

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE INGENIERIA Y ARQUITECTURA

DEPARTAMENTO DE MATEMATICA

RESOLUCION DE SISTEMAS DE ECUACIONES
LINEALES MEDIANTE METODOS DIRECTOS

TRABAJO DE GRADUACION PRESENTADO POR:

SARA CECILIA CASTELLANOS CASTILLO

JOSE FRANCISCO RIVERA ZAVALETA

PARA OPTAR EL TITULO DE

LICENCIADO EN MATEMATICA

MAYO DE 1985

SAN SALVADOR, EL SALVADOR, CENTRO AMERICA



T
512.53
C348r

UNIVERSIDAD DE EL SALVADOR

RECTOR : DR. MIGUEL ANGEL PARADA

SECRETARIO GENERAL: DRA. ANA GLORIA CASTANEDA DE MONTOYA

FACULTAD DE INGENIERIA Y ARQUITECTURA

DECANO : ING. MANUEL ANTONIO CAÑAS LAZO

SECRETARIO: ING. RENE MAURICIO MEJIA MENDEZ

DEPARTAMENTO DE MATEMATICA

JEFE DEL DEPARTAMENTO: LIC. JOSE JAVIER RIVERA LAZO

ASESOR: ING. GABRIEL MELENDEZ MAYORGA



P R O L O G O

Nuestro propósito al elaborar el presente trabajo de investigación, es dar lineamientos que ayuden a escoger acertadamente un método directo que resuelva un sistema de ecuaciones lineales dado; dicho sistema no debe tener muchas ecuaciones, excepto que tenga una configuración especial ; si este fuera el caso que se presentara, se recomienda el empleo de los Métodos Iterativos.

En el capítulo I, se recuerdan algunos conocimientos básicos del Algebra Matricial que facilitan la comprensión de lo expuesto en este trabajo; así mismo tratamos de familiarizar al lector con la notación que emplearemos, también conoceremos las diferentes situaciones que se pueden presentar al querer resolver un sistema de ecuaciones lineales; y además se definen conceptos que serán de mucha utilidad en el desarrollo de los siguientes capítulos.

En el capítulo II se describen: *el Método de Eliminación Gaussiana* y *el Método de Gauss-Jordan*; que son Métodos Directos que resuelven un sistema de ecuaciones lineales. Para dichos métodos se dan diversos algoritmos que se aplicarán tomando en cuenta la naturaleza del sistema planteado.

Para el Método de Eliminación Gaussiana empleamos las estrategias de pivotamiento completo y parcial para Sistemas de Ecuaciones Lineales que se representan por medio de matrices de orden n . Finalmente se establece una comparación entre el Método de Eliminación Gaussiana con sustitución hacia atrás y

el Método de Gauss-Jordan.

En el capítulo III se exponen otros métodos directos, que se basan en la factorización de matrices; dichos métodos se emplean para resolver Sistemas de Ecuaciones Lineales que involucren una matriz de orden n . Estos métodos son: *Doolittle*, *Crout* y *Cholesky*. Para la reducción de Crout se ha empleado pivoteamiento parcial. En este capítulo se explica, cómo los algoritmos de factorización pueden ser simplificados para el caso de Sistemas Tridiagonales de ecuaciones.

En el capítulo IV se analiza la estabilidad de los algoritmos; para ello comenzamos definiendo los conceptos de normas de vectores y normas de matrices que servirán para el estudio de la estimación del error en el cálculo de la solución de un Sistema de Ecuaciones Lineales; obtenida la solución por las técnicas estudiadas anteriormente, en este capítulo damos a conocer una técnica de refinamiento de soluciones; la cuál consiste en buscar una solución que se aproxime a la solución verdadera.

Al final de la mayoría de las secciones de cada capítulo se encuentran varios ejercicios adicionales, los cuáles constituyen ampliaciones a la teoría incluida en este trabajo.

INDICE

Página

Prólogo

CAPITULO I

FUNDAMENTACION TEORICA

1. Espacios vectoriales	1
2. Independencia lineal, subespacio y bases.	7
3. Tipos de matrices	13
4. Operaciones con matrices.	25
5. Propiedades de las matrices	33
6. Transformaciones lineales y matrices.	41
7. Ecuaciones lineales	46
8. Errores, aritmética y estabilidad	54

CAPITULO II

METODOS DIRECTOS PARA RESOLVER SISTEMAS DE ECUACIONES LINEALES (primera parte)

1. Matrices triangulares y sistemas.	69
2. Eliminación Gaussiana	79
3. Método de Gauss-Jordan	115

CAPITULO III

METODOS DIRECTOS PARA RESOLVER SISTEMAS DE ECUACIONES LINEALES (segunda parte)

1. Factorización de matrices	137
2. Resolución de sistemas de ecuaciones lineales utilizando la factorización	159

CAPITULO IV

ESTABILIDAD

1. Normas de vectores.	168
2. Normas de matrices.	182
3. Estimación de error, condicionamiento de una matriz y <u>re</u> finamiento iterativo.	194

BIBLIOGRAFIA

LISTA DE ALGUNOS SIMBOLOS USADOS

a, b, \dots : escalares

E, K, \dots : conjuntos

\bar{x}, \bar{y}, \dots : vectores columna

A, B, \dots : matrices

a_{ij} : elementos de la matriz A

\bar{A}, \bar{B}, \dots : espacios o subespacios

\oplus : suma directa de subespacios

B_{ij} : submatriz de B

$f_l(e)$: la expresión e evaluada en aritmética de --
punto flotante.

$=$: operador para reemplazamiento

\leftarrow : operador para sobre-escribir

\longleftrightarrow : operador para intercambios

$\kappa(A)$: número de condicionamiento de la matriz A .

CAPITULO I

FUNDAMENTACION TEORICA

1. ESPACIOS VECTORIALES

DEFINICION 1.1

Sea E un conjunto y K un cuerpo conmutativo, decimos que sobre E se define la estructura de *Espacio Vectorial* si se tiene definidas dos operaciones, una interna

$$+ : E \times E \longrightarrow E$$

$$(x, y) \rightsquigarrow + (x, y) = x + y$$

tal que $(E, +)$ es un grupo conmutativo; y otra operacion externa

$$\cdot : K \times E \longrightarrow E$$

$$(a, x) \rightsquigarrow \cdot (a, x) = ax$$

tal que, si x é y son elementos cualesquiera de E y $a, b \in K$, entonces

- i) $1 \cdot x = x$ (1 : identidad multiplicativa del cuerpo)
- ii) $a(b \cdot x) = (ab) \cdot x$
- iii) $(a+b) \cdot x = ax + bx$
- iv) $a \cdot (x+y) = ax + ay$

EJEMPLO 1.2 (algunos espacios vectoriales).

1- \mathbb{R} es un espacio vectorial sobre si mismo.

2- Todo cuerpo conmutativo K es un espacio vectorial sobre si mismo.

3- Si K es un cuerpo conmutativo, el conjunto K , donde

$K^n = \{X/X = (x_1, x_2, \dots, x_n), x_i \in K\}$, forman un espacio vectorial con las operaciones definidas por:

Sean $X, Y \in K^n$. Entonces

$$X + Y = (x_1+y_1, x_2+y_2, \dots, x_n+y_n)$$

$$aX = (ax_1, ax_2, \dots, ax_n), a \in K$$

En particular \mathbb{R}^n es un espacio vectorial sobre \mathbb{R} .

En efecto:

I $(\mathbb{R}^n, +)$ cumple ser un grupo conmutativo.

$$\begin{aligned} \text{i) } X + Y &= (x_1, x_2, x_3, \dots, x_n) + (y_1, y_2, \dots, y_n) \\ &= (x_1+y_1, x_2+y_2, \dots, x_n+y_n) \end{aligned}$$

ii) Las propiedades asociativas y conmutativas se cumplen, ya que los componentes de los elementos de \mathbb{R}^n son elementos de \mathbb{R} .

iii) El elemento neutro en \mathbb{R}^n respecto a la suma es aquel - cuyos n componentes son ceros.

$$X + 0 = (x_1, x_2, \dots, x_n) + (0, 0, \dots, 0) = (x_1, x_2, \dots, x_n) = X$$

iv) $\forall X = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ existe $-X = (-x_1, -x_2, \dots, -x_n)$ tal que $X + (-X) = 0$, $0 \in \mathbb{R}^n$.

II $\forall X, Y \in \mathbb{R}^n$, $a, b \in \mathbb{R}$ se cumple:

i) $\exists 1 \in \mathbb{R}$ tal que $1 \cdot X = X$, $\forall X \in \mathbb{R}^n$

$$\begin{aligned} 1 \cdot X &= 1(x_1, x_2, \dots, x_n) = (1 \cdot x_1, 1 \cdot x_2, \dots, 1 \cdot x_n) \\ &= (x_1, x_2, \dots, x_n) = X. \end{aligned}$$

Las otras propiedades tambien son fáciles de comprobar.

La demostración se basa en el hecho de que las componen-

tes de los elementos de \mathbb{R}^n pertenecen a \mathbb{R} los cuales cumplen la asociatividad y distributividad con la operación multiplicación.

$\therefore \mathbb{R}^n$ es un espacio vectorial sobre \mathbb{R} .

DEFINICION 1.3

A los elementos de un espacio vectorial se le llaman *vectores*. En este trabajo adoptaremos la siguiente notación; hablaremos de n-uplas ordenadas como vectores de \mathbb{R}^n pero arreglados en columna, así:

$$\bar{x} \in \mathbb{R}^n \Rightarrow \bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

\bar{x}^t nos denotará el vector fila, $\bar{x}^t = [x_1, x_2, \dots, x_n]$

DEFINICION 1.4

Sea K un cuerpo arbitrario. Una ordenación rectangular de la forma

$$\begin{bmatrix} a_{11} & a_{12} \dots a_{1n} \\ a_{21} & a_{22} \dots a_{2n} \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ \cdot & \cdot \quad \cdot \\ a_{m1} & a_{m2} \dots a_{mn} \end{bmatrix}$$

donde a_{ij} son elementos de K ; se llama una *Matriz* sobre K ó simplemente una *Matriz*, si K está implícito.

La matriz anterior se denota por $[a_{ij}]$, $i = 1, 2, \dots, m$;

$j = 1, 2, \dots, n$; ó simplemente por $[a_{ij}]_{m \times n}$

Las m n -uplas horizontales $[a_{11}, a_{12}, \dots, a_{1n}], [a_{21}, a_{22}, \dots, a_{2n}], \dots$
 $\dots [a_{m1}, a_{m2}, \dots, a_{mn}]$, son las filas de la matriz y las n m -uplas
 verticales

$$\begin{bmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ a_{m1} \end{bmatrix}, \begin{bmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ a_{m2} \end{bmatrix}, \dots, \begin{bmatrix} a_{1n} \\ a_{2n} \\ \cdot \\ \cdot \\ a_{mn} \end{bmatrix}, \text{ son sus columnas.}$$

Obsérvese que el elemento a_{ij} llamado la componente i - j ocupa -
 la i -ésima fila y la j -ésima columna. Una matriz con m filas y
 n columnas se llama *Matriz* m por n ó *Matriz* de orden $m \times n$.

Una matriz de orden $m \times n$, donde $m = n$ se llama *Matriz*
cuadrada y se dice que es de orden n .

EJEMPLO 1.5

La siguiente es una matriz 2×3

$$\begin{bmatrix} 1 & -4 & 5 \\ 0 & 3 & -2 \end{bmatrix}$$

DEFINICION 1.6

Dos matrices A y B son *iguales*, esto es $A = B$, si
 tienen la misma forma y sus elementos correspondientes son igua-
 les. Por tanto, la igualdad de dos matrices $m \times n$ es equivalen-
 te a un sistema de $m \cdot n$ igualdades, una para cada par de elemen-
 tos.

EJEMPLO 1.7

Es fácil verificar que el conjunto M formado por las matrices de orden $m \times n$ definidas sobre \mathbb{R} forman un espacio vectorial con las operaciones:

Sean $A, B \in M$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \text{ y } B = \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ b_{m1} & b_{m2} & \dots & b_{mn} \end{bmatrix}$$

$$A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \dots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \dots & a_{2n} + b_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \dots & a_{mn} + b_{mn} \end{bmatrix}$$

$$cA = \begin{bmatrix} ca_{11} & ca_{12} & \dots & ca_{1n} \\ ca_{21} & ca_{22} & \dots & ca_{2n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ ca_{m1} & ca_{m2} & \dots & ca_{mn} \end{bmatrix}$$

EJERCICIOS

1.1 Mostrar que el conjunto de todas las funciones reales

$f: \mathbb{R} \rightarrow \mathbb{R}$, definidas para $0 \leq x \leq 1$, y tales que $f(1/2) = 0$, forman un espacio vectorial sobre los reales, considerándose válidas las definiciones de adición y multiplicación escalar como sigue:

$$(f + g)(x) = f(x) + g(x)$$

$$(af)(x) = af(x), \quad \forall x \in \mathbb{R}$$

1.2 Mostrar que el conjunto de los vectores de coordenadas en \mathbb{R}^3 , cuya primera coordenada es la unidad, no es un espacio vectorial.

1.3 Verificar que el conjunto de las ternas de números reales $[x_1, x_2, x_3]^t$ en el cual se han definido la adición y la multiplicación por un escalar como:

$$[x_1, x_2, x_3]^t + [y_1, y_2, y_3]^t = [x_1 + y_1, x_2 + y_2, x_3 + y_3]^t \quad \text{y}$$

$$a[x_1, x_2, x_3]^t = [0, 0, 0]^t, \quad \text{en donde } a \in \mathbb{R},$$

no es un espacio vectorial.

1.4 El conjunto de las parejas de números reales $[x_1, x_2]^t$, definida la adición por: $[x_1, x_2]^t + [y_1, y_2]^t = [x_1 + y_1, x_2 + y_2]^t$, y la multiplicación por un escalar como:

$$a[x_1, x_2]^t = [a^2 x_1, a^2 x_2]^t, \quad \text{en donde } a \in \mathbb{R}. \text{ Muestre que dicho conjunto no satisface todas las propiedades de un espacio vectorial.}$$

1.5 Mostrar que el conjunto de las parejas de números reales

$[x_1, x_2]^t$ en donde se ha definido la adición y la multiplicación por un escalar como:

$$[x_1, x_2]^t + [y_1, y_2]^t = [x_1 + y_1, x_2 + y_2]^t$$

$$b[x_1, x_2]^t = [5bx_1, 5bx_2] ,$$

no es un espacio vectorial.

1.6 Si V es un espacio vectorial sobre los reales, v pertenece a V y $v + v = 0$, muestre que $v = 0$.

1.7 Si V es un espacio vectorial, v pertenece a V y a, b son escalares, muestre que $av = bv$ y $a \neq b$, entonces $v = 0$

1.8 Considere el conjunto de todas las matrices $n \times n$ que tienen todos los elementos $a_{ij} = 0$, cuando $i = j$.

Muestre que satisface todas las propiedades de un espacio vectorial.

2. INDEPENDENCIA LINEAL, SUBESPACIO Y BASES

DEFINICION 2.1

Sean $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m \in \mathbb{R}^n$ y $a_1, a_2, \dots, a_m \in \mathbb{R}$.

Entonces el vector $B = a_1\bar{x}_1 + a_2\bar{x}_2 + \dots + a_m\bar{x}_m$ es llamado *combinación lineal* de $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$. Los números a_1, a_2, \dots, a_m son llamados coeficientes de la combinación lineal.

Si $a_1 = a_2 = \dots = a_m = 0$, la combinación se dice que es *trivial*, en caso contrario se dice que es *no trivial*.

EJEMPLO 2.2

Si \bar{e}_i ($i = 1, 2, \dots, n$), denota el n -vector cuya i -ésima

ma componente es la unidad y las otras componentes son ceros. -
Entonces para cualquier vector $\bar{x} = [x_1, x_2, \dots, x_n]^t$ de \mathbb{R}^n se tie-
ne:

$$\bar{x} = x_1 \bar{e}_1 + x_2 \bar{e}_2 + \dots + x_n \bar{e}_n \quad (2.1)$$

En otras palabras cualquier n-vector es una combinación lineal
de $\bar{e}_1, \bar{e}_2, \bar{e}_3, \dots, \bar{e}_n$. Estos vectores continuaran apareciendo y re-
servaremos los símbolos \bar{e}_i ($i = 1, 2, \dots, n$) para referirnos a -
ellos.

DEFINICION 2.3

Sea $A \subset \mathbb{R}^n$. Los elementos de A son *linealmente in-*
dependientes (l.i) si $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_m \in A$ y

$$a_1 \bar{x}_1 + a_2 \bar{x}_2 + \dots + a_m \bar{x}_m = \bar{0}, \text{ entonces } a_1 = a_2 = \dots = a_m = 0.$$

En este caso se dice tambien que A es un conjunto linealmente -
independiente.

Un conjunto A de vectores que no es linealmente independiente -
se denomina *Linealmente dependiente (l.d)*

EJEMPLO 2.4

En \mathbb{R}^n supóngase que se tiene la siguiente combina---
ción lineal

$$\begin{aligned} & a_1 \bar{e}_1 + a_2 \bar{e}_2 + \dots + a_n \bar{e}_n = \bar{0} \\ & = a_1 [1, 0, \dots, 0]^t + a_2 [0, 1, \dots, 0]^t + \dots + a_n [0, 0, \dots, 0, 1]^t = \bar{0} \\ & = [a_1, 0, \dots, 0]^t + [0, a_2, \dots, 0]^t + \dots + [0, 0, \dots, a_n]^t = \bar{0} \\ & = [a_1, a_2, a_3, \dots, a_n]^t = \bar{0} \end{aligned}$$

entonces $a_1 = a_2 = a_3 = \dots = a_n = 0$

$\therefore \bar{e}$ ($i = 1, 2, \dots, n$) son *l.i.*

EJEMPLO 2.5

El conjunto de vectores:

$$\bar{x}_1 = \begin{bmatrix} 5 \\ 2 \\ 1 \end{bmatrix}, \quad \bar{x}_2 = \begin{bmatrix} -1 \\ 3 \\ 3 \end{bmatrix}, \quad \bar{x}_3 = \begin{bmatrix} 9 \\ 7 \\ 5 \end{bmatrix}, \quad \bar{x}_4 = \begin{bmatrix} 3 \\ 8 \\ 7 \end{bmatrix}$$

es linealmente dependiente puesto que podemos formar la combinación lineal no trivial igual a cero:

$$4\bar{x}_1 - \bar{x}_2 - 3\bar{x}_3 + 2\bar{x}_4 = \bar{0}$$

TEOREMA 2.6

Sean $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m \in \mathbb{R}^n$ linealmente independientes; su póngase que

$$\bar{y} = a_1\bar{x}_1 + a_2\bar{x}_2 + \dots + a_m\bar{x}_m \quad (2.2)$$

$$\text{y} \quad \bar{y} = b_1\bar{x}_1 + b_2\bar{x}_2 + \dots + b_m\bar{x}_m \quad (2.3)$$

$$\text{entonces} \quad a_i = b_i \quad (i = 1, 2, \dots, m) \quad (2.4)$$

PRUEBA

De (2.2) y (2.3) se sigue que

$$\bar{0} = \bar{y} - \bar{y} = (a_1 - b_1)\bar{x}_1 + (a_2 - b_2)\bar{x}_2 + \dots + (a_m - b_m)\bar{x}_m$$

Como los vectores $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$ son linealmente independientes, se tiene que $a_i - b_i = 0$ ($i = 1, 2, \dots, n$), lo cual es justamente (2.4).

Este teorema nos garantiza la unicidad de la representación de un vector mediante vectores linealmente independientes.

DEFINICION 2.7

Sea A un subconjunto no vacío de \mathbb{R}^n . Entonces A es un *subespacio* de \mathbb{R}^n si:

1. $\bar{x}, \bar{y} \in A \Rightarrow \bar{x} + \bar{y} \in A$
2. $\bar{x} \in A \wedge a \in \mathbb{R} \Rightarrow a\bar{x} \in A$

Se sigue inmediatamente de la definición que si A es un subespacio de \mathbb{R}^n , combinaciones lineales de elementos de A están también en A . Recíprocamente, cualquier subconjunto no vacío de \mathbb{R}^n que es cerrado bajo combinaciones lineales es un subespacio.

EJEMPLO 2.8

El espacio \mathbb{R}^n es subespacio de sí mismo.

El conjunto $\{0\}$ que es solamente el vector cero, es el subespacio trivial de \mathbb{R}^n .

DEFINICION 2.9

Sean $B \subset A \subset \mathbb{R}^n$, se dice que B genera a A si todo vector de A puede obtenerse como combinación lineal de los elementos de B .

DEFINICION 2.10

Sea $A \subset \mathbb{R}^n$ un subespacio y $B \subset A$.

Entonces B es una *base* de A si:

- 1- Los elementos de B son linealmente independientes
- 2- B genera a A .

EJEMPLO 2.11

Por el ejemplo 2.4, los vectores $\bar{e}_1, \bar{e}_2, \dots, \bar{e}_n \in \mathbb{R}^n$ son linealmente independientes; por el ejemplo 2.2, ellos generan \mathbb{R}^n . Por tanto el conjunto $\{\bar{e}_1, \bar{e}_2, \bar{e}_3, \dots, \bar{e}_n\}$ es una base para \mathbb{R}^n .

TEOREMA 2.12

Cualquier base para \mathbb{R}^n debe tener exactamente n elementos.

DEFINICION 2.13

Si A es un subespacio no trivial de \mathbb{R}^n , la *dimensión* de A , denotada por $\dim(A)$ es el número de elementos de una base de A . La dimensión del subespacio trivial es cero.

Del teorema 2.12 y la definición 2.13 se tiene que $\dim(\mathbb{R}^n) = n$. En particular $\dim(\mathbb{R}^3) = 3$.

DEFINICION 2.14

Sea $A \subset \mathbb{R}^n$ un subespacio y sean U y V subespacios de A . Entonces U y V son llamados *complementos* en A si:

- i) $U \cap V = \{0\}$
- ii) A es generado por $U \cup V$

Escribimos:

$$A = U \oplus V$$

TEOREMA 2.15

Sea A un subespacio de \mathbb{R}^n .

1. si $U \subset A$ es un subespacio, entonces U tiene un complemento - en A .
2. si $A = U \oplus V$ y $\{u_1, u_2, \dots, u_k\}$ y $\{v_1, v_2, \dots, v_\ell\}$ son bases para U y V , entonces $\{u_1, \dots, u_k, v_1, \dots, v_\ell\}$ es una base para A .
Consecuentemente $\dim(A) = \dim(U) + \dim(V)$
3. si $A = U \oplus V$ y $x \in A$, entonces x puede ser escrito univocamente en la forma $x = u + v$, donde $u \in U$ y $v \in V$.

▲

EJERCICIOS

- 2.1 Sea P el subconjunto de todos los elementos de \mathbb{R}^3 , cuya tercera coordenada es el número cero. Verificar si este conjunto P es un subespacio de \mathbb{R}^3 .
- 2.2 Probar que en \mathbb{R}^n
 $c\bar{x} = \bar{0} \Rightarrow c = 0 \text{ ó } \bar{x} = \bar{0}$, con $c \in \mathbb{R}$.
- 2.3 Probar que cualquier conjunto que contenga el vector cero - es linealmente dependiente.
- 2.4 Verificar que los vectores $[2,3]^t$, $[3,5]^t$ son linealmente independientes.
- 2.5 Verificar que los vectores $[2,3]^t$, $[3,5]^t$, $[1,1]^t$, son linealmente dependientes.
- 2.6 Sean A y B subespacios de \mathbb{R}^n . Pruebe que si $A \subset B$ y $\dim(A) = \dim(B)$, entonces $A = B$

3. TIPOS DE MATRICES

DEFINICION 3.1

Sea A una matriz $m \times n$ y sea $k = \min\{m, n\}$.

Los elementos a_{ii} ($i = 1, 2, \dots, k$) se dice que estan en la diagonal de A ; a_{ii} es llamado el i -ésimo elemento de la diagonal de A .

Los elementos $a_{i, i+1}$ se dice que estan en la *su-*
perdiagonal de A . Los elementos $a_{i, i-1}$ en la *subdiagonal* de A

Si A es cuadrada, los elementos $a_{n-i+1, i}$ ($i = 1, 2, \dots, n$), se dice que estan en la diagonal secundaria de A .

DEFINICION 3.2

Una matriz A de orden n , se dice que es *diagonal* si $a_{ij} = 0$, $\forall i \neq j$. Es decir

$$\begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}$$

Notación que emplearemos: $A = \text{diag} [a_1, a_2, \dots, a_n]$

DEFINICION 3.3

Una matriz A de orden $m \times n$ es *Trapezoidal Supe-*
rior si

$$i > j \Rightarrow a_{ij} = 0$$

es *Trapezoidal Inferior* si

$$i < j \Rightarrow a_{ij} = 0$$

Una matriz cuadrada que sea trapezoidal superior (inferior) se llama *Triangular Superior* (*Inferior*).

OBSERVACION 3.4

Algunos tipos especiales de matrices triangulares son:

- a) Si T es triangular superior (inferior) con elementos ceros en la diagonal, entonces T es llamada *triangular estrictamente superior* (*inferior*).

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

Triangular estrictamente

Inferior

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

Triangular estrictamente

Superior

- b) Si los elementos de la diagonal de T son unos, T es llamada *triangular unitaria superior* (*inferior*).

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

Triangular unitaria

Inferior

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

Triangular unitaria

Superior.

EJEMPLO 3.5

Una matriz 3 x 5 trapezoidal superior tiene la forma:

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

DEFINICION 3.6

Una matriz cuadrada A es *Hessenberg Superior* si

$$i > j + 1 \Rightarrow a_{ij} = 0$$

es *Hessenberg inferior* si

$$i < j - 1 \Rightarrow a_{ij} = 0$$

es *tridiagonal* si al mismo tiempo es *Hessenberg superior* é *inferior*.

EJEMPLO 3.7

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

Hessenberg Superior

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

Tridiagonal

DEFINICION 3.8

Una matriz de orden n es llamada *Matriz Banda* si existen p y q; $1 < p, q < n$ con la propiedad que $a_{ij} = 0$ cuando $i + p \leq j$ ó $j + q \leq i$. El ancho de la banda para una matriz -

de este tipo está definida así:

$$w = p + q - 1$$

EJEMPLO 3.9

Una matriz banda de orden 8, con $p = 2$ y $q = 3$, tiene la forma siguiente:

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

$$w = p + q - 1$$

$$w = 2 + 3 - 1$$

$$\Rightarrow w = 4$$

EJERCICIO 3.10

Mostrar que si A es una matriz banda con $p = q = 2$, entonces es simultaneamente Hessenberg Superior y Hessenberg Inferior, además tambien es tridiagonal.

PRUEBA

Como A es una matriz banda con $p = q = 2$, debe cumplirse - que:

$$a_{ij} = 0 \text{ cuando } i+2 \leq j \quad \text{ó} \quad a_{ij} = 0 \text{ cuando } j+2 \leq i \quad (\text{def. 3.8})$$

$$\Leftrightarrow a_{ij} = 0 \text{ cuando } i \leq j-2 \quad \text{ó} \quad a_{ij} = 0 \text{ cuando } i \leq j+2$$

$$\Leftrightarrow a_{ij} = 0 \text{ cuando } i < j-1 \quad \text{ó} \quad a_{ij} = 0 \text{ cuando } i > j+1$$

Se tiene así que la matriz A es Hessenberg superior y - Hessenberg inferior.

Por lo tanto la matriz A es tridiagonal. (def. 3.6)

DEFINICION 3.11

Cualquier matriz donde todos sus elementos son - ceros es llamada *Matriz Cero* (o *Matriz Nula*) y la denotaremos - por el símbolo 0.

DEFINICION 3.12

Una matriz A de orden n es *Simétrica* si

$$a_{ij} = a_{ji} \quad (i, j = 1, 2, \dots, n).$$

OBSERVACION 2.13

La denominación "simétrica" proviene del hecho de que los elementos de A estan dispuestos simetricamente con - relación a la diagonal principal. Así:

$$A = \begin{bmatrix} 5 & 0 & 3 \\ 0 & 1 & -1 \\ 3 & -1 & 4 \end{bmatrix}$$

OBSERVACION 3.14

Dada una matriz A y supóngase que ciertas filas

y columnas de A han sido seleccionadas. El arreglo rectangular de elementos de A que están situados en la intersección de estas filas y columnas es también una matriz y es llamada *submatriz* de A . Por ejemplo, el arreglo

$$S = \begin{bmatrix} a_{22} & a_{23} & a_{25} \\ a_{42} & a_{43} & a_{45} \end{bmatrix}$$

es una submatriz obtenida de A por selección de las filas 2 y 4, y las columnas 2, 3 y 5 como sigue

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ -a_{21}- & a_{22}- & a_{23}- & a_{24}- & a_{25}- \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ -a_{41}- & a_{42}- & a_{43}- & a_{44}- & a_{45}- \end{bmatrix}$$

Los elementos que se encuentran en la intersección de las líneas punteadas componen los elementos de S . La siguiente es una definición más precisa de la idea de una submatriz.

DEFINICION 3.15

Sea A una matriz $m \times n$ y sea $1 < i_1 < i_2 \dots < i_k < m$ y $1 < j_1 < j_2 \dots < j_\ell < n$. La matriz S de orden $k \times \ell$ cuyo (u,v) -elemento es

$$s_{uv} = a_{i_u, j_v}$$

es llamada *Submatriz* de A . Si $k = \ell$ y $i_1 = j_1, i_2 = j_2, \dots, i_k = j_k$, entonces S es llamada *submatriz principal* de A .

Si $i_1 = 1, i_2 = 2, \dots, i_k = k$ y $j_1 = 1, j_2 = 2, \dots, j_\ell = \ell$, entonces S es llamada *submatriz directriz* de A . Consecuentemente, una submatriz principal es una submatriz cuadrada en la cual fi las y columnas son iguales. Por ejemplo, la matriz

$$\begin{bmatrix} a_{22} & a_{24} \\ a_{42} & a_{44} \end{bmatrix}$$

es una submatriz principal 2×2 de A .

DEFINICION 3.16

Una *matriz directriz* de A , es una submatriz que se halla en la esquina superior izquierda de A , es decir

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}$$

OBSERVACION 3.17

Las submatrices directrices principales de una matriz ocurren frecuentemente y por lo cual requieren de su propia notación. Denotaremos las submatrices directrices principales de A de orden k así $A^{[k]}$. Por ejemplo

$$A^{[3]} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

DEFINICION 3.18

La matriz de orden n .

$$I_n = \text{diag}[1, 1, 1, \dots, 1]$$

es llamada *Matriz Identidad* (o *Matriz Unidad*) de orden n .

Note que $I_n = [\bar{e}_1, \bar{e}_2, \dots, \bar{e}_n]$, donde los vectores \bar{e}_i constituyen la base canónica de \mathbb{R}^n ; \bar{e}_i es la i -ésima columna de la matriz identidad.

Usaremos I en lugar de I_n para referirnos a la matriz identidad.

La importante propiedad de la matriz identidad es que

$$IA = AI = A$$

para cualquier matriz A siempre que la multiplicación sea posible.

DEFINICION 3.19

Sea A una matriz de orden $m \times n$. La *traspuesta* de A es la matriz B de orden $n \times m$ donde sus elementos están dados por

$$b_{ij} = a_{ji} \quad (i = 1, 2, \dots, n \quad ; \quad j = 1, 2, \dots, m)$$

Escribimos

$$B = A^t$$

EJEMPLO 3.20

Dado el vector columna

$$\begin{bmatrix} 1 \\ -2 \\ 4 \end{bmatrix}$$

puede ser escrito en una sola línea como: $[1, -2, 4]^t$.

TEOREMA 3.21

Sean las matrices A y B , y $a \in \mathbb{R}$. Entonces;

1- $(A)^t = A$

2- $(A + B)^t = A^t + B^t$

3- $(aA)^t = aA^t$

4- $(AB)^t = B^t A^t$

DEFINICION 3.22

Sea A una matriz cuadrada. La *Inversa* de A es -- una matriz cuadrada, denotada por A^{-1} que satisface la propie-- dad

$$AA^{-1} = A^{-1}A = I$$

Una matriz que tiene inversa es llamada *No singu*lar, en caso contrario es llamada *Singular*.

TEOREMA 3.23

Sea A no singular. Entonces A^{-1} es no singular y - $(A^{-1})^{-1} = A$.

Igualmente A^t es no singular y $(A^t)^{-1} = (A^{-1})^t$.

PRUEBA

La no singularidad para A^{-1} se da implícitamente. Probaremos que $(A^{-1})^{-1} = A$.

Por definición de A^{-1} se tiene inmediatamente que



$$AA^{-1} = A^{-1}A = I$$

Si se considera A^{-1} como la matriz dada, en lugar de A y como A es la inversa única de A^{-1} se tiene que

$$(A^{-1})^{-1} = A$$

Demostraremos ahora que $(A^t)^{-1} = (A^{-1})^t$.

Partimos de

$$AA^{-1} = A^{-1}A = I$$

Tomando la traspuesta y notando que $I^t = I$, se obtiene

$$(A^{-1})^t A^t = A^t (A^{-1})^t = I$$

Por tanto $(A^{-1})^t$ es la inversa de A^t . Entonces

$$(A^t)^{-1} = (A^{-1})^t.$$

A

TEOREMA 3.24

Sean A y B de orden n . Entonces AB es no singular si y solamente si A y B son no singulares y

$$(AB)^{-1} = B^{-1}A^{-1}$$

COROLARIO 3.25

Sea A de orden n . Entonces A^n es no singular si y solamente si A es no singular y

$$(A^n)^{-1} = (A^{-1})^n$$

NOTA:

La discusión de la inversa fue originada al tratar de re--

resolver la ecuación $A\bar{x} = \bar{b}$, que al multiplicarla por A^{-1} se obtiene $\bar{x} = A^{-1}\bar{b}$ (donde \bar{x} y \bar{b} representaran vectores - columna).

DEFINICION 3.26

La matriz A de orden n , es llamada *Diagonal estrictamente dominante* cuando

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{ij}|$$

para cada $i = 1, 2, \dots, n$.

EJEMPLO 3.27

Sea la matriz A de orden 3.

$$A = \begin{bmatrix} 7 & 2 & 0 \\ 3 & 5 & -1 \\ 0 & 5 & -6 \end{bmatrix} \quad A^t = \begin{bmatrix} 7 & 3 & 0 \\ 2 & 5 & 5 \\ 0 & -1 & -6 \end{bmatrix}$$

La matriz A es diagonal estrictamente dominante porque:

$$|7| > |2| + |0|$$

$$|5| > |3| + |-1|$$

$$|-6| > |0| + |5|$$

Es interesante notar que A^t no es diagonal estrictamente dominante.

EJERCICIOS

- 3.1 Compruebe que la traspuesta de una matriz triangular inferior es una matriz triangular superior.
- 3.2 Verificar que una matriz es diagonal si y solo si es a la vez triangular superior é inferior.
- 3.3 Pruebe que la traspuesta de una matriz simétrica escrita como A^t es igual a A . Es decir $A^t = A$.
- 3.4 Demostrar que si $A = [a_{ij}]$ es una matriz cuadrada, la matriz $B = [b_{ij}] = A + A^t$ es simétrica.
- 3.5 Dadas las matrices $n \times n$ simétricas A y B . Pruebe que $A + B$ es simétrica.
- 3.6 Demostrar que

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 5 & 7 \\ -2 & -4 & -5 \end{bmatrix} \text{ es la inversa de } \begin{bmatrix} 3 & -2 & -1 \\ -4 & 1 & -1 \\ 2 & 0 & 1 \end{bmatrix}$$

- 3.7 Demostrar que $(ABC)^t = C^t B^t A^t$
- 3.8 Demostrar que si una matriz cuadrada A de orden m es simétrica y que si P es de orden $m \times n$, $B = P^t A P$ es otra matriz simétrica.

4. OPERACIONES CON MATRICES

En el ejemplo 1.7 se definió la suma de matrices y la multiplicación de una matriz por un escalar. Utilizando el criterio de la suma de matrices, se enuncia el siguiente teorema.

TEOREMA 4.1

La suma de dos matrices diagonales (trapezoidal superior, trapezoidal inferior, Hessenberg superior, Hessenberg inferior, tridiagonal, simétrica) es una matriz diagonal (trapezoidal superior, trapezoidal inferior, Hessenberg superior, Hessenberg inferior, tridiagonal, simétrica),

Mostraremos solamente que la suma de dos matrices diagonales también es matriz diagonal.

Sean A y B dos matrices diagonales y C la matriz que resulta de sumar A y B ; sean además a_{ij} , b_{ij} , c_{ij} , los elementos de estas matrices. Debe cumplirse que:

$$\begin{aligned} a_{ij} &= b_{ij} = 0, \text{ para todo } i \neq j \\ \Rightarrow a_{ij} + b_{ij} &= 0, \text{ para todo } i \neq j \\ \Rightarrow c_{ij} &= a_{ij} + b_{ij} = 0, \text{ para todo } i \neq j \end{aligned}$$

$\therefore C$ es una matriz diagonal.

DEFINICION 4.2

Sea A una matriz $\ell \times m$ y B una matriz $m \times n$. El producto de A y B es la matriz C de orden $\ell \times n$ cuyos elementos están dados por:

$$c_{ij} = \sum_{k=1}^m a_{ik} b_{kj} \quad (i = 1, 2, \dots, \ell; j = 1, 2, \dots, n) \quad (4.1)$$

Escribimos

$$C = AB \quad (4.2)$$

OBSERVACION 4.3

Note que el producto AB está definido solamente cuando el número de columnas de A es igual al número de filas de B (en este caso se dice que A y B son *conformables* para el producto). El producto tiene el mismo número de filas de A y el mismo número de columnas de B .

TEOREMA 4.4

El producto de una matriz diagonal y una matriz diagonal (trapezoidal, Hessenberg, tridiagonal) es diagonal (trapezoidal, Hessenberg, tridiagonal).

El producto de dos matrices trapezoidales superiores (inferiores) es trapezoidal superior (inferior).

PRUEBA

Probaremos solamente que el producto de dos matrices trapezoidales superiores es trapezoidal superior.

Sean A y B matrices trapezoidales superiores de orden $\ell \times m$ y $m \times n$ respectivamente. Probaremos que AB es una matriz C trapezoidal superior.

Por hipótesis tenemos que

$$i > j \Rightarrow a_{ij} = b_{ij} = 0 \quad (\text{ver def. 3.3})$$

demostraremos que

$$i > j \Rightarrow c_{ij} = 0$$

Por definición de producto de matrices

$$c_{ij} = \sum_{k=1}^m a_{ik} b_{kj} \quad (i = 1, 2, \dots, \ell; j = 1, 2, \dots, n) \quad (4.3)$$

Consideremos $i > j$

se tiene que

$$a_{ik} = 0, \quad \forall k = 1, 2, \dots, i-1 \quad (4.4)$$

Luego (4.3) se puede escribir como:

$$c_{ij} = \sum_{k=1}^{i-1} a_{ik} b_{kj} + \sum_{k=i}^m a_{ik} b_{kj} \quad (4.5)$$

$$\Rightarrow c_{ij} = 0 + \sum_{k=i}^m a_{ik} b_{kj} \quad , \text{ por (4.4)}$$

$$\Rightarrow c_{ij} = \sum_{k=i}^m a_{ik} b_{kj}$$

$$\text{Pero } b_{kj} = 0, \quad \forall k = i, i+1, i+2, \dots, m.$$

$$\therefore c_{ij} = 0 \quad \text{cuando } i > j$$

▲

En esta sección describiremos una técnica usual que emplearemos frecuentemente para facilitar las manipulaciones matriciales.

EJEMPLO 4.5

Supóngase que particionamos una matriz A en submatrices como sigue:

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & | & a_{14} & | & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & | & a_{24} & | & a_{25} & a_{26} \\ \hline a_{31} & a_{32} & a_{33} & | & a_{34} & | & a_{35} & a_{36} \end{bmatrix} \quad (4.6)$$

podemos escribir esto mas simplemente como:

$$A = \begin{bmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \end{bmatrix} \quad (4.7)$$

TEOREMA 4.6

Podemos sumar y restar matrices particionadas como -- si las submatrices fueran elementos (escalares) ordinarios, -- siempre que las matrices esten particionadas de tal manera que las operaciones puedan efectuarse.

EJEMPLO 4.7

Sean

$$B = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix}, \quad C = \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix}$$

Si B_i es de orden $m_i \times n$ ($i = 1, 2, 3$) con $m_1 + m_2 + m_3 = m$ y las -- submatrices C_i son del mismo orden que las correspondientes B_i , para cada i , entonces:

$$B + C = \begin{bmatrix} B_1 \\ B_2 \\ B_3 \end{bmatrix} + \begin{bmatrix} C_1 \\ C_2 \\ C_3 \end{bmatrix} = \begin{bmatrix} B_1 + C_1 \\ B_2 + C_2 \\ B_3 + C_3 \end{bmatrix} \quad (4.8)$$

TEOREMA 4.8

Podemos multiplicar matrices particionadas como si las submatrices fueran elementos (escalares) ordinarios, siempre que las matrices esten particionadas de tal forma que los productos correspondientes puedan ser efectuados.

EJEMPLO 4.9

Sean las matrices A y B particionadas de la siguiente manera:

$$A = \left[\begin{array}{cc|c} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \end{array} \right] = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

$$B = \left[\begin{array}{cc} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \hline b_{31} & b_{32} \end{array} \right] = \begin{bmatrix} B_{11} \\ B_{21} \end{bmatrix}$$

Su producto sería:

$$C = \begin{bmatrix} C_{11} \\ C_{21} \end{bmatrix} = \begin{bmatrix} A_{11}B_{11} + A_{12}B_{21} \\ A_{21}B_{11} + A_{22}B_{21} \end{bmatrix} \quad (4.9)$$

Es fácil verificar que los productos correspondientes son conformables.

EJEMPLO 4.10

Sea $A = [A_1, A_2, A_3, \dots, A_n]$ una matriz $m \times n$ particionada en columnas. Si $\bar{x} \in \mathbb{R}^n$, entonces:

$$A\bar{x} = [A_1, A_2, \dots, A_n] \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_n \end{bmatrix}$$

$$A\bar{x} = x_1A_1 + x_2A_2 + \dots + x_nA_n \quad (4.10)$$

así $A\bar{x}$ es una combinación lineal de las columnas de A . Reciprocamente si \bar{y} es una combinación lineal de las columnas de A , es decir

$$\bar{y} = x_1A_1 + x_2A_2 + \dots + x_nA_n \quad (4.11)$$

entonces

$$\bar{y} = A\bar{x} \quad (4.12)$$

donde $\bar{x} = [x_1, x_2, \dots, x_n]^t$. Así el conjunto de todos los productos $A\bar{x}$ es el mismo que el conjunto de combinaciones lineales de columnas de A .

EJEMPLO 4.11

Sea $B = \text{diag} [b_1, b_2, \dots, b_n]$. El producto AB es la matriz obtenida de multiplicar la i -ésima columna de A por b_i , para cada i . El producto BA es la matriz obtenida de multiplicar la i -ésima fila de A por b_i , para cada i .

EJEMPLO 4.12

Sea B una matriz triangular superior. Entonces la primera columna de AB es un múltiplo de la primera columna de A . La segunda columna de AB es una combinación lineal de las dos columnas de A . En general, la i -ésima columna de AB es una combinación lineal de las primeras i columnas de A . En particular

Sean

$$A = \begin{bmatrix} 2 & 3 & 1 \\ 4 & 2 & 2 \\ 1 & 2 & 1 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 3 & 1 & 2 \\ 0 & 2 & 1 \\ 0 & 0 & 4 \end{bmatrix}$$

$$AB = \begin{bmatrix} 6 & 8 & 11 \\ 12 & 8 & 18 \\ 3 & 5 & 8 \end{bmatrix}$$

Para $i = 2$:

$$[8, 8, 5]^t = 1[2, 4, 1]^t + 2[3, 2, 2]^t$$

EJEMPLO 4.13

Muchas discusiones de matrices se centran alrededor de las *operaciones elementales en fila*. Estas son operaciones simples que pueden ser efectuadas en las filas de una matriz A para obtener una nueva matriz.

