^{(k)}| \end{aligned}$$

donde $\epsilon^{(k)} = \max_i |\epsilon^{(k)}|$.

Para asegurar la estabilidad numérica de la iteración (4.6) se requiere que el modulo del factor

de amplificación del error $\xi_{ij}^{(k)}$ no exceda de 1. Por lo tanto, es necesario que

$$\frac{1}{p} \left| (p-1) - \sum_{r=1}^{p-1} k_2(A_n)^{1-p+r} \right| \leq 1, \quad i, j = 1:n. \quad (4.25)$$

$$k_2(A_n) = \|A_n\|_2 \|A_n^{-1}\|_2$$

resolviendo y simplificando (4.25) se tiene

$$\left(\sum_{r=0}^{p-2} k_2(A_n)^{-r} \right) \leq 2p-1 \quad (4.26)$$

a partir de experimentos numéricos es fácil notar que cuando $p \rightarrow \infty$, $k_2(A_n) \rightarrow 1$

por ejemplo, para la raíz cubica la condición para la estabilidad es $k_2(A_n)^{-1} + 1 \leq 5$

resolviendo tenemos que $k_2(A_n) \geq 0.25$.

Para la raíz sesentava la condición para la estabilidad es $k_2(A_n) \geq 0.978$

Con lo anterior queda demostrado que (5.6) es numéricamente estable, ya que para una matriz

A el mínimo valor que se puede presentar para $k_2(A_n)$ es la unidad.

CAPÍTULO 5 EXPERIMENTOS NUMÉRICOS

Por medio de experimentos numéricos con matrices, se comparó la convergencia y estabilidad numérica del (MSNFSPR) con respecto a los siguientes algoritmos: (MN), (MSN), [18, iter. (3.6) (HWA)] y [19, iter. (5.10)], con matrices utilizadas en [18], [33] y [8]. Los resultados se presentan en las tablas 5.1, 5.2, 5.3, 5.4 y en los gráficos 5.1, 5.2, 5.3 y 5.4. Para cada experimento realizado se reporta: El tamaño n y el número de condición $k_2(A)$ de la matriz, el método utilizado, el número de iteraciones y el error relativo residual (err) que se calculará utilizando la norma de Frobenius.

$$err = \frac{\|Y^p - A\|_F}{\|A\|_F}, \quad \text{o bien} \quad err = \frac{\|X^p - A\|_F}{\|A\|_F}$$

Los cálculos se llevaron a cabo utilizando MATLAB Versión 8.3.0.532(R2014a) en una Laptop con procesador Intel(R) Core(TM) i5-4210U CPU @ 1.7 GHz 2.4 GHz y 8 GB de Ram. El costo computacional de cada iteración se calculó utilizando la información de [14, p. 336], y se reporta en la tabla 5.5.

$$\left. \begin{aligned} X_0 &= I, \quad N_0 = \frac{A}{\|A\|_F} \\ X_{k+1} &= X_k \left(\frac{(p-1)I + N_k}{p} \right), \\ N_{k+1} &= \left(\frac{-1)I + N_k}{p} \right)^{-p} N_k. \\ A^{\frac{1}{p}} &= X(\|A\|_F)^{\frac{1}{p}} \end{aligned} \right\} [18, \text{iteración (3.6), HWA}]$$

$$\left. \begin{aligned} X_0 &= I, \quad N_0 = \frac{A}{\|A\|_F} \\ X_{k+1} &= X_k \left((p+1)I + (p-1)N_k \right)^{-1} \left((p-1)I + (p+1)N_k \right), \\ N_{k+1} &= N_k \left(\left((p+1)I + (p-1)N_k \right)^{-1} \left((p-1)I + (p+1)N_k \right) \right)^{-p} \\ A^{\frac{1}{p}} &= X(\|A\|_F)^{\frac{1}{p}} \end{aligned} \right\} [19, \text{iteración (5.10)}]$$

Experimento No. 1. Se calculará la raíz quinta de la matriz de Lehmer simétrica definida positiva de tamaño $n = 2 \times 2$.

