



“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

---

---

## ALGORITMOS PARA EL CÁLCULO DE LA RAÍZ CUADRADA DE UNA MATRIZ NO SINGULAR

TRABAJO DE TESIS  
QUE PARA OBTENER EL TÍTULO DE:  
INGENIERO INDUSTRIAL Y DE SISTEMAS

PRESENTAN:

LUCERO JOSEFINA CASTILLO GARCIA  
CARLOS ADOLFO HUERTA RIVERA

NAVOJOA, SONORA

NOVIEMBRE DEL 2015

# 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 13 de noviembre de 2015

**Ing. Ma. del Rosario Castrejón Lemus**  
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

**LUCERO JOSEFINA CASTILLO GARCIA**  
**CARLOS ADOLFO HUERTA RIVERA**

el día Lunes 23 de noviembre de 2015 en la Sala de Juntas del edificio B a las 12:00 horas.

**A T E N T A M E N T E**

**MIEMBROS DEL JURADO**

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

**FIRMA**

The block contains three handwritten signatures written over horizontal lines. The top signature is the most legible, appearing to be 'Alfredo Mendoza Mexia'. The middle signature is less legible but appears to be 'Rafael Verdugo Miranda'. The bottom signature is also less legible but appears to be 'Ignacio Yocupicio Villegas'. There is a date '23/11/15' written near the middle signature.

## AGRADECIMIENTOS

**Lucero:**

A **Dios** por tanto amor que me refleja en bendiciones, por permitirme cumplir mi primer gran meta en la vida.

A mi familia, especialmente a mis padres **Miguel Castillo Castro** y **Rosa Idalia García Espinoza** por darme los medios necesarios para llegar a este día, todo el apoyo y cariño, por educarme de manera adecuada para ser lo que hoy soy, sin ustedes no podría haberlo logrado.

A mi gran compañero **Carlos Adolfo Huerta Rivera**, que este sea el primer logro juntos, de muchos que compartiremos en la vida.

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

GRACIAS.

## AGRADECIMIENTOS

### **Carlos:**

A **Dios** por estar vivo y tener salud y permitirme llegar a este día.

A mi abuela **Dolores Jacobo Escalante** por atenderme toda mi vida.

A mis tíos y tías por su apoyo recibido de una u otra forma, motivándome o financiándome en la necesidad.

A mi hermano **Cesar Andrés Huerta Rivera** por ser mi hermano y entretenerme en la existencia.

A mi novia **Lucero** por arrastrarme en cosas necesarias.

Al **C. D. Alfredo Mendoza Mexía** por sus enseñanzas y su apoyo en la elaboración de esta tesis además de ser una inspiración a seguir aprendiendo.

## DEDICATORIAS

### **Lucero:**

A mi familia, ya que sin ellos nada de esto sería posible

A mi madre **Rosa Idalia García Espinoza**, por siempre apoyarme y estar a mi lado en todo momento, todos mis triunfos son también tuyos TE AMO MAMI

A mi padre **Miguel Castillo Castro** por darme los medios necesarios para cumplir esta meta, espero hacerte sentir orgulloso, TE AMO

A mis hermanos **Miguel y Gabriel**, por que pase lo que pase siempre estamos juntos LOS QUIERO MUCHO

### **Carlos:**

A mi abuela **Dolores Jacobo Escalante** por cuidar de mí todos estos años, TE QUIERO MUCHO ABUELA.

## ÍNDICE

<b>Agradecimientos</b>	i
<b>Dedicatorias</b>	iii
<b>Índice</b>	iv
<b>Introducción</b>	1
<b>Antecedentes</b>	2
<b>Definición del problema</b>	4
<b>Justificación</b>	5
<b>Objetivo principal</b>	6
<b>CAPÍTULO 1 ÁLGEBRA MATRICIAL</b>	7
1.1 Matrices	7
1.2 Operaciones básicas con matrices	8
1.2.1 Igualdad de matrices	8
1.2.2 Suma de matrices	8
1.2.3 Multiplicación por un escalar	8
1.2.4 Resta de matrices	8
1.2.5 Multiplicación de matrices	9
1.2.6 Transposición de matrices	9
1.2.7 Traza de una matriz	10
1.3 Determinante de una matriz	10
1.4 Matriz inversa	11
1.5 Matrices cuadradas especiales	13

1.5.1 Matriz diagonal	13
1.5.2 Matriz identidad	13
1.5.3 Matriz escalar	13
1.5.4 Matriz triangular inferior	13
1.5.5 Matriz triangular superior	14
1.5.6 Matriz nula	14
1.5.7 Matriz simétrica	14
1.5.8 Matriz idempotente	14
1.5.9 Matriz ortogonal	14
1.6 Autovalores de una matriz	14
1.7 Diagonalización de matrices	15
<b>CAPÍTULO 2 FORMA CANÓNICA DE JORDAN PARA CARACTERIZAR LA RAÍZ CUADRADA DE UNA MATRIZ</b>	<b>17</b>
2.1 Existencia de raíces cuadradas de matrices	17
2.2 Forma canónica de Jordan	18
2.3 El número de raíces cuadradas de una matriz	22
<b>CAPÍTULO 3 LOS MÉTODOS DE NEWTON PARA EL CASO MATRICIAL</b>	<b>28</b>
3.1 El método de Newton para el cálculo de la raíz cuadrada principal de una matriz singular	28
3.2 Análisis de convergencia del método simplificado de Newton	30
3.3 Análisis de estabilidad numérica del método simplificado de Newton	34

<b>CAPÍTULO 4 ALGUNOS ALGORITMOS PARA EL CÁLCULO DE LA RAÍZ CUADRADA PRINCIPAL DE UNA MATRIZ NO SINGULAR</b>	<b>38</b>
4.1 Un método simplificado de Newton para calcular la raíz cuadrada de una matriz simétrica definida positiva (IASMN)	38
4.1.1 Iteración alternativa simplificada del método de Newton para el cálculo de raíces cuadradas principales de matrices reales simétricas definidas positivas	39
4.1.2 Análisis de convergencia del (IASMN)	40
4.1.3 Análisis de estabilidad numérica del (IASMN)	43
4.1.4 Factor de escala para la (IASMN)	46
4.2 Un método simplificado de Newton con factorizaciones sucesivas para el cálculo de la raíz cuadrada principal de una matriz no singular (MSNFSRC)	47
4.2.1 El método simplificado de Newton	47
4.2.2 El método simplificado de Newton con factorizaciones sucesivas para el cálculo de la raíz cuadrada de una matriz	48
4.2.3 Análisis de convergencia del (MSNFSRC)	49
4.2.4 Análisis de estabilidad numérica del (MSNFSRC)	50
4.2.5 Factor de escala para el (MSNFSRC)	52
4.3 Un método simplificado de Newton acoplado para el cálculo de la raíz cuadrada principal de una matriz no singular (MSNARC) propuesta original	53
4.3.1 Análisis de convergencia del (MSNARC)	53
4.3.2 Análisis de estabilidad numérica del (MSNARC)	55
<b>CAPÍTULO 5 EXPERIMENTOS NUMÉRICOS</b>	<b>60</b>

**LISTA DE TABLAS**

<b>TABLA</b>	<b>DESCRIPCIÓN</b>	<b>PÁGINA</b>
5.1	Comportamiento de la convergencia para el experimento No.1	61
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

**LISTA DE GRÁFICAS**

<b>GRÁFICA</b>	<b>DESCRIPCIÓN</b>	<b>PÁGINA</b>
5.1	Comportamiento de la convergencia para el experimento No.1	61
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

En este trabajo de tesis se presenta una modificación novedosa al método simplificado de Newton que elimina su inestabilidad numérica para el cálculo de la raíz cuadrada principal de una matriz no singular. El algoritmo que se obtiene es atractivo por ser computacionalmente sencillo de implementar, robusto, computacionalmente económico, convergente, y para propósitos prácticos, numéricamente estable.

En el primer capítulo de esta tesis se tratan temas relacionados con el álgebra matricial como: operaciones elementales entre matrices, tipos de matrices, la inversa de una matriz, transpuesta de una matriz, determinante de una matriz y valores propios de una matriz.

En el segundo capítulo se tratan temas relacionados con la raíz cuadrada de una matriz cuadrada no singular, existencia de la raíz cuadrada de una matriz, la representación canónica de Jordan de una matriz, y el número de raíces cuadradas de una matriz.

En el tercer capítulo se analizan las variantes más actuales y relevantes del Método de Newton (MN) para el cálculo de la raíz cuadrada de una matriz cuadrada no singular y se presenta una modificación obteniendo el Método Simplificado de Newton (MSN) del cual se estudia su convergencia y estabilidad numérica.

En el cuarto capítulo se presentan algunos algoritmos para el cálculo de la raíz cuadrada principal de una matriz no singular, como es el (IASMN) (Iteración alternativa simplificada del Método de Newton), y el (MSNFSRC) (Método Simplificado de Newton con factorizaciones sucesivas para el cálculo de la raíz cuadrada), y se analiza de su convergencia y estabilidad numérica, además se propone el Método Simplificado de Newton Acoplado para el cálculo de la raíz cuadrada principal de una matriz no singular (MSNARC) y se hace un análisis de su convergencia y estabilidad numérica, este algoritmo es atractivo por ser computacionalmente sencillo de implementar, robusto, económico, convergente y numéricamente estable para cualquier matriz no singular.

Por último, en el quinto capítulo se presentan una serie de experimentos numéricos donde se compara el desempeño del (MSNARC) con respecto a los métodos tratados.

## ANTECEDENTES

*Para la elaboración de los antecedentes, se revisaron los siguientes libros y artículos [5], [6], [3], [25], [17], [8], [28], [27] y [29]*

El estudio de raíces cuadradas de matrices comenzó con Cayley [5], cuando en 1958 en uno de sus primeros artículos sobre teoría de matrices publicó fórmulas para el cálculo de raíces cuadradas de matrices de  $2 \times 2$  y de  $3 \times 3$ . Más tarde proporcionó más detalles sobre ambos casos, todo lo anterior motivado por una pregunta de Tait que fue “¿cómo calcular la raíz cuadrada de un tensor?” [6].

Un método natural para el cálculo de la raíz cuadrada de una matriz cuadrada no singular es el Método de Newton (MN), pero este algoritmo presenta el problema que en cada iteración se resuelve una ecuación matricial de Sylvester para matrices densas  $XH + HX = A - X^2$ , lo cual requiere la utilización del método de Bartels-Stewart [3] con el consiguiente alto costo computacional lo cual lo hace poco atractivo.

Simplificando al (MN), surgen dos iteraciones que representan el Método Simplificado de Newton (MSN) que son  $X_{k+1} = 0.5(X_k + X_k^{-1}A)$  y  $Y_{k+1} = 0.5(Y_k + AY_k^{-1})$  ambas presentan muy buenas propiedades de convergencia, pero su estabilidad numérica es muy pobre. Esta inestabilidad fue notada primeramente por Laasonen [25], quien en 1958 afirmó sin demostrarlo que para matrices con valores propios reales positivos, que el (MSN) “si el proceso iterativo es llevado a cabo indefinidamente, no es estable cuando el número de condición de la matriz en cuestión excede el valor de nueve”. Higham [17], en 1986 realizó un análisis rigurosamente matemático de la estabilidad numérica del (MSN) y probó lo que Laasonen había observado.

Además del (MN) y del (MSN), existen otros métodos que son convergentes y numéricamente estables para calcular la raíz cuadrada de una matriz no singular, como por ejemplo Denman y beavers [8], en 1976 presentaron un par de iteraciones acopladas para el cálculo de la raíz cuadrada de una matriz no singular. Meini [28], en el 2004 desarrolló el método de reducción cíclica. Long [27], en el 2008 desarrolló el algoritmo de Newton con incorporación de búsqueda lineal exacta, y Mendoza [29], en el 2010 desarrolló el método (IASMN) para el cálculo de raíces cuadradas de matrices simétricas positivas definidas.

La rica variedad de métodos para el cálculo de la raíz cuadrada de una matriz no singular, junto con el análisis de su convergencia y estabilidad numérica, son en la actualidad temas de estudio e investigación de gran interés.

La raíz cuadrada de una matriz es una de las funciones de matrices más comúnmente utilizada, mostrándose más comúnmente en el contexto de matrices positivas definidas simétricas. El cálculo de la raíz cuadrada de una matriz desempeña un papel especial, por ejemplo: en la modelación estocástica de acuíferos, en la generación sintética de variables aleatorias, en el cálculo de la función signo matricial, en el problema general de los valores propios definidos, en la descomposición polar de una matriz y en el cálculo de la media geométrica. Todo lo anterior la convierte en una útil herramienta computacional y teórica. La necesidad del cálculo de la raíz cuadrada de matrices surge a partir de problemas de la ciencia y la ingeniería, como, por ejemplo: en la modelación estocástica, en la construcción de funciones de matrices, en la solución de ecuaciones diferenciales, en control automático, en robótica, entre otros.

## DEFINICIÓN DEL PROBLEMA

Algunos métodos actuales para calcular la raíz cuadrada de una matriz (como por ejemplo el Método de Newton (MN) y el Método Simplificado de Newton (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 cuadrada de una matriz definida positiva es que su número de condición debe ser menor que nueve, esto representa una 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 de resolver un sistema lineal de  $n^2 \times n^2$ , lo cual lo hace poco atractivo.

## JUSTIFICACIÓN

Los métodos actuales para el cálculo de la raíz cuadrada de una matriz definida positiva, 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 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 Método Simplificado de Newton (MSN) ambas presentan muy buenas propiedades de convergencia, pero su estabilidad numérica es muy pobre, por tal motivo Mendoza [29] realiza ciertas modificaciones al (MSN) obteniendo el (MSNFSRC) el cual es convergente y numéricamente estable para fines prácticos.

Nos podemos dar cuenta de la rica variedad de métodos que existen para el cálculo de la raíz cuadrada de una matriz no singular, 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 se presenta como propuesta un nuevo método el cual es atractivo por ser computacionalmente sencillo de implementar, robusto, económico, convergente y numéricamente estable para cualquier matriz definida positiva denominado (MSNARC).

## **OBJETIVO PRINCIPAL**

Este trabajo de tesis, tiene como objetivo exponer y analizar los diferentes métodos para el cálculo de la raíz cuadrada principal de una matriz no singular, y la propuesta de un nuevo método que es convergente y numéricamente estable.

## CAPÍTULO 1 ÁLGEBRA MATRICIAL

Para la elaboración de este capítulo, se revisaron los siguientes libros [2] y [1]

En ingeniería el álgebra matricial es utilizada en el planteamiento y solución de problemas, por ejemplo, en el control de recursos humanos y materiales en un sistema orientado a la optimización del uso de estos.

El cálculo de matrices nos ayuda a solucionar problemas de la ciencia e ingeniería, tales como: la modelación estocástica de acuíferos, en la construcción de funciones de matrices, en la solución de ecuaciones diferenciales, en control automático, en robótica, entre otros.

A continuación, se abordarán conceptos básicos del álgebra matricial, que serán necesarios para la comprensión del cálculo de la raíz cuadrada de matrices tema de la presente tesis.

### 1.1 Matrices

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

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

Las matrices se representan por letras mayúsculas en negrita  $A$ . El elemento de la fila  $i$ -ésima y de la columna  $j$ -ésima se representa por una letra minúscula con un par de subíndices,  $a_{ij}$ . De aquí, un modo abreviado de escribir una matriz es  $A = [a_{ij}]$  para  $i = 1, 2, \dots, m$  y  $j = 1, 2, \dots, n$ . 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 números de cualquier clase. Se consideran aquí matrices de números reales,  $a_{ij} \in R$ .

Por ejemplo, la matriz

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

es una matriz rectangular de orden  $3 \times 4$ ; el elemento de la fila 3 y columna 3 es 11.

## 1.2 Operaciones básicas con matrices

### 1.2.1 Igualdad de 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 todo  $i = 1, 2, \dots, m$  y  $j = 1, 2, \dots, n$ .