Ellas son:

1. Multiplicar la i -ésima fila de A por $c \in \mathbb{R}$.

EJERCICIOS

4.1 Efectuar las siguientes operaciones:

$$\text{a) } \begin{bmatrix} 1 & 6 & 4 \\ -4 & 2 & 8 \end{bmatrix} + \begin{bmatrix} -2 & 0 & 1 \\ 2 & -3 & 4 \end{bmatrix} \quad \text{b) } \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + 4 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} + 8 \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}$$

$$\text{c) } \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 4 & 3 \\ 2 & 1 \end{bmatrix} \quad \text{d) } \begin{bmatrix} 8 & 0 & -1 \\ 2 & 4 & 1 \\ -3 & -2 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \\ 3 \end{bmatrix}$$

$$\text{e) } \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & 2 & -1 \\ 1 & 4 & 6 \\ 3 & -5 & 4 \end{bmatrix} \quad \text{f) } \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 4 & -1 \\ 2 & 5 & -2 \\ 3 & 6 & 3 \end{bmatrix}$$

4.2 Muestre que $aA = \text{diag}[a, a, \dots, a]A$; $a \in \mathbb{R}$.

4.3 Sea $A \in \mathbb{R}^{\ell \times m}$, $B \in \mathbb{R}^{m \times n}$ y $a, b \in \mathbb{R}$. Muestre que
 $(aA)(bB) = abAB$.

4.4 Sea A simétrica, $M \in \mathbb{R}^{m \times n}$. Muestre que MAM^t es simétrica.

4.5 Muestre que para todo $\bar{x} \in \mathbb{R}^n$, $\bar{x}^t \bar{x} \geq \bar{0}$; además
 $\bar{x}^t \bar{x} = \bar{0}$ si y solo si $\bar{x} = \bar{0}$.

4.6 Sea $A, B \in \mathbb{R}^{m \times n}$. Pruebe que $A = B$ si y solo si $A\bar{x} = B\bar{x}$ para todo $\bar{x} \in \mathbb{R}^n$.

5. PROPIEDADES DE LAS MATRICES

Hemos dicho que el conjunto M de las matrices definidas sobre \mathbb{R} forman un espacio vectorial, por tanto las matri-

ces cumplen las propiedades de espacio vectorial (def. 1.1).

La sustracción de matrices se define en términos de operaciones ya consideradas:

$$A - B = A + (-1)B$$

El producto entre matrices no obedece todas las propiedades del producto ordinario entre números reales.

TEOREMA 5.1

Sean A una matriz $\ell \times m$, B una matriz $m \times n$, C una matriz $\ell \times n$ y $a \in \mathbb{R}$. Entonces:

1. $(AB)C = A(BC)$
2. $A(B+C) = AB + AC$
3. $(A+B)C = AC + BC$
4. $a(AB) = (aA)B = A(aB)$

EJEMPLO 5.2

El teorema anterior muestra que el producto entre matrices satisface las propiedades asociativas y distributiva. Sin embargo *no es conmutativo* como puede apreciarse en el siguiente ejemplo:

Sean

$$A = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

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

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

luego $AB \neq BA$

EJEMPLO 5.3

El siguiente producto matricial

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

muestra que es posible que al multiplicar dos matrices diferentes de cero resulte la matriz cero. En particular esto significa que de la expresión

$$AX = AY, \quad A \neq 0$$

no podemos concluir que $X = Y$; esto es, la *Ley Cancelativa* para números reales no se puede extender a las matrices.

TEOREMA 5.4

Si la matriz A de orden n posee inversa, ésta es única.

PRUEBA

Sea B la matriz inversa de A . Entonces tenemos

$$AB = BA = I$$

Supongamos que existe otra matriz inversa C tal que

$$AC = CA = I$$

COMO

$$AB = BA$$

$$\Rightarrow C(AB) = C(BA)$$

$$\Rightarrow (CA)B = C(BA)$$

$$\Rightarrow I B = C I$$

$$\text{luego } B = C$$

▲

DEFINICION 5.5

Cualquier matriz B obtenida al realizar una sola operación elemental en fila en la matriz unidad I , se conoce como una *Matriz elemental*.

DEFINICION 5.6

Dos matrices A y B se denominan *Equivalentes* ($A \sim B$), si una de ellas se deduce de la otra como consecuencia de efectuar operaciones elementales en fila.

DEFINICION 5.7

Una matriz simétrica A de orden n , es llamada *Definida positiva* si

$$\bar{x}^t A \bar{x} > 0$$

para cada vector columna n -dimensional $\bar{x} \neq \bar{0}$.

EJEMPLO 5.8

La matriz

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

es una matriz definida positiva. Supóngase que \bar{x} es cualquier vector columna 3-dimensional, entonces:

$$\bar{x}^t A \bar{x} = [x_1, x_2, x_3] \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

$$= [x_1, x_2, x_3] \begin{bmatrix} 2x_1 - x_2 \\ -x_1 + 2x_2 - x_3 \\ -x_2 + 2x_3 \end{bmatrix}$$

$$= [2x_1^2 - 2x_1x_2 + 2x_2^2 - 2x_2x_3 + 2x_3^2]$$

$$= [x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2]$$

$$y \quad x_1^2 + (x_1 - x_2)^2 + (x_2 - x_3)^2 + x_3^2 > 0$$

a menos que

$$x_1 = x_2 = x_3 = 0$$

DEFINICION 5.9

La suma de los elementos de la diagonal principal de cualquier matriz cuadrada A , se llama *Traza de A* y se denota por $tr[A]$.

$$\text{tr}[A] = a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^n a_{ii}$$

DEFINICION 5.10

Sea la matriz A de orden $m \times n$. El *espacio columna* denotado por $R(A)$ es el subespacio de \mathbb{R}^m generado por las *columnas* de A . El *espacio fila* de A es $R(A^t) \subset \mathbb{R}^n$. El *rango* de A , denotado por $\text{rang}(A)$ es la dimensión de $R(A)$, o sea

$$\text{rang}(A) = \dim [R(A)].$$

DEFINICION 5.11

Sea A una matriz $m \times n$. El *espacio Nulo* de A , denotado por $N(A)$ se define como sigue

$$N(A) = \{\bar{x} \in \mathbb{R}^n / A\bar{x} = \bar{0}\}$$

La *nulidad* de A , denotada por $\text{Null}(A)$, es la dimensión de $N(A)$, o sea

$$\text{Null}(A) = \dim [N(A)]$$

EJEMPLO 5.12

Sea

$$A = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} ; \quad A^t = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

$$\bar{y}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} ; \quad \bar{y}_1^t = [1, 0]$$

$$\bar{y}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad ; \quad \bar{y}_2^t = [0, 1]$$

$$\bar{y}_3 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Los vectores $\{\bar{y}_1, \bar{y}_2\}$ son l.i y $\bar{y}_3 = \bar{y}_1 + \bar{y}_2$, por lo que es l.d.;
los vectores $\{\bar{y}_1^t, \bar{y}_2^t\}$ son l.i y $\bar{y}_3^t = \bar{y}_1^t + \bar{y}_2^t$ es también l.d.

Entonces

$$\text{rang}(A) = 2 \quad \text{y} \quad \text{rang}(A^t) = 2.$$

Por definición, $N(A) = \{\bar{x} \in \mathbb{R}^3 / A\bar{x} = \bar{0}\}$

$$A\bar{x} = \begin{bmatrix} x_1 + x_3 \\ x_2 + x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{array}{l} x_1 + x_3 = 0 \\ x_2 + x_3 = 0 \end{array} \Rightarrow \begin{array}{l} x_1 = -x_3 \\ x_2 = -x_3 \end{array}$$

$$\text{luego } N(A) = \{\bar{x} \in \mathbb{R}^3 / \bar{x} = \begin{bmatrix} a \\ a \\ -a \end{bmatrix}, a \in \mathbb{R}\}$$

$$\dim [N(A)] = 1$$

$$\Rightarrow \text{Null}(A) = 1$$

Por definición, $N(A^t) = \{\bar{x} \in \mathbb{R}^2 / A^t\bar{x} = \bar{0}\}$

$$A^t\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_1 + x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow x_1 = x_2 = 0$$

$$\text{luego } N(A^t) = \{ \bar{x} \in \mathbb{R}^2 / \bar{x} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \}$$

$$\dim [N(A^t)] = 0$$

$$\Rightarrow \text{Null } (A^t) = 0$$

Nótese que $\text{Null } (A) \neq \text{Null } (A^t)$

OBSERVACION 5.13

En el ejemplo anterior hemos visto que a pesar de que $\text{rang } (A) = \text{rang } (A^t)$, no necesariamente $\text{Null } (A)$ tiene que ser igual a $\text{Null } (A^t)$.

EJERCICIOS

5.1 Dadas las matrices

$$A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 1 & 1 \\ 2 & 0 & 2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 3 & 0 \\ 0 & 4 & -1 \\ 2 & 3 & 0 \end{bmatrix}, \quad C = \begin{bmatrix} 6 & 5 & 7 \\ 2 & 2 & 4 \\ 3 & 3 & 6 \end{bmatrix}$$

verifique que $AC = BC$, sin embargo, $A \neq B$.

5.2 Demostrar que

$$\text{a) } \text{Traza } [A + B] = \text{traza } [A] + \text{traza } [B]$$

$$\text{b) } \text{Traza } [cA] = c \text{ traza } [A]$$

5.3 Sea $A \in \mathbb{R}^{m \times n}$. Demostrar que si $n > m$, entonces

$$\text{Null } (A) \neq 0.$$

5.4 Probar que si A es definida positiva, entonces A^{-1} es definida positiva.

5.5 Construya una matriz A no simétrica pero que $\bar{x}^t A \bar{x} > 0$

para todo $\bar{x} \neq \bar{0}$.

5.6 Muestre que la matriz

$$A = \begin{bmatrix} 4 & -1 & 1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{bmatrix}$$

es definida positiva

5.7 Demostrar que si U es triangular superior, entonces

$(AU)^{[k]} = A^{[k]} U^{[k]}$ para todo k siempre que $A^{[k]}$ esté definida. (Sugerencia: considerar las particiones

$$A = \begin{bmatrix} A^{[k]} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}, \quad U = \begin{bmatrix} U^{[k]} & U_{12} \\ 0 & U_{22} \end{bmatrix}$$

las cuales son conformables con respecto a la multiplicación).

5.8 Demostrar si L es triangular inferior, entonces

$(LA)^{[k]} = L^{[k]} A^{[k]}$ para todo k , siempre que $A^{[k]}$ esté definida.

6. TRANSFORMACIONES LINEALES Y MATRICIALES

DEFINICION 6.1

Sea $f: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ una función. Se dice que f es una transformación lineal si

$$f(a\bar{x} + b\bar{y}) = af(\bar{x}) + bf(\bar{y}), \quad \forall \bar{x}, \bar{y} \in \mathbb{R}^n; a, b \in \mathbb{R} \quad (6.1)$$

EJEMPLO 6.2

La función $T: \mathbb{R}^2 \longrightarrow \mathbb{R}$, tal que $T([x, y]^t) = x$ es una transformación lineal.

En efecto. Sean $[x_1, y_1]^t, [x_2, y_2]^t \in \mathbb{R}^2$ y $a, b \in \mathbb{R}$.

Entonces:

$$\begin{aligned} T(a[x_1, y_1]^t + b[x_2, y_2]^t) &= T([ax_1, ay_1]^t + [bx_2, by_2]^t) \\ &= T([ax_1 + bx_2, ay_1 + by_2]^t) \\ &= ax_1 + bx_2, \quad \text{por def. de } T \\ &= aT([x_1, y_1]^t) + bT([x_2, y_2]^t). \end{aligned}$$

EJEMPLO 6.3

Sea $f: \mathbb{R}^3 \longrightarrow \mathbb{R}^3$ una transformación lineal. Esta transformación lineal está determinada en forma única por los vectores de la base $\{\bar{e}_1, \bar{e}_2, \bar{e}_3\}$. En efecto:

$$\text{Sea } f(\bar{e}_1) = A_1 = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \end{bmatrix} \quad ; \quad f(\bar{e}_2) = A_2 = \begin{bmatrix} a_{12} \\ a_{22} \\ a_{32} \end{bmatrix} \quad \text{y} \quad f(\bar{e}_3) = A_3 = \begin{bmatrix} a_{13} \\ a_{23} \\ a_{33} \end{bmatrix}$$

entonces si $\bar{x} = x_1\bar{e}_1 + x_2\bar{e}_2 + x_3\bar{e}_3$

se tiene que:

$$\begin{aligned} \bar{y} = f(\bar{x}) &= f(x_1\bar{e}_1 + x_2\bar{e}_2 + x_3\bar{e}_3) \\ &= x_1f(\bar{e}_1) + x_2f(\bar{e}_2) + x_3f(\bar{e}_3) \\ &= x_1A_1 + x_2A_2 + x_3A_3 \end{aligned}$$

En términos de componentes individuales, si $\bar{y} = f(\bar{x})$, entonces:

$$\begin{aligned}
 y_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\
 y_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\
 y_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3
 \end{aligned}
 \tag{6.2}$$

Las ecuaciones (6.2) muestran que la transformación lineal f es tá completamente determinada por la matriz

$$\begin{bmatrix}
 a_{11} & a_{12} & a_{13} \\
 a_{21} & a_{22} & a_{23} \\
 a_{31} & a_{32} & a_{33}
 \end{bmatrix}
 \tag{6.3}$$

El arreglo (6.3) es llamado *representación matricial* de la transformación lineal f .

Ampliando lo anterior, enunciaremos el siguiente teorema.

TEOREMA 6.4

Sea $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es una transformación lineal. Entonces hay una única matriz de orden $m \times n$ tal que

$$f(\bar{x}) = A_f \bar{x}, \tag{6.4}$$

para todo $\bar{x} \in \mathbb{R}^n$. Recíprocamente, si A_f es una matriz $m \times n$, entonces la función f definida por (6.4) es una transformación lineal de \mathbb{R}^n a \mathbb{R}^m .

PRUEBA

Primero observemos que si B_f es cualquier otra matriz satisfaciendo (6.4), entonces $\forall \bar{x} \in \mathbb{R}^n$

$$B_f \bar{x} = f(\bar{x}) = A_f \bar{x}$$

Como f es una transformación lineal y (6.4) es válida para cualquier $\bar{x} \in \mathbb{R}^n$, tomemos $\bar{x} = \bar{e}_i$; luego tendríamos que

$$B_f \bar{e}_i = A_f \bar{e}_i \quad ,$$

entonces $B_i = A_i$, es decir que

$$b_{1i} = a_{1i}$$

$$b_{2i} = a_{2i}$$

$$\vdots$$

$$\vdots$$

$$\vdots$$

$$b_{ni} = a_{ni}$$

Por tanto $B_f = A_f$. Así, si A_f existe, es única.

Para construir A_f , sea $A_i = f(\bar{e}_i)$, ($i = 1, 2, \dots, n$) $\in \mathbb{R}^m$ y
 $A_f = (A_1, A_2, \dots, A_n)$

Entonces,

$$\begin{aligned} f(\bar{x}) &= f(x_1 \bar{e}_1 + x_2 \bar{e}_2 + \dots + x_n \bar{e}_n) \\ &= x_1 f(\bar{e}_1) + x_2 f(\bar{e}_2) + \dots + x_n f(\bar{e}_n) \\ &= x_1 A_1 + x_2 A_2 + \dots + x_n A_n \end{aligned}$$

así que A_f satisface (6.4).

Es fácil probar el recíproco.



EJERCICIOS

6.1 Mostrar que si $f, g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ son transformaciones lineales, entonces la función $f + g: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ definida por

$$(f + g)(\bar{x}) = f(\bar{x}) + g(\bar{x})$$

es una transformación lineal. Además

$$A_{f+g} = A_f + A_g$$

6.2 Muestre que si la función $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ es una transformación lineal y $a \in \mathbb{R}$. Entonces la función af definida por

$$(af)(\bar{x}) = af(\bar{x}) \quad , \quad \bar{x} \in \mathbb{R}^n$$

es una transformación lineal. Además

$$A_{af} = aA_f$$

6.3 Sea $f: \mathbb{R}^n \longrightarrow \mathbb{R}^m$ y $g: \mathbb{R}^m \longrightarrow \mathbb{R}^l$ transformaciones lineales. Mostrar que la función $g \circ f$ definida por

$$(g \circ f)(\bar{x}) = g(f(\bar{x})) \quad , \quad \bar{x} \in \mathbb{R}^n$$

es una transformación lineal. Además

$$A_{g \circ f} = A_g A_f$$

6.4 Muestre que la función $T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ definida por

$$T([x_1, x_2]^t) = [x_1 + x_2, 5x_2]^t$$

es una transformación lineal.

6.5 Muestre que la función $f: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ definida por

$$f([x_1, x_2]^t) = x_1 x_2$$

no es una transformación lineal.

6.6 Verificar que $T: \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ definida por

$$T \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} 2 & -1 & 3 \\ 0 & 2 & 1 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{bmatrix} 2x_1 - x_2 + 3x_3 \\ 2x_2 + x_3 \end{bmatrix}$$

es una transformación lineal.

7. ECUACIONES LINEALES

El sistema de m ecuaciones lineales

$$\begin{aligned} a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n &= b_1 \\ a_{21} x_1 + a_{22} x_2 + \dots + a_{2n} x_n &= b_2 \\ \vdots &\vdots \\ a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mn} x_n &= b_m \end{aligned} \tag{7.1}$$

con las variables x_1, x_2, \dots, x_n puede ser escrito en la forma matricial

$$A\bar{x} = \bar{b} \tag{7.2}$$

Dado un sistema de la forma (7.1) podemos formar la matriz

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & & \vdots & \vdots \\ \vdots & \vdots & & \vdots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & b_m \end{array} \right] = [A, \bar{b}]$$

que es llamada la matriz aumentada del sistema.

En esta sección se discutirá condiciones bajo las que --- (7.2) tiene una solución. Los resultados de esta sección no -- son constructivos; éstos no nos dicen cómo calcular una solu-- ción \bar{x} de (7.2). Métodos computacionales para resolver siste--

mas lineales seran tratados en el capítulo siguiente.

Hay alguna terminología asociada con el sistema (7.1) y la ecuación matricial (7.2). Sea $A \in \mathbb{R}^{m \times n}$. Si $m > n$, el sistema es llamado *sobredeterminado* (hay más ecuaciones que variables). Si $m < n$, el sistema se dice que es *bajodeterminado* (hay más variables que ecuaciones). Si $\bar{b} = 0$, el sistema es llamado *Homogéneo*.

La primera pregunta trata de la existencia de una solución para (7.2). Un criterio clásico está contenido en el siguiente teorema.

TEOREMA 7.1

La ecuación (7.2) tiene una solución \bar{x} si y solo si

$$\text{rang}([A, \bar{b}]) = \text{rang}(A) \quad (7.3)$$

PRUEBA

Supongamos que existe un \bar{x} tal que $A\bar{x} = \bar{b}$. Entonces \bar{b} es una combinación lineal de las columnas de A , por lo tanto está en el espacio columna $R(A)$. Se sigue que $R([A, \bar{b}]) = R(A)$, y

$$\text{rang}([A, \bar{b}]) = \text{rang}(A)$$

Recíprocamente, note que $R(A) \subset R([A, \bar{b}])$. Por tanto si (7.3) es satisfecha, $R(A)$ y $R([A, \bar{b}])$ tienen la misma dimensión y son además iguales (ejercic. 2.6)

Por tanto $\bar{b} \in R(A)$; esto es, \bar{b} es una combinación lineal de las columnas de A . Por el ejemplo (4.10) esto significa que $\bar{b} = A\bar{x}$ para algún $\bar{x} \in \mathbb{R}^n$.

COROLARIO 7.2

Si $\text{rang}(A) = m$, entonces (7.2) siempre tiene una solución.

PRUEBA

Como las columnas de A y $[A, \bar{b}]$ son m -vectores $m = \text{rang}(A) \leq \text{rang}([A, \bar{b}]) \leq m$, lo cual implica que $\text{rang}(A) = \text{rang}([A, \bar{b}]) = m$.



OBSERVACION 7.3

El teorema hace fácil la construcción de sistemas que no tienen soluciones. Por ejemplo, tomar

$$A = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad \text{y} \quad \bar{b} = [1, 0]^t$$

Entonces $\text{rang}(A) = 1 \neq 2 = \text{rang}([A, \bar{b}])$

TEOREMA 7.4

Sea \bar{x} una solución de (7.2). Entonces el conjunto de todas las soluciones de (7.2) es $\bar{x} + N(A)$, donde $\bar{x} + N(A) = \{\bar{z} \in \mathbb{R}^n / \bar{z} = \bar{x} + \bar{y} \wedge A\bar{y} = 0\}$.

PRUEBA

Si \bar{y} satisface (7.2), es decir, $A\bar{y} = \bar{b}$, entonces $A\bar{y} - A\bar{x} = \bar{0} \Rightarrow A(\bar{y} - \bar{x}) = 0$

$$\Rightarrow (\bar{y} - \bar{x}) \in N(A)$$

además $\bar{y} = \bar{x} + (\bar{y} - \bar{x}) \in \bar{x} + N(A)$

Recíprocamente, si $\bar{y} \in \bar{x} + N(A)$, entonces $\bar{y} = \bar{x} + \bar{z}$, para algún \bar{z} tal que $A\bar{z} = \bar{0}$

Por lo tanto, $A\bar{y} = A(\bar{x} + \bar{z})$

$$A\bar{y} = A\bar{x} + A\bar{z}$$

$$A\bar{y} = A\bar{x} = \bar{b}$$

▲

El teorema anterior nos permite obtener otras soluciones de (7.2), sumando vectores en el espacio nulo de A . Solamente existe una manera para que la solución sea única y es cuando el espacio nulo de A conste solamente del vector cero.

COROLARIO 7.5

Una solución de (7.2) es única si y solo si $\text{Null}(A) = \{0\}$.

TEOREMA 7.6

Un sistema bajodeterminado de ecuaciones homogéneas siempre tiene una solución no trivial.

PRUEBA

La ecuación homogénea $A\bar{x} = \bar{0}$ siempre tiene la solución trivial $\bar{x} = \bar{0}$. Por el teorema (7.3), el conjunto de todas las soluciones del sistema homogéneo sería el subespacio

$$\bar{0} + N(A) = N(A).$$

Si $\text{Null}(A) = \{0\} = \dim[N(A)]$, entonces $N(A) = \{\bar{0}\}$. En este caso $A\bar{x} = \bar{0}$ solo tendría la solución trivial $\bar{x} = \bar{0}$; como el

sistema es bajodeterminado, es decir que $n > m$, entonces $\text{Null}(A) \neq 0$ (por ejercicio 5.3); por consiguiente $\dim[N(A)] \neq 0$. Por tanto el sistema bajodeterminado $A\bar{x} = \bar{0}$ tiene una solución no trivial



Si $a \neq 0$, la ecuación escalar $ax_1 = b$ puede ser resuelta dividiendo por "a" y obtener $x_1 = a^{-1}b$. Esto sugiere que es posible expresar la solución de (7.2) en la forma

$$\bar{x} = A^{-1}\bar{b} \quad (7.4)$$

El requerimiento de que (7.2) tenga una solución única limita la dimensión de A:

- 1°) Si la solución es única, tenemos que $\text{Null}(A) = 0$, de lo que se deduce que $m \geq n$.
- 2°) La representación (7.4) implica que (7.2) tiene una solución para todo $b \in \mathbb{R}^m$. Esto significa que las n columnas de A generan \mathbb{R}^m y por lo tanto que $n \geq m$. Así $n = m$.

EJEMPLO 7.7

La restricción que $n \geq m$ no puede ser liberada, puesto que (7.4) implica que (7.2) debe siempre tener una solución. Sin embargo, podemos liberar la restricción de que la solución sea única y requerir solamente que (7.4) produzca una de las soluciones. Por ejemplo, sea

$$A = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} \quad \text{y} \quad C = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

Entonces $AC = I$. Por tanto, si

$$\bar{x} = C\bar{b}$$

tenemos $A\bar{x} = AC\bar{b} = I\bar{b} = \bar{b}$. Así la matriz C hace el papel de A^{-1} en (7.4). Sin embargo, C no es la única matriz que hace esto. Por ejemplo, la matriz

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

también satisface $AD = I$, y servirá igualmente de A^{-1} en (7.4).

Del ejemplo anterior, vemos que si A es una matriz de orden $m \times n$ ($m \neq n$), existen matrices que hacen el papel de inversa sólo a la derecha ó sólo a la izquierda de A

Si A es una matriz cuadrada, no necesariamente posee inversa. Condiciones necesarias y suficientes para que A tenga inversa están dadas en el siguiente teorema.

TEOREMA 7.8

Sea A una matriz cuadrada de orden n . Entonces las siguientes proposiciones son equivalentes:

1. $\text{rang}(A) = n$
2. $\text{Null}(A) = 0$
3. Las columnas de A son linealmente independientes
4. Las filas de A son linealmente independientes

5. Existe una matriz A^{-1} de orden n satisfaciendo

$$A^{-1}A = AA^{-1} = I_n \quad (7.5)$$

PRUEBA

Las condiciones 1, 2 y 3 son obviamente equivalentes. Para demostrar que la parte 3 implica 5:

Sea $f: \mathbb{R}^n \longrightarrow \mathbb{R}^n$, la transformación lineal representada por la matriz A . Ahora por la parte 1 y el corolario 7.2, la ecuación $A\bar{x} = \bar{b}$ siempre tiene una solución. Por la parte 2 y corolario 7.5, la solución es única. Esto equivale a decir que f es uno a uno y sobre (ejercicio 7.6). Por lo tanto f tiene una inversa g satisfaciendo

$$g \circ f = f \circ g = i_n \quad (7.6)$$

afirmamos que g es lineal. Para mostrar esto, sean $\bar{x}_1, \bar{x}_2 \in \mathbb{R}^n$ y $\bar{y}_1 = g(\bar{x}_1)$, $\bar{y}_2 = g(\bar{x}_2)$, además que

$$\bar{x}_i = f(\bar{y}_i) \quad (i = 1, 2)$$

entonces

$$\begin{aligned} f(a\bar{y}_1 + b\bar{y}_2) &= af(\bar{y}_1) + bf(\bar{y}_2) \\ &= a\bar{x}_1 + b\bar{x}_2 \end{aligned}$$

Por tanto

$$\begin{aligned} ag(\bar{x}_1) + bg(\bar{x}_2) &= a\bar{y}_1 + b\bar{y}_2 \\ &= g[f(a\bar{y}_1 + b\bar{y}_2)] \\ &= g(a\bar{x}_1 + b\bar{x}_2) \end{aligned}$$

Como g es una transformación lineal de \mathbb{R}^n en \mathbb{R}^n es re

presentada por una matriz A^{-1} de orden n . Entonces la ecuación (7.6) es equivalente a $A^{-1}A = AA^{-1} = I_n$, lo cual es justamente (7.5).

Para mostrar que la parte 5 implica parte 3, supóngase que $A\bar{x} = \bar{0}$. Entonces

$$\bar{0} = A^{-1}\bar{0} = A^{-1}A\bar{x} = I_n\bar{x} = \bar{x}$$

Así, si $A\bar{x} = \bar{0}$, entonces $\bar{x} = 0$, lo cual es equivalente a decir que las columnas de A son linealmente independientes.

Argumentos idénticos con las filas en lugar de las columnas establecerá la equivalencia de las partes 4 y 5. ▲

EJERCICIOS

- 7.1 Muestre que la matriz $A \in \mathbb{R}^{n \times n}$ es no singular si y solamente si la ecuación homogénea $A\bar{x} = \bar{0}$ tiene la solución trivial.
- 7.2 Sea $A \in \mathbb{R}^{m \times n}$. Muestre que $A^t A$ es no singular si y solamente si A tiene columnas linealmente independientes.
- 7.3 Muestre que si A tiene filas linealmente independientes y columnas linealmente independientes, entonces A es cuadrada y no singular.
- 7.4 Sean

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{y} \quad \bar{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

Probar que el sistema homogéneo $A\bar{x} = \bar{0}$ tiene soluciones no triviales.

7.5 Las siguientes matrices representan la matriz aumentada $[A, \bar{b}]$ de un sistema de ecuaciones lineales. En cada caso determinar cuales ecuaciones tienen infinito número de soluciones, solución única ó no posee solución.

$$\text{a) } \begin{bmatrix} 0 & 1 & 0 & 0 & 4 \\ 0 & 0 & 1 & 2 & 5 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\text{b) } \begin{bmatrix} 3 & 2 & 1 & 6 \\ 0 & 1 & 2 & 4 \\ 0 & 0 & 1 & 1 \end{bmatrix}$$

$$\text{c) } \begin{bmatrix} 1 & 3 & 4 & 9 \\ 0 & 2 & 3 & 7 \\ 0 & 0 & 2 & 6 \\ 0 & 0 & 1 & 4 \end{bmatrix}$$

$$\text{d) } \begin{bmatrix} 1 & 0 & 0 & 0 & 4 \\ 0 & 1 & 2 & 0 & 5 \\ 0 & 0 & 0 & 1 & 6 \end{bmatrix}$$

7.6 Si el sistema $A\bar{x} = \bar{b}$ tiene solución única; probar que la transformación lineal $f: \mathbb{R}^n \longrightarrow \mathbb{R}^n$ representada por la matriz A , es biyectiva.

8. ERRORES, ARITMETICA Y ESTABILIDAD

En esta sección daremos conceptos que serán de mucha utilidad para estudiar los efectos del error por redondeo en los algoritmos que aparecieran en los capítulos siguientes.

Como casi siempre tratamos con cantidades aproximadas, el primer problema es dar una definición satisfactoria de error en una aproximación.

Sea b una aproximación de a . Una medida natural de la exactitud de b es el número $|a - b|$. Sin embargo, este número no tiene sentido a menos que el tamaño de a sea conocido; resulta ser

además una cantidad engañosa. Por ejemplo, consideremos dos casos

$$\begin{aligned} a &= 1.234 & , & & a &= 0.002, \\ b &= 1.233 & , & & b &= 0.001, \\ |a - b| &= 10^{-3} & , & & |a - b| &= 10^{-3} \end{aligned}$$

En ambos casos $|a - b|$ es 10^{-3} , pero solamente en el primer caso diremos que b es una buena aproximación de a .

En el segundo caso, a es tan pequeño que la diferencia $|a - b| = 10^{-3}$ entre a y b representa un cambio significativo. Así, es claro que lo que es importante; es la relación de $|a - b|$ con a . Este ejemplo motiva las siguientes definiciones de error.

DEFINICION 8.1

Sean $a, b \in \mathbb{R}$, con b considerada como una aproximación de a .

El *residual* de b es el número

$$a - b$$

El *error absoluto* ó error en b es el número

$$e = |a - b|$$

Si $a \neq 0$, el *residual relativo* de b es el número

$$\frac{a - b}{a}$$

así, si el residual de b es ξ , es decir,

$$a - b = \xi$$

$$\Rightarrow b = a - \xi$$

En otras palabras, si b tiene error absoluto pequeño, entonces b puede ser obtenido de a sustrayendo una cantidad pequeña. Si similarmente si b tiene residual relativo $\tilde{\delta}$, es decir,

$$\frac{a - b}{a} = \tilde{\delta}$$

$$\Rightarrow 1 - \frac{b}{a} = \tilde{\delta}$$

$$\Rightarrow \frac{b}{a} = 1 - \tilde{\delta}$$

$$\therefore b = a(1 - \tilde{\delta}) = a(1 + \delta) \text{ donde } \delta = -\tilde{\delta}$$

En otras palabras, si b tiene un error relativo pequeño, entonces b puede ser obtenido de a multiplicando por una cantidad muy cercana a la unidad.

Muchos cálculos numéricos en computadora son efectuados en aritmética de punto flotante. La forma de un número en punto flotante varía de computador en computador. Para ilustrar su diferencia daremos la siguiente definición.

DEFINICION 8.2

Un número decimal con t dígitos expresado en punto flotante tiene la forma

$$y = \pm \cdot d_1 d_2 \dots d_t 10^e, \quad (8.1)$$

donde t es la precisión del sistema, es decir, el número de ci

fras que podemos representar*. En la expresión (8.1) observamos dos partes:

- 1) Fracción ó Mantisa: $\pm \cdot d_1 d_2 \dots d_t$
- 2) Exponente ó Característica e . El número "e" es un entero -- tal que $m < e < M$, donde m es un entero negativo y M un entero positivo.

Si en la expresión (8.1) se da $0.1 \leq | \cdot d_1 d_2 \dots d_t | < 1$, entonces se dice que el número está *normalizado*.

Si el sistema es de base β , (8.1) se transforma en

$$y = \pm \cdot d_1 d_2 \dots d_t \beta^e, \quad 0 < d_i < \beta \quad (8.2)$$

Un sistema de punto flotante queda determinado de la manera siguiente

$$f(\beta, t, m, M), \quad (8.3)$$

donde β es la base del sistema, t la precisión, m y M es el -- rango del exponente.

EJEMPLO 8.3

Sea el sistema representado por $f(10, 4, -63, 63)$. -- Tres maneras diferentes de expresar el número $y = 58.73$ en forma de punto flotante serían:

* En muchas computadoras los números en punto flotante tienen precisión fija, otras tienen la capacidad para manipular números en punto flotante con el doble de la precisión usual. Tales números en punto flotante son llamados *números en doble -- precisión* y los cálculos involucrados se dice que están en *doble precisión*.

- i) $y = 0.5873 \times 10^2$
- ii) $y = 0.0587 \times 10^3$
- iii) $y = 0.0058 \times 10^4$

Los números ii) y iii) no están normalizados y no son convenientes por la pérdida de cifras significativas que nos ocasiona. Por esta razón en lo sucesivo consideraremos únicamente números normalizados.

En todo número real puede ser expresado en la forma:

$$y = \pm \cdot d_1 d_2 \dots d_t d_{t+1} d_{t+2} \dots \times 10^e,$$

es decir, $y = (\pm \cdot d_1 d_2 \dots d \pm 0.0 \dots 0 d_{t+1} d_{t+2} \dots) 10^e$

$$y = \pm \cdot d_1 d_2 \dots d \cdot 10^e \pm \cdot d_{t+1} d_{t+2} \dots \times 10^{e-t},$$

lo que simbolizaremos por

$$y = \pm p_y 10^e \pm q_y 10^{e-t} \quad (8.4)$$

Si queremos representar en el sistema $f(10, 4, -63, 63)$ el número $y = 0.4678722\bar{2} \times 10^3$, o sea

$$y = 0.4678 \times 10^3 + 0.722\bar{2} \times 10^{-1}$$

debemos redondearlo. Para ello existen dos formas:

i) Redondeo por Corte:

$$y_c = 0.4678 \times 10^3$$

ii) Redondeo Simétrico:

$$y_s = 0.4679 \times 10^3$$

Es decir que un número de la forma (8.4) representado con redondeo por corte sería

$$y_c = \pm p_y 10^e$$

y por medio de redondeo simétrico

$$Y_s = \begin{cases} \pm p_y 10^e, & \text{si } |q_y| < \frac{1}{2} \\ \pm p_y 10^e \pm 10^{e-t}, & \text{si } |q_y| \geq \frac{1}{2}. \end{cases}$$

Para el análisis de los errores absolutos y relativos que se cometen al representar un número en un sistema de punto flotante, consideraremos únicamente los números positivos.

Si un número y está expresado en la forma (8.4), entonces

$$Y_c = p_y 10^e.$$

Luego el error absoluto sería

$$\begin{aligned} e_y &= |y - Y_c| = |p_y 10^e + q_y 10^{e-t} - p_y 10^e| \\ \therefore e_y &= q_y 10^{e-t}. \end{aligned} \quad (8.5)$$

La cota para el error absoluto es

$$e_y \leq 10^{e-t}, \quad \text{ya que } 0 \leq q_y < 1.$$

El error relativo a Y_c sería

$$\begin{aligned} \delta_y &= \frac{|e_y|}{|Y|} = \frac{q_y 10^{e-t}}{p_y 10^e + q_y 10^{e-t}} \\ \delta_y &= \frac{q_y 10^{e-t}}{\cdot d_1 d_2 \dots d_t d_{t+1} \dots 10^e}. \end{aligned} \quad (8.6)$$

Luego una cota para (8.6) sería

$$\frac{q_y 10^{e-t}}{\cdot d_1 d_2 \dots d_t d_{t+1} \dots 10^e} \leq \frac{10^{e-t}}{10^{-1} 10^e} = \frac{10^{e-t}}{10^{e-1}} = 10^{1-t}$$

DEFINICION 8.4

El número $\mu_c = 10^{1-t}$, es la *Unidad de Redondeo - por Corte* en la aritmética decimal de punto flotante con t dígitos significativos de precisión.

Si un número y está expresado como

$$y = p_y 10^e + q_y 10^{e-t}$$

entonces el error absoluto de y_s cuando $q_y < \frac{1}{2}$, sería

$$e_y = |y - y_s| = |p_y 10^e + q_y 10^{e-t} - p_y 10^e|$$

$$\therefore e_y = q_y 10^{e-t} .$$

Como $q_y < \frac{1}{2}$ se tiene que

$$e_y = q_y 10^{e-t} < \frac{1}{2} 10^{e-t} \quad (8.7)$$

El error absoluto de y_s cuando $q_y \geq \frac{1}{2}$ es:

$$e_y = |y - y_s| = |p_y 10^e + q_y 10^{e-t} - p_y 10^e - 10^{e-t}|$$

$$= |q_y 10^{e-t} - 10^{e-t}|$$

$$= |(q_y - 1) 10^{e-t}|$$

$$\therefore e_y = |1 - q_y| 10^{e-t} . \quad (8.8)$$

Luego una cota para (8.8) sería:

$$e_y = |1 - q_y| 10^{e-t} \leq \frac{1}{2} 10^{e-t} .$$

El error relativo para y_s es:

$$\delta_y = \left| \frac{e}{y} \right| \frac{|1 - q_y| 10^{e-t}}{|.d_1 d_2 \dots d_t d_{t+1} \dots| 10^e}$$

$$\delta_y = \frac{|1 - q_y| 10^{e-t}}{.d_1 d_2 \dots d_t d_{t+1} \dots 10^e} \quad (8.9)$$

Una cota para (8.9) es:

$$\frac{|1 - q_y| 10^{e-t}}{.d_1 d_2 \dots d_t d_{t+1} \dots 10^e} \leq \frac{1/2 10^{e-t}}{10^{-1} 10^e} = \frac{1}{2} 10^{-t}.$$

NOTA:

De la misma manera se hacen los análisis de error cuando se trate de números negativos y llegaríamos a los mismos resultados.

DEFINICION 8.5

Se define $\mu = \frac{1}{2} 10^{1-t} = 5 \times 10^{-t}$ como la *Unidad de Redondeo Simétrico* en la aritmética de punto flotante con t dígitos.

OBSERVACION 8.6

- Ya sea que se ocupe en los cálculos redondeo por corte ó redondeo simétrico, la unidad de redondeo no depende de la magnitud del número que representemos en el sistema, sino que unicamente depende de las características del sistema.
- Notemos que $\mu_c = 2\mu$.

NOTA:

En los cálculos que esten involucrados en este trabajo, se ocupará siempre redondeo simétrico; en lo sucesivo usaremos

el símbolo μ en lugar de μ_s .

Sea $\text{fl}(a)$ la expresión en punto flotante del número real a ; hemos comprobado que

$$\text{fl}(a) = a(1 + \delta) \quad , \quad \text{donde } \delta = -\tilde{\delta} \leq \mu$$

Si $\text{fl}(ab)$, $\text{fl}(a/b)$, $\text{fl}(a + b)$ denotan el producto, cociente y suma de los números reales a y b donde éstos están representados en punto flotante, entonces:

$$\begin{aligned} \text{i) } \text{fl}(ab) &= ab(1 + \delta) \quad , \quad \text{donde } |\delta| \leq \mu \\ \text{ii) } \text{fl}(a/b) &= (a/b)(1 + \delta) \quad , \quad \text{donde } |\delta| \leq \mu \\ \text{iii) } \text{fl}(a + b) &= a(1 + \delta_1) + b(1 + \delta_2) \quad , \quad \text{donde} \\ &|\delta_1| \quad , \quad |\delta_2| \leq \mu \end{aligned} \tag{8.10}$$

Estas expresiones dicen que las operaciones de redondeo, multiplicación y división introducen solamente pequeños errores relativos.

En nuestra definición de un número en punto flotante no observamos limitaciones en el tamaño de la característica. Sin embargo, si la característica está entre -100 y 100, entonces los números mayores que 10^{100} ó menores que 10^{-100} no tienen una representación en punto flotante. Cuando una operación produce un número con una característica bastante grande o muy pequeña, se dice que sufre de *sobrecarga* ó *bajacarga* respectivamente.

Daremos ahora dos ejemplos de cálculos en los cuales el error por redondeo produce resultados inexactos. Todos los cálculos están dados en aritmética de 4-dígitos.

EJEMPLO 8.7

La menor raíz r_1 de la ecuación cuadrática

$$x^2 - 6.433x + 0.009474 = 0 \quad (8.11)$$

es, con cinco cifras, 1.4731×10^{-3} . Una expresión numérica para r_1 está dada por la fórmula cuadrática

$$r_1 = \frac{6.433 - \sqrt{(6.433)^2 - 4(0.009474)}}{2} \quad (8.12)$$

En aritmética de 4-dígitos, una evaluación de (8.12) con redondeo simétrico sería:

- 1) $a = fl[(6.433)^2] = 41.38$
- 2) $b = fl[4(0.009474)] = 0.03800$
- 3) $c = fl(a - b) = 41.34$
- 4) $d = fl(\sqrt{c}) = 6.430$
- 5) $e = fl(6.433 - d) = 0.003$
- 6) $r_1' = fl(e/2) = 0.0015$

así el valor calculado r_1' no está de acuerdo con r_1 en su segunda cifra significativa.

EJEMPLO 8.8

El siguiente sistema de ecuaciones lineales

$$\begin{aligned} 3.000 x_1 + 4.127 x_2 &= 15.41 \\ 1.000 x_1 + 1.374 x_2 &= 5.147 \end{aligned} \quad (8.13)$$

tiene por solución $x_1 = 13.6658$ y $x_2 = -6.2$. Si la primera ecuación es dividida por 3.000 y restada de la segunda, el resultado es

$$\left(1.374 - \frac{4.127}{3}\right) x_2 = 5.147 - \frac{15.41}{3}$$

así

$$x_2 = \frac{5.147 - \frac{15.41}{3}}{1.374 - \frac{4.127}{3}} \quad (8.14)$$

Una evaluación de (8.14) es la siguiente

- 1) $a = fl\left(\frac{15.41}{3}\right) = 5.140$
- 2) $b = fl(5.147 - a) = 0.07$
- 3) $c = fl\left(\frac{4.127}{3}\right) = 1.376$
- 4) $d = fl(1.374 - c) = -0.002$
- 5) $x'_1 = fl\left(\frac{b}{d}\right) = -3.500$

De nuevo el valor calculado no está de acuerdo con el valor -- verdadero. En ambos ejemplos, el fracaso se encuentra en la -- resta de cantidades bastante iguales que cancela mas cifras -- significativas (paso 5 en el ejemplo 8.7 y pasos 2 y 4 en ejemplo 8.8).