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

A está bien condicionada, con $k_2(A) = \|A\|_F \|A^{-1}\|_F = 3.0$ y valores propios (0.5, 1.5).

Tabla 5.1. Comportamiento de la convergencia

Método	Iterac.	Residual
(NM), (3.8)	6	2.46E-016
(MSN), (3.12)	6	2.56E-016
(MSNFSPR), (4.6)	6	3.85E-016
(HWA)	6	8.51E-016
[19, Iteración (5.10)]	4	3.06E-016

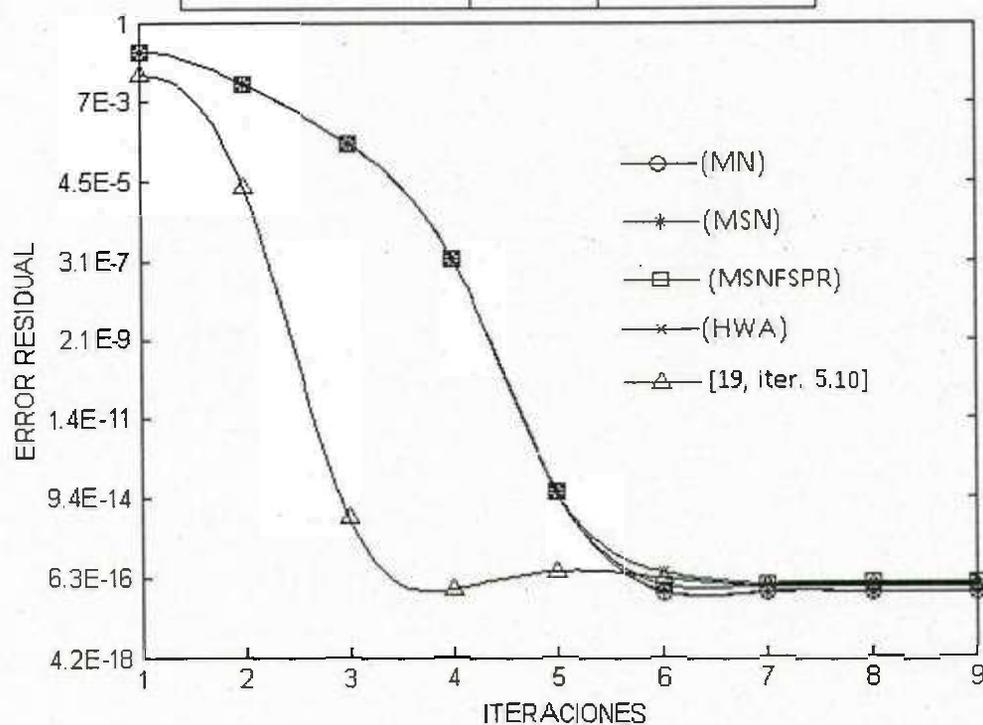


FIG. 5.1. Comportamiento de la convergencia del (MN), (MSN), (MSNFSPR), (HWA) y [19, iteración 5.10]

En la figura 5.1 se observa que todos los métodos proporcionan excelentes resultados, [19, Iteración 5.10] converge en 4 iteraciones (pero tiene un costo computacional mayor que los demás métodos, ver tabla 5.5), mientras que los otros métodos presentan un comportamiento muy similar en cuanto a la manera de converger y lo hacen hasta la sexta iteración. Todos los algoritmos presentan entre ellos una aproximación muy similar. El (MSN) es numéricamente estable debido a que el número de condición de la matriz de Lehmer es menor 4.52.