### 1.2.2 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 todo  $i = 1, 2, \dots, m$  y  $j = 1, 2, \dots, n$ .

La suma de matrices cumple las propiedades

- a) Conmutativa:  $\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$
- b) Asociativa:  $(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$
- c) Existencia de elemento neutro o matriz nula  $\mathbf{0} = [0]$ :  $\mathbf{A} + \mathbf{0} = \mathbf{0} + \mathbf{A} = \mathbf{A}$
- d) Existencia de matriz opuesta:  $\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$

### 1.2.3 Multiplicación por un escalar

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

### 1.2.4 Resta de matrices

La resta 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 todo  $i = 1, 2, \dots, m$  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.

### 1.2.5 Multiplicación de matrices

Sean  $A = [a_{ij}]$  y  $B = [b_{ij}]$  dos matrices de órdenes  $m \times n$  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  $c_{ij} = \sum_{k=1}^n a_{ik}b_{kj}$  para  $i = 1, 2, \dots, m$  y  $j = 1, 2, \dots, n$ .

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 con las propiedades:

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

La multiplicación de matrices no cumple la propiedad conmutativa la propiedad conmutativa

$$AB \neq BA$$

### 1.2.6 Trasposición de matrices

La matriz traspuesta de una matriz  $A$  de orden  $m \times n$  es la matriz  $A'$  de orden  $n \times m$  se obtiene permutando las filas por las columnas

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

La trasposición de matrices cumple las siguientes propiedades:

- a) Reflexiva:  $(A')' = A$ ,
- b)  $(A + B)' = A' + B'$ , la traspuesta de la suma de dos matrices es la suma de las matrices traspuestas.

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

### 1.2.7 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

$$tr(A) = a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^m a_{ii}$$

Es claro que se cumplen las siguientes propiedades:

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

### 1.3 Determinante de una matriz

Dada una matriz  $A$  de orden  $n$ , se define el determinante de  $A$ , que se notara indistintamente por  $|A|$  o  $\det(A)$ , como la suma de los  $n!$  productos signados de  $n$  factores que se obtienen considerando los elementos de la matriz, de forma que cada producto contenga un elemento y solo uno de cada fila y cada columna de  $A$ .

Analíticamente esto significa que:

$$|A| = \sum_k^{n!} (-1)^{s_k} a_{1j_1} a_{2j_2} \dots a_{nj_n}$$

Dónde:

- $\{j_1, j_2, \dots, j_n\}$  es una de las  $n!$  permutaciones de los elementos del conjunto de números naturales  $\{1, 2, 3, \dots, n\}$ .

- $S_k$  es el número de trasposiciones o cambios necesarios para reordenar la permutación  $\{j_1, j_2, \dots, j_n\}$  en el orden de  $\{1, 2, 3, \dots, n\}$ .

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

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

Sea  $A = [a_{ij}]$  una matriz de orden  $m \times n$  y sea  $M_{ij} = [m_{ij}]$  la submatriz de orden  $m - 1 \times n - 1$  que se obtiene de eliminar la fila  $i$  y la columna  $j$  de  $A$ . Se denomina menor del elemento  $a_{ij}$  al determinante de la matriz  $M_{ij}$  y cofactor del elemento  $a_{ij}$  a la cantidad  $(-1)^{i+j}|M_{ij}|$

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 (1.2) Sea la matriz  $A$

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

El determinante de  $A$  es

$$\begin{aligned} |A| &= \begin{vmatrix} 1 & 1 & 3 \\ 1 & 1 & 0 \\ 3 & 1 & 2 \end{vmatrix} = 3 \begin{vmatrix} 1 & 1 \\ 3 & 1 \end{vmatrix} - 0 \begin{vmatrix} 1 & 1 \\ 3 & 1 \end{vmatrix} + 2 \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix} = (3x - 2) - (0x - 2) + (2x - 0) \\ &= -6 \end{aligned}$$

#### 1.4 Matriz inversa

Sean  $A$  y  $B$  dos matrices cuadradas de orden  $n$  de forma que  $AB = BA = I$ ; en estas condiciones,  $B$  es la matriz inversa de  $A$ , ( $B = A^{-1}$ ), y recíprocamente,  $A$  es la inversa de  $B$ , ( $A = B^{-1}$ ).

Conocer si el valor de  $|A|$  es cero o no permite determinar no solo la existencia de la inversa de  $A$ , sino que, en caso de que exista  $A^{-1}$  es posible también obtener su expresión en función de  $|A|$ .

Algunas propiedades de la matriz inversa son las siguientes:

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

La inversa de una matriz  $A$  se calcula del siguiente modo

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

Donde  $\text{adj}(A)$  es la matriz adjunta o traspuesta de la matriz de cofactores de  $A$

$$\text{adj}(A) = \begin{pmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{pmatrix}$$

Ejemplo (1.3) Dada la matriz  $A$

$$A = \begin{pmatrix} 4 & 1 & -1 \\ 0 & 3 & 2 \\ 3 & 0 & 7 \end{pmatrix}$$

Con un determinante  $|A| = 99$ , para hallar su inversa, en primer lugar se calcula  $\text{adj}(A)$  que es:

$$\text{adj}(A) = \begin{pmatrix} \begin{vmatrix} 3 & 2 \\ 0 & 7 \end{vmatrix} & -\begin{vmatrix} 1 & -1 \\ 0 & 7 \end{vmatrix} & \begin{vmatrix} 1 & -1 \\ 3 & 2 \end{vmatrix} \\ -\begin{vmatrix} 0 & 2 \\ 3 & 7 \end{vmatrix} & \begin{vmatrix} 4 & -1 \\ 3 & 7 \end{vmatrix} & -\begin{vmatrix} 4 & -1 \\ 0 & 2 \end{vmatrix} \\ \begin{vmatrix} 0 & 3 \\ 3 & 0 \end{vmatrix} & -\begin{vmatrix} 4 & 1 \\ 3 & 0 \end{vmatrix} & \begin{vmatrix} 3 & 1 \\ 0 & 3 \end{vmatrix} \end{pmatrix} = \begin{pmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{pmatrix}$$

Por tanto

$$A^{-1} = \frac{1}{99} \begin{pmatrix} 21 & -7 & 5 \\ 6 & 31 & -8 \\ -9 & 3 & 12 \end{pmatrix}$$

## 1.5 Matrices cuadradas especiales

### 1.5.1 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,  $a_{ij} = 0 \forall i \neq j$ ,

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

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

### 1.5.2 Matriz identidad

Es una matriz diagonal cuyos elementos de la diagonal principal son todos iguales a uno, se denota por  $I_m$ .

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

### 1.5.3 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  $\lambda$  por una matriz identidad,  $\lambda I_m$ .

### 1.5.4 Matriz triangular inferior

Es una matriz cuadrada cuyos elementos por encima de la diagonal principal son todos nulos,  $a_{ij} = 0 \forall i < j$ .

$$A = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{pmatrix}$$

### 1.5.5 Matriz triangular superior

Es una matriz cuadrada cuyos elementos por debajo de la diagonal principal son todos nulos,  $a_{ij} = 0 \forall i > j$ .

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ 0 & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{mm} \end{pmatrix}$$

### 1.5.6 Matriz nula

Es una matriz (cuadrada o rectangular) cuyos elementos son todos iguales a cero, se denota por  $0$ .

### 1.5.7 Matriz simétrica

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

### 1.5.8 Matriz idempotente

Es una matriz cuadrada que cumple  $A^2 = AA = A$ .

### 1.5.9 Matriz ortogonal

Es una matriz cuadrada que cumple  $AA' = I_m$ .

## 1.6 Autovalores de una matriz

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

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

Que es una ecuación polinomial en  $\lambda$  de orden  $m$

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

Ejemplo (1.4) La ecuación característica de la matriz  $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$  es:

$$|A - \lambda I| = \begin{vmatrix} 1 - \lambda & 2 \\ 2 & 1 - \lambda \end{vmatrix} = (1 - \lambda)^2 - 4 = \lambda^2 - 2\lambda - 3 = 0$$

Los autovalores de la matriz  $A$  son las raíces  $\lambda_1 = -1$  y  $\lambda_2 = 3$

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 autovalores de una matriz simétrica pertenecen al cuerpo de los números reales.

### 1.7 Diagonalización de matrices

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

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

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

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 de  $P$ , son los correspondientes autovectores  $p_1, \dots, p_m$ .

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

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

La descomposición espectral de una matriz simétrica  $A$  de orden  $m$  es

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

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

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 reacciones

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

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

Y los elementos de las siguientes columnas

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

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

## CAPÍTULO 2 FORMA CANÓNICA DE JORDAN PARA CARACTERIZAR LA RAÍZ CUADRADA DE UNA MATRIZ

### 2.1. Existencia de raíces cuadradas de matrices

Para la elaboración de este capítulo, se revisaron los libros y artículos [14], [24], [10], [13] y [18].

Dada una matriz  $A$   $n \times n$  con elementos complejos  $A \in \mathbb{C}^{n \times n}$ , una solución  $X \in \mathbb{C}^{n \times n}$  de la ecuación cuadrática matricial.

$$F(X) \equiv X^2 - A = 0 \quad (2.1)$$

es una raíz cuadrada de  $A$ , y la raíz cuadrada particular  $A^{1/2}$  es definida por:

$$(A^{1/2})^2 = A \quad \text{y} \quad \operatorname{Re} \lambda_k(A^{1/2}) > 0 \quad \text{para toda } k$$

donde  $\lambda_k(A)$  es un valor propio de  $A$ .

Cualquier solución de (2.1) proporciona una raíz cuadrada de  $A \in \mathbb{C}^{n \times n}$ . Cuando  $n = 1$ , entonces estamos en el plano complejo además es bien sabido que la raíz cuadrada de cualquier número complejo existe. Todo número complejo diferente de cero tiene dos raíces cuadradas distintas. A diferencia de la raíz cuadrada de un escalar, la raíz cuadrada de una matriz  $A \in \mathbb{C}^{n \times n}$  puede no existir.

Por ejemplo, la matriz singular:

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

no tiene raíz cuadrada. Sin embargo la matriz singular:

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

Tiene raíces cuadradas  $X_1 = A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$  y  $X_2 = -A = \begin{pmatrix} -1 & 0 \\ 0 & 0 \end{pmatrix}$

por lo tanto la existencia de la raíz cuadrada de una matriz no es un tema tan sencillo de resolver y abordar. Para caracterizar las raíces cuadradas de una matriz, primeramente,

presentaremos la forma canónica de Jordan, la cual puede ser usada para obtener las condiciones de existencia de una raíz cuadrada de una matriz no singular.

## 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

$$A = PJP^{-1} \in \mathbb{C}^{n \times n}, \quad (2.2)$$

Donde

$$J = \text{diag}(J_1, J_2, \dots, J_p), \quad J_i = J_i(\lambda_i) = \begin{pmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \lambda_i & 1 \\ \lambda_i & & & \lambda_i \end{pmatrix} \in \mathbb{C}^{m_i \times m_i}. \quad (2.3)$$

Además también se cumple que:

$$f(A) = Pf(J)P^{-1} \in \mathbb{C}^{n \times n} \quad (2.4)$$

Los  $J_i(\lambda_i)$  son los llamados bloques de Jordan y están dispuestos en diagonal

La forma (2.3) 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  $\lambda_i$  es la matriz cuadrada  $k \times k$  que tiene el número  $\lambda_i$  en la diagonal principal; El número 1 en la supradiagonal principal; y el número 0 en los restantes términos.

Por ejemplo, el bloque de Jordan de tamaño  $3 \times 3$  y valor propio 5 es

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

El bloque de Jordan de tamaño  $2 \times 2$  y valor propio  $-4$  es

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

El bloque de Jordan de tamaño  $1 \times 1$  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  $k$ ,  $\lambda_1$  será su único valor propio.

Un suprabloque de Jordan de tamaño  $m \times m$  y valor propio  $\lambda_1$  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  $\lambda_1$ , 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  $6 \times 6$  de valor propio 7, formado por tres bloques de Jordan de tamaños  $2 \times 2$ ,  $3 \times 3$  y  $1 \times 1$  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 supra bloque  $2 \times 2$  formado por dos bloques de valor propio 7, un suprabloque  $3 \times 3$  formado por dos bloques de valor propio cero y un supra bloque  $1 \times 1$  de valor propio 3, es:

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

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 multiplicidad 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 diagonal (es decir, todos sus términos fuera de la diagonal principal son nulos) si y solo si los bloques de Jordan tienen todos tamaño  $1 \times 1$ , y esto ocurre si y solo 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 las formas canónicas 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 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  $\lambda_i$ , tamaño  $m_i \times m_i$ , y formado por  $r_i$  bloques.

Por ejemplo, sea la matriz  $A$ :

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

La forma canónica de Jordan de  $A$  será:

El polinomio característico de  $A$   $\det(A - \lambda I) = -(\lambda - 4)(\lambda + 1)^3 \Rightarrow \lambda_1 = 4$  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  $1 \times 1$ , y otro de valor propio -1 y tamaño  $3 \times 3$ . Para determinar completamente la matriz  $J$  necesitamos saber cuantos bloques forman el suprabloque de tamaño  $3 \times 3$ , o sea la multiplicidad geométrica  $r_2$  del valor propio -1. Para ello hallemos el subespacio propio.

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

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

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

Si la matriz dada  $A$  ya es una forma canónica de Jordan, es necesario hacer cálculos para determinar  $J$  y  $P$  tales que  $A = PJP^{-1}$ . En efecto basta tomar  $J = A$  y  $P = I$

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

1. La forma canónica de Jordan es 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 Jordan es 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  $2 \times 2$  y tiene un solo valor propio doble  $\lambda_1$  con multiplicidad geométrica igual a 1, entonces su forma canónica de Jordan es  $J = \begin{bmatrix} \lambda_1 & 1 \\ 0 & \lambda_1 \end{bmatrix}$

Existen procedimientos, que no veremos en general, para encontrar una matriz  $P$  de cambio de base Jordan: es decir una matriz invertible  $P$  tal que  $A = PJP^{-1}$

### 2.3. El número de raíces cuadradas de una matriz

El siguiente teorema caracteriza las raíces cuadradas de una matriz cuadrada compleja y determina cuáles de ellas corresponden a funciones de matrices en el sentido de (2.4)

*Teorema: Si  $A \in \mathbb{C}^{n \times n}$  es una matriz no singular con la descomposición de Jordan*

$$A = PJP^{-1} = P \text{diag}(J_{\lambda_1}, J_{\lambda_2}, \dots, J_{\lambda_r}) P^{-1} \quad (2.5)$$

Donde  $J_{\lambda_r} \in \mathbb{C}^{n \times n}$ ,  $r$  es el número de bloques de Jordan,  $\sum_{k=1}^r n_k = n$  que es el número de valores propios y  $s$  es el número de valores propios diferentes de  $A$ . Entonces si ( $s \leq r$ ) se verifica lo siguiente:

Que la matriz  $A$  tiene precisamente  $2^s$  raíces cuadradas que son funciones de  $A$  en el sentido de (2.4)

Ahora sabemos que cualquier matriz  $A \in \mathbb{C}^{n \times n}$  no singular tiene raíz cuadrada, eso significa que la ecuación (2.1) tiene solución. Sin embargo, el número de raíces cuadradas varía de 2 al infinito. Por ejemplo analicemos la matriz

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

De (2.5) La forma canónica de Jordan de  $A$  será:  $\text{diag}(J_{\lambda_1}, J_{\lambda_2}, \dots, J_{\lambda_r}) = Q^{-1}AQ$ , por lo tanto utilizando la función de Matlab  $[Q, J]=\text{jordan}(A)$  con una aproximación de 3 decimales tenemos

$$J = \text{diag}(J_{\lambda_1}, J_{\lambda_2}, \dots, J_{\lambda_r}) = \begin{bmatrix} 0.139 & 0 \\ 0 & 28.861 \end{bmatrix}$$