A pesar de sus similitudes, sin embargo, los cálculos fallan por razones diferentes. La falla en el ejemplo 8.7 fué -- causada por la desafortunada representación (8.12) de r_1 , cuyo cálculo requiere la cancelación de cifras significativas. Esta cancelación es evitada con la expresión alternativa

$$r_1 = \frac{2(0.009474)}{6.433 + \sqrt{(6.433)^2 - 4(0.09474)}} \quad (8.15)$$

cuyo resultado con 4 dígitos es $r_1' = 1.473$, que es una respuesta bastante satisfactoria.

Si el lector está, por el momento, queriendo usar la palabra "cercana" sin especificar exactamente lo que significa, podemos hacer esta distinción mas específicamente. Muchos problemas numéricos pueden ser descritos de la siguiente manera.

Dada una función definida matemáticamente $f: \mathbb{C}^m \rightarrow \mathbb{R}^n$. Las m componentes del argumento \bar{x} de f son los datos para determinar el problema y las n componentes de $f(\bar{x})$ son las respuestas. El problema numérico consiste entonces en calcular la aproximación a $f(\bar{x})$ dado un \bar{x} .

Por ejemplo, la DATA en el ejemplo 8.7 es el vector $\bar{x} = (b, c)$ de coeficientes de la ecuación $x^2 + bx + c = 0$.

La naturaleza de la función f limita la exactitud alcanzada por los cálculos, pero en la práctica solamente una aproximación \bar{x}^* a la DATA es conocida, y siempre solamente podemos calcular $f(\bar{x}^*)$.

Ahora para algunas clases de problemas, $f(\bar{x})$ y $f(\bar{x}^*)$ pueden diferir grandemente. Tal problema se dice que está *mal condicionado*. Si intentamos resolver un problema mal condicionado comenzando con la data inexacta, la solución será inexacta no importando cómo sea calculada. El problema del ejemplo 8.8 está mal condicionado.

La selección e implementación de un algoritmo para resolver el problema matemático asociado con f es importante para definir una nueva función f^* que, dada una data \bar{x} resulta una

solución aproximada $f^*(\bar{x})$. No esperamos que f^* resuelva problemas mal condicionados con más exactitud que lo que la data garantiza, sin embargo, estaríamos muy preocupados si f^* introduce inexactitudes más grandes que la propia función f . Esto sugiere que requerimos de f^* lo siguiente: Para cualquier $\bar{x} \in E$ existe valor próximo $\bar{x}^* \in E$ tal que $f(\bar{x}^*)$ está cerca de $f^*(\bar{x})$. En otras palabras requerimos que el algoritmo dé una solución que esté cerca de la solución exacta de un problema poco perturbado. Algoritmos con esta propiedad son llamados *estables*. La ecuación (8.12) representa una inestable manera de calcular la raíz r_1 de (8.11).

EJERCICIOS

8.1 Sean $A, B \in \mathbb{R}^{m \times n}$. Mostrar que $\ell(A + B) = A' + B'$, donde

$$a'_{ij} = a_{ij}(1 + \partial_{ij}), \quad |\partial_{ij}| \leq \mu,$$

y

$$b'_{ij} = b_{ij}(1 + \delta_{ij}), \quad |\delta_{ij}| \leq \mu.$$

8.2 Mostrar que si $|\partial_1|, |\partial_2|, \dots, |\partial_n| \leq \mu$, entonces

$$(1 + \partial_1)(1 + \partial_2) \dots (1 + \partial_n) = (1 + \partial)^n,$$

donde

$$|\partial| \leq \mu.$$

8.3 Sea $|\partial| \leq \mu$. Mostrar que si $n\mu < 0.1$, entonces

$$(1 + \partial)^n = 1 + n\partial', \quad \text{donde } |\partial'| \leq 1.06 \mu.$$

Así si

$$\mu' = 1.06 \mu$$

y $|\partial_1|, |\partial_2|, \dots, |\partial_n| \leq \mu$, se sigue que

$$(1 + \partial_1)(1 + \partial_2) \dots (1 + \partial_n) = 1 + n\mu,$$

donde $|\partial| < \mu'$.

8.4 Mostrar que

$$\ell(a_1 + a_2 + \dots + a_n) = a_1(1 + \partial_1) + a_2(1 + \partial_2) + \dots + a_n(1 + \partial_n)$$

donde

$$|\partial_1| \leq (n - 1) \mu'$$

y

$$|\partial_i| \leq (n - i + 1) \mu' \quad (i = 2, 3, \dots, n)$$

CAPITULO I I

METODOS DIRECTOS PARA RESOLVER SISTEMAS DE ECUACIONES LINEALES

PRIMERA PARTE

En el capítulo anterior se dió a conocer las condiciones que debe cumplir un sistema de ecuaciones lineales para que -- tenga solución.

En este capítulo estaremos concentrados con algoritmos - que resolverán la ecuación $A\bar{x} = \bar{b}$ (1)

Los algoritmos tienen en común la idea de factorar A en - la forma

$$A = R_1 R_2 \cdots R_k$$

La ecuación (1) puede ser facilmente resuelta tomando $\bar{y}_0 = \bar{b}$ y para $i = 1, 2, \dots, k$ calculando \bar{y}_i como la solución de la ecuación

$$R_i \bar{y}_i = \bar{y}_{i-1}$$

Entonces se tiene

$$R_1 R_2 R_3 \cdots R_k \bar{x} = \bar{b}$$

$$R_1 R_2 R_3 \cdots R_k \bar{x} = \bar{y}_0$$

$$R_2 R_3 \cdots R_k \bar{x} = \bar{y}_1 = R_1^{-1} \bar{y}_0$$

\vdots

$$R_i \cdots R_k \bar{x} = \bar{y}_{i-1} = R_{i-1}^{-1} \cdots \bar{y}_0$$

Obviamente \bar{y}_k es la solución deseada \bar{x} .

Las matrices R_i que resultan de la factorización de A son a menudo triangulares. Por consiguiente estudiaremos los algoritmos para resolver sistemas triangulares.

Cuando A es no singular, los métodos de este capítulo pueden ser adaptados para dar algoritmos para el cálculo de A^{-1} ; sin embargo, el cálculo de A^{-1} es innecesario para la resolución de (1) porque consume demasiado tiempo; los métodos que daremos aquí serán mas cortos que primero calcular A^{-1} y luego formar $A^{-1}\vec{b}$.

1. MATRICES TRIANGULARES Y SISTEMAS

En esta sección daremos algoritmos para obtener la inversa de matrices triangulares y para resolver sistemas de ecuaciones triangulares. Hay varias razones para considerar este primer caso especial. En primer lugar, nuestro algoritmos subsiguientes facilitan la resolución de sistemas triangulares. Segundo, los algoritmos para sistemas triangulares son comparativamente simples. Tercero, los algoritmos proveen ejemplos buenos y simplificados de algunas consideraciones prácticas relacionadas con el almacenamiento y la cantidad de operaciones efectuadas. Restringiremos nuestra atención a matrices triangulares superiores, las modificaciones para matrices triangulares inferiores serían obvias.

TEOREMA 1.1

Sean A y C de orden ℓ y m respectivamente, y B una matriz de orden $\ell \times m$. Entonces la matriz

$$T = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}$$

es no singular si y solo si A y C son no singulares. En este caso

$$T^{-1} = \begin{bmatrix} A & B \\ 0 & C \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & -A^{-1}BC^{-1} \\ 0 & C^{-1} \end{bmatrix} \quad (1.1)$$

PRUEBA

"<="

Si A y C son no singulares, la matriz del lado derecho de (1.1) está bien definida y es la inversa de T.

En efecto:

Sean X, Z de orden ℓ y m respectivamente, y Y de orden $\ell \times m$.

$$\begin{aligned} \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} \begin{bmatrix} X & Y \\ 0 & Z \end{bmatrix} &= \begin{bmatrix} AX + BY & AY + BZ \\ 0X + CY & 0Y + CZ \end{bmatrix} \\ &= \begin{bmatrix} AX & AY + BZ \\ 0 & CZ \end{bmatrix} \\ &= \begin{bmatrix} I_{\ell} & 0 \\ 0 & I_m \end{bmatrix} \end{aligned}$$

entonces

$$\begin{aligned} AX = I_\ell &\Rightarrow X = A^{-1}I_\ell ; \text{ como } A \text{ es no singular} \\ &\Rightarrow X = A^{-1} \end{aligned}$$

$$\begin{aligned} CZ = I_m &\Rightarrow Z = C^{-1}I_m ; \text{ como } C \text{ es no singular} \\ &\Rightarrow Z = C^{-1} \end{aligned}$$

$$\begin{aligned} AY + BZ = 0 &\Rightarrow AY + BC^{-1} = 0 \\ &\Rightarrow AY = -BC^{-1} \\ &\Rightarrow Y = -A^{-1}BC^{-1} \end{aligned}$$

luego T es no singular y (1.1) es su inversa.

" \Rightarrow "

Supóngase que C es singular. Entonces existe un vector diferente de cero \bar{x} , con m componentes, tal que $\bar{x}^t C = \bar{0}^t$.

Si $\bar{0}^t$ denota el cero vector fila con ℓ componentes, entonces $[\bar{0}^t, \bar{x}^t] \neq 0$ y

$$\begin{aligned} [\bar{0}^t, \bar{x}^t] T &= [\bar{0}^t, \bar{x}^t] \begin{bmatrix} A & B \\ 0 & C \end{bmatrix} \\ &= [\bar{0}^t A + \bar{x}^t 0, \bar{0}^t B + \bar{x}^t C] = \bar{0} \end{aligned}$$

Por lo tanto T es singular. El caso donde A es singular es tratado similarmente.

▲

El teorema 1.1 tiene mejor uso en consecuencias teóricas. Note que la matriz A es submatriz directriz principal de orden ℓ de T ; esto es $A = T^{[\ell]}$ (ver la discusión siguiendo la definición 1.3.15).

De (1.1) se tiene que la matriz A^{-1} es la submatriz directriz principal de orden ℓ de T^{-1} . En otras palabras, si T es triangular superior, entonces $(T^{[\ell]})^{-1} = (T^{-1})^{[\ell]}$

El teorema 1.1 nos permite describir la inversa de una matriz triangular.

TEOREMA 1.2

Sea T una matriz triangular superior. Entonces T es no singular si y solo si sus elementos de la diagonal son diferentes de cero. En este caso T^{-1} es triangular superior.

PRUEBA

La prueba es por inducción en el orden de T . La afirmación es obviamente cierta para matrices de orden 1. Supóngase que el teorema es cierto para todas las matrices de orden $n > 1$ y T es de orden n . Entonces T puede ser particionada en la forma

$$T = \begin{bmatrix} T_{n-1} & \bar{t}_n \\ 0 & t_{nn} \end{bmatrix}$$

donde T_{n-1} es la submatriz directriz principal de T de orden $n-1$ y $\bar{t}_n = [t_{1n}, t_{2n}, \dots, t_{n-1,n}]^t$. Ahora T_{n-1} y t_{nn} son matrices triangulares de orden menor que n . Ellas tienen elementos diferentes de ceros en la diagonal. Por la hipótesis inductiva ellas son no singulares si y solo si tienen elementos diferentes de cero en la diagonal. Finalmente por el teorema 1.1, T es no singular si y solo si T_{n-1} y t_{nn} son no singulares.

Para ver que T^{-1} es triangular superior, note que por teorema 1.1

$$T^{-1} = \begin{bmatrix} T_{n-1}^{-1} & -T_{n-1}^{-1} \bar{t}_n t_{nn}^{-1} \\ 0 & t_{nn}^{-1} \end{bmatrix} \quad (1.2)$$

Por hipótesis inductiva T_{n-1}^{-1} es triangular superior. Por lo tanto T^{-1} es triangular superior.

▲

Regresemos al problema de resolución del sistema lineal

$$T \bar{x} = \bar{b} \quad (1.3)$$

donde T es una matriz triangular superior no singular de orden n . La última ecuación del sistema (1.3) es

$$t_{nn} x_n = b_n$$

del que se tiene

$$x_n = t_{nn}^{-1} b_n \quad (1.4)$$

En general, si conocemos $x_n, x_{n-1}, \dots, x_{i+1}$, podemos resolver la i -ésima ecuación de (1.3), que es

$$t_{ii} x_i + t_{i,i+1} x_{i+1} + \dots + t_{in} x_n = b_i,$$

para obtener

$$x_i = t_{ii}^{-1} (b_i - \sum_{j=i+1}^n t_{ij} x_j) \quad (1.5)$$

Las fórmulas (1.4) y (1.5) definen un algoritmo para el cálculo de la solución de (1.3).

ALGORITMO 1.3:

Sea $T \in \mathbb{R}^{n \times n}$ una matriz triangular superior y no singular, y sea $\bar{b} \in \mathbb{R}^n$. Este algoritmo calcula la solución de la ecuación $T\bar{x} = \bar{b}$

$$1) \left\{ \begin{array}{l} \text{Para } i = n, n-1, \dots, 1 \\ 1) \ x_i = t_{ii}^{-1} \left(b_i - \sum_{j=i+1}^n t_{ij} x_j \right) \end{array} \right.$$

El algoritmo 1.3 para resolver un sistema triangular es algunas veces llamado *Método de sustitución hacia atrás*. Matemáticamente, el algoritmo puede fallar solamente cuando algún $t_{ii} = 0$, en cuyo caso t_{ii}^{-1} no estaría definido. Por el teorema 1.2 esto puede suceder solamente cuando T es singular. En otras palabras, si T es no singular, el algoritmo 1.3 puede ser completado.

Una importante consideración práctica es la posibilidad -- que el algoritmo falle por sobre-carga del computador, aunque T sea no singular. Por ejemplo, si

$$T = \begin{bmatrix} 10^{-60} & 1 \\ 0 & 10^{-60} \end{bmatrix}$$

y $\bar{b} = [0, 1]^t$, entonces $\bar{x} = [-10^{120}, 10^{60}]^t$. En un computador cuyo orden de punto flotante tiene "caracteres" limitado por 100, la primera componente de \bar{x} no puede ser representada como un número de punto flotante y cualquier intento de ejecutar el algoritmo 1.3 para este dato resultará una sobre-carga.

El algoritmo 1.3 no ocupa los elementos de la parte infe--

rior del arreglo T . Así esta parte está libre para otros propósitos, es decir, para almacenar los elementos de una matriz estrictamente triangular inferior.

Finalmente note que el algoritmo 1.3 requiere alrededor de $\frac{n^2}{2}$ multiplicaciones.

El algoritmo 1.3 puede ser adaptado para calcular la inversa S de una matriz triangular superior T . Específicamente si S la particionamos en columnas, es decir, $S = [S_1, S_2, \dots, S_n]$, entonces la ecuación $TS = I$ es equivalente al conjunto de ecuaciones

$$TS_k = \bar{e}_k \quad (k = 1, 2, \dots, n)$$

Cada ecuación puede ser resuelta por el algoritmo 1.3 para obtener la inversa S . Sin embargo este es un procedimiento ineficiente, para la mitad de los elementos de S que se sabe que son ceros y no tiene sentido calcularlos. Para evitar esta dificultad, sea $\bar{S}'_k = [s_{1k}, s_{2k}, \dots, s_{kk}]^t$. Entonces es fácil verificar que

$$T^{[k]} \bar{S}'_k = \bar{e}_k \quad (1.6)$$

donde \bar{e}_k es ahora un k -vector. La ecuación (1.6) es un sistema triangular superior de orden k para los elementos distintos de cero de S_k , el cual puede ser resuelto por el algoritmo 1.3. Si esto es dado para cada k , resulta aquí un algoritmo para invertir una matriz triangular superior. Por razones que serán obvias mas tarde, resolveremos las ecuaciones en el orden $k = n, n-1, \dots, 1$ (conf. al algoritmo 1.5).

ALGORITMO 1.4

Sea $T \in \mathbb{R}^{n \times n}$ una matriz triangular superior y no singular. Este algoritmo calcula T^{-1} .

$$1) \left\{ \begin{array}{l} \text{Para } k = n, n-1, \dots, 1 \\ 1) s_{kk} = t_{kk}^{-1} \\ 2) s_{ik} = -t_{ii}^{-1} \sum_{j=i+1}^n t_{ij} s_{jk} \quad (i = k-1, k-2, \dots, 1). \end{array} \right.$$

En este algoritmo hemos tomado ventaja del hecho que todas las componentes excepto las últimas de \bar{e}_k son cero. Como el algoritmo 1.3, este algoritmo está matemáticamente bien definido cuando T es *no singular*. Practicamente la sobre carga puede impedir su ejecución completa. Alguna precisión puede ser ganada por la acumulación de productos internos en doble precisión.

Cuando no es requerido guardar los elementos de la matriz T , ellos pueden ser sobre-escritos en el lugar de los elementos de $S = T^{-1}$. En efecto, en el algoritmo 1.4, una vez t_{ik} ha sido usado para calcular s_{ik} éste ya no se necesita, y por lo tanto podemos escribir s_{ik} en el lugar de t_{ik} . Esto lo vemos en el siguiente algoritmo para invertir T en su propio arreglo.

ALGORITMO 1.5

Sea $T \in \mathbb{R}^{n \times n}$ una matriz no singular triangular superior. Este algoritmo sobre escribe T con su inversa S .

$$1) \left\{ \begin{array}{l} \text{Para } k = n, n-1, \dots, 1 \\ 1) t_{kk} \leftarrow s_{kk} = t_{kk}^{-1} \\ 2) t_{ik} \leftarrow s_{ik} = -t_{ii}^{-1} \sum_{j=i+1}^n t_{ij} s_{jk} \quad (i=k-1, k-2, \dots, 1) \end{array} \right.$$

Ambos algoritmos 1.4 y 1.5 requieren alrededor de $n^3/6$ multiplicaciones las cuales difieren en la cantidad de multiplicaciones para el algoritmo 1.3 por un factor de $\frac{n}{3}$.

Este hecho tiene consecuencias importantes para la solución de sistemas triangulares.

Un simple esquema para resolver (1.3) podría ser como sigue:

$$1) \quad S = T^{-1}$$

$$2) \quad \bar{x} = S\bar{b}$$

cuando el primer paso es realizado por el algoritmo 1.5. Sin embargo, la expresión 1 requiere $n^3/6$ multiplicaciones, además que las $\frac{n^2}{2}$ multiplicaciones de la expresión 2 son insignificantes. Obviamente, cuando n sea grande, este proceso es mas caro que el algoritmo 1.3. En otras palabras, si solamente se requiere la solución de (1.3), no es necesario calcular T^{-1} .

La inversa de una matriz raras veces es requerida en computaciones matriciales. Cada vez que solicitemos calcular

$$\bar{x} = T^{-1}\bar{b}$$

podemos alternativamente calcular la solución de

$$T\bar{x} = \bar{b}$$

reconstruyendolo en tal forma se puede ahorrar un buen tiempo de máquina, como se muestra en el siguiente ejemplo.

EJEMPLO 1.6

Deseamos calcular la expresión

$$a = \bar{x}^t T^{-1} \bar{y} ,$$

donde T es una matriz triangular superior de orden n . En vez de invertir T , puede procederse como sigue:

- 1) Resolver la ecuación $T \bar{z} = \bar{y}$
- 2) Calcular $a = \bar{x}^t \bar{z}$

El simple hecho de calcular T^{-1} requiere $\frac{n^3}{6}$ multiplicaciones, - además los cálculos que se efectúan son insignificantes. El proceso alternativo, por otro lado, requiere solamente $\frac{n^2}{2}$ multiplicaciones.

EJEMPLO 1.7

Deseamos calcular

$$C = T^{-1} B$$

donde T es de orden n y $B \in \mathbb{R}^{n \times m}$. Si particionamos B y C por columnas, entonces

$$[C_1, C_2, \dots, C_m] = [T^{-1} B_1, T^{-1} B_2, \dots, T^{-1} B_m]$$

donde $C_i = T^{-1} B_i$; así las columnas de C pueden ser encontradas resolviendo los sistemas

$$T C_i = B_i \quad (i = 1, 2, \dots, m)$$

Si T es triangular superior, el simple cálculo requiere $\frac{n^3}{6}$ multiplicaciones para la inversión de T y otras $\frac{mn}{2}$ multiplicaciones para calcular $C = T^{-1} B$, dando un total de $(\frac{1}{6}n + \frac{1}{2}m)n^2$ multiplicaciones. El proceso alternativo requiere $\frac{mn^2}{2}$ multiplicaciones. Comparando esta cantidad de operaciones, vemos que el procedimiento que involucra T^{-1} siempre requiere mas trabajo que -

el proceso alternativo, aunque la diferencia se hace insignificante para $m \gg n$.

EJERCICIOS

1.1 Elaborar un algoritmo para encontrar la inversa de una matriz triangular inferior.

1.2 Utilizar el algoritmo 1.3 para resolver el sistema

$$\begin{aligned}2x_1 + 4x_2 - 2x_3 &= 1 \\- 3x_2 + 6x_3 &= 9 \\- 13x_3 &= 12\end{aligned}$$

1.3 Aplicar el algoritmo 1.4 para encontrar la inversa de la matriz

$$\begin{bmatrix} 2 & 4 & -2 \\ 0 & -3 & 6 \\ 0 & 0 & -12 \end{bmatrix}$$

1.4 Elaborar un algoritmo para resolver un sistema triangular inferior.

1.5 Escribir un algoritmo para encontrar la inversa de una matriz triangular superior T de otra manera diferente al algoritmo 1.5 [sugerencia: Resolver los sistemas

$$\bar{S}_k^t T = \bar{e}_k^t (k = 1, 2, \dots, n)]$$

2. ELIMINACION GAUSSIANA.

En esta sección daremos algoritmos para reducir una matriz a la forma trapezoidal superior. El algoritmo básico, llamado -

Eliminación Gaussiana, minimiza el proceso de eliminación de -- las variables en un sistema de ecuaciones lineales. Además, los algoritmos pueden ser descritos en términos de operaciones ma-- triciales, y estas descripciones nos guían a las factorizacio-- nes que se verán en el capítulo III.

Para ilustrar estas ideas consideremos el siguiente siste-- ma de ecuaciones lineales:

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 &= 6 \\ x_1 - x_2 + 5x_3 &= 0 \\ 4x_1 + x_2 - 2x_3 &= 2 \end{aligned} \tag{2.1}$$

Si $1/2$ veces la primera ecuación es restada de la segunda y 2 - veces la primera restada de la tercera, resulta el sistema

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 &= 6 \\ - 3x_2 + 6x_3 &= -3 \\ - 7x_2 + 2x_3 &= -10 \end{aligned} \tag{2.2}$$

en el cual la variable x_1 no aparece en la segunda y tercera -- ecuación. Además la variable x_2 puede ser eliminada de la terce-- ra ecuación sustrayendo $7/3$ veces la segunda ecuación de la ter-- cera y obtenemos

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 &= 6 \\ - 3x_2 + 6x_3 &= -3 \\ - 12x_3 &= -3 \end{aligned} \tag{2.3}$$

El sistema (2.3) es triangular superior y puede ser resuelto - por las técnicas de la sección 1. Obviamente esta idea de la - eliminación puede ser extendida para dar un método general para

resolver sistemas de cualquier orden.

Sean $A = A_1$, A_2 y A_3 las matrices de los sistemas (2.1), (2.2), y (2.3), respectivamente; por ejemplo,

$$A_2 = \begin{bmatrix} 2 & 4 & -2 \\ 0 & -3 & 6 \\ 0 & -7 & -2 \end{bmatrix}$$

Entonces A_2 es obtenida de A_1 restando $1/2$ la primera fila de A_1 de la segunda fila y 2 veces la primera fila de la tercera. De la discusión de multiplicación matricial en la sección 4 del capítulo anterior, sabemos que A_2 puede ser obtenida premultiplicando A_1 por una matriz adecuada, y en efecto es fácil verificar que

$$A_2 = M_1 A_1 \tag{2.4}$$

donde

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$$

similarmente

$$A_3 = M_2 A_2 \tag{2.5}$$

donde

$$M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -7/3 & 1 \end{bmatrix}$$

Combinando (2.4) y (2.5), vemos que

$$A_3 = M_2 M_1 A \quad (2.6)$$

Así para resolver el sistema $A\bar{x} = \bar{b}$ necesitamos solamente calcular el vector $\bar{c} = M_2 M_1 \bar{b}$ y resolver el sistema triangular superior

$$A_3 \bar{x} = M_2 M_1 A \bar{x} = M_2 M_1 \bar{b} = \bar{c}$$

Además, la expresión (2.6) puede ser rearmada para dar una factorización de A en una matriz triangular inferior y una matriz triangular superior. Como las M_i son no singulares, tenemos de (2.6) que

$$A = M_1^{-1} M_2^{-1} A_3 \quad (2.7)$$

Como M_1 y M_2 son matrices triangulares unitarias inferiores, el producto de sus inversas también lo es. Por lo tanto, si hacemos $L = M_1^{-1} M_2^{-1}$ y $U = A_3$, la ecuación (2.7) se convierte en

$$A = LU,$$

donde L es triangular unitaria inferior y U es triangular superior. Como

$$M_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1/2 & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 7/3 & 1 \end{bmatrix},$$

L es el producto de matrices triangulares muy simples.

La discusión de arriba depende esencialmente de las matrices M_i , las cuales son llamadas *Matrices Elementales Triangulares Inferiores*. Comencemos nuestra exposición formal de *Eliminación*

ción Gaussiana con la discusión de sus propiedades.

DEFINICION 2.1

Una matriz elemental triangular inferior de orden n y de índice k , es una matriz de la forma

$$M = I_n - \bar{m} \bar{e}_k^t,$$

donde

$$\bar{e}_i^t \bar{m} = 0 \quad (i = 1, 2, \dots, k) \quad (2.8)$$

Las condiciones (2.8) dicen que las primeras k componentes de \bar{m} son ceros; esto es, \bar{m} tiene la forma

$\bar{m} = [0, 0, \dots, 0, m_{k+1}, m_{k+2}, \dots, m_n]^t$. En general una matriz elemental triangular inferior tiene la forma

$$M_k = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & \dots & 0 \\ 0 & 0 & \dots & -m_{k+1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & -m_n & \dots & 1 \end{bmatrix}$$

Así una matriz elemental triangular inferior es una matriz identidad con algunos elementos adicionales distintos de cero en la k -ésima columna bajo la diagonal.

Las matrices elementales triangulares inferiores son fáciles de invertir.

TEOREMA 2.2:

Sea $M = I - \bar{m} \bar{e}_k^t$ una matriz elemental triangular inferior. Entonces

$$M^{-1} = I + \bar{m} \bar{e}_k^t$$

PRUEBA

Sea $X = I + \bar{m} \bar{e}_k^t$. Entonces

$$\begin{aligned} MX &= (I - \bar{m} \bar{e}_k^t) (I + \bar{m} \bar{e}_k^t) \\ &= I - \bar{m} \bar{e}_k^t + \bar{m} \bar{e}_k^t - \bar{m} \bar{e}_k^t \bar{m} \bar{e}_k^t \\ &= I - \bar{m} (\bar{e}_k^t \bar{m}) \bar{e}_k^t \end{aligned}$$

Como M es una matriz elemental triangular inferior, $\bar{e}_k^t \bar{m} = 0$. -
Por lo tanto $MX = I$ y X es la inversa de M .

▲

El significado computacional de las matrices elementales triangulares inferiores es que pueden ser usadas para introducir componentes cero en un vector.

TEOREMA 2.3

Sea $\bar{e}_k^t \bar{x} = x_k \neq 0$. Entonces existe una única matriz elemental triangular inferior M de índice k tal que

$$M \bar{x} = [x_1, x_2, \dots, x_k, 0, \dots, 0]^t \quad (2.9)$$

PRUEBA

Tenemos M de la forma $M = I - \bar{m} \bar{e}_k^t$. Como M es de índice k , debemos tener que

$$m_i = 0 \quad (i = 1, 2, \dots, k) \quad (2.10)$$

Ahora

$$\begin{aligned} M \bar{x} &= (I - \bar{m} \bar{e}_k^t) \bar{x} \\ &= \bar{x} - (\bar{m} \bar{e}_k^t) \bar{x} \\ &= \bar{x} - \bar{m} (\bar{e}_k^t \bar{x}). \end{aligned}$$

Como las $n-k$ componentes de $M\bar{x}$ son ceros debemos tener que

$$x_i - x_k m_i = 0 \quad (i = k+1, k+2, \dots, n)$$

y si $x_k \neq 0$

$$m_i = \frac{x_i}{x_k} \quad (i = k+1, k+2, \dots, n) \quad (2.11)$$

así, si M existe, está determinada en forma única por (2.10) y (2.11). Por otro lado, si $x_k \neq 0$, entonces el vector \bar{m} puede -- ser determinado de (2.10) y (2.11), y es fácil verificar que la matriz asociada $M = I - \bar{m}\bar{e}_k^t$ satisface (2.9)

▲

Así cuando $x_k \neq 0$, podemos multiplicar \bar{x} por una matriz elemental triangular inferior de tal manera que las últimas componentes de \bar{x} sean reemplazadas por ceros y las otras componentes se dejan inalteradas.

Cuando $x_k = 0$, tal matriz no existe, a menos que x_{k+1}, \dots, x_n sean también ceros, en cuyo caso cualquier matriz elemental triangular inferior de índice k hará el trabajo. Será notorio que las ecuaciones (2.10) y (2.11) definen un algoritmo para calcular \bar{m} y por consiguiente M .

En el método de eliminación Gaussiana para reducir una matriz a la forma trapezoidal superior, la matriz es premultiplicada por una secuencia de matrices elementales triangulares inferiores, cada una escogida para introducir una columna con ceros bajo la diagonal.

Sea $A_1 = A$ una matriz $m \times n$. Si $a_{11}^{(1)} = a_{11}$ es distinto de -

cero, entonces por el teorema 2.3 existe una única matriz elemental triangular inferior M_1 de índice 1 que elimina los $m-1$ elementos de la primera columna de A_1 . Si A_1 es premultiplicada por M_1 , resulta una matriz A_2 de la forma

$$A_2 = M_1 A_1 = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \dots & a_{2n}^{(2)} \\ 0 & a_{32}^{(2)} & a_{33}^{(2)} & \dots & a_{3n}^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & a_{m2}^{(2)} & a_{m3}^{(2)} & \dots & a_{mn}^{(2)} \end{bmatrix}$$

El número que aparece entre paréntesis en la parte superior de cada elemento de una fila, nos indicará las transformaciones realizadas.

Note que la primera fila de A_2 es la misma que la primera fila de A_1 .

En general supóngase que $A_k = M_{k-1} \dots M_1 A_1$ tiene la forma -

$$A_k = \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & A_{22}^{(k)} \end{bmatrix}$$

donde $A_{11}^{(k)}$ es una matriz triangular superior de orden $k-1$ y para $i = 1, 2, \dots, k-1$ la i -ésima fila de A_k es igual a la i -ésima fila de A_i . Por ejemplo, con $m = n = 5$ tenemos

$$A_3 = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & a_{14}^{(1)} & a_{15}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & a_{24}^{(2)} & a_{25}^{(2)} \\ 0 & 0 & a_{33}^{(3)} & a_{34}^{(3)} & a_{35}^{(3)} \\ 0 & 0 & a_{43}^{(3)} & a_{44}^{(3)} & a_{45}^{(3)} \\ 0 & 0 & a_{53}^{(3)} & a_{54}^{(3)} & a_{55}^{(3)} \end{bmatrix}$$

Si $a_{kk}^{(k)} \neq 0$, entonces existe una matriz elemental triangular inferior M_k tal que anula los últimos $m-k$ elementos de la k -ésima columna de A_k . Una matriz semejante puede escribirse en la forma

$$M_k = \begin{bmatrix} I_{k-1} & 0 \\ 0 & M'_k \end{bmatrix}$$

donde M'_k es una matriz elemental triangular inferior de índice 1. Ahora

$$\begin{aligned} A_{k+1} = M_k A_k &= \begin{bmatrix} I_{k-1} & 0 \\ 0 & M'_k \end{bmatrix} \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & A_{22}^{(k)} \end{bmatrix} \\ &= \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & M'_k A_{22}^{(k)} \end{bmatrix} \end{aligned}$$

Así las primeras $k-1$ filas de A_{k+1} son iguales a las A_k . Como la primera fila de $A_{22}^{(k)}$ y $M'_k A_{22}^{(k)}$ son las mismas, las primeras k filas de A_k y A_{k+1} son iguales. También por construcción de M_k , $M'_k A_{22}^{(k)}$ tiene la forma

$$\begin{bmatrix} x & x & \dots & x \\ 0 & x & \dots & x \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ 0 & x & \dots & x \end{bmatrix}$$

Por lo tanto, la submatriz directriz principal de orden k de A_{k+1} es triangular superior. En otras palabras, A_{k+1} es un paso en el transcurso de la reducción.

La reducción terminará cuando nos salgamos de las filas o columnas. Cuando $m > n$, esto sucede después del n -ésimo paso. Cuando $m \leq n$, la matriz es de la forma trapezoidal inferior después del $(m-1)$ -ésimo paso. Por ejemplo cuando $m = 1$, no se requiere reducción, ya que un vector fila es trivialmente de la forma trapezoidal superior. Así la reducción termina con la matriz A_{r+1} , donde

$$r = \min\{m-1, n\}$$

La reducción de A es constructiva en el sentido que podemos especificar un algoritmo para el cálculo de A_k y M_k . Específicamente, la k -ésima columna de A_k es el vector

$$\bar{c}_k^{(k)} = \left[a_{1k}^{(1)}, a_{2k}^{(2)}, \dots, a_{kk}^{(k)}, \dots, a_{mk}^{(k)} \right]^t.$$

y M_k es la matriz elemental triangular inferior de índice k tal que

$$M_k \bar{c}_k^{(k)} = \left[a_{1k}^{(1)}, a_{2k}^{(2)}, \dots, a_{kk}^{(k)}, 0, \dots, 0 \right]^t$$

Por el teorema 2.3, $M_k = I - \bar{m}_k \bar{e}_k^t$, donde

$$\bar{m}_k = [0, 0, \dots, 0, m_{k+1,k}, \dots, m_{mk}]^t$$

$$y \quad m_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} \quad (i = k+1, k+2, \dots, m) \quad (2.12)$$

Así M_k está determinada en forma única cuando $a_{kk}^{(k)} \neq 0$. En principio A_{k+1} puede ser calculado formando M_k y calculando el producto $M_k A_k$. Sin embargo, considerables ahorros en operaciones y almacenamientos pueden ser obtenidos tomando ventaja de la forma especial de M_k . En efecto,

$$\begin{aligned} A_{k+1} &= M_k A_k = (I - \bar{m}_k \bar{e}_k^t) A_k \\ &= A_k - \bar{m}_k \bar{e}_k^t A_k \end{aligned}$$

Ahora $\bar{e}_k^t A_k$ es la k -ésima fila de A_k . Por lo tanto la i -ésima fila de A_{k+1} puede ser formada restando m_{ik} veces la k -ésima fila de A_k de la i -ésima fila de A_k . Como $m_{1k} = m_{2k} = \dots = m_{kk} = 0$, solamente las filas $k+1, k+2, \dots, m$ son alteradas. En términos de elementos tenemos

$$a_{ik}^{(k+1)} = 0 \quad (i = k+1, k+2, \dots, m) \quad (2.13)$$

$$\begin{aligned} y \quad a_{ij}^{(k+1)} &= a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)} \\ &(i = k+1, k+2, \dots, m ; j = k+1, k+2, \dots, n) \quad (2.14) \end{aligned}$$

Las ecuaciones (2.12) al (2.14) especifican completamente los cálculos involucrados en la reducción. En cuanto al almacenamiento concierne, podemos hacer que los elementos de A_{k+1} se sobrescriban en los lugares correspondientes a los elementos de A_k . En aplicaciones tales como resolver sistemas lineales, -

es importante conocer las matrices M_k .

Después que el $a_{ik}^{(k)}$ es usado para calcular M_{ik} , éste no es utilizado de nuevo, y en efecto es reducida a cero en la transformación de A_k y A_{k+1} . Por lo tanto, si queremos almacenar solamente los elementos diferentes de cero de A_{k+1} , el número m_{ik} puede ser colocado en el lugar originalmente ocupado por a_{ik} .

Estas consideraciones pueden ser resumidas en el siguiente algoritmo.

ALGORITMO 2.4 (Eliminación Gaussiana)

Sea A una matriz de orden $m \times n$ con $m > 1$ y sea $r = \min\{m - 1, n\}$. Este algoritmo sobrescribe A con la matriz trapezoidal superior A_{r+1} . Los multiplicadores m_{ik} se sobrescriben en los a_{ik} .

$$1) \left\{ \begin{array}{l} \text{Para } k = 1, 2, \dots, r \\ 1) a_{ik} \leftarrow m_{ik} = \frac{a_{ik}}{a_{kk}} \quad (i = k+1, k+2, \dots, m) \\ 2) a_{ij} \leftarrow a_{ij} - m_{ik} a_{kj} \quad (i = k+1, \dots, m; j = k+1, \dots, n) \end{array} \right.$$

El número de multiplicaciones para este algoritmo está dado por

$$\begin{aligned} \sum_{j=1}^r (m-j)(n-j) &= \sum_{j=1}^r (mn - mj - nj + j^2) \\ &= \frac{r^3}{3} - (m+n) \frac{r^2}{2} + mn r - \left(\frac{3m+3n-3r-1}{6} \right) r \end{aligned}$$

Cuando r es bastante grande el número de multiplicaciones es -- aproximadamente

$$\frac{r^3}{3} - (m+n) \frac{r^2}{2} + mn r$$

Si $m = n$, entonces $r = n - 1$, y el número de multiplicaciones sería.

$$\sum_{j=1}^{n-1} (n-j)(n-j) = \sum_{j=1}^{n-1} (n-j)^2$$

desarrollando el miembro derecho de la expresión anterior, obtenemos

$$\sum_{j=1}^{n-1} (n-j)^2 = \sum_{j=1}^{n-1} j^2 \quad ,$$

$$\text{luego} \quad \sum_{j=1}^{n-1} j^2 = \frac{n^3}{3} - \frac{n^2}{2} + \frac{n}{6} \quad .$$

Cuando n es considerablemente grande, el número de multiplicaciones es aproximadamente $\frac{n^3}{3}$.

La generalización para resolver sistema de ecuaciones lineales por el método de eliminación Gaussiana con sustitución hacia atrás está dada en el siguiente algoritmo.

ALGORITMO 2.5

Sea el sistema lineal

$$A \bar{x} = \bar{b} \tag{2.15}$$

es decir

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 = a_{1,n+1}$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 = a_{2,n+1}$$

⋮

⋮

⋮

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m = a_{m,n+1}$$

este algoritmo resuelve el sistema (2.15).

1) Construir la matriz aumentada

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} & a_{2,n+1} \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ \cdot & \cdot & \cdot & & & \cdot \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} & a_{m,n+1} \end{array} \right]$$

2) Hacer $r = \min\{m - 1, n\}$

3) Para $k = 1, 2, \dots, r$

$$1) a_{ik} \leftarrow m_{ik} = \frac{a_{ik}}{a_{kk}} \quad (i = k+1, k+2, \dots, m)$$

$$2) a_{ij} \leftarrow a_{ij} - m_{ik} a_{kj} \quad (i = k+1, \dots, m; j = k+1, \dots, n, n+1)$$

4) Si $m > n$, ir a paso 6)

5) Si $a_{mm} \neq 0$, ir a paso 9)

6) Si $a_{i,n+1} = 0$ para $i = r+1, r+2, \dots, m$, imprimir el mensaje "el sistema no tiene solución" e ir al paso 10)

7) Hacer para $i = r, r-1, \dots, 1$

$$1) x_i = a_{ii}^{-1} (a_{i,n+1} - \sum_{j=i+1}^r a_{ij} x_j)$$

8) Ir a paso 10)

9) Para $i = m, m-1, \dots, 1$

$$1) x_i = a_{ii}^{-1} (a_{i,n+1} - \sum_{j=i+1}^m a_{ij} x_j)$$

10) El proceso se ha completado.

Nota: Este algoritmo va hacer aplicado a un sistema lineal de orden 3 en el ejemplo 4.1 de este capítulo.

Los números $a_{kk}^{(k)}$ ($k = 1, 2, \dots, r$) son llamados elementos de la reducción. Así como están dados los algoritmos (2.4) y (2.5) pueden ser completados si y solamente si todos los elementos pivotes son diferentes de cero. Es conveniente que la matriz original tenga una condición bajo la que los elementos pivotes estén garantizados que sean diferentes de cero. Tal condición está dada en el siguiente teorema.

TEOREMA 2.6

Los elementos pivotes $a_{ii}^{(i)}$ ($i = 1, 2, \dots, k$) son diferentes de cero si y solamente si las submatrices directrices -- principales $A^{[i]}$ ($i = 1, 2, \dots, k$) son no singulares.

PRUEBA

Esta prueba es por inducción en k . Para $k = 1$, el teorema es trivialmente cierto cuando $A^{[1]} = a_{11}^{(1)}$. Para el paso inductivo es suficiente asumir que $A^{[1]}, \dots, A^{[k-1]}$ son no singulares -- y mostrar que $A^{[k]}$ es no singular si y solo si $a_{kk}^{(k)} \neq 0$.

Por hipótesis inductiva, $a_{11}^{(1)}, a_{22}^{(2)}, \dots, a_{k-1, k-1}^{(k-1)} \neq 0$. Por -- tanto el algoritmo 2.4 puede ser completado hasta su $(k-1)$ -ésimo paso para obtener matrices triangulares inferiores M_i de índice i ($i = 1, 2, \dots, k-1$) tal que

$$A_k = M_{k-1}M_{k-2}\dots M_1A = \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & A_{22}^{(k)} \end{bmatrix}$$

donde $A_{11}^{(k)}$ es triangular superior con elementos en la diagonal $a_{11}^{(1)}, \dots, a_{k-1, k-1}^{(k-1)}$. Se deduce que $A_k^{[k]}$ es triangular superior --

con elementos en la diagonal $a_{11}^{(1)}, \dots, a_{k-1, k-1}^{(k-1)}, a_{kk}^{(k)}$. Como

$a_{11}^{(1)}, \dots, a_{k-1, k-1}^{(k-1)} \neq 0$, $A_k^{[k]}$ es no singular si y solo si

$a_{kk}^{(k)} \neq 0$. Ahora como M_1, \dots, M_{k-1} son triangulares inferiores --

$$A_k^{[k]} = M_{k-1}^{[k]} \dots M_1^{[k]} A^{[k]} \quad (\text{ejercicio 1.5.8}).$$

Como M_i es triangular inferior unitaria, lo es tambien -- $M_i^{[k]}$. Así $A^{[k]}$ es no singular si y solo si $A_k^{[k]}$ es no singular, lo que es cierto sí y solo sí $a_{kk}^{(k)} = 0$



En su k -ésimo paso el método de eliminación Gaussiana produce matrices triangulares inferiores M_i de índice $i (i=1, 2, \dots, k)$ tal que $A_{k+1} = M_k M_{k-1} \dots M_1 A$ tiene ceros debajo de los primeros k -elementos de la diagonal. Una vez iniciado, el método procede siempre que los elementos pivotes sean diferentes de cero. Esto sugiere que las matrices M_1, M_2, \dots, M_k son determinadas en forma única por el requerimiento que $M_k M_{k-1} \dots M_1 A$ tiene ceros bajo -- los primeros k elementos de la diagonal. El siguiente teorema -- nos da la unicidad de estas matrices elementales.