Experimento No. 2. Para ilustrar la inestabilidad numérica del (MSN) en la vecindad de la solución, se considera la matriz S de tamaño $n = 3 \times 3$, se calculará la raíz quinta de $A = S^5$,

$$S = \begin{bmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{bmatrix}$$

cuyos valores propios son (0.00215549, 1 y 14.49784450), la matriz A está cerca

de ser singular. El número de condición de A es $k_2(A) = 6,726$

Tabla 5.2. Comportamiento de la convergencia

Método	Iterac.	Residual
(NM), (3.8)	12	4.05E-016
(MSN), (3.12)	09	1.17E-005
(MSNFSPR), (4.6)	12	1.43E-015
(HWA)	12	1.39E-015
[19, Iteración (5.10)]	07	2.12E-015

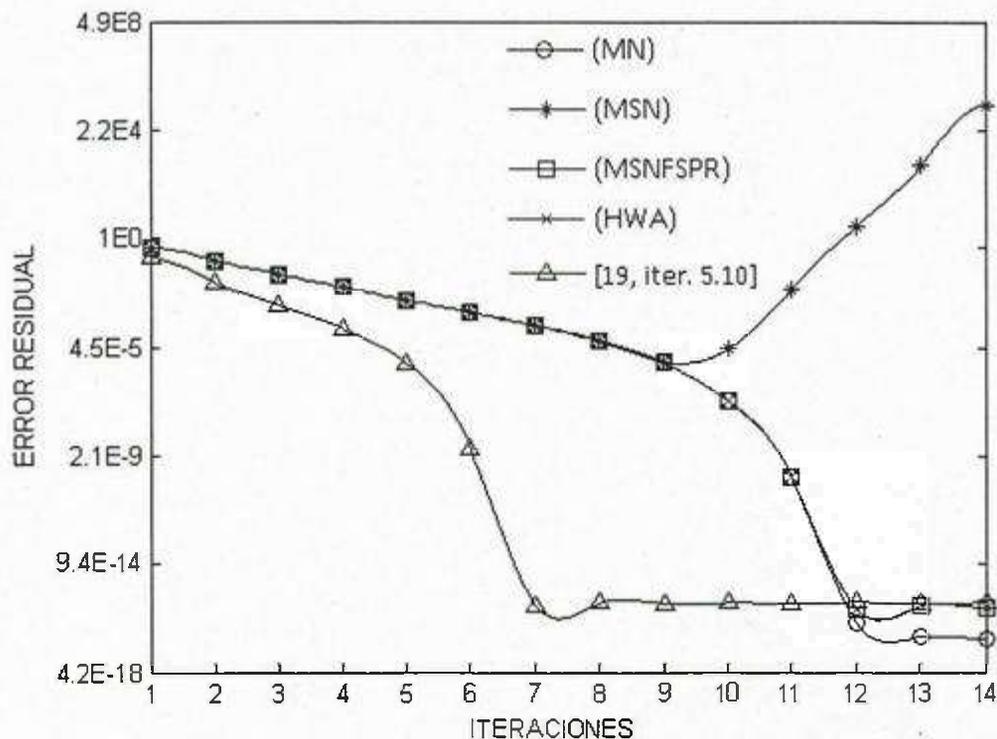


FIG. 5.2. Comportamiento de la convergencia del (MN), (MSN), (MSNFSPR), (HWA) y [19, iteración 5.10]

En la figura 5.2 se observa que el (MSN) diverge a partir de la novena iteración este comportamiento es debido a que el número de condición de la matriz es mucho mayor que 4.52. También puede observarse que los algoritmos restantes proporcionan excelentes resultados. Se observa que, al inicio, aparentemente los algoritmos tienen una convergencia del tipo lineal, pero esto se debe a la escala logarítmica utilizada en el gráfico, en realidad todas las iteraciones tienen una convergencia del tipo cuadrática. [19, Iteración 5.10] converge en 7 iteraciones, mientras que (MN), (MSNFSPR) y (HWA) lo hacen en 12 iteraciones. El grado de aproximación que presentan los algoritmos que son numéricamente estables es bastante similar.

