UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE ESTUDIOS SUPERIORES ACATLÁN

Material de apoyo a la docencia, para la asignatura de Métodos Numéricos I del plan de estudios 2006, de la licenciatura en Matemáticas Aplicadas y Computación

TESINA

QUE PARA OBTENER EL TÍTULO DE
LICENCIATURA EN MATEMÁTICAS APLICADAS Y COMPUTACIÓN

PRESENTA
Claudia Sánchez Vera

Asesor:
Mayra Olguín Rosas

Agosto 2007

Material de apoyo a la docencia, para la asignatura de Métodos Numéricos I del plan de estudios 2006, de la licenciatura en Matemáticas Aplicadas y Computación

Claudia Sánchez Vera

1 de agosto de 2007

Agradecimientos

A mis papás José Eugenio y María Eugenia, por todo el apoyo que me han brindado, por sus consejos día a día de la situaciones vividas a lo largo de este trabajo, por ser los mejores y estar conmigo incondicionalmente, gracias porque sin ellos y sus enseñanzas no estaría aquí ni sería quien soy ahora, a ellos les dedico este trabajo... los quiero mucho.

A mi hermano Eugenio, por aguantar todas mis manías de estudio, por confiar en mí y por el apoyo brindado en la culminación de ésta etapa tan importante de mi vida.

A mi asesora Mayra, por los conocimientos que compartió conmigo y por su valioso tiempo dedicado a éste trabajo.

A mis sinodales y profesores, por sus valiosas enseñanzas.

A mis amigos, por todos los momentos vividos durante la universidad, y porque siempre están, estuvieron y seguirán estando, brindándome cariño y soporte.

A la UNAM y a la FES Acatlán, por contribuir a mi formación personal y académica.

A Dios, por bendecidme para llegar hasta donde he llegado.

A todas aquellas personas que de alguna manera hicieron posible la terminación de este trabajo y que no las mencione… gracias a todos.

Índice general

Objetivo

Desarrollar un material de apoyo a la docencia dirigido principalmente a los estudiantes de la licenciatura en Matemáticas Aplicadas y Computación, como complemento para la enseñanza de la asignatura de Métodos Numéricos I, del plan de estudios 2006.

Introducción

Al leer el título de este trabajo tal vez te preguntes: ¿porqué elaborar un material de apoyo a la docencia de Métodos Numéricos I?, si en la actualidad existen muchos libros y páginas de Internet relacionados con este tema. Efectivamente lo anterior es muy cierto; sin embargo, se han modificado los planes de estudio de la Licenciatura en Matemáticas Aplicadas y Computación, con el fin de mejorar y actualizar la carrera, siendo el más actual el plan de estudios 2006; por lo que creo que el generar un material didáctico de apoyo a la docencia enfocado en la asignatura de Métodos Numéricos I del plan 2006, me permite contribuir a este fin.

Así mismo recuerdo platicando con compañeros durante la carrera, el como nos gustaría encontrar en un sólo libro todos los temas explicados de forma clara y sencilla referente a las asignaturas que llevábamos; debido a que en muchas ocasiones en algunos libros el tema no estaba explicado tan claramente o en los ejemplos no se desarrollaba a detalle como se obtenían los resultados, y no entendíamos como pasaban de un punto a otro.

Por lo anterior, me decidí a realizar un material de apoyo a la docencia, en el cual se explican de forma clara y sencilla los temas de la asignatura de Métodos Numéricos I del plan de estudios 2006; así mismo en los ejemplos se explican a detalle como se obtienen los resultados.

Éste material esta dirigido principalmente a: alumnos que están cursando dicha asignatura, alumnos que están preparando el extraordinario y a los profesores que la imparten.

En el capítulo I, se explicará que son los métodos numéricos, el análisis de error, así como las herramientas disponibles para el análisis numérico.

En el capítulo II, se expondrán algunos métodos numéricos para hallar raíces reales de ecuaciones no lineales de una variable, que satisfacen a una ecuación del tipo $f(x)=0$, incluyendo los métodos de bisección, falsa posición, de Newton y finalmente de la secante.

En el capítulo III, se describirán algunos métodos numéricos para hallar los valores $x_1, x_2, \ldots, x_n$ que en forma simultánea satisfacen a un sistema de ecuaciones lineales.

En el capítulo IV, como complemento del capítulo III, se tratarán algunos métodos numéricos para factorizar una matriz A, en la forma LU; mismos que también sirven para hallar los valores $x_1, x_2, \ldots, x_n$ que en forma simultánea satisfacen a un sistema de ecuaciones lineales.

Capítulo 1

Análisis del error

1.1 Introducción

Un método numérico es un procedimiento mediante el cual se obtiene, casi siempre de manera aproximada, la solución de ciertos problemas realizando cálculos puramente aritméticos y lógicos (operaciones aritméticas elementales, cálculo de funciones, consulta de una tabla de valores, etc.).

Cuando se emplean los métodos numéricos para resolver algún tipo de ecuación se debe a que:

1. No existe solución exacta.

2. Resolverla de manera analítica es demasiado costoso en cuanto a trabajo, ó incluso imposible.

Al aplicar los métodos numéricos se sigue un procedimiento que consiste en una lista finita de instrucciones precisas que especifican una secuencia de operaciones algebraicas y lógicas (algoritmo), que producen una aproximación de la solución del problema (**solución numérica**) ó un mensaje. La eficiencia en el cálculo de dicha aproximación depende, en parte, de la facilidad de implementación del algoritmo y de las características especiales y limitaciones de los instrumentos de cálculo (las calculadoras o computadoras). En general, al emplear estos instrumentos de cálculo se introducen errores; los cuales se desean cuantificar y minimizar.

Existen dos tipos de errores que están relacionados directamente con los métodos numéricos:

1. Errores de redondeo.

2. Errores de truncamiento.

Sin embargo existen otro tipo de errores como las equivocaciones, errores de formulación o del modelo, incertidumbre en la obtención de datos, entre otros; aunque los métodos numéricos pueden estudiarse en su mayoría de los casos, en forma independiente de estos errores. Por consiguiente se supondrá que no hay errores por equivocaciones, que el modelo es adecuado y que se está trabajando sin errores en las mediciones de los datos.

1.2. Errores de redondeo: aritmética del punto flotante, errores de truncamiento, absoluto y relativo

Los métodos numéricos dan una aproximación a la solución analítica (solución exacta), es decir existe cierto error en la solución numérica; por lo que es muy importante el estudio de los diferentes tipos de error.

Antes de analizar los errores asociados con los métodos numéricos, es útil repasar algunos conceptos básicos referentes a la representación aproximada a los números mismos.

El concepto de cifras o dígitos significativos se ha desarrollado para designar formalmente la confiabilidad de un valor numérico. Las ***cifras significativas*** de un número son aquellas que pueden utilizarse en forma confiable. El dígito principal (ó más significativo) de un número real x no nulo, es el dígito no nulo más a la izquierda de su expansión decimal. Todos los dígitos incluyendo los ceros a la derecha del dígito significativo principal, son significativos y el 'ultimo desplegado se llama dígito menos significativo. Sin embargo los ceros a la izquierda del dígito significativo principal no son significativos. Así el dígito significativo principal de 0.067 es 6 (no cero), pero el dígito menos significativo de 1.090 es el cuarto dígito significativo, cero (no el tercero 9).

Error absoluto y relativo

Se denomina ***error absoluto*** de un número x_n que aproxima a otro x, a la distancia entre ellos. Si sólo se dispone de que el error es por ejemplo 1m, no se sabe nada de la fiabilidad del resultado, ya que no es lo mismo decir que se ha cometido un error de un metro al medir la altura de una persona que al medir la distancia entre dos galaxias. Se debe reflejar de alguna manera "lo que se esta evaluando" en el dato del error. Para esto se utiliza el ***error relativo*** que es el cociente entre el error absoluto y el objeto evaluado.

Cuando se manejan cantidades "muy grandes" o "muy pequeñas", el error absoluto puede ser engañoso, mientras que el error relativo es más significativo en estos casos.

Si se cuenta con la solución analítica se puede calcular el error verdadero en forma exacta:

Definición 1.2.1 Error absoluto, relativo y relativo porcentual cuando se conoce la solución analítica.

Si x_n es una aproximación a x,

1. el error absoluto verdadero se define como:

$$\varepsilon_{a,v} = |x - x_n| \tag{1.1}$$

2. el error relativo verdadero como:

$$\varepsilon_{r,v} = \left| \frac{x - x_n}{x} \right| \qquad \text{siempre que} \, x \neq 0 \tag{1.2}$$

3. y el error relativo porcentual verdadero como:

$$\varepsilon_{r,v}(\%) = \left|\frac{x - x_n}{x}\right| 100\% \qquad \text{siempreque} x \neq 0 \tag{1.3}$$

■**Ejemplo1.2.1** Se desea medir la longitud de un puente y la de un tornillo, y se obtiene $5999\,\text{cm}$ y $5\,\text{cm}$, respectivamente. Si los valores verdaderos son $6000\,\text{cm}$ y $6\,\text{cm}$, calcular el error absoluto verdadero y el error relativo porcentual verdadero; y compare los resultados.

Con los datos anteriores, se sabe que para el puente $x = 6000\,\text{cm}$ y $x_n = 5999\,\text{cm}$, y para el tornillo $x = 6\,\text{cm}$ y $x_n = 5\,\text{cm}$. Se comienza calculando el error absoluto verdadero con la ecuación (1); obteniéndose:

$$\begin{aligned} \varepsilon_{a,v} &= \left|6000\text{cm} - 5999\text{cm}\right| = 1\text{cm} && \text{paraelpuente} \\ \varepsilon_{a,v} &= \left|6\text{cm} - 5\text{cm}\right| = 1\text{cm} && \text{paraeltornillo} \end{aligned}$$

Después se calcula el error relativo verdadero porcentual con la ecuación (1.3); obteniéndose:

$$\varepsilon_{r,v} = \left|\tfrac{6000-5999}{6000}\right| 100\% = 0.01666667\% \qquad \text{paraelpuente}$$

$$\varepsilon_{r,v} = \left|\tfrac{6-5}{6}\right| 100\% = 16.66666667\% \qquad \text{paraeltornillo}$$

$\therefore$ Aunque ambas medidas tienen un error absoluto verdadero de 1cm, no dice nada sobre la fiabilidad del resultado; a diferencia del error relativo porcentual verdadero del tornillo que es mucho mayor; con lo que se puede concluir que se ha hecho un buen trabajo en la medición del puente; mientras que la medición para el tornillo se realizó de manera deficiente.

En la realidad la mayoría de las veces no se cuenta con la solución analítica, por lo que sólo se tiene una aproximación del error.

Ciertos métodos numéricos usan un método iterativo para calcular la solución numérica; en tales métodos ésta se calcula considerando la aproximación anterior. Éste proceso se efectúa varias veces, esperando cada vez una mejor aproximación a la solución, en tales casos el error se puede calcular como se indica a continuación:

Definición 1.2.2 Error absoluto, relativo y relativo porcentual.

Si x_{n-1} es una aproximación a x_n,

1. el error absoluto se define como:

$$\varepsilon_a = \left|x_n - x_{n-1}\right| \tag{1.4}$$

2. el error relativo como:

$$\varepsilon_r = \left|\frac{x_n - x_{n-1}}{x_n}\right| \qquad \text{siempreque} x_n \neq 0 \tag{1.5}$$

3. y el error relativo porcentual como:

$$\varepsilon_r(\%) = \left|\frac{x_n - x_{n-1}}{x_n}\right| 100\% \qquad \text{siempreque} x_n \neq 0 \tag{1.6}$$

En los métodos iterativos, el error relativo (por ser más significativo) se utiliza como uno de los criterios de paro. (Ver capítulo 2: secciones 2.1, 2.2, 2.3 y 2.4; y capítulo 3: sección 3.4).

Error de redondeo

Los ***errores de redondeo*** se producen ya que los números pueden requerir para su representación decimal, de una infinidad de dígitos. En la práctica, para su manejo sólo debe considerarse un número finito de dígitos en su representación, procediéndose a su determinación mediante un adecuado redondeo. Un caso típico lo presentan las computadoras que, en su memoria, almacenan sólo representaciones finitas de los números.

Las ***reglas para redondear un número*** son las siguientes:

1. Si el primero de los dígitos descartado $(n+1)$ es menor que 5, no se altera el n-ésimo dígito.

2. Si el primero de los dígitos descartado $(n+1)$ es mayor que 5, se suma uno al dígito de la n-ésima posición.

3. Si el primero de los dígitos descartado $(n+1)$ es igual que 5, se redondea como en el caso de la segunda regla.

■**Ejemplo1.2.2** Redondear el número 4179.3650002 a 4, 5, 6 y 7 dígitos; y el número 3.14159 a 3, 4 y 5 dígitos; siguiendo las reglas de redondeo.

Se redondea el número 4179.3650002:

1. A 4 dígitos; como el primero de los dígitos descartado es 3 y es <5 no cambian los dígitos restantes: 4179

2. A 5 dígitos; como el primero de los dígitos descartado es 6 y es >5 se le agrega un uno al último de los dígitos: 4179.4

3. A 6 dígitos; como el primero de los dígitos descartado es 5 y es $=5$ se le agrega un uno al último de los dígitos: 4179.37

4. A 7 dígitos; como el primero de los dígitos descartado es 0 y es <5 no cambian los dígitos restantes: 4179.365

Después se redondea el número 3.14159:

1. A 3 dígitos; como el primero de los dígitos descartado es 1 y es <5 no cambian los dígitos restantes: 3.14

2. A 4 dígitos; como el primero de los dígitos descartado es 5 y es $=5$ se le agrega un uno al último de los dígitos: 3.142

3. A 5 dígitos; como el primero de los dígitos descartado es 9 y es >5 se le agrega un uno al último de los dígitos: 3.1416

Al resolver un problema matemático por medio de la computadora el número limitado de dígitos provoca errores de redondeo. En algunos casos, estos errores causan efectos muy serios y hacen que los resultados de los cálculos carezcan por completo de sentido. Por esta razón es importante saber como se representan los números en la computadora.

Aritmética del punto flotante

El sistema numérico que se usa cotidianamente se llama sistema decimal. La base del sistema numérico decimal es 10. Sin embargo, las computadoras no usan el sistema decimal en los cálculos ni en la memoria, sino que usan el sistema binario (base 2).

La base de un número se denota por medio de un subíndice; por ejemplo: 3.224_{10}, es 3.224 en base 10 ó 1001.11_2, es 1001.11 en base 2.

Un esquema de almacenamiento ampliamente usado para la forma binaria es el estándar de la ***IEEE Standard for Binary Floating-Point Arithmetic***. Esta forma fue definida por el *Institute of Electrical and Electronic Engineers*, IEEE (Instituto de Ingenieros Eléctricos y Electrónicos). En general los fabricantes de las microcomputadoras utilizan este estándar para el hardware de punto flotante.

Para almacenar un número en forma binaria, primero se *normaliza*; es decir, se vuelve a escribir el número, en un equivalente binario de notación científica convencional, desplazando el punto decimal inmediatamente a la derecha del primer bit significativo del número (es decir después del primer 1 de izquierda a derecha). Luego el resultado se multiplica por 2^n, si el punto se desplazó n posiciones a la izquierda, o por 2^{-n} si el punto se desplazó n posiciones a la derecha; la multiplicación es necesaria para recuperar el valor original del número.

Un número de punto flotante se puede escribir en la forma normalizada general, de la siguiente manera:

$$V = (-1)^s (1.f)_2 (2^t)_{10} \tag{1.7}$$

donde V es el número de punto flotante, s es el **bit de signo**, cuyo valor puede ser 0 ó 1 (si $s=1$ indica un número negativo, si $s=0$ indica un número positivo). Los bits significativos del número están contenidos en la cantidad $(1.f)$, que se denomina **mantisa**. La potencia t al final de la representación es la **característica** ó **exponente**.

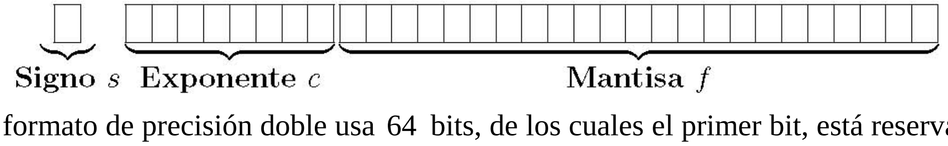
El valor binario del exponente t no se almacena directamente; en vez de ello, se almacena en forma *sesgada* o *equivalente* como un valor binario no negativo c. La relación del exponente real t en términos del valor almacenado c y el sesgo b es:

$$t = c - b \tag{1.8}$$

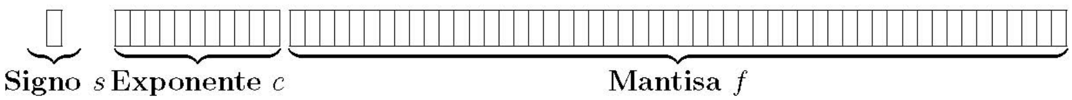
$$\Rightarrow c = t + b \tag{1.9}$$

donde b, toma un valor 127 para precisión simple (32 bits) y 1023 para precisión doble (64 bits).

El formato de precisión simple usa 32 bits, de los cuales el primer bit (llamado bit más significativo), está reservado para el bit del signo s; los 8 bits siguientes se usan para almacenar un patrón de bits que representan el exponente t en forma sesgada (c) y los 23 bits restantes se usan para la mantisa f. Sólo se almacena la parte de la mantisa denotada por f porque el dígito principal siempre es igual a 1 y se entiende que forma parte del número.



El formato de precisión doble usa 64 bits, de los cuales el primer bit, está reservado para el bit del signo s; los 11 bits siguientes se usan para almacenar un patrón de bits que representan el exponente t en forma sesgada (c) y los 52 bits restantes se usan para la mantisa f. Sólo se almacena la parte de la mantisa denotada por f porque el dígito principal siempre es igual a 1 y se entiende que forma parte del número.