TEOREMA 2.7

Sea $A \in \mathbb{R}^{m \times n}$ y supóngase que $A^{[1]}, A^{[2]}, \dots, A^{[k]}$ son no singulares. Para $i = 1, 2, \dots, k$ sea M_i y N_i matrices elementales triangulares inferiores de índice i . Si $M_k M_{k-1} \dots M_1 A$ y $N_k N_{k-1} \dots N_1 A$ tienen ambas ceros debajo de sus primeros k elementos de la diagonal, entonces $M_i = N_i (i = 1, 2, \dots, k)$

PRUEBA

La prueba es por inducción en k . Para $k = 1$, M_1 y N_1 son matrices elementales triangulares inferiores de índice 1 que introduce ceros en los últimos $m - 1$ elementos de la primera columna de A . Como $a_{11} \neq 0$ estos requerimientos determinan en forma única a M_1 y N_1 (Teorema 2.3) las cuales son indudablemente iguales.

Para el paso inductivo, asumimos que $A^{[1]}, \dots, A^{[k]}$ son no singulares y que $M_k \dots M_1 A$ y $N_k \dots N_1 A$ tienen ambas ceros debajo sus primeros k -elementos de la diagonal.

Como M_k es una matriz elemental triangular inferior de índice k , también lo es M_k^{-1} , y es fácil verificar que

$$M_{k-1} \dots M_1 A = M_k^{-1} N_k \dots N_1 A$$

tiene ceros bajo sus primeros $k-1$ elementos de la diagonal. ---

Igualmente $N_{k-1} \dots N_1 A$ tiene ceros bajo sus primeros $k-1$ elementos de la diagonal. Por la hipótesis inductiva

$M_i = N_i$ ($i = 1, 2, \dots, k-1$). Por lo tanto $M_{k-1} \dots M_1 A = N_{k-1} \dots N_1 A = A_k$,

donde A_k es la matriz resultante de los $k-1$ pasos de la eliminación

Gaussiana. Por el teorema 2.6, $a_{kk}^{(k)} \neq 0$. Por tanto M_k y N_k

son las únicas matrices elementales triangulares inferiores de

índice k que introducen ceros debajo de la k -ésima elemento de

la diagonal de A_k . Se sigue que $M_k = N_k$.

▲

El algoritmo 2.4 se interrumpe en el paso 1.1 cuando el pivote $a_{kk} = 0$. Si algunos $a_{ik} \neq 0$ ($i = k+1, \dots, m$), entonces la -

reducción no puede ser continuada. Sin embargo, si $a_{ik} = 0$ ($i = k+1, \dots, m$), entonces A_k es ya cero bajo su k -ésimo elemento de la diagonal, y cualquier matriz elemental triangular inferior de índice k puede ser usada como M_k , así, si surge un elemento pivote que es cero en el paso k , es posible continuar la reducción, pero M_k, M_{k+1}, \dots, M_r no serían únicas.

EJEMPLO 2.8

Ilustremos varias situaciones que pueden surgir en la reducción en las siguientes matrices 2×2 . Note que $r=1$ y por lo tanto a_{11} es el factor determinante en la reducción.

1. $a_{11} \neq 0$

$$A_1 = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix}, \quad M_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

2. $a_{11} \neq 0$

$$A_1 = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}, \quad M_1 = \begin{bmatrix} 1 & 0 \\ -1 & 1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 2 \\ 0 & 0 \end{bmatrix}$$

Aquí A_1 es singular. Nótese que M_1 está determinada en forma única.

3. $a_{11} = 0$

$$A_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$

La reducción no puede continuarse.

4. $a_{11} = 0$

$$A_1 = \begin{bmatrix} 0 & 1 \\ 0 & 2 \end{bmatrix}$$

Aquí la reducción puede terminarse, pero no es única. En efecto para cualquier c podemos construir

$$M_1 = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}$$

dando

$$A_2 = \begin{bmatrix} 0 & 1 \\ 0 & 2+c \end{bmatrix}$$

Hemos visto arriba que la reducción de una matriz a la forma trapezoidal falla cuando el elemento pivote $a_{kk}^{(k)}$ es cero.

El siguiente ejemplo ilustra las consecuencias de calcular con un elemento pivote pequeño.

EJEMPLO 2.9

Sea

$$A_1 = \begin{bmatrix} 0.001 & 2.000 & 3.000 \\ -1.000 & 3.712 & 4.623 \\ -2.000 & 1.072 & 5.643 \end{bmatrix}$$

Luego utilizando solamente cuatro dígitos en la reducción de A a la forma triangular superior producirá las siguientes matrices.

$$M_1 = \begin{bmatrix} 1.000 & 0.000 & 0.000 \\ 1.000 & 1.000 & 0.000 \\ 2.000 & 0.000 & 1.000 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 0.001 & 2.000 & 3.000 \\ 0.000 & 2004. & 3005. \\ 0.000 & 4001. & 6006. \end{bmatrix} ,$$

$$M_2 = \begin{bmatrix} 1.000 & 0.000 & 0.000 \\ 0.000 & 1.000 & 0.000 \\ 0.000 & -1.997 & 1.000 \end{bmatrix} ,$$

$$A_3 = \begin{bmatrix} 0.001 & 2.000 & 3.000 \\ 0.000 & 2004. & 2005. \\ 0.000 & 0.000 & 5.000 \end{bmatrix} .$$

Hay razón de dudar en este cálculo, por una rigurosa cancelación ocurrida en el cálculo del elemento (3,3) de A_3 por la fórmula

$$\{ \ell [6006. - (1.997)(3005.)] \}$$

En efecto el valor correcto del elemento (3,3) con cuatro cifras es 5.922, lo que no está de acuerdo con el valor calculado en el segundo cálculo. Si este valor encontrado es empleado en siguientes cálculos, el resultado será en general exacto solamente en una o dos cifras.

El ejemplo anterior sugiere que el apareamiento de un pivote pequeño en el curso del algoritmo 2.4 puede causar problemas serios. Cuando se tiene un problema de esta naturaleza el sistema puede ser arreglado de nuevo, así que la reducción es efectuada con una matriz diferente para el cual no tengamos elementos pivotes pequeños. Por ejemplo, supóngase que la matriz en el ejemplo 2.9 ha surgido relacionada con la solución del --

sistema lineal.

$$\begin{aligned} 0.001 x_1 + 2.000 x_2 + 3.000 x_3 &= 1.000 \\ -1.000 x_1 + 3.712 x_2 + 4.623 x_3 &= 2.000 \\ -2.000 x_1 + 1.072 x_2 + 5.643 x_3 &= 3.000 \end{aligned}$$

Si la primera y la segunda ecuación son intercambiadas, resulta un sistema lineal cuya matriz

$$\begin{bmatrix} -1.000 & 3.712 & 4.623 \\ 0.001 & 2.000 & 3.000 \\ -2.000 & 1.072 & 5.643 \end{bmatrix}$$

puede ser reducida sin dificultad.

Las consideraciones anteriores sugieren que modifiquemos el algoritmo 2.4 para evitar los elementos pivotes pequeños intercambiando las filas y columnas de la matriz A . Específicamente si en la k -ésima etapa de la reducción, el elemento $a_{kk}^{(k)}$ es sumamente pequeño, podemos escoger algún otro elemento, es decir $a_{y_k, z_k}^{(k)} \neq 0$ como pivote. Si intercambiamos las filas k y y_k , y también las columnas k y z_k , el resultado es para mover $a_{y_k, z_k}^{(k)}$ a la (k, k) posición, y la reducción puede ser continuada con el nuevo pivote. Obviamente, debemos tener que $y_k \geq k$ y $z_k \geq k$, de lo contrario los intercambios distorcionarán los ceros previamente introducidos en la reducción.

En términos matriciales el algoritmo modificado puede ser descrito como sigue. En la k -ésima etapa, A_k es premultiplicada y posmultiplicada, respectivamente, por las permutaciones ele--

mentales I_{k,y_k} y I_{k,z_k} (conforme ejem. 1.4.13) y también pre--
multiplicada por una matriz elemental triangular inferior M_k de
índice k escogida para introducir una nueva columna de ceros. -
Así

$$A_{k+1} = M_k I_{k,y_k} A_k I_{k,z_k}$$

Por simplicidad de notación, denotaremos la permutación --
elemental I_{k,y_k} y I_{k,z_k} por P_k y Q_k , respectivamente, así que

$$A_{k+1} = M_k P_k A_k Q_k \quad (2.16)$$

La incorporación de intercambios en el algoritmo 2.4 com--
plica la expresión para A_k . Se pensaría que sería más difícil -
analizar el algoritmo modificado. Afortunadamente, como muestra
el siguiente teorema, el algoritmo modificado es equivalente a
hacer todos los intercambios primero y entonces aplicar el algo--
ritmo 2.4.

TEOREMA 2.10

Sea el algoritmo modificado aplicado a A produciendo
matrices Q_i, P_i, M_i , y A_{i+1} ($i = 1, 2, \dots, k$).

Sea

$$A' = P_k P_{k-1} \dots P_1 A Q_1 Q_2 \dots Q_k$$

El algoritmo 2.4 puede ser aplicado a A' por medio de su -
 k -ésimo paso. Si M'_i y A'_{i+1} ($i = 1, 2, \dots, k$), son las matrices -
producidas por el algoritmo 2.4, entonces

$$A'_{k+1} = A_{k+1}$$

$$Y \quad M'_i = P_k P_{k-1} \dots P_{i+1} M_i P_{i+1} \dots P_{k-1} P_k \quad (2.17)$$

PRUEBA

De la ecuación (2.16) se sigue que

$$A_{k+1} = M'_k P_k M'_{k-1} P_{k-1} \dots M'_1 P_1 A Q_1 Q_2 \dots Q_k$$

ya que $P_i^2 = I$, es decir que $P_i^{-1} = P_i$ por lo tanto

$$A_{k+1} = M'_k M'_{k-1} \dots M'_1 A' \quad (2.18)$$

donde M'_i está definida por (2.17). Ahora es fácil verificar que si M_k es una matriz elemental triangular inferior de índice k y $i, j > k$, entonces $I_{ij} M_k I_{ij}$ es también una matriz elemental triangular inferior de índice k . Se sigue que M'_i es una matriz elemental triangular inferior de índice k . Los primeros k elementos de la diagonal de A_{k+1} son diferentes de cero y por lo tanto las primeras k submatrices principales directrices de A' son no singulares. Así por el teorema 2.6, M'_1, M'_2, \dots, M'_k y A_{k+1} son el resultado de aplicar k pasos del algoritmo 2.4 a A'



El ejemplo 2.9 sugiere que debemos escoger el pivote tan grande como sea posible. En otras palabras deseamos escoger y_k y z_k , ambos no menores que k , así que

$$\left| a_{y_k, z_k}^{(k)} \right| \geq \left| a_{ij}^{(k)} \right| \quad (i = k, k+1, \dots, m ; j = k, k+1, \dots, n)$$

Existe la posibilidad que este máximo elemento sea cero; sin embargo, esto puede suceder solamente cuando las últimas $m-k$ filas de A_k sean ceros, y en este caso A_k ya está en la --

forma trapezoidal superior. El algoritmo modificado con los pivotes escogidos anteriormente es llamado eliminación Gaussiana con pivotamiento completo.

ALGORITMO 2.11 (Eliminación Gaussiana con pivotamiento completo)

Sea $A \in \mathbb{R}^{m \times n}$ con $m > 1$. Este algoritmo calcula la solución del sistema (2.15). Los índices de las filas y columnas permutadas son y_k y z_k , respectivamente.

1) Construir la matriz aumentada

$$\left[\begin{array}{cccccc|c} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} & a_{2,n+1} \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & \cdot & & \cdot & \cdot \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} & a_{m,n+1} \end{array} \right]$$

2) Hacer $r = \min\{m - 1, n\}$

3) Para $k = 1, 2, \dots, r$

1) Encontrar $y_k, z_k \geq k$ tal que

$$|a_{y_k, z_k}| = \max \{|a_{ij}| : i, j \geq k\}$$

2) $x_k \leftarrow x_{z_k}$

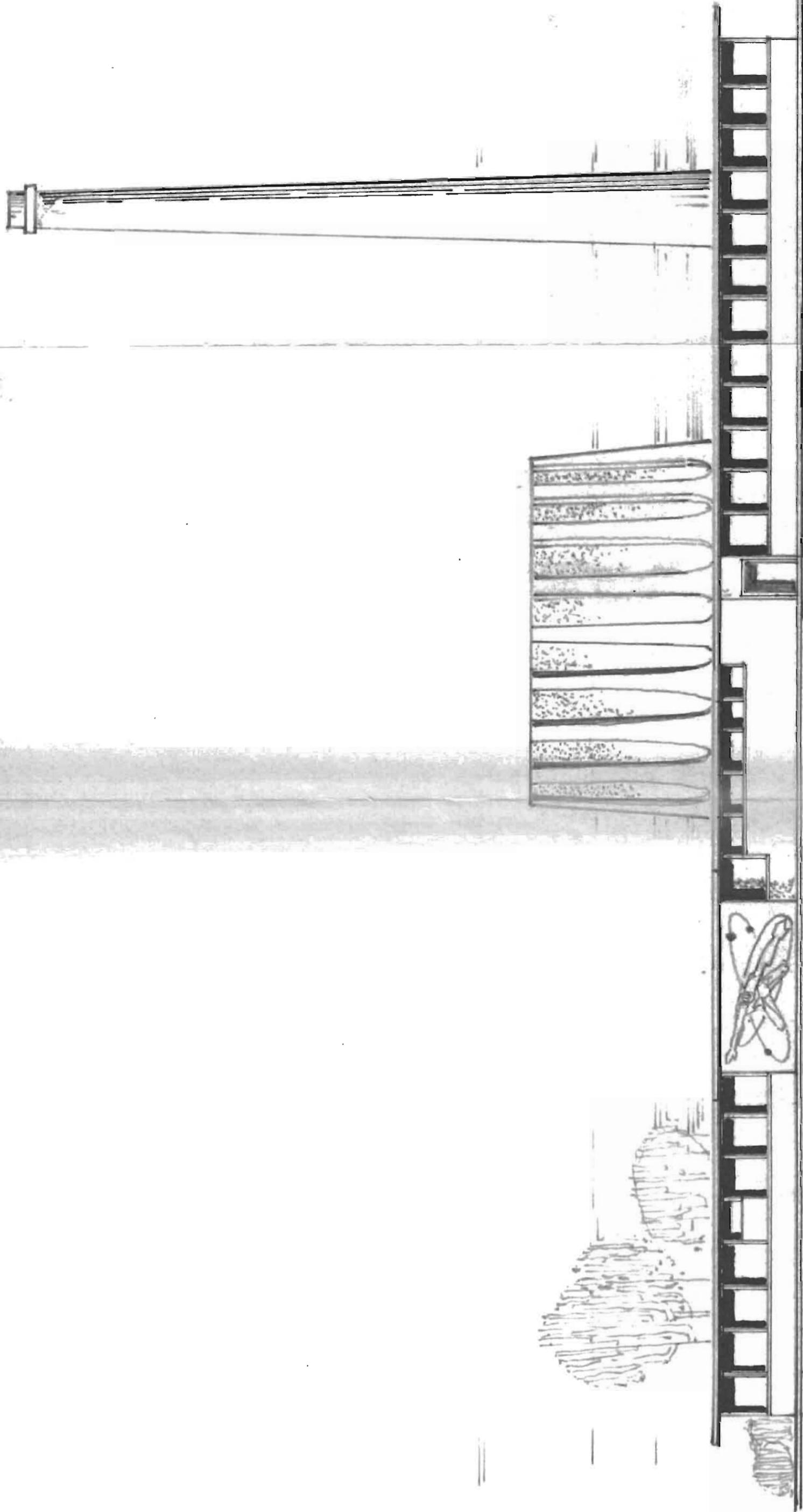
3) Si $a_{y_k, z_k} = 0$, hacer $r = k - 1$ e ir a paso 10.

4) $a_{kj} \leftrightarrow a_{y_k, j}$ ($j = k, k+1, \dots, n+1$)

5) $a_{ik} \leftrightarrow a_{i, z_k}$ ($i = 1, 2, \dots, m$)

6) $a_{ik} \leftarrow m_{ik} = \frac{a_{ik}}{a_{kk}}$ ($i = k+1, k+2, \dots, m$)

7) $a_{ij} \leftarrow a_{ij} - m_{ik} a_{kj}$ ($i = k+1, \dots, m ; j = k+1, \dots, n+1$)



ELEVACION PRINCIPAL . ESCALA 1:200

- 4) Si $m > n$, ir a paso (6)
- 5) Si $a_{mm} \neq 0$, ir a paso (9).
- 6) Si $a_{i,n+1} \neq 0$ para $i = r+1, r+2, \dots, m$, imprimir el mensaje "el sistema no tiene solución" e ir a paso (10).
- 7) Hacer para $i = r, r-1, \dots, 1$

$$\left\{ \begin{array}{l} 1) x_i = a_{ii}^{-1} (a_{i,n+1} - \sum_{j=i+1}^r a_{ij} x_j) \\ 2) x_{z_i} \leftarrow x_i \end{array} \right.$$

- 8) Ir a paso (10)
- 9) Para $i = m, m-1, \dots, 1$

$$\left\{ \begin{array}{l} 1) x_i = a_{ii}^{-1} (a_{i,n+1} - \sum_{j=i+1}^m a_{ij} x_j) \\ 2) x_{z_i} \leftarrow x_i \end{array} \right.$$

- 10) El proceso ha terminado

▲

Para mostrar la efectividad del algoritmo anterior resolveremos un sistema de ecuaciones lineales que tiene el número de ecuaciones menor que el número de variables, para el caso, el sistema es de orden 3×4 .

EJEMPLO 2.12

Aplicando el algoritmo 2.11, resolver el siguiente sistema:

$$3x_1 + 6x_2 - x_3 + 8x_4 = 10$$

$$2x_1 + 4x_2 + 4x_3 + 3x_4 = 9$$

$$x_1 + 2x_2 - x_3 + 3x_4 = 3$$

1) La matriz aumentada es

$$\left[\begin{array}{cccc|c} 3 & 6 & -1 & 8 & 10 \\ 2 & 4 & 4 & 3 & 9 \\ 1 & 2 & -1 & 3 & 3 \end{array} \right]$$

2) $r = \min \{m - 1, n\}$

$$r = \min \{3 - 1, 4\}$$

$$r = 2$$

3) Para $k = 1, 2$.

$$\boxed{k = 1}$$

1) $y_1, z_1 \geq 1$ tal que $|a_{y_1 z_1}| = \max \{|a_{ij}|; i, j \geq 1\} = a_{14} = 8$

2) $x_1 \leftarrow x_4$

3) $a_{14} = 8 \neq 0$

4) $a_{1j} \longleftrightarrow a_{1j}$ ($j = 1, 2, 3, 4, 5$); (la primera fila no cambia).

5) $a_{i1} \longleftrightarrow a_{i4}$ ($i = 1, 2, 3$)

$$\left. \begin{array}{l} a_{11} \longleftrightarrow a_{14} \\ a_{21} \longleftrightarrow a_{24} \\ a_{31} \longleftrightarrow a_{34} \end{array} \right\} \Rightarrow \left[\begin{array}{cccc|c} 8 & 6 & -1 & 3 & 10 \\ 3 & 4 & 4 & 2 & 9 \\ 3 & 2 & -1 & 1 & 3 \end{array} \right]$$

6) $a_{i1} \longleftrightarrow m_{i1} = \frac{a_{i1}}{a_{11}}$ ($i = 2, 3$)

$$\left. \begin{array}{l} a_{21} \longleftrightarrow m_{21} = \frac{a_{21}}{a_{11}} = \frac{3}{8} \\ a_{31} \longleftrightarrow m_{31} = \frac{a_{31}}{a_{11}} = \frac{3}{8} \end{array} \right\} \Rightarrow \left[\begin{array}{cccc|c} 8 & 6 & -1 & 3 & 10 \\ \frac{3}{8} & 4 & 4 & 2 & 9 \\ \frac{3}{8} & 2 & -1 & 1 & 3 \end{array} \right]$$

7) $a_{ij} \longleftrightarrow a_{ij} - m_{i1} a_{1j}$ ($i = 2, 3$; $j = 2, 3, 4, 5$)

$$i = 2$$

$$a_{22} \longleftarrow a_{22} - m_{21}a_{12} = 4 - \frac{3}{8}(6) = \frac{7}{4}$$

$$a_{23} \longleftarrow a_{23} - m_{21}a_{13} = 4 - \frac{3}{8}(-1) = \frac{35}{8}$$

$$a_{24} \longleftarrow a_{24} - m_{21}a_{14} = 2 - \frac{3}{8}(3) = \frac{7}{8}$$

$$a_{25} \longleftarrow a_{25} - m_{21}a_{15} = 9 - \frac{3}{8}(10) = \frac{21}{4}$$

$$i = 3$$

$$a_{32} \longleftarrow a_{32} - m_{31}a_{12} = 2 - \frac{3}{8}(6) = -\frac{1}{4}$$

$$a_{33} \longleftarrow a_{33} - m_{31}a_{13} = -1 - \frac{3}{8}(-1) = -\frac{5}{8}$$

$$a_{34} \longleftarrow a_{34} - m_{31}a_{14} = 1 - \frac{3}{8}(3) = -\frac{1}{8}$$

$$a_{35} \longleftarrow a_{35} - m_{31}a_{15} = 3 - \frac{3}{8}(10) = -\frac{3}{4}$$

$$\Rightarrow \left[\begin{array}{cccc|c} 8 & 6 & -1 & 3 & 10 \\ \textcircled{3/8} & 7/4 & 35/8 & 7/8 & 21/4 \\ \textcircled{3/8} & -1/4 & -5/8 & -1/8 & -3/4 \end{array} \right]$$

$$\boxed{k = 2}$$

$$1) y_2, z_2 \geq 2 \text{ tal que } |a_{y_2, z_2}| = \max\{|a_{ij}| : i, j \geq 2\} = a_{23} = \frac{35}{8}$$

$$2) x_2 \longleftarrow x_3$$

$$3) a_{23} = \frac{35}{8} \neq 0$$

$$4) a_{2j} \longleftarrow a_{2j} \quad (j = 2, 3, 4, 5) ; \text{ (la segunda fila no cambia)}$$

$$5) a_{i2} \longleftarrow a_{i3} \quad (i = 1, 2, 3)$$

$$\left. \begin{array}{l} a_{12} \longleftarrow a_{13} \\ a_{22} \longleftarrow a_{23} \\ a_{32} \longleftarrow a_{33} \end{array} \right\} \Rightarrow \left[\begin{array}{cccc|c} 8 & -1 & 6 & 3 & 10 \\ \textcircled{3/8} & 35/8 & 7/4 & 7/8 & 21/4 \\ \textcircled{3/8} & -5/8 & -1/4 & -1/8 & -3/4 \end{array} \right]$$

$$6) a_{i2} \longleftarrow m_{i2} = \frac{a_{i2}}{a_{22}} \quad (i = 3)$$

$$a_{32} \longleftarrow m_{32} = \frac{a_{32}}{a_{22}} = \frac{-5/8}{35/8} = -\frac{1}{7}$$

$$7) a_{ij} \longleftarrow a_{ij} - m_{i2} a_{2j} \quad (i = 3 ; j = 3, 4, 5)$$

$$a_{33} \longleftarrow a_{33} - m_{32} a_{23} = -\frac{1}{4} - \left(-\frac{1}{7}\right) \left(\frac{7}{4}\right) = 0$$

$$a_{34} \longleftarrow a_{34} - m_{32} a_{24} = -\frac{1}{8} - \left(-\frac{1}{7}\right) \left(\frac{7}{8}\right) = 0$$

$$a_{35} \longleftarrow a_{35} - m_{32} a_{25} = -\frac{3}{4} - \left(-\frac{1}{7}\right) \left(\frac{21}{4}\right) = 0$$

$$\Rightarrow \left[\begin{array}{cccc|c} 8 & -1 & 6 & 3 & 10 \\ \textcircled{3/8} & 35/8 & 7/4 & 7/8 & 21/4 \\ \textcircled{3/8} & \textcircled{-1/7} & 0 & 0 & 0 \end{array} \right]$$

$$4) m = 3 \wedge n = 4 \Rightarrow 3 \neq 4$$

$$5) a_{mm} = a_{33} = 0$$

$$6) a_{i, n+1} = 0 \quad \text{cuando } i = 3 \\ a_{35} = 0$$

$$7) \text{ Para } i = r, r-1, \dots, 1$$

$$1) i = 2, 1$$

$$x_2 = \frac{1}{a_{22}} (a_{25}) = \frac{21/4}{35/8} = \frac{42}{35} = \frac{6}{5}$$

$$2) x_3 \longleftarrow x_2$$

$$1) x_1 = \frac{1}{a_{11}} (a_{15} - a_{12} x_2)$$

$$x_1 = \frac{1}{8} \left[10 - (-1) \left(\frac{42}{35} \right) \right] = \frac{1}{8} \cdot \frac{56}{5} = \frac{7}{5}$$

$$2) x_4 \longleftarrow x_1$$

$$\implies x_3 = \frac{6}{5} \wedge x_4 = \frac{7}{5}$$

NOTA: Los números encerrados en círculos en cada matriz son los números de los m_{ik} .

En la práctica, el paso 3.1 del algoritmo 2.11 puede -- consumir una buena porción de tiempo después que busca el mas -- grande alrededor de los $(m - k + 1)(n - k + 1)$ elementos de A . Una alternativa frecuentemente usada es buscar solamente la k -ésima columna para el elemento mayor y efectuar intercambio de fila -- para llevarla a la posición pivotal. Si en el k -ésimo paso el -- pivote es cero, los ceros requeridos estan ya en la k -ésima co- -- lumna y tomamos $M_k = I$. Esta estrategia de pivotamiento es lla- -- mado Eliminación Gaussiana con pivotamiento parcial.

ALGORITMO 2.13 (Eliminación Gaussiana con pivotamiento parcial)

Sea A una matriz $m \times n$. Este algoritmo calcula la solución del sistema (2.15) intercambiando filas en la matriz -- aumentada.

1) Construir la matriz aumentada

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & \dots & a_{2n} & a_{2,n+1} \\ \vdots & & & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} & a_{m,n+1} \end{array} \right]$$

2) Hacer $r = \min \{m - 1, n\}$

- 3) Para $k = 1, 2, \dots, r$
- 1) Hallar $y_k \geq k$ tal que $|a_{y_k, k}| = \max \{|a_{ik}| : i \geq k\}$
 - 2) Si $a_{y_k, k} = 0$, tomar siguiente k .
 - 3) $a_{kj} \longleftrightarrow a_{y_k, j} \quad (j = k, k+1, \dots, n+1)$
 - 4) $a_{ik} \longleftarrow m_{ik} = \frac{a_{ik}}{a_{kk}} \quad (i = k+1, k+2, \dots, m)$
 - 5) $a_{ij} \longleftarrow a_{ij} - m_{ik} a_{kj} \quad (i = k+1, \dots, m ; j = k+1, \dots, n+1)$

4) Si $m > n$ ir a paso 6)

5) Si $a_{mm} \neq 0$ ir a paso 9)

6) Si $a_{i, n+1} \neq 0$ para $i = r+1, r+2, \dots, m$, imprimir el mensaje --
"el sistema no tiene solución" e ir al paso 10)

7) Hacer para $i = r, r-1, \dots, 1$

$$1) x_i = a_{ii}^{-1} (a_{i, n+1} - \sum_{j=i+1}^r a_{ij} x_j)$$

8) Ir a paso 10

9) Para $i = m, m-1, \dots, 1$

$$1) x_i = a_{ii}^{-1} a_{i, n+1} - \sum_{j=i+1}^m a_{ij} x_j$$

10) El proceso ha terminado.

▲

EJEMPLO 2.14:

Empleando eliminación Gaussiana con pivotamiento --
parcial, resolver el sistema

$$\begin{aligned}x_1 + 2x_2 + 2x_3 &= 2 \\3x_1 - 2x_2 - x_3 &= 5 \\2x_1 - 5x_2 + 3x_3 &= -4 \\x_1 + 4x_2 + 6x_3 &= 0\end{aligned}$$

1) La matriz aumentada es

$$\left[\begin{array}{ccc|c} 1 & 2 & 2 & 2 \\ 3 & -2 & -1 & 5 \\ 2 & -5 & 3 & -4 \\ 1 & 4 & 6 & 0 \end{array} \right]$$

2) $r = \min \{4 - 1, 3\}$
 $r = 3$

3) Para $k = 1, 2, 3$.

$$\boxed{k = 1}$$

1) $y_1 \geq 1$ tal que $|a_{y_1, i}| = \max \{|a_{i1}| : i \geq 1\} = a_{21} = 3$

2) $a_{y_1, 1} = a_{21} = 3 \neq 0$

3) $a_{ij} \longleftrightarrow a_{y_1 j} \quad (j = 1, 2, 3, 4)$

$$\left. \begin{array}{l} a_{11} \longleftrightarrow a_{21} \\ a_{12} \longleftrightarrow a_{22} \\ a_{13} \longleftrightarrow a_{23} \\ a_{14} \longleftrightarrow a_{24} \end{array} \right\} \Rightarrow \left[\begin{array}{ccc|c} 3 & -2 & -1 & 5 \\ 1 & 2 & 2 & 2 \\ 2 & -5 & 3 & -4 \\ 1 & 4 & 6 & 0 \end{array} \right]$$

4) $a_{i1} \longleftarrow m_{i1} = \frac{a_{i1}}{a_{11}} \quad (i = 2, 3, 4)$

$$\left. \begin{array}{l} a_{21} \longleftarrow m_{21} = \frac{a_{21}}{a_{11}} = \frac{1}{3} \\ a_{31} \longleftarrow m_{31} = \frac{a_{31}}{a_{11}} = \frac{2}{3} \\ a_{41} \longleftarrow m_{41} = \frac{a_{41}}{a_{11}} = \frac{1}{3} \end{array} \right\} \Rightarrow \left[\begin{array}{ccc|c} 3 & -2 & -1 & 5 \\ 1/3 & 2 & 2 & 2 \\ 2/3 & -5 & 3 & -4 \\ 1/3 & 4 & 6 & 0 \end{array} \right]$$

$$5) a_{ij} \leftarrow a_{ij} - m_{i1} a_{1j} \quad (i = 2, 3, 4 \quad ; \quad j = 2, 3, 4).$$

$$\boxed{i = 2}$$

$$a_{22} \leftarrow a_{22} - m_{21} a_{12} = 2 - \frac{1}{3}(-2) = \frac{8}{3}$$

$$a_{23} \leftarrow a_{23} - m_{21} a_{13} = 2 - \frac{1}{3}(-1) = \frac{7}{3}$$

$$a_{24} \leftarrow a_{24} - m_{21} a_{14} = 2 - \frac{1}{3}(5) = \frac{1}{3}$$

$$\boxed{i = 3}$$

$$a_{32} \leftarrow a_{32} - m_{31} a_{12} = -5 - \frac{2}{3}(-2) = -\frac{11}{3}$$

$$a_{33} \leftarrow a_{33} - m_{31} a_{13} = 3 - \frac{2}{3}(-1) = \frac{11}{3}$$

$$a_{34} \leftarrow a_{34} - m_{31} a_{14} = -4 - \frac{2}{3}(5) = -\frac{22}{3}$$

$$\boxed{i = 4}$$

$$a_{42} \leftarrow a_{42} - m_{41} a_{12} = 4 - \frac{1}{3}(-2) = \frac{14}{3}$$

$$a_{43} \leftarrow a_{43} - m_{41} a_{13} = 6 - \frac{1}{3}(-1) = \frac{19}{3}$$

$$a_{44} \leftarrow a_{44} - m_{41} a_{14} = 0 - \frac{1}{3}(5) = -\frac{5}{3}$$

$$\Rightarrow \left[\begin{array}{cccc|c} 3 & -2 & -1 & & 5 \\ 1/3 & 8/3 & 7/3 & & 1/3 \\ 2/3 & -11/3 & 11/3 & & -22/3 \\ & & & & \\ 1/3 & 14/3 & 19/3 & & -5/3 \end{array} \right]$$

$$\boxed{k = 2}$$

$$1) y_2 \geq 2 \text{ tal que } |a_{y_2, 2}| = \max \{|a_{i, 2}| : i \geq 2\} = a_{42} = \frac{14}{3}$$

$$2) a_{y_2, 2} = a_{42} = \frac{14}{3} \neq 0$$

$$3) a_{2j} \leftrightarrow a_{4j} \quad (j = 2, 3, 4)$$

$$\left. \begin{array}{l} a_{22} \longleftrightarrow a_{42} \\ a_{23} \longleftrightarrow a_{43} \\ a_{24} \longleftrightarrow a_{44} \end{array} \right\} \Rightarrow \begin{bmatrix} 3 & -2 & -1 & 5 \\ \textcircled{1/3} & 14/3 & 19/3 & -5/3 \\ \textcircled{2/3} & -11/3 & 11/3 & -22/3 \\ \textcircled{1/3} & 8/3 & 7/3 & 1/3 \end{bmatrix}$$

$$4) \left. \begin{array}{l} a_{i2} \longleftarrow m_{i2} = \frac{a_{i2}}{a_{22}} \quad (i = 3, 4) \\ a_{32} \longleftarrow m_{32} = \frac{a_{32}}{a_{22}} = \frac{-\frac{11}{3}}{\frac{14}{3}} = -\frac{11}{14} \\ a_{42} \longleftarrow m_{42} = \frac{a_{42}}{a_{22}} = \frac{8/3}{14/3} = \frac{4}{7} \end{array} \right\} \Rightarrow \begin{bmatrix} 3 & -2 & -1 & 5 \\ \textcircled{1/3} & 14/3 & 19/3 & -5/3 \\ \textcircled{2/3} & -11/14 & 11/3 & -22/3 \\ \textcircled{1/3} & \textcircled{4/7} & 7/3 & 1/3 \end{bmatrix}$$

$$5) a_{ij} \longleftarrow a_{ij} - m_{i2} a_{2j} \quad (i = 3, 4 ; j = 3, 4)$$

$$\boxed{i = 3}$$

$$a_{33} \longleftarrow a_{33} - m_{32} a_{23} = \frac{11}{3} - \left(-\frac{11}{14}\right) \left(\frac{19}{3}\right) = \frac{121}{14}$$

$$a_{34} \longleftarrow a_{34} - m_{32} a_{24} = -\frac{22}{3} - \left(-\frac{11}{14}\right) \left(-\frac{5}{3}\right) = -\frac{121}{14}$$

$$\boxed{i = 4}$$

$$a_{43} \longleftarrow a_{43} - m_{42} a_{23} = \frac{7}{3} - \frac{4}{7} \left(\frac{19}{3}\right) = -\frac{9}{7}$$

$$a_{44} \longleftarrow a_{44} - m_{42} a_{24} = \frac{1}{3} - \frac{4}{7} \left(\frac{-5}{3}\right) = \frac{9}{7}$$

$$\begin{bmatrix} 3 & -2 & -1 & 5 \\ \textcircled{1/3} & 14/3 & 19/3 & -5/3 \\ \textcircled{2/3} & -11/14 & 121/14 & -121/14 \\ \textcircled{1/3} & \textcircled{4/7} & -9/7 & 9/7 \end{bmatrix}$$

$$\boxed{k = 3}$$

$$1) y_3 \geq 3 \text{ tal que } |a_{y_3, 3}| = \max \{|a_{i3}| : i \geq 3\} = a_{33} = \frac{121}{14}$$

$$2) a_{y_3, 3} = a_{33} = \frac{121}{14} \neq 0$$

$$3) a_{3j} \longleftrightarrow a_{3j} \quad (j = 3, 4)$$

$$a_{33} \longleftrightarrow a_{33}$$

$$a_{34} \longleftrightarrow a_{34}$$

$$4) a_{i3} \longleftarrow m_{i3} = \frac{a_{i3}}{a_{33}} \quad (i = 4)$$

$$a_{43} \longleftarrow m_{43} = \frac{a_{43}}{a_{33}} = \frac{-\frac{9}{7}}{\frac{121}{14}} = -\frac{18}{121}$$

$$\left[\begin{array}{ccc|c} 3 & -1 & -1 & 5 \\ \textcircled{1/3} & 14/3 & 19/3 & -5/3 \\ \textcircled{2/3} & -11/14 & 121/14 & -121/14 \\ \textcircled{1/3} & \textcircled{4/7} & -18/121 & 9/7 \end{array} \right]$$

$$5) a_{ij} \longleftarrow a_{ij} - m_{i3} a_{3j} \quad (i = 4 ; j = 4)$$

$$a_{44} \longleftarrow a_{44} - m_{43} a_{34} = \frac{9}{7} - \left(-\frac{18}{121}\right) \left(-\frac{121}{14}\right) = 0$$

$$\implies \left[\begin{array}{ccc|c} 3 & -2 & -1 & 5 \\ \textcircled{1/3} & 14/3 & 19/3 & -5/3 \\ \textcircled{2/3} & -11/14 & 121/14 & -121/14 \\ \textcircled{1/3} & \textcircled{4/7} & -18/121 & 0 \end{array} \right]$$

4) Si $m > n$ ($4 > 3$) ir al paso 6)

6) Para $i = 4$

$$a_{i, n+1} = a_{44} = 0$$

7) Para $i = 3, 2, 1$.

$$\underline{\underline{i = 3}}$$

$$x_3 = \frac{1}{a_{33}} (a_{34}) = \frac{14}{121} \left(-\frac{121}{14}\right)$$

$$\boxed{x_3 = -1}$$

$$\underline{\underline{i = 2}}$$

$$x_2 = \frac{1}{a_{22}} (a_{24} - a_{23}x_3)$$

$$= \frac{3}{14} \left[-\frac{5}{3} - \left(\frac{19}{3}\right)(-1) \right]$$

$$= \frac{3}{14} \left(\frac{14}{3}\right)$$

$$\boxed{x_2 = 1}$$

$$\underline{\underline{i = 1}}$$

$$x_1 = \frac{1}{a_{11}} [a_{14} - (a_{12}x_2 + a_{13}x_3)]$$

$$= \frac{1}{3} \{5 - [(-2)(1) + (-1)(-1)]\}$$

$$= \frac{1}{3} (5 + 1)$$

$$\boxed{x_1 = 2}$$

Otra alternativa para obtener la solución de un sistema de la forma $A\bar{x} = \bar{b}$ a través de la eliminación Gaussiana se da en el siguiente algoritmo.

ALGORITMO 2.15

Sea P_i y M_i ($i = 1, 2, \dots, n-1$), y A_n las matrices

obtenidas aplicando eliminación Gaussiana con pivotamiento parcial a A (pasos 1, 2 y 3 del algoritmo 2.13). Dado el n -vector \bar{b} , este algoritmo calcula la solución del sistema $A\bar{x} = \bar{b}$.

$$1) \bar{x} = \bar{b}$$

$$2) \left\{ \begin{array}{l} \text{Para } k = 1, 2, \dots, n-1 \\ 1) \bar{x} \leftarrow M_k P_k \bar{x} \end{array} \right.$$

$$3) \bar{x} \leftarrow A_n^{-1} \bar{x}.$$

OBSERVACION 2.16:

Notemos que el algoritmo anterior para poder llevarse a cabo, combina varios algoritmos vistos anteriormente.

EJERCICIOS

2.1 Elabore un algoritmo para resolver un sistema de ecuaciones con sustitución hacia adelante.

2.2 Escriba un algoritmo con sustitución hacia adelante utilizando las siguientes estrategias: a) con pivotamiento completo y b) con pivotamiento parcial.

2.3 Resolver los siguientes sistemas lineales utilizando el método de Eliminación Gaussiana.

$$\begin{array}{ll} \text{a) } x_1 + 3x_2 - x_3 + x_4 = 5 & \text{b) } x_1 + x_2 - x_3 = 3 \\ & 2x_1 - x_2 + 3x_3 = 0 \\ & 2x_2 + x_3 + x_4 = 7 & -x_1 - 2x_2 + x_3 = -5 \\ & x_3 + 2x_4 = 7 & \\ & 3x_4 = 9 & \end{array}$$

$$\begin{aligned} \text{c) } x_1 - 2x_2 &= 3 \\ 2x_1 + x_2 &= 4 \end{aligned}$$

$$\begin{aligned} \text{d) } 3x - y + 2z &= -3 \\ x + y + z &= -4 \\ 2x + y - z &= -3 \end{aligned}$$

2.4 Use el algoritmo 2.5 para resolver los siguientes sistemas lineales:

$$\begin{aligned} \text{a) } x_1 - x_2 + 3x_3 &= 2 \\ 3x_1 - 3x_2 + x_3 &= -1 \\ x_1 + x_2 &= 3 \end{aligned}$$

$$\begin{aligned} \text{b) } 2x + y &= -1 \\ x + y &= 2 \\ x - 3y &= 5 \end{aligned}$$

$$\begin{aligned} \text{c) } \frac{1}{4}x_1 + \frac{1}{5}x_2 + \frac{1}{6}x_3 &= 9 \\ \frac{1}{3}x_1 + \frac{1}{4}x_2 + \frac{1}{5}x_3 &= 8 \\ \frac{1}{2}x_1 + x_2 + 2x_3 &= 8 \end{aligned}$$

$$\begin{aligned} \text{d) } 2x + y + z &= 1 \\ 2x + 4y - z &= -1 \end{aligned}$$

2.5 Mostrar que el sistema lineal

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= 4 \\ x_1 - 2x_2 + x_3 &= 6 \\ -x_1 - 12x_2 + 5x_3 &= 10 \end{aligned}$$

tiene infinito número de soluciones.

3. METODO DE GAUSS-JORDAN

En esta sección trataremos el método de Gauss-Jordan que resuelve la ecuación lineal $A\bar{x} = \bar{b}$, donde A es de orden n y \bar{b} es un n -vector; el método puede ser descrito de la siguiente manera:

Sea el sistema

$$\begin{aligned}
E_1: & a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = a_{1,n+1} \\
E_2: & a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = a_{2,n+1} \\
& \vdots \\
& \vdots \\
E_n: & a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n = a_{n,n+1}
\end{aligned} \tag{3.1}$$

El método de Gauss-Jordan usa la i -ésima ecuación para eliminar no solamente la variable x_i de las ecuaciones $E_{i+1}, E_{i+2}, \dots, E_n$, como se hizo en el método de eliminación de Gauss, sino que también la elimina de las ecuaciones E_1, E_2, \dots, E_{i-1} . Resultando la matriz

$$\left[\begin{array}{cccc|c}
a_{11}^{(1)} & 0 & \dots & 0 & a_{1,n+1}^{(n)} \\
0 & a_{22}^{(2)} & 0 & \dots & a_{2,n+1}^{(n)} \\
\vdots & & & & \vdots \\
\vdots & & & & \vdots \\
0 & \dots & \dots & \dots & a_{n,n+1}^{(n)}
\end{array} \right] \tag{3.2}$$

Luego la solución es obtenida haciendo

$$x_i = \frac{a_{i,n+1}^{(n)}}{a_{ii}^{(i)}} \tag{3.3}$$

para cada $i = 1, 2, \dots, n$. Este proceso evita la sustitución hacia atrás en la eliminación Gaussiana.

Ilustraremos esta idea considerando el mismo sistema (2.1) empleado en la sección 2.