Como puede observarse la forma canónica de Jordan de  $A$  está compuesta de 2 bloques de Jordan o sea  $r = 2$ , además tiene 2 valores propios diferentes o sea  $s = 2$ , dado que se cumple que  $(s \leq r)$  entonces  $A$  tiene  $2^2 = 4$  raíces cuadradas que con una aproximación de 3 decimales son las siguientes;

$$X_1, X_2 = \pm \begin{pmatrix} 1.567 & 1.741 \\ 2.611 & 4.178 \end{pmatrix} \quad X_3, X_4 = \pm \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

Analicemos otro ejemplo, calculemos las raíces cuadradas de la matriz

$$A = \begin{pmatrix} -3 & -3 \\ 3 & 5 \\ -3 & -3 & 1 \end{pmatrix}$$

De (2.5) La forma canónica de Jordan de  $A$  será:

$$J = \text{diag}(J_{\lambda_1}, J_{\lambda_2}, \dots, J_{\lambda_r}) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}$$

Como puede observarse la forma canónica de Jordan de  $A$  está compuesta de 2 bloques de Jordan o sea  $r = 2$ , además tiene 2 valores propios diferentes o sea  $s = 2$ , dado que se cumple que  $(s \leq r)$  entonces  $A$  tiene  $2^2 = 4$  raíces cuadradas que con una aproximación de 3 decimales son las siguientes;

$$X_1, X_2 = \mp \begin{pmatrix} -1.606 & -1.242 \\ 1.607 & 3.021 & 1.234 \\ -1.242 & -1.243 & 1.000 \end{pmatrix} \quad X_3, X_4 = \mp \begin{pmatrix} -19.607 & -7.243 \\ 19.607 & 7.243 \\ -18.192 & -7.243 & -1 \end{pmatrix}$$

Por último desarrollaremos detalladamente paso por paso el siguiente ejemplo y calculemos las raíces cuadradas de la matriz  $A$ .

$$A = \begin{pmatrix} 322 & -323 & -323 & 322 \\ 325 & -326 & -325 & 323 \\ -259 & 261 & 261 & -260 \\ -237 & 237 & 238 & -237 \end{pmatrix}$$

Primero obtendremos el polinomio característico de  $A$  de la siguiente manera

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

$$(322 - \lambda) \begin{vmatrix} -326 - \lambda & -325 & 326 \\ 261 & 261 - \lambda & -260 \\ 237 & 238 & -237 - \lambda \end{vmatrix} + 323 \begin{vmatrix} 325 & -325 & 326 \\ -259 & 261 - \lambda & -260 \\ -237 & 238 & -237 - \lambda \end{vmatrix} -$$

$$323 \begin{vmatrix} 325 & -326 - \lambda & 326 \\ -259 & 261 & -260 \\ -237 & 237 & -237 - \lambda \end{vmatrix} - 322 \begin{vmatrix} 325 & -326 - \lambda & -325 \\ -259 & 261 & 261 - \lambda \\ -237 & 237 & 238 \end{vmatrix} = 0$$

Desarrollando y simplificando tendremos el polinomio característico de  $A$  :

$$(\lambda - 5)^4 = 0$$

Por lo tanto  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = 5$  o sea el único valor propio de  $A$  es cinco y aparece con multiplicidad igual a cuatro lo cual nos indica que tendremos que checar los subespacios propios del valor propio  $\lambda_i = 5$  para determinar de cuantos bloques de Jordan estará compuesta la forma canónica de Jordan de la matriz  $A$ . El número de bloques lo proporciona la siguiente fórmula.

**No. De bloques = tamaño de  $A$  - rango  $(A - 5I)$**

$$(A - 5I) = \begin{pmatrix} 317 & -323 & -323 & 322 \\ 325 & -331 & -325 & 323 \\ -259 & 261 & 256 & -260 \\ -237 & 237 & 238 & -242 \end{pmatrix}$$

El rango se define como el número de columnas o renglones que son linealmente independientes, en este caso utilizando la función  $rank(A - 5I)$  de Matlab, se obtiene un valor del rango igual a 3, por lo tanto el número de bloques de Jordan será:  $(4-3)=1$ , entonces la forma canónica de Jordan tendrá la siguiente forma.

$$J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 5 & 1 & 0 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 5 \end{pmatrix}$$

Dado que  $r = 1$  y  $s = 1$  por lo tanto el número de raíces cuadradas será  $2^1 = 2$ .

Aplicando (2.4)

$$\sqrt{A} = P\sqrt{J}P^{-1}$$

Ahora se tienen dos problemas: el primero es calcular la raíz cuadrada de  $\sqrt{J}$  que puede decirse que es relativamente sencillo, y el segundo que es calcular  $P$ . Las columnas de  $P$  serán los vectores propios de  $A$ . Calcularemos primero  $\sqrt{J}$ .

Como es sabido, la raíz cuadrada de una matriz triangular superior, también será otra matriz diagonal superior, por lo tanto debe existir una matriz triangular superior  $X$  tal que  $X^2 = J$ .

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ 0 & x_{22} & x_{23} & x_{24} \\ 0 & 0 & x_{33} & x_{34} \\ 0 & 0 & 0 & x_{44} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ 0 & x_{22} & x_{23} & x_{24} \\ 0 & 0 & x_{33} & x_{34} \\ 0 & 0 & 0 & x_{44} \end{pmatrix} = \begin{pmatrix} 5 & 1 & 0 & 0 \\ 0 & 5 & 1 & 0 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 5 \end{pmatrix}$$

Desarrollando tendremos que:  $x_{11} = x_{22} = x_{33} = x_{44} = \sqrt{5}$  luego

$$x_{12} = \frac{1}{x_{11} + x_{22}} = \frac{1}{2\sqrt{5}} = x_{23} = x_{34}; \quad x_{13} = \frac{-x_{12}x_{23}}{(x_{33} + x_{11})} = \frac{-1}{40\sqrt{5}} = x_{24}$$

$$x_{14} = -\frac{x_{12}x_{24} + x_{13}x_{34}}{(x_{11} + x_{44})} = -\frac{\frac{1}{2\sqrt{5}} \cdot \frac{-1}{40\sqrt{5}} + \frac{-1}{40\sqrt{5}} \cdot \frac{1}{2\sqrt{5}}}{2\sqrt{5}} = -\frac{1}{400\sqrt{5}}$$

Por lo tanto: 
$$\sqrt{J} = \begin{pmatrix} \sqrt{5} & \frac{1}{2\sqrt{5}} & \frac{-1}{40\sqrt{5}} & -\frac{1}{400\sqrt{5}} \\ 0 & \sqrt{5} & \frac{1}{2\sqrt{5}} & \frac{-1}{40\sqrt{5}} \\ 0 & 0 & \sqrt{5} & \frac{1}{2\sqrt{5}} \\ 0 & 0 & 0 & \sqrt{5} \end{pmatrix}$$
 Ahora calcularemos  $P$

Una forma sencilla de obtener los  $(n - 1)$  vectores propios sin resolver un sistema de ecuaciones lineales se basa en el teorema de Cayley-Hamilton que establece que cada matriz cuadrada satisface su propio polinomio característico. Así, si  $\lambda_1, \lambda_2, \dots, \lambda_n$  son los valores propios de  $A$  se cumple que

$$(A - \lambda_1 I)(A - \lambda_2 I) \dots (A - \lambda_n I) = 0,$$

por lo que los vectores columna de

$$(A - \lambda_2 I) \dots (A - \lambda_n I)$$
 son vectores propios de  $\lambda_1$

Por lo tanto el polinomio característico de  $A$  tiene la forma

$$(\lambda_1 - 5)(\lambda_2 - 5)(\lambda_3 - 5)(\lambda_4 - 5) = 0$$

Aplicando el teorema de Cayley Hamilton la primer columna del producto

$(A - 5I)(A - 5I)(A - 5I)$  Nos proporciona la primera columna de  $P$

Desarrollando tendremos:

$$(A - 5I)(A - 5I)(A - 5I) = \begin{pmatrix} \mathbf{5922} & -9198 & -9828 & 6048 \\ \mathbf{4230} & -6570 & -7020 & 4320 \\ -3572 & 5548 & 5928 & -3648 \\ -5170 & 8030 & 8580 & -5280 \end{pmatrix}$$

La columna en negritas es la primer columna ( $v_1$ ) de  $P$ , luego tenemos

$$(A - 5I)(A - 5I) = \begin{pmatrix} \mathbf{2857} & -3467 & -3468 & 2832 \\ \mathbf{2363} & -2977 & -3012 & 2352 \\ -1962 & 2462 & 2488 & -1952 \\ -2392 & 2868 & 2858 & -2368 \end{pmatrix}$$

La columna en negritas es la segunda columna ( $v_2$ ) de  $P$ , continuando tenemos

$$(A - 5I) = \begin{pmatrix} \mathbf{317} & -323 & -323 & 322 \\ \mathbf{325} & -331 & -325 & 326 \\ -259 & 261 & 256 & -260 \\ -237 & 237 & 238 & -242 \end{pmatrix}$$

La columna en negritas es la tercer columna ( $v_3$ ) de  $P$ , continuando tenemos que se deben de cumplir las siguientes relaciones para matrices con 4 valores propios.

$$(A - \lambda_1 I)v_2 = v_1 \quad y \quad (A - \lambda_2 I)v_4 = v_3$$

Ya hemos calculado  $v_1$ ,  $v_2$  y  $v_3$ , y como tenemos 4 valores propios iguales, por lo tanto:

$$v_4 = (A - \lambda_2 I) \setminus v_3$$

El símbolo " $\setminus$ " significa el proceso de eliminación gaussiana, sustituyendo valores tenemos.

$$v_4 = \begin{pmatrix} \mathbf{317} & -323 & -323 & 322 \\ \mathbf{325} & -331 & -325 & 326 \\ -259 & 261 & 256 & -260 \\ -237 & 237 & 238 & -242 \end{pmatrix} \setminus \begin{bmatrix} 317 \\ 325 \\ -259 \\ -237 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Finalmente la forma que tendrá la matriz de transformación  $P$  es:

$$P = \begin{pmatrix} 5922 & 2857 & 317 & 1 \\ 4230 & 2363 & 325 & 0 \\ -3572 & -1962 & -259 & 0 \\ -5170 & -2392 & -237 & 0 \end{pmatrix}$$

Bien, ahora ya estamos en condiciones de calcular la  $\sqrt{A}$

$$\sqrt{A} = P\sqrt{J}P^{-1}$$

Sustituyendo valores y ejecutando operaciones con una aproximación de 4 decimales se llega a que

$$\sqrt{A} = \pm \begin{pmatrix} 47.7982 & -43.7464 & -44.4396 & 47.1005 \\ 50.9823 & -45.8394 & -46.8456 & 51.4296 \\ -39.9720 & 37.0382 & 38.2904 & -40.3923 \\ -32.0317 & 29.9074 & 30.8577 & -31.3049 \end{pmatrix}$$

Como puede verse, el cálculo de la raíz cuadrada de una matriz 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 raíces cuadradas de matrices, sólo es útil para su caracterización. Debido a lo anteriormente expuesto, lo más adecuado para el cálculo de raíces cuadradas de matrices, 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 El método de Newton para el cálculo de la raíz cuadrada principal de una matriz singular

Para la elaboración de este capítulo, se revisaron los siguientes artículos y libros [14], [18], [30] [26] [19] y [7]

El Método de Newton (MN) es uno de los métodos numéricos más poderosos y bien conocidos para resolver un sistema no lineal  $f(x) = 0$ . Para una función escalar  $f$ , el (MN) está definido por la siguiente iteración

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}, \quad f'(x_k) \neq 0, \quad k = 0, 1, 2, \dots \quad (3.1)$$

Una aproximación natural al calcular una raíz cuadrada de una matriz  $A$  es aplicar el (MN), consideremos una función general  $F: \mathbb{C}^{n \times n} \rightarrow \mathbb{C}^{n \times n}$  para dar solución a  $F(X) = 0$  el (MN) propone una solución inicial  $X_0$  y el siguiente proceso iterativo [30].

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

donde  $F'$  denota la derivada de Fréchet de  $F$  [26], como se define en (3.3)

$$F(X_k + H) - F(X_k) = F'(X_k)H + \omega(X_k, H) \quad \text{para toda } \|H\| < \varepsilon, \quad \text{para algun } \varepsilon > 0 \quad (3.3)$$

Donde  $\frac{\|\omega(X_k, H)\|}{\|H\|} \rightarrow 0$  cuando  $\|H\| \rightarrow 0$

Expandiendo  $F(X + H)$  en series de Taylor a primer orden se tiene

$$F(X + H) = F(X) + F'(X)H + \frac{1}{2}F''(X)H^2 + \dots$$

$$F'(X)H = F(X + H) - F(X) \quad (3.4)$$

Para  $F(X) = X^2 - A$  tenemos que

$$F(X + H) = (X + H)^2 - A$$

$$F(X + H) - (X^2 - A) = F(X + H) - F(X) = XH + HX + H^2, \quad H^2 \rightarrow 0$$

Y comparando con (3.4) vemos que

$$F'(X)H = XH + HX \quad (3.5)$$

De (3.2) se tiene

$$F'(X_k)(X_{k+1} - X_k) = F'(X_k)H_k = -F(X_k) \quad (3.6)$$

Donde  $H_k = X_{k+1} - X_k$ . Combinando (3.5) y (3.6) se obtiene el (MN) para el cálculo de la raíz cuadrada de una matriz no singular  $A$

$$\left. \begin{array}{l} X_0 = A, \quad \text{o} \quad X_0 = I \\ X_k H_k + H_k X_k = A - X_k^2 \\ X_{k+1} = X_k + H_k \end{array} \right\} \quad k = 0, 1, 2, \dots \quad (3.7)$$

El teorema de convergencia local estándar para el (MN) [30] dice que, siempre  $\|X - X_0\|$  sea suficientemente pequeño y la transformación lineal  $F'(X)$  sea no singular la iteración de Newton (3.7) converge cuadráticamente a una raíz cuadrada de  $A$ .

Para calcular la raíz cuadrada de  $A$  nos obliga a resolver (3.7) para  $H_k$ , esto se puede llevar a cabo con la ayuda del operador  $\text{vec}$  que para  $A$  se define como  $\text{vec}(A)$  (que significa el apilamiento de las columnas de  $A$  una debajo de otra iniciando de izquierda a derecha), junto con el producto de Kronecker  $A \otimes B = (a_{ij}B)$ . Aplicando el operador  $\text{vec}$  a (3.7) y utilizando la propiedad  $\text{vec}(AXB) = (B^* \otimes A)\text{vec}(X)$  [7] y [19] se obtiene

$$(I \otimes X + X^* \otimes I)\text{vec}(H) = \text{vec}(A - X^2). \quad (3.8)$$

El sistema lineal (3.8) puede ser resuelto utilizando cualquier método estándar, siempre que la matriz de coeficientes sea no singular. Sin embargo (3.8) es un sistema lineal de  $n^2 \times n^2$ , lo que implica un alto costo computacional para su solución que de acuerdo a [18], es del orden de  $2n^6 \vartheta \log_2 p$ ,  $1 \leq \vartheta \leq 2$  operaciones aritméticas por iteración, lo cual hace poco atractivo al (MN) para el cálculo de la raíz cuadrada una matriz no singular.

Una consideración razonable para simplificar (3.7) es que el producto  $X_k H_k$  conmute, es decir

$$X_k H_k = H_k X_k$$

Entonces la iteración (3.7) puede ser escrita como

$$\left. \begin{aligned} X_0 = A, \text{ o } X_0 = I \\ 2X_k H_k = A - X_k^2 \\ X_{k+1} = X_k + H_k \end{aligned} \right\} \quad k = 0, 1, 2, \dots$$

Obteniéndose dos nuevas iteraciones conocidas como el Método Simplificado de Newton (MSN)

$$\left. \begin{aligned} Y_0 = A, \text{ o } Y_0 = I \\ Y_{k+1} = 0.5(Y_k + Y_k^{-1}A) \end{aligned} \right\} \quad (3.9)$$

$$\left. \begin{aligned} Z_0 = A, \text{ o } Z_0 = I \\ Z_{k+1} = 0.5(Z_k + AZ_k^{-1}) \end{aligned} \right\} \quad (3.10)$$

### 3.2 Análisis de convergencia del Método Simplificado de Newton (MSN)

Teorema 1. Sea  $An = A\|A\|_F^{-1}$  una matriz normalizada no singular con valores propios  $\lambda_i$ ,  $\{Re(\lambda_i) > 0, 0 < |\lambda_i| \leq 1\}$  el (MSN) en la forma (3.11) converge cuadráticamente a una raíz cuadrada principal  $Y$  de  $An$  siempre que  $\|Y - Y_0\|$  sea lo suficientemente pequeño y  $Y_k^{-1}$  sea no singular. Es decir  $Y = \sqrt{An} = \sqrt{A}\|A\|_F^{-1/2}$  o sea  $\sqrt{A} = Y\|A\|_F^{1/2}$ .