El exponente es desplazado antes de ser almacenado, ajustando su valor para ponerlo dentro de un rango sin signo adaptable a una comparación. Así, para un número en precisión simple, un exponente en el rango -126 (-1111110_2) a $+127$ (1111111_2) es desplazado mediante la suma de 127 para obtener un valor en el rango 1 a 254 (0 y 255 tienen valores especiales descritos más adelante). Para un número en precisión doble, un exponente en el rango -1022 (-1111111110_2) a $+1023$ (1111111111_2) es desplazado mediante la suma de 1023 para obtener un valor en el rango 1 a 2046 (0 y 2047 tienen valores especiales descritos más adelante).

Antes de ver como se almacena un número en una computadora en una palabra de 32 y 64 bits, es necesario saber como realizar las conversiones del sistema decimal al binario y viceversa; tanto para números enteros, como para fraccionarios.

Conversión de números enteros del sistema decimal al sistema binario

Para convertir un número n del sistema decimal al sistema binario, se divide el número n entre 2 y se registra el cociente c_1 y el residuo r_1 resultantes; se divide c_1 entre 2 y se anotan el nuevo cociente c_2 y el nuevo residuo r_2. Este procedimiento se repite hasta obtener un cociente c_i igual a cero con un residuo r_i. El número equivalente a n en el sistema binario queda formado así: $r_i, r_{i-1}, r_{i-2}, \ldots, r_1$.

■**Ejemplo1.2.3** Convertir -118_{10} al sistema binario.

Se comienza dividiendo 118 entre 2; lo que resulte se divide nuevamente entre 2. Se sigue el mismo procedimiento hasta que el cociente sea igual a cero. Los residuos son los que forman el número en sistema binario. (El signo no se toma en cuenta para realizar las divisiones, sólo se coloca al final de realizar la conversión).

118	$\div$	2	=	59	ysobra	0	donde	$c_1 = 59$	y	$r_1 = 0$	
59	$\div$	2	=	29	ysobra	1		$c_2 = 29$	y	$r_2 = 1$	
29	$\div$	2	=	14	ysobra	1		$c_3 = 14$	y	$r_3 = 1$	
14	$\div$	2	=	7	ysobra	0		$c_4 = 7$	y	$r_4 = 0$	
7	$\div$	2	=	3	ysobra	1		$c_5 = 3$	y	$r_5 = 1$	
3	$\div$	2	=	1	ysobra	1		$c_6 = 1$	y	$r_6 = 1$	
1	$\div$	2	=	0	ysobra	1		$c_7 = 0$	y	$r_7 = 1$	$\uparrow$

$\therefore\ -118_{10} = -1110110_2$

Conversión de números enteros del sistema binario al sistema decimal

Para convertir un número n del sistema binario al sistema decimal, se multiplica cada dígito de n por la base 2 elevado a una potencia igual a la posición del dígito, tomando como posición cero el dígito más a la derecha. La suma da el equivalente en decimal.

■**Ejemplo1.2.4** Convertir -1110110_2 al sistema decimal.

Se comienza multiplicando cada dígito de 1110110 por 2 elevado a una potencia igual a la posición del dígito, tomando como posición cero el dígito más a la derecha; sumando después todos los términos que formarán el número en sistema decimal. (El signo no se toma en cuenta para realizar las multiplicaciones, sólo se coloca al final de realizar la conversión).

$$-1110110_2 = (1\times 2^6 + 1\times 2^5 + 1\times 2^4 + 0\times 2^3 + 1\times 2^2 + 1\times 2^1 + 0\times 2^0) = -118_{10}$$

$\therefore\ -1110110_2 = -118_{10}$

Conversión de números fraccionarios del sistema decimal al sistema binario

Para convertir un número n fraccionario del sistema decimal al sistema binario, se multiplica dicho número por 2; el resultado tiene una parte entera e_1 y una parte fraccionaria f_1. Se multiplica ahora f_1 por 2 y se obtiene un nuevo producto con parte entera e_2 y fraccionaria f_2. Este procedimiento se repite un número suficiente de veces ó hasta que se presenta $f_i = 0$. El equivalente de n en sistema binario queda así: $0.e_1e_2e_3e_4\ldots$

■**Ejemplo1.2.5** Convertir 0.625_{10} al sistema binario.

Se comienza multiplicando 0.625 por 2; la parte fraccionaria que resulte se multiplica nuevamente por 2. Se sigue el mismo procedimiento un número suficiente de veces ó hasta que la parte fraccionaria sea igual a cero. La parte entera que resulta de estas multiplicaciones es la que forma el número en sistema binario.

$$\begin{array}{ccccccccc} 0.625 & \times & 2 & = & 1.25 & \text{donde} & e_1 = 1 & \text{y} & f_1 = 0.25 & \downarrow \\ 0.25 & \times & 2 & = & 0.5 & & e_2 = 0 & \text{y} & f_2 = 0.5 \\ 0.5 & \times & 2 & = & 1.0 & & e_3 = 1 & \text{y} & f_3 = 0.0 \end{array}$$

$\therefore\ 0.625_{10} = 0.101_2$

Conversión de números fraccionarios del sistema binario al sistema decimal

Para convertir un número n fraccionario del sistema binario al sistema decimal, se multiplica cada dígito de n por la base 2 elevado a una potencia igual a la posición del dígito, tomando como posición inicial -1, a partir del punto. La suma da el equivalente en decimal.

■**Ejemplo1.2.6** Convertir 0.101_2 al sistema decimal.

Se comienza multiplicando cada dígito de 0.101 por 2 elevado a una potencia igual a la posición del dígito, tomando como posición inicial -1, a partir del punto; sumando después todos los términos que formarán el número en sistema decimal.

$$0.101_2 = (1\times 2^{-1} + 0\times 2^{-2} + 1\times 2^{-3}) = 0.625_{10}$$

$\therefore\ 0.101_2 = 0.625_{10}$

Ya que se sabe como realizar las conversiones del sistema decimal al binario y viceversa, a continuación se verán ejemplos de como se almacena un número en una computadora.

■**Ejemplo1.2.7** Representar el número -118.625 en una palabra de 32 bits usando el sistema de la IEEE.

Para poder representar -118.625 en un tamaño de palabra de 32 bits, se tienen que identificar cada una de sus partes (el signo s, el exponente t en forma sesgada (c) y la mantisa f).

Dado que es un número negativo, el signo se representa con $\mathbf{s}=1$.

Primero se escribe el número -118.625 sin signo en base 2, obteniendo primero la parte entera y después la fraccionaria. El resultado es 1110110.101; resultados obtenidos en los ejemplos 1.2.3 y 1.2.5

Ya que se tiene el número en base 2, se normaliza desplazando el punto decimal inmediatamente a la derecha del primer bit significativo del número (es decir después del primer 1 de izquierda a derecha), obteniendo: $1110110.101=(1.110110101)_2(2^6)_{10}$.

Se escribe éste número de punto flotante en la forma normalizada general, usando la ecuación (1.7); obteniéndose:

$$V=(-1)^s(1.f)_2(2^t)_{10}=(-1)^1(1.110110101)_2(2^6)_{10}$$

El 2 se eleva a la 6, ya que el punto decimal se desplazó 6 veces a la izquierda.

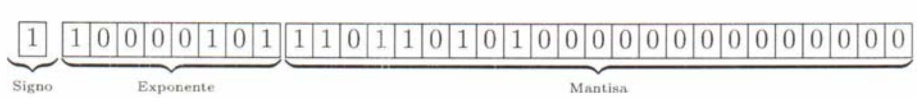
La mantisa es la parte a la derecha del punto decimal, y se rellena con ceros a la derecha hasta que se obtengan los 23 bits. Es decir $\mathbf{f}=11011010100000000000000$.

El exponente es $t=6$, se necesita desplazarlo para obtener el valor de c, usando la ecuación (1.9). Para el formato IEEE de 32 bits, el desplazamiento es $b=127$: $\Rightarrow \mathbf{c}=6+127=133$. Éste valor se convierte en binario.

133	$\div$	2	=	66	ysobra	1	donde	$c_1=66$	y	$r_1=1$	
66	$\div$	2	=	33	ysobra	0		$c_2=33$	y	$r_2=0$	
33	$\div$	2	=	16	ysobra	1		$c_3=16$	y	$r_3=1$	
16	$\div$	2	=	8	ysobra	0		$c_4=8$	y	$r_4=0$	
8	$\div$	2	=	4	ysobra	0		$c_5=4$	y	$r_5=0$	
4	$\div$	2	=	2	ysobra	0		$c_6=2$	y	$r_6=0$	
2	$\div$	2	=	1	ysobra	0		$c_7=1$	y	$r_7=0$	
1	$\div$	2	=	0	ysobra	1		$c_8=0$	y	$r_8=1$	$\uparrow$

$\Rightarrow 133_{10}=10000101_2$

$\therefore$ -118.625 se representa en una palabra de 32 bits usando el sistema de la IEEE, de la siguiente manera:



Comprobación:

Éste número de punto flotante se puede escribir en la forma normalizada general, de la siguiente manera; (para obtener el valor de t, se utiliza la ecuación (1.8)):

$$
\begin{aligned}
1\ \ 10000101\ \ 11011010100000000000000 &= \\
(-1)^1(1.11011010100000000000000)_2(2^{133-127})_{10} &= \\
(-1)(1.11011010100000000000000)_2(2^6)_{10} &= \\
-1110110.10100000000000000_2 &= -118.625
\end{aligned}
$$

Lo anterior se obtuvo desplazando 6 posiciones a la derecha el punto decimal de la mantisa, ya que $t = 6$. Después se escribe el número sin signo en base 10, obteniendo primero la parte entera y después la fraccionaria. El resultado es -118.625, ver los resultados obtenidos en los ejemplos 1.2.4 y 1.2.6.

■**Ejemplo1.2.8** Representar el número 0.2 en una palabra de 32 bits usando el sistema de la IEEE.

Para poder representar 0.2 en un tamaño de palabra de 32 bits, se tienen que identificar cada una de sus partes (el signo s, el exponente t en forma sesgada (c) y la mantisa f).

Dado que es un número positivo, el signo se representa con $\mathbf{s} = 0$.

Primero se escribe el número en base 2, como no tiene parte entera, se obtiene sólo la parte fraccionaria. Se comienza multiplicando 0.2 por 2; la parte fraccionaria que resulte se multiplica nuevamente por 2. Se sigue el mismo procedimiento un número suficiente de veces ó hasta que la parte fraccionaria sea igual a cero. La parte entera que resulta de estas multiplicaciones es la que forma el número en sistema binario.

$$
\begin{array}{ccccccccccc}
0.2 & \times & 2 & = & 0.4 & \text{donde} & e_1 = 0 & \text{y} & f_1 = 0.4 & \downarrow \\
0.4 & \times & 2 & = & 0.8 & & e_2 = 0 & \text{y} & f_2 = 0.8 & \\
0.8 & \times & 2 & = & 1.6 & & e_3 = 1 & \text{y} & f_3 = 0.6 & \\
0.6 & \times & 2 & = & 1.2 & & e_4 = 1 & \text{y} & f_4 = 0.2 & \\
0.2 & \times & 2 & = & 0.4 & & e_5 = 0 & \text{y} & f_5 = 0.4 & \\
0.4 & \times & 2 & = & 0.8 & & e_6 = 0 & \text{y} & f_6 = 0.8 & \\
0.8 & \times & 2 & = & 1.6 & & e_7 = 1 & \text{y} & f_7 = 0.6 & \\
0.6 & \times & 2 & = & 1.2 & & e_8 = 1 & \text{y} & f_8 = 0.2 &
\end{array}
$$

$$\Rightarrow\ 0.2_{10} = 0.0011001100110011001100110011\ldots_2$$

No se sigue multiplicando, debido a que se vuelve a obtener la misma parte fraccionaria.

Ya que se tiene el número en base 2, se normaliza desplazando el punto decimal inmediatamente a la derecha del primer bit significativo del número (es decir después del primer 1 de izquierda a derecha), obteniendo: $0.0011001100110011001100110011 = (1.1001100110011001100110011)_2(2^{-3})_{10}$.

Se escribe éste número de punto flotante en la forma normalizada general, usando la ecuación (1.7); obteniéndose:

$$V = (-1)^s (1.f)_2 (2^t)_{10} = (-1)^0 (1.10011001100110011001100110011)_2 (2^{-3})_{10}$$

El 2 se eleva a la -3, ya que el punto decimal se desplazó 3 veces a la derecha.

La mantisa es la parte a la derecha del punto decimal, y se colocan solamente 23 bits. Es decir $\mathbf{f} = 10011001100110011001100$.

El exponente $t = -3$, se necesita desplazarlo para obtener el valor de c, usando la ecuación (1.9). Para el formato IEEE de 32 bits, el desplazamiento es $b = 127$: $\Rightarrow \mathbf{c} = -3 + 127 = 124$. Éste valor se convierte en binario.

124	$\div$	2	=	62	ysobra	0	donde	$c_1 = 62$	y	$r_1 = 0$	
62	$\div$	2	=	31	ysobra	0		$c_2 = 31$	y	$r_2 = 0$	
31	$\div$	2	=	15	ysobra	1		$c_3 = 15$	y	$r_3 = 1$	
15	$\div$	2	=	7	ysobra	1		$c_4 = 7$	y	$r_4 = 1$	
7	$\div$	2	=	3	ysobra	1		$c_5 = 3$	y	$r_5 = 1$	
3	$\div$	2	=	1	ysobra	1		$c_6 = 1$	y	$r_6 = 1$	
1	$\div$	2	=	0	ysobra	1		$c_7 = 0$	y	$r_7 = 1$	$\uparrow$

$\Rightarrow\ 124_{10} = 1111100_2$

$\therefore\ 0.2$ se representa en una palabra de 32 bits usando el sistema de la IEEE, de la siguiente manera:

$$\underbrace{\boxed{0}}_{\text{Signo}}\ \underbrace{\boxed{0\,1\,1\,1\,1\,1\,0\,0}}_{\text{Exponente}}\ \underbrace{\boxed{1\,0\,0\,1\,1\,0\,0\,1\,1\,0\,0\,1\,1\,0\,0\,1\,1\,0\,0\,1\,1\,0\,0}}_{\text{Mantisa}}$$

Comprobación:

Éste número de punto flotante se puede escribir en la forma normalizada general, de la siguiente manera; (para obtener el valor de t, se utiliza la ecuación (1.8)):

$$\begin{aligned}
&1\ 01111100\ 10011001100110011001100 = \\
&(-1)^0(1.10011001100110011001100)_2(2^{124-127})_{10} = \\
&(1)(1.10011001100110011001100)_2(2^{-3})_{10} = \\
&0.00110011001100110011001100_2 = 0.199999988079_{10}
\end{aligned}$$

Lo anterior se obtuvo desplazando 3 posiciones a la izquierda el punto decimal de la mantisa, ya que $t = -3$. Después se escribe el número en base 10, multiplicando cada dígito de $0.00110011001100110011001100$ por 2 elevado a una potencia igual a la posición del dígito, tomando como posición inicial -1, a partir del punto; sumando después todos los términos que formarán el número en sistema decimal.

$$\begin{aligned}
0.00110011001100110011001100_2 &= \\
(1\times 2^{-3} + 1\times 2^{-4} + 1\times 2^{-7} + 1\times 2^{-8} + 1\times 2^{-11} + 1\times 2^{-12} + & \\
1\times 2^{-15} + 1\times 2^{-16} + 1\times 2^{-19} + 1\times 2^{-20} + 1\times 2^{-23} + 1\times 2^{-24}) &= \\
0.199999988079_{10} &\approx 0.2_{10}
\end{aligned}$$

Con ejemplo 1.2.8 se puede ejemplificar el error de redondeo, ya que el número 0.2, para su representación requiere de una infinidad de dígitos, y la memoria de la computadora en éste caso sólo puede almacenar 32 bits, por lo que al pasar el número almacenado en la computadora al sistema decimal no resulta exactamente 0.2, sino 0.199999988079.

■**Ejemplo 1.2.9** Representar el número 0.2 en una palabra de 64 bits usando el sistema de la IEEE.

Para poder representar 0.2 en un tamaño de palabra de 64 bits, se tienen que identificar cada una de sus partes (el signo s, el exponente t en forma sesgada (c) y la mantisa f).

Dado que es un número positivo, el signo se representa con $\mathbf{s} = 0$.

Primero se escribe el número en base 2, como no tiene parte entera, se obtiene sólo la parte fraccionaria. Se comienza multiplicando 0.2 por 2; la parte fraccionaria que resulte se multiplica nuevamente por 2. Se sigue el mismo procedimiento un número suficiente de veces ó hasta que la parte fraccionaria sea igual a cero. La parte entera que resulta de estas multiplicaciones es la que forma el número en sistema binario.

$$
\begin{array}{ccccccccccc}
0.2 & \times & 2 & = & 0.4 & \text{donde} & e_1 = 0 & \text{y} & f_1 = 0.4 & \downarrow \\
0.4 & \times & 2 & = & 0.8 & & e_2 = 0 & \text{y} & f_2 = 0.8 & \\
0.8 & \times & 2 & = & 1.6 & & e_3 = 1 & \text{y} & f_3 = 0.6 & \\
0.6 & \times & 2 & = & 1.2 & & e_4 = 1 & \text{y} & f_4 = 0.2 & \\
0.2 & \times & 2 & = & 0.4 & & e_5 = 0 & \text{y} & f_5 = 0.4 & \\
0.4 & \times & 2 & = & 0.8 & & e_6 = 0 & \text{y} & f_6 = 0.8 & \\
0.8 & \times & 2 & = & 1.6 & & e_7 = 1 & \text{y} & f_7 = 0.6 & \\
0.6 & \times & 2 & = & 1.2 & & e_8 = 1 & \text{y} & f_8 = 0.2 &
\end{array}
$$

$$\Rightarrow$$

$$0.2_{10} = 0.00110011001100110011001100110011001100110011001100110011\ldots_2$$

No se sigue multiplicando, debido a que se vuelve a obtener la misma parte fraccionaria.