Dado el sistema lineal

$$\begin{aligned}
2x_1 + 4x_2 - 2x_3 &= 6 \\
x_1 - x_2 + 5x_3 &= 0 \\
4x_1 + x_2 - 2x_3 &= 2
\end{aligned} \tag{3.4}$$

Si $\frac{1}{2}$ veces la primera ecuación es restada de la segunda y 2 veces la primera es restada de la tercera, obtenemos el sistema

$$\begin{aligned} 2x_1 + 4x_2 - 2x_3 &= 6 \\ -3x_2 + 6x_3 &= -3 \\ -7x_2 + 2x_3 &= -10 \end{aligned} \quad (3.5)$$

Si $-\frac{4}{3}$ veces la segunda ecuación es restada de la primera y $\frac{7}{3}$ veces la segunda es restada de la tercera, resulta el sistema

$$\begin{aligned} 2x_1 + 6x_3 &= 2 \\ -3x_2 + 6x_3 &= -3 \\ -12x_3 &= -3 \end{aligned} \quad (3.6)$$

Podemos eliminar x_3 del sistema (3.6) multiplicando por $-\frac{1}{2}$ la tercera ecuación restársela a la primera y lo mismo restarlo a la segunda

$$\begin{aligned} 2x_1 &= \frac{1}{2} \\ -3x_2 &= -\frac{9}{2} \\ -12x_3 &= -3 \end{aligned} \quad (3.7)$$

vemos que el sistema (3.7) es diagonal y se resuelve en la forma (3.3).

Sean $A = A_1, A_2, A_3$ y A_4 . Las matrices del sistema (3.4), (3.5), (3.6) y (3.7) respectivamente. Entonces A_2 es obtenida de A_1 , premultiplicando A_1 por la matriz

$$R_1 = \begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ -2 & 0 & 1 \end{bmatrix}$$

es decir que $A_2 = R_1 A_1$ (3.8)

Similarmente

$$A_3 = R_2 A_2 \tag{3.9}$$

donde

$$R_2 = \begin{bmatrix} 1 & 4/3 & 0 \\ 0 & 1 & 0 \\ 0 & -7/3 & 1 \end{bmatrix}$$

y $A_4 = R_3 A_3$ (3.10)

donde

$$R_3 = \begin{bmatrix} 1 & 0 & 1/2 \\ 0 & 1 & 1/2 \\ 0 & 0 & 1 \end{bmatrix}$$

combinando (3.8), (3.9) y (3.10), vemos que

$$A_4 = R_3 R_2 R_1 A \tag{3.11}$$

luego de la misma manera que en la eliminación de Gauss, para resolver el sistema $A \bar{x} = \bar{b}$, necesitamos calcular el vector $\bar{c} = R_3 R_2 R_1 \bar{b}$ y resolver el sistema diagonal

$$A_4 \bar{x} = R_3 R_2 R_1 A \bar{x} = R_3 R_2 R_1 \bar{b} = \bar{c}$$

En la discusión anterior intervienen matrices R_i , las cuales son llamadas R-matrices elementales, las cuales se describen en la siguiente definición.

$$R = I - \bar{r} \bar{e}_k^t$$

donde

$$\bar{e}_k^t \bar{r} = 0. \tag{3.12}$$

De la condición (3.12) se deduce que todas las componentes de \bar{r} son diferentes de cero, excepto la k -ésima componente; es decir \bar{r} tiene la forma $\bar{r} = [r_1, r_2, \dots, r_{k-1}, 0, r_{k+1}, r_{k+2}, \dots, r_n]^t$.

En general una R-matriz elemental tiene la forma

$$R_k = \begin{bmatrix} 1 & 0 & 0 & \dots & r_1 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & -r_2 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & -r_{k-1} & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & 1 & 0 & \dots & 0 \\ 0 & \dots & \dots & \dots & -r_{k+1} & \dots & \dots & \dots \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & \dots & -r_n & \dots & \dots & 1 \end{bmatrix}$$

TEOREMA 3.2:

Sea $R = I - \bar{r} \bar{e}_k^t$, donde $\bar{e}_k^t \bar{r} = 0$. Entonces

- 1) $R^{-1} = I + \bar{r} \bar{e}_k^t$
- 2) $\bar{e}_k^t \bar{x} = 0 \Rightarrow R \bar{x} = \bar{x}$

PRUEBA

- 1) Sea $Y = I + \bar{r} \bar{e}_k^t$. Entonces

$$\begin{aligned}
RY &= (I - \bar{r} \bar{e}_k^t) (I + \bar{r} \bar{e}_k^t) \\
&= I - \bar{r} \bar{e}_k^t + \bar{r} \bar{e}_k^t - (\bar{r} \bar{e}_k^t) (\bar{r} \bar{e}_k^t) \\
&= I - (\bar{r} \bar{e}_k^t) (\bar{r} \bar{e}_k^t) \\
&= I - \bar{r} (\bar{e}_k^t \bar{r}) \bar{e}_k^t \\
&= I, \quad \text{ya que } \bar{e}_k^t \bar{r} = 0
\end{aligned}$$

Por lo tanto $RY = I$ y Y es la inversa de R .

Sea $R = I - \bar{r} \bar{e}_k^t$

$$\begin{aligned}
R\bar{x} &= (I - \bar{r} \bar{e}_k^t) \bar{x} \\
&= \bar{x} - (\bar{r} \bar{e}_k^t) \bar{x} \\
&= \bar{x} - \bar{r} (\bar{e}_k^t \bar{x}) \\
&= \bar{x} - 0, \quad \text{ya que por hipótesis } \bar{e}_k^t \bar{x} = 0 \\
&= \bar{x}
\end{aligned}$$

▲

El significado computacional de las R-matrices elementales, al igual que las matrices M , es que se emplean para hacer cero las componentes de un vector.

TEOREMA 3.3:

Si $\bar{e}_k^t \bar{x} \neq 0$, entonces existe una R-matriz elemental de índice k tal que

$$R\bar{x} = x_k \bar{e}_k = [0, 0, \dots, 0, x_k, 0, \dots, 0]^t \quad (3.13)$$

PRUEBA

Se tiene que R es de la forma $R = I - \bar{r} \bar{e}_k^t$.

Como R es de índice k , debemos tener que

$$r_k = 0 \quad (3.14)$$

luego

$$\begin{aligned} R\bar{x} &= (I - \bar{r} \bar{e}_k^t) \bar{x} \\ &= \bar{x} - (\bar{r} \bar{e}_k^t) \bar{x} \\ &= \bar{x} - \bar{r} (\bar{e}_k^t \bar{x}) \end{aligned}$$

$$= \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_k \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{bmatrix} - \begin{bmatrix} r_1 x_k \\ r_2 x_k \\ \cdot \\ \cdot \\ \cdot \\ r_{k-1} x_k \\ 0 \\ r_{k+1} x_k \\ \cdot \\ \cdot \\ \cdot \\ r_n x_k \end{bmatrix} = [0, 0, \dots, 0, x_k, 0, \dots, 0]^t$$

Luego $x_i - x_k r_i = 0$ ($i = 1, 2, \dots, k-1, k+1, k+2, \dots, n$)

como $\bar{e}_k^t \bar{x} = x_k \neq 0$

Por tanto

$$r_i = \frac{x_i}{x_k} \quad (i = 1, 2, \dots, k-1, k+1, k+2, \dots, n) \quad (3.15)$$

así, si R existe, está determinada en forma única por (3.14) y (3.15). Por otra parte si $x_k \neq 0$, entonces el vector \bar{r} puede ser determinado de (3.14) y (3.15) y se puede verificar que la

matriz asociada $R = I - \bar{r} \bar{e}_k^t$ satisface (3.13)

A

Podemos observar que la discusión anterior nos define un algoritmo para calcular \bar{r} y por lo tanto R .

En este método de Gauss-Jordan que reduce una matriz a la forma diagonal, la matriz es premultiplicada por una secuencia de R -matrices elementales, cada una escogida para introducir ceros bajo y sobre la diagonal.

Sea $A_1 = A$ una matriz de orden n . Si $a_{11}^{(1)} = a_{11}$ es distinto de cero, entonces por el teorema 3.3, existe una única R -matriz elemental R_1 de índice 1 que elimina los $n-1$ elementos de la primera columna de A_1 . Si A_1 es premultiplicada por R_1 , resulta una matriz A_2 de la forma

$$A_2 = R_1 A_1 = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & \dots & a_{1n}^{(1)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \dots & a_{2n}^{(2)} \\ 0 & a_{32}^{(2)} & a_{33}^{(2)} & \dots & a_{3n}^{(2)} \\ \vdots & & & & \\ 0 & a_{n2}^{(2)} & a_{n3}^{(2)} & \dots & a_{nn}^{(2)} \end{bmatrix}$$

Si A_2 es premultiplicada por R_2 , entonces resulta la matriz A_3 de la forma

$$A_3 = R_2 A_2 = \begin{bmatrix} a_{11}^{(1)} & 0 & a_{13}^{(2)} & \dots & a_{1n}^{(2)} \\ 0 & a_{22}^{(2)} & a_{23}^{(2)} & \dots & a_{2n}^{(2)} \\ 0 & 0 & a_{32}^{(3)} & \dots & a_{3n}^{(3)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & a_{n3}^{(3)} & \dots & a_{nn}^{(3)} \end{bmatrix}$$

En general supóngase que A_{k+1} tiene la forma

$$A_{k+1} = \begin{bmatrix} D_{k+1} & A_{12}^{(k+1)} \\ 0 & A_{22}^{(k+1)} \end{bmatrix}$$

donde D_{k+1} es diagonal de orden k . Así A_{n+1} es diagonal. La matriz A se reducirá a la forma diagonal después del n -ésimo paso.

Podemos construir un algoritmo para el cálculo de A_k y R_k . Específicamente la k -ésima columna de A_k es el vector

$$\bar{c}_k = [a_{1k}^{(k)}, a_{2k}^{(k)}, a_{3k}^{(k)}, \dots, a_{kk}^{(k+1)}, \dots, a_{nk}^{(k+1)}]^t$$

y R_k es la R -matriz elemental de índice k tal que

$$R_k \bar{c}_k = [0, 0, \dots, a_{kk}^{(k+1)}, 0, \dots, 0]^t$$

Por el teorema 3.3, $R_k = I - \bar{r}_k \bar{e}_k^t$, donde

$$\bar{r}_k = [r_{1k}, r_{2k}, \dots, r_{k-1,k}, r_{k+1,k}, \dots, r_{nk}]^t$$

$$y \quad r_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}} \quad (i = 1, 2, \dots, k-1, k+1, \dots, n) \quad (3.16)$$

Así R_k está determinada en forma única si $a_{kk}^{(k)} \neq 0$.

En base a todo lo discutido anteriormente establecemos el

siguiente algoritmo.

ALGORITMO 3.4 (Método de Gauss-Jordan)

Sea A una matriz de orden n . Este algoritmo calcula la solución \bar{x} del sistema lineal (3.1) y sobrescribe los elementos r_{ik} en los correspondientes elementos a_{ik}

1) Construir la matría aumentada

$$\left[\begin{array}{cccc|c} a_{11} & a_{12} & \dots & a_{1n} & a_{1,n+1} \\ a_{21} & a_{22} & \dots & a_{2n} & a_{2,n+1} \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn} & a_{n,n+1} \end{array} \right]$$

- 2) Para $k = 1, 2, \dots, n$
- 1) $a_{ik} \leftarrow r_{ik} = \frac{a_{ik}}{a_{kk}} \quad (i = 1, 2, \dots, k-1, k+1, \dots, n)$
- 2) Para $i = 1, 2, \dots, k-1, k+1, \dots, n$
- 1) $a_{ij} \leftarrow a_{ij} - r_{ik} a_{kj} \quad (j = k+1, k+2, \dots, n+1)$
- 3) Hacer $x_i = \frac{a_{i,n+1}}{a_{ii}} \quad (i = 1, 2, \dots, n).$

NOTA:

Para comprobar el funcionamiento del algoritmo 3.4, puede verse en el desarrollo del ejemplo 4.2 que se encuentra en la próxima sección.

OBSERVACION 3.5

El algoritmo de Gauss-Jordan funciona bajo las -

mismas condiciones tratadas en la sección 2 de este capítulo sobre la Eliminación Gaussiana.

EJERCICIOS

3.1 Elaborar un algoritmo general de Gauss-Jordan para resolver sistemas de ecuaciones de orden $m \times n$.

3.2 Modificar el algoritmo 3.4 utilizando las estrategias de pivotamiento:

a) parcial y b) completo.

3.3 Resuelva el siguiente sistema por el método de Gauss-Jordan

$$3.333 x_1 + 15920 x_2 - 10.333 x_3 = 15913$$

$$2.222 x_1 + 16.71 x_2 + 9.612 x_3 = 28.544$$

$$1.5611 x_1 + 5.1791 x_2 + 1.6852 x_3 = 8.454$$

3.4 Use el algoritmo 3.4 para resolver los sistemas lineales a) y c) del ejercicio 2.4 de la sección anterior.

4. COMPARACION DE LOS METODOS

Para concluir este capítulo haremos una comparación entre el método de Eliminación de Gauss y el método de Gauss-Jordan en lo que concierne a su aplicación al problema de la búsqueda de la solución de un sistema lineal no homogéneo.

El criterio generalmente aceptado de la "superioridad" de un método dado, consiste en el número necesario de operaciones de cálculo, es decir, el número de multiplicaciones y divisiones, así como también el de adiciones y sustracciones necesarias.

Cuando comparamos técnicas para resolver sistemas lineales necesitamos considerar el efecto de error por redondeo y el tiempo requerido para completar los cálculos. Ambos aspectos dependen del número de operaciones aritméticas necesarias para encontrar la solución al problema.

Solamente haremos el cálculo del número de multiplicaciones y divisiones que se realizan al aplicar el método de Eliminación de Gauss, así también, cuando se emplea el de Gauss-Jordan. Contar solo estas operaciones y no tener en cuenta adiciones y sustracciones, se justifica por la observación de que estas son las partes del algoritmo que mas tiempo consumen y mas errores producen. Las operaciones que contaremos serán las que resuelvan sistemas lineales de n ecuaciones con n variables. Al aplicar el algoritmo 2.5 (Eliminación Gaussiana con sustitución hacia atrás) a un sistema lineal, la cantidad de multiplicaciones y divisiones que se efectúan para encontrar la solución se obtiene de la siguiente forma:

Para el paso 3:

1) $(n - j)$ divisiones.

2) $(n - j)(n - j + 1)$ multiplicaciones.

Nº de multiplicaciones y divisiones:

$$(n - j) + (n - j)(n - j + 1) = (n - j)(n - j + 2).$$

El número total de multiplicaciones y divisiones para este paso es:

$$\begin{aligned}
\sum_{j=1}^{r=n-1} (n-j)(n-j+2) &= \sum_{j=1}^{n-1} (n^2 - 2nj + 2n + j^2 - 2j) \\
&= \sum_{j=1}^{n-1} [n^2 - 2j(n+1) + 2n + j^2] \\
&= \sum_{j=1}^{n-1} n^2 - 2(n+1) \sum_{j=1}^{n-1} j + \sum_{j=1}^{n-1} 2n + \sum_{j=1}^{n-1} j^2 \\
&= (n-1)n^2 - 2(n+1) \frac{(n-1)n}{2} + (n-1)2n + \frac{n(n-1)(2n-1)}{6} \\
&= n(n-1) \left[n - (n+1) + 2 + \frac{(2n-1)}{6} \right] \\
&= n(n-1) \left[1 + \left(\frac{n}{3} - \frac{1}{6} \right) \right] \\
&= n(n-1) \left[\frac{5}{6} + \frac{1}{3}n \right] \\
&= \frac{5}{6}n^2 + \frac{1}{3}n^3 - \frac{5}{6}n - \frac{1}{3}n^2 \\
&= \frac{n^3}{3} + \frac{1}{2}n - \frac{5}{6}n = \frac{2n^3 + 3n^2 - 5n}{6}
\end{aligned}$$

Otro paso del algoritmo que requiere operaciones aritméticas es el paso 8 en el cual se verifican:

- a) Primera sustitución hacia atrás: 1 división
- b) Sigüientes sustituciones hacia atrás: 1 división y $(n-j)$ multiplicaciones.

El total de multiplicaciones y divisiones en este paso es:

$$\begin{aligned}
1 + \sum_{j=1}^{n-1} [1 + (n-j)] &= 1 + \sum_{j=1}^{n-1} [(n+1) - j] \\
&= 1 + \sum_{j=1}^{n-1} (n+1) - \sum_{j=1}^{n-1} j \\
&= 1 + (n-1)(n+1) - \frac{n(n-1)}{2} \\
&= 1 + n^2 - 1 - \frac{n^2}{2} + \frac{n}{2} \\
&= \frac{n^2 + n}{2}
\end{aligned}$$

Entonces, el total de multiplicaciones y divisiones en el algoritmo 2.5 es

$$\begin{aligned}
\frac{2n^3 + 3n^2 - 5n}{6} + \frac{n^2 + n}{2} &= \frac{2n^3 + 6n^2 - 2n}{6} \\
&= \frac{n^3 + 3n^2 - n}{3} \quad (4.1)
\end{aligned}$$

Cuando se aplica el algoritmo 3.4 (Método de Gauss-Jordan) la cantidad de multiplicaciones y divisiones que se realizan se obtienen de la siguiente forma:

En el paso 2:

- 1) Para cada columna se efectúan $(n-1)$ divisiones.
- 2) En cada fila se efectúan $(n-j+1)$ multiplicaciones.

El total de filas que se multiplican es de $(n-1)$.

El total de productos por columna es: $(n-1)(n-j+1)$.

Total de multiplicaciones y divisiones para este paso es:

$$\begin{aligned}
\sum_{j=1}^n [(n-1) + (n-1)(n-j+1)] &= \sum_{j=1}^n (n-1)(n-j+2) \\
&= \sum_{j=1}^n (n^2 - jn + n + j - 2) \\
&= \sum_{j=1}^n [(n^2 + n - 2) + j(1-n)] \\
&= \sum_{j=1}^n (n^2 + n - 2) + (1-n) \sum_{j=1}^n j \\
&= n(n^2 + n - 2) + (1-n) \frac{n(n+1)}{2} \\
&= n^3 + n^2 - 2n + \frac{n - n^3}{2} \\
&= \frac{2n^3 + 2n^2 - 4n + n - n^3}{2} \\
&= \frac{n^3 + 2n^2 - 3n}{2} \\
&= \frac{n^3}{2} + n^2 - \frac{3n}{2}
\end{aligned}$$

En el paso 3, se efectúan n divisiones.

Entonces, el total de multiplicaciones y divisiones en el algoritmo 3.4 es:

$$\frac{n^3}{2} + n^2 - \frac{3n}{2} + n = \frac{n^3}{2} + n^2 - \frac{n}{2}. \quad (4.2)$$

La siguiente tabla nos demuestra cómo la cantidad de operaciones varía entre los dos métodos, a medida se cambia el tamaño de n , utilizando las expresiones (4.1) y (4.2).

NUMERO TOTAL DE MULTIPLICACIONES Y DIVISIONES		
Número de Ecuaciones	Eliminación de Gauss con sustitución hacia atrás	Método de Gauss-Jordan
n	$\frac{n^3 + 3n^2 - n}{3}$	$\frac{n^3}{2} + n^2 - \frac{n}{2}$
3	17	21
7	161	217
10	430	595
50	40,150	64,975

De la tabla anterior notamos que en el método de Gauss-Jordan a medida que el valor de n se incrementa el número de operaciones es mayor que en el método de Eliminación Gaussiana.

EJEMPLO 4.1:

Comprobaremos que al resolver un sistema lineal de orden 3×3 aplicando el algoritmo 2.5, el número de multiplicaciones y divisiones es de 17.

Resolver el sistema lineal

$$2x_1 + 4x_2 - 2x_3 = 6$$

$$x_1 - x_2 + 5x_3 = 0$$

$$4x_1 + x_2 - 2x_3 = 2$$

1) La matriz aumentada es

$$\left[\begin{array}{ccc|c} 2 & 4 & -2 & 6 \\ 1 & -1 & 5 & 0 \\ 4 & 1 & -2 & 2 \end{array} \right]$$

2) $r = \min \{3 - 1, 3\} = 2$

3) Para $k = 1, 2$

$$\boxed{k = 1}$$

$$1) a_{i1} \longleftarrow m_{i1} = \frac{a_{i1}}{a_{11}} \quad (i = 2, 3)$$

$$a_{21} \longleftarrow m_{21} = \frac{a_{21}}{a_{11}} = \frac{1}{2} \checkmark$$

$$a_{31} \longleftarrow m_{31} = \frac{a_{31}}{a_{11}} = \frac{4}{2} = 2$$

$$2) a_{ij} \longleftarrow a_{ij} - m_{i1} a_{1j} \quad (i = 2, 3 ; j = 2, 3, 4)$$

$$a_{22} \longleftarrow a_{22} - m_{21} a_{12} = -1 - \frac{1}{2}(4) = -3 \checkmark$$

$$a_{23} \longleftarrow a_{23} - m_{21} a_{13} = 5 - \frac{1}{2}(-2) = 6 \checkmark$$

$$a_{24} \longleftarrow a_{24} - m_{21} a_{14} = 0 - \frac{1}{2}(6) = -3 \checkmark$$

$$a_{32} \longleftarrow a_{32} - m_{31} a_{12} = 1 - 2(4) = -7 \checkmark$$

$$a_{33} \longleftarrow a_{33} - m_{31} a_{13} = -2 - 2(-2) = 2 \checkmark$$

$$a_{34} \longleftarrow a_{34} - m_{31} a_{14} = 2 - 2(6) = -10 \checkmark$$

$$\implies \begin{bmatrix} 2 & 4 & -2 & | & 6 \\ 1/2 & -3 & 6 & | & -3 \\ 2 & -7 & 2 & | & -10 \end{bmatrix}$$

$$\boxed{k = 2}$$

$$1) a_{i2} \longleftarrow m_{i2} = \frac{a_{i2}}{a_{22}} \quad (i = 3)$$

$$a_{32} \longleftarrow m_{32} = \frac{a_{32}}{a_{22}} = \frac{-7}{-3} = \frac{7}{3} \checkmark$$

$$\begin{aligned}
 2) \quad a_{ij} &\leftarrow a_{ij} - m_{i2} a_{2j} \quad (i = 3 ; j = 3, 4) \\
 a_{33} &\leftarrow a_{33} - m_{32} a_{23} = 2 - \left(\frac{7}{3}\right)(6) = -12 \checkmark \\
 a_{34} &\leftarrow a_{34} - m_{32} a_{24} = -10 - \left(\frac{7}{3}\right)(-3) = -3 \checkmark \\
 &\implies \begin{bmatrix} 2 & 4 & -2 & | & 6 \\ 1/2 & -3 & 6 & | & -3 \\ 2 & 7/3 & -12 & | & -3 \end{bmatrix}
 \end{aligned}$$

$$4) \quad m = n = 3$$

$$8) \quad \text{Para } i = 3, 2, 1$$

$$x_3 = \frac{1}{a_{33}} (a_{34}) = \frac{1}{-12} (-3) = \frac{1}{4} \checkmark$$

$$x_2 = \frac{1}{a_{22}} (a_{24} - a_{23}x_3)$$

$$x_2 = \frac{1}{-3} (-3 - 6 \left(\frac{1}{4}\right)) = \frac{3}{2} \checkmark \checkmark$$

$$x_1 = \frac{1}{a_{11}} (a_{14} - \sum_{j=2}^3 a_{1j}x_j) = \frac{1}{a_{11}} \left[a_{14} - [a_{12}x_2 + a_{13}x_3] \right]$$

$$x_1 = \frac{1}{2} \left[6 - \left[4 \cdot \frac{3}{2} + (-2) \cdot \frac{1}{2} \right] \right]$$

$$x = \frac{1}{2} \left(6 - \frac{11}{2} \right) = \frac{1}{4} \quad \checkmark \checkmark \checkmark$$

Total: 17 multip. y divisiones

Nota:

La marca (\checkmark) identifica a cada una de las operaciones (multiplicaciones y divisiones) efectuadas en el algoritmo.

EJEMPLO 4.2:

Si aplicamos el algoritmo 3.4 (Método de Gauss-Jordan) al mismo sistema lineal de orden 3×3 , el número de multiplicaciones y divisiones es de 21.

1) La matriz aumentada es:

$$\left[\begin{array}{ccc|c} 2 & 4 & -2 & 6 \\ 1 & -1 & 5 & 0 \\ 4 & 1 & -2 & 2 \end{array} \right]$$

2) Para $k = 1, 2, 3$.

$$\boxed{k = 1}$$

$$1) a_{i1} \longleftarrow r_{i1} = \frac{a_{i1}}{a_{11}} \quad (i = 2, 3)$$

$$a_{21} \longleftarrow r_{21} = \frac{a_{21}}{a_{11}} = \frac{1}{2} \checkmark$$

$$a_{31} \longleftarrow r_{31} = \frac{a_{31}}{a_{11}} = \frac{4}{2} = 2 \checkmark$$

2) Para $i = 2, 3$.

$$i = 2$$

$$a_{2j} \longleftarrow a_{2j} - r_{21} a_{1j} \quad (j = 2, 3, 4)$$

$$a_{22} \longleftarrow a_{22} - r_{21} a_{12} = -1 - \frac{1}{2}(4) = -3 \checkmark$$

$$a_{23} \longleftarrow a_{23} - r_{21} a_{13} = 5 - \frac{1}{2}(-2) = 6 \checkmark$$

$$a_{24} \longleftarrow a_{24} - r_{21} a_{14} = 0 - \frac{1}{2}(6) = -3 \checkmark$$

$$i = 3$$

$$a_{32} \longleftarrow a_{32} - r_{31}a_{12} = 1 - 2(4) = -7 \checkmark$$

$$a_{33} \longleftarrow a_{33} - r_{31}a_{13} = -2 - 2(-2) = 2 \checkmark$$

$$a_{34} \longleftarrow a_{34} - r_{31}a_{14} = 2 - 2(6) = -10 \checkmark$$

$$\implies \begin{bmatrix} 2 & 4 & -2 & | & 6 \\ 1/2 & -3 & 6 & | & -3 \\ 2 & -7 & 2 & | & -10 \end{bmatrix}$$

$$\boxed{k = 2}$$

$$1) a_{i2} \longleftarrow r_{i2} = \frac{a_{i2}}{a_{22}} \quad (i = 1, 3)$$

$$a_{12} \longleftarrow r_{12} = \frac{a_{12}}{a_{22}} = \frac{4}{-3} \checkmark$$

$$a_{32} \longleftarrow r_{32} = \frac{a_{32}}{a_{22}} = \frac{-7}{-3} = \frac{7}{3} \checkmark$$

2) Para $i = 1, 3$

$$i = 1$$

$$a_{1j} \longleftarrow a_{1j} - r_{12}a_{2j} \quad (j = 3, 4)$$

$$a_{13} \longleftarrow a_{13} - r_{12}a_{23} = -2 - \left(-\frac{4}{3}\right)(6) = 6 \checkmark$$

$$a_{14} \longleftarrow a_{14} - r_{12}a_{24} = 6 - \left(-\frac{4}{3}\right)(-3) = 2 \checkmark$$

$$i = 3$$

$$a_{3j} \leftarrow a_{3j} - r_{32}a_{2j} \quad (j = 3, 4)$$

$$a_{33} \leftarrow a_{33} - r_{32}a_{23} = 2 - \left(\frac{7}{3}\right)(6) = -12\checkmark$$

$$a_{34} \leftarrow a_{34} - r_{32}a_{24} = -10 - \left(\frac{7}{3}\right)(-3) = -3\checkmark$$

$$\implies \begin{bmatrix} 2 & -4/3 & 6 & | & 2 \\ 1/2 & -3 & 6 & | & -3 \\ 2 & 7/3 & -12 & | & -3 \end{bmatrix}$$

$$\boxed{k = 3}$$

$$1) a_{i3} \leftarrow r_{i3} = \frac{a_{i3}}{a_{33}} \quad (i = 1, 2)$$

$$a_{13} \leftarrow r_{13} = \frac{a_{13}}{a_{33}} = \frac{6}{-12} = -\frac{1}{2} \checkmark$$

$$a_{23} \leftarrow r_{23} = \frac{a_{23}}{a_{33}} = \frac{6}{-12} = -\frac{1}{2} \checkmark$$

2) Para $i = 1, 2$

$$i = 1$$

$$a_{1j} \leftarrow a_{1j} - r_{13}a_{3j} \quad (j = 4)$$

$$a_{14} \leftarrow a_{14} - r_{13}a_{34} = 2 - \left(-\frac{1}{2}\right)(-3) = \frac{1}{2} \checkmark$$

$$i = 2$$

$$a_{2j} \leftarrow a_{2j} - r_{23}a_{3j} \quad (j = 4)$$

$$a_{24} \leftarrow a_{24} - r_{23}a_{34} = -3 - \left(-\frac{1}{2}\right)(-3) = \frac{-9}{2} \checkmark$$

$$\implies \begin{bmatrix} 2 & -4/3 & -1/2 & | & 1/2 \\ 1/2 & -3 & -1/2 & | & -9/2 \\ 2 & 7/3 & -12 & | & -3 \end{bmatrix}$$

3) Hacer

$$x_i = \frac{a_{i,4}}{a_{ii}} \quad (i = 1, 2, 3)$$

$$x_1 = \frac{a_{14}}{a_{11}} = \frac{1/2}{2} = \frac{1}{4} \quad \checkmark$$

$$x_2 = \frac{a_{24}}{a_{22}} = \frac{-9/2}{-3} = \frac{3}{2} \quad \checkmark$$

$$x_3 = \frac{a_{34}}{a_{33}} = \frac{-3}{-12} = \frac{1}{4} \quad \checkmark$$

Total: 21 multipl. y divisiones.

CAPITULO III

METODOS DIRECTOS PARA RESOLVER SISTEMAS DE ECUACIONES LINEALES

SEGUNDA PARTE

1. FACTORIZACION DE MATRICES

En esta sección consideraremos algoritmos para factorar -- una matriz cuadrada A en el producto LU de una matriz triangu-- lar inferior y una matriz triangular superior. Tal factoriza--- ción es a menudo llamada *descomposición LU* de la matriz A . Como una descomposición LU expresa a una matriz como el producto de matrices muy simples, puede ser usada para simplificar cálculos que involucran a la matriz. Por ejemplo si se quiere obtener la inversa de A , puede ser calculada en la forma $A^{-1} = U^{-1}L^{-1}$ por las técnicas del capítulo II, sección 1.

EJEMPLO 1.1:

Consideremos el siguiente sistema lineal.

$$\begin{aligned}x_1 + x_2 + 3x_4 &= 4 \\2x_1 + x_2 - x_3 + x_4 &= 1 \\3x_1 - x_2 - x_3 + 2x_4 &= -3 \\-x_1 + 2x_2 + 3x_3 - x_4 &= 4\end{aligned}\tag{1.1}$$

que aplicando eliminación Gaussiana se reduce al sistema equivalente.

$$\begin{bmatrix} 1 & 1 & 0 & 3 & 4 \\ 0 & -1 & -1 & -5 & -7 \\ 0 & 0 & 3 & 13 & 13 \\ 0 & 0 & 0 & -13 & -13 \end{bmatrix} \quad (1.2)$$

Sea U la matriz triangular superior de orden 4.

$$U = \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} \quad (1.3)$$

la cual es el resultado de la eliminación Gaussiana efectuada sobre A . Si se define L como la matriz triangular inferior cuyos elementos ℓ_{ij} están dados por

$$\ell_{ij} = \begin{cases} 0 & , \text{ cuando } i = 1, 2, \dots, j-1 \\ 1 & , \text{ cuando } i = j \\ m_{ij} & , \text{ cuando } i = j+1, j+2, \dots, n \end{cases}$$

donde m_{ij} son los multiplicadores empleados en la eliminación Gaussiana. Se puede verificar.

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix}$$

además es fácil verificar que

$$\begin{aligned}
 LU &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & 4 & 1 & 0 \\ -1 & -3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 3 \\ 0 & -1 & -1 & -5 \\ 0 & 0 & 3 & 13 \\ 0 & 0 & 0 & -13 \end{bmatrix} \\
 &= \begin{bmatrix} 1 & 1 & 0 & 3 \\ 2 & 1 & -1 & 1 \\ 3 & -1 & -1 & 2 \\ -1 & 2 & 3 & -1 \end{bmatrix} = A
 \end{aligned}$$

Los resultados obtenidos en el ejemplo (1.1) son ciertos - en general y estan dados en el siguiente teorema.

TEOREMA 1.2:

Si el procedimiento de eliminación Gaussiana (algoritmo 2.5) puede ser efectuado en el sistema $A\bar{x} = \bar{b}$ sin intercambios de filas, es decir sin pivotamiento, entonces, la matriz A puede ser factorada en el producto de una matriz triangular inferior con una triangular superior.

$$A = LU,$$

donde $U = (u_{ij})$ y $L = (\ell_{ij})$ estan definidas por

$$u_{ij} = \begin{cases} a_{ij}^{(i)}, & \text{cuando } i = 1, 2, \dots, j \\ 0 & , \text{cuando } i = j+1, j+2, \dots, n \end{cases}$$

$$\ell_{ij} = \begin{cases} 0 & , \text{ si } i = 1, 2, \dots, j-1 \\ 1 & , \text{ si } i = j \\ m_{ij} & , \text{ si } i = j+1, j+2, \dots, n \end{cases}$$

donde $a_{ij}^{(i)}$ es el elemento (i,j) de A_n , y m_{ij} es el multiplicador.

PRUEBA

Como la eliminación Gaussiana se puede efectuar sin pivoteo

$$A_n = M_{n-1}M_{n-2}\dots M_2M_1A$$

donde A_n es triangular superior con elementos

$$a_{ij} = \begin{cases} a_{ij}^{(i)}, & i = 1, 2, \dots, j \\ 0 & , i = j+1, \dots, n \end{cases}$$

$$\text{Sea } U = M_{n-1}M_{n-2}\dots M_2M_1A$$

$$\text{entonces } A = M_1^{-1}M_2^{-1}\dots M_{n-2}^{-1}M_{n-1}^{-1}U$$

$$\text{Haciendo } L = M_1^{-1}M_2^{-1}\dots M_{n-1}^{-1} = M^{-1}$$

se tienen que $A = LU$

verificaremos que $L = M^{-1}$ es triangular unitaria inferior.

Se sabe por teorema 2.2 que

$$M_1^{-1} = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ m_{21} & 1 & \dots & \dots & \dots & \vdots \\ m_{31} & 0 & 1 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ m_{n1} & 0 & \dots & \dots & \dots & 1 \end{bmatrix}; \text{ en general } M_k^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & \dots & 0 \\ 0 & 1 & \dots & \dots & \dots & \dots & 0 \\ \dots & \dots & \dots & 0 & \dots & \dots & \vdots \\ \dots & \dots & \dots & \dots & 1 & \dots & \vdots \\ \dots & \dots & \dots & \dots & \dots & m_{k+1,k} & 0 \\ 0 & \dots & \dots & \dots & m_{n,k} & 0 & \dots \vdots 1 \end{bmatrix}$$

entonces

$$M_1^{-1}M_2^{-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ m_{21} & 1 & 0 & \dots & 0 \\ m_{31} & m_{32} & 1 & \dots & 0 \\ \vdots & \vdots & 0 & \dots & \vdots \\ m_{n1} & m_{n2} & 0 & \dots & 1 \end{bmatrix}$$

$$M_1^{-1}M_2^{-1}M_3^{-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ m_{21} & 1 & 0 & \dots & 0 \\ m_{31} & m_{32} & 1 & \dots & \vdots \\ \vdots & \vdots & m_{43} & \dots & \vdots \\ \vdots & \vdots & \vdots & \dots & \vdots \\ m_{n1} & m_{n2} & m_{n3} & \dots & 1 \end{bmatrix}$$

continuando este proceso llegamos a

$$M^{-1} = L^{-1} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ m_{21} & 1 & 0 & \dots & 0 & 0 \\ m_{31} & m_{32} & 1 & \dots & \vdots & \vdots \\ \vdots & \vdots & m_{43} & \dots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \dots & 1 & 0 \\ m_{n1} & m_{n2} & m_{n3} & \dots & m_{n,n-1} & 1 \end{bmatrix}$$

Una condición importante para llevar a cabo la eliminación Gaussiana sin intercambio de filas está dada en el siguiente -- teorema.

TEOREMA 1.3:

Si A es una matriz estrictamente diagonal dominante

de orden n (def. 1.3.26), entonces A es no singular. Además. - la eliminación Gaussiana puede ser efectuada en cualquier sistema lineal de la forma $A\bar{x} = \bar{b}$, para obtener su única solución -- sin intercambio de filas o columnas.

Para mostrar que A es no singular, considerar el sistema lineal $A\bar{x} = \bar{0}$ y supóngase que existe un $\bar{x} \neq \bar{0}$ que es solución de este sistema. En este caso, para algún k , $0 < |x_k| = \max_{1 \leq j \leq n} |x_j|$

y como

$$\sum_{j=1}^n a_{ij} x_j = 0 \quad (\forall i = 1, 2, \dots, n)$$

se tiene que

$$a_{kk} x_k = - \sum_{\substack{j=1 \\ j \neq k}}^n a_{kj} x_j$$

esto implica que

$$|a_{kk}| |x_k| \leq \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| |x_j|$$

o sea

$$|a_{kk}| \leq \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}| \frac{|x_j|}{|x_k|} \leq \sum_{\substack{j=1 \\ j \neq k}}^n |a_{kj}|$$

lo que es contradictorio a la definición de matriz estrictamente dominante. Consecuentemente, la única solución para $A\bar{x} = \bar{0}$ es $\bar{x} = \bar{0}$, lo que equivale a decir que A es no singular.

Es fácil hacer la prueba que la eliminación Gaussiana puede ser efectuada sin hacer pivotamiento.

Para ilustrar un procedimiento para el cálculo de las matrices L y U consideraremos el siguiente ejemplo.

EJEMPLO 1.4

Considerar la matriz estrictamente diagonal dominante de orden 4.

$$A = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 2 & 4 & 1 & 0 \\ 1 & 1 & 4 & -1 \\ -1 & 0 & -1 & 3 \end{bmatrix}$$

Los teoremas 1.2 y 1.3 nos garantizan que A puede ser factorada en la forma $A = LU$, donde:

$$L = \begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} \text{ y } U = \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix}$$

Los dieciseis elementos conocidos de A pueden ser usados para determinar parcialmente los diez elementos desconocidos de L y el mismo número en U .

El método usado en este ejemplo arbitrario requiere $l_{11} = l_{22} = l_{33} = l_{44} = 1$; este es conocido como el método de Doolittle. Más adelante en esta sección veremos un método en el cual los elementos de la diagonal de U son unos y es llamado método de Crout. El requerimiento que $l_{ii} = u_{ii}$ para cada valor de i es llamado método de Choleski.

La parte de la multiplicación de L por U

$$L \cdot U = \begin{bmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{bmatrix} \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix}$$

la cual la determina la primera fila de A, resulta en las cuatro ecuaciones

$$u_{11} = 6, \quad u_{12} = 2, \quad u_{13} = 1, \quad u_{14} = -1$$

La parte que determina los restantes elementos en la primera columna de A prevee las ecuaciones

$$l_{21} u_{11} = 2 \quad l_{21} = 1/3$$

$$l_{31} u_{11} = 1 \quad \text{y} \quad l_{31} = 1/6$$

$$l_{41} u_{11} = -1 \quad l_{41} = -1/6$$

En este paso las matrices L y U asumen la forma

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1/3 & 1 & 0 & 0 \\ 1/6 & l_{32} & 1 & 0 \\ -1/6 & l_{42} & l_{43} & 1 \end{bmatrix} \quad \text{y} \quad U = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix}$$

La parte de la multiplicación que determina los restantes elementos en la segunda fila de A da las ecuaciones

$$\frac{2}{3} + u_{22} = 4 \quad u_{22} = \frac{10}{3}$$

$$\frac{1}{3} + u_{23} = 1 \quad \text{así} \quad u_{23} = \frac{2}{3}$$

$$\frac{-1}{3} + u_{24} = 0 \quad u_{24} = \frac{1}{3}$$

y lo que determina los restantes elementos en la segunda columna de A da

$$\begin{aligned} \frac{2}{3} + \frac{10}{3} l_{32} &= 1 & l_{32} &= \frac{1}{5} \\ \frac{-2}{6} + \frac{10}{3} l_{42} &= 0 & \text{así} & l_{42} = \frac{1}{10} \end{aligned}$$

Este proceso es continuado alternando diagonalmente columnas y filas para finalmente obtener:

$$L = \begin{bmatrix} 1 & 0 & 0 & 0 \\ \frac{1}{3} & 1 & 0 & 0 \\ \frac{1}{6} & \frac{1}{5} & 1 & 0 \\ \frac{-1}{6} & \frac{1}{10} & \frac{-9}{27} & 1 \end{bmatrix} \quad \text{y} \quad U = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 0 & \frac{10}{3} & \frac{2}{3} & \frac{1}{2} \\ 0 & 0 & \frac{37}{10} & \frac{-9}{10} \\ 0 & 0 & 0 & \frac{191}{74} \end{bmatrix}$$

Un procedimiento general para factorar matrices en un producto de matrices triangulares está contenido en el siguiente algoritmo.

ALGORITMO 1.5 (Factorización Directa)

Para factorar la matriz

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

en matrices triangulares

$$L = \begin{bmatrix} l_{11} & 0 & \dots & \dots & 0 \\ & \ddots & & & \vdots \\ l_{21} & l_{22} & \dots & & \vdots \\ \vdots & \vdots & & \ddots & 0 \\ l_{n1} & l_{n2} & \dots & \dots & l_{nn} \end{bmatrix} \text{ y } U = \begin{bmatrix} u_{11} & u_{12} & \dots & \dots & u_{1n} \\ 0 & u_{22} & \dots & \dots & u_{2n} \\ \vdots & 0 & \dots & \dots & \vdots \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & \dots & \dots & 0 & u_{nn} \end{bmatrix}$$

1) Para $k = 1, 2, \dots, n$ seleccionar los elementos l_{kk} y u_{kk} de manera que

$$1) \quad l_{kk} u_{kk} = a_{kk} - \sum_{p=1}^{k-1} l_{kp} u_{pk}$$

$$2) \quad l_{ik} = u_{kk}^{-1} \left(a_{ik} - \sum_{p=1}^{k-1} l_{ip} u_{pk} \right) \quad (i = k, k+1, \dots, n)$$

$$3) \quad u_{kj} = l_{kk}^{-1} \left(a_{kj} - \sum_{p=1}^{k-1} l_{kp} u_{pj} \right) \quad (j = j+1, k+2, \dots, n)$$

El algoritmo 1.5 se puede particularizar para los métodos de Doolittle y Crout.

Al factorar una matriz, aunque nuevas matrices L y U son construidas, los valores actuales generados pueden ser colocados en los correspondientes elementos de A , los cuales ya no son necesarios.

Así a_{ij} y l_{ij} , para cada $i = 2, 3, \dots, n$ y $j = 1, 2, \dots, i-1$; $l_{ii} = 1$, para $i = 1, 2, \dots, n$ y $a_{ij} = u_{ij}$ para cada $i = 1, 2, \dots, n$ y $j = i, i+1, \dots, n$, sería el resultado.

A continuación presentaremos el algoritmo para factorar una matriz A de orden n , haciendo los $l_{kk} = 1$ para $k = 1, 2, \dots, n$.

ALGORITMO 1.6: (Método de Doolittle)

Este algoritmo sobre escribe los elementos de U y

L en los correspondientes elementos de A.

$$\begin{array}{l}
 1) \quad \left\{ \begin{array}{l}
 \text{Para } k = 1, 2, \dots, n \\
 1) \quad a_{kj} \longleftarrow u_{kj} = a_{kj} - \sum_{p=1}^{k-1} \ell_{kp} u_{pj} \quad (j = 1, 2, \dots, n) \\
 2) \quad a_{ik} \longleftarrow \ell_{ik} = u_{kk}^{-1} (a_{ik} - \sum_{p=1}^{k-1} \ell_{ip} u_{pk}) \quad (i = k+1, \dots, n)
 \end{array} \right.
 \end{array}$$

El siguiente algoritmo factora una matriz A de orden n atilizando el método de Crout.