Partiendo del (MSN) (3.9) y considerando que  $An \in \mathbb{C}^{n \times n}$  es una matriz no singular

$$Y_{k+1} = \frac{1}{2}(Y_k + Y_k^{-1}An) \quad (3.11)$$

Se analizara su convergencia a través de la forma canónica de Jordan  $An = PJP^{-1}$  o bien  $J = P^{-1}AnP$  obteniendo la diagonalización de (3.11) de la siguiente manera.

De acuerdo con la forma canónica de Jordan para una matriz  $An$  no singular, existe una matriz de paso  $P$  no singular de tal manera:

$$\tau = P^{-1}AnP = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_i], \quad n \geq i \geq 1 \quad (3.12)$$

$\tau$  es una matriz diagonal cuyos elementos  $(\lambda_i)$  son los valores propios de  $An$ .

Aplicando la representación (3.12) se tiene que:

$$P^{-1}Y_{k+1}P = D_{k+1} = \text{diag}[d_1^{k+1}, d_2^{k+1}, \dots, d_i^{k+1}]$$

$$P^{-1}Y_kP = D_k = \text{diag}[d_1^k, d_2^k, \dots, d_i^k]$$

$$P^{-1}Y_k^{-1}P = D_k^{-1} = \text{diag}\left[\frac{1}{d_1^k}, \frac{1}{d_2^k}, \dots, \frac{1}{d_i^k}\right]$$

Sustituyendo las matrices diagonales equivalentes en (3.11)

$$D_{k+1} = \frac{1}{2}(D_k + D_k^{-1}\tau)$$

Suponiendo que se trata de una matriz  $A$  de  $3 \times 3$

$$\begin{bmatrix} d_1^{k+1} & 0 & 0 \\ 0 & d_2^{k+1} & 0 \\ 0 & 0 & d_3^{k+1} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} d_1^k & 0 & 0 \\ 0 & d_2^k & 0 \\ 0 & 0 & d_3^k \end{bmatrix} + \begin{bmatrix} (d_1^k)^{-1} & 0 & 0 \\ 0 & (d_2^k)^{-1} & 0 \\ 0 & 0 & (d_3^k)^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

Desacoplando se tiene que

$$d_1^{k+1} = \frac{1}{2} \left( d_1^k + \frac{\lambda_1}{d_1^k} \right)$$

$$d_2^{k+1} = \frac{1}{2} \left( d_2^k + \frac{\lambda_2}{d_2^k} \right)$$

$$d_3^{k+1} = \frac{1}{2} \left( d_3^k + \frac{\lambda_3}{d_3^k} \right)$$

Lo anterior puede representarse de la siguiente manera

$$d_i^{k+1} = \frac{1}{2} \left( d_i^k + \frac{\lambda_i}{d_i^k} \right) \quad (3.13)$$

Manipulando algebraicamente (3.13) se tiene

$$2d_i^k(d_i^{k+1}) = (d_i^k)^2 + \lambda_i \quad (3.14)$$

$$2d_i^k(d_i^{k+1}) - 2\lambda_i = (d_i^k)^2 - \lambda_i \quad (3.15)$$

Dividiendo (3.15) por (3.14) se tiene que

$$\frac{d_i^k(d_i^{k+1}) - \lambda_i}{d_i^k(d_i^{k+1})} = \frac{(d_i^k)^2 - \lambda_i}{(d_i^k)^2 + \lambda_i} \quad \text{Para } k = 0, 1, 2, \dots \quad (3.16)$$

$$G(d_i^k, \lambda_i) = \frac{(d_i^k)^2 - \lambda_i}{(d_i^k)^2 + \lambda_i} \quad (3.17)$$

Sustituyendo como valor inicial  $d_i^0 = 1$  en (3.17) se obtiene

$$G(\lambda_i) = (1 - \lambda_i) \frac{1}{1 + \lambda_i} \quad (3.18)$$

Para calcular  $d_i^1$  se sustituye  $d_i^0$  en (3.13) y se obtiene

$$d_i^1 = \frac{1}{2} \left( 1 + \frac{\lambda_i}{1} \right) = \frac{1 + \lambda_i}{2} \quad (3.19)$$

Sustituyendo (3.19) en (3.17) se obtiene

$$G(d_i^1, \lambda_i) = (1 - \lambda_i)^2 \frac{1}{1 + 6\lambda_i + \lambda_i^2} \quad (3.20)$$

Para calcular  $d_i^2$  se sustituye (3.19) en (3.13) y se obtiene

$$d_i^2 = \frac{1}{2} \left( \frac{2\lambda_i}{1 + \lambda_i} + \frac{1 + \lambda_i}{2} \right) \quad (3.21)$$

Sustituyendo (3.21) en (3.17) se obtiene

$$G(d_i^2, \lambda_i) = (1 - \lambda_i)^4 \frac{1}{1 + 28\lambda_i + 70\lambda_i^2 + 28\lambda_i^3 + \lambda_i^4} \quad (3.22)$$

Para calcular  $d_i^3$  se sustituye (3.21) en (3.13) y se obtiene

$$d_i^3 = \frac{1}{2} \left( \frac{2\lambda_i}{\frac{1+\lambda_i}{2} + \frac{1+\lambda_i}{2}} + \frac{1}{2} \left( \frac{2\lambda_i}{1+\lambda_i} + \frac{1+\lambda_i}{2} \right) \right) \quad (3.23)$$

Sustituyendo (3.23) en (3.17) se obtiene

$$G(d_i^k, \lambda_i) = (1 - \lambda_i)^8 \frac{1}{1+120\lambda_i+1820\lambda_i^2+8008\lambda_i^3+12870\lambda_i^4+8008\lambda_i^5+1820\lambda_i^6+120\lambda_i^7+\lambda_i^8} \quad (3.24)$$

Analizando las ecuaciones desde (3.18) a (3.24) se observa que presenta un comportamiento que se ajusta a la siguiente expresión

$$G(d_i^k, \lambda_i) = (1 - \lambda_i)^{(2^k)} \left( \frac{1}{P(\lambda_i)} \right) \quad (3.25)$$

Donde  $P(\lambda_i)$  es un polinomio de grado  $2^k$  que representa un número complejo cuyo módulo es mayor que la unidad.

Igualando (3.17) con (3.25) se obtiene la siguiente relación

$$\frac{(d_i^k)^2 - \lambda_i}{(d_i^k)^2 + \lambda_i} = (1 - \lambda_i)^{(2^k)} \left( \frac{1}{P(\lambda_i)} \right) \quad (3.26)$$

Esta relación por su forma nos permitirá probar que  $d_i^k = \sqrt{\lambda_i}$ , cuando  $k \rightarrow \infty$ , aplicando límites en ambos lados

$$\lim_{k \rightarrow \infty} \left| \frac{(d_i^k)^2 - \lambda_i}{(d_i^k)^2 + \lambda_i} \right| = \lim_{k \rightarrow \infty} \left| (1 - \lambda_i)^{(2^k)} \frac{1}{P(\lambda_i)} \right|$$

Desarrollando

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

Debido a las características de  $An$   $0 < |1 - \lambda_i| \leq 1$  además  $\left| \frac{1}{p(\lambda_i)} \right| < 1$  para toda  $i$ , por lo tanto

$$\lim_{k \rightarrow \infty} \left| (d_i^k)^2 - \lambda_i \right| = 0 \quad (3.27)$$

La expresión (3.27) será verdadera si y solo si  $(d_i^k)^2 = \lambda_i$  cuando  $k \rightarrow \infty$ , lo cual implica que:

$$d_i^k = \sqrt{\lambda_i} \text{ cuando } k \rightarrow \infty \quad (3.28)$$

Acoplando (3.28)

$$D_k = \sqrt{\tau} \text{ cuando } k \rightarrow \infty \quad (3.29)$$

luego tenemos que

$$ZD_kZ^{-1} = Z\sqrt{\tau}Z^{-1} \text{ cuando } k \rightarrow \infty \quad (3.30)$$

$$Y_k = Y = \sqrt{An} = \sqrt{A}\|A\|_F^{-1/2} \text{ cuando } k \rightarrow \infty \quad (3.31)$$

finalmente

$$\sqrt{A} = Y\|A\|_F^{1/2}.$$

De esta manera queda demostrado que el (MSN) es convergente.

### 3.3 Análisis de estabilidad numérica del (MSN)

Por conveniencia se trabajará con la forma (3.11) del (MSN):

Si  $\hat{Y}_k$  representa la  $k$ -ésima iteración perturbada con la matriz de perturbaciones  $\Delta_k$  entonces (Higham, 1987)[14].

$$\hat{Y}_k \approx Y_k \quad (3.32)$$

$$\Delta_k = \hat{Y}_k - Y_k \quad (3.33)$$

Perturbaremos (3.11) con (3.33), para analizar cómo se comporta la propagación de esta perturbación en la  $(k+1)$ -ésima iteración perturbada representada por la matriz  $\widehat{Y}_{k+1}$ . Considerando que no se introducen nuevos errores de redondeo en el cálculo de  $\widehat{Y}_{k+1}$  se tiene que

$$\begin{aligned}\widehat{Y}_{k+1} &= \frac{1}{2}(\widehat{Y}_k + \widehat{Y}_k^{-1}An) \\ &= \frac{1}{2}(Y_k + \Delta_k + (Y_k + \Delta_k)^{-1}An)\end{aligned}\quad (3.34)$$

Desarrollando

$$\widehat{Y}_{k+1} = \frac{1}{2}(Y_k + \Delta_k + Y_k^{-1}An - Y_k^{-1}\Delta_k Y_k^{-1}An)\quad (3.35)$$

Restando (3.11) de (3.35) se obtiene el error  $\Delta_{k+1}$

$$\Delta_{k+1} = \frac{1}{2}(\Delta_k - Y_k^{-1}\Delta_k Y_k^{-1}A) + O\|\Delta_k^2\|\quad (3.36)$$

Diagonalizando (3.36) se tiene que

$$P^{-1}\Delta_{k+1}P = \widehat{\Delta}^{k+1} = [\widehat{\delta}_{ij}^{k+1}]\quad (3.37)$$

$$P^{-1}\Delta_k P = \bar{\Delta}^k = [\widehat{\delta}_{ij}^k]\quad (3.38)$$

$$P^{-1}Y_k^{-1}P = D_k^{-1} = \text{diag}\left[\frac{1}{d_1^k}, \frac{1}{d_2^k}, \dots, \frac{1}{d_i^k}\right]\quad (3.39)$$

$$P^{-1}AP = \tau = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_i]\quad (3.40)$$

Sustituyendo en (3.36) desde (3.37) hasta (3.40) se tiene

$$\widehat{\delta}_{k+1} = \frac{1}{2}\left(\widehat{\delta}_{ij}^k - \frac{\widehat{\delta}_{ij}^k \cdot \lambda_j}{(d_i^k) \cdot (d_j^k)}\right)\quad (3.41)$$

Simplificando (3.41)

$$\hat{\delta}_{k+1} = \frac{1}{2} \hat{\delta}_{ij}^k \left( 1 - \frac{\lambda_j}{(d_i^k) \cdot (d_j^k)} \right)$$

Como quedó demostrado en el análisis de convergencia

Si  $k \rightarrow \infty$ ,  $d_i^k = \sqrt{\lambda_i}$  y  $d_j^k = \sqrt{\lambda_j}$  por lo tanto

$$\hat{\delta}_{k+1} = \frac{1}{2} \hat{\delta}_{ij}^k \left( 1 - \frac{(\lambda_j)^{1/2}}{(\lambda_i)^{1/2}} \right)$$

$$\hat{\delta}_{k+1} = \hat{\delta}_{ij}^k \frac{1}{2} \left( 1 - \left( \frac{\lambda_j}{\lambda_i} \right)^{\frac{1}{2}} \right)$$

Dónde debe cumplirse que  $\frac{1}{2} \left| 1 - \left( \frac{\lambda_j}{\lambda_i} \right)^{\frac{1}{2}} \right| < 1$  para asegurar la no amplificación de los errores

y por ende la convergencia del (MSN). Se tiene que  $\max \left( \frac{\lambda_i}{\lambda_j} \right)^{\frac{1}{2}} = K_2(\mathbf{An})^{\frac{1}{2}}$  donde  $K_2(\mathbf{An})$  es el número de condición de  $\mathbf{An}$

$$\frac{1}{2} \left| 1 - K_2(\mathbf{An})^{\frac{1}{2}} \right| \leq 1 \quad (3.42)$$

Resolviendo (3.42)

$$2 \geq -1 + K_2(\mathbf{An})^{\frac{1}{2}} \geq -2$$

$$3 \geq K_2(\mathbf{An})^{\frac{1}{2}} \geq -1$$

$$9 \geq K_2(\mathbf{An}) \geq 1 \quad (3.43)$$

De acuerdo a (3.43) el (MSN) es numéricamente estable para matrices con número de condición mayor igual que 1 pero menor igual que 9.

Lo anterior representa una fuerte restricción que hace poco atractivo al (MSN) para el cálculo de la raíz cuadrada principal de una matriz no singular.

## **CAPÍTULO 4 ALGUNOS ALGORITMOS EXISTENTES PARA EL CÁLCULO DE LA RAÍZ CUADRADA PRINCIPAL DE UNA MATRIZ NO SINGULAR.**

*Para la elaboración de este capítulo, se revisaron los siguientes artículos y libros [17], [8], [28], [27], [29], [14], [13], [19], [23], [32], [12], [11], [15], [25], [4], [16], [22], [21], [20], [9], [31] y [33].*

### **4.1 Un método simplificado de Newton para calcular la raíz cuadrada de una matriz real simétrica definida positiva (IASMN)**

Este algoritmo fue desarrollado por el C. D. Alfredo Mendoza Mexía, Maestro de tiempo completo del Departamento de Física Matemáticas e Ingeniería de la Universidad de Sonora Unidad Regional Sur.

Este trabajo de investigación original fue publicado por la Revista Internacional de Métodos Numéricos de la Universidad Politécnica de Catalunya de España en marzo del 2010. A continuación, se reproducen las secciones más importantes.

Un método usual para calcular la raíz cuadrada de una matriz es el Método de Newton (MN)[23][11], además de este método existen muchos otros, por ejemplo: Higham[15] analizó el comportamiento de un conjunto de métodos para calcular la raíz cuadrada de matrices en general, contemplando también el caso hermitiano. Para el cálculo de raíces cuadradas de matrices en general, Meini[28] desarrolló el método de reducción cíclica y, más recientemente, Long[27] desarrolló el algoritmo de Newton con incorporación de búsqueda lineal exacta.

El problema de calcular la raíz cuadrada de una matriz se puede encontrar a menudo en problemas de la ciencia y la ingeniería [12], por ejemplo: en la modelación estocástica de acuíferos, en la construcción de funciones de matrices, en la solución de ecuaciones diferenciales, en control automático, en robótica, entre otros. La teoría de existencia de la raíz cuadrada de una matriz y del número de raíces está bien documentada en [19],[14]. En este trabajo se desarrolla una iteración alternativa simplificada del Método de Newton (IASMN) para calcular la raíz cuadrada de una matriz real simétrica definida positiva que es computacionalmente económica, convergente y numéricamente estable.

#### 4.1.1 Iteración alternativa simplificada del método de Newton (IASMN) para el cálculo de raíces cuadradas principales de matrices reales simétricas definidas positivas.