Experimento No.3. Se considera la matriz A con elementos complejos, de tamaño $n = 3 \times 3$ con número de condición $k_2(A) = 32.4602$. Los valores propios de A son $(19.9990+15.3502i, 2.9634+4.7164i$ y $0.0376+0.9334i)$, se calculará la raíz veinteva de A .

$$A = \begin{bmatrix} 1 + 15i & 1 + i & 1 - i \\ 2 + 5i & 2 + 5i & 1 - i \\ 1 - i & -2 + 3i & 1 + i \end{bmatrix}$$

Tabla 5.3. Comportamiento de la convergencia

Método	Itera c.	Residual
(NM), (3.8)	9	2.29E-015
(MSN), (3.12)	9	5.63E-011
(MSNFSPR), (4.6)	9	1.02E-014
(HWA)	9	1.00E-014
[19, Iteración (5.10)]	6	7.51E-015

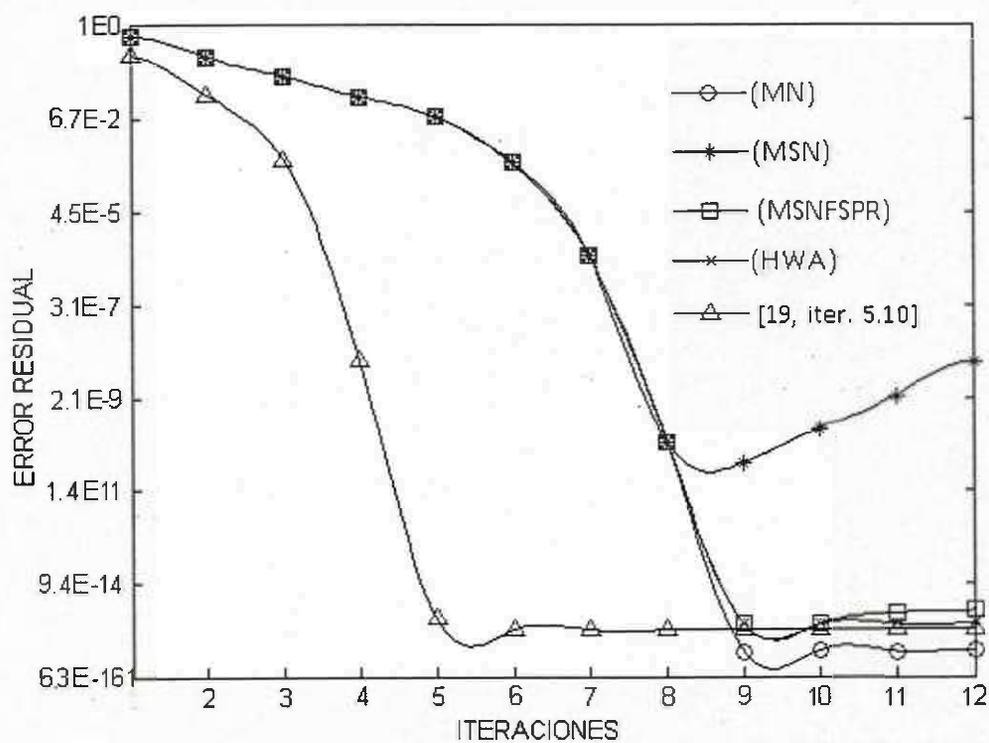


FIG. 5.3. Comportamiento de la convergencia del (MN), (MSN), (MSNFSPR), (HWA) y [19, iteración 5.10]

En la figura 5.3 se observa que el (MSN) diverge a partir de la octava iteración. [19, Iteración 5.10] converge en 6 iteraciones. El comportamiento de la convergencia de los demás algoritmos

es muy similar y convergen en 9 iteraciones. Todos los algoritmos que son numéricamente estables proporcionan excelentes resultados.