ALGORITMO 1.7: (Método de Crout)

En este algoritmo los elementos de la diagonal de U se escogen que sean unos y también los elementos de L y U se sobre escriben en los correspondientes elementos de A.

$$\begin{array}{l}
 1) \quad \left\{ \begin{array}{l}
 \text{Para } k = 1, 2, \dots, n \\
 1) \quad a_{ik} \longleftarrow \ell_{ik} = a_{ik} - \sum_{p=1}^{k-1} \ell_{ip} u_{pk} \quad (i = k, \dots, n) \\
 2) \quad a_{kj} \longleftarrow u_{kj} = \ell_{kk}^{-1} (a_{kj} - \sum_{p=1}^{k-1} \ell_{kp} u_{pj}) \quad (j = k+1, k+2, \dots, n)
 \end{array} \right.
 \end{array}$$

Para su completación, este algoritmo requiere aproximadamente $\frac{n^3}{3}$ multiplicaciones.

Una dificultad puede surgir cuando se usa este algoritmo para obtener la factorización de la matriz de coeficientes de un sistema lineal de ecuaciones que es originada por el hecho que no es empleado el pivotamiento para reducir el efecto del error por redondeo.

El siguiente ejemplo ilustra los cálculos involucrados en el método de Crout de una matriz 3 x 3. También se muestra los -

efectos desastrosos de un elemento pequeño en la diagonal.

EJEMPLO 1.8:

$$\text{Sea } A = \begin{bmatrix} 0.001 & 2.000 & 3.000 \\ -1.000 & 3.712 & 4.623 \\ -2.000 & 1.072 & 5.643 \end{bmatrix}$$

(A es la matriz del ejemplo 2.2.9). En aritmética de cuatro dígitos, la reducción de Crout es como sigue

$$l_{11} = 0.001 \quad u_{11} = u_{22} = u_{33} = 1$$

$$l_{21} = -1.000$$

$$l_{31} = -2.000$$

$$u_{12} = \delta l \left(\frac{2.000}{0.001} \right) = 2000,$$

$$u_{13} = \delta l \left(\frac{3.000}{0.001} \right) = 3000,$$

$$l_{22} = \delta l [3.712 + (1.000)(2000)] = 2004,$$

$$l_{32} = \delta l [1.072 + (2.000)(2000)] = 4001,$$

$$u_{23} = \delta l \left[\frac{4.623 + (1.000)(3.000)}{2004} \right] = 1.500,$$

$$l_{33} = \delta l [5.643 + (2.000)(3000) - (4001)(1.500)] = 4.000$$

Note que el cálculo de l_{33} envuelve la cancelación de tres cifras significativas. Por lo tanto podemos esperar que el valor calculado de l_{33} sea inexacto, y en verdad, el verdadero valor para l_{33} , redondeado a cuatro cifras es 5.922.

El ejemplo 1.8 indica la necesidad de evitar elementos pequeños en la diagonal mediante la reducción de Crout. Como hici

mos en la eliminación Gaussiana, buscaremos un remedio para el problema permutando filas y columnas en la matriz original A . - No es conveniente emplear el pivotamiento completo para la reducción de Crout; por lo tanto usaremos el pivotamiento parcial, que es el intercambio de filas.

El principal obstáculo para incorporar intercambios en el algoritmo de Crout es que no podemos saber que ℓ_{kk} es pequeño - hasta que lo hemos calculado. Sin embargo, en este paso un intercambio en las filas de A cambiará la descomposición de Crout, la cual tendría que ser recalculada. Afortunadamente existe una relación muy simple entre la descomposición de Crout de una matriz y la matriz obtenida intercambiando dos de sus filas. Esta relación es mejor derivada considerando lo que sucede al aplicar la eliminación Gaussiana a A cuando dos de sus filas son - permutadas.

Supóngase que $A_1 = A$ es de orden 5 y la tercera y quinta - fila de A son intercambiadas para obtener una nueva matriz

$$A_1' = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & a_{14}^{(1)} & a_{15}^{(1)} \\ a_{21}^{(1)} & a_{22}^{(1)} & a_{23}^{(1)} & a_{24}^{(1)} & a_{25}^{(1)} \\ a_{51}^{(1)} & a_{52}^{(1)} & a_{53}^{(1)} & a_{54}^{(1)} & a_{55}^{(1)} \\ a_{41}^{(1)} & a_{42}^{(1)} & a_{43}^{(1)} & a_{44}^{(1)} & a_{45}^{(1)} \\ a_{31}^{(1)} & a_{32}^{(1)} & a_{33}^{(1)} & a_{34}^{(1)} & a_{35}^{(1)} \end{bmatrix}.$$

Examinando el algoritmo 2.5, es fácil verificar que des--- pues de un paso de eliminación Gaussiana, A_1' se convierte en

$$A'_2 = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & a_{14}^{(1)} & a_{15}^{(1)} \\ m_{21} & a_{22}^{(2)} & a_{23}^{(2)} & a_{24}^{(2)} & a_{25}^{(2)} \\ m_{51} & a_{52}^{(2)} & a_{53}^{(2)} & a_{54}^{(2)} & a_{55}^{(2)} \\ m_{41} & a_{42}^{(2)} & a_{43}^{(2)} & a_{44}^{(2)} & a_{45}^{(2)} \\ m_{31} & a_{32}^{(2)} & a_{33}^{(2)} & a_{34}^{(2)} & a_{35}^{(2)} \end{bmatrix}$$

Otro paso mas nos resulta

$$A'_3 = \begin{bmatrix} a_{11}^{(1)} & a_{12}^{(1)} & a_{13}^{(1)} & a_{14}^{(1)} & a_{15}^{(1)} \\ m_{21} & a_{22}^{(2)} & a_{23}^{(2)} & a_{24}^{(2)} & a_{25}^{(2)} \\ m_{51} & m_{52} & a_{53}^{(3)} & a_{54}^{(3)} & a_{55}^{(3)} \\ m_{41} & m_{42} & a_{43}^{(3)} & a_{44}^{(3)} & a_{45}^{(3)} \\ m_{31} & m_{32} & a_{33}^{(3)} & a_{34}^{(3)} & a_{35}^{(3)} \end{bmatrix}$$

si continuáramos aplicando el algoritmo 1.7 de esta sección, a la matriz A'_1 , después del paso 1.1 cuando $k = 3$, habremos obtenido el arreglo.

$$\begin{bmatrix} l_{11} & u_{12} & u_{13} & u_{14} & u_{15} \\ l_{21} & l_{22} & u_{23} & u_{24} & u_{25} \\ l_{51} & l_{52} & l_{53} & a_{54} & a_{55} \\ l_{41} & l_{42} & l_{43} & a_{44} & a_{45} \\ l_{31} & l_{32} & l_{33} & a_{34} & a_{35} \end{bmatrix}$$

En otras palabras, después del paso 1.1 cuando $k = 3$, la diferencia entre los arreglos obtenidos de A y A' solamente es que la tercera y la quinta fila han sido intercambiadas.

Este es generalmente el caso. Si despues del paso 1.1 en el algoritmo 1.7 intercambiamos las filas k y ℓ ($\ell > k$), el efecto será el mismo que si intercambiamos las filas k y ℓ en la matriz original antes de comenzar. Podemos efectuar intercambios de filas como está dado en el siguiente algoritmo.

ALGORITMO 1.9

Sea A de orden n . En este algoritmo sobrescribe A con la descomposición de Crout de $I_{n-1, y_{n-1}} \dots I_{1, y_1} A$.

- 1) Para $k = 1, 2, \dots, n$
 - 1) $a_{ik} \leftarrow l_{ik} = a_{ik} - \sum_{p=1}^{k-1} l_{ip} u_{pk}$ ($i = k, \dots, n$)
 - 2) Encontrar y_k tal que $|l_{y_k, k}| \geq |l_{ik}|$ ($i = k, k+1, \dots, n$)
 - 3) $a_{kj} \longleftrightarrow a_{y_k j}$ ($j = 1, 2, \dots, n$)
 - 4) $a_{kj} \leftarrow u_{kj} = l_{kk}^{-1} (a_{kj} - \sum_{p=1}^{k-1} l_{kp} u_{pj})$ ($j = k+1, \dots, n$)

Cuando se conoce que la matriz A es definida positiva (def. 1.5.7), un mejoramiento significativo en la técnica de la factorización de una matriz puede ser analizado con respecto al número de operaciones requeridas.

TEOREMA 1.10:

Si A es una matriz definida positiva de orden n , entonces A es no singular. Además la eliminación Gaussiana puede ser efectuada en cualquier sistema lineal $A\bar{x} = \bar{b}$ para obtener su única solución sin intercambios de filas o columnas.

PRUEBA

Si $\bar{x} \neq \bar{0}$ es un vector que satisface $A\bar{x} = \bar{0}$, entonces $\bar{x}^t A \bar{x} = \bar{0}$, lo que contradice que A sea definida positiva. Consecuentemente, $A\bar{x} = \bar{0}$ tiene una única solución que es el vector $\bar{0}$ y esto implica que A es no singular.

Para mostrar la parte relacionada con la eliminación Gaussiana, sea $A^{[k]}$ la submatriz directriz principal de A para un arbitrario k , $k = 1, 2, \dots, n$.

Si $A^{[k]}$ no es definida positiva, entonces existe un k -vector $\hat{x} = \hat{0}$ tal que $\hat{x}^t A^{[k]} \hat{x} \leq 0$. Si construimos un n -vector $\bar{x} \neq \bar{0}$ de \hat{x} colocando ceros en los últimos $n-k$ espacios. Como

$$\bar{x}^t A \bar{x} = \hat{x}^t A \hat{x} \leq 0$$

A no es definida positiva, y tenemos una contradicción. Consecuentemente, cada submatriz directriz principal de A es definida positiva y por lo tanto no singular.



LEMA 1.11:

Una submatriz principal de una matriz definida positiva es definida positiva.

PRUEBA

Sea A' la submatriz principal formada por la intersección de filas y columnas $i_1 < i_2 < \dots < i_r$. Sea $\bar{x}' \neq 0$ un r -vector. Sea \bar{x} el n -vector definido por

$$\begin{aligned} x_{ik} &= x'_k & (k = 1, 2, \dots, r) \\ x_j &= 0 & (j \neq i_1, i_2, \dots, i_r) \end{aligned}$$

entonces $\bar{x} \neq 0$, y en estas condiciones se tiene que

$$\bar{x}^t A \bar{x} = \bar{x}'^t A \bar{x}' \quad ;$$

luego como A es definida positiva, $\bar{x}^t A \bar{x} = \bar{x}'^t A \bar{x}' > 0$, y A' es -
definida positiva.

TEOREMA 1.12:

Si A es definida positiva, entonces existe una úni-
ca matriz triangular inferior L con elementos positivos en la --
diagonal tal que $A = LL^t$.

PRUEBA

La prueba es por inducción en el orden de A. Si A es defi-
nida positiva, entonces $a_{11} > 0$, y L está definida unicamente -
por $l_{11} = \sqrt{a_{11}}$.

Supóngase que la aseveración es verdadera para matrices -
de orden n-1 y A' es definida positiva de orden n. Como A' es si-
métrica, puede ser particionada en la forma

$$A' = \begin{bmatrix} A & \bar{y} \\ \bar{y}^t & a \end{bmatrix}$$

Por el lema 1.11, A' es definida positiva. Buscaremos una matriz
triangular inferior L' que satisfaga $L'L'^t = A'$.

Sea L' particionada en la forma

$$L' = \begin{bmatrix} L & 0 \\ \bar{x}^t & l \end{bmatrix}$$

Luego efectuando

$$A' = L'L'^t$$

$$\begin{bmatrix} A & \bar{y} \\ \bar{y}^t & a \end{bmatrix} = \begin{bmatrix} L & 0 \\ \bar{x}^t & \ell \end{bmatrix} \begin{bmatrix} L^t & \bar{x} \\ 0 & \ell \end{bmatrix}$$

$$= \begin{bmatrix} LL^t & L\bar{x} \\ \bar{x}^t L^t & \bar{x}^t \bar{x} + \ell^2 \end{bmatrix}$$

Entonces requerimos que:

$$A = LL^t \quad (1.4)$$

$$L\bar{x} = \bar{y} \quad (1.5)$$

$$\bar{x}^t L^t = \bar{y}^t \quad (1.6)$$

$$\bar{x}^t \bar{x} + \ell^2 = a \quad (1.7)$$

Por hipótesis inductiva, existe una única matriz triangular inferior L con elementos positivos en la diagonal satisfaciendo (1.4). Como L es no singular, $\bar{x} = L^{-1}\bar{y}$ es el único vector satisfaciendo (1.5) y (1.6). Finalmente, si $a - \bar{x}^t \bar{x} > 0$, entonces ℓ estará definida en forma única por

$$\ell = \sqrt{a - \bar{x}^t \bar{x}}$$

Para mostrar que $a - \bar{x}^t \bar{x} > 0$, es de notar que la no singularidad de L implica que A es no singular. Sea $\bar{b} = A^{-1}\bar{y}$. Entonces, como A es definida positiva

$$\begin{aligned}
0 &< [\bar{\mathbf{b}}^t, -1] \begin{bmatrix} \mathbf{A} & \bar{\mathbf{y}} \\ \bar{\mathbf{y}}^t & a \end{bmatrix} \begin{bmatrix} \bar{\mathbf{b}} \\ -1 \end{bmatrix} \\
&= [\bar{\mathbf{b}}^t \mathbf{A} \bar{\mathbf{y}}^t, \bar{\mathbf{b}}^t \bar{\mathbf{y}} - a] \begin{bmatrix} \bar{\mathbf{b}} \\ -1 \end{bmatrix} \\
&= \bar{\mathbf{b}}^t \mathbf{A} \bar{\mathbf{b}} - \bar{\mathbf{y}}^t \bar{\mathbf{b}} - \bar{\mathbf{b}}^t \bar{\mathbf{y}} + a \\
&= \bar{\mathbf{b}}^t \mathbf{A} \bar{\mathbf{b}} - 2\bar{\mathbf{b}}^t \bar{\mathbf{y}} + a \\
&= \bar{\mathbf{b}}^t \mathbf{A} \mathbf{A}^{-1} \bar{\mathbf{y}} - 2\bar{\mathbf{b}}^t \bar{\mathbf{y}} + a, \quad \text{ya que } \bar{\mathbf{b}} = \mathbf{A}^{-1} \bar{\mathbf{y}} \\
&= \bar{\mathbf{b}}^t \bar{\mathbf{y}} - 2\bar{\mathbf{b}}^t \bar{\mathbf{y}} + a \\
&= a - \bar{\mathbf{y}}^t (\mathbf{L} \mathbf{L}^t)^{-1} \bar{\mathbf{y}} = a - (\mathbf{L}^{-1} \bar{\mathbf{y}})^t (\mathbf{L}^{-1} \bar{\mathbf{y}}) \\
&= a - \bar{\mathbf{x}}^t \bar{\mathbf{x}}
\end{aligned}$$



La prueba del teorema 1.12 permite construir la descomposición de Cholesky de una matriz calculando sucesivamente las descomposiciones de sus submatrices directrices principales. Sea A_k la submatriz directriz principal de orden k y sea $A_k = L_k L_k^t$ la descomposición de Cholesky de A_k . Entonces si

$$A_k = \begin{bmatrix} A_{k-1} & \bar{\mathbf{y}}_k \\ \bar{\mathbf{y}}_k^t & a_{kk} \end{bmatrix}$$

de la prueba del teorema sigue que

$$L_k = \begin{bmatrix} L_{k-1} & 0 \\ \bar{\mathbf{x}}_k^t & \ell_{kk} \end{bmatrix}$$

donde

$$\bar{x}_k = L_{k-1}^{-1} \bar{y}_k \quad (1.8)$$

$$y \quad \ell_{kk} = \sqrt{a_{kk} - \bar{x}_k^t x_k}$$

Construyendo un algoritmo para calcular la descomposición de Cholesky de una matriz definida positiva A , notaremos que, como A es simétrica, necesitamos solamente trabajar con la mitad inferior. Además los elementos de L pueden sobre escribir los correspondientes elementos de A . Por supuesto, calculando \bar{x}_k de (1.8) no haremos la forma L_{k-1}^{-1} ; mejor resolveremos el sistema

$$L_{k-1} \bar{x}_k = \bar{y}_k$$

ALGORITMO 1.13: (Reducción de Cholesky)

Sea A definida positiva de orden n . Este algoritmo escribe la matriz L de la reducción de Cholesky de A en la mitad inferior del arreglo de A .

$$\begin{array}{l}
 1) \quad \left| \begin{array}{l}
 \text{Para } k = 1, 2, \dots, n \\
 1) \quad \left| \begin{array}{l}
 \text{Para } i = 1, 2, \dots, k-1 \\
 1) \quad a_{ik} \leftarrow \ell_{ki} = \ell_{ii}^{-1} \left(a_{ki} - \sum_{j=1}^{i-1} \ell_{ij} \ell_{kj} \right) \\
 2) \quad a_{kk} \leftarrow \ell_{kk} = \left(a_{kk} - \sum_{j=1}^{k-1} \ell_{kj}^2 \right)^{1/2}
 \end{array} \right. \\
 \end{array} \right.
 \end{array}$$

Si A es definida positiva, el algoritmo siempre puede ser corrido por completo. Requiere cerca de $\frac{n^3}{6}$ multiplicaciones, la mitad del número requerido para la eliminación Gaussiana ó la

reducción de Crout. Esto era de esperarse, ya que el algoritmo toma ventaja de la simetría de A para reducir los cálculos involucrados.

A continuación daremos un ejemplo para ilustrar el algoritmo 1.13.

EJEMPLO

Sea la matriz definida positiva

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

Factorar dicha matriz en la forma $A = LL^t$ donde L es triangular inferior.

Paso 1: $k = 1, 2, 3.$

$$\boxed{k = 1}$$

$$1) \begin{array}{l} i = 1 - 1 = 0 \\ 1) \end{array}$$

$$2) a_{11} \leftarrow l_{11} = \sqrt{a_{11}} = \sqrt{2}$$

$$\boxed{k = 2}$$

$$1) \begin{array}{l} i = 1 \\ 1) \end{array}$$

$$1) a_{21} \leftarrow l_{21} = \frac{1}{l_{11}} (a_{21})$$

$$l_{21} = \frac{1}{\sqrt{2}} (-1) = -\frac{\sqrt{2}}{2}$$

$$2) a_{22} \leftarrow l_{22} = \left(a_{22} - l_{21}^2 \right)^{1/2}$$

$$l_{22} = \left(2 - \left(\frac{\sqrt{2}}{2} \right)^2 \right)^{1/2} = \frac{\sqrt{6}}{2}$$

$$\boxed{k = 3}$$

$$1) i = 1, 2$$

$$1) a_{31} \leftarrow l_{31} = \frac{1}{l_{11}} (a_{31})$$

$$l_{31} = \frac{1}{\sqrt{2}} (0) = 0$$

$$a_{32} \leftarrow l_{32} = \frac{1}{l_{22}} (a_{32} - l_{21} l_{31})$$

$$l_{32} = \frac{2}{\sqrt{6}} \left[-1 - \left(-\frac{\sqrt{2}}{2} \right) (0) \right] = -\frac{\sqrt{6}}{3}$$

$$2) a_{33} \leftarrow l_{33} = \left(a_{33} - l_{31}^2 - l_{32}^2 \right)^{1/2}$$

$$l_{33} = \left(2 - 0^2 - \left(-\frac{\sqrt{6}}{3} \right)^2 \right)^{1/2} = \frac{2}{3} \sqrt{3}$$

Luego A queda factorada así:

$$A = LL^t = \begin{bmatrix} \sqrt{2} & 0 & 0 \\ -\frac{\sqrt{2}}{2} & \frac{\sqrt{6}}{2} & 0 \\ 0 & -\frac{\sqrt{6}}{3} & \frac{2}{3}\sqrt{3} \end{bmatrix} \begin{bmatrix} \sqrt{2} & -\frac{\sqrt{2}}{2} & 0 \\ 0 & \frac{\sqrt{6}}{2} & -\frac{\sqrt{6}}{3} \\ 0 & 0 & \frac{2}{3}\sqrt{3} \end{bmatrix}$$

EJERCICIOS

1.1 Elaborar un algoritmo que sobrescriba una matriz definida positiva A con su inversa. Trabajar solamente con la mitad inferior del arreglo A. [Sugerencia: Usar el algoritmo 1.13

para calcular L^{-1} de la descomposición de Cholesky; luego calcular L^{-1} y $L^{-t}L^{-1}$].

1.2 Mostrar que la siguiente matriz es definida positiva y obtenga una factorización de la matriz en la forma $A = LL^t$ usando método de Cholesky.

$$A = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 2 & 4 & 1 & 0 \\ 1 & 1 & 4 & -1 \\ -1 & 0 & -1 & 3 \end{bmatrix}$$

1.3 Mostrar que la matriz

$$A = \begin{bmatrix} 4 & -1 & 1 \\ -1 & 4.25 & 2.75 \\ 1 & 2.75 & 3.5 \end{bmatrix},$$

es definida positiva.

2. RESOLUCION DE SISTEMAS DE ECUACIONES LINEALES UTILIZANDO LA FACTORIZACION.

En esta sección daremos algoritmos para resolver el sistema

$$A\bar{x} = \bar{b} \quad (2.1)$$

donde A es una matriz de orden n , utilizando la factorización de matrices tratadas en la sección anterior.

La reducción de Crout de A con pivotamiento da permutaciones elementales P_i ($i = 1, 2, \dots, n-1$) tal que

$$P_{n-1}P_{n-2}\dots P_1A = LU,$$

donde L es triangular inferior y U es triangular unitaria superior. Por lo tanto la solución de (2.1) está dada por

$$\bar{x} = U^{-1}L^{-1}P_{n-1}P_{n-2}\dots P_1\bar{b}$$

Esto nos guía al siguiente algoritmo.

ALGORITMO 2.1:

Sea LU la descomposición de Crout de $P_{n-1}P_{n-2}\dots P_1A$, y sea un vector \bar{b} dado. Este algoritmo calcula la solución de (2.1).

- 1) $\bar{x} = \bar{b}$
- 2) $\left\{ \begin{array}{l} \text{Para } k = 1, 2, \dots, n-1 \\ 1) \bar{x} \leftarrow P_k \bar{x} \end{array} \right.$
- 3) $\bar{x} \leftarrow L^{-1}\bar{x}$
- 4) $\bar{x} \leftarrow U^{-1}\bar{x}$

▲

Este algoritmo puede ser corrido hasta el final si A es no singular y requiere alrededor de n^2 multiplicaciones. En la parte 4 un mínimo ahorro puede ser efectuado tomando ventaja del hecho que U es triangular unitaria superior.

Cuando A es simétrica y definida positiva, tiene una descomposición Cholesky en la forma $A = LL^t$ donde L es triangular inferior. Así la solución de (2.1) está dada por

$$\bar{x} = L^{-t}L^{-1}\bar{b}$$

y obtenemos el siguiente algoritmo.

ALGORITMO 2.2:

Sea LL^t la descomposición Cholesky de la matriz A definida positiva, y sea el vector \bar{b} dado. Este algoritmo calcula la solución de (2.1).

- 1) $\bar{x} = L^{-1}\bar{b}$
- 2) $\bar{x} \leftarrow L^{-t}\bar{x}$

Las descomposiciones anteriores pueden ser usadas para encontrar la inversa de A , ya sea invirtiendo y multiplicando las matrices en la descomposición ó resolviendo las n ecuaciones lineales

$$A\bar{x}_i = e_i$$

para las columnas de la inversa. Otra vez hacemos énfasis que la inversa de una matriz es rara vez requerida. Los algoritmos de factorización pueden ser simplificados considerablemente en el caso de matrices banda debida a la gran cantidad de ceros que aparecen en estas matrices.

Es particularmente interesante observar la forma que asumen el método de Crout ó el de Doolittle en este caso. Para ilustrar la situación, supóngase que una matriz tridiagonal

$$A = \begin{bmatrix} a_{11} & a_{12} & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ a_{21} & a_{22} & a_{23} & \cdot & \cdot & \cdot & \cdot & 0 \\ 0 & a_{32} & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & a_{n-1,n} \\ \cdot & a_{n-1,n} \\ 0 \dots \dots 0 & \cdot & a_{n,n-1} & \cdot & \cdot & \cdot & \cdot & a_{nn} \end{bmatrix}$$

puede ser factorada en las siguientes matrices triangulares

$$\bar{L} = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{bmatrix} \quad U = \begin{bmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & u_{nn} \end{bmatrix}$$

Como A tiene solamente (3n-2) elementos diferentes de cero, existen solamente (3n-2) condiciones para ser aplicadas para determinar los de L y U. Supóngase que las matrices pueden actualmente ser encontradas en la forma

$$L = \begin{bmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & & \vdots \\ 0 & \cdot & & 0 \\ \vdots & \cdot & & \vdots \\ 0 & \cdot & \cdot & 0 \\ & & & l_{n,n-1} & l_{nn} \end{bmatrix} \quad \text{y} \quad U = \begin{bmatrix} 1 & u_{12} & 0 & \dots & 0 \\ 0 & 1 & \cdot & \cdot & \cdot \\ \vdots & \cdot & \cdot & \cdot & 0 \\ \vdots & \cdot & \cdot & \cdot & u_{n-1,n} \\ \vdots & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \cdot & 1 \end{bmatrix}$$

En esta forma, existen (2n-1) elementos no determinados de L y (n-1) elementos no determinados de U, lo cual totaliza el número de condiciones mencionadas arriba, y los elementos ceros de A son obtenidos automáticamente.

La multiplicación involucrada con A = LU da, además de los elementos ceros, las ecuaciones

$$a_{11} = l_{11}$$

$$a_{i,i-1} = l_{i,i-1} \quad \text{Para cada } i = 1, 2, \dots, n$$

$$u_{i,i+1} = \frac{a_{i,i+1}}{l_{ii}} \quad \text{Para cada } i = 1, 2, \dots, n-1$$

$$l_{i+1,i+1} = a_{i+1,i+1} - l_{i+1,i} u_{i,i+1} \quad (i = 1, 2, \dots, n-1)$$

Un completo algoritmo para resolver un sistema de ecuacio--

nes lineales cuya matriz de coeficientes es tridiagonal es presentado a continuación.

ALGORITMO 2.3: (Reducción de Crout para Sistemas lineales Tridimensionales)

Para resolver el sistema de ecuaciones lineales:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 &= b_1 \equiv a_{1,n+1} \\ a_{21}x_1 + a_{22}x_2 + a_{23}x_3 &= b_2 \equiv a_{2,n+1} \\ a_{32}x_2 + a_{33}x_3 + a_{34}x_4 &= b_3 \equiv a_{3,n+1} \\ &\vdots \\ &\vdots \\ a_{n,n-1}x_{n-1} + a_{nn}x_n &= b_n \equiv a_{n,n+1} \end{aligned}$$

Construir la matriz aumentada $[A,b]$

1) Hacer $\ell_{11} = a_{11}$

2) Para $i = 2, 3, \dots, n$
 1) $\ell_{i,i-1} = a_{i,i-1}$

3) Para $i = 1, 2, \dots, n-1$
 1) $u_{i,i+1} = \frac{a_{i,i+1}}{\ell_{ii}}$
 2) $\ell_{i+1,i+1} = a_{i+1,i+1} - \ell_{i+1,i}u_{i,i+1}$

4) $z_1 = \frac{a_{1,n+1}}{\ell_{11}}$

5) $z_i = \ell_{ii}^{-1}(a_{i,n+1} - \ell_{i,i-1}z_{i-1})$ (para $i = 2, 3, \dots, n$)

6) Hacer $x_n = z_n$

7) $x_i = z_i - u_{i,i+1}x_{i+1}$ (para $i = n-1, n-2, \dots, 1$)

Este algoritmo puede ser adaptado para almacenar los elementos

tos de U y L, en los correspondientes elementos de A.

También se tiene que el algoritmo requiere solamente $(5n-4)$ multiplicaciones y consecuentemente tiene considerables ventajas computacionales sobre los métodos que no consideran la tridiagonalidad de la matriz, especialmente para valores grandes de n.

EJEMPLO 2.4:

Para ilustrar el proceso involucrado en el algoritmo 2.3, consideremos el sistema tridiagonal de ecuaciones

$$\begin{aligned} 2x_1 - x_2 &= 1 \\ -x_1 + 2x_2 - x_3 &= 0 \\ -x_2 + 2x_3 - x_4 &= 0 \\ -x_3 + 2x_4 &= 1 \end{aligned}$$

cuya matriz aumentada es

$$\left[\begin{array}{cccc|c} 2 & -1 & 0 & 0 & 1 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 2 & 1 \end{array} \right]$$

1) $l_{11} = 2$

2) Para $i = 2, 3, 4$.

$$l_{21} = a_{21} = -1 ; l_{32} = a_{32} = -1 ; l_{43} = a_{43} = -1$$

3) Para $i = 1, 2, 3$.

$$\left| \begin{aligned} u_{12} &= \frac{a_{12}}{l_{11}} = -\frac{1}{2} ; l_{22} = a_{22} - l_{21}u_{12} = 2 - (-1)\left(-\frac{1}{2}\right) = \frac{3}{2} ; \\ u_{23} &= \frac{a_{23}}{l_{22}} = \frac{-1}{3/2} = -\frac{2}{3} ; l_{33} = a_{33} - l_{32}u_{23} = 2 - (-1)\left(-\frac{2}{3}\right) = \frac{4}{3} ; \end{aligned} \right.$$

$$\left\{ \begin{array}{l} u_{34} = \frac{a_{34}}{\ell_{33}} = \frac{-1}{4/3} = -\frac{3}{4} \quad ; \quad \ell_{44} = a_{44} - \ell_{43}u_{34} = 2 - (-1)\left(-\frac{3}{4}\right) = \frac{5}{4} . \end{array} \right.$$

$$4) \quad z_1 = \frac{a_{1,n+1}}{\ell_{11}} = \frac{1}{2}$$

5) Para $i = 2, 3, 4$.

$$z_2 = \ell_{22}^{-1}(a_{25} - \ell_{21}z_1) = \frac{2}{3}[0 - (-1)\left(\frac{1}{2}\right)] = \frac{1}{3}$$

$$z_3 = \ell_{33}^{-1}(a_{35} - \ell_{32}z_2) = \frac{3}{4}[0 - (-1)\left(\frac{1}{3}\right)] = \frac{1}{4}$$

$$z_4 = \ell_{44}^{-1}(a_{45} - \ell_{43}z_3) = \frac{4}{5}[1 - (-1)\left(\frac{1}{4}\right)] = 1$$

$$6) \quad x_4 = z_4 = 1$$

7) Para $i = 3, 2, 1$.

$$x_3 = z_3 - u_{34}x_4 = \frac{1}{4} - \left(-\frac{3}{4}\right)(1) = 1$$

$$x_2 = z_2 - u_{23}x_3 = \frac{1}{3} - \left(-\frac{2}{3}\right)(1) = 1$$

$$x_1 = z_1 - u_{12}x_2 = \frac{1}{2} - \left(-\frac{1}{2}\right)(1) = 1$$

El algoritmo ha factorado la matriz involucrada en:

$$\begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & \frac{3}{2} & 0 & 0 \\ 0 & -1 & \frac{4}{3} & 0 \\ 0 & 0 & -1 & \frac{5}{4} \end{bmatrix} \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & 0 \\ 0 & 0 & 1 & -\frac{3}{4} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

y da la solución correcta, $x_1 = x_2 = x_3 = x_4 = 1$.

El algoritmo 2.3 puede ser aplicado siempre que $\ell_{ii} \neq 0$ para cada $i = 1, 2, \dots, n$. Cualquiera de las dos condiciones que aseguran esto son:

- a) que la matriz de coeficientes sea definida positiva ó
 b) que la matriz sea estrictamente diagonal dominante.

Una condición adicional que asegura que este algoritmo pueda ser aplicado, está dada en el siguiente teorema.

TEOREMA 2.5:

Si $|a_{11}| > |a_{12}|$ y $|a_{ii}| \geq |a_{i,i-1}| + |a_{i,i+1}|$ para cada $i = 2, 3, \dots, n-1$, y $|a_{nn}| > |a_{n,n-1}|$, entonces A es no singular y los valores de ℓ_{ii} descritos en el algoritmo 2.3 son diferentes de cero para cada $i = 1, 2, \dots, n$.

OBSERVACION 2.6:

El teorema anterior es un caso particular de las matrices estrictamente diagonal dominante.

EJERCICIOS

2.1 Resolver el sistema $A\bar{x} = \bar{b}$ utilizando los algoritmos 1.13 y

2.2 de este capítulo

$$A = \begin{bmatrix} 6 & 2 & 1 & -1 \\ 2 & 4 & 1 & 0 \\ 1 & 1 & 4 & -1 \\ -1 & 0 & -1 & 3 \end{bmatrix} \quad \text{y} \quad \bar{b} = \begin{bmatrix} 1 \\ 2 \\ -1 \\ 3 \end{bmatrix}$$

2.2 Resolver el sistema

$$\begin{aligned} 4x_1 - x_2 + x_3 &= 2 \\ -x_1 + 4.25x_2 + 2.75x_3 &= 1 \\ x_1 + 2.75x_2 + 3.5x_3 &= -1, \end{aligned}$$

utilizando los algoritmos 1.7 y 2.1

2.3 Elaborar un algoritmo para resolver sistemas de ecuaciones -

lineales cuya matriz de coeficientes sea Hessenberg superior.

2.4 Sea la factorización de Crout.

$$LU = \begin{bmatrix} 4 & 0 & 0 \\ 1 & 5/2 & 0 \\ -1 & 1/2 & 7/5 \end{bmatrix} \begin{bmatrix} 1 & 1/2 & -1/2 \\ 1 & 1 & 1/5 \\ 0 & 0 & 1 \end{bmatrix}$$

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

Resolver el sistema $A\bar{x} = \bar{b}$ utilizando el algoritmo 2.1, dado $\bar{b} = [2, 1, -1]^t$.

CAPITULO IV

E S T A B I L I D A D

1. NORMAS DE VECTORES

La noción de norma vectorial surge del intento de generalizar la idea de la longitud de un vector en \mathbb{R} ó \mathbb{R}^2 . El ejemplo mas simple de norma es la función *valor absoluto*, la cual mide la distancia de un escalar al origen. El valor absoluto $|x|$ de un número x puede ser definido por la ecuación

$$|x| = \sqrt{x^2}$$

y tiene las tres bien conocidas propiedades:

1. $x \neq 0 \Rightarrow |x| > 0$,
2. $|ax| = |a||x|$, (1.1)
3. $|x+y| \leq |x| + |y|$

Según (1.1.1), el valor absoluto se dice que es definido positivo. De (1.1.2), se dice que es homogéneo. La tercera condición es llamada *desigualdad triangular* por razones que seran mas claras adelante.

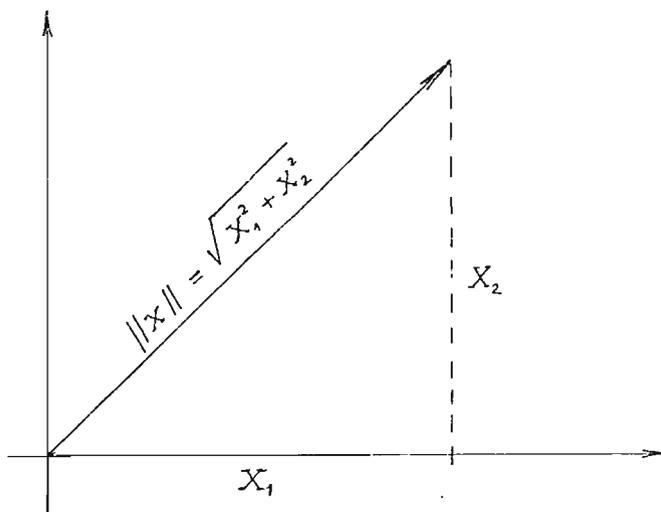


Fig. 1

Un ejemplo poco trivial de norma es la longitud euclídea-
na de un vector en \mathbb{R}^2 . Para cualquier $\bar{x} \in \mathbb{R}^2$, es denotada por

$$\|\bar{x}\|_2 = \sqrt{x_1^2 + x_2^2} \quad (1.2)$$

Por el teorema de Pitágoras, $\|\bar{x}\|_2$ es la distancia de \bar{x} al origen (Fig. 1).

La norma Euclídea $\|\cdot\|_2$ satisface tres condiciones que corresponden a las propiedades (1.1) satisfechas por el valor absoluto. Específicamente,

1. $\bar{x} \neq 0 \Rightarrow \|\bar{x}\|_2 > 0$
2. $\|a\bar{x}\|_2 = |a| \|\bar{x}\|_2$
3. $\|\bar{x} + \bar{y}\|_2 \leq \|\bar{x}\|_2 + \|\bar{y}\|_2$

Las primeras dos condiciones son claras de la definición de $\|\cdot\|_2$. La tercera condición puede ser interpretado geométricamente del hecho que la longitud del lado de un triángulo no

es mayor que la suma de las longitudes de los otros dos lados, de donde proviene el nombre de desigualdad triangular (Fig. 2).

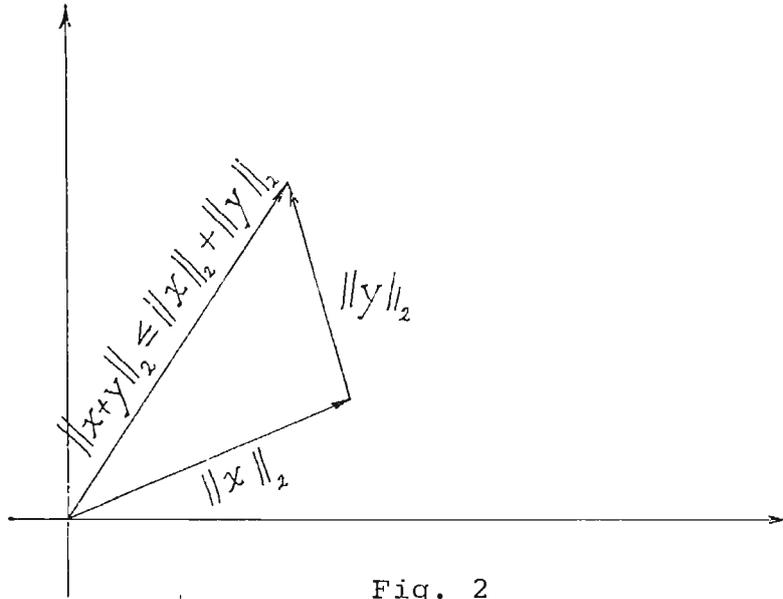


Fig. 2

Una manera natural de extender estas ideas a \mathbb{R}^n es generalizar la fórmula (1.2) en la manera obvia de obtener una longitud Euclídeana de un vector en \mathbb{R}^n . Sin embargo, esto es algo limitado, pues existen otras funciones que miden la longitud de un vector. Por ejemplo, el número $\psi_1(\bar{x})$ definido para $\bar{x} \in \mathbb{R}^2$ por

$$\psi_1(x) = \text{máx} \{ |x_1|, |x_2| \} \quad (1.3)$$

es una medida razonable de la longitud de \bar{x} y es computacionalmente mas conveniente que (1.2).

DEFINICION 1.1

La norma de un vector (ó simplemente norma) en \mathbb{R}^n , es una función $\psi: \mathbb{R}^n \rightarrow \mathbb{R}$ que satisface las siguientes condiciones:

1. $\bar{x} \neq 0 \Rightarrow \psi(\bar{x}) > 0$
2. $\psi(a\bar{x}) = |a|\psi(\bar{x})$
3. $\psi(\bar{x} + \bar{y}) \leq \psi(\bar{x}) + \psi(\bar{y})$

El valor absoluto es una norma en \mathbb{R}^1 , y la función $\|\cdot\|_2$ definida por (1.2) es una norma en \mathbb{R}^2 . La función ψ definida por (1.3) es también una norma en \mathbb{R}^2 .

La definición de norma tiene algunas consecuencias elementales. Por la propiedad 1.1.1, la función ψ es positiva, excepto en el vector cero. Por propiedad 1.1.2,

$$\psi(0) = \psi(0 \cdot \bar{x}) = 0\psi(\bar{x}) = 0,$$

así que la norma del vector cero es cero. También de 1.1.2, se sigue que

$$\psi(-\bar{x}) = \psi(-1 \cdot \bar{x}) = |-1|\psi(\bar{x}) = \psi(\bar{x})$$

Una importante desigualdad es la siguiente:

$$|\psi(\bar{x}) - \psi(\bar{y})| \leq \psi(\bar{x} - \bar{y}) \quad (1.4)$$

lo cual en términos de la norma euclídea dice que la longitud de un lado de un triángulo no es menor que la diferencia de la longitud de los otros dos lados (Fig. 3). Para establecerlo, note que

$$\psi(\bar{x}) = \psi[(\bar{x} - \bar{y}) + \bar{y}] \leq \psi(\bar{x} - \bar{y}) + \psi(\bar{y})$$

Por lo tanto

$$\psi(\bar{x}) - \psi(\bar{y}) \leq \psi(\bar{x} - \bar{y}) \quad (1.5)$$

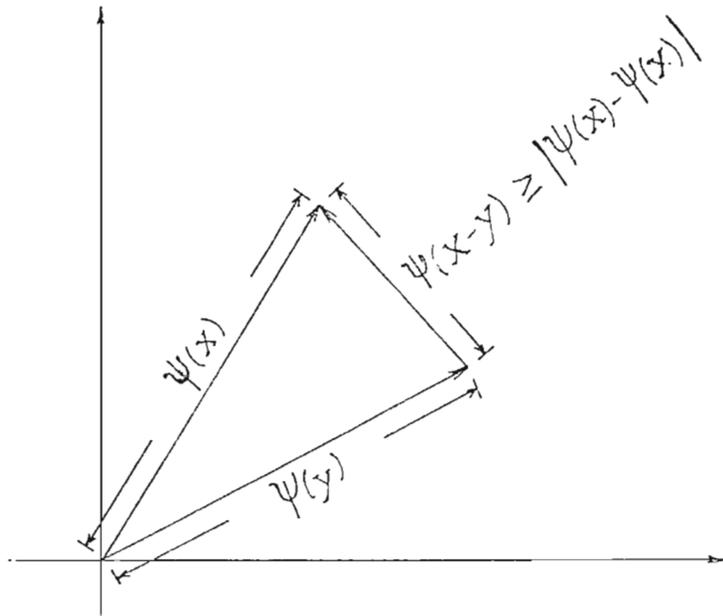


Fig. 3

Si \bar{x} e \bar{y} son intercambiados en (1.5), obtenemos

$$\psi(\bar{y}) - \psi(\bar{x}) \leq \psi(\bar{y} - \bar{x}) = \psi(\bar{x} - \bar{y}) \quad (1.6)$$

Las desigualdades (1.5) y (1.6) implican a (1.4).

Si \bar{y} es colocado en lugar de $-\bar{y}$ en (1.4), resulta la siguiente desigualdad

$$|\psi(\bar{x}) - \psi(\bar{y})| \leq \psi(\bar{x} + \bar{y})$$

Regresemos ahora a la construcción de normas de vectores. Comencemos con tres normas en \mathbb{R}^n que son frecuentemente usadas en procesos de análisis matricial.