Ya que se tiene el número en base 2, se normaliza desplazando el punto decimal inmediatamente a la derecha del primer bit significativo del número (es decir después del primer 1 de izquierda a derecha), obteniendo:

$$
\begin{array}{lr}
0.0011001100110011001100110011001100110011001100110011001100110011 & = \\
(1.1001100110011001100110011001100110011001100110011001100110011)_2(2^{-3})_{10} & =
\end{array}
$$

Se escribe éste número de punto flotante en la forma normalizada general, usando la ecuación (1.7); obteniéndose:

$$V = (-1)^s(1.f)_2(2^t)_{10} = (-1)^0(1.1001100110011001100110011001100110011001100110011001100110011)_2(2^{-3})_{10}$$

El 2 se eleva a la -3, ya que el punto decimal se desplazó 3 veces a la derecha.

La mantisa es la parte a la derecha del punto decimal, y se colocan solamente 52 bits. Es decir $\mathbf{f} = 1001100110011001100110011001100110011001100110011001$.

El exponente $t=-3$, se necesita desplazarlo para obtener el valor de c, usando la ecuación (1.9). Para el formato IEEE de 64 bits, el desplazamiento es $b=1023$: $\Rightarrow \mathbf{c}=-3+1023=1020$. Éste valor se convierte en binario.

1020	$\div$	2	=	510	ysobra	0	donde	$c_1=510$	y	$r_1=0$		
510	$\div$	2	=	255	ysobra	0		$c_2=255$	y	$r_2=0$		
255	$\div$	2	=	127	ysobra	1		$c_3=127$	y	$r_3=1$		
127	$\div$	2	=	63	ysobra	1		$c_4=63$	y	$r_4=1$		
63	$\div$	2	=	31	ysobra	1		$c_5=31$	y	$r_5=1$		
31	$\div$	2	=	15	ysobra	1		$c_6=15$	y	$r_6=1$		
15	$\div$	2	=	7	ysobra	1		$c_7=7$	y	$r_7=1$		
7	$\div$	2	=	3	ysobra	1		$c_8=3$	y	$r_8=1$		
3	$\div$	2	=	1	ysobra	1		$c_9=1$	y	$r_9=1$		
1	$\div$	2	=	0	ysobra	1		$c_{10}=0$	y	$r_{10}=1$	$\uparrow$	

$\Rightarrow\ 1020_{10}=1111111100_2$

$\therefore$ 0.2 se representa en una palabra de 64 bits usando el sistema de la IEEE, de la siguiente manera:

$$\underbrace{\boxed{0}}_{\text{Signo}}\ \underbrace{\boxed{01111111100}}_{\text{Exponente}}\ \underbrace{\boxed{1001100110011001100110011001100110011001100110011001}}_{\text{Mantisa}}$$

Comprobación:

Éste número de punto flotante se puede escribir en la forma normalizada general, de la siguiente manera; (para obtener el valor de t, se utiliza la ecuación (1.8)):

$$\begin{aligned}
&0\ \ 01111111100\ \ 1001100110011001100110011001100110011001100110011001 =\\
&(-1)^0(1.1001100110011001100110011001100110011001100110011001)_2(2^{1020-1023})_{10}=\\
&(1)(1.1001100110011001100110011001100110011001100110011001)_2(2^{-3})_{10}=\\
&0.0011001100110011001100110011001100110011001100110011001_2=\\
&0.200000000001_{10}
\end{aligned}$$

Lo anterior se obtuvo desplazando 3 posiciones a la izquierda el punto decimal de la mantisa, ya que $t=-3$. Después se escribe el número en base 10, multiplicando cada dígito de 0.0011001100110011001100110011001100110011001100110011001 por 2 elevado a una potencia igual a la posición del dígito, tomando como posición inicial -1, a partir del punto; sumando después todos los términos que formarán el número en sistema decimal.

$$\begin{aligned}
&0.0011001100110011001100110011001100110011001100110011001_2=\\
&(1\times2^{-3}+1\times2^{-4}+1\times2^{-7}+1\times2^{-8}+1\times2^{-11}+1\times2^{-12}+\\
&1\times2^{-15}+1\times2^{-16}+1\times2^{-19}+1\times2^{-20}+1\times2^{-23}+1\times2^{-24}+\\
&1\times2^{-27}+1\times2^{-28}+1\times2^{-31}+1\times2^{-32}+1\times2^{-35}+1\times2^{-36}+\\
&1\times2^{-39}+1\times2^{-40}+1\times2^{-43}+1\times2^{-44}+1\times2^{-47}+1\times2^{-48}+\\
&1\times2^{-51}+1\times2^{-52}+1\times2^{-55})=\\
&0.200000000001_{10}\approx0.2_{10}
\end{aligned}$$

Con los ejemplos 1.2.8 y 1.2.9 se puede observar que el error de redondeo disminuye con la precisión doble, ya que con ésta se pueden almacenar un número mayor de bits.

Algunos casos especiales para 32 bits

- Si $c = 255$ y f es no nulo, entonces $V = NaN$ ("Not a number")

$$0\ \ 11111111\ \ 00000100000000000000000 = \text{NaN}$$
$$1\ \ 11111111\ \ 00100010010101010101010 = \text{NaN}$$

- Si $c = 255$ y f es cero y s es 1, entonces $V = -Infinito$

$$1\ \ 11111111\ \ 00000000000000000000000 = -\text{Infinito}$$

- Si $c = 255$ y f es cero y s es 0, entonces $V = Infinito$

$$0\ \ 11111111\ \ 00000000000000000000000 = \text{Infinito}$$

- Si $c = 0$ y f es cero y s es 0, entonces $V = 0$

$$0\ \ 00000000\ \ 00000000000000000000000 = 0$$

- Si $c = 0$ y f es cero y s es 1, entonces $V = -0$

$$1\ \ 00000000\ \ 00000000000000000000000 = -0$$

- Si $0 < c < 255$ entonces $V = (-1)^s(1.f)_2(2^{c-127})_{10}$

$$\begin{aligned} 0\ \ 10000000\ \ 00000000000000000000000 &= \\ (-1)^0(1.00000000000000000000000)_2(2^{128-127})_{10} &= \\ (1)(1.00000000000000000000000)_2(2^1)_{10} &= \\ 10.0000000000000000000000_2 &= 2_{10} \end{aligned}$$

$$\begin{aligned} 0\ \ 10000001\ \ 10100000000000000000000 &= \\ (-1)^0(1.10100000000000000000000)_2(2^{129-127})_{10} &= \\ (1)(1.10100000000000000000000)_2(2^2)_{10} &= \\ 110.100000000000000000000_2 &= 6.5_{10} \end{aligned}$$

$$\begin{aligned} 1\ \ 10000001\ \ 10100000000000000000000 &= \\ (-1)^1(1.10100000000000000000000)_2(2^{129-127})_{10} &= \\ (-1)(1.10100000000000000000000)_2(2^2)_{10} &= \\ -110.100000000000000000000_2 &= -6.5_{10} \end{aligned}$$

$$\begin{aligned} 0\ \ 00000001\ \ 00000000000000000000000 &= \\ (-1)^0(1.00000000000000000000000)_2(2^{1-127})_{10} &= \\ (1)(1.00000000000000000000000)_2(2^{-126})_{10} &= (1 * 10^{-126})_{10} \end{aligned}$$

- Si $c = 0$ y f es no nulo, entonces $V = (-1)^s(0.f)_2(2^{-126})_{10}$

$$\begin{aligned} 0\ \ 00000000\ \ 10000000000000000000000 &= \\ (-1)^0(0.10000000000000000000000)_2(2^{-126})_{10} &= \\ (1)(0.10000000000000000000000)_2(2^{-126})_{10} &= (1 * 10^{-127})_{10} \end{aligned}$$

$$\begin{aligned} 0\ \ 00000000\ \ 00000000000000000000001 &= \\ (-1)^0(0.00000000000000000000001)_2(2^{-126})_{10} &= \\ (1)(0.00000000000000000000001)_2(2^{-126})_{10} &= (1 * 10^{-149})_{10} \end{aligned}$$

Algunos casos especiales para 64 bits

- Si $c = 2047$ y f es no nulo, entonces $V = NaN$ ("Not a number")

  0 11111111111 0000010000000000000000000000100000000000000000000011=NaN
  1 11111111111 0010001001010101010101000100010010101010101010101010=NaN

- Si $c = 2047$ y f es cero y s es 1, entonces $V = -Infinito$

  1 11111111111 0000000000000000000000000000000000000000000000000000=−Infinito

- Si $c = 2047$ y f es cero y s es 0, entonces $V = Infinito$

  0 11111111111 0000000000000000000000000000000000000000000000000000=Infinito

- Si $c = 0$ y f es cero y s es 0, entonces $V = 0$

  0 00000000000 0000000000000000000000000000000000000000000000000000=0

- Si $c = 0$ y f es cero y s es 1, entonces $V = -0$

  1 00000000000 0000000000000000000000000000000000000000000000000000=−0

- Si $0 < c < 2047$ entonces $V = (-1)^s(1.f)_2(2^{c-1023})_{10}$

$$
\begin{aligned}
&0\ 10000000000\ 0000000000000000000000000000000000000000000000000000 =\\
&(-1)^0(1.0000000000000000000000000000000000000000000000000000)_2(2^{1024-1023})_{10}=\\
&(1)(1.0000000000000000000000000000000000000000000000000000)_2(2^1)_{10}=\\
&10.000000000000000000000000000000000000000000000000000_2=2_{10}
\end{aligned}
$$

- Si $c = 0$ y f es no nulo, entonces $V = (-1)^s(0.f)_2(2^{-1022})_{10}$

$$
\begin{aligned}
&0\ 00000000000\ 1000000000000000000000000000000000000000000000000000 =\\
&(-1)^0(0.1000000000000000000000000000000000000000000000000000)_2(2^{-1022})_{10}=\\
&(1)(0.1000000000000000000000000000000000000000000000000000)_2(2^{-1022})_{10}=\\
&(1 * 10^{-1023})_{10}
\end{aligned}
$$

$$
\begin{aligned}
&0\ 00000000000\ 0000000000000000000000000000000000000000000000000001 =\\
&(-1)^0(0.0000000000000000000000000000000000000000000000000001)_2(2^{-1022})_{10}=\\
&(1)(0.0000000000000000000000000000000000000000000000000001)_2(2^{-1022})_{10}=\\
&(1 * 10^{-1074})_{10}
\end{aligned}
$$

Error de truncamiento

Los ***errores de truncamiento*** corresponden al caso de calcular por ejemplo $\operatorname{sen}(x)$, $\cos(x)$, e^x, etc., usando series de potencias. Estas series tienen un número infinito de elementos, pero evidentemente para calcular su valor sólo se puede considerar un número finito de ellos. Es el caso de constantes que no pueden representarse exactamente, como por ejemplo, los números pi y e, o fracciones como $\frac{1}{3}$, $\frac{2}{3}$, etc. Estos errores también se presentan cuando se hacen transformaciones del sistema decimal al binario y viceversa.

Para entender más a que se refiere el error de truncamiento es necesario conocer el desarrollo de funciones en series de potencias; ya que en las matemáticas con frecuencia las funciones se representan mediante series infinitas.

Si se toman en cuenta más términos de la serie (es decir mientras menos términos se trunquen), casi siempre será mejor la aproximación.

Definición 1.2.3 Serie de Taylor. Es una serie infinita de potencias, que representa de manera exacta a una función $f(x)$ alrededor de un punto $x=c$; se define como:

$$f(x)=f(c)+f'(c)(x-c)+\frac{f''(c)(x-c)^2}{2!}+\frac{f'''(c)(x-c)^3}{3!}+\ldots+\frac{f^{(n)}(c)(x-c)^n}{n!} \quad (1.10)$$

Si $c=0$ se trata de la serie de Maclaurin.

■**Ejemplo 1.2.10** Obtener los primeros 6 términos de la serie de Taylor para $f(x)=e^x$ alrededor de 0.

Se comienza calculando las 5 primeras derivadas de la función, obteniéndose:

$$f(x)=e^x$$

$$f'(x)=e^x$$

$$f''(x)=e^x$$

$$f'''(x)=e^x$$

$$f^{IV}(x)=e^x$$

$$f^{V}(x)=e^x$$

Después se evalúan en torno a 0:

$$f(0)=e^0=1$$

$$f'(0)=e^0=1$$

$$f''(0)=e^0=1$$

$$f'''(0)=e^0=1$$

$$f^{IV}(0)=e^0=1$$

$$f^{V}(0)=e^0=1$$

Finalmente se sustituyen los resultados anteriores en la ecuación (1.10):

$$f(x) = f(0) + f'(0)(x-0) + \frac{f''(0)(x-0)^2}{2!} + \frac{f'''(0)(x-0)^3}{3!} + \frac{f^{IV}(0)(x-0)^4}{4!} + \frac{f^{V}(0)(x-0)^5}{5!} + \dots$$

$$f(x) = 1 + 1(x-0) + \frac{1(x-0)^2}{2!} + \frac{1(x-0)^3}{3!} + \frac{1(x-0)^4}{4!} + \frac{1(x-0)^5}{5!} + \dots$$

$$f(x) = 1 + 1x + \frac{1x^2}{2!} + \frac{1x^3}{3!} + \frac{1x^4}{4!} + \frac{1x^5}{5!} + \dots$$

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \dots$$

$\therefore$ Los primeros seis términos de la serie de Taylor para e^x son:

$$e^x \approx 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} \qquad (1.11)$$

■**Ejemplo 1.2.11** Si se sabe que la raíz exacta de $e^{(0.5)} = 1.6487212707$, calcular el error relativo porcentual que se esta cometiendo si se trunca su representación en series de potencias para $1, 2, 3, \dots, 6$ términos. Utilizar la ecuación (1.11); con 6 decimales.

Se comienza calculando $e^{(0.5)}$ trucando la serie de potencias (1.11) a un término:

$$e^{(0.5)} = 1$$

Después se calcula $e^{(0.5)}$ trucando la serie de potencias (1.11) a dos términos:

$$e^{(0.5)} = 1+(0.5)=1.5$$

Después se calcula $e^{(0.5)}$ trucando la serie de potencias (1.11) a tres términos:

$$e^{(0.5)} = 1 + (0.5) + \frac{(0.5)^2}{2!} = 1.625$$

Para obtener $e^{(0.5)}$ trucando la serie de potencias (1.11) a cuatro, cinco y seis términos, se repite el proceso anterior. Los resultados se muestran en el cuadro 1.1

Términos	Resultados	$\varepsilon_r(\%)$
1	1.000000	——
2	1.500000	33.333333
3	1.625000	7.692308
4	1.645833	1.265803
5	1.648437	0.157968
6	1.648697	0.015770

Cuadro 1.1: Error de truncamiento

A continuación se indica como calcular el error relativo porcentual $\left|\frac{x_n - x_{n-1}}{x_n}\right| 100\%$.

Para la primera iteración el error relativo porcentual no se calcula, porque no se cuenta con una aproximación inicial.

Para la segunda iteración el error relativo porcentual se calcula de la siguiente manera:

$$\left|\frac{x_2 - x_1}{x_2}\right| 100\% = \left|\frac{1.5-1}{1.5}\right| 100\% = 33.333333\%$$

Para la tercera iteración el error relativo porcentual se calcula:

$$\left|\frac{x_3 - x_2}{x_3}\right| 100\% = \left|\frac{1.625-1.5}{1.625}\right| 100\% = 7.692308\%$$

Para obtener el error relativo porcentual de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 1.1

$\therefore$ Se puede observar que mientras más términos de la serie se tomen; casi siempre será mejor la aproximación a la raíz.

■**Ejemplo 1.2.12** Obtener los primeros 5 términos de la serie de Taylor para $f(x) = \cos(x)$ alrededor de 0.

Se comienza calculando las 8 primeras derivadas de la función, obteniéndose:

$$f(x) = \cos(x)$$

$$f'(x) = -\text{sen(x)}$$

$$f''(x) = -\cos(x)$$

$$f'''(x) = \text{sen(x)}$$

$$f^{IV}(x) = \cos(x)$$

$$f^{V}(x) = -\text{sen(x)}$$

$$f^{VI}(x) = -\cos(x)$$

$$f^{VII}(x) = \text{sen(x)}$$

$$f^{VIII}(x) = \cos(x)$$

Después se evalúan en torno a 0:

$$f(0) = \cos(0) = 1$$

$$f'(0) = -\text{sen}(0)=0$$

$$f''(0) = -\cos(0) = -1$$

$$f'''(0) = \text{sen}(0)=0$$

$$f^{IV}(0) = \cos(0) = 1$$

$$f^{V}(0) = -\text{sen}(0)=0$$

$$f^{VI}(0) = -\cos(0) = -1$$

$$f^{VII}(0) = \text{sen}(0)=0$$

$$f^{VIII}(0) = \cos(0) = 1$$

Finalmente se sustituyen los resultados anteriores en la ecuación (1.10):

$$f(x) = f(0) + f'(0)(x-0) + \frac{f''(0)(x-0)^2}{2!} + \frac{f'''(0)(x-0)^3}{3!} + \frac{f^{IV}(0)(x-0)^4}{4!} + \frac{f^{V}(0)(x-0)^5}{5!} + \frac{f^{VI}(0)(x-0)^6}{6!} + \frac{f^{VII}(0)(x-0)^7}{7!} + + \frac{f^{VIII}(0)(x-0)^8}{8!} \dots$$

$$f(x) = 1 + 0(x-0) + \frac{-1(x-0)^2}{2!} + \frac{0(x-0)^3}{3!} + \frac{1(x-0)^4}{4!} + \frac{0(x-0)^5}{5!} + \frac{-1(x-0)^6}{6!} + \frac{0(x-0)^7}{7!} + \frac{1(x-0)^8}{8!} \dots$$

$$f(x) = 1 - \frac{1x^2}{2!} + \frac{1x^4}{4!} - \frac{1x^6}{6!} + \frac{1x^8}{8!} \dots$$

$$f(x) = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} + \dots$$

$\therefore$ Los primeros cinco términos de la serie de Taylor para $\cos(x)$ son:

$$\cos(x) \approx 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} \tag{1.12}$$

■**Ejemplo 1.2.13** Si se sabe que la raíz exacta de $\cos(0.01) = 0.999950000417$, calcular el error relativo porcentual que se esta cometiendo si se trunca su representación en series de potencias para $1, 2, 3, \dots, 5$ términos. Utilizar la ecuación (1.12); con 12 decimales.