Experimento No.4. Se considera la matriz $A = S^{15}$ de tamaño $n = 3 \times 3$ con número de condición $k_2(A) = 1.56E + 10$, este valor indica un alto grado de mal condicionamiento. Los valores propios de A son [14348907, 32768 y 1], se calculará la raíz quinceava de A .

$$S = \begin{bmatrix} -1 & -2 & 2 \\ -4 & -6 & 6 \\ -4 & -16 & 13 \end{bmatrix}$$

Tabla 5.4. Comportamiento de la convergencia

Método	Iterac.	Residual
(NM), (3.8)	23	3.14E-014
(MSN), (3.12)	12	9.79E-005
(MSNFSPR), (4.6)	25	3.21E-013
(HWA)	23	1.67E-013
[19, Iteración (5.10)]	12	3.50E-012

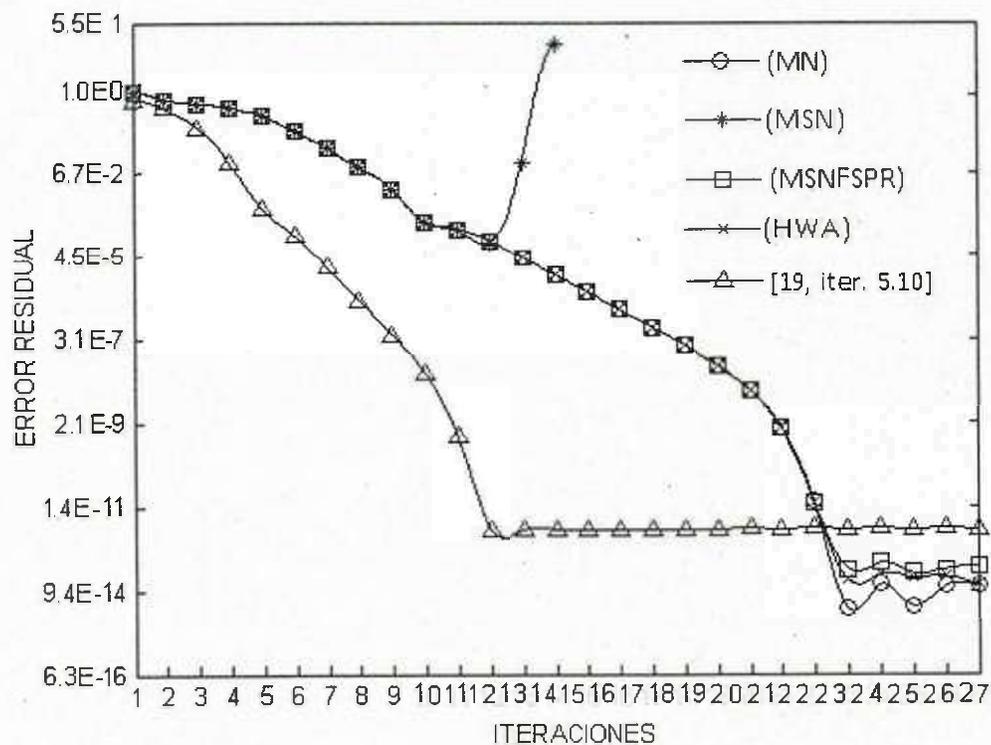


FIG. 5.4. Comportamiento de la convergencia del (MN), (MSN), (MSNFSPR), (HWA) y [19, iteración 5.10]

En la figura 5.4 se observa que el (MSN) diverge a partir de la doceava iteración. [19, Iteración 5.10] converge en 12 iteraciones. El comportamiento de la convergencia de los algoritmos numéricamente estables es muy similar y convergen en un promedio de 24 iteraciones. Los algoritmos que presentan una mejor aproximación son el (MN) y (MSNFSPR). En general los algoritmos que son numéricamente estables proporcionan excelentes resultados.

Tabla 5.5. Costo computacional por iteración para el cálculo de la de la p -ésima raíz de una matriz no singular.