Sea  $A \in \mathbb{R}^{n \times n}$  una matriz real simétrica definida positiva, es decir, satisface que  $A = A^T$  y  $q^T A q > 0$  para todo vector  $q \neq 0$  en  $\mathbb{R}^n$ , además todos los valores propios de  $A$  son positivos[23].

La matriz real simétrica definida positiva  $X \in \mathbb{R}^{n \times n}$  es una raíz cuadrada de  $A$  si satisface la siguiente ecuación cuadrática matricial

$$F(X) = X^2 - A = 0 \quad (4.1)$$

Por simetría de  $X$  se tiene que

$$X = X^T \quad (4.2)$$

$$X(X) = X^T(X) \quad (4.3)$$

Sustituyendo (4.3) en (4.1)

$$F(X) = X^T X - A = 0 \quad (4.4)$$

Construyamos las iteraciones del MN a partir de (4.4)

$$\lim_{\|H\| \rightarrow 0} [F(X+H) - F(X)] = X^T H + H^T X \quad (4.5)$$

de la siguiente expresión

$$\lim_{H \rightarrow 0} \frac{\|F(X+H) - F(X) - F'(X)H\|}{\|H\|} = 0$$

Donde  $F'$  representa la derivada de Fréchet [32] de  $F$ , se tiene que

$$F'(X)H = F(X+H) - F(X) \quad (4.6)$$

La perturbación  $H_k$  se define por

$$H_k = X_{k+1} - X_k \quad (4.7)$$

Combinando (4.7) con (4.4), (4.5) y (4.6) se llega a las iteraciones del MN

$$\begin{aligned} \text{Dado } X_0 &= \alpha I_n, \quad \alpha > 0 [5] \\ X_k^T H_k + H_k^T X_k &= A - X_k^T X_k \\ X_{k+1} &= X_k + H_k \\ \text{Para } k &= 0, 1, 2, \dots \end{aligned} \quad (4.8)$$

#### 4.1.2 Análisis de convergencia del (IASMN)

*Teorema 1.* Sea  $A \in \mathbb{R}^{n \times n}$  una matriz real simétrica definida positiva, si  $\|X - X_0\|$  es lo suficientemente pequeño y la transformación lineal  $F(X)$  es no singular entonces las iteraciones  $\{X_k\}$  proporcionadas por (4.8) convergen a  $X$  cuando  $k \rightarrow \infty$ .

De acuerdo con (4.7) las iteraciones (4.8) se representan por

$$X_k^T (X_{k+1} - X_k) + (X_{k+1} - X_k)^T X_k = A - X_k^T X_k \quad (4.9)$$

Desarrollando (4.9) y simplificando

$$X_k^T X_{k+1} + X_{k+1}^T X_k = A + X_k^T X_k = X_k^T X_k + A \quad (4.10)$$

Para  $k = 0$ , la solución inicial en (4.10) es  $X_0 = \alpha I_n$ , con  $\alpha > 0$ , entonces

$$X_0^T X_1 + X_1^T X_0 = X_0^T X_0 + A \quad (4.11)$$

Donde  $X_0$  y  $(X_0^T X_0 + A)$  son matrices reales simétricas definidas positivas, entonces los productos  $X_0^T X_1$  y  $X_1^T X_0$  también representan matrices reales definidas positivas, por consiguiente  $X_1$  y  $X_1^T$  son también matrices reales definidas positivas.

Por las características de  $A$ , existe una matriz  $Q$  no singular tal que

$$Q^{-1}AQ = \Sigma = \text{diag}(\lambda_1, \dots, \lambda_n) \quad (4.12)$$

Donde  $\lambda_1, \dots, \lambda_n$  son los valores propios de la matriz  $A$ . También se tiene que

$$Q^{-1}X_1Q = D_1 = \text{diag}(d_1^k, \dots, d_n^k) \quad (4.13)$$

$$Q^{-1}X_1^TQ = D_1^T = D_1 \quad (4.14)$$

$$Q^{-1}X_0Q = D_0 = D_0^T \quad (4.15)$$

Diagonalizando y simplificando (4.11) se tiene

$$D_1 = \frac{1}{2}(D_0 + D_0^{-T}\Sigma) \quad (4.16)$$

Como puede verse  $D_1$  es una matriz diagonal, utilizando el proceso inverso veremos que

$$X_1 = QD_1Q^{-1} = \frac{1}{2}(QD_0Q^{-1} + (QD_0Q^{-1})^T Q\Sigma Q^{-1}) = \frac{1}{2}(X_0 + X_0^{-T}A)$$

Es fácil comprobar que  $X_1$ , además de ser una matriz real definida positiva, también es simétrica.

De nuevo en (4.10) hacemos  $k=1$  y, utilizando el mismo procedimiento para  $k=0$  se encuentra que  $D_2$  también será diagonal y, que  $X_2$ , es una matriz real simétrica definida positiva, y así sucesivamente, por lo tanto, generalizando se tiene

$$D_{k+1} = \frac{1}{2}(D_k + D_k^{-T}\Sigma) \quad (4.17)$$

La ecuación (4.17) representa esencialmente  $n$  iteraciones escalares desacopladas de Newton para las raíces cuadradas  $\sqrt{\lambda_i}$ ,  $1 \leq i \leq n$ , o sea

$$d_i^{k+1} = \frac{1}{2}\left(d_i^k + \frac{\lambda_i}{d_i^k}\right), \quad 1 \leq i \leq n \quad (4.18)$$

Utilizando la siguiente iteración escalar [13]

$$z_{k+1} = \frac{1}{2} \left( z_k + \frac{a}{z_k} \right) \quad (4.19)$$

$$z_{k+1} \pm \sqrt{a} = \frac{(z_k \pm \sqrt{a})^2}{2z_k} \quad (4.20)$$

$$\frac{z_{k+1} - \sqrt{a}}{z_{k+1} + \sqrt{a}} = \left( \frac{z_0 - \sqrt{a}}{z_0 + \sqrt{a}} \right)^{2^{k+1}} \equiv \gamma^{2^{k+1}} \quad (4.21)$$

Para  $z_0$  y  $a$  positivos se tiene que  $|\gamma| < 1$ , entonces:

$$\lim_{k \rightarrow \infty} z_k = \sqrt{a} \quad (4.22)$$

Debido a que los valores propios  $\lambda_i$  y los valores de la solución inicial  $a_i^{(0)} = \alpha > 0$  tienen la misma forma de  $a$  y de  $z_0$  respectivamente, entonces se cumple que:

$$\lim_{k \rightarrow \infty} D_k = \sqrt{\Sigma} = \text{diag} \left( \lambda_i^{\frac{1}{2}} \right) \quad (4.23)$$

Por lo tanto,

$$\lim_{k \rightarrow \infty} QD_k Q^{-1} = \lim_{k \rightarrow \infty} X_k = Q\sqrt{\Sigma}Q^{-1} = \sqrt{A} = X. \quad (4.24)$$

La IASMN, se obtiene a partir de (4.17) de la siguiente manera

$$QD_{k+1}Q^{-1} = \frac{1}{2} (QD_kQ^{-1} + QD_k^{-T}Q^{-1}Q\Sigma Q^{-1}) \quad (4.25)$$

$$X_{k+1} = \frac{1}{2} (X_k + X_k^{-T}A) \quad (4.26)$$

Utilizando en (4.26) la eliminación gaussiana ( $\setminus$ ) en lugar de la inversa por ser más eficaz en cuanto al tiempo de ejecución y exactitud numérica, se llega a la IASMN para calcular la raíz cuadrada  $X$  de una matriz  $A$  real simétrica definida positiva, la cual tiene la siguiente expresión.

$$\begin{aligned} \text{Dado } X_0 &= \alpha I_n, \alpha > 0 \quad [5] \\ X_{k+1} &= \frac{1}{2} (X_k + X_k^T \setminus A) \quad (4.27) \\ \text{Para } k &= 0, 1, 2, \dots \end{aligned}$$

*Corolario.* Sea  $A \in \mathbb{R}^{n \times n}$  una matriz real simétrica definida positiva, si la solución inicial en (4.24) es de la forma  $X_0 = \alpha I_n$  con  $\alpha > 0$ , entonces todas las iteraciones  $\{X_k\}$  proporcionadas por (4.24) son reales simétricas definidas positivas.

### 4.1.3 Análisis de estabilidad numérica de la (IASMN)

Por conveniencia en la IASMN se utilizará la inversa en lugar de la eliminación gaussiana, es decir:

$$X_{k+1} = \frac{1}{2} \left( X_k + (X_k^T)^{-1} A \right) \quad (4.28)$$

Si  $\hat{X}_k$  representa la  $k$ -ésima iteración perturbada con la matriz de perturbaciones  $\Delta_k$  entonces [27]

$$\begin{aligned} \hat{X}_k &\approx X_k \\ \Delta_k &= \hat{X}_k - X_k \end{aligned} \quad (4.29)$$

Perturbaremos (4.28) con (4.29), para analizar la propagación de estas perturbaciones en la  $(k+1)$ -ésima iteración perturbada representada por la matriz  $\hat{X}_{k+1}$ . Considerando que no se introducen nuevos errores de redondeo en el cálculo de  $\hat{X}_{k+1}$  se tiene

$$\hat{X}_{k+1} = \frac{1}{2} \left( \hat{X}_k + (\hat{X}_k^T)^{-1} A \right) = \frac{1}{2} \left( X_k + \Delta_k + \left[ (X_k + \Delta_k)^T \right]^{-1} A \right) \quad (4.30)$$

Utilizando una aproximación de primer orden en series de Taylor para  $(X_k^T + \Delta_k^T)^{-1}$  y considerando que  $\|\Delta_k^T\|$  es una cantidad muy pequeña [12]

$$(X_k^T + \Delta_k^T)^{-1} = (X_k^T)^{-1} - (X_k^T)^{-1} \Delta_k^T (X_k^T)^{-1} + \mathcal{O}(\|\Delta_k^T\|^2) \quad (4.31)$$

Sustituyendo (4.31) en (4.30) y desarrollando

$$\hat{X}_{k+1} = \frac{1}{2} \left( X_k + \Delta_k + (X_k^T)^{-1} A - (X_k^T)^{-1} \Delta_k^T (X_k^T)^{-1} A \right) + \mathcal{O}(\|\Delta_k^T\|^2) \quad (4.32)$$

Restando (4.28) a (4.32)

$$\Delta_{k+1} = \frac{1}{2} \left( \Delta_k - (X_k^T)^{-1} \Delta_k^T (X_k^T)^{-1} A \right) + \mathcal{O}(\|\Delta_k^T\|^2) \quad (4.33)$$

Utilizando la notación (4.12) para diagonalizar (4.33)

$$\bar{\Delta}_{k+1} = \frac{1}{2} \left( \bar{\Delta}_k - (D_k^T)^{-1} \bar{\Delta}_k^T (D_k^T)^{-1} \Sigma \right) + \mathcal{O}(\|\bar{\Delta}_k^T\|^2) \quad (4.34)$$

Por lo tanto tenemos que

$$\begin{aligned} D_k^T &= D_k = \text{diag}(d_i^k) \\ \bar{\Delta}_k &= \delta_{ij}^k \\ \bar{\Delta}_k^T &= \delta_{ji}^k \end{aligned} \quad (4.35)$$

Expresando a (4.34) como  $n$  iteraciones desacopladas

$$\delta_{ij}^{k+1} = \frac{1}{2} \left( \delta_{ij}^k - \frac{\delta_{ji}^k}{d_j^k d_i^k} \lambda_i \right) + \mathcal{O}(\|\bar{\Delta}_k^T\|^2) \quad 1 \leq i, j \leq n. \quad (4.36)$$

Si la matriz  $A$  está bien condicionada y el número de iteraciones  $k$  tiende a infinito entonces la matriz de perturbaciones  $\Delta_k$  conserva su simetría, es decir,  $\bar{\Delta}_k^T = \bar{\Delta}_k$  y como consecuencia

$$\delta_{ji}^k = \delta_{ij}^k \quad (4.37)$$

Rescribiendo (4.36)

$$\delta_{ij}^{k+1} = \frac{1}{2} \left( \delta_{ij}^k - \frac{\delta_{ij}^k}{d_j^k d_i^k} \lambda_i \right) + \mathcal{O} \left( \|\bar{\Delta}_k\|^2 \right) \quad (4.38)$$

Factorizando

$$\delta_{ij}^{k+1} = \frac{1}{2} \delta_{ij}^k \left( 1 - \frac{\lambda_i}{d_j^k d_i^k} \right) + \mathcal{O} \left( \|\bar{\Delta}_k\|^2 \right) \quad (4.39)$$

Si  $k \rightarrow \infty$  entonces se tiene que  $d_j^k = \lambda_j^{1/2} + \varepsilon_j^k$ ,  $d_i^k = \lambda_i^{1/2} + \varepsilon_i^k$ , con  $\varepsilon_i^k$  y  $\varepsilon_j^k \rightarrow 0$

Rescribiendo (4.39)

$$\delta_{ij}^{k+1} = \frac{1}{2} \delta_{ij}^k \left( 1 - \frac{\lambda_i^{1/2}}{\lambda_j^{1/2}} \right) + \mathcal{O} \left( \|\bar{\Delta}_k\|^2 \right) \quad 1 \leq i, j \leq n \quad (4.40)$$

Para asegurar la estabilidad numérica de (4.25), y por ende la de la IASMN se debe de cumplir que

$$\frac{1}{2} \left| 1 - \frac{\lambda_i^{1/2}}{\lambda_j^{1/2}} \right| \leq 1 \quad (4.41)$$

El número de condición [17] de  $A$  es  $k_2(A) = \|A\|_2 \|A^{-1}\|_2 = \frac{\lambda_j}{\lambda_i}$   $1 \leq i, j \leq n$ , por lo tanto,

$$\frac{\lambda_i}{\lambda_j} = k_2(A)^{-1}$$

Rescribiendo (4.41)

$$\frac{1}{2} \left| 1 - k_2(A)^{-1/2} \right| \leq 1 \quad (4.42)$$

Resolviendo (4.42) para  $k_2(A)$  se obtienen dos resultados

$$k_2(A) \geq 1 \quad (4.43)$$

$$k_2(A) \geq \frac{1}{9} \quad (4.44)$$

Dado que la matriz  $A$  es real simétrica definida positiva, el mínimo valor para el número de condición que se espera es la unidad, por lo tanto, el resultado correcto para este caso es (4.43). Con esto se demuestra que la IASMN es, para propósitos prácticos, numéricamente estable para números de condición mayores que la unidad, siempre y cuando este número de condición no represente un mal condicionamiento.

#### 4.1.4 Factor de escala para la (IASMN)

Con la finalidad de disminuir el tiempo de convergencia de la IASMN, cada una de las iteraciones  $X_k$  será multiplicada por un factor de escala  $\alpha_k > 0$ , de tal manera que se minimice

$$\|A - \alpha_k^2 X_k^T X_k\|_F^2 \quad \text{para } k=0,1,2,\dots \quad (4.45)$$

El valor de  $\alpha_k$  que minimiza (4.42) está dado por

$$\alpha_k = \frac{\sqrt{\text{traza}(A)}}{\|X_k\|_F} \quad \alpha_k > 0 \quad \text{para } k=0,1,2,\dots \quad (4.46)$$

El costo computacional de calcular el factor de escala  $\alpha_k$  puede considerarse despreciable.

Entonces la IASMN con factor de escala tiene la siguiente expresión

$$\begin{aligned} & \text{Dado } X_0 = I_n \\ & \alpha_k = \frac{\sqrt{\text{traza}(A)}}{\|X_k\|_F} \quad \alpha_k > 0 \\ & X_{k+1} = \frac{1}{2} \left( \alpha_k X_k + (\alpha_k X_k)^T A \right) \\ & k=0,1,2,\dots \end{aligned} \quad (4.47)$$

Debido a que la IASMN es convergente, entonces el  $\lim_{k \rightarrow \infty} \alpha_k = 1$

## 4.2 Un método simplificado de Newton con factorizaciones sucesivas para el cálculo de la raíz cuadrada principal de una matriz no singular (MSNFSRC).