Se comienza calculando $\cos(0.01)$ trucando la serie de potencias (1.12) a un término:

$$\cos(0.01) = 1$$

Después se calcula $\cos(0.01)$ trucando la serie de potencias (1.12) a dos términos:

$$\cos(0.01) = 1 - \tfrac{(0.01)^2}{2!} = 0.999950000000$$

Después se calcula $\cos(0.01)$ trucando la serie de potencias (1.12) a tres términos:

$$\cos(0.01) = 1 - \tfrac{(0.01)^2}{2!} + \tfrac{(0.01)^4}{4!} = 0.999950000417$$

Para obtener $\cos(0.01)$ trucando la serie de potencias (1.12) a cuatro y cinco términos, se repite el proceso anterior. Los resultados se muestran en el cuadro 1.2

Términos	Resultados	$\varepsilon_r(\%)$
1	1.000000000000	——
2	0.999950000000	0.005000250013
3	0.999950000417	0.000000041702
4	0.999950000417	0.000000000000
5	0.999950000417	0.000000000000

Cuadro 1.2: Error de truncamiento

A continuación se indica como calcular el error relativo porcentual $\left|\tfrac{x_n - x_{n-1}}{x_n}\right| 100\%$.

Para la primera iteración el error relativo porcentual no se calcula, porque no se cuenta con una aproximación inicial.

Para la segunda iteración el error relativo porcentual se calcula de la siguiente manera:

$$\left|\frac{x_2 - x_1}{x_2}\right| 100\% = \left|\frac{0.999950000000 - 1}{0.999950000000}\right| 100\% = 0.005000250013\%$$

Para la tercera iteración el error relativo porcentual se calcula:

$$\left|\frac{x_3 - x_2}{x_3}\right| 100\% = \left|\frac{0.999950000417 - 0.999950000000}{0.999950000417}\right| 100\% = 0.000000041702\%$$

Para obtener el error relativo porcentual de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 1.

$\therefore$ Se puede observar en este caso, que con los tres primeros términos de la serie se obtiene la solución exacta.

1.3 Propagación del error en distintas operaciones aritméticas

Cuando se resuelve un problema matemático por métodos numéricos y aunque las operaciones se lleven a cabo exactamente, obtenemos una aproximación a la solución del problema; por lo que es importante tratar de conocer el efecto que sobre el resultado final del problema tiene cada una de las operaciones realizadas.

Para estudiar como se propaga el error, se necesita saber cual es el efecto que cada una de las operaciones básicas tiene sobre el error final cuando se aplican sobre dos números $x_1 \pm \varepsilon(x_1)$ y $x_2 \pm \varepsilon(x_2)$.

■**Ejemplo 1.3.1** Suponer que una calculadora o computadora opera con 5 decimales en su sistema numérico de punto flotante. Realizar las cuatro operaciones aritméticas básicas $(+,-,\times,\div)$ utilizando primero todos los decimales y después sólo 5, con los dos siguientes números de máquina: $x = 0.31426 \times 10^3$ y $y = 0.92577 \times 10^5$; y calcular los errores relativos verdaderos para cada operación.

Se comienza realizando las cuatro operaciones aritméticas básicas $(+,-,\times,\div)$ utilizando todos los decimales:

Cuando se suman o restan dos números de punto flotante, los dígitos del número de menor exponente deben desplazarse para alinear los puntos decimales con el número de mayor exponente, en este caso el de menor exponente es x, por lo que se desplaza a manera de que el exponente sea igual que el de y: $x = 0.0031426 \times 10^5$

Para la multiplicación y división primero se normalizan [1] los números, en este caso no es necesario que los exponentes sean iguales; en el caso de la multiplicación se suman los exponentes y en el caso de la división se restan. Si el resultado no esta normalizado, se normaliza.

$$x + y = 0.0031426 \times 10^5 + 0.92577 \times 10^5 = 0.9289126 \times 10^5$$
$$x - y = 0.0031426 \times 10^5 - 0.92577 \times 10^5 = -0.9226274 \times 10^5$$
$$x \times y = 0.31426 \times 10^3 \times 0.92577 \times 10^5 = 0.2909324802 \times 10^8$$
$$x \div y = 0.31426 \times 10^3 \div 0.92577 \times 10^5 = 0.339457964721 \times 10^{-2}$$

La calculadora o computadora con 5 decimales los almacenará en forma redondeada de la siguiente manera:

$$x + y = 0.00314 \times 10^5 + 0.92577 \times 10^5 = 0.92891 \times 10^5$$
$$x - y = 0.00314 \times 10^5 - 0.92577 \times 10^5 = -0.92263 \times 10^5$$

---

[1] En el sistema decimal, cualquier número real puede expresarse mediante la denominada notación científica normalizada. Para expresar un número en notación científica normalizada se multiplica o divide por 10 tantas veces sea necesario para que todos los dígitos aparezcan a la derecha del punto decimal y de modo que el primer dígito después del punto no sea cero.

$$x\times y = 0.31426\times10^{3}\times0.92577\times10^{5} = 0.29093\times10^{8}$$
$$x\div y = 0.31426\times10^{3}\div0.92577\times10^{5} = 0.33946\times10^{-2}$$

n	x	x_n	$\varepsilon_{r,v}$
1	0.9289126×10^{5}	0.92891×10^{5}	2.8×10^{-6}
2	-0.9226274×10^{5}	-0.92263×10^{5}	2.8×10^{-6}
3	0.2909324802×10^{8}	0.29093×10^{8}	8.5×10^{-6}
4	$0.339457964721\times10^{-2}$	0.33946×10^{-2}	6.0×10^{-6}

Cuadro 1.3 Error relativo verdadero

A continuación se calcula el error relativo verdadero $\left|\frac{x-x_n}{x}\right|$.

Para la suma el error relativo verdadero se calcula de la siguiente manera:

$$\left|\frac{x-x_1}{x}\right| = \left|\frac{0.9289126\times10^{5}-0.92891\times10^{5}}{0.9289126\times10^{5}}\right| = 0.0000028 = 2.8\times10^{-6}$$

Para la resta el error relativo verdadero se calcula:

$$\left|\frac{x-x_2}{x}\right| = \left|\frac{-0.9226274\times10^{5}-(-0.92263\times10^{5})}{-0.9226274\times10^{5}}\right| = 0.0000028 = 2.8\times10^{-6}$$

Para obtener el error relativo verdadero de la multiplicación y de la división se repite el proceso anterior. Los errores se muestran en el cuadro 1.3.

$\therefore$ Se puede observar que al realizar operaciones básicas con dos números que han sido redondeados el resultado final también tendrá un error; es decir éste se propaga.

Cuando el problema consiste en calcular el resultado $y = f(x)$, se tiene la siguiente fórmula aproximada de propagación del error:

$$\varepsilon(y) = |f'(x)|\,\varepsilon(x) \tag{1.13}$$

En el caso más general, en que una función depende de más de una variable $y = f(x_1, x_2, \ldots, x_n)$, la fórmula aproximada de propagación del error es:

$$\varepsilon(y) = \sum_{i=1}^{n}\left|\frac{\partial}{\partial x_i} f(x_1, x_2, \ldots, x_n)\right|\varepsilon(x_i) \tag{1.14}$$

■**Ejemplo 1.3.2** Dados dos valores $x_1 = 2.0\pm0.1$ y $x_2 = 3.0\pm0.2$, estimar el error resultante en la función $y = x_1x_2^2$

El error cometido, de acuerdo con la ecuación (1.14), se puede calcular mediante:

$$\varepsilon(y) = \left|\tfrac{\partial}{\partial x_1}(x_1x_2^2)\right|\varepsilon(x_1) + \left|\tfrac{\partial}{\partial x_2}(x_1x_2^2)\right|\varepsilon(x_2)$$

$$\varepsilon(y) = \left|x_2^2\right|\varepsilon(x_1) + \left|2x_1x_2\right|\varepsilon(x_2)$$

Sustituyendo los valores de x_1, $\varepsilon(x_1)$, x_2 y $\varepsilon(x_2)$ se obtiene:

$$y = (2.0)(3.0)^2 = 18$$

$$\varepsilon(y) = \left|3.0^2\right|(0.1) + \left|2(2.0)(3.0)\right|(0.2) = 3.3$$

$\therefore\ y = 18 \pm 3.3$

ó que el valor verdadero se encuentra entre 14.7 y 21.3.

■**Ejemplo 1.3.3** Dado el siguiente sistema de ecuaciones lineales, estimar el error resultante del producto xy.

$$x + ay = 5$$

$$bx + 2y = d \tag{1.15}$$

donde $a = 1.000 \pm 0.002$; $b = \frac{1}{a}$ y $d = b - a$

Se comienza resolviendo el sistema de ecuaciones lineales por reducción. Se despeja x de la primera ecuación de (1) y se sustituye en la segunda para obtener el valor de y; obteniéndose:

$$x = 5 - ay$$

$$b(5 - ay) + 2y = \mathrm{d}$$

$$5b - aby + 2y = \mathrm{d}$$

$$-aby + 2y = \mathrm{d{-}5b}$$

$$(2 - ab)y = \mathrm{d{-}5b}$$

$$\Rightarrow y = \frac{d-5b}{2-ab}$$

Después se sustituye y en x:

$$x = 5 - a\left(\frac{d-5b}{2-ab}\right)$$

$$x = 5 + \frac{-ad+5ab}{2-ab}$$

$$x = \frac{5(2-ab)-ad+5ab}{2-ab}$$

$$x = \frac{10-5ab-ad+5ab}{2-ab}$$

$$\Rightarrow x = \frac{10-ad}{2-ab}$$

Ecuaciones que conducen a la siguiente expresión para el producto:

$$xy = \frac{\overbrace{(10-ad)}^{A}\overbrace{(d-5b)}^{B}}{\underbrace{(2-ab)^2}_{C}} \tag{1.16}$$

Se resolverá el problema por dos métodos.

El primer método consistirá en calcular el error asociado a cada una de las variables y los términos de la expresión anterior con las ecuaciones (1.13) y (1.14), obteniéndose:

$$a = 1.000 \qquad \varepsilon(a) = 0.002$$

$$b = \frac{1}{a} = \frac{1}{1} = 1 \qquad \varepsilon(b) = \left|\frac{d}{da}\left(\frac{1}{a}\right)\right|\varepsilon(a) = |-a^{-2}|\varepsilon(a) = \frac{\varepsilon(a)}{a^2} = \frac{\varepsilon(a)}{1} = \varepsilon(a)$$

$$\begin{aligned} d &= b - a & \varepsilon(d) &= \left|\frac{\partial}{\partial b}(b-a)\right|\varepsilon(b) + \left|\frac{\partial}{\partial a}(b-a)\right|\varepsilon(a) \\ &= 1 - 1 = 0 & &= |1|\varepsilon(b) + |-1|\varepsilon(a) = \varepsilon(a) + \varepsilon(a) = 2\varepsilon(a) \end{aligned}$$

$$\begin{aligned} A &= (10 - ad) & \varepsilon(A) &= \left|\frac{\partial}{\partial a}(10-ad)\right|\varepsilon(a) + \left|\frac{\partial}{\partial d}(10-ad)\right|\varepsilon(d) \\ &= 10 - (1)(0) = 10 & &= |-d|\varepsilon(a) + |-a|\varepsilon(d) = 0\varepsilon(a) + \varepsilon(d) = \varepsilon(d) = 2\varepsilon(a) \end{aligned}$$

$$\begin{aligned} B &= (d - 5b) & \varepsilon(B) &= \left|\frac{\partial}{\partial d}(d-5b)\right|\varepsilon(d) + \left|\frac{\partial}{\partial b}(d-5b)\right|\varepsilon(b) \\ &= 0 - 5(1) = -5 & &= |1|\varepsilon(d) + |-5|\varepsilon(b) = \varepsilon(d) + 5\varepsilon(a) \\ & & &= 2\varepsilon(a) + 5\varepsilon(a) = 7\varepsilon(a) \end{aligned}$$

$$\begin{aligned} C &= (2 - ab)^2 & \varepsilon(C) &= \left|\frac{\partial}{\partial a}(4-4ab+a^2b^2)\right|\varepsilon(a) + \left|\frac{\partial}{\partial b}(4-4ab+a^2b^2)\right|\varepsilon(b) \\ &= (2-1)^2 = 1 & &= |(4b+2ab^2)|\varepsilon(a) + |(4a+2a^2b)|\varepsilon(b) \\ & & &= |-4(1)+2(1)(1)^2|\varepsilon(a) + |-4(1)+2(1)^2(1)|\varepsilon(b) = \\ & & &= 2\varepsilon(a) + 2\varepsilon(b) = 2\varepsilon(a) + 2\varepsilon(a) = 4\varepsilon(a) \end{aligned}$$

$$\begin{aligned} AB &= (10)(-5) & \varepsilon(AB) &= \left|\frac{\partial}{\partial A}(AB)\right|\varepsilon(A) + \left|\frac{\partial}{\partial B}(AB)\right|\varepsilon(B) \\ &= -50 & &= |B|\varepsilon(A) + |A|\varepsilon(B) = |-5|\varepsilon(A) + |10|\varepsilon(B) \\ & & &= 5(2\varepsilon(a)) + 10(7\varepsilon(a)) = \\ & & &= 10\varepsilon(a) + 70\varepsilon(a) = 80\varepsilon(a) \end{aligned}$$

$$\begin{aligned} xy &= \frac{AB}{C} & \varepsilon(xy) &= \left|\frac{\partial}{\partial A}\left(\frac{AB}{C}\right)\right|\varepsilon(A) + \left|\frac{\partial}{\partial B}\left(\frac{AB}{C}\right)\right|\varepsilon(B) + \left|\frac{\partial}{\partial C}\left(\frac{AB}{C}\right)\right|\varepsilon(C) \\ &= \frac{(10)(-5)}{1} = -50 & &= \left|\frac{B}{C}\right|\varepsilon(A) + \left|\frac{A}{C}\right|\varepsilon(B) + \left|-\frac{AB}{C^2}\right|\varepsilon(C) = \\ & & &= \frac{|CB|\varepsilon(A) + |CA|\varepsilon(B) + |-AB|\varepsilon(C)}{|C^2|} \\ & & &= \frac{|(1)(-5)|2\varepsilon(a) + |(1)(10)|7\varepsilon(a) + |-(10)(-5)|4\varepsilon(a)}{|1^2|} \\ & & &= \frac{10\varepsilon(a) + 70\varepsilon(a) + 200\varepsilon(a)}{1} = 280\varepsilon(a) \end{aligned}$$

Sustituyendo el valor de $\varepsilon(a)$ se obtiene:

$$xy = -50$$

$$\varepsilon(xy) = 280\varepsilon(a) = 280(0.002) = 0.56$$

$$\Rightarrow \; xy = -50 \pm 0.56$$

ó que el valor verdadero se encuentra entre -50.56 y -49.44.

El segundo método consiste en sustituir en la ecuación (1.16) los valores de b $(b = \frac{1}{a})$ y d $(d = b - a)$ por sus correspondientes expresiones en función de a, obteniéndose:

$$xy = \frac{(10 - ad)(d - 5b)}{(2 - ab)^2}$$
$$xy = \frac{(10 - a(b-a))\left((b-a) - 5\left(\frac{1}{a}\right)\right)}{\left(2 - a\left(\frac{1}{a}\right)\right)^2}$$
$$xy = \frac{\left(10 - a\left(\frac{1}{a} - a\right)\right)\left(\left(\frac{1}{a} - a\right) - 5\left(\frac{1}{a}\right)\right)}{\left(2 - a\left(\frac{1}{a}\right)\right)^2}$$
$$xy = \frac{\left(10 - \frac{a}{a} + a^2\right)\left(\frac{1}{a} - a - \frac{5}{a}\right)}{\left(2 - \frac{a}{a}\right)^2}$$
$$xy = \frac{\left(10 - 1 + a^2\right)\left(\frac{1}{a} - a - \frac{5}{a}\right)}{(2-1)^2}$$
$$xy = \frac{\left(9 + a^2\right)\left(\frac{1 - a^2 - 5}{a}\right)}{(1)^2}$$
$$xy = \left(9 + a^2\right)\left(\frac{-4 - a^2}{a}\right)$$
$$xy = \frac{-36 - 13a^2 - a^4}{a}$$
$$xy = -\frac{36}{a} - 13a - a^3$$

El error cometido, de acuerdo con la ecuación (1.13), se puede calcular mediante:

$$\varepsilon(xy) = \left|\frac{d}{da}\left(-\frac{36}{a} - 13a - a^3\right)\right|\varepsilon(a)$$
$$\varepsilon(xy) = \left|\frac{36}{a^2} - 13 - 3a^2\right|\varepsilon(a)$$

Sustituyendo los valores de a y $\varepsilon(a)$ se obtiene:

$$xy = -\frac{36}{a} - 13a - a^3 = -\frac{36}{1} - 13(1) - (1)^3 = -50$$
$$\varepsilon(xy) = \left|\frac{36}{(1)^2} - 13 - 3(1)^2\right|(0.002) = 0.04$$

$$\Rightarrow \; xy = -50 \pm 0.04$$

ó que el valor verdadero se encuentra entre -50.04 y -49.96.

$\therefore$ Aunque los dos resultados sean correctos, con el segundo método el error es menor, debido a que se han eliminado operaciones intermedias, permitiendo que algunos errores se cancelen. En general, cuanto menor sea el número de pasos intermedios que se efectúen para alcanzar la solución, menor será el error cometido.