Iteración	No. de operaciones aritméticas por iteración	
(NM), (3.8)	$2n^6 \vartheta \log_2 p,$	$1 \leq \vartheta \leq 2$
(MSN), (3.12)	$2n^3(2 + \vartheta \log(p)),$	$1 \leq \vartheta \leq 2$
(MSNFSPR), (4.6)	$2n^3(3 + \vartheta \log(p)),$	$1 \leq \vartheta \leq 2$
[18, iter. (3.6)]	$2n^3(3 + \vartheta \log(p)),$	$1 \leq \vartheta \leq 2$
[19, iter. (5.10)]	$2n^3(5 + \vartheta \log(p)),$	$1 \leq \vartheta \leq 2$

CONCLUSIONES

En este trabajo de tesis se ha presentado una modificación novedosa al Método Simplificado de Newton que elimina su bien conocida inestabilidad numérica para el cálculo de la p -ésima raíz principal de una matriz no singular. Al llevar a cabo el análisis de convergencia para el (MSN) y el (MSNFSPR), este se llevó a cabo utilizando una metodología muy diferente a las utilizadas en la literatura, pero más sencilla, llegándose al mismo resultado, *“que bajo condiciones especiales que son fáciles de cumplir el (MSN) y el (MSNFSPR) son convergentes para el cálculo de la p -ésima raíz de una matriz no singular, pero solamente el (MSNFSPR) es numéricamente estable.”*

Otro punto importante de resaltar es que la forma original del Método simplificado de Newton que es la siguiente:

$$\left. \begin{aligned} \text{dado } Y_0 = I \\ Y_{k+1} = \frac{1}{p} \left((p-1)Y_k + Y_k^{1-p} An \right) \\ A^{\frac{1}{p}} = Y(\|A\|_F)^{\frac{1}{p}} \end{aligned} \right\}, \quad k = 0, 1, 2, \dots,$$

no se modificó de manera drástica, sólo se expresó la matriz An como el producto de dos matrices B_k y C_k^{p-1} que conmutan entre si, y utilizando un poco de álgebra toma finalmente la siguiente forma:

$$\left. \begin{aligned} B_0 = An, \quad C_0 = I, \quad Y_0 = I \\ \text{for } k = 0 : w \\ Y_{k+1} = \frac{1}{p} \left((p-1)Y_k + B_k (Y_k^{-1} C_k)^{p-1} \right) \\ C_k = Y_k \\ B_{k+1} = pY_{k+1} - (p-1)Y_k \\ C_{k+1} = C_k \\ \text{end} \\ Y = Y_{k+1} = Y_k \\ A^{\frac{1}{p}} = Y(\|A\|_F)^{\frac{1}{p}} \end{aligned} \right\}$$

La cual tiene un costo computacional de $2n^3(3 + \vartheta \log(p))$ operaciones aritméticas por iteración.

BIBLIOGRAFÍA

- [1] Ayres, F. (1992). Schaum's outline of matrices. *Matrices/. Serie Schaums*.
- [2] Bini, D. A., Higham, N. J., & Meini, B. (2005). Algorithms for the matrix pth root. *Numerical Algorithms*, 39(4), 349-378.
- [3] Björck, Å., & Hammarling, S. (1983). A Schur method for the square root of a matrix. *Linear algebra and its applications*, 52, 127-140.
- [4] Cayley, P. (1879). Desiderata and suggestions: no. 3. The Newton-Fourier imaginary problem. *American Journal of Mathematics*, 2(1), 97-97.
- [5] Demmel, J. W. (1997). *Applied numerical linear algebra*. Siam., Philadelphia, PA.
- [6] Dennis Jr, J. E., & Schnabel, R. B. (1996). *Numerical methods for unconstrained optimization and nonlinear equations (Vol. 16)*. Siam.
- [7] García, R. B., & Álvaro, P. S. (1998). *Algebra lineal y teoría de matrices*. Prentice Hall.
- [8] Guo, C. H. (2010). On Newton's method and Halley's method for the principal pth \square root of a matrix. *Linear algebra and its applications*, 432(8), 1905-1922.
- [9] Higham, N. J. (1986). Computing the polar decomposition-with applications. *SIAM Journal on Scientific and Statistical Computing*, 7(4), 1160-1174.
- [10] Higham, N. J. (1986). Newton's method for the matrix square root. *Mathematics of Computation*, 46(174), 537-549.
- [11] Higham, N. J. (1987). Computing real square roots of a real matrix. *Linear Algebra and its applications*, 88, 405-430.
- [12] Higham, N. J. (1997). Stable iterations for the matrix square root. *Numerical Algorithms*, 15(2), 227-242.