Ellas son las normas 1, 2 y ∞ definidas por

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

$$\|\bar{x}\|_2 = \sqrt{\sum_{i=1}^n x_i^2} = \sqrt{\bar{x}^t \bar{x}},$$

y

$$\|\bar{x}\|_\infty = \text{máx} \{|x_i| ; i = 1, 2, \dots, n\}.$$

La norma-2 es la generalización de la longitud euclidea de vectores de \mathbb{R}^2 y \mathbb{R}^3 en \mathbb{R}^n , también es llamada norma euclidea. La norma ∞ es algunas veces llamada la norma máxima ó la norma *Chevychev*.

Las tres funciones $\|\cdot\|_1$, $\|\cdot\|_2$ y $\|\cdot\|_\infty$ son obviamente definidas positivas y homogéneas. Además es fácil mostrar que $\|\cdot\|_1$ y $\|\cdot\|_\infty$ satisface la desigualdad triangular. En efecto,

$$\begin{aligned} 1) \|\bar{x} + \bar{y}\|_1 &= \sum_{i=1}^n |x_i + y_i| \\ &\leq \sum_{i=1}^n (|x_i| + |y_i|) \\ &= \sum_{i=1}^n |x_i| + \sum_{i=1}^n |y_i| \\ &= \|\bar{x}\|_1 + \|\bar{y}\|_1 \end{aligned}$$

$$\begin{aligned} 2) \|\bar{x} + \bar{y}\|_\infty &= \text{máx} \{|x_i + y_i| : i = 1, 2, \dots, n\} \\ &\leq \text{máx} \{|x_i| : i = 1, 2, \dots, n\} + \text{máx} \{|y_i| : i = 1, 2, \dots, n\} \\ &= \|\bar{x}\|_\infty + \|\bar{y}\|_\infty \end{aligned}$$

Para establecer la desigualdad triangular para la norma-2,

requerimos de una desigualdad que es importante y es la -- bien conocida desigualdad de Cauchy.

TEOREMA 1.2:

Para todo $\bar{x}, \bar{y} \in \mathbb{R}^n$,

$$|\bar{x}^t \bar{y}| \leq \|\bar{x}\|_2 \|\bar{y}\|_2 \quad (1.7)$$

se cumple la igualdad sí y solo sí \bar{x} é \bar{y} son linealmente de-- pendientes.

PRUEBA

Si \bar{x} ó \bar{y} es cero, el teorema es trivialmente cierto. Por lo tanto supóngase que \bar{x} é \bar{y} son distintas de cero. Entonces $\|\bar{x}\|_2$ y $\|\bar{y}\|_2$ son diferentes de cero.

Sean $\bar{x}_1 = \frac{\bar{x}}{\|\bar{x}\|_2}$ y $\bar{y}_1 = \frac{\bar{y}}{\|\bar{y}\|_2}$. Entonces $\|\bar{x}_1\|_2 = \|\bar{y}_1\|_2 = 1$,

y la desigualdad (1.7) es equivalente a la desigualdad

$$|\bar{x}_1^t \bar{y}_1| \leq 1$$

Ahora supóngase que $\bar{x}_1^t \bar{y}_1 \geq 0$. Entonces

$$\begin{aligned} 0 \leq \|\bar{x}_1 - \bar{y}_1\|_2^2 &= (\bar{x}_1 - \bar{y}_1)^t (\bar{x}_1 - \bar{y}_1) \\ &= \bar{x}_1^t \bar{x}_1 + \bar{y}_1^t \bar{y}_1 - 2\bar{x}_1^t \bar{y}_1 \\ &= \|\bar{x}_1\|_2^2 + \|\bar{y}_1\|_2^2 - 2\bar{x}_1^t \bar{y}_1 \\ &= 1 + 1 - 2\bar{x}_1^t \bar{y}_1 \\ &= 2 - 2\bar{x}_1^t \bar{y}_1 \end{aligned} \quad (1.8)$$

Por lo tanto $\bar{x}_1^t \bar{y}_1 \leq 1$.

Si $\bar{x}_1 \bar{y}_1 < 0$. Entonces

$$\begin{aligned} 0 \leq \|\bar{x}_1 + \bar{y}_1\|_2^2 &= (\bar{x}_1 + \bar{y}_1)^t (\bar{x}_1 + \bar{y}_1) \\ &= \bar{x}_1^t \bar{x}_1 + \bar{y}_1^t \bar{y}_1 + 2\bar{x}_1^t \bar{y}_1 \\ &= \|\bar{x}_1\|_2^2 + \|\bar{y}_1\|_2^2 + 2\bar{x}_1^t \bar{y}_1 \\ &= 2 + 2\bar{x}_1^t \bar{y}_1 \end{aligned}$$

de aquí que $\bar{x}_1^t \bar{y}_1 \leq 1$.

Ahora supóngase que $|\bar{x}^t \bar{y}| = \|\bar{x}\|_2 \|\bar{y}\|_2$ con $\bar{x}^t \bar{y} \geq 0$. Esto puede suceder cuando la desigualdad en (1.8) es una igualdad. Por lo tanto $\|\bar{x}_1 - \bar{y}_1\|_2 = 0$, y $\bar{x}_1 = \bar{y}_1$. Esto implica que

$$\begin{aligned} \frac{\bar{x}}{\|\bar{x}\|_2} &= \frac{\bar{y}}{\|\bar{y}\|_2} \\ \bar{x} &= \frac{\|\bar{x}\|_2}{\|\bar{y}\|_2} \bar{y} \end{aligned}$$

y por lo tanto \bar{x} é \bar{y} son linealmente dependientes.

El caso $\bar{x}^t \bar{y} < 0$ es tratado similarmente.

Finalmente, supóngase que \bar{x} é \bar{y} son linealmente dependientes. Como $\bar{y} \neq 0$, $\bar{x} = a\bar{y}$ para algún escalar a .

Entonces $\|\bar{x}\|_2 = |a| \|\bar{y}\|_2$, y

$$\begin{aligned} |\bar{x}^t \bar{y}| &= |(a\bar{y})^t \bar{y}| \\ &= |a \bar{y}^t \bar{y}| \\ &= |a| \bar{y}^t \bar{y} \\ &= |a| \|\bar{y}\|_2 \|\bar{y}\|_2 \\ &= \|\bar{x}\|_2 \|\bar{y}\|_2 \end{aligned}$$



La prueba anterior hace fuerte el uso de la identidad $\|\bar{x}\|_2^2 = \bar{x}^t \bar{x}$. Esta caracterización es también usada para establecer que $\|\cdot\|_2$ es una norma.

TEOREMA 1.3

La función $\|\cdot\|_2$ satisface la desigualdad triangular.

PRUEBA

$$\begin{aligned} \|\bar{x} + \bar{y}\|_2^2 &= (\bar{x} + \bar{y})^t (\bar{x} + \bar{y}) \\ &= \bar{x}^t \bar{x} + 2\bar{x}^t \bar{y} + \bar{y}^t \bar{y} \\ &\leq \|\bar{x}\|_2^2 + 2\|\bar{x}\|_2 \|\bar{y}\|_2 + \|\bar{y}\|_2^2 \text{ por desig. de Cauchy} \\ &= (\|\bar{x}\|_2 + \|\bar{y}\|_2)^2 \end{aligned}$$

▲

Hemos usado el mismo símbolo $\|\cdot\|_2$ para denotar la norma 2 en $\mathbb{R}^1, \mathbb{R}^2, \mathbb{R}^3, \dots$. Mientras no sea un abuso de notación, puede evitarse introduciendo alguna terminología que nos será útil más tarde en relación con la norma de matrices.

DEFINICION 1.4

Sea $\psi: \bigcup_{n=1}^{\infty} \mathbb{R}^n \longrightarrow \mathbb{R}$. Entonces ψ es una familia de normas de vectores si para cada $n = 1, 2, \dots$, la restricción de ψ a \mathbb{R}^n es una norma.

Así cada una de las funciones $\|\cdot\|_1, \|\cdot\|_2, \|\cdot\|_{\infty}$ es una familia de normas de vectores. Incidentalmente, estas normas son casos especiales de normas de Holder ó p -normas de-

finidas por

$$\|\bar{x}\|_p = \sqrt[p]{\sum_{i=1}^n |x_i|^p}, \quad 1 \leq p < \infty$$

Podemos demostrar que $\|\bar{x}\|_\infty$ es $\lim_{p \rightarrow \infty} \|\bar{x}\|_p$,

observemos que

$$|x_1|^p + |x_2|^p + \dots + |x_n|^p \leq n \max |x_i|^p \text{ para } i = 1, 2, \dots, n$$

luego

$$\|\bar{x}\|_p \leq \sqrt[p]{n \max |x_i|^p} = \sqrt[p]{n} \max |x_i| \text{ para } i = 1, 2, \dots, n$$

bastaría demostrar que $\lim_{p \rightarrow \infty} \sqrt[p]{n} = 1$

$$\text{Sea } y = \sqrt[p]{n} \implies \text{Ln } y = \text{Ln } \sqrt[p]{n}$$

$$\implies \text{Ln } (y) = \frac{1}{p} \text{Ln } (n)$$

$$\implies \lim_{p \rightarrow \infty} \text{Ln } (y) = \lim_{p \rightarrow \infty} \frac{\text{Ln } (n)}{p}$$

$$\implies \lim_{p \rightarrow \infty} \text{Ln } (y) = 0$$

$$\text{Luego } \lim_{p \rightarrow \infty} y = 1$$

$$\therefore \lim_{p \rightarrow \infty} \|\bar{x}\|_p = \max \{x_i : i = 1, 2, \dots, n\} = \|\bar{x}\|_\infty$$

Una técnica útil para construir nuevas normas de las ya conocidas está contenida en el siguiente teorema.

TEOREMA 1.5

Sea ψ una norma en \mathbb{R}^m y $A \in \mathbb{R}^{m \times n}$ con columnas linealmente independientes. Entonces la función $\mu: \mathbb{R}^m \longrightarrow \mathbb{R}$ defini-

da por

$$\mu(\bar{x}) = \psi(A\bar{x})$$

es una norma en \mathbb{R}^n .

PRUEBA

Como A tiene columnas linealmente independientes y ψ es una norma,

$$\begin{aligned} \text{i)} \quad \bar{x} \neq 0 &\implies A\bar{x} \neq 0 \\ &\implies \psi(A\bar{x}) > 0 \\ &\implies \mu(\bar{x}) > 0 \end{aligned}$$

Por lo tanto μ es definida positiva.

$$\begin{aligned} \text{ii)} \quad \mu(a\bar{x}) &= \psi(aA\bar{x}) \\ &= |a| \psi(A\bar{x}) \\ &= |a| \mu(\bar{x}) \end{aligned}$$

Luego μ es homogénea. Finalmente

$$\begin{aligned} \text{iii)} \quad \mu(\bar{x} + \bar{y}) &= \psi[A(\bar{x} + \bar{y})] \\ &= \psi(A\bar{x} + A\bar{y}) \\ &\leq \psi(A\bar{x}) + \psi(A\bar{y}) \\ &= \mu(\bar{x}) + \mu(\bar{y}) \end{aligned}$$

así que μ satisface la desigualdad triangular.

▲

El teorema 1.5 puede ser aplicado para obtener una clase importante de norma.

TEOREMA 1.6:

Sea $A \in \mathbb{R}^{n \times n}$ una matriz definida positiva. Entonces la función $\psi: \mathbb{R}^n \rightarrow \mathbb{R}$ definida por

$$\psi(\bar{x}) = \sqrt{\bar{x}^t A \bar{x}}$$

es una norma en \mathbb{R}^n .

PRUEBA

Por el teorema 3.1.12, A puede ser escrita en la forma $A = LL^t$, donde L es una matriz triangular inferior no singular. Entonces

$$\begin{aligned} \psi(\bar{x}) &= \sqrt{\bar{x}^t A \bar{x}} \\ &= \sqrt{\bar{x}^t (LL^t) \bar{x}} \\ &= \sqrt{(\bar{x}^t L) (L^t \bar{x})} \\ &= \sqrt{(L^t \bar{x})^t (L^t \bar{x})} \\ &= \|L^t \bar{x}\|_2 \end{aligned}$$

Por lo tanto, por el teorema 1.5, ψ es una norma.

▲

Note que cuando $A = I$, la función ψ del teorema 1.6 se reduce a la norma 2.

Ilustraremos el teorema 1.6 con el siguiente ejemplo.

EJEMPLO 1.7

Encontrar la norma del vector $[2, 1, 3]^t$ mediante la norma definida en el teorema 1.6.

Sea la matriz definida positiva

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

$$\begin{aligned} \psi(\bar{x}) &= \left[\begin{matrix} [2, 1, 3] \\ \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \end{matrix} \right]^{1/2} \\ &= \left[\begin{matrix} [3, -3, 5] \\ \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix} \end{matrix} \right]^{1/2} = \sqrt{18} = 3\sqrt{3}. \end{aligned}$$

Como la norma de un vector da una medida para la "distancia" entre el vector y el origen, la distancia entre dos vectores puede ser definida como la norma de la diferencia de los vectores.

DEFINICION 1.8

Si $\bar{x} = [x_1, x_2, \dots, x_n]^t$ y $\bar{y} = [y_1, y_2, \dots, y_n]^t$ son vectores en \mathbb{R}^n , las distancias ℓ_1 y ℓ_∞ entre \bar{x} e \bar{y} están definidas por

$$\|\bar{x} - \bar{y}\|_2 = \left(\sum_{i=1}^n |x_i - y_i|^2 \right)^{1/2} \quad (1.9)$$

$$\|\bar{x} - \bar{y}\|_\infty = \max_{1 \leq i \leq n} |x_i - y_i| \quad (1.10)$$

EJEMPLO 1.9

El sistema lineal

$$3.3330 x_1 + 15920 x_2 - 10.333 x_3 = 15913$$

$$2.2220 x_1 + 16.710 x_2 + 9.612 x_3 = 28.544$$

$$1.5611 x_1 + 5.1791 x_2 + 1.6852 x_3 = 8.4254$$

tiene por solución $[x_1, x_2, x_3]^t = [1.0000, 1.0000, 1.0000]^t$. Si la eliminación Gaussiana es efectuada en aritmética de 5-dígitos, usando pivotamiento completo (algoritmo 2.2.11, pag 102), la solución obtenida es

$$\bar{x}' = [x_1', x_2', x_3']^t = [1.2001, 0.99991, 0.92538]^t$$

El tamaño de $\bar{x} - \bar{x}'$ está dado por

$$\begin{aligned} \|\bar{x} - \bar{x}'\|_\infty &= \text{máx} \{ |1.0000 - 1.2001|, |1.0000 - 0.99991|, |1.0000 - 0.92538| \} \\ &= \text{máx} \{ 0.2001, 0.00009, 0.07462 \} \\ &= 0.2001 \end{aligned}$$

Aunque las componentes x_2' y x_3' son muy buenas aproximaciones de x_2 y x_3 , note que la componente x_1' es una mala aproximación a x_1 .

▲

EJERCICIOS

1.1 Mostrar que la función ψ definida por ecuación (1.3) es una norma.

1.2 Probar la identidad de polarización

$$\bar{x}^t \bar{y} = \frac{\|\bar{x} + \bar{y}\|_2^2 - \|\bar{x} - \bar{y}\|_2^2}{4}$$

1.3 Encontrar $\|\bar{x}\|_\infty$, $\|\bar{x}\|_1$ y $\|\bar{x}\|_2$ para los siguientes vectores

a) $\bar{x} = [3, -4, 0, 3/4]^t$

b) $\bar{x} = \left[\frac{4}{k+1}, \frac{2}{k^2}, k^2 e^{-k} \right]^t$, para un k entero positivo fijo.

1.4 Encontrar vectores linealmente independientes \bar{x} e \bar{y} tal que

$$\|\bar{x} + \bar{y}\|_{\infty} = \|\bar{x}\|_{\infty} + \|\bar{y}\|_{\infty}$$

1.5 Mostrar que $\|\bar{x} + \bar{y}\|_2 = \|\bar{x}\|_2 + \|\bar{y}\|_2$ sí y solo sí \bar{x} e \bar{y} son linealmente dependientes y $\bar{x}^t \bar{y} \geq 0$.

1.6 Sean a_1, a_2, \dots, a_n positivos. Muestre que la función

$\psi: \mathbb{R}^n \longrightarrow \mathbb{R}$ definida por

$$\psi(\bar{x}) = \left(\sum_{i=1}^n a_i x_i^2 \right)^{1/2}$$

sea una norma.

2. NORMAS DE MATRICES

En esta sección consideraremos el problema de extender la idea de norma de un vector a matrices. Hemos visto en el ejemplo 1.1.7 que el conjunto de matrices en $\mathbb{R}^{m \times n}$ es un espacio vectorial, lo que es esencialmente idéntico con \mathbb{R}^{mn} . Consecuentemente cualquier norma de vectores (vectorial) en \mathbb{R}^{mn} induce una función definida positiva y homogénea en $\mathbb{R}^{m \times n}$ que satisface la desigualdad triangular, y es natural llamar a tal función una norma matricial. Por ejemplo, la norma -2 en \mathbb{R}^{mn} induce la norma Frobenius $\|\cdot\|_F$ en $\mathbb{R}^{m \times n}$ definida por

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2} \quad (2.1)$$

La misma prueba que $\|\cdot\|_2$ es una norma vectorial, muestra que $\|\cdot\|_F: \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}$ es definida positiva, función homogénea

y que satisface la desigualdad triangular

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

Adoptaremos esta definición general de norma matricial. Sin embargo, normas matriciales definidas de esta manera no son muy usuales. La razón es que esta definición de norma matricial no da una relación entre las normas de dos matrices y la norma de su producto. Consecuentemente restringiremos nuestra atención a normas de matrices satisfaciendo una condición de consistencia, la cual para la norma Frobenius indica

$$\|AB\|_F \leq \|A\|_F \|B\|_F. \quad (2.2)$$

Comencemos nuestro desarrollo con la definición general de norma matricial.

DEFINICION 2.1:

Una función $\psi: \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}$ es una norma matricial en $\mathbb{R}^{m \times n}$ si

- 1) $A \neq 0 \Rightarrow \psi(A) > 0, A \in \mathbb{R}^{m \times n}$.
- 2) $\psi(aA) = |a| \psi(A), A \in \mathbb{R}^{m \times n}, a \in \mathbb{R}$.
- 3) $\psi(A+B) \leq \psi(A) + \psi(B), A, B \in \mathbb{R}^{m \times n}$.

Esta definición tiene algunas consecuencias inmediatas. Como una norma matricial en $\mathbb{R}^{m \times n}$ es esencialmente una norma vectorial en \mathbb{R}^{mn} , se sigue que una norma matricial goza de todas las propiedades de una norma vectorial.

DEFINICION 2.2:

La función $\psi: \bigcup_{m, n=1}^{\infty} \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}$ es una familia

de normas matriciales si para cada $m, n \geq 1$ la restricción de ψ a $\mathbb{R}^{m \times n}$ es una norma matricial.

Se sigue de esta definición y de los comentarios anteriores que si ψ es una familia de normas matriciales, entonces la restricción de ψ a $\cup_{n=1}^{\infty} \mathbb{R}^{n \times 1}$ es una familia de normas vectoriales.

EJEMPLO 2.3:

Las normas de Frobenius definida para cualquier $A \in \mathbb{R}^{m \times n}$ por (2.1) forman una familia de normas matriciales. La restricción de $\|\cdot\|_F$ a $\cup_{n=1}^{\infty} \mathbb{R}^{n \times 1}$ es la familia de normas vectoriales; esto es $\|\bar{x}\|_F = \|\bar{x}\|_2$ para todo vector \bar{x} .

La norma Frobenius es una generalización natural de la norma vectorial Euclídeana. Puesto que es relativamente fácil calcular y satisface la desigualdad (2.2), es usada frecuentemente en cálculos matriciales. Sin embargo, todas las normas matriciales que satisfacen la definición 2.1 no necesariamente satisfacen una desigualdad tal como (2.2). Por ejemplo, una generalización natural de la norma $-\infty$ es la función ψ definida por

$$\psi(A) = \text{máx} \{ |a_{ij}| : i = 1, 2, \dots, m ; j = 1, 2, \dots, n \}, \quad (2.3)$$

con $A \in \mathbb{R}^{m \times n}$

Es fácil verificar que ψ es una norma matricial según la definición 2.1. Sin embargo, si

$$A = B = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix},$$

entonces $\psi(A)\psi(B) = 1$ mientras que $\psi(AB) = 2$, así que no es cier-

to que $\psi(AB) \leq \psi(A)\psi(B)$. Por esta razón la norma matricial (2.3) es rara vez utilizada.

Estas consideraciones sugieren que consideremos las propiedades de normas matriciales que satisfacen una relación como (2.2).

Como A, B y AB en general tendrán diferentes dimensiones, usaremos una norma diferente para cada miembro de (2.2).

DEFINICION 2.4

Sean $\mu: \mathbb{R}^{\ell \times m} \rightarrow \mathbb{R}$; $\psi: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ y $\vartheta: \mathbb{R}^{\ell \times n} \rightarrow \mathbb{R}$

normas matriciales. Entonces μ , ψ y ϑ son consistentes si

$$\vartheta(AB) \leq \mu(A) \psi(B)$$

para todo $A \in \mathbb{R}^{\ell \times m}$ y $B \in \mathbb{R}^{m \times n}$. Una norma matricial en $\mathbb{R}^{n \times n}$ es consistente si es consistente consigo misma. Una familia de normas matriciales ψ es consistente si

$$\psi(AB) \leq \psi(A) \psi(B)$$

siempre que el producto AB esté definido.

Como un ejemplo de familia de normas matriciales consistentes, mostraremos que la norma de Frobenius del ejemplo 2.3 es consistente.

TEOREMA 2.5:

La familia $\|\cdot\|_F$ es consistente.

PRUEBA

Mostremos primero que

$$\|A\bar{x}\|_2 \leq \|A\|_F \|\bar{x}\|_2 \quad (2.4)$$

para cualquier $A \in \mathbb{R}^{\ell \times m}$ y $\bar{x} \in \mathbb{R}^m$. Sea A particionada en filas:

$A^t = [A_1, A_2, \dots, A_\ell]$. Entonces

$$A\bar{x} = \begin{bmatrix} A_1^t \bar{x} \\ A_2^t \bar{x} \\ \cdot \\ \cdot \\ A_\ell^t \bar{x} \end{bmatrix}.$$

Por tanto

$$\|A\bar{x}\|_2^2 = \sum_{i=1}^{\ell} |A_i^t \bar{x}|^2.$$

Por la desigualdad de Cauchy $|A_i^t \bar{x}| \leq \|A_i\|_2 \|\bar{x}\|_2$. Por lo --
tanto

$$\|A\bar{x}\|_2^2 \leq \|\bar{x}\|_2^2 \sum_{i=1}^{\ell} \|A_i\|_2^2.$$

Observemos que $\|A\|_F^2 = \sum_{i=1}^{\ell} \sum_{j=1}^m a_{ij}^2$

$$= \sum_{i=1}^{\ell} (a_{i1}^2 + a_{i2}^2 + \dots + a_{im}^2)$$

$$= \sum_{i=1}^{\ell} \|A_i\|_2^2$$

$$= \sum_{i=1}^{\ell} \|A_i\|_2^2.$$

Por lo tanto (2.4) es satisfecha.

Ahora, sea $C = AB$, donde $A \in \mathbb{R}^{\ell \times m}$ y $B \in \mathbb{R}^{m \times n}$. Si

$B = [B_1, B_2, \dots, B_n]$ es particionada en columnas, entonces

$$\|C\|_F^2 = \|AB\|_F^2 = \|[AB_1, AB_2, \dots, AB_n]\|_F^2$$

$$\sum_{j=1}^n \|AB_j\|_2^2 \leq \|A\|^2 \sum_{j=1}^n \|B_j\|_2^2 = \|A\|_F^2 \|B\|_F^2$$

la cual es la relación de consistencia,

▲

En particular, la definición 2.4 incluye la noción de consistencia de una norma vectorial y una norma matricial.

Por ejemplo, si $\|\cdot\|: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ es una norma matricial y $\psi: \mathbb{R}^n \rightarrow \mathbb{R}$ es una norma vectorial, entonces ψ y $\|\cdot\|$ son consistentes si

$$\psi(A\bar{x}) \leq \|A\| \psi(\bar{x}).$$

La desigualdad (2.4) en la prueba del teorema 2.5, muestra que la norma matricial Frobenius es consistente con la norma vectorial Euclídeana.

Algunas veces sucede que dada una norma matricial consistente en $\mathbb{R}^{n \times n}$ se requiere de una norma vectorial consistente. La prueba del siguiente teorema muestra cómo construir tal norma.

TEOREMA 2.6:

Sea $\|\cdot\|: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ una norma matricial consistente. Entonces existe una norma en \mathbb{R}^n que es consistente con $\|\cdot\|$.

PRUEBA

Sea $\bar{y} \neq 0$ un n -vector. Definimos la función $\psi: \mathbb{R}^n \rightarrow \mathbb{R}$ por

$$\psi(\bar{x}) = \|\bar{x}\bar{y}^t\|, \quad \bar{x} \in \mathbb{R}^n \quad (2.5)$$

Es fácil verificar que ψ es una norma vectorial. Además,

$$\psi(A\bar{x}) = \|(A\bar{x})\bar{y}^t\| = \|A(\bar{x}\bar{y}^t)\| \leq \|A\| \|\bar{x}\bar{y}^t\| = \|A\| \psi(\bar{x}),$$

así que ψ es consistente con $\|\cdot\|$.

▲

Regresemos ahora a un proceso importante para la construcción de normas matriciales consistentes. La idea para la construcción es la siguiente:

Sea $A \in \mathbb{R}^{m \times n}$ y sean las normas $\psi: \mathbb{R}^n \rightarrow \mathbb{R}$ y $\mu: \mathbb{R}^m \rightarrow \mathbb{R}$. Si $\psi(\bar{x}) = 1$ é $\bar{y} = a\bar{x}$, entonces $\psi(\bar{y}) = |a|$ y

$$\mu(A\bar{y}) = \mu(aA\bar{x}) = |a|\mu(A\bar{x}) = \psi(\bar{x})\mu(A\bar{x}).$$

En otras palabras, el número $\mu(A\bar{x})$ mide, en cuánto la transformación lineal aumenta (o disminuye), a cualquier vector que es un múltiplo de \bar{x} . Si existe tal número mayor $\mu(A\bar{x})$, esta constante es un candidato natural para una norma de A .

TEOREMA 2.7

Si $\|\cdot\|$ es cualquier norma vectorial en $\mathbb{R}^{m \times n}$ entonces $\|A\| = \max_{\|\bar{x}\|=1} \|A\bar{x}\|$, con $\bar{x} \in \mathbb{R}^n$,

define una norma matricial en el conjunto de matrices reales de orden $m \times n$, la cual es llamada *norma natural*.

PRUEBA

Probaremos que satisface las tres condiciones de una norma ma

$$\begin{aligned}
 \text{i)} \quad \|A\| &= \max_{\|\bar{x}\|=1} \|A\bar{x}\| \\
 &= \max_{\|\bar{x}\|=1} \left\| \begin{pmatrix} A_1 \bar{x} \\ A_2 \bar{x} \\ \vdots \\ A_m \bar{x} \end{pmatrix} \right\|, \text{ con } A = [A_1, A_2, \dots, A_m] \text{ particionada en} \\
 &\hspace{20em} \text{filas}
 \end{aligned}$$

$$= \|A_k \bar{x}\| \leq \|A_k\| \|\bar{x}\| = \|A\| > 0$$

$$\text{ii)} \quad \|aA\| \leq \| \| = \max_{\|\bar{x}\|=1} \|(aA)\bar{x}\|, \quad a \in \mathbb{R}$$

$$= \max_{\|\bar{x}\|=1} \|a(A\bar{x})\|$$

$$= \max_{\|\bar{x}\|=1} |a| \|A\bar{x}\|$$

$$= |a| \max_{\|\bar{x}\|=1} \|A\bar{x}\|$$

$$= |a| \|A\|$$

$$\text{iii)} \quad \|A + B\| = \max_{\|\bar{x}\|=1} \|(A + B)\bar{x}\|, \quad A, B \in \mathbb{R}^{m \times n}$$

$$= \max_{\|\bar{x}\|=1} \|A\bar{x} + B\bar{x}\|$$

$$\leq \max_{\|\bar{x}\|=1} \left(\|A\bar{x}\| + \|B\bar{x}\| \right)$$

$$\leq \max_{\|\bar{x}\|=1} \|A\bar{x}\| + \max_{\|\bar{x}\|=1} \|B\bar{x}\|$$

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

A

TEOREMA 2.9

Sea $A \in \mathbb{R}^{m \times n}$. Entonces

$$\|A\|_1 = \max \left\{ \sum_{i=1}^m |a_{ij}| : j = 1, 2, 3, \dots, n \right\} \quad (2.9)$$

y

$$\|A\|_\infty = \max \left\{ \sum_{j=1}^n |a_{ij}| : i = 1, 2, 3, \dots, m \right\} \quad (2.10)$$

PRUEBA

La estrategia de la prueba es la misma en ambos casos.

Llamemos λ a la parte derecha de (2.9) y (2.10). Mostraremos - primero que para cualquier vector \bar{x} tenemos que

$$\|A\bar{x}\|_p \leq \lambda \|\bar{x}\|_p \quad (p = 1, \infty). \text{ Esto implica que}$$

$$\|A\|_p \leq \lambda. \quad (2.11)$$

después encontramos un vector particular \bar{x} con $\|\bar{x}\|_p = 1$ tal - que

$$\|A\bar{x}\|_p = \lambda$$

$$\lambda = \|A\bar{x}\|_p \leq \|A\|_p \|\bar{x}\|_p \leq \|A\|_p$$

esto muestra que

$$\|A\|_p \geq \lambda, \quad (2.12)$$

de (2.11) y (2.12) se deduce que $\|A\|_p = \lambda$

Específicamente, para la norma-1 sea $A = [A_1, A_2, \dots, A_n]$ particionada en columnas. Entonces $\lambda = \text{máx} \{ \|A_j\|_1 : j=1, 2, \dots, n \}$

y

$$\begin{aligned} \|A\bar{x}\|_1 &= \|x_1 A_1 + x_2 A_2 + \dots + x_n A_n\|_1 \\ &\leq \|x_1 A_1\|_1 + \|x_2 A_2\|_1 + \dots + \|x_n A_n\|_1 \\ &\leq |x_1| \|A_1\|_1 + |x_2| \|A_2\|_1 + \dots + |x_n| \|A_n\|_1 \\ &\leq (|x_1| + |x_2| + \dots + |x_n|) \text{máx} \{ \|A_j\|_1 \} = \lambda \| \bar{x} \|_1 \end{aligned}$$

Por otro lado si $\lambda = \|A_k\|_1$, entonces $\|\bar{e}_k\|_1 = 1$ y

$$\|A\bar{e}_k\|_1 = \|A_k\|_1 = \lambda$$

Esto establece (2.9)

Para la norma $-\infty$, $\lambda = \text{máx} \{ \sum_{j=1}^n |a_{ij}| : i = 1, 2, \dots, n \}$. Para

el cual \bar{x} tenemos:

$$\begin{aligned} \|A\bar{x}\| &= \text{máx}_i \{ \sum_j |a_{ij} x_j| \} \\ &\leq \text{máx}_i \{ \sum_j |a_{ij}| |x_j| \} \\ &\leq \text{máx}_i \{ \sum_j |a_{ij}| \} \text{máx}_j \{ |x_j| \} = \lambda \| \bar{x} \|_\infty \end{aligned}$$

Por otro lado si $\lambda = \sum_{j=1}^n |a_{kj}|$ y $\bar{x} = [\text{sign}(a_{k1}), \text{sign}(a_{k2}), \dots$

$\dots, \text{sign}(a_{kn})]^t$, entonces $\|\bar{x}\|_\infty = 1$ y $\|A\bar{x}\|_\infty = \lambda$.

Así la norma-1 de una matriz es igual al máximo de la norma-1 de sus columnas. Por esta razón la norma matricial-1 es -

frecuentemente llamada *suma de normas de columna*. Similarmente la norma- ∞ es llamada *suma de normas de fila*. Como estas normas son fáciles de calcular, junto con la norma Frobenius, son a menudo usadas en algoritmos matriciales.

EJEMPLO 2.10

$$\text{Si } A = \begin{bmatrix} 1 & 2 & -1 \\ 0 & 3 & -1 \\ 5 & -1 & 1 \end{bmatrix}$$

Entonces

$$\sum_{j=1}^3 |a_{1j}| = |1| + |2| + |-1| = 4$$

$$\sum_{j=1}^3 |a_{2j}| = |0| + |3| + |-1| = 4$$

$$\sum_{j=1}^3 |a_{3j}| = |5| + |-1| + |1| = 7$$

$$\text{Así } \|A\|_{\infty} = \max \{4, 4, 7\} = 7$$

Computacionalmente no es conveniente la norma matricial-2. Sin embargo, esta norma tiene algunas propiedades muy buenas que la hacen útil para propósitos teóricos. Algunas de estas propiedades están contenidas en el siguiente teorema:

TEOREMA 2.11

Sea $A \in \mathbb{R}^{m \times n}$. Entonces

$$1. \quad \|A\|_2 = \max_{\|\bar{x}\|_2 = \|\bar{y}\|_2 = 1} |\bar{y}^t A \bar{x}|,$$

$$2. \|A^t\|_2 = \|A\|_2,$$

$$3. \|A^t A\|_2 = \|A\|_2^2.$$

PRUEBA

Para la parte 1, sea $\|\bar{x}\|_2 = \|\bar{y}\|_2 = 1$. Entonces por la desigualdad de Cauchy,

$$|\bar{y}^t A \bar{x}| \leq \|\bar{y}\|_2 \|A \bar{x}\|_2 \leq \|\bar{y}\|_2 \|\bar{x}\|_2 \|A\|_2 = \|A\|_2$$

Por otro lado, sea $\|\bar{x}\|_2 = 1$ y $\|A \bar{x}\|_2 = \|A\|_2$. Sea $\bar{y} = \frac{A \bar{x}}{\|A \bar{x}\|_2}$.

Entonces $\|\bar{y}\|_2 = 1$ y

$$|\bar{y}^t A \bar{x}| = \frac{\bar{x}^t A^t A \bar{x}}{\|A \bar{x}\|_2} = \frac{(A \bar{x})^t A \bar{x}}{\|A \bar{x}\|_2} = \frac{\|A \bar{x}\|_2^2}{\|A \bar{x}\|_2} = \|A \bar{x}\|_2 = \|A\|_2.$$

así que, para esta selección de \bar{x} é \bar{y} , igualmente es logrado.

Para la parte 2., note que

$$\begin{aligned} \|A^t\|_2 &= \max_{\|\bar{x}\|_2 = \|\bar{y}\|_2 = 1} |\bar{y}^t A^t \bar{x}| \\ &= \max_{\|\bar{x}\|_2 = \|\bar{y}\|_2 = 1} |(A \bar{y})^t \bar{x}| = \max_{\|\bar{x}\|_2 = \|\bar{y}\|_2 = 1} |\bar{x}^t A \bar{y}| = \|A\|_2 \end{aligned}$$

Finalmente, para la parte 3 tenemos:

$$\|A^t A\|_2 \leq \|A^t\|_2 \|A\|_2 \leq \|A\|_2^2$$

Por otro lado, sea $\|\bar{x}\|_2 = 1$ y $\|A \bar{x}\|_2 = \|A\|_2$. Entonces por la parte 1 con $\bar{y} = \bar{x}$

$$\|A^t A\|_2 \geq |\bar{x}^t A^t A \bar{x}| = \|A \bar{x}\|_2^2 = \|A\|_2^2.$$

A

EJERCICIOS

2.1 Verificar que la función ψ definida por (2.5) es una norma vectorial.

2.2 Probar que si $A = [A_1, A_2, \dots, A_n]$ está particionada en columnas, entonces

$$\|A\|_F^2 = \|A_1\|_2^2 + \|A_2\|_2^2 + \dots + \|A_n\|_2^2$$

2.3 Mostrar que para $\bar{x} \in \mathbb{R}^n$, $\|\bar{x}\|_2 = \|\bar{x}^t\|_2$.

2.4 Sea $\psi: \mathbb{R}^{n \times n} \rightarrow \mathbb{R}$ definida por

$$\psi(A) = n \{ \text{máx} |a_{ij}| : i, j = 1, 2, \dots, n \}$$

muestre que ψ es una norma matricial consistente.

2.5 Muestre que $\|\cdot\|_{\odot}$ y $\|\cdot\|_{\oplus}$ son normas matriciales, donde

$$\|A\|_{\odot} = \left(\sum_{i=1}^n \sum_{j=1}^n |a_{ij}|^2 \right)^{1/2} \quad \text{y} \quad \|A\|_{\oplus} = \sum_{i=1}^n \sum_{j=1}^n |a_{ij}|$$

para cualquier matriz A de orden $n \times n$.

2.6 Encontrar $\|\cdot\|_{\infty}$ y $\|\cdot\|_1$, para las siguientes matrices:

$$\text{a) } A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \quad \text{b) } B = \begin{bmatrix} -3 & 2 \\ 3/2 & 7/2 \end{bmatrix} \quad \text{c) } C = \begin{bmatrix} 10 & 15 \\ 0 & 1 \end{bmatrix}$$

3. ESTIMACION DE ERROR, CONDICIONAMIENTO DE UNA MATRIZ Y REFINAMIENTO ITERATIVO.

En esta sección discutiremos los efectos de los errores por redondeo de los algoritmos, así como también la exactitud de sistemas lineales y por último el mejoramiento de soluciones. Con relación a los errores por redondeo mostraremos que -

si los algoritmos de los capítulos II y III son usados para --
calcular una solución de la ecuación

$$A\bar{x} = \bar{b} \quad (3.1)$$

la solución calculada \tilde{x} satisface

$$(A + H) \tilde{x} = \bar{b}, \quad (3.2)$$

y daremos cotas rigurosas para los elementos de H . Si puede --
ser mostrado que H es pequeño, entonces el algoritmo es esta--
ble en el sentido de la sección 8 del capítulo I.

Tal análisis de error tiene importantes limitaciones. En
primer lugar, las cotas del error estan a menudo más lejos que
los errores observados. En segundo lugar un resultado estable
tal como (3.2) no puede asegurar la exactitud de la solución,
a menos que el problema esté bien condicionado, pero si el siste
ma (3.1) está mal condicionado aún para un μ pequeña corres-
ponderá una gran desviación para \tilde{x} .

Una cuestión importante es que un análisis de error puede
sugerir como arreglar los detalles de un algoritmo para una mayo
r estabilidad. Por ejemplo, usando el hecho que la cota en μ
de (3.2) contiene factores que dependen de la estrategia de pivo
tamiento.

Un riguroso análisis del error por redondeo procede de --
aplicaciones repetidas de las cotas del error por redondeo del
capítulo I, sección 8. Expondremos solamente los resultados de
los análisis del error por redondeo. Asumiremos que todos los
cálculos han sido efectuados en aritmética de punto flotante -

de t -dígitos, satisfaciendo las cotas del capítulo I, sección 8. Además, asumiremos que la *baja-carga* y *sobre-carga* no ha ocurrido en el cálculo. Finalmente asumiremos que el orden del problema considerado, es decir n , está también restringido por $n\mu < 0.1$

Regresemos primero a la solución de un sistema triangular por medio del algoritmo (2.1.3)

TEOREMA 3.1

Sea $T \in \mathbb{R}^{n \times n}$ y $\bar{b} \in \mathbb{R}^n$. Si \tilde{x} denota la solución calculada de la ecuación $T\bar{x} = \bar{b}$, obtenida por el algoritmo 1.3 Entonces \tilde{x} satisface la ecuación

$$(T + E)\tilde{x} = \bar{b}$$

donde $|e_{ij}| \leq (n+1) \Pi |t_{ij}| 10^{-t}$ ($i, j = 1, 2, \dots, n$) (3.3)

Aquí Π es una constante de orden unitaria que dependen de los detalles de la aritmética.

Este teorema es en muchos aspectos bastante satisfactorio. Dice que la solución calculada es una solución exacta de un problema en el cual T ha sido perturbada ligeramente. En efecto, los elementos de $T+E$ difieren de los elementos correspondientes de T en los pequeños errores relativos. Para un n moderado, esos errores relativos son comparables con los errores hechos en el redondeo de T misma, y si los elementos de T se calculan, los errores e_{ij} pueden ser considerablemente más pequeños que la incerteza en T . Si productos internos son acumulados en doble precisión, el factor n puede ser eliminado de

(3.3), así que la matriz de error E es bastante comparable con el error hecho en el redondeo de T .

Nada menos, que los resultados anteriores ilustran las limitaciones de la primera parte de esta sección. En primer lugar, los detalles del análisis muestra que

$$|e_{ij}| \leq (j - i + 2) \prod |t_{ij}| 10^{-t},$$

lo cual puede ser considerado mas pequeño que (3.3). Mas importante, sin embargo, es el hecho que las *soluciones de sistemas triangulares son usualmente calculadas con alta precisión*. Este hecho no puede ser probado en general ya que existen problemas para los que la afirmación anterior no es verdadera.

La pregunta de que si el algoritmo 2.1.4 para invertir -- una matriz triangular superior es estable, está abierta. Como cada columna S_i de la inversa fué formada resolviendo el sistema $TS_i = \bar{e}_i$, se sigue del teorema 3.1 que S_i es la i -ésima columna de una matriz ligeramente perturbada $T + E_i$. Sin embargo, como la perturbación es diferente para cada i , no se sigue que $S = [S_1, S_2, \dots, S_n]$ sea la inversa de alguna matriz perturbada $T + E$ aunque S esté cercana a la inversa de tal matriz.

En la práctica, las inversas calculadas de matrices triangulares, que usualmente son encontradas, son bastante exactas.

Ahora consideraremos los análisis de error de la eliminación Gaussiana y de los algoritmos para la descomposición triangular. Hemos visto ya, que la estrategia para seleccionar los pivotes en estos algoritmos pueden tener un efecto marcado en sus propiedades numéricas y esperaríamos que el análisis de --

error refleje la estrategia de pivotamiento. Se deduce que la selección de pivotes afecta el resultado limitando el crecimiento de los elementos calculados en el curso de la reducción. Por el teorema 2.2.10 podemos asumir que cualquier intercambio requerido por la estrategia de pivotamiento ha sido ya efectuado. Así, sea el algoritmo de eliminación Gaussiana efectuado en aritmética de punto flotante sobre la matriz $A = A_1$ dando las matrices M_1, M_2, \dots, M_{n-1} , y A_2, A_3, \dots, A_n . Sea

$$\beta_k = \max \{ |a_{ij}^{(k)}| \} \quad (k = 1, 2, \dots, n)$$

y

$$\gamma = \frac{\max \{ \beta_k : k = 1, 2, \dots, n \}}{\beta_1}$$

así γ , la relación entre el mayor elemento de A_1, A_2, \dots, A_n con el mayor elemento de A_1 , es una medida del crecimiento de las matrices generadas por el algoritmo. Con estas definiciones tenemos el siguiente resultado.

TEOREMA 3.2

Las matrices $M_1, M_2, \dots, M_{n-1}, A_n$ calculadas por el algoritmo 2.2.4 satisfacen

$$M_1^{-1} M_2^{-1} \dots M_{n-1}^{-1} A_n = A + E \quad (3.4)$$

donde

$$|e_{ij}| \leq n \Pi \beta_1 \gamma 10^{-t} \quad (3.5)$$

para alguna constante Π de orden unitario que es independiente de A .