Algunas estrategias para minimizar el error de redondeo

1. **Estrategia de la mantisa completa.**

Para minimizar el error de redondeo, se deben de introducir los valores de entrada con tantos dígitos significativos como puedan almacenarse en el dispositivo (calculadora o computadora). Por ejemplo: en un dispositivo de 7 dígitos significativos, pi se debe proporcionar como 3.141593 (no como 3.1416).

2. Estrategia de respuesta final.

A menudo, los valores de entrada de un cálculo son conocidos con menos dígitos exactos que la precisión del dispositivo de cálculo, se sugiere redondear la respuesta final a una exactitud conocida. Por ejemplo: si un cálculo bien condicionado resulta 23.3876 y el dato de entrada menos exacto se conoce solo con exactitud de tres dígitos significativos, entonces la respuesta debe anotarse como 23.4 es decir, redondeada a 3 dígitos. Es aconsejable una nota de que incluso 23.4 podría ser inexacto si se sabe que el cálculo estuvo mal condicionado.

3. Estrategia de operaciones mínimas.

Para ayudar a minimizar el error de redondeo, evaluar las expresiones matemáticas en una forma que requiera el menor número de operaciones aritméticas. Ejemplo:

$$y^6 = \mathrm{y} \times \mathrm{y} \times \mathrm{y} \times \mathrm{y} \times \mathrm{y} \times \mathrm{y}$$

$$y^2 = \mathrm{y} \times \mathrm{y}$$

$$y^4 = y^2 \times y^2$$

$$y^6 = y^4 \times y^2$$

4. Estrategia de la multiplicación anidada.

Evaluar los polinomios de forma anidada. Ejemplo:

$$f(x) = 2x^4 - 19x^3 + 56.98x^2 - 56.834x + 5.1324$$

su equivalente en forma anidada es:

$$f(x) = (((2\mathrm{x}-19)\mathrm{x}+56.98)\mathrm{x}-56.834)\mathrm{x}+5.1324$$

Ambas formas requieren cuatro adiciones/sustracciones y cuatro multiplicaciones por una potencia de x, sin embargo, la evaluación en la primera ecuación requiere el cálculo adicional de x^4, x^3 y x^2.

1.4 Orden de convergencia

Al aplicar los métodos numéricos se siguen algoritmos, que se definen como secuencias de operaciones algebraicas y lógicas que producen una aproximación a la solución del problema. Generalmente, se dispone de varios algoritmos para resolver un problema en particular; uno de los criterios de selección es la estabilidad del algoritmo;

esto es que a pequeños errores de los valores manejados se obtengan pequeños errores en los resultados finales.

Definición 1.4.1 Supóngase que un error ε se introduce en algún paso en los cálculos y que el error de propagación de n operaciones subsiguientes se denote por ε_n. Si:

1. $|\varepsilon_n| \approx C_n$, donde C es una constante independiente de n; se dice entonces que la ***propagación del error es lineal***.

2. $|\varepsilon_n| \approx k^n$, para $k > 1$; se dice entonces que la ***propagación del error es exponencial***.

La propagación lineal de los errores suele ser inevitable; cuando C y ε son pequeños, los resultados finales normalmente son aceptables. Por otro lado la propagación exponencial debe evitarse, ya que el término k^n crece con rapidez para valores relativamente pequeños de n. Esto conduce a resultados finales muy poco exactos, sea cual sea el tamaño de ε. Como consecuencia, se dice que un algoritmo con crecimiento lineal del error es estable, mientras que un algoritmo con una propagación exponencial es inestable. (Ver figura 1.1).

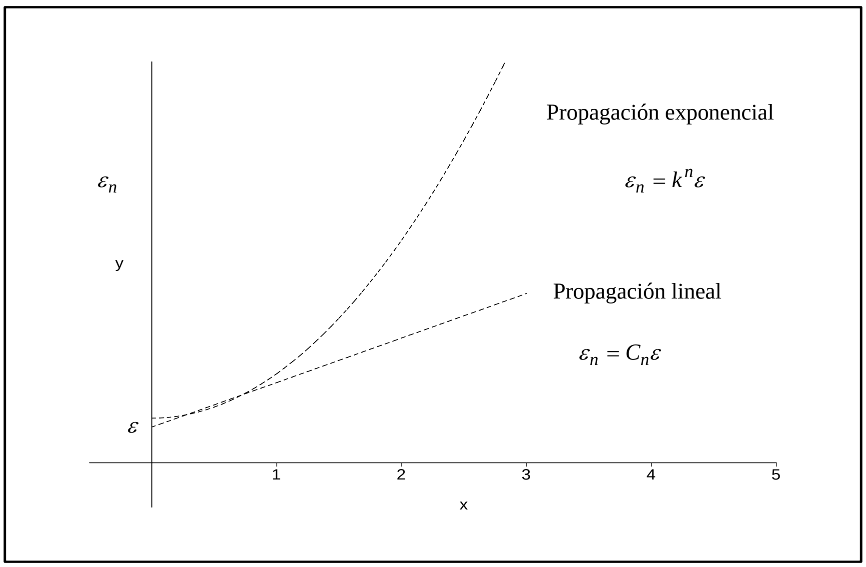


Figura 1.1 Propagación lineal y propagación exponencial de errores.

1.5 Herramientas disponibles para el análisis numérico. (Matlab, Maple, Mathematica, etc.)

Algunas herramientas disponibles para el análisis matemático que pueden facilitar enormemente el trabajo y ahorrar una gran cantidad de tiempo son los programas como **Matlab**, **Maple** y **Mathematica**. Lo anterior no significa que estos programas vayan a desplazar la manera de aprender las matemáticas de la forma tradicional; simplemente se deben ver estos programas como una herramienta que será de gran ayuda.

Matlab es un lenguaje de programación de alto nivel, con un enfoque directo hacia la computación científica. Más concretamente, es un programa de cálculo numérico con instrucciones dirigidas a la resolución de problemas científicos. Desde este punto de vista puede ser considerado entonces como una gran calculadora científica programable y muy potente. El nombre de Matlab proviene de las palabras inglesas *Matrix Laboratory* (laboratorio de matrices), que evidentemente dan una idea de la utilidad primordial de dicho programa, orientado a matrices y vectores.

Maple es un programa matemático de propósito general capaz de realizar cálculos simbólicos, algebraicos y de álgebra computacional. Fue originalmente desarrollado en 1981 por el Grupo de Cálculo Simbólico en la Universidad de Waterloo en Waterloo, Ontario, Canadá. Su nombre proviene de *MAthematical PLEasure* (Placer Matemático).

Mathematica es el primer programa para la computación y visualización numérica, simbólica y gráfica; incluye un amplio rango de funciones matemáticas, soporta operaciones de álgebra lineal, realiza todo tipo de operaciones algebraicas, opera con funciones, derivadas e integrales, entre otras muchas cosas.

En los cuadros 1.4, 1.5, 1.6, 1.7, 1.8 , 1.9 y 1.10, se muestran algunas de las funciones que se pueden realizar en estos programas; que serán de gran utilidad, para los temas que se verán más adelante.

En la interfase de Matlab, en la parte superior de la hoja de trabajo en blanco aparece un símbolo con el siguiente aspecto EDU>>. Esta cadena de caracteres es el indicador de petición de órdenes de Matlab en la edición de estudiante de Matlab. Si se teclea al final de una orden ";" ésta se ejecuta pero el resultado no se visualiza por pantalla. Para que se ejecute una operación, se debe pulsar la tecla *ENTER*.

En la interfase de Maple, en la parte superior de la hoja de trabajo en blanco aparece un símbolo con el siguiente aspecto [>. Este símbolo es el prompt de comandos e indica que lo que espera el editor es una instrucción del sistema Maple: cualquier cosa que se escriba a continuación aparecerá en rojo, color reservado a los comandos, mientras que el texto utiliza el color negro. Las instrucciones de Maple han de finalizar con ";" (característica común con el lenguaje de programación C)o con ":". La diferencia entre ambas opciones es que la primera genera una salida en la pantalla (en azul) mientras que la segunda evita que ésta aparezca aunque, por supuesto, en ambos casos el comando se ejecuta cuando se pulsa la tecla *ENTER*.

La interfase de Mathematica ofrece una forma interactiva de operar con el programa, de forma que cuando se introduce una operación, Mathematica la ejecuta y devuelve un valor. Al mismo tiempo, numera los *Input* (**In**[] entradas) y los *Output* (**Out**[] salidas), secuencialmente. Al igual que en Matlab, si se teclea al final de una orden ";" ésta se ejecuta pero el resultado no se visualiza por pantalla. Para que se ejecute una operación, se debe pulsar la tecla *ENTER*.

1.6. Aplicaciones y ejercicios

Ejercicio1.1 Redondear y trucar cada uno de los siguientes casos utilizando 5 cifras significativas. **1.** e=2.718281828, **2.** π=3.141592653, **3.** x=-123456789, **4.** y=0.0000213475 y 5. z= $\frac{2}{3}$.

Solución:

1. $e = 2.718281828$
   Redondeando: 2.7183
   Truncando: 2.7182
2. $\pi = 3.141592653$
   Redondeando: 3.1416
   Truncando: 3.1415
3. $x = -123456789$
   Redondeando: -12346
   Truncando: -12345
4. $y = 0.0000213475$
   Redondeando: 0.000021348
   Truncando: 0.000021347
5. $z = \frac{2}{3} = 0.66666666667$
   Redondeando: 0.66667
   Truncando: 0.66666

Ejercicio1.2 Encontrar el error absoluto y el error relativo porcentual de x con respecto a xn, en cada uno de los siguientes casos: **1.** $x=.50\times10^{2}$, $xn=.51\times10^{2}$; **2.** $x=.50\times10^{-3}$, $xn=.51\times10^{-3}$ y **3.** $x=.50\times10^{6}$, $xn =.51\times10^{6}$.

Solución:

1. $x = .50 \times 10^2$, $x_n = .51 \times 10^2$, entonces:

$$\begin{aligned}
\varepsilon_{a,v} &= \left|.50 \times 10^2 - .51 \times 10^2\right| = \left|-.01 \times 10^2\right| = .1 \times 10^1 = 1.0 \\
\varepsilon_{r,v}(\,\%) &= \left|\frac{.50 \times 10^2 - .51 \times 10^2}{.50 \times 10^2}\right| 100\,\% = \frac{1.0 \times 10^0}{.50 \times 10^2} 100\,\% \\
&= 2 \times 10^{-2} \times 100\,\% = 2\,\%
\end{aligned}$$

2. $x = .50 \times 10^{-3}$, $x_n = .51 \times 10^{-3}$, entonces:

$$\begin{aligned}
\varepsilon_{a,v} &= \left|.50 \times 10^{-3} - .51 \times 10^{-3}\right| = \left|-.01 \times 10^{-3}\right| = .1 \times 10^{-4} = 0.00001 \\
\varepsilon_{r,v}(\,\%) &= \left|\frac{.50 \times 10^{-3} - .51 \times 10^{-3}}{.50 \times 10^{-3}}\right| 100\,\% = \frac{.1 \times 10^{-4}}{.50 \times 10^{-3}} 100\,\% \\
&= .2 \times 10^{-1} \times 100\,\% = 2\,\%
\end{aligned}$$

3. $x = .50 \times 10^6$, $x_n = .51 \times 10^6$, entonces:

$$\begin{aligned}\varepsilon_{a,v} &= \left|.50 \times 10^6 - .51 \times 10^6\right| = \left|-.01 \times 10^6\right| = .1 \times 10^5 = 10000 \\ \varepsilon_{r,v}(\%) &= \left|\frac{.50 \times 10^6 - .51 \times 10^6}{.50 \times 10^6}\right| 100\,\% = \frac{.1 \times 10^5}{.50 \times 10^6} 100\,\% \\ &= .2 \times 10^{-1} \times 100\,\% = 2\,\%\end{aligned}$$

Con este ejercicio se puede observar que el error relativo es invariante al cambio de escala. Cuando se manejan cantidades muy grandes o muy pequeñas, el error absoluto puede ser engañoso, mientras que el error relativo es más significativo en estos casos.

Ejercicio1.3 Obtener una aproximación para e^{-5} y e^{5}, redondeando a 4 cifras significativas usando la serie de Taylor.

$$e^x \approx 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \ldots + \frac{x^n}{n!}$$

Solución:

Las aproximaciones correspondientes a e^{-5} y e^5 aparecen en el cuadro 1.11.

$\therefore$ Al redondear a 4 cifras significativas, $e^{-5} \approx 0.006738$ (la suma $\sum_{k=0}^{n} \frac{(-5)^k}{k!}$ se estabilizó en $n = 24$) y $e^5 \approx 148.4$, la suma ($\sum_{k=0}^{n} \frac{5^k}{k!}$ se estabilizó en $n = 17$).

Ejercicio1.4 Usar la serie de Taylor de e^x con 9 términos redondeando a 4 cifras significativas para obtener e^{-5} mediante:

$$e^{-5} = \frac{1}{e^5} \approx \frac{1}{\sum_{k=0}^{9} \frac{5^k}{k!}}$$

y compararlos resultados obtenidos con el ejercicio1.3.(Valor exacto de e^{-5} $=6.737946999... \times 10^{-3}$).

Solución:

Las aproximaciones correspondientes a e^{-5} aparecen en el cuadro 1.12.

Al redondear a 4 cifras significativas y utilizando 9 términos de la serie, e^{-5} ≈ 0.007230. Es decir con el método utilizado en el ejercicio 1.4 se obtiene una mejor aproximación que con el utilizado en el ejercicio1.3. Esto se debe a que hay términos relativamente grandes con respecto a un número pequeño e^{-5}, los cuales al ser sumados producen pérdidas de cifras significativas. Por lo que una forma más adecuada de calcular e^{-5} es aumentando la precisión y calculando $\frac{1}{e}$

en lugar de e^{-5}.

Ejercicio1.5 La pérdida de cifras significativas se puede evitar a veces reordenando los términos de la función usando una identidad conocida del álgebra o de la trigonometría. Encontrar, en cada uno de los siguientes casos, una formula equivalente a la dada que evite la pérdida de cifras significativas y la resta ó suma de dos números casi iguales.

$$
\begin{array}{lll}
1. & \ln(x+1)-\ln(x) & \text{para } x \text{ grande} \\
2. & \cos^2(x)-\operatorname{sen}^2(x) & \text{para } x=\frac{\pi}{4} \\
3. & \sqrt{\frac{1+\cos(x)}{2}} & \text{para } x=\pi \\
4. & \frac{1}{1-\cos(x)} & \text{para } x=0
\end{array}
$$

Solución:

$$
\begin{array}{lrl}
1. & \ln(x+1)-\ln(x) & = \ln(\frac{x+1}{x}) = \ln(1+\frac{1}{x}) \\
2. & \cos^2(x)-\operatorname{sen}^2(x) & = \cos(2x) \\
3. & \sqrt{\frac{1+\cos(x)}{2}} & = \cos(\frac{x}{2}) \\
4. & \frac{1}{1-\cos(x)} & = \frac{1}{2\operatorname{sen}^2(\frac{x}{2})} \\
 & \operatorname{sen}(\frac{x}{2}) & = \sqrt{\frac{1-\cos(x)}{2}} = \frac{\sqrt{1-\cos(x)}}{\sqrt{2}} \\
 & \sqrt{2}\operatorname{sen}(\frac{x}{2}) & = \sqrt{1-\cos(x)} \\
 & (\sqrt{2}\operatorname{sen}(\frac{x}{2}))^2 & = (\sqrt{1-\cos(x)})^2 \\
 & 2\operatorname{sen}^2(\frac{x}{2}) & = 1-\cos(x)
\end{array}
$$

Al reordenar los términos de la función usando una identidad conocida, se puede evitar la pérdida de cifras significativas y la resta ó suma de dos números casi iguales.

Ejercicio1.6 Representar el número 1.32421875 en una palabra de 64 bits usando el sistema de la IEEE.

Solución:

Para poder representar 1.32421875 en un tamaño de palabra de 64 bits, se tienen que identificar cada una de sus partes (el signo s, el exponente t en forma sesgada (c) y la mantisa f).

Dado que es un número positivo, el signo se representa con s = 0.

Primero se escribe el número 1.32421875 sin signo en base 2, obteniendo primero la parte entera y después la fraccionaria. El resultado es 1.010100110; (resultado obtenido en Maple).

Debido a que el número ya esta normalizado se escribe en la forma normalizada general, usando la ecuación (1.7); obteniéndose:

$$V = (-1)^s(1.f)_2(2^t)_{10} = (-1)^0(1.010100110)_2(2^0)_{10}$$

La mantisa es la parte a la derecha del punto decimal, y se rellena con ceros a la derecha hasta que se obtengan los 52 bits. Es decir: **f=010100110000000000000000000000000000000000000.**

El exponente t=0, se necesita desplazarlo para obtener el valor de c, usando la ecuación (1.9). Para el formato IEEE de 64 bits, el desplazamiento es b=1023: c =0+1023 = **1023**. Este valor se convierte en binario, obteniéndose: 1111111111; (resultados obtenidos en Maple). 1.32421875 se representa en una palabra de 64 bits usando el sistema de la IEEE, de la siguiente manera:

$$\underbrace{0}_{\text{Signo}}\ \underbrace{01111111111}_{\text{Exponente}}\ \underbrace{0101001100000000000000000000000000000000000000000000}_{\text{Mantisa}}$$

Ejercicio1.7 Usar el sistema de la IEEE para determinar el decimal equivalente al siguiente número de máquina de punto flotante en una palabra de 64 bits.

$$\boxed{1\,10000001010\,1001001100000000000000000000000000000000000000000000}$$

Solución:

Este número de punto flotante se puede escribir en la forma normalizada general, de la siguiente manera; (para obtener el valor de t, se utiliza la ecuación (1.8)), donde c=100000010102=103410; (resultado obtenido en Maple).

$$\begin{aligned}
&1\ \ 1000001010\ \ 10010001000000000000000000000000000000000000 =\\
&(-1)^1(1.100100010000000000000000000000000000000000000)_2(2^{1034-1023})_{10}=\\
&(-1)(1.1001001100000000000000000000000000000000000)_2(2^{11})_{10}=\\
&-110010011000.00000000000000000000000000000000000_2=\\
&-3224_{10}
\end{aligned}$$

Lo anterior se obtuvo desplazando 11 posiciones a la derecha el punto decimal de la mantisa, ya que t=11. Después se escribe el número en base 10.