- [13] Higham, N. J. (2002). The matrix computation toolbox for MATLAB (version 1.0). Manchester Centre for Computational Mathematics.
- [14] Higham, N. J. (2008). Functions of matrices: theory and computation. Siam.
- [15] Horn, R. A., & Johnson, C. R. Topics in matrix analysis, 1991. Cambridge University Press, Cambridge, 37, 39.
- [16] Hoskins, W. D., & Walton, D. J. (1979). A faster, more stable method for computing the p th roots of positive definite matrices. *Linear Algebra and its Applications*, 26, 139-163.
- [17] Iannazzo, B. (2003). A note on computing the matrix square root. *Calcolo*, 40(4), 273-283.
- [18] Iannazzo, B. (2006). On the Newton method for the matrix p th root. *SIAM journal on matrix analysis and applications*, 28(2), 503-523.
- [19] Iannazzo, B. (2008). A family of rational iterations and its application to the computation of the matrix p th root. *SIAM Journal on Matrix Analysis and Applications*, 30(4), 1445-1462.
- [20] Israel, R. B., Rosenthal, J. S., & Wei, J. Z. (2001). Finding generators for Markov chains via empirical transition matrices, with applications to credit ratings. *Mathematical finance*, 11(2), 245-265.
- [21] Julia, G. (1918). Mémoire sur l'itération des fonctions rationnelles. *Journal de mathématiques pures et appliquées*, 47-246.
- [22] Kalbfleisch, J. D., & Lawless, J. F. (1985). The analysis of panel data under a Markov assumption. *Journal of the American Statistical Association*, 80(392), 863-871.
- [23] Koç, Ç. K., & Bakkaloğlu, B. (1995). Halley's method for the matrix sector function. *Automatic Control, IEEE Transactions on*, 40(5), 944-949.
- [24] Kolman, B., & Hill, D. R. (2006). Álgebra lineal. Pearson Educación.

- [25] Kreinin, A., & Sidelnikova, M. (2001). Regularization algorithms for transition matrices. *Algo Research Quarterly*, 4(1/2), 23-40.
- [26] Laasonen, P. (1958). On the iterative solution of the matrix equation $AX^2 - I = 0$ ", *Mathematical Tables and Other Aids to Computation*, 109-116.
- [27] Lebedev, L. P., Vorovich, I. I., & Cloud, M. J. (2012). *Functional analysis in mechanics*. Springer Science & Business Media.
- [28] Long, J. H., Hu, X. Y., & Zhang, L. (2008, February). Newton's method with exact line search for the square root of a matrix. In *Journal of Physics: Conference Series* (Vol. 96, No. 1, p. 012034). IOP Publishing.
- [29] Meini, B. (2004). The matrix square root from a new functional perspective: theoretical results and computational issues. *SIAM journal on matrix analysis and applications*, 26(2), 362-376.
- [30] Mexía, A.M., & Aldama, O. G. (2010). Un método simplificado de Newton para calcular la raíz cuadrada de una matriz real simétrica definida positiva. *Métodos numéricos para cálculo y diseño en ingeniería: Revista internacional*, 26(1), 47-53.
- [31] Ortega, J. M. (1990). *Numerical analysis: a second course* (Vol. 3). Siam.
- [32] Singer, B., & Spilerman, S. (1976). The representation of social processes by Markov models. *American Journal of Sociology*, 1-54.
- [33] Smith, M. I. (2003). A Schur algorithm for computing matrix p th roots. *SIAM journal on matrix analysis and applications*, 24(4), 971-989.
- [34] Tisseur, F. (2001). Newton's method in floating point arithmetic and iterative refinement of generalized eigenvalue problems. *SIAM Journal on Matrix Analysis and Applications*, 22(4), 1038-1057.
- [35] Waugh, F. V., & Abel, M. E. (1967). On fractional powers of a matrix. *Journal of the American Statistical Association*, 62(319), 1018-1021.