Este algoritmo fue desarrollado por el C. D. Alfredo Mendoza Mexía, Maestro de tiempo completo del Departamento de Física Matemáticas e Ingeniería de la Universidad de Sonora Unidad Regional Sur.

Este trabajo de investigación original fue publicado en línea en el año 2012 por la Revista de Ingeniería y Medio Ambiente de la División de Ciencias e Ingeniería de la Universidad de Sonora Unidad Regional Sur. A continuación, se reproducen las secciones más importantes.

El método de Newton (MN) para calcular la raíz cuadrada de una matriz se ha estudiado durante casi cincuenta años y actualmente esta metodología ya está bien comprendida. Desde que Laasonen [25] demostró sus propiedades de convergencia, y también observó que para ciertos casos era numéricamente inestable, varios autores han obtenido un conjunto de variantes del método de Newton y han demostrado que son convergentes y numéricamente estables, entre los que podemos destacar por ejemplo a Denman y Beavers [8], Björck y Hammarling [4], Higham [17], [14], [16], [15], He [11], Iannazzo [22], Meini [28], Long [27] y Mendoza y Gómez [29].

El problema de calcular la raíz cuadrada de una matriz se puede encontrar a menudo en problemas de la ciencia y la ingeniería [12], por ejemplo: en la modelación estocástica de acuíferos, en la construcción de funciones de matrices, en la solución de ecuaciones diferenciales, en control automático, en robótica, entre otros. La teoría de existencia de la raíz cuadrada de una matriz y del número de raíces está bien documentada en [19], [14].

En este trabajo de investigación se desarrolla un método simplificado de Newton con factorizaciones sucesivas para el cálculo de la raíz cuadrada de una matriz (MSNFSRC). Al introducir la variante de ir factorizando sucesivamente a la matriz en cuestión, se elimina la inestabilidad numérica del método simplificado de Newton (MSN). El algoritmo que se obtiene es atractivo por ser computacionalmente sencillo de implementar, robusto, económico, convergente y, para propósitos prácticos, numéricamente estable.

### 4.2.1. El método simplificado de Newton

Una simplificación razonable a las iteraciones del método de Newton (MN) es suponer que el producto  $H_k X_k$  conmuta [17], es decir,

$$H_k X_k = X_k H_k \quad (4.48)$$

Sustituyendo (4.48) en (3.7) se obtiene el bien conocido par de iteraciones simplificadas del método de Newton [25], [4], [21], [20], [9], [31].

$$X_{k+1} = \frac{1}{2}(X_k + X_k^{-1}A) \quad (4.49)$$

$$Y_{k+1} = \frac{1}{2}(Y_k + AY_k^{-1}) \quad (4.50)$$

Las iteraciones (4.49) y (4.50) en caso de converger, lo hacen en forma cuadrática hacia una raíz cuadrada de  $A$ . En [25] Higham demuestra que (4.49) es numéricamente inestable para números de condición de  $A$  mayores que nueve, haciendo esto extensivo para (4.50).

#### 4.2.2. El método simplificado de Newton con factorizaciones sucesivas para el cálculo de la raíz cuadrada de una matriz.

Para desarrollar el algoritmo se utilizará la ecuación (4.49), factorizando sucesivamente a la matriz  $A$ . Con lo anterior se elimina la inestabilidad numérica de (4.49).

De hecho la inestabilidad numérica de (4.49) demostrada en [17] se debe principalmente a la pre o post multiplicación de la inversa de  $X_k$  por la matriz  $A$ . Por otro lado, dado que  $X_k$  conmuta con  $A$  [22], la iteración (4.49) puede reescribirse como

$$X_{k+1} = \frac{1}{2}(X_k + \sqrt{A} X_k^{-1} \sqrt{A}) \quad (4.51)$$

Aplicando la metodología tratada en [17] se puede probar que (4.51) es convergente y numéricamente estable, pero evidentemente (4.51) no nos es útil, ya que involucra a  $\sqrt{A}$  que es el problema en cuestión, pero se puede utilizar un argumento similar que es el siguiente: en cada  $(k + 1)$ -ésima iteración descomponer a la matriz  $A$  como el producto de dos matrices  $B_k$  y  $C_k$ , de tal manera que se cumpla lo siguiente:

$$A = B_k C_k = C_k B_k \quad (4.52)$$

En otras palabras  $B_k$  y  $C_k$  conmutan, con esto se garantiza que  $X_k$  commute con  $B_k$  y  $C_k$ , lo cual implica que

$$X_{k+1} = \frac{1}{2}(X_k + B_k X_k^{-1} C_k) \quad (4.53)$$

$$X_{k+1} = \frac{1}{2}(X_k + X_k^{-1} B_k C_k) \quad (4.54)$$

Ahora el problema se reduce a encontrar en cada  $(k + 1)$ -ésima iteración las matrices  $B_k$  y  $C_k$  que satisfagan a (4.52); lo anterior se resuelve de la siguiente manera: Dado  $B_0 = A$ ,  $C_0 = I_n$ ,  $X_0 = I_n$  y habiéndose realizado a partir de (4.53) la  $(k + 1)$ -ésima iteración, se hace  $C_k = X_k$  y por consiguiente  $B_k = 2X_{k+1} - X_k$ , con lo anterior se garantiza que en cada  $(k + 1)$ -ésima iteración se cumple con (4.52).

Utilizando en (4.53) la eliminación gaussiana (1) en lugar de la matriz inversa por ser más eficiente en tiempo de ejecución y exactitud numérica, se obtiene el MSNFSRC para calcular la raíz cuadrada de la matriz  $A$ . Con la finalidad de reducir el tiempo de convergencia del MSNFSRC se utiliza un factor de escala  $\alpha_k$ . El algoritmo es el siguiente.

$$\left. \begin{array}{l} \text{Dado } B_0 = A, C_0 = I_n \text{ y } X_0 = I_n \\ \alpha_k = \frac{\sqrt{\text{traza}(A)}}{\sqrt{\text{traza}(X_k^2)}} \text{ (no sim.)} \\ \alpha_k = \frac{\sqrt{\text{traza}(A)}}{\|X_k\|_F} \text{ (sim.)} \\ \text{for } i = 0:w \\ \quad X_k = \alpha_k X_k \\ \quad X_{k+1} = \frac{1}{2}(X_k + B_k(X_k \setminus C_k)) \\ \quad C_k = X_k \\ \quad B_{k+1} = 2X_{k+1} - X_k \\ \quad C_{k+1} = C_k \\ \text{end} \\ X_{k+1} = X_k = B_k = C_k = X = \sqrt{A} \end{array} \right\} \quad (4.55)$$

### 4.2.3. Análisis de convergencia del (MSNFSRC)

**Teorema 1.** Sea  $A \in C^{n \times n}$  una matriz no singular y diagonalizable con valores propios  $\lambda_i$ ,  $\{\lambda_i \in C, |\lambda_i| > 0, \text{Re}\lambda_i > 0, n \geq i \geq 1\}$ . Si  $\{X_0 = \beta I_n, \beta > 0\}$ ,  $\|X - X_0\|$  es lo suficientemente pequeño y  $X_k^{-1}$  es no singular, entonces el proceso iterativo (4.55) converge a  $X = B = C = \sqrt{A}$  cuando  $k \rightarrow \infty$ , es decir

$$\lim_{k \rightarrow \infty} \{X_k\} = \lim_{k \rightarrow \infty} \{B_k\} = \lim_{k \rightarrow \infty} \{C_k\} = X = B = C = \sqrt{A}$$

Se utilizará el enfoque de Henrici [13]. Por las características de la matriz  $A$ , existe una matriz no singular  $Z$  tal que [23]

$$\tau = Z^{-1} A Z = \text{diag}(\lambda_1, \dots, \lambda_n), \quad n \geq i \geq 1 \quad (4.56)$$

Donde  $\lambda_1, \dots, \lambda_n$  son los valores propios de la matriz  $A$ , también se tiene que

$$D_k = Z^{-1} X_k Z = \text{diag}(d_1^k, \dots, d_n^k) \quad (4.57)$$

$$F_k = Z^{-1} B_k Z = \text{diag}(f_1^k, \dots, f_n^k) \quad (4.58)$$

$$G_k = Z^{-1} C_k Z = \text{diag}(g_1^k, \dots, g_n^k) \quad (4.59)$$

De (4.55), diagonalizando a

$$X_{k+1} = \frac{1}{2}(X_k + B_k(X_k \setminus C_k)) = \frac{1}{2}(X_k + B_k X_k^{-1} C_k)$$

Se tiene que

$$D_{k+1} = \frac{1}{2}(D_k + F_k D_k^{-1} G_k) \quad (4.60)$$

De acuerdo con (4.53), (4.53) y (4.54) se tiene que

$$D_{k+1} = \frac{1}{2}(D_k + D_k^{-1} F_k G_k) = \frac{1}{2}(D_k + D_k^{-1} \tau) \quad (4.61)$$

La ecuación (4.61) es idéntica a la ecuación (3.13) del capítulo 3, por lo tanto se concluye que el (MSNFSCR) es convergente.

#### 4.2.4. Análisis de estabilidad numérica del (MSNFSCR)

Para llevar a cabo el análisis de estabilidad numérica por conveniencia utilizaremos a (4.53) es decir

$$X_{k+1} = \frac{1}{2}(X_k + B_k X_k^{-1} C_k) \quad (4.62)$$

Si  $\hat{X}_k$  representa la  $k$ -ésima iteración perturbada con la matriz de perturbaciones  $\Delta_k$  entonces [27]

$$\begin{aligned} \hat{X}_k &\approx X_k \\ \Delta_k &= \hat{X}_k - X_k \end{aligned} \quad (4.63)$$

Perturbaremos (4.62) con (4.63), para analizar cómo se comporta la propagación de esta perturbación en la  $(k+1)$ -ésima iteración perturbada representada por la matriz  $\hat{X}_{k+1}$ . Considerando que no se introducen nuevos errores de redondeo en el cálculo de  $\hat{X}_{k+1}$  se tiene que

$$\bar{X}_{k+1} = \frac{1}{2} \left( \bar{X}_k + B_k \bar{X}_k^{-1} C_k \right) = \left. \begin{aligned} & \frac{1}{2} (X_k + \Delta_k + B_k (X_k + \Delta_k)^{-1} C_k) \end{aligned} \right\} \quad (4.64)$$

Desarrollando  $(X_k + \Delta_k)^{-1}$  en series de Taylor a primer orden [33], [9] y considerando que  $\Delta_k$  es muy pequeña.

$$(X_k + \Delta_k)^{-1} = X_k^{-1} - X_k^{-1} \Delta_k X_k^{-1} + O(\|\Delta_k\|^2) \quad (4.65)$$

Sustituyendo (4.65) en (4.64) y desarrollando

$$\bar{X}_{k+1} = \frac{1}{2} \left( X_k + \Delta_k + B_k X_k^{-1} C_k - \right. \left. B_k X_k^{-1} \Delta_k X_k^{-1} C_k + O(\|\Delta_k\|^2) \right) \quad (4.66)$$

Restando (4.62) de (4.66)

$$\Delta_{k+1} = \frac{1}{2} (\Delta_k - B_k X_k^{-1} \Delta_k X_k^{-1} C_k) + O(\|\Delta_k\|^2) \quad (4.67)$$

Utilizando la notación (4.42) para diagonalizar (4.67) se obtiene

$$\tilde{\Delta}_{k+1} = \frac{1}{2} (\tilde{\Delta}_k - F_k D_k^{-1} \tilde{\Delta}_k D_k^{-1} G_k) + O(\|\tilde{\Delta}_k\|^2) \quad (4.68)$$

$$\tilde{\Delta}_k = \delta_{ij}^k$$

Expresando a (4.68) como  $n$  iteraciones escalares desacopladas

$$\delta_{i,j}^{k+1} = \frac{1}{2} \left( \delta_{i,j}^k - \frac{f_i^k \delta_{i,j}^k g_j^k}{d_i^k d_j^k} \right) + O(\|\tilde{\Delta}_k\|^2), \quad \left. \begin{aligned} & 1 \leq i, j \leq n \end{aligned} \right\} \quad (4.69)$$

$$\delta_{i,j}^{k+1} = \frac{1}{2} \delta_{i,j}^k \left( 1 - \frac{f_i^k g_j^k}{d_i^k d_j^k} \right) + O(\|\tilde{\Delta}_k\|^2) \quad (4.70)$$

Si  $k \rightarrow \infty$  como se ha demostrado anteriormente  $d_i^k = f_i^k = \sqrt{\lambda_i} + \epsilon_i^k$ ,  $d_j^k = g_j^k = \sqrt{\lambda_j} + \epsilon_j^k$ , con  $\epsilon_i^k$  y  $\epsilon_j^k \rightarrow 0$

Reescribiendo (4.70)

$$\delta_{i,j}^{k+1} = \frac{1}{2} \delta_{i,j}^k \left( 1 - \frac{\sqrt{\lambda_i} \sqrt{\lambda_j}}{\sqrt{\lambda_i} \sqrt{\lambda_j}} \right) + O(\|\epsilon^k\|^2), \quad \left. \begin{aligned} & 1 \leq i, j \leq n \end{aligned} \right\} \quad (4.71)$$

Para asegurar la estabilidad numérica del MSNFSRC se debe cumplir que

$$\frac{1}{2} \left( 1 - \frac{\sqrt{\lambda_i} \sqrt{\lambda_j}}{\sqrt{\lambda_i} \sqrt{\lambda_j}} \right) \leq 1 \quad (4.72)$$

Como puede notarse la condición expuesta en (4.72) se cumple cabalmente, por lo tanto el MSNFSRC es numéricamente estable.

#### 4.2.5. Factor de escala ( $\alpha_k$ ) para el (MSNFSRC)

Con la finalidad de disminuir el tiempo de convergencia del MSNFSRC, en cada una de las iteraciones,  $X_k$  será multiplicada por un factor de escala  $\alpha_k > 0$  de tal manera que se minimice

$$\|\sqrt{A} - \alpha_k X_k\|_F^2 \quad k = 0, 1, 2, \dots \quad (4.73)$$

El valor de  $\alpha_k$  que minimiza (4.73) está dado por

$$\alpha_k = \sqrt{\frac{\text{traza}(A)}{\text{traza}(X_k^2)}}, \quad k = 0, 1, 2, \dots \text{ (no simét.)} \quad (4.74)$$

$$\alpha_k = \frac{\sqrt{\text{traza}(A)}}{\|X_k\|_F}, \quad k = 0, 1, 2, \dots \text{ (simét.)} \quad (4.75)$$

El costo computacional de calcular el factor de escala  $\alpha_k$  se puede considerar despreciable. Debido a que el MSNFSRC es convergente, entonces

$$\lim_{k \rightarrow \infty} \alpha_k = 1$$

4.3. Un método simplificado de Newton acoplado para el cálculo de la raíz cuadrada principal de una matriz no singular (MSNARC) (propuesta original)

$$\left. \begin{array}{l} X_0 = Y_0 = I \\ \text{for } i = 1:w \\ \quad X_{k+1} = 0.5(X_k + Y_k \setminus A) \\ \quad Y_{k+1} = 0.5(Y_k + (X_k' \setminus A)') \\ \quad X_k = X_{k+1} \\ \quad Y_k = Y_{k+1} \\ \text{end} \\ X_k = \sqrt{A} \\ Y_k = \sqrt{A} \end{array} \right\} \quad (4.76)$$

#### 4.3.1 Análisis de convergencia del (MSNARC)

Partiendo del (MSNARC) (4.76) y considerando que  $A \in \mathbb{C}^{n \times n}$  es una matriz no singular

$$X_{k+1} = 0.5(X_k + Y_k \setminus A) \quad (4.77)$$

$$Y_{k+1} = 0.5(Y_k + (X_k' \setminus A)') \quad (4.78)$$

Se analizara su convergencia a través de la forma canónica de Jordan  $A = PJP^{-1}$  o bien  $J = P^{-1}AP$  obteniendo la diagonalización de (4.78) de la siguiente manera.