Por el teorema, la descomposición triangular calculada a través de la eliminación Gaussiana es la descomposición triangular exacta de una matriz perturbada. La perturbación será pequeña comparada con β_1 (esto es, comparado con el mayor elemento de A_1) con tal de que el factor de crecimiento γ sea pequeño. Así, es importante establecer cotas para el crecimiento de elementos en el curso de la reducción. Para el pivotamiento -- parcial y completo las cotas superiores pueden ser dadas por γ que no dependen de A . La cota para el pivotamiento parcial es

$$\gamma \leq \{n \cdot 2^1 \cdot 3^{1/2} \cdot 4^{1/3} \dots n^{1/n-1}\}^{1/2}$$

Esta cota se incrementa un poco lenta con respecto a n y además la prueba que se establece muestra que no puede ser alcanzada. Así, *la eliminación Gaussiana con pivotamiento completo es un algoritmo estable.*

La cota para el pivotamiento parcial es

$$\gamma \leq 2^{n-1} \quad (3.6)$$

Esta es una función de crecimiento un poco rápida y su uso en la cota (3.5) sugiere que el orden de las matrices que pueden ser descompuestas por eliminación Gaussiana con pivotamiento parcial es estrictamente limitado. En este caso los elementos de E pueden ser del mismo tamaño que los elementos de A . *No se puede afirmar que la eliminación Gaussiana con pivotamiento parcial sea incondicionalmente estable.*

En la práctica, sin embargo, la cota (3.6) es rara vez alcanzada. Usualmente los elementos de las matrices reducidas --

A_1, A_2, \dots, A_n quedan del mismo orden de magnitud ó muestran --- siempre un decrecimiento progresivo en el tamaño. En la práctica la eliminación Gaussiana con pivotamiento parcial debe ser considerado un algoritmo estable.

El análisis de error de la reducción de Crout da resultados similares. Las matrices L y U calculadas satisfacen

$$LU = A + E$$

donde los elementos de E satisfacen (3.5). El factor de crecimiento γ en las cotas para la reducción de Crout es el mismo - que el factor de crecimiento para la eliminación Gaussiana.

No se puede afirmar que la reducción de Crout sea estable en general. Sin embargo, para propósitos prácticos puede ser - considerado un algoritmo estable.

La reducción de Crout permite la acumulación de productos internos en doble precisión. Hasta donde el análisis de error está interesado, el efecto de ésto es quitar el factor n de la cota (3.5). Así, si γ es también de orden unitaria, los elementos de la matriz error E son del mismo tamaño que los errores por redondeo de A y en muchos son considerados más pequeños.

El análisis de error del algoritmo de Cholesky para matrices definidas positivas difiere de las otras en que no existe un factor de crecimiento (esto es también cierto para la eliminación Gaussiana y reducción de Crout aplicados a matrices definidas positivas). Este algoritmo calcula una matriz triangular inferior L satisfaciendo

$$LL^t = A + E$$

donde los elementos de E satisfacen (3.5) con $\gamma = 1$. Cuando los productos internos no pueden ser acumulados en doble precisión, la reducción de Crout y la eliminación Gaussiana tienen las mismas propiedades numéricas; en cambio, si los productos internos pueden ser acumulados en doble precisión, la reducción de Crout es superior.

Las cotas para la solución del sistema lineal $A\bar{x} = \bar{b}$, pueden ser obtenidas combinando los teoremas 3.1 y 3.2. Para definir las supondremos que la matriz A ha sido descompuesta por la eliminación Gaussiana con pivotamiento parcial en el producto

$$A = (M_1^{-1}M_2^{-1}\dots M_{n-1}^{-1})A_n = LU$$

Los elementos de L son los multiplicadores m_{ij} del algoritmo 2.2.4, los cuales son menores que la unidad en magnitud. Por lo tanto

$$|l_{ij}| \leq 1$$

Los elementos de U son tomados de las matrices A_1, A_2, \dots, A_n de la reducción y por lo tanto satisfacen

$$|u_{ij}| \leq \gamma \beta_1,$$

donde β_1 es la magnitud del mayor elemento de A_1 y γ es el factor de crecimiento. Finalmente del teorema 3.2 sabemos que

$$LU = A + E \tag{3.7}$$

donde los elementos de E satisfacen (3.5)

Ahora supóngase que el algoritmo 2.2.15, es usado para re

resolver el sistema $A\bar{x} = \bar{b}$. El paso 2 del algoritmo es numéricamente equivalente a resolver el sistema triangular inferior

$$L\bar{y} = \bar{b} \quad (3.8)$$

mientras el paso 3, equivale a resolver el sistema triangular superior

$$U\bar{x} = \bar{y}$$

Hemos observado que los sistemas triangulares son usualmente resueltos con alta precisión. Por lo tanto no es sorprendente que la solución calculada \tilde{x} está a menudo muy cerca de la verdadera solución \bar{x}' del sistema

$$(A + E)\bar{x}' = \bar{b} \quad (3.9)$$

En otras palabras, la solución calculada \tilde{x} de un sistema de ecuaciones lineales está a menudo cerca de una solución \bar{x}' de (3.9), donde la matriz de error E es independiente de \bar{b} .

El siguiente teorema muestra que si dijamos el requerimiento de que la matriz de error sea independiente de \bar{b} , podemos obtener un resultado rigurosamente estable.

TEOREMA 3.3

Si \tilde{x} denota la solución calculada para el sistema $A\bar{x} = \bar{b}$ por medio del algoritmo 2.2.15. Entonces \tilde{x} satisface

$$(A + H)\tilde{x} = \bar{b} \quad (3.10)$$

donde

$$|h_{ij}| < (n\pi + 2n^2\theta + n^3\theta^2 10^{-t}) \gamma \epsilon_1 10^{-t}, \quad (3.11)$$

con π y θ constantes de orden unitaria.

PRUEBA

Por el teorema 3.1, la solución calculada \tilde{y} de (3.8) satisface

$$(L + F)\tilde{y} = \bar{b}$$

donde

$$|f_{ij}| \leq n \delta |\ell_{ij}| 10^{-t} \leq n \delta 10^{-t}$$

con δ constante de orden unitaria. La última desigualdad se sigue del hecho que el pivotamiento parcial asegura que $|\ell_{ij}| < 1$. En el paso 3 del algoritmo 2.2.15, resolvemos el sistema $U\bar{x} = \tilde{y}$, dando una solución calculada \tilde{x} que satisface

$$(U + G)\tilde{x} = \tilde{y}$$

donde

$$|g_{ij}| \leq n \delta |u_{ij}| 10^{-t} \leq n \delta \beta_1 \gamma 10^{-t}$$

luego

$$(L + F)\tilde{y} = \bar{b}$$

$$\implies (L + F)(U + G)\tilde{x} = \bar{b}$$

$$\implies (LU + LG + FU + FG)\tilde{x} = \bar{b}$$

$$\implies (A + E + LG + FU + FG)\tilde{x} = \bar{b}$$

$$\therefore (A + H)\tilde{x} = \bar{b}, \text{ donde } H = E + LG + FU + FG ;$$

ahora el mayor elemento del producto de dos matrices de orden n está acotado por n veces el producto del mayor elemento de cada matriz. Por lo tanto, usando las cotas para los elementos de E, F, G, L y U , obtenemos

$$A + H \leq n \Pi \beta_1 \gamma 10^{-t} + n^2 \partial \beta_1 \gamma 10^{-t} + n^2 \partial \beta_1 \gamma 10^{-t} + n^3 \partial^2 \beta_1 \gamma 10^{-2t}$$

$$A + H \leq (n \Pi + 2n^2 \partial + n^3 \partial^2 10^{-t}) \gamma \beta_1 10^{-t}$$

▲

Como teoremas no generales en las inversas calculadas de ecuaciones triangulares han sido establecidas, no es extraño - que poco puede ser dicho acerca de las inversas calculadas de matrices generales. Supóngase por ejemplo, que las columnas de la inversa X , son calculadas, como las soluciones de los sistemas $AX_i = \bar{e}_i$ ($i = 1, 2, \dots, n$). Entonces del teorema 3.3 sabemos que cada columna calculada \tilde{x}_i satisface $(A + H_i) \tilde{x}_i = \bar{e}_i$; sin embargo las matrices de error H_i pueden ser diferentes para cada columna, y no podemos concluir que existe una matriz particular pequeña H tal que $(A + H) \tilde{x} = I$.

No menos se sigue de la observación presedida del teorema 3.3 que cada \tilde{x}_i estará a menudo cerca de la solución \bar{x}'_i de la ecuación $(A + E) \bar{x}'_i = \bar{e}_i$, donde E es la matriz error de la descomposición LU de A . Como E es independiente de \bar{e}_i , se sigue - que \tilde{X} está cercana a la inversa de $A + E$.

Consideraremos ahora el siguiente problema. Sea $A \in \mathbb{R}^{n \times n}$ - no singular y sea E una matriz de orden $n \times n$ que se presume -- sea pequeña. ¿Qué tan pequeña debe ser E para que la matriz - perturbada $A + E$ sea también singular, y cuánto difiere --- $(A + E)^{-1}$ de A^{-1} ?. En contestación a estas preguntas usaremos - la teoría de normas desarrollada en las dos secciones anteriores.

Una aplicación de nuestros resultados es el cálculo de --

las inversas de matrices. Antes hemos indicado que la inversa calculada de una matriz A estaría a menudo cerca de la inversa exacta de una matriz ligeramente perturbada $A + E$. Sin embargo, este resultado no garantiza que la inversa calculada para A^{-1} y $(A + E)^{-1}$ no pueden diferir grandemente. En la terminología de la sección 8 del capítulo I, tal matriz sería llamada *mal condicionada con respecto a la inversión*.

Nuestro análisis no solamente da condiciones bajo las cuales A está mal condicionada, sino que también asocia a A un número de condicionamiento que mide el grado de su mal condicionamiento.

Además para hablar rigurosamente acerca del tamaño de errores que involucran vectores y matrices introduciremos las siguientes generalizaciones de error absoluto y relativo.

DEFINICION 3.4

Sean $A, B \in \mathbb{R}^{m \times n}$ con B considerada como una aproximación a A . El *residual* de B es la matriz

$$A - B$$

Si $\psi: \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ es una norma, el error en B con respecto a ψ es el número

$$\psi(A - B)$$

Si $A \neq 0$, el error relativo en B con respecto a ψ es el número

$$\frac{\psi(A - B)}{\psi(A)}$$

Donde no haya confusión, quitaremos la frase "con respecto a ψ " y nos referiremos simplemente a error ó error relativo. Las nociones de residual, error y error relativo están definidos para n -vectores considerados como matrices de orden $n \times 1$.

EJEMPLO 3.5

En un computador dado supóngase que el valor redondeado de un número b es $\ell(b) = b(1 + \delta)$, donde $|\delta| \leq 10^{-t}$. Sea $A \in \mathbb{R}^{m \times n}$ y sea a la magnitud del mayor elemento de A . Sea B obtenida de A redondeando sus elementos. Entonces $B = A + E$, donde

$$\sum_{j=1}^n |e_{ij}| \leq |\delta_{ij}| |a_{ij}| \leq 10^{-t} \|A\|_{\infty}$$

Por tanto

$$\|E\|_{\infty} \leq \|A\|_{\infty} 10^{-t},$$

y

$$\frac{\|A - B\|}{\|A\|_{\infty}} \leq 10^{-t}$$

Entonces, el error relativo con respecto a la norma- ∞ en una matriz redondeada con t cifras no es mayor que 10^{-t} .

Cuando no existe crecimiento en el proceso de eliminación, el teorema 3.3 puede ser interpretado diciendo que la solución calculada \tilde{x} de (3.1) satisface

$$(A + H) \tilde{x} = \bar{b}, \quad (3.12)$$

donde

$$\|H\|_{\infty} \leq \phi(n) \|A\|_{\infty} 10^{-t} \quad (3.13)$$

donde $\phi(n)$ es una función de n cuya forma depende de los detalles aritméticos en el cálculo, pero que en cualquier evento no es muy grande.

Una manera de estimación de la exactitud de una solución aproximada para un sistema lineal es determinar entonces la -- cercana aproximación que viene satisfaciendo el sistema.

DEFINICION 3.6

Si $\tilde{x} \in \mathbb{R}^n$ es una aproximación a la solución del sistema lineal (3.1), el vector residual para \tilde{x} con respecto a este sistema está definido por

$$\bar{r} = \bar{b} - A\tilde{x}.$$

Sería intuitivamente esperado que cuando $\|\bar{r}\|$ es pequeño para algún vector, entonces $\|x - \tilde{x}\|$ también sea pequeño. Aunque este sea a menudo el caso, ciertos sistemas especiales que ocurren con bastante frecuencia no poseen esta propiedad.

EJEMPLO 3.7

El sistema lineal $A\bar{x} = \bar{b}$ dado por

$$\begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 3 \\ 3.0001 \end{bmatrix}$$

tiene la única solución $\bar{x} = [1, 1]^t$. La aproximación a esta solución $\tilde{x} = [3, 0]^t$ tiene como vector residual

$$\bar{r} = \bar{b} - A\tilde{x} = \begin{bmatrix} 3 \\ 3.0001 \end{bmatrix} - \begin{bmatrix} 1 & 2 \\ 1.001 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 \\ 0.0002 \end{bmatrix} ,$$

así $\|\bar{r}\|_{\infty} = 0.0002$. Aunque la norma del vector residual es pequeña, la aproximación $\tilde{x} = [3, 0]^t$ es obviamente muy mala; en efecto, $\|\bar{x} - \tilde{x}\|_{\infty} = 2$

▲

La dificultad que surgió en el ejemplo puede ser explicada simplemente notando que la solución del sistema representa la intersección de las rectas

$$L_1: x_1 + 2x_2 = 3 \quad \text{y} \quad L_2: 1.0001x_1 + 2x_2 = 3.0001$$

El punto $(3, 0)$ está situado sobre L_1 y las rectas son casi paralelas. Esto implica que $(3, 0)$ también está bastante cerca de L_2 , aunque difiere significativamente del punto de intersección $(1, 1)$. (ver Figura 4).

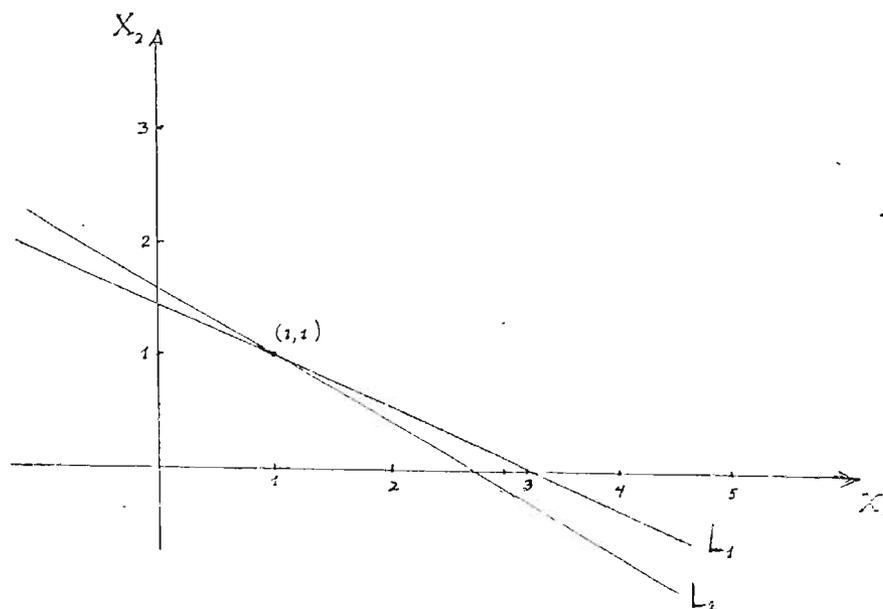


Fig. 4

En general, no podemos contar con la geometría del sistema para dar una indicación de cuando podría ocurrir el problema. Podemos, sin embargo, obtener esta información considerando la norma de A y la de su inversa.

TEOREMA 3.8

Si \tilde{x} es una aproximación a la solución de $A\bar{x} = \bar{b}$ y A es una matriz no singular, entonces para cualquier norma natural,

$$\|\bar{x} - \tilde{x}\| \leq \|\bar{r}\| \|A^{-1}\| \quad (3.14)$$

y

$$\frac{\|\bar{x} - \tilde{x}\|}{\|\bar{x}\|} \leq \|A\| \|A^{-1}\| \frac{\|\bar{r}\|}{\|\bar{b}\|}, \text{ siempre que } \bar{x} \neq 0 \text{ y } \bar{b} \neq 0 \quad (3.15)$$

donde \bar{r} es el vector residual para \tilde{x} con respecto al sistema (3.1).

PRUEBA

$$\text{Como } \bar{r} = \bar{b} - A\tilde{x}$$

$$\bar{r} = A\bar{x} - A\tilde{x}$$

$$\bar{r} = A(\bar{x} - \tilde{x}),$$

$$\text{luego } \bar{x} - \tilde{x} = A^{-1}\bar{r}.$$

Entonces

$$\|\bar{x} - \tilde{x}\| = \|A^{-1}\bar{r}\| \leq \|A^{-1}\| \|\bar{r}\|$$

Además como $\bar{b} = A\bar{x}$, $\|\bar{b}\| \leq \|A\| \|\bar{x}\|$, es decir

$$\frac{1}{\|\bar{x}\|} \leq \frac{\|A\|}{\|\bar{b}\|},$$

así que

$$\frac{\|\bar{x} - \tilde{x}\|}{\|\bar{x}\|} \leq \frac{\|A\| \|A^{-1}\|}{\|\bar{b}\|} \|\tilde{r}\|$$

Las condiciones (3.14) y (3.15) implican que las cantidades $\|A^{-1}\| = \|A\| \|A^{-1}\|$ pueden ser usadas para dar una indicación de la relación entre el residual y la exactitud de la aproximación. En general, el error relativo ($\|\bar{x} - \tilde{x}\| / \|\bar{x}\|$) es de mucho interés y, por la ecuación (3.15), este error es acotado por el producto de ($\|A\| \|A^{-1}\|$) con el residual relativo para esta aproximación, ($\|\tilde{r}\| / \|\bar{b}\|$). Cualquier norma conveniente puede ser usada para esta aproximación, siendo el único requisito que sea usado desde el principio.

DEFINICION 3.9

El número de condicionamiento $\mathbb{K}(A)$ de la matriz no singular A , relativo a la norma $\|\cdot\|$ es definido por

$$\mathbb{K}(A) = \|A\| \|A^{-1}\|$$

con esta notación las desigualdades en el teorema 3.8 se convierten en

$$\|\bar{x} - \tilde{x}\| \leq \mathbb{K}(A) \frac{\|\tilde{r}\|}{\|A\|}$$

y

$$\frac{\|\bar{x} - \tilde{x}\|}{\|\bar{x}\|} \leq \mathbb{K}(A) \frac{\|\tilde{r}\|}{\|\bar{b}\|}$$

Puesto que, para cualquier matriz no singular A ,

$$1 = \|I\| = \|AA^{-1}\| \leq \|A\| \|A^{-1}\| = \mathbb{K}(A),$$

la expectación es que la matriz A estará *bien condicionada* si $\kappa(A)$ es cercano a uno y *mal condicionada* cuando $\kappa(A)$ es significativamente mayor que uno.

EJEMPLO 3.10

La matriz para el sistema considerado en el ejemplo 3.7 fué

$$A = \begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix}$$

que tiene $\|A\|_{\infty} = 3.0001$. Esta norma sería considerada grande, sin embargo, como

$$A^{-1} = \begin{bmatrix} -10000 & 10000 \\ 5000.5 & -5000 \end{bmatrix}$$

$$\|A^{-1}\|_{\infty} = 20,000 ,$$

y para la norma- ∞ , $\kappa(A) = (20,000)(3.0001) = 60.002$. El tamaño del número de condicionamiento para este ejemplo nos guardaría de hacer decisiones precipitadas de exactitud basada en el residual de una aproximación.

Mientras en la teoría del número de condicionamiento de una matriz dependa totalmente de las matrices y sus inversas, en la práctica el cálculo de la inversa estaría sujeto a error por redondeo, dependiendo de la exactitud con la que los cálculos son efectuados.

Mas profundidad en la naturaleza del residual puede ser ganada considerando el residual de una solución redondeada \bar{x}' de

(3.1). Del ejemplo 3.5 conocemos que \bar{x}' puede ser escrito en la forma

$$\bar{x}' = \bar{x} + \bar{e}, \text{ donde } \|\bar{e}\|_{\infty} \leq \|\bar{x}\|_{\infty} 10^{-t}.$$

Por lo tanto

$$\bar{r} = \bar{b} - A\bar{x}' = \bar{b} - A\bar{x} - A\bar{e} = -A\bar{e}$$

y

$$\|\bar{r}\|_{\infty} \leq \|A\|_{\infty} \|\bar{e}\|_{\infty} \leq \|A\|_{\infty} \|\bar{x}\|_{\infty} 10^{-t}. \text{ Así}$$

$$\frac{\|\bar{r}\|_{\infty}}{\|A\|_{\infty}} \leq \|\bar{x}\|_{\infty} 10^{-t} \quad (3.16)$$

Ahora, si A está mal condicionada y \bar{x} refleja el mal condicionamiento de A , de modo que $\|\bar{x}\|_{\infty}$ sea mayor, entonces \bar{r} será grande. En otras palabras, la solución redondeada de una ecuación puede tener un residual grande. Este fenómeno lo ilustraremos en el siguiente ejemplo.

EJEMPLO 3.11

Sea

$$A = \begin{bmatrix} 1.000 & 1.001 \\ 1.000 & 1.000 \end{bmatrix}$$

y $\bar{b} = [2.0001, 2.000]^t$. Entonces la solución exacta de la ecuación $A\bar{x} = \bar{b}$ es $\bar{x} = [1, 1]^t$. Sin embargo, el vector $\bar{x} = [2, 0]^t$, que no está muy cerca de \bar{x} , tiene un vector residual muy pequeño $\bar{r} = [10^{-3}, 0]^t$. Por otro lado, la solución de $A\bar{x} = \bar{b} = [1, 0]^t$, es $\bar{x} = [-1000, 1000]^t$, y el vector $\tilde{x} = [-1001, 1000]^t$ está muy cerca de \bar{x} . Sin embargo, el vector residual de \tilde{x} es $\bar{r} = [0, 1]^t$,

el cual es tan grande como \bar{b} .

Es instructivo comparar el residual de una solución redondeada \bar{x}' de (3.1) con el residual de la solución \tilde{x} calculada por eliminación Gaussiana. Sabemos que \tilde{x} satisface (3.12), donde H satisface (3.13). Por lo tanto

$$\bar{r} = \bar{b} - A\tilde{x} = H\bar{x} \quad , \quad 0$$

$$\frac{\|\bar{r}\|_{\infty}}{\|A\|_{\infty}} \leq \phi(n) 10^{-t} \|\tilde{x}\|_{\infty} \quad (3.17)$$

Si $\phi(n)$ no es tan grande, la cota (3.17) es comparable con (3.16). Así, el residual de la solución calculada será aproximadamente del mismo tamaño que el residual de la solución exacta redondeada con t cifras.

Si hacemos la suposición de que la solución aproximada del sistema lineal está siendo determinada usando aritmética de t -dígitos y eliminación Gaussiana, el vector residual \bar{r} de la aproximación \tilde{x} tiene la propiedad

$$\|\bar{r}\| \approx 10^{-t} \|A\| \|\tilde{x}\| \quad (3.18)$$

De esta ecuación aproximada, una estimación para el efectivo número de condicionamiento en aritmética de t -dígitos puede ser obtenida sin necesidad de invertir la matriz A . En la actualidad, la aproximación en (3.18) asume que todas las operaciones aritméticas en la técnica de la eliminación Gaussiana son efectuadas usando aritmética de t -dígitos, aunque las operaciones que son necesitadas para determinar el residual están hechas en aritmética de doble precisión.

La aproximación para el número de condicionamiento $\mathbb{K}(A)$ con t -dígitos viene de la consideración del sistema lineal

$$A\bar{y} = \bar{r}$$

La solución para este sistema puede ser fácilmente aproximado dado que los multiplicadores para el método de eliminación -- Gaussiana han sido ya calculados.

En efecto, \tilde{y} , la solución aproximada de $A\bar{y} = \bar{r}$, satisface

$$\tilde{y} \approx A^{-1}\bar{r} = A^{-1}(\bar{b} - A\tilde{x}) = A^{-1}\bar{b} - A^{-1}A\tilde{x} = \bar{x} - \tilde{x}, \quad (3.19)$$

así \tilde{y} es una estimación del error en la aproximación de la solución del sistema original. La ecuación (3.18) puede ser consecuentemente usada para deducir que

$$\begin{aligned} \|\tilde{y}\| &\approx \|\bar{x} - \tilde{x}\| = \|A^{-1}r\| \leq \|A^{-1}\| \|r\| \approx \|A^{-1}\| (10^{-t} \|A\| \|\tilde{x}\|) \\ &= 10^{-t} \|\tilde{x}\| \mathbb{K}(A). \end{aligned}$$

Esto da una aproximación para el número de condicionamiento relacionado con la solución del sistema $A\bar{x} = \bar{b}$ usando eliminación Gaussiana y el tipo de aritmética de t -dígitos descrito -- arriba

$$\mathbb{K}(A) \approx \frac{\|\tilde{y}\|}{\|\tilde{x}\|} 10^t \quad (3.20)$$

EJEMPLO 3.12

El sistema lineal dado por

$$\begin{bmatrix} 3.3330 & 15920 & -10.333 \\ 2.2220 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 15913 \\ 28.544 \\ 8.4254 \end{bmatrix}$$

tiene la solución exacta $\bar{x} = [1, 1, 1]^t$.

Usando eliminación Gaussiana y aritmética de 5-dígitos resulta la matriz aumentada

$$\left[\begin{array}{ccc|c} 3.3330 & 15920 & -10.333 & 15913 \\ 0 & -10596 & 16.501 & -10580 \\ 0 & 0 & -5.0790 & -4.700 \end{array} \right]$$

y la solución aproximada de este sistema es

$$\tilde{x} = [1.2001, 0.99991, 0.92538]^t.$$

El vector residual correspondiente a x es calculado en doble precisión

$$\bar{r} = \bar{b} - A\tilde{x} = \begin{bmatrix} -0.00518 \\ 0.27413 \\ -0.18616 \end{bmatrix}$$

$$\|\bar{r}\|_{\infty} = 0.27413.$$

Usando aritmética de 5-dígitos para los cálculos da la aproximación para la inversa de A ,

$$A^{-1} = \begin{bmatrix} -1.1701 \times 10^{-4} & -1.4983 \times 10^{-1} & 8.5416 \times 10^{-1} \\ 6.2782 \times 10^{-5} & 1.2124 \times 10^{-4} & -3.0662 \times 10^{-4} \\ -8.6631 \times 10^{-5} & 1.3846 \times 10^{-1} & -1.9689 \times 10^{-1} \end{bmatrix}$$

y utilizando el teorema 2.9 resulta que $\|A^{-1}\|_{\infty} = 1.0041$ y $\|A\|_{\infty} = 15934$. El número de condicionamiento en este caso es exactamente grande:

$$\kappa(A) \approx (1.0041)(15934) = 14999. \quad (3.21)$$

La estimación para el número de condicionamiento dada en la

discusión precedente es obtenida resolviendo primero el sistema

$$\begin{bmatrix} 3.3330 & 15920 & -10.333 \\ 2.2220 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} -0.00518 \\ 0.27413 \\ -0.18616 \end{bmatrix}$$

lo cual implica que $\tilde{y} = [-0.20008, 8.9987 \times 10^{-5}, 0.074607]^t$, y entonces, usando la estimación dada en (3.20):

$$\mathbb{K}(A) \approx 10^5 \frac{\|\tilde{y}\|_\infty}{\|\tilde{x}\|_\infty} = \frac{10^5 (0.20008)}{1.2001} = 16672$$

Esta estimación está bastante relacionada al valor verdadero y requiere considerablemente menos trabajo computacional.

Dado que la solución real $\bar{x} = [1, 1, 1]^t$ es conocida para este sistema, podemos calcular ambas

$$\|\bar{x} - \tilde{x}\|_\infty = 0.2001 \quad \text{y} \quad \frac{\|\bar{x} - \tilde{x}\|}{\|\bar{x}\|} = \frac{0.2001}{1} = 0.2001$$

Las cotas del error dadas en el teorema 3.8 para estos valores serían

$$\|\bar{x} - \tilde{x}\|_\infty \leq \mathbb{K}(A) \frac{\|\bar{r}\|_\infty}{\|A\|_\infty} = \frac{(15999)(0.27413)}{15934} = 0.27525$$

y

$$\frac{\|x - \tilde{x}\|_\infty}{\|\bar{x}\|_\infty} \leq \mathbb{K}(A) \frac{\|\bar{x}\|_\infty}{\|\bar{b}\|_\infty} = \frac{(15999)(0.27431)}{15913} = 0.27561$$

▲

En la ecuación (3.19) usamos la estimación $\tilde{y} \approx \bar{x} - \tilde{x}$, donde y es la solución aproximada al sistema $A\bar{y} = \bar{r}$.

Sería razonable sospechar de este resultado que $\tilde{x} + \tilde{y}$ sería una aproximación más exacta a la solución del sistema lineal $A\bar{x} = \bar{b}$ que la original aproximación \tilde{x} . El método que utiliza esta suposición es llamado *refinamiento iterativo* o *mejoramiento iterativo*, y consiste en efectuar iteraciones en el sistema cuya parte derecha es el vector residual para aproximaciones sucesivas hasta obtener resultados satisfactoriamente exactos. El proceso es generalmente usado solamente en sistemas en los cuales se sospecha que la matriz relacionada esté mal condicionada aunque la aproximación para un sistema bien condicionado no se mejoraría significativamente por esta técnica.

El método comienza con una solución aproximada \tilde{x}_1 y produce una nueva solución aproximada \tilde{x}_2 que está más cercana a \tilde{x} que \tilde{x}_1 . Obviamente uno puede aplicar de nuevo el método a la solución aproximada \tilde{x}_2 para obtener otra mejor aproximación \tilde{x}_3 . Procediendo de este modo, obtenemos una sucesión $\langle \tilde{x}_k \rangle$ de soluciones aproximadas que convergen a la verdadera solución \bar{x} .

La idea del método es simple. Sea \tilde{x}_1 una solución aproximada de (3.1), obtenida presumiblemente por las técnicas de los capítulos II y III. Sea

$$\bar{r}_1 = \bar{b} - A\tilde{x}_1$$

el residual correspondiente a \tilde{x}_1 . Entonces si resolvemos el sistema

$$A\tilde{y}_1 = \bar{r}_1 \tag{3.22}$$

y calculamos

$$\tilde{x}_2 = \tilde{x}_1 + \tilde{y}_1,$$

se sigue que

$$\begin{aligned} \tilde{x}_2 &= \tilde{x}_2 + A^{-1} r_1 \\ &= \tilde{x}_1 + A^{-1} (\bar{b} - A\tilde{x}) \\ &= \tilde{x}_1 + A^{-1} \bar{b} - \tilde{x}_1 \\ &= A^{-1} \bar{b} \\ &= \bar{x} \end{aligned}$$

así que \tilde{x}_2 es la solución exacta. Note que este proceso es ventajoso, pero si la matriz A ha sido descompuesta para calcular \tilde{x}_1 ; es decir por reducción de Crout, entonces no se necesita descomponerla de nuevo para resolver (3.22). Como el cálculo del residual requiere solamente n^2 multiplicaciones, el proceso completo puede ser finalizado con $2n^2$ multiplicaciones.

ALGORITMO 3.13

Sea \tilde{x} una solución aproximada de la ecuación $A\tilde{x} = \bar{b}$, donde A es de orden n . Si A no es demasiada mal condicionada, este algoritmo cuando es ejecutado en aritmética de t -dígitos sobre escribe en \tilde{x} una solución aproximada que es cercanamente igual a la solución exacta redondeada a t cifras.

- 1) Calcular $\bar{r} = \bar{b} - A\tilde{x}$ en doble precisión y redondeada en precisión simple.
- 2) Calcular en simple precisión la solución de la ecuación $A\bar{y} = \bar{r}$.
- 3) Si $\|\bar{y}\|_{\infty} / \|\tilde{x}\|_{\infty} < 10^{-t}$, el proceso ha terminado.
- 4) $\tilde{x} \leftarrow \tilde{x} + \bar{y}$
- 5) Ir a paso 1.

Así como está construido el algoritmo, puede ser que se encuentre en un lazo indefinido; si A está bastante mal condicionada para que la iteración sea convergente. En este caso las correcciones \bar{y}_i no mostrarán un decrecimiento estable, ahora decreciendo, ahora creciendo. Así es suficiente terminar el algoritmo cuando

$$\frac{\|\bar{y}_i\|}{\|\tilde{x}_{i+1}\|} > \frac{\|y_{i-1}\|}{\|\tilde{x}_i\|}$$

EJEMPLO 3.14

Utilizando el algoritmo 3.13 encontraremos un mejoramiento de la solución obtenida en el ejemplo 3.12.

La solución aproximada que se obtuvo fué

$$\tilde{x}^{(1)} = [1.2001, 0.99991, 0.92538]^t$$

1) El vector residual sería:

$$\bar{r} = \begin{bmatrix} 15913 \\ 28.544 \\ 8.4254 \end{bmatrix} - \begin{bmatrix} 3.333 & 15920 & -10.333 \\ 2.222 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{bmatrix} \begin{bmatrix} 1.2001 \\ 0.99991 \\ 0.92538 \end{bmatrix}$$

$$\bar{r} = \begin{bmatrix} 15913 - 15913.00518 \\ 28.544 - 28.26987086 \\ 8.4254 - 8.611560367 \end{bmatrix}$$

$$\bar{r} = \begin{bmatrix} -0.005182 \\ 0.27412914 \\ -0.186160367 \end{bmatrix}, \text{ en doble precisión.}$$

$$\Rightarrow \bar{r}^{(1)} = \begin{bmatrix} -0.0051820 \\ 0.27413 \\ -0.18616 \end{bmatrix}, \text{ en precisión simple.}$$

2) Calcular la solución de $A\bar{y}^{(1)} = \bar{r}^{(1)}$

$$[A_1 \bar{r}^{(1)}] = \begin{bmatrix} 3.333 & 15920 & -10.333 & -0.005182 \\ 2.222 & 16.710 & 9.612 & 0.27413 \\ 1.5611 & 5.1791 & 1.6852 & -0.18616 \end{bmatrix}$$

$$= \begin{bmatrix} 3.333 & 15920 & -10.333 & -0.005182 \\ 0 & -10596 & 16.501 & 0.27758 \\ 0 & -7451.4 & 6.525 & -0.18373 \end{bmatrix}$$

$$= \begin{bmatrix} 3.333 & 15920 & -10.333 & -0.005182 \\ 0 & -10596 & 16.501 & 0.27758 \\ 0 & 0 & -5.0790 & -0.37893 \end{bmatrix}$$

$$y_3^{(1)} = \frac{-0.37893}{-5.0790} \implies y_3^{(1)} = 0.074607$$

$$y_2^{(1)} = \frac{0.27758 - 16.501(0.074607)}{-10596}$$

$$y_2^{(1)} = \frac{0.27758 - 1.2311}{-10596} = \frac{-0.95352}{-10596}$$

$$\implies y_2^{(1)} = 0.000089989$$

$$y_1^{(1)} = \frac{-0.005182 + 10.333(0.074607) - 15920(0.000089989)}{3.333}$$

$$y_1^{(1)} = \frac{-0.005182 + 0.77091 - 1.4326}{3.333} = \frac{-0.66687}{3.333}$$

$$\implies y_1^{(1)} = -0.20008$$

$$\tilde{y}^{(1)} = \begin{bmatrix} -0.2008 \\ 0.000089989 \\ 0.074607 \end{bmatrix}$$

$$3) \frac{\|y^{(1)}\|_{\infty}}{\|\tilde{x}^{(1)}\|_{\infty}} = \frac{0.2008}{1.2001} = 0.16732 > 10^{-5}$$

$$4) \tilde{x}^{(1)} \longleftarrow \tilde{x}^{(1)} + \tilde{y}^{(1)}$$

$$\tilde{x}^{(2)} = \begin{bmatrix} 1.2001 \\ 0.99991 \\ 0.92538 \end{bmatrix} + \begin{bmatrix} -0.20008 \\ 0.000089989 \\ 0.074607 \end{bmatrix}$$

$$\tilde{x}^{(2)} = \begin{bmatrix} 1.0000 \\ 1.0000 \\ 0.99999 \end{bmatrix}$$

1) El vector residual sería : $\bar{r}^{(2)} = \bar{b} - A\tilde{x}^{(2)}$

$$\bar{r}^{(2)} = \begin{bmatrix} 15913 \\ 28.544 \\ 8.4254 \end{bmatrix} - \begin{bmatrix} 3.333 & 15920 & -10.333 \\ 2.222 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{bmatrix} \begin{bmatrix} 1.0000 \\ 1.0000 \\ 0.99999 \end{bmatrix}$$

$$\bar{r}^{(2)} = \begin{bmatrix} 15913 - 15913.00010333 \\ 28.544 - 28.54390388 \\ 8.4254 - 8.425383148 \end{bmatrix}$$

$$\bar{r}^{(2)} = \begin{bmatrix} -0.00010333 \\ 0.00009612 \\ 0.000016852 \end{bmatrix}$$

2) Calcular la solución de $A\bar{y}^{(2)} = \bar{r}^{(2)}$

$$[A, \bar{r}^{(2)}] = \begin{bmatrix} 3.333 & 15920 & -10.333 & -0.00010333 \\ 2.222 & 16.710 & 9.6120 & 0.00009612 \\ 1.5611 & 5.1791 & 1.6852 & 0.000016852 \end{bmatrix}$$

$$= \begin{bmatrix} 3.333 & 15920 & -10.333 & -0.0001033 \\ 0 & -10596 & 16.501 & 0.00016501 \\ 0 & -7451.4 & 6.525 & 0.00006525 \end{bmatrix}$$

$$= \begin{bmatrix} 3.333 & 15920 & -10.333 & -0.00010333 \\ 0 & -10596 & 16.501 & 0.00016501 \\ 0 & 0 & -5.0790 & -0.00005079 \end{bmatrix}$$

$$y_3^{(2)} = \frac{-0.00005079}{-5.0790} \implies y_3^{(2)} = 0.00001$$

$$y_2^{(2)} = \frac{0.00016501 - 16.501(0.00001)}{-10596}$$

$$y_2^{(2)} = \frac{0.00016501 - 0.00016501}{-10596} = \frac{0}{-10596}$$

$$\implies y_2^{(2)} = 0$$

$$y_1^{(2)} = \frac{-0.00010333 + 10.333(0.00001) - 15920(0)}{3.333}$$

$$y_1^{(2)} = \frac{-0.00010333 + 0.00010333}{3.333} = \frac{0}{3.333}$$

$$\implies y_1^{(2)} = 0$$

$$\tilde{\mathbf{y}}^{(2)} = \begin{bmatrix} 0 \\ 0 \\ 0.00001 \end{bmatrix}$$

$$3) \frac{\|\tilde{\mathbf{y}}^{(2)}\|_{\infty}}{\|\tilde{\mathbf{x}}^{(2)}\|_{\infty}} = \frac{0.0001}{1.00000} = 0.00001 \neq 10^{-5}$$

$$4) \tilde{\mathbf{x}}^{(3)} = \begin{bmatrix} 1.0000 \\ 1.0000 \\ 0.99999 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0.00001 \end{bmatrix}$$

$$\tilde{\mathbf{x}}^{(3)} = \begin{bmatrix} 1.0000 \\ 1.0000 \\ 1.0000 \end{bmatrix}$$

1) Calcular el vector residual $\bar{\mathbf{r}}^{(3)} = \bar{\mathbf{b}} - A\tilde{\mathbf{x}}^{(3)}$

$$\bar{\mathbf{r}}^{(3)} = \begin{bmatrix} 15913 \\ 28.544 \\ 8.4254 \end{bmatrix} - \begin{bmatrix} 3.333 & 15920 & -10.333 \\ 2.222 & 16.710 & 9.6120 \\ 1.5611 & 5.1791 & 1.6852 \end{bmatrix} \begin{bmatrix} 1.0000 \\ 1.0000 \\ 1.0000 \end{bmatrix}$$

$$\bar{\mathbf{r}}^{(3)} = \begin{bmatrix} 15913 - 15913 \\ 28.544 - 28.544 \\ 8.4254 - 8.4254 \end{bmatrix}$$

$$\bar{\mathbf{r}}^{(3)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

EJERCICIOS

3.1 Resolver los siguientes sistemas lineales usando Eliminación Gaussiana y mejoramiento iterativo. También estimar $K(A)$

$$a) \quad x_1 + \frac{1}{2}x_2 + \frac{1}{3}x_3 = \frac{11}{6},$$

$$5x_1 + \frac{10}{3}x_2 + \frac{5}{2}x_3 = \frac{65}{6},$$

$$\frac{100}{3}x_1 + 25x_2 + 20x_3 = \frac{225}{3}. \quad (\text{aritmética de 3 dígitos})$$

$$b) \quad 1.003x_1 + 58.09x_2 = 68.12,$$

$$5.550x_1 + 321.8x_2 = 3.77.8. \quad (\text{aritmética de 4 dígitos})$$

$$c) \quad 3.9x_1 + 1.6x_2 = 5.5$$

$$6.8x_1 + 2.9x_2 = 9.7 \quad (\text{aritmética de 2 dígitos})$$

$$e) \quad 4.56x_1 + 2.18x_2 = 6.74,$$

$$2.79x_1 + 1.38x_2 = 4.13. \quad (\text{aritmética de 3 dígitos})$$

3.2 Calcular los números de condicionamiento de las siguientes matrices relativos a $\|\cdot\|_\infty$

$$a) \quad \begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix}$$

$$b) \quad \begin{bmatrix} 3.9 & 1.6 \\ 6.8 & 2.9 \end{bmatrix}$$

$$c) \quad \begin{bmatrix} 1.0003 & 58.09 \\ 5.550 & 321.8 \end{bmatrix}$$

comparar con 3.1(c) comparar con 3.1(b)

3.3 Mostrar que, si B es singular, entonces

$$\frac{1}{K(A)} \leq \frac{\|A - B\|}{\|A\|}$$

Sugerencia: existe un vector $\bar{x} \neq 0$, con $\|\bar{x}\| = 1$, tal que $B\bar{x} = \bar{0}$. Derivar la estimación usando

$$\|A\bar{x}\| \geq \frac{\|A\|}{\|A^{-1}\|}]$$

3.4 Usando el ejercicio 3.3, estimar los números de condicionamiento para las siguientes matrices

a) $\begin{bmatrix} 1 & 2 \\ 1.0001 & 2 \end{bmatrix}$

comparar con 3.2(a)

b) $\begin{bmatrix} 3.9 & 1.6 \\ 6.8 & 2.9 \end{bmatrix}$

comparar con 3.1(c)

B I B L I O G R A F I A

1. G. W. Stewart
"INTRODUCTION TO MATRIX COMPUTATIONS"
Academic Press. Inc., New York.
2. Richard L. Burden, J. Douglas Faires and Albert C. Reynolds.
"NUMERICAL ANALYSIS"
Prindle Weber Schmidt
3. Ben Noble
"APPLIED LINEAR ALGEBRA"
Prentice - Hall, Inc., New Jersey. 1969
4. Marvin Marcus y Henryk Mink
"ELEMENTOS DE ALGEBRA LINEAL"
Editorial Limusa - Wiley, S.A.
México, 1971
5. Mina S. de Carakushansky y Guilherme de la Penha.
"ALGEBRA LINEAL"
Editorial Mc Graw - Hill, 1976
6. Richard E. Johnson
"ALGEBRA LINEAL"
Compañía Editorial Continental, S.A.
México, 1969.
7. G. Hadley
"ALGEBRA LINEAL"
Fondo Educativo Interamericano, S.A.
Bogotá - Cali - México - Panamá - San Juan - Santiago - 1969