El número decimal equivalente es -3224.

Ejercicio1.8 Evaluar $f(x)= x^3-6.1x^2+3.2x+1.5$ en x=4.71; redondeando a tres cifras significativas y calcular el error relativo, (valor exacto f(4.71)=-14.263899).

Solución:

En el cuadro 1.13 se dan los resultados intermedios de los cálculos.

$$f(4.71) = 104 - 135 + 15.1 + 1.5 = -14.4$$

El error relativo verdadero es:

$$\left| \frac{-14.263899 - (-14.4)}{-14.263899} \right| = 0.00954$$

$f(4.71) = -14.4$

Ejercicio1.9 Evaluar f(x)= x^3-6.1x^2+3.2x+1.5 en x=4, escribiendo la función en forma anidada; redondeando a tres cifras significativas y calcular el error relativo; comparar con el resultado obtenido en el ejercicio 1.8.

Solución:

Se escribe f(x) de forma anidada, obteniéndose:

$$f(4.71) = ((4.71 - 6.1)4.71 + 3.2)4.71 + 1.5 = -14.3$$

El error relativo verdadero es:

$$\left|\frac{-14.263899 - (-14.3)}{-14.263899}\right| = 0.00253$$

f(4.71)=-14.3. Se puede observar con los ejercicios 1.8 y 1.9 que al evaluar los polinomios en forma anidada se reduce el error relativo.

Ejercicio1.10 Una cámara instalada en una sonda espacial tomó fotografías del planeta Marte y las envió en forma de señales de radio hacia la tierra, en donde una computadora recibió las "fotografías" enforma de numerales binarios formados por seis bits. El sombreado de cada punto, en la fotografía final, fue determinado por esos seis bits. El numeral 0000002 indicaba un punto blanco y el numeral 1111112 denotaba un punto negro. Los 62 numerales intermedios representaban distintos sombreados que iban del gris al negro. Para obtener una fotografía completa se necesitaban 40.000 puntos.

1 Si uno de los numerales recibidos fue el 1101112 ¿indicar a que numeral decimal corresponde?
2 ¿El punto que corresponde al numeral recibido en el inciso anterior representa una sombra de color gris cercana al blanco o al negro?
3 ¿Cuál es el numeral binario que representaría el gris más claro que no llegue al blanco?
4 ¿Cuál es el numeral binario que constituiría el gris más oscuro que no llegue al negro?
5 ¿Indicar que numeral binario representaría el sombreado correspondiente al número 31?

Solución:

1. $110111_2 = 55_{10}$
2. El 55 es un número que representa una sombra color gris cercana al negro
3. 000001_2
4. 111110_2
5. 11111_2

Descripción	Matlab
Número de cifras significativas con las que se desea trabajar	EDU>> vpa(pi, 6) ans= 3.14159
Aproximación a la solución de una ecuación de una variable	EDU>> vpa(allvalues(solve('$-2*x$^8 $+ 5*x$^6 $- 22*x$^5 $- 11 = 0$','x')), 6) ans= [-2.59307] [.848435] [.292025 $-$ .797110*i] [$-$.292025+.797110*i] [.726027 $-$ .544071*i] [.726027+.544071*i] [1.28675 $-$ 1.59957*i] [1.28675+1.59957*i]
Gráfica de una función	EDU>> x=linspace(-5, 5); EDU>> y= -2*x.^8 + 5*x.^6 $-$ 22*x.^5 $-$ 11; EDU>> plot(x, y) EDU>> axis([-4, 1, -300, 800])
Definir una matriz A	EDU>> A = [1 1 1; 2 -1 1; 3 2 1] A= 1 1 1 2 -1 1 3 2 1
Determinante de A	EDU>> det(A) ans= 5
Algún elemento de A	EDU>> A(2, 2) ans= -1
Inversa de A	EDU>> inverse(A) ans= [$-3/5$, $1/5$, $2/5$] [$1/5$, $-2/5$, $1/5$] [$7/5$, $1/5$, $-3/5$]
Transpuesta de A	EDU>> transpose(A) ans= [1, 2, 3] [1, -1, 2] [1, 1, 1]
Resolver un sistema $Ax = b$	EDU>> b = [11; 5; 24] b = 11 5 24 EDU>> x = linsolve(A, b) x= [4] [5] [2]

Cuadro 1.4: Algunas funciones en Matlab

Descripción	Matlab
Multiplicación de un escalar por A	EDU>> 2*A ans= $\begin{matrix} 2 & 2 & 2 \\ 4 & -2 & 2 \\ 6 & 4 & 2 \end{matrix}$
Multiplicación de matrices	EDU>> A*b ans= $\begin{matrix} 40 \\ 41 \\ 67 \end{matrix}$
Eigenvalores de A	EDU>> eig(A) ans= $\begin{matrix} 3.6044 \\ -0.7468 \\ -1.8576 \end{matrix}$
Factorización de la matriz A en la forma LU	EDU>> lu(A) ans= $\begin{matrix} 3.0000 & 2.0000 & 1.0000 \\ -0.6667 & -2.3333 & 0.3333 \\ -0.3333 & 0.1429 & 0.7143 \end{matrix}$
Factorización de la matriz A en la forma LL^t	EDU>> A = matrix[5 1 2; 1 7 0; 2 0 5] A= $\begin{matrix} 5 & 1 & 2 \\ 1 & 7 & 0 \\ 2 & 0 & 5 \end{matrix}$ EDU>> vpa(chol(A),4) ans= [2.236, .4472, .8944] [0, 2.608, −.1534] [0, 0, 2.044]

Cuadro 1.5: Algunas funciones en Matlab

Descripción	Maple
Número de cifras significativas con las que se desea trabajar	> **Digits** := 7; $Digits := 7$
Aproximación a la solución de una ecuación de una variable	> f := −2 * x^8 + 5 * x^6 − 22 * x^5 − 11; $f := -2x^8 + 5x^6 - 22x^5 - 11$ > **fsolve**(f, x); $-2.593, -0.8484$
Gráfica de una función	> **plot**(f, x=−4..1, y=−300..800, color=[black]);
Se carga la librería linalg	> **with(linalg)**;
Definir una matriz A	> A := **matrix**([[1, 1, 1], [2, −1, 1], [3, 2, 1]]); $A := \begin{bmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 3 & 2 & 1 \end{bmatrix}$
Determinante de A	> **det**(A); 5

Cuadro 1.6: Algunas funciones en Maple

Descripción	Maple
Menores de A	> **minor**(A, 2, 2); $\begin{bmatrix} 1 & 1 \\ 3 & 1 \end{bmatrix}$
Algún elemento de A	> A[2, 2]; -1
Inversa de A	> **inverse**(A); $\begin{bmatrix} \frac{-3}{5} & \frac{1}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{-2}{5} & \frac{1}{5} \\ \frac{7}{5} & \frac{1}{5} & \frac{-3}{5} \end{bmatrix}$
Transpuesta de A	> **transpose**(A); $\begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 2 \\ 1 & 1 & 1 \end{bmatrix}$
Resolver un sistema $Ax = b$	> b := matrix([[11], [5], [24]]); $b := \begin{bmatrix} 11 \\ 5 \\ 24 \end{bmatrix}$ > x := **linsolve**(A, b); $x := \begin{bmatrix} 4 \\ 5 \\ 2 \end{bmatrix}$
Multiplicación de un escalar por A	> **evalm**$(2 * A)$; $\begin{bmatrix} 2 & 2 & 2 \\ 4 & -2 & 2 \\ 6 & 4 & 2 \end{bmatrix}$
Multiplicación de matrices	> **multiply**(A, b); $\begin{bmatrix} 40 \\ 41 \\ 67 \end{bmatrix}$
Transformar la matriz $[A\|b]$ en la matriz identidad I	> Ab := **matrix**([[1, 1, 1, 11], [2, −1, 1, 5], [3, 2, 1, 24]]); $\text{Ab}:= \begin{bmatrix} 1 & 1 & 1 & 11 \\ 2 & -1 & 1 & 5 \\ 3 & 2 & 1 & 24 \end{bmatrix}$ > **gaussjord**(Ab); $\begin{bmatrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 2 \end{bmatrix}$
Transformar la matriz $[A\|b]$ en matriz triangular superior	> **gausselim**(Ab); $\begin{bmatrix} 1 & 1 & 1 & 11 \\ 0 & -3 & -1 & -17 \\ 0 & 0 & \frac{-5}{3} & \frac{-10}{3} \end{bmatrix}$
Eigenvalores de A	> evalf(**eigenvalues**(A)); $3.604+0.\,I, \quad -1.858\text{-}0.0008660\,I, \quad -0.7465+0.0008660\,I$

Cuadro 1.7: Algunas funciones en Maple

Descripción	Maple
Factorización de la matriz A en la forma LU	> **LUdecomp**(A, L='l', U='u'); $$\begin{bmatrix} 1 & 1 & 1 \\ 0 & -3 & -1 \\ 0 & 0 & \frac{-5}{3} \end{bmatrix}$$ > **evalm**(l); $$\begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 3 & \frac{1}{3} & 1 \end{bmatrix}$$ > **evalm**(u); $$\begin{bmatrix} 1 & 1 & 1 \\ 0 & -3 & -1 \\ 0 & 0 & \frac{-5}{3} \end{bmatrix}$$
Factorización de la matriz A en la forma LL^t	> A := **matrix**([[5, 1, 2], [1, 7, 0], [2, 0, 5]]); $$A := \begin{bmatrix} 5 & 1 & 2 \\ 1 & 7 & 0 \\ 2 & 0 & 5 \end{bmatrix}$$ > evalf(**cholesky**(A)); $$\begin{bmatrix} 2.236 & 0. & 0. \\ 0.4472 & 2.608 & 0. \\ 0.8944 & -0.1534 & 2.043 \end{bmatrix}$$

Cuadro 1.8: Algunas funciones en Maple

Descripción	Mathematica
Número de cifras significativas con las que se desea trabajar	In[1]:= N[Pi,20] Out[1] = 3.1415926535897932385
Aproximación a la solución de una ecuación	In[2]:= N[**Solve**[$-2*x\hat{}8+5*x\hat{}6-22*x\hat{}5-11 == 0, x$]] Out[2]= {{x-> −2.59307}, {x-> −0.848435}, > {x-> −0.292025 − 0.79711 I}, > {x-> −0.292025 + 0.79711 I}, > {x-> −0.726027 − 0.544071 I}, > {x-> −0.726027 + 0.544071 I}, > {x-> −1.28675 − 1.59957 I}, > {x-> −1.28675 + 1.59957 I}}
Gráfica de una función	In[3]:= **Plot**[{$-2*x\hat{}8+5*x\hat{}6-22*x\hat{}5-11$}, {$x, -4, 1$}, PlotRange-> {−300, 800}] Out[3]= -Graphics-
Definir una matriz A	In[4]:= MatrixForm[A = {{1,1,1},{2,−1,1},{3,2,1}}] Out[4]//MatrixForm= $\begin{matrix} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 3 & 2 & 1 \end{matrix}$
Determinante de A	In[5]:= **Det**[A] Out[5]=5
Menores de A	In[6]:= **Minors**[A][[2,2]] Out[6]= -2
Algún elemento de A	In[7]:= A[[2,2]]; Out[7]= -1
Inversa de A	In[8]:= MatrixForm[**Inverse**[A]] Out[8]//MatrixForm= $\begin{matrix} -(\frac{3}{5}) & \frac{1}{5} & \frac{2}{5} \\ \frac{1}{5} & -(\frac{2}{5}) & \frac{1}{5} \\ \frac{7}{5} & \frac{1}{5} & -(\frac{3}{5}) \end{matrix}$
Transpuesta de A	In[9]:= MatrixForm[**Transpose**[A]] Out[9]//MatrixForm= $\begin{matrix} 1 & 2 & 3 \\ 1 & -1 & 2 \\ 1 & 1 & 1 \end{matrix}$
Resolver un sistema $Ax = b$	In[10]:= MatrixForm[b = {{11},{5},{24}}] Out[10]//MatrixForm= $\begin{matrix} 11 \\ 5 \\ 24 \end{matrix}$ In[11]:= x=LinearSolve[A,b] Out[11]={{4}, {5}, {2}}
Multipliación de un escalar por A	In[12]:= MatrixForm[A ∗ 2] Out[12]//MatrixForm= $\begin{matrix} 2 & 2 & 2 \\ 4 & -2 & 2 \\ 6 & 4 & 2 \end{matrix}$

Cuadro 1.9: Algunas funciones en Mathematica

Descripción	Mathematica
Multiplicación de matrices	In[13]:= MatrixForm[A . b] Out[13]//MatrixForm= 40 41 67
Transformar la matriz $[A\|b]$ en la matriz identidad I	In[14]:= MatrixForm[Ab = $\{\{1, 1, 1, 11\}, \{2, -1, 1, 5\}, \{3, 2, 1, 24\}\}$] Out[14]//MatrixForm= $\begin{matrix} 1 & 1 & 1 & 11 \\ 2 & -1 & 1 & 5 \\ 3 & 2 & 1 & 24 \end{matrix}$ In[15]:= MatrixForm[RowReduce[Ab]] Out[15]//MatrixForm= $\begin{matrix} 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 5 \\ 0 & 0 & 1 & 2 \end{matrix}$
Eigenvalores de A	In[16]:= N[Eigenvalues[A]] Out[16]= $\{3.60438-1.11022\ 10^{-16}$ I, $-0.74676+4.44089\ 10^{-16}$ I, $-1.85762-2.22045\ 10^{-16}$ I$\}$
Factorización de la matriz A en la forma LU	In[17]:= **LUDecomposition**[A] Out[17]= $\{\{\{1, 1, 1\}, \{3, -1, -2\}, \{2, 3, 5\}\}, \{1, 3, 2\}, 1\}$

Cuadro 1.10: Algunas funciones en Mathematica

Términos	Resultados para $e^{-5} = \sum_{k=0}^{n} \frac{(-5)^k}{k!}$	Resultados para $e^{5} = \sum_{k=0}^{n} \frac{5^k}{k!}$
1	1.000	1.000
2	-4.000	6.000
3	8.500	18.50
4	-12.33	39.33
5	13.71	65.38
6	-12.33	91.42
7	9.368	113.1
8	-6.133	128.6
9	3.555	138.3
10	-1.827	143.7
11	0.8640	146.4
12	-0.3592	147.6
13	0.1505	148.1
14	-0.04556	148.3
15	0.02446	148.4
16	0.001120	148.4
17	0.008413	148.4
18	0.006268	
19	0.006864	
20	0.006707	
21	0.006746	
22	0.006737	
23	0.006739	
24	0.006739	

Cuadro 1.11: Ejercicio para obtener una aproximación de e^{-5} y e^{5}

Términos	Resultados para $e^{-5} = \sum_{k=0}^{9} \frac{(-5)^k}{k!}$	Resultados para $\frac{1}{e^5 = \sum_{k=0}^{9} \frac{5^k}{k!}}$
1	1.000	1.000
2	-4.000	0.1667
3	8.500	0.05405
4	-12.33	0.02542
5	13.71	0.01530
6	-12.33	0.01094
7	9.368	0.008840
8	-6.133	0.007775
9	3.555	0.007230

Cuadro 1.12: Ejercicio para obtener una aproximación de e^{-5} utilizando dos métodos

x	x^2	x^3	$6.1x^2$	$3.2x$
4.71	22.2	104	135	15.1

Cuadro 1.13: Cálculos intermedios de la función $f(x) = x^3 - 6.1x^2 + 3.2x + 1.5$

Capítulo 2

Solución numérica de ecuaciones

En muchas aplicaciones de la ingeniería y la ciencia se presentan con frecuencia problemas que no se pueden resolver analíticamente [1], por lo que se recurre a los métodos numéricos para poder encontrar una solución.
El objetivo principal de este capítulo será aplicar algunos métodos numéricos para hallar raíces reales [2] de ecuaciones no lineales de una variable, que satisfacen a una ecuación del tipo:

$$f(x) = 0 \tag{2.1}$$

Los valores que hacen que una función $f(x) = 0$, se conocen con el nombre de raíces o ceros de f.
Se analizarán los siguientes métodos, por medio de los cuales se obtendrá un valor x_n tal que:

$$x_n \approx x \tag{2.2}$$

$$f(x_n) \approx 0 \tag{2.3}$$

1. **Métodos cerrados**: requieren dos valores iniciales x_0 y x_1 [3], que deben encerrar a la raíz.
   (a) Método de bisección.
   (b) Método de falsa posición.
2. **Métodos abiertos**: requieren un sólo valor de inicio x_0 o dos de ellos x_0 y x_1, pero que no necesariamente deben encerrar a la raíz.
   (a) Método de Newton.
   (b) Método de la secante.

---

[1] Los métodos analíticos consisten en despejar la variable x de la función $f(x)$.
[2] Sólo se estudiarán raíces complejas para ecuaciones polinómicas.
[3] Los valores iniciales también se pueden denominar como a y b.

Cabe mencionar que no existe un método universal de resolución de sistemas de ecuaciones no lineales. Algunos de ellos funcionarán sobre ciertos sistemas y no servirán para resolver otros. Los métodos que presenten un buen comportamiento sobre algunos sistemas pueden no ser los mejores para resolver otros sistemas diferentes.

2.1 Método de bisección

El *método de bisección* es el más antiguo y sencillo para determinar las raíces reales de una ecuación; también denominado método de Bolzano, ya que éste fue el primero en proponerlo. Comienza con un intervalo $[a,b]$ donde $f(a)$ y $f(b)$ son de signos opuestos, garantizando así la existencia de al menos una raíz en el intervalo $[a,b]$. Ésta es una consecuencia del teorema del valor medio para funciones continuas.