De acuerdo con la forma canónica de Jordan para una matriz  $A$  no singular, existe una matriz de paso  $P$  no singular de tal manera:

$$\tau = P^{-1}AP = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_i], \quad n \geq i \geq 1 \quad (4.79)$$

$\tau$  es una matriz diagonal cuyos elementos  $(\lambda_i)$  son los valores propios de  $A$ .

Aplicando la representación (4.79) en (4.77) y (4.78) se tiene que:

$$P^{-1}X_{k+1}P = D_{k+1} = \text{diag}[d_1^{k+1}, d_2^{k+1}, \dots, d_i^{k+1}]$$

$$P^{-1}Y_{k+1}P = E_{k+1} = \text{diag}[e_1^{k+1}, e_2^{k+1}, \dots, e_i^{k+1}]$$

$$P^{-1}X_kP = D_k = \text{diag}[d_1^k, d_2^k, \dots, d_i^k]$$

$$P^{-1}Y_kP = E_k = \text{diag}[e_1^k, e_2^k, \dots, e_i^k]$$

$$P^{-1}X'_k P = D'_k = D_k = \text{diag}[d_1^k, d_2^k, \dots, d_i^k]$$

Sustituyendo las matrices diagonales equivalentes en (4.77) y (4.78)

$$D_{k+1} = \frac{1}{2}(D_k + (E_k \setminus \tau)) = \frac{1}{2}(D_k + (E_i^{-1} \tau)) \quad (4.80)$$

$$E_{k+1} = \frac{1}{2}(E_k + (D_k \setminus \tau)) = \frac{1}{2}(E_k + (D_k^{-1} \tau)) \quad (4.81)$$

Suponiendo que se trata de una matriz  $A$  de  $3 \times 3$  para (4.80)

$$\begin{bmatrix} d_1^{k+1} & 0 & 0 \\ 0 & d_2^{k+1} & 0 \\ 0 & 0 & d_3^{k+1} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} d_1^k & 0 & 0 \\ 0 & d_2^k & 0 \\ 0 & 0 & d_3^k \end{bmatrix} + \begin{bmatrix} (e_1^k)^{-1} & 0 & 0 \\ 0 & (e_2^k)^{-1} & 0 \\ 0 & 0 & (e_3^k)^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

Desacoplando (4.80) se tiene que

$$d_1^{k+1} = \frac{1}{2} \left( d_1^k + \frac{\lambda_1}{e_1^k} \right)$$

$$d_2^{k+1} = \frac{1}{2} \left( d_2^k + \frac{\lambda_2}{e_2^k} \right)$$

$$d_3^{k+1} = \frac{1}{2} \left( d_3^k + \frac{\lambda_3}{e_3^k} \right)$$

Lo anterior puede representarse de la siguiente manera

$$d_i^{k+1} = \frac{1}{2} \left( d_i^k + \frac{\lambda_i}{e_i^k} \right) \quad (4.82)$$

De igual manera para (4.81) tenemos que

$$\begin{bmatrix} e_1^{k+1} & 0 & 0 \\ 0 & e_2^{k+1} & 0 \\ 0 & 0 & e_3^{k+1} \end{bmatrix} = \frac{1}{2} \begin{bmatrix} e_1^k & 0 & 0 \\ 0 & e_2^k & 0 \\ 0 & 0 & e_3^k \end{bmatrix} + \begin{bmatrix} (d_1^k)^{-1} & 0 & 0 \\ 0 & (d_2^k)^{-1} & 0 \\ 0 & 0 & (d_3^k)^{-1} \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

Desacoplando (4.81)

$$e_1^{k+1} = \frac{1}{2} \left( e_1^k + \frac{\lambda_1}{d_1^k} \right)$$

$$e_2^{k+1} = \frac{1}{2} \left( e_2^k + \frac{\lambda_2}{d_2^k} \right)$$

$$e_3^{k+1} = \frac{1}{2} \left( e_3^k + \frac{\lambda_3}{d_3^k} \right)$$

Lo anterior puede representarse de la siguiente manera

$$e_i^{k+1} = \frac{1}{2} \left( e_i^k + \frac{\lambda_i}{d_i^k} \right) \quad (4.83)$$

Debido a la convergencia anteriormente probada  $d_i^k = e_i^k$ , (4.82) y (4.83) serán iguales a (3.13) por lo tanto la convergencia del MSNARC queda comprobada.

#### 4.3.2 Análisis de estabilidad numérica del (MSNARC)

Se trabajará con la forma (4.77) y (4.78) del (MSNARC) y considerando que  $Y_k \setminus A = Y_k^{-1} A$  y  $(X_k' \setminus A')' = A X_k^{-1}$  se tiene que

$$X_{k+1} = \frac{1}{2} (X_k + Y_k^{-1} A) \quad (4.84)$$

$$Y_{k+1} = 0.5 (Y_k + A X_k^{-1}) \quad (4.85)$$

Si  $\hat{X}_k$  y  $\hat{Y}_k$  representa la  $k$ -ésima iteración perturbada con la matriz de perturbaciones  $\Delta_k$  entonces (Higham, 1987)[14].

$$\hat{X}_k \approx X_k \quad (4.86)$$

$$\Delta_{Xk} = \hat{X}_k - X_k \quad (4.87)$$

Y

$$\hat{Y}_k \approx Y_k \quad (4.88)$$

$$\Delta_{Yk} = \widehat{Y}_k - Y_k \quad (4.89)$$

Perturbaremos (4.84) con (4.87), y (4.85) con (4.89) para analizar cómo se comporta la propagación de esta perturbación en la  $(k+1)$ -ésima iteración perturbada representada por la matriz  $\widehat{X}'_{k+1}$  y  $\widehat{Y}'_{k+1}$ . Considerando que no se introducen nuevos errores de redondeo en el cálculo de  $\widehat{X}'_{k+1}$  y  $\widehat{Y}'_{k+1}$  se tiene que

$$\begin{aligned} \widehat{X}_{k+1} &= \frac{1}{2} \left( \widehat{X}_k + \left( \widehat{Y}_k^{-1} A \right) \right) \\ &= \frac{1}{2} \left( (X_k + \Delta_{Xk}) + ((Y_k + \Delta_{Yk})^{-1} A) \right) \end{aligned} \quad (4.90)$$

Desarrollando (4.90)

$$\widehat{X}_{k+1} = \frac{1}{2} (X_k + \Delta_{Xk} + Y_k^{-1} A - Y_k^{-1} \Delta_{Yk} Y_k^{-1} A) \quad (4.91)$$

Restando (4.84) de (4.91) se obtiene el error  $\Delta_{Xk+1}$

$$\Delta_{Xk+1} = \frac{1}{2} (\Delta_{Xk} - Y_k^{-1} \Delta_{Yk} Y_k^{-1} A) + O\|\Delta_k^2\| \quad (4.92)$$

Y para

$$\begin{aligned} \widehat{Y}_{k+1} &= \frac{1}{2} \left( \widehat{Y}_k + \left( A \widehat{X}_k^{-1} \right) \right) \\ &= \frac{1}{2} \left( (Y_k + \Delta_{Yk}) + (A(X_k + \Delta_{Xk})^{-1}) \right) \end{aligned} \quad (4.93)$$

Desarrollando (4.93)

$$\widehat{Y}_{k+1} = \frac{1}{2} (Y_k + \Delta_{Yk} + A X_k^{-1} - A X_k^{-1} \Delta_{Xk} X_k^{-1}) \quad (4.94)$$

Restando (4.85) de (4.94) se obtiene el error  $\Delta_{Yk+1}$

$$\Delta_{Yk+1} = \frac{1}{2} (\Delta_{Yk} - A X_k^{-1} \Delta_{Xk} X_k^{-1}) + O\|\Delta_k^2\| \quad (4.95)$$

De acuerdo con la forma canónica de Jordan para una matriz  $A$  no singular, existe una matriz de paso  $P$  no singular de tal manera:

$$\tau = P^{-1}AP = \text{diag}[\lambda_1, \lambda_2, \dots, \lambda_i], \quad n \geq i \geq 1 \quad (4.96)$$

$\tau$  es una matriz diagonal cuyos elementos  $(\lambda_i)$  son los valores propios de  $A$ .

Aplicando la representación (4.96) a (4.92) y (4.95) se tiene que:

$$P^{-1}\Delta_{X_{k+1}}P = \hat{\Delta}_X^{k+1} = [\hat{\delta}_{ij}^{k+1}] \quad (4.97)$$

$$P^{-1}\Delta_{Y_{k+1}}P = \hat{\Delta}_Y^{k+1} = [\hat{\varepsilon}_{ij}^{k+1}] \quad (4.98)$$

$$P^{-1}\Delta_{X_k}P = \hat{\Delta}_X^k = [\hat{\delta}_{ij}^k] \quad (4.99)$$

$$P^{-1}\Delta_{Y_k}P = \hat{\Delta}_Y^k = [\hat{\varepsilon}_{ij}^k] \quad (4.100)$$

$$P^{-1}X_k^{-1}P = D_k^{-1} = \text{diag} \left[ \frac{1}{d_1^k}, \frac{1}{d_2^k}, \dots, \frac{1}{d_i^k} \right] \quad (4.101)$$

$$P^{-1}Y_k^{-1}P = E_k^{-1} = \text{diag} \left[ \frac{1}{e_1^k}, \frac{1}{e_2^k}, \dots, \frac{1}{e_i^k} \right] \quad (4.102)$$

Sustituyendo en (4.92) y (4.95) las matrices diagonales equivalentes desde (4.97) hasta (4.102) se tiene

$$\hat{\delta}_{k+1} = \frac{1}{2} \left( \hat{\delta}_{ij}^k - \frac{\hat{\varepsilon}_{ij}^k \cdot \lambda_j}{(d_i^k) \cdot (d_j^k)} \right) \quad (4.103)$$

$$\hat{\varepsilon}_{k+1} = \frac{1}{2} \left( \hat{\varepsilon}_{ij}^k - \frac{\lambda_i \cdot \hat{\delta}_{ij}^k}{(e_i^k) \cdot (e_j^k)} \right) \quad (4.104)$$

Debido a que como se probó anteriormente  $d^k = e^k$  y  $d^k \rightarrow \sqrt{\lambda}$  se tiene que

$$\hat{\delta}_{k+1} = \frac{1}{2} \left( \hat{\delta}_{ij}^k - \frac{\hat{\varepsilon}_{ij}^k \cdot \lambda_j}{(\lambda_i)^{\frac{1}{2}} \cdot (\lambda_j)^{\frac{1}{2}}} \right) \quad (4.105)$$

$$\hat{\varepsilon}_{k+1} = \frac{1}{2} \left( \hat{\varepsilon}_{ij}^k - \frac{\lambda_i \cdot \delta_{ij}^k}{(\lambda_i)^{\frac{1}{2}} \cdot (\lambda_j)^{\frac{1}{2}}} \right) \quad (4.106)$$

Desarrollando (4.105) y (4.106)

$$\hat{\delta}_{k+1} = \frac{1}{2} \left( \delta_{ij}^k - \frac{\hat{\varepsilon}_{ij}^k \cdot (\lambda_j)^{\frac{1}{2}}}{(\lambda_i)^{\frac{1}{2}}} \right) \quad (4.107)$$

$$\hat{\varepsilon}_{k+1} = \frac{1}{2} \left( \hat{\varepsilon}_{ij}^k - \frac{(\lambda_i)^{\frac{1}{2}} \cdot \delta_{ij}^k}{(\lambda_j)^{\frac{1}{2}}} \right) \quad (4.108)$$

Como  $\left(\frac{\lambda_j}{\lambda_i}\right)^{\frac{1}{2}} = \sqrt{K_2(A)}$  se tiene que

$$\hat{\delta}_{k+1} = \frac{1}{2} \left( \delta_{ij}^k - \hat{\varepsilon}_{ij}^k \sqrt{K_2(A)} \right) \quad (4.109)$$

$$\hat{\varepsilon}_{k+1} = \frac{1}{2} \left( \hat{\varepsilon}_{ij}^k - \frac{\delta_{ij}^k}{\sqrt{K_2(A)}} \right) \quad (4.110)$$

Debido a la convergencia anteriormente probada  $\hat{\delta}_{ij}^k = \hat{\varepsilon}_{ij}^k$  por lo cual

$$\tilde{\Delta}_{k+1}^{ij} = \frac{1}{2} \delta_{ij}^k \left( 1 - \frac{1}{\sqrt{K_2(A)}} \right) \quad (4.111)$$

Para asegurar la estabilidad numérica de MSNARC se debe cumplir lo siguiente

$$\left| 1 - \frac{1}{\sqrt{K_2(A)}} \right| \leq 1$$

Desarrollando se tiene que

$$2 \geq (K_2(A))^{\frac{1}{2}} \geq 0$$

$$K_2(A) \geq \frac{1}{4}$$

Como el valor mínimo de  $K_2(\mathbf{A})$  es 1 se cumple que el (MSNARC) será numéricamente estable para cualquier matriz no singular.

## CAPÍTULO 5 EXPERIMENTOS NUMÉRICOS

Por medio de experimentos numéricos con matrices, se comparó la convergencia y estabilidad numérica del (MSNARC) con respecto a los siguientes algoritmos: (MSN), (MSNF SRC) y (IASMN). 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.

$$err = \frac{\|Y^2 - A\|_F}{\|A\|_F}, \quad \text{o bien} \quad err = \frac{\|X^2 - A\|_F}{\|A\|_F}$$

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. Las expresiones matemáticas de los algoritmos de prueba son las siguientes. Se trabajará con la matriz normalizada  $An = A\|A\|_F^{-1}$ .

$$\left. \begin{array}{l} \text{dado } Y_0 = I \\ Y_{k+1} = 0.5(Y_k + Y_k^{-1} An) \\ \sqrt{A} = \sqrt{(\|A\|_F)} Y_k \end{array} \right\}, \quad k = 0, 1, 2, \dots, \quad (MSN)$$

$$\left. \begin{array}{l} X_0 = Y_0 = I \\ \text{for } i = 1:w \\ \quad X_{k+1} = 0.5(X_k + Y_k \backslash An) \\ \quad Y_{k+1} = 0.5(Y_k + (X_k' \backslash An)') \\ \quad X_k = X_{k+1} \\ \quad Y_k = Y_{k+1} \\ \text{end} \\ \sqrt{A} = \sqrt{(\|A\|_F)} Y_k \end{array} \right\} \quad (MSNARC)$$

$$\left. \begin{array}{l} \text{dado } X_0 = I \\ X_{k+1} = 0.5(X_k + X_k \backslash An) \\ \sqrt{A} = \sqrt{(\|A\|_F)} X_k \end{array} \right\}, \quad k = 0, 1, 2, \dots, \quad (IASMN)$$

$$\left. \begin{array}{l} \text{dado } B_0 = An, C_0 = I, X_0 = I \\ X_{k+1} = 0.5(X_k + B_k(X_k \backslash C_k)) \\ \quad C_k = X_k \\ B_{k+1} = 2X_{k+1} - X_k \\ \quad C_{k+1} = C_k \\ \sqrt{A} = \sqrt{(\|A\|_F)} X_k \end{array} \right\}, \quad k = 0, 1, 2, \dots, (MSNF SRC)$$

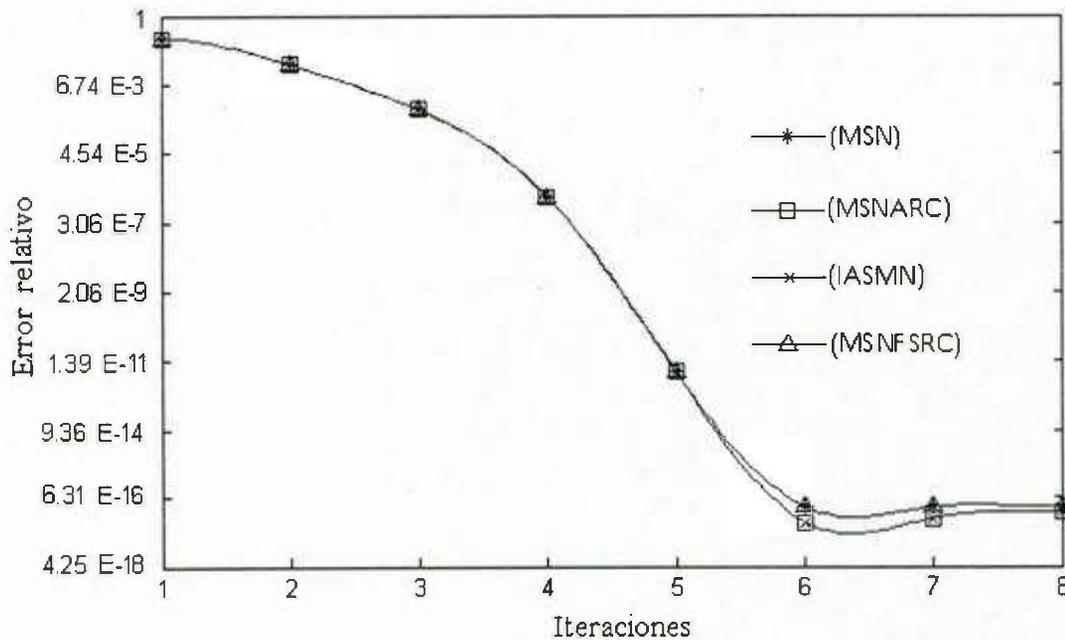
Experimento No. 1. Se calculará la raíz cuadrada de la matriz de Lehmer simétrica definida positiva de tamaño  $n = 3 \times 3$ .