Teorema 2.1.1 Si f es una función continua definida en el intervalo $[a,b]$, y se satisface $f(a) * f(b) < 0$, entonces existe al menos un número x en (a,b) tal que $f(x) = 0$.

Para simplificar, asumimos que $[a,b]$ contiene exactamente una raíz x. (Aunque el método también se puede aplicar cuando hay más de una raíz en $[a,b]$; pero cuando esto sucede no se garantiza que el método converja a la solución buscada).
Éste método consiste en dividir sucesivamente el intervalo $[a,b]$ por la mitad, hasta que la longitud del subintervalo tienda a cero.
En la figura 2.1 se muestra la representación gráfica del método de bisección.

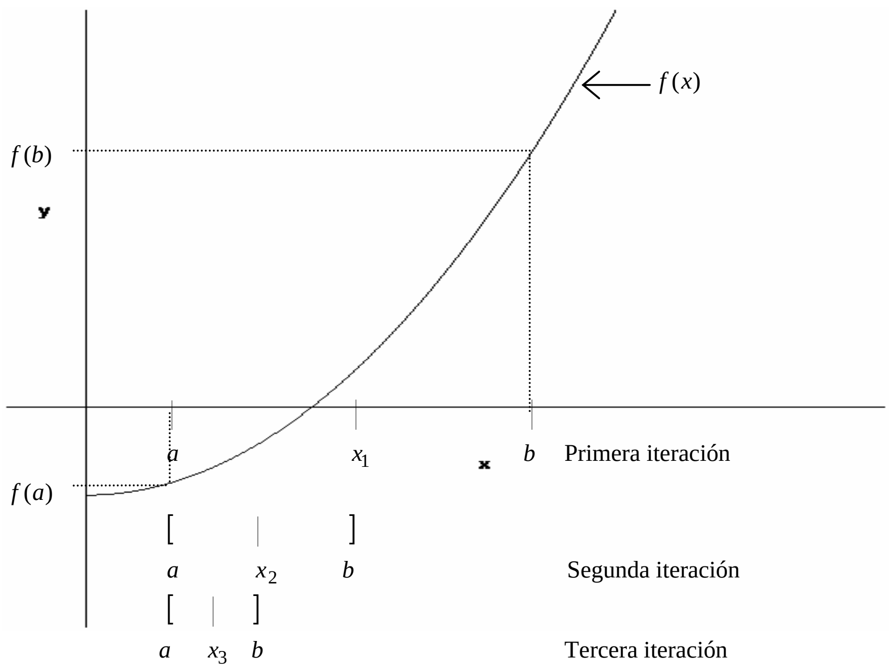


Figura 2.1. Representación gráfica del método de bisección

El primer paso antes de aplicar el método de bisección para garantizar que converja a la solución, es comprobar que f es continua en el intervalo $[a,b]$ [4] y que en éste intervalo se encuentra exactamente una raíz [5]; si se cumplen estos dos requisitos se garantiza la convergencia del método; pero no significa que si no se cumple cualquiera de estos el método vaya a diverger.

A continuación, se calcula para la primera iteración $(n=1)$ el punto medio del intervalo, que será la primera aproximación de la raíz. Este valor esta dado por:

$$x_1 = a + \frac{b-a}{2} = \frac{2a+b-a}{2} = \frac{a+b}{2}$$

Para obtener el n-ésimo punto medio, la ecuación sería:

$$x_n = \frac{a+b}{2} \qquad \textit{fórmuladeiteraciónparaelmétododebisección} \tag{2.4}$$

Después se verifican los signos de $f(a)*f(x_n)$ y $f(x_n)*f(b)$, para saber en que subintervalo se encuentra la raíz:

1. Si $f(a)*f(x_n)<0$, entonces en el subintervalo $[a,x_n]$ se encuentra la raíz, renombrando $a=a$ y $b=x_n$.
2. Si $f(a)*f(x_n)>0$, entonces en el subintervalo $[x_n,b]$ se encuentra la raíz, renombrando $a=x_n$ y $b=b$.
3. Si $f(a)*f(x_n)=0$, entonces x_n [6] es la raíz.

Si ocurre alguno de los dos primeros casos se calcula para la segunda iteración el punto medio del nuevo intervalo con la ecuación (5), siendo éste la aproximación más actualizada de la raíz.

Se repite el proceso hasta que se cumplan los siguientes criterios de convergencia:

$$\left|\frac{x_n - x_{n-1}}{x_n}\right| \le Tol \tag{2.5}$$

$$\left|f(x_n)\right| \le Tol \tag{2.6}$$

$$n > N \tag{2.7}$$

donde Tol es alguna tolerancia definida >0 y N es un número máximo de iteraciones.

Se deben cumplir simultáneamente los criterios (2.5) y (2.6), para asegurar que se ha encontrado una aproximación de la raíz. En su defecto se debe cumplir el criterio (2.7), el cual nos indicará que no se ha logrado la convergencia, en un cierto número de iteraciones.

---

[4]Cuando no hay información previa acerca de los valores aproximados de las raíces, una forma sencilla para hallar intervalos $[a,b]$ que contengan una raíz, es escribir una tabla de la función para valores de x con separación uniforme; ó graficar la función con ayuda de un programa.

[5]Una forma de verificar estos dos requisitos es graficando la función $f(x)$.

[6]Debido a los errores de redondeo, es poco probable que $f(x_n)=0$, por lo que se deben de utilizar otros criterios para saber cuando detener el proceso.

Teorema 2.1.2 Si f es una función continua definida en el intervalo $[a,b]$, y se satisface $f(a)*f(b)<0$. El método de bisección genera una sucesión $\{x_n\}_{n=1}^{\infty}$ que aproxima una raíz de f, tal que:

$$|x_n - x_{n-1}| \leq \frac{b-a}{2^n} \leq Tol \qquad \text{para} n \geq 1 \tag{2.8}$$

Aplicando el teorema 2.1.2, se puede despejar n de la ecuación (2.8/, para saber cuántas iteraciones se necesitan realizar con una tolerancia (Tol); obteniéndose:

$$\Rightarrow \frac{b-a}{2^n} \leq \text{Tol}$$

$$b-a \leq 2^n Tol$$

$$\frac{b-a}{Tol} \leq 2^n$$

$$\log\left(\frac{b-a}{Tol}\right) \leq \log 2^n$$

$$\log\left(\frac{b-a}{Tol}\right) \leq n \log 2$$

$$\frac{\log\left(\frac{b-a}{Tol}\right)}{\log 2} \leq \text{n}$$

$$N > \frac{\log\left(\frac{b-a}{Tol}\right)}{\log 2} \tag{2.9}$$

Nota: Es necesario utilizar más iteraciones de las obtenidas al calcular $\frac{\log\left(\frac{b-a}{Tol}\right)}{\log 2}$; ya que se deben cumplir simultáneamente los criterios: $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq Tol$ y $f(x_n) \leq Tol$, para asegurar que se ha encontrado una aproximación de la raíz; es decir no sólo se esta tomando en cuenta un error determinado, sino que la función $f(x) \approx Tol$; y la ecuación (2.9) sólo toma en cuenta el intervalo y el error.

Algoritmo 2.1.1 Método de bisección. Para encontrar una aproximación de la raíz, de una ecuación $f(x)=0$ donde f es una función continua en $[a,b]$ y $f(a)*f(b)<0$. (Para garantizar la convergencia del método).

Datos: $f(x)$, el intervalo $[a,b]$, una tolerancia Tol y un número máximo de iteraciones N.

Resultados: Una raíz aproximada o un mensaje de falla.

- **Paso 1:** Calcular $\frac{\log\left(\frac{b-a}{Tol}\right)}{\log 2} + 10$, *redondear a* enteros e igualar con N.
- **Paso 2:** Hacer $n=1$.
- **Paso 3:** Mientras $n \leq N$, repetir los pasos 4 a 15.

- **Paso 4:** Calcular $f(a)$ y $f(b)$.
- **Paso 5:** Si $n=1$ entonces realizar las siguientes comparaciones. De lo contrario CONTINUAR.
    * **Paso 6:** Si $f(a)*f(b)=0$ entonces IMPRIMIR “Una raíz de la ecuación dada es uno de los extremos del intervalo” y TERMINAR. De lo contrario CONTINUAR.
    * **Paso 7:** Si $f(a)*f(b)>0$, entonces IMPRIMIR “Intervalo incorrecto, no hay raíz o hay varias raíces” y TERMINAR. De lo contrario CONTINUAR.
    * **Paso 8:** Si $f(a)*f(b)<0$, entonces IMPRIMIR “Intervalo correcto” y CONTINUAR. De lo contrario TERMINAR.
- **Paso 9:** Calcular $x_n = \frac{a+b}{2}$.
- **Paso 10:** Calcular $f(x_n)$.
- **Paso 11:** Si $f(a)*f(x_n)=0$ entonces IMPRIMIR “Una raíz de la ecuación dada es x_n” y TERMINAR. De lo contrario CONTINUAR.
- **Paso 12:** Si $f(a)*f(x_n)<0$, entonces hacer $a=a$ y $b=x_n$, de lo contrario hacer $a=x_n$ y $b=b$.
- **Paso 13:** Si $n>1$ entonces calcular $\varepsilon = \left|\frac{x_n - x_{n-1}}{x_n}\right|$ y realizar la siguiente comparación. De lo contrario CONTINUAR.
    * **Paso 14:** Si $\varepsilon \le Tol$ y $\left|f(x_n)\right| \le Tol$ entonces IMPRIMIR “Una raíz aproximada de la ecuación dada es x_n” y TERMINAR. De lo contrario CONTINUAR.
- **Paso 15:** Hacer $n=n+1$.

• **Paso 16:** IMPRIMIR mensaje de falla “El método no converge a una raíz” y TERMINAR.

Ejemplo 2.1.1 Usar el método de bisección para encontrar una aproximación de la raíz de la siguiente ecuación; en $[0,1]$, con una tolerancia de 10^{-5}; utilizar 6 decimales.

$$f(x) = e^x - x^2 + 3x - 2 \tag{2.10}$$

Antes de aplicar el método de bisección se grafica la función $f(x)$, obteniendo la figura 2.2; donde se puede observar que la función es continua en el intervalo $[0,1]$ y que en éste intervalo se encuentra exactamente una raíz; por lo que se empieza a aplicar el método de bisección, ya que se puede garantizar que va ha converger.

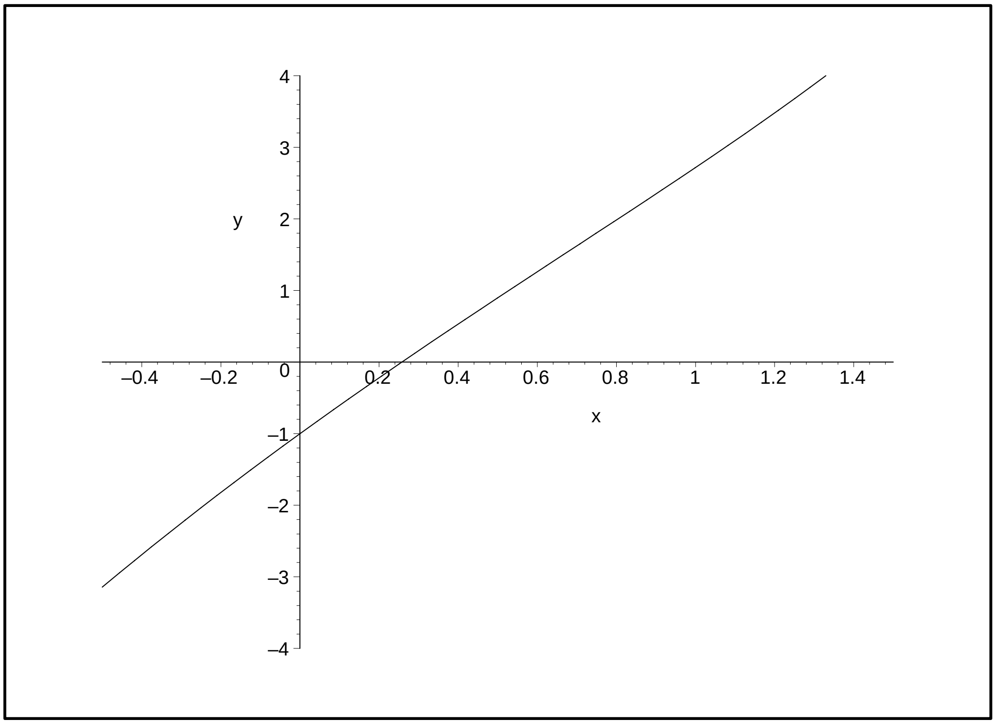


Figura 2.2 Gráfica de la ecuación $f(x) = e^x - x^2 + 3x - 2$

Utilizando el algoritmo 2.1.1, sin tomar en cuenta el **Paso 1**; se empieza calculando $f(0)$ y $f(1)$, obteniéndose:

$$f(0) = e^{(0)} - (0)^2 + 3(0) - 2 = -1.000000$$
$$f(1) = e^{(1)} - (1)^2 + 3(1) - 2 = 2.718282$$

Como $f(0) * f(1) < 0$, el intervalo es correcto ya que encierra una raíz (teorema 2.1.1); a continuación se calcula x_1 con la ecuación (2.4):

$$x_1 = \frac{0+1}{2} = 0.500000$$

después se calcula $f(x_1)$:

$$f(0.500000) = e^{(0.500000)} - (0.500000)^2 + 3(0.500000) - 2 = 0.898721$$

y se multiplica $f(0) * f(0.500000) = -0.898721$; como es < 0, el intervalo para la segunda iteración sería: $a = 0$ y $b = 0.500000$.
Luego se calcula x_2 con la ecuación (2.4):

$$x_2 = \frac{0+0.500000}{2} = 0.250000$$

después se calcula $f(x_2)$:

$$f(0.250000) = e^{(0.250000)} - (0.250000)^2 + 3(0.250000) - 2 = -0.028475$$

y se multiplica $f(0) * f(0.250000) = 0.028475$; como es > 0, el intervalo para la tercera iteración sería: $a = 0.250000$ y $b = 0.500000$.

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq 0.00001$ y que $|f(x_n)| \leq 0.00001$; esto se cumple en la iteración 19: $0.000007 \leq 0.00001$ y $0.000006 \leq 0.00001$.

Para obtener $x_3, x_4, x_5, x_6, \ldots, x_{19}$; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.1.

n	a	b	$f(a)$	$f(b)$	x_n	$f(x_n)$	$f(a)*f(x_n)$	ε
1	0.000000	1.000000	-1.000000	2.718282	0.500000	0.898721	−	——
2	0.000000	0.500000	-1.000000	0.898721	0.250000	-0.028475	+	1.000000
3	0.250000	0.500000	-0.028475	0.898721	0.375000	0.439366	−	0.333333
4	0.250000	0.375000	-0.028475	0.439366	0.312500	0.206682	−	0.200000
5	0.250000	0.312500	-0.028475	0.206682	0.281250	0.089433	−	0.111111
6	0.250000	0.281250	-0.028475	0.089433	0.265625	0.030564	−	0.058824
7	0.250000	0.265625	-0.028475	0.030564	0.257813	0.001066	−	0.030303
8	0.250000	0.257813	-0.028475	0.001066	0.253906	-0.013698	+	0.015385
9	0.253906	0.257813	-0.013698	0.001066	0.255859	-0.006315	+	0.007634
10	0.255859	0.257813	-0.006315	0.001066	0.256836	-0.002625	+	0.003802
11	0.256836	0.257813	-0.002625	0.001066	0.257324	-0.000780	+	0.001898
12	0.257324	0.257813	-0.000780	0.001066	0.257568	0.000144	−	0.000948
13	0.257324	0.257568	-0.000780	0.000144	0.257446	-0.000318	+	0.000474
14	0.257446	0.257568	-0.000318	0.000144	0.257507	-0.000088	+	0.000237
15	0.257507	0.257568	-0.000088	0.000144	0.257538	0.000028	−	0.000119
16	0.257507	0.257538	-0.000088	0.000028	0.257522	-0.000030	+	0.000059
17	0.257522	0.257538	-0.000030	0.000028	0.257530	-0.000001	+	0.000030
18	0.257530	0.257538	-0.000001	0.000028	0.257534	0.000013	−	0.000015
19	0.257530	0.257534	-0.000001	0.000013	0.257532	0.000006	−	0.000007

Cuadro 2.1: Ejemplo del método de bisección cuando converge

A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.

Para la primera iteración el error relativo no se calcula, porque no se cuenta con una aproximación inicial.

Para la segunda iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{0.250000 - 0.500000}{0.250000}\right| = 1.000000$$

Para la tercera iteración el error relativo se calcula:

$$\left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{0.375000 - 0.250000}{0.375000}\right| = 0.333333$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.1.

$$x \approx 0.257532 \qquad \text{Con una tolerancia } \text{Tol=0.00001}$$

Comprobación:

$$\text{f}(0.257532)=\text{e}^{(0.257532)} - (0.257532)^2 + 3(0.257532) - 2 = 0.000006 \approx 0$$

La raíz verdadera de x, con diez cifras decimales , es $x = 0.2575302854$. Se puede observar que $x_{17} = 0.257530$ es una mejor aproximación de x que x_{19}, ya que $|f(x_{17})| < |f(x_{19})|$, es decir $f(x_{17})$ se aproxima más a 0 (cero); se puede verificar que esto es cierto ya que se conoce la raíz verdadera. Por lo tanto $x \approx 0.257530$. Cabe aclarar que aplicando el algoritmo 2.1.1 el resultado que se obtiene es, $x \approx 0.257532$.
Con el ejemplo 2.1.1, se puede observar que el hecho de que el método de bisección no tome en cuenta la magnitud de los valores de la función en la aproximaciones calculadas x_n, y que sólo tome en cuenta el signo de $f(x_n)$;puede resultar que una aproximación intermedia (x_{17}), se aproxime más a la raíz que la respuesta final (x_{19}).