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

$A$  está bien condicionada, con  $k_2(A) = \|A\|_F \|A^{-1}\|_F = 6.6633$  y valores propios (0.3020, 0.6855, 2.0124).

Tabla 5.1. Comportamiento de la convergencia

Método	Iterac.	Residual
(MSN)	6	9.55 E-017
(MSNARC)	6	9.94 E-017
(IASMN)	6	9.94 E-017
(MSNFSRC)	6	3.11 E-016



Gráfica 5.1. Comportamiento de la convergencia de los algoritmos

En la gráfica 5.1 se observa que todos los métodos proporcionan excelentes resultados, lo anterior debido a que la matriz está muy bien condicionada. Todos los métodos presentan un comportamiento muy similar en cuanto a la manera de converger, y lo hacen hasta la sexta iteración. El (MSN) es numéricamente estable debido a que el número de condición de la matriz de Lehmer es menor que nueve.

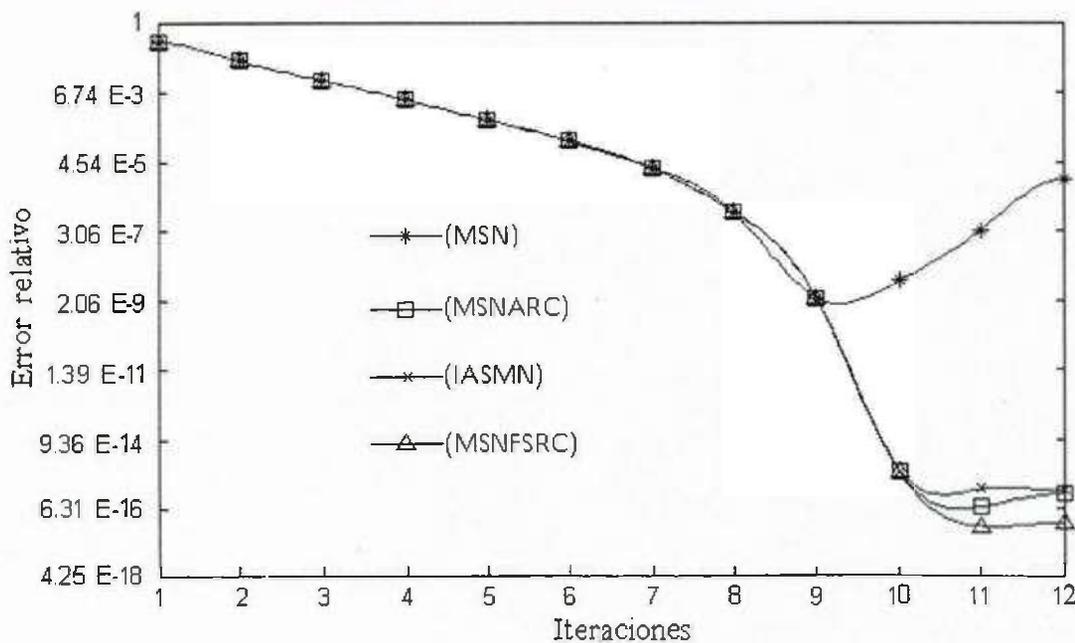
Experimento No. 2. Para ilustrar la inestabilidad numérica del (MSN) en la vecindad de la solución, se considera la matriz  $S$  de tamaño  $n = 3 \times 3$ , se calculará la raíz cuadrada de  $A = S^5$ , cuyos valores propios son  $(0.00215549, 1 \text{ y } 14.49784450)$ , la matriz  $A$  está cerca

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

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
(MSN)	9	2.29 E-009
(MSNARC)	10	8.81 E-015
(IASMN)	10	9.53 E-015
(MSNF SRC)	10	8.73 E-015



Gráfica 5.2. Comportamiento de la convergencia de los algoritmos

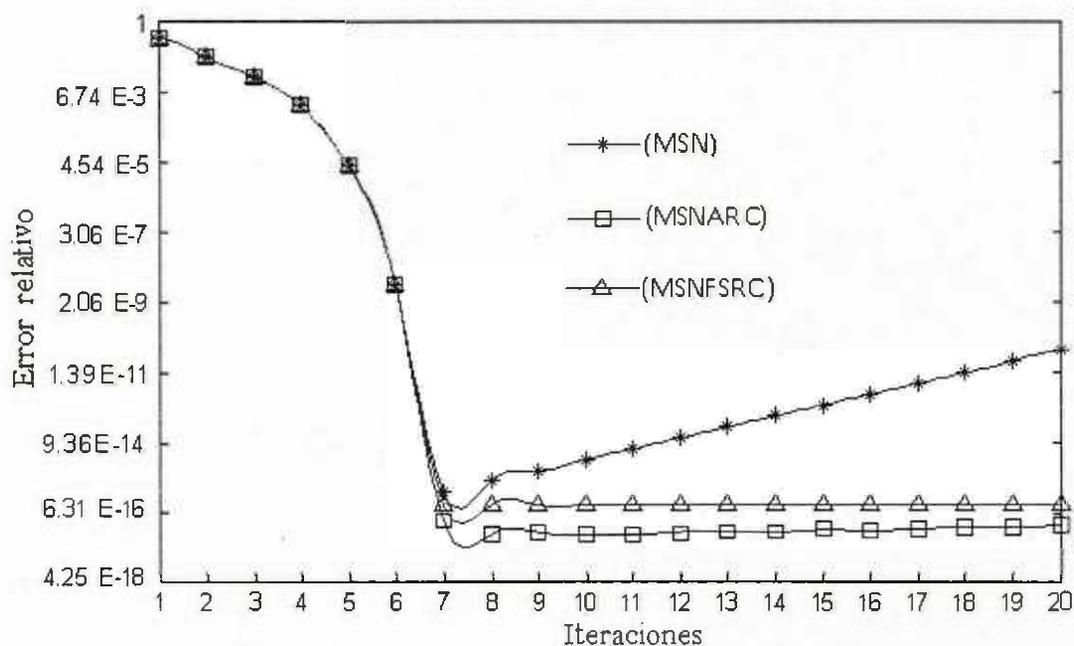
En la gráfica 5.2 se observa que el (MSN) diverge a partir de la novena iteración este comportamiento es debido a que el número de condición de la matriz es mucho mayor que nueve. Se observa que, al inicio, aparentemente los algoritmos tienen una convergencia del tipo lineal, pero esto se debe a la escala logarítmica utilizada en el gráfico, en realidad todas las iteraciones que son numéricamente estables tienen una convergencia del tipo cuadrática. Los algoritmos numéricamente estables convergen a partir de la décima iteración. 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 cuadrada de  $A$ .

$$A = \begin{bmatrix} 20 + 1Si & 1 + i & 1 - i \\ 2 + Si & 2 + 5i & 1 - i \\ 1 - i & -2 + 3i & 1 + i \end{bmatrix}$$

Tabla 5.3. Comportamiento de la convergencia

Método	Iterac.	Residual
(MSN)	7	2.80 E-015
(MSNARC)	7	3.40 E-016
(MSNF SRC)	7	1.12 E-015



Gráfica 5.3. Comportamiento de la convergencia de los algoritmos

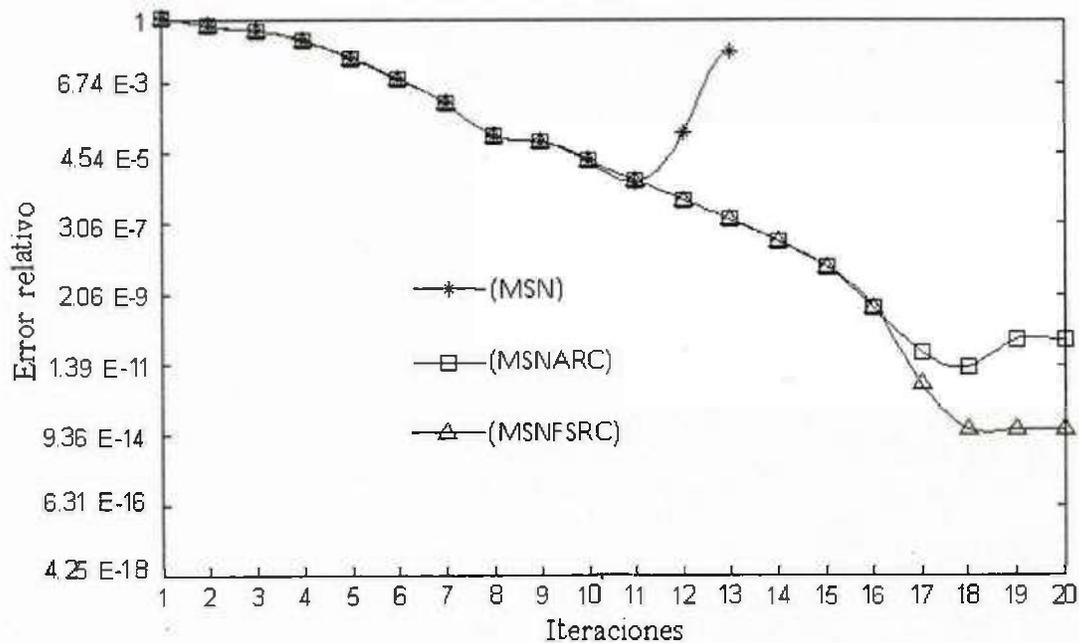
En la gráfica 5.3 se observa que el (MSN) diverge a partir de la octava iteración. El comportamiento de la convergencia de los demás algoritmos es muy similar y convergen en ocho iteraciones. Todos los algoritmos que son numéricamente estables proporcionan excelentes resultados.

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 cuadrada de  $A$ .

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

Tabla 5.4. Comportamiento de la convergencia

Método	Iterac.	Residual
(MSN)	11	5.56 E -006
(MSNARC)	17	3.50 E -011
(MSNF SRC)	18	1.36 E -013



Gráfica 5.4. Comportamiento de la convergencia de los algoritmos

En la gráfica 5.4 se observa que el (MSN) diverge a partir de la onceava iteración. El comportamiento de la convergencia de los algoritmos numéricamente estables es muy similar y convergen en 18 iteraciones. En general los algoritmos que son numéricamente estables proporcionan buenos resultados.

## BIBLIOGRAFÍA

- [1] Ayres, F. (1992). *Matrices*. México: MacGraw-Hill.
- [2] Barbolla, R., & Sanz P. (1998). *Algebra lineal y teoría de matrices*. Madrid: PRENTICE HALL.
- [3] Bartels R. H. and Stewart G. W. (1972). "Algorithm 432, Solution of the Matrix Equation  $AX+XB=C$ ", *Communications of the ACM*, Vol. 15, N° 9, pp. 820-913
- [4] Björck Å., Hammarling S. (1983) A Schur method for the square root of a matrix. *Linear Algebra Appl.* 52:127-140
- [5] Cayley A. (1858). A memoir on the theory of matrices. *Philos. Trans. Roy. Soc. London*, 148:17-37
- [6] Cayley A. (1872). On the extraction of the square root of a matrix of the third order. *Proc. Roy. Soc. Edinburgh*, 7:675-682
- [7] Demmel J. W. (1997). *Applied Numerical Linear Algebra*, SIAM, Philadelphia, PA
- [8] Denman E. and Alex N. Beavers Jr. (1976). The matrix sign function and computations in systems. *Appl. Math. Comput.*, 2:63-94
- [9] Golub G. H. Van Loan C. F. (1983). *Matrix computations*. Johns Hopkins Univ. Press, Baltimore, Maryland
- [10] Golub, Gene H., and Charles F. Van Loan. (2012). *Matrix computations*. Vol. 3. JHU Press
- [11] He J.H. (2000). "Improvement of Newton iteration method", *International journal of nonlinear sciences and numerical simulation*, Vol. 1, pp. 239-240
- [12] He J.H. (2003). "A new iteration method for solving algebraic equations", *Appl. Math. Comput.*, Vol. 135, pp. 81-84
- [13] Henrici P. (1964). "Elements of Numerical Analysis, Wiley", New York
- [14] Higham N. J. (1987). "Computing Real Square Roots of a Real Matrix". *Linear Algebra and Appl.* 88/89, pp. 405-430
- [15] Higham N. J. (1997). "Stable iterations for the matrix square root". *Numerical algorithms*, Vol. 15, pp. 227-242
- [16] Higham N. J. (1989). Computing the polar decomposition-with applications. *SIAM J. Sci. Statist. Comput.* 7:1160-1174
- [17] Higham N.J. (1986). Newton's method for the matrix square root. *Math. Comp.* 46:537-549

- [18] Higham N. J. (2008). *Functions of Matrices: Theory and Computation*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA
- [19] Horn R. A. and C. R. Johnson. (1991). "Topics in matrix analysis", Cambridge university press, Cambridge
- [20] Hoskins W. D., Walton D. J. (1979). A faster, more stable method for computing the  $p$ th roots of positive definite matrices. *Linear Algebra Appl.* 26:139-163
- [21] Hoskins W. D., Walton D. J. (1978). A faster method of computing the square root of a matrix. *IEEE Trans. Automat. Control.* AC-23:494-495
- [22] Iannazzo B. (2003). A note on computing the matrix square root. *Calcolo.* 40:273-283
- [23] Kolman B. (2006). "Algebra lineal", Octava Edición, Pearson, México
- [24] Kolman B. (2005). *Introductory linear algebra*. Prentice-Hall
- [25] Laasonen P. (1958). On the iterative solution of the matrix equation  $AX^2 - I = 0$ . *M.T.A.C.* 12:109-116
- [26] Lebedev, L. P., Vorovich, I. I., & Cloud, M. J. (2012). *Functional analysis in mechanics*. Springer Science & Business Media.
- [27] Long J.H., Hu X.Y., Zhang L. (2008). Newton's method with exact line search for the square root of a matrix. *J. Phys., Conf. Ser.* 96:012034
- [28] Meini B. (2004). The matrix square root from a new functional perspective: Theoretical results and computational issues. *SIAM J. Matrix Anal. Appl.* 26:362-376
- [29] Mendoza A., Gómez O. R. (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.* 26:47-53
- [30] Ortega J. M. (1972). *Numerical Analysis: A Second Course*, Academic Press, New York
- [31] Schatzman M. (2002). *Numerical analysis: A mathematical introduction*. Clarendon Press/Oxford University Press, Oxford/New York
- [32] Selberherr S. (1984) "Analysis and simulation of semiconductor devices", Wien, Springer-Verlag, New York
- [33] Stewart G. W. (1973) *Introduction to Matrix Computations*. Academic Press, New York