Ejemplo 2.1.2 Determinar el número de iteraciones necesarias para resolver la ecuación (2.10); en $[0,1]$, con una tolerancia de 10^{-5}.
Utilizando la ecuación (2.9); se calcula el valor de N, de la siguiente manera:

$$\begin{aligned} n &\geq \frac{\log\left(\frac{1-0}{10^{-5}}\right)}{\log 2} \\ n &\geq \frac{\log\left(\frac{1}{10^{-5}}\right)}{\log 2}(1) \\ n &\geq \frac{\log 100000}{0.301029995664}(2) \\ n &\geq 16.609640474436 \end{aligned}$$

Redondeando a enteros, se obtiene:

$$N = 17$$

Por lo tanto se necesitan 17 iteraciones para lograr una aproximación con una tolerancia de 10^{-5}. Esto se puede verificar con el ejercicio 2.1.1. Cabe mencionar que en este ejemplo sólo se esta tomando en cuenta el error de tolerancia Tol.
Observando los ejemplos 2.1.1 y 2.1.2, es conveniente utilizar la ecuación (2.9) para calcular el número máximo de iteraciones N, en lugar de designarlo arbitrariamente. Por lo que es importante tomar en cuenta el **Paso 1** en algoritmo 2.1.1.
Así mismo en el ejemplo 2.1.2, se puede observar que para calcular el valor de N, no se toma en cuenta la función $f(x)$, sólo el intervalo y el error de tolerancia; por lo que al calcular el valor de N aplicando el teorema 2.1.2 y la ecuación (2.9); no significa que en esa iteración ya se encontró un aproximación de la raíz. Se tienen que cumplir simultáneamente los siguientes criterios: $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq Tol$ y $|f(x_n)| \leq Tol$, para asegurar que se ha encontrado una aproximación de la raíz.

Ejemplo 2.1.3 Usar el método de bisección para encontrar una aproximación de la raíz de la siguiente ecuación; en $[2,5]$, con una tolerancia de 10^{-5}; utilizar 6 decimales.

$$f(x) = \frac{2}{x-4}$$

Antes de aplicar el método de bisección se grafica la función $f(x)$, obteniendo la figura 2.3; donde se puede observar que la función no es continua en el intervalo $[2,5]$ y que en éste intervalo no se encuentra alguna raíz. La línea que intersecta en $x=4$, es una singularidad [7].

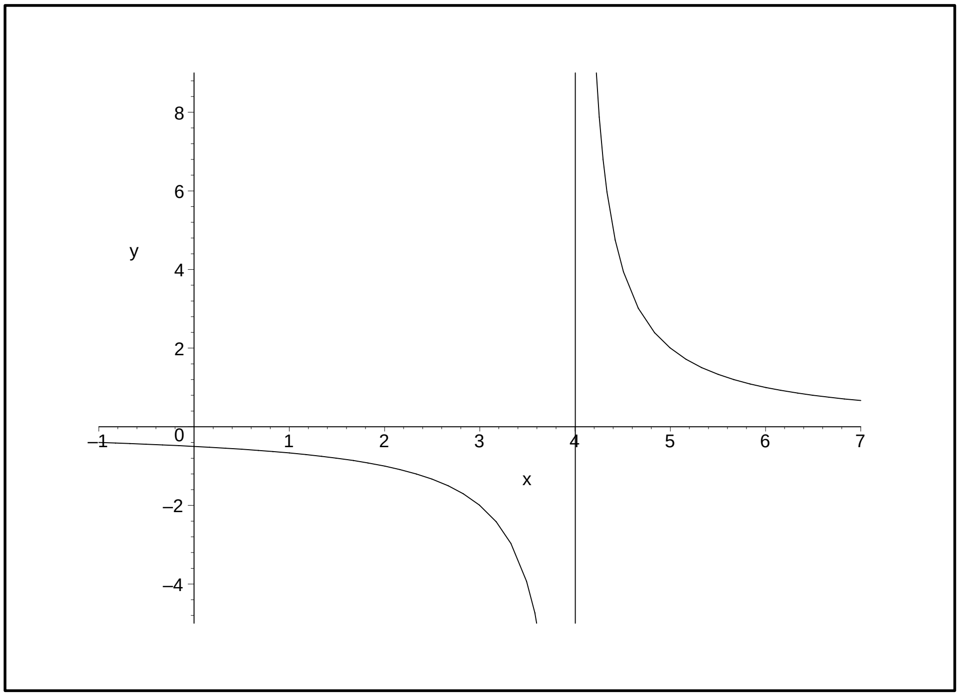


Figura 2.3 Gráfica de la ecuación $f(x)=\frac{2}{x-4}$

Supóngase que no se grafica la función y se empieza a aplicar el método. Utilizando el algoritmo 2.1.1, se empieza calculando el valor de N:

$$
\begin{aligned}
N &\geq \frac{\log\left(\frac{5-2}{10^{-5}}\right)}{\log 2}+10 \\
N &\geq \frac{\log\left(\frac{3}{10^{-5}}\right)}{\log 2}+10(3) \\
N &\geq \frac{\log 300000}{0.301029995664}+10(4) \\
N &\geq 18.194602975158+10(5) \\
N &\geq 28.194602975158
\end{aligned}
$$

Redondeando a enteros, se obtiene:

$$\Rightarrow N=28$$

Después se calcula $f(2)$ y $f(5)$, obteniéndose:

$$f(2)=\frac{2}{3-4}=-1.000000$$

[7]Un punto singular es aquel en el que el valor de la función tiende a infinito.

$$f(5) = \frac{2}{3-4} = 2.000000$$

Como $f(2) * f(5) < 0$, se supondría que el intervalo es correcto y que encierra una raíz; a continuación se calcula x_1 con la ecuación (2.4):

$$x_1 = \frac{2+5}{2} = 3.500000$$

después se calcula $f(x_1)$:

$$f(3.500000) = \frac{2}{3.500000-4} = -4.000000$$

y se multiplica $f(2) * f(3.500000) = 4.000000$; como es > 0, el intervalo para la segunda iteración sería: $a = 3.500000$ y $b = 5$.

Luego se calcula x_2 con la ecuación (2.4):

$$x_2 = \frac{3.500000+5}{2} = 4.250000$$

después se calcula $f(x_2)$:

$$f(4.250000) = \frac{2}{4.250000-4} = 8.000000$$

y se multiplica $f(3.500000) * f(4.250000) = -32.000000$; como es < 0, el intervalo para la tercera iteración sería: $a = 3.500000$ y $b = 4.250000$.

Para saber cuando dejar de iterar, se debe de cumplir que $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq 0.00001$ y que $\left|f(x_n)\right| \leq 0.00001$; lo cual no se cumple, ya que aunque el error si se va haciendo más pequeño, $f(x_n)$ en cada iteración va creciendo, (ver cuadro 2.2).

Para obtener $x_3, x_4, x_5, x_6, \ldots$; se repite el proceso anterior. Algunas iteraciones se muestran en el cuadro 2.2.

n	a	b	$f(a)$	$f(b)$	x_n	$f(x_n)$	$f(a) * f(x_n)$	ε
1	2.000000	5.000000	-1.000000	2.000000	3.500000	-4.000000	+	——
2	3.500000	5.000000	-4.000000	2.000000	4.250000	8.000000	−	0.176471
3	3.500000	4.250000	-4.000000	8.000000	3.875000	-16.000000	+	0.096774
4	3.875000	4.250000	-16.000000	8.000000	4.062500	32.000000	−	0.046154
5	3.875000	4.062500	-16.000000	32.000000	3.968750	-64.000000	+	0.023622
6	3.968750	4.062500	-64.000000	32.000000	4.015625	128.000000	−	0.011673
7	3.968750	4.015625	-64.000000	128.000000	3.992188	-256.016385	+	0.005871
8	3.992188	4.015625	-256.016385	128.000000	4.003907	511.901715	−	0.002927
9	3.992188	4.003907	-256.016385	511.901715	3.998048	-1024.590164	+	0.001465
10	3.998048	4.003907	-1024.590164	511.901715	4.000978	2044.989775	−	0.000732
11	3.998048	4.000978	-1024.590164	2044.989775	3.999513	-4106.776181	+	0.000366
12	3.999513	4.000978	-4106.776181	2044.989775	4.000246	8130.081301	−	0.000183
13	3.999513	4.000246	-4106.776181	8130.081301	3.999880	-16666.666667	+	0.000092
				⋮				

Cuadro 2.2: Ejemplo del método de bisección cuando encuentra una singularidad

A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.

Para la primera iteración el error relativo no se calcula, porque no se cuenta con una aproximación inicial.
Para la segunda iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{4.250000 - 3.500000}{4.250000}\right| = 0.176471$$

Para la tercera iteración el error relativo se calcula:

$$\left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{3.875000 - 4.250000}{3.875000}\right| = 0.096774$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.2.
El método de bisección no converge a una raíz.

El método de bisección puede atrapar una singularidad como si fuera una raíz, debido a que no reconoce la diferencia entre una raíz y una singularidad. Para evitar este problema es necesario graficar la función; pero si se llegara a aplicar el algoritmo 2.1.1, se tiene que verificar que se cumplan simultáneamente los siguientes criterios: $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq Tol$ y $|f(x_n)| \leq Tol$; para asegurar que se ha encontrado una aproximación de la raíz.

Ejemplo 2.1.4 Usar el método de bisección para encontrar una aproximación de la primera raíz positiva de la siguiente ecuación; en $[0,7]$, con una tolerancia de 10^{-5}; utilizar 7 decimales.

$$f(x) = -19(x-0.5)(x-1) + e^x - e^{-2x}$$

Antes de aplicar el método de bisección se grafica la función $f(x)$, obteniendo la figura 2.4; donde se puede observar que la función es continua en el intervalo $[0,7]$, pero en éste intervalo se encuentran tres raíces; por lo que se tiene que tomar otro intervalo que contenga exactamente una raíz [8]. Como se busca la primera raíz positiva se toma el intervalo $[0,1]$, empezando a aplicar el método de bisección, ya que se puede garantizar que va ha converger, a la solución buscada.

[8]Aunque el método también se puede aplicar cuando hay más de una raíz en $[a,b]$; pero cuando esto sucede no se puede garantizar que el método converja a la solución buscada, (ver ejercicio 2.2).

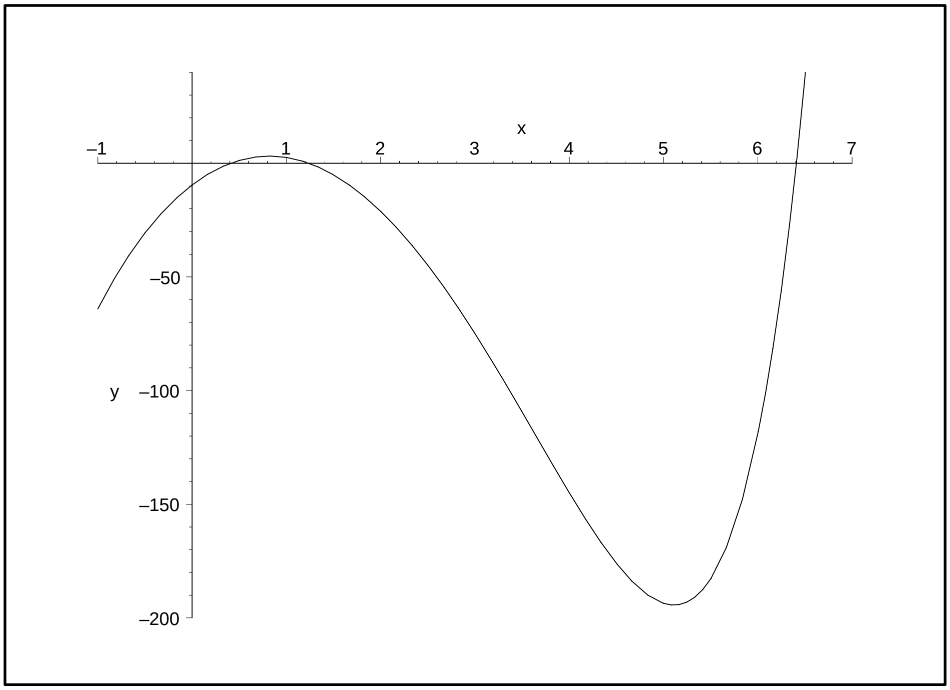


Figura 2.4 Gráfica de la ecuación $f(x) = -19(x-0.5)(x-1)+e^{x}-e^{-2x}$

Utilizando el algoritmo 2.1.1, se empieza calculando el valor de N :

$$
\begin{aligned}
N &\geq \frac{\log\left(\frac{1-0}{10^{-5}}\right)}{\log 2} + 10 \\
N &\geq \frac{\log\left(\frac{1}{10^{-5}}\right)}{\log 2} + 10(6) \\
N &\geq \frac{\log 100000}{0.301029995664} + 10(7) \\
N &\geq 16.609640474436 + 10(8) \\
N &\geq 26.609640474436
\end{aligned}
$$

Redondeando a enteros, se obtiene:

$$\Rightarrow N = 27$$

Después se calcula $f(0)$ y $f(1)$, obteniéndose:

$$f(0) = -19(0-0.5)(0-1)+e^{0}-e^{-2(0)} = -9.5000000$$

$$f(1) = -19(1-0.5)(1-1)+e^{1}-e^{-2(1)} = 2.5829465$$

Como $f(0) * f(1) < 0$, el intervalo es correcto ya que encierra una raíz (teorema 2.1.1); a continuación se calcula x_1 con la ecuación (2.4):

$$x_1 = \frac{0+1}{2} = 0.5000000$$

después se calcula $f(x_1)$:

$$f(0.5000000) = -19(0.5000000 - 0.5)(0.5000000 - 1) + e^{0.5000000} - e^{-2(0.5000000)}$$
$$= 1.2808418$$

y se multiplica $f(0) * f(0.5000000) = -12.1679971$; como es < 0, el intervalo para la segunda iteración sería: $a = 0$ y $b = 0.5000000$.
Luego se calcula x_2 con la ecuación (2.4):

$$x_2 = \frac{0 + 0.5000000}{2} = 0.2500000$$

después se calcula $f(x_2)$:

$$f(0.2500000) = -19(0.2500000 - 0.5)(0.2500000 - 1) + e^{0.2500000} - e^{-2(0.2500000)}$$
$$= -2.8850052$$

y se multiplica $f(0) * f(0.2500000) = 27.4075494$; como es > 0, el intervalo para la tercera iteración sería: $a = 0.2500000$ y $b = 0.5000000$.
Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq 0.00001$ y que $|f(x_n)| \leq 0.00001$; esto se cumple en la iteración 18: $0.0000094 \leq 0.00001$ y $0.0000090 \leq 0.00001$.
Para obtener $x_3, x_4, x_5, x_6, \ldots, x_{18}$; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.3.

n	a	b	$f(a)$	$f(b)$	x_n	$f(x_n)$	$f(a) * f(x_n)$	ε
1	0.0000000	1.0000000	-9.5000000	2.5829465	0.5000000	1.2808418	−	——
2	0.0000000	0.5000000	-9.5000000	1.2808418	0.2500000	-2.8850052	+	1.0000000
3	0.2500000	0.5000000	-2.8850052	1.2808418	0.3750000	-0.5017501	+	0.3333333
4	0.3750000	0.5000000	-0.5017501	1.2808418	0.4375000	0.4639995	−	0.1428571
5	0.3750000	0.4375000	-0.5017501	0.4639995	0.4062500	-0.0001867	+	0.0769231
6	0.4062500	0.4375000	-0.0001867	0.4639995	0.4218750	0.2365690	−	0.0370370
7	0.4062500	0.4218750	-0.0001867	0.2365690	0.4140625	0.1193580	−	0.0188679
8	0.4062500	0.4140625	-0.0001867	0.1193580	0.4101563	0.0598783	−	0.0095237
9	0.4062500	0.4101563	-0.0001867	0.0598783	0.4082032	0.0299195	−	0.0047846
10	0.4062500	0.4082032	-0.0001867	0.0299195	0.4072266	0.0148847	−	0.0023982
11	0.4062500	0.4072266	-0.0001867	0.0148847	0.4067383	0.0073535	−	0.0012005
12	0.4062500	0.4067383	-0.0001867	0.0073535	0.4064942	0.0035853	−	0.0006005
13	0.4062500	0.4064942	-0.0001867	0.0035853	0.4063721	0.0016996	−	0.0003005
14	0.4062500	0.4063721	-0.0001867	0.0016996	0.4063111	0.0007573	−	0.0001501
15	0.4062500	0.4063111	-0.0001867	0.0007573	0.4062806	0.0002861	−	0.0000751
16	0.4062500	0.4062806	-0.0001867	0.0002861	0.4062653	0.0000497	−	0.0000377
17	0.4062500	0.4062653	-0.0001867	0.0000497	0.4062577	-0.0000677	+	0.0000187
18	0.4062577	0.4062653	-0.0000677	0.0000497	0.4062615	-0.0000090	+	0.0000094

Cuadro 2.3: Ejemplo del método de bisección cuando converge

A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.

Para la primera iteración el error relativo no se calcula, porque no se cuenta con una aproximación inicial.
Para la segunda iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{0.2500000 - 0.5000000}{0.2500000}\right| = 1.0000000$$

Para la tercera iteración el error relativo se calcula:

$$\left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{0.3750000 - 0.2500000}{0.3750000}\right| = 0.3333333$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.3.
La primera raíz positiva es:

$$x \approx 0.4062615 \qquad \text{Con una tolerancia } \text{Tol}=0.00001$$

Comprobación:

$$\begin{aligned} f(0.4062615) &= -19(0.4062615 - 0.5)(0.4062615 - 1) + e^{0.4062615} - e^{-2(0.4062615)} \\ &= -0.0000090 \approx 0 \end{aligned}$$

No fue necesario realizar las 27 iteraciones, ya que se cumplieron antes simultáneamente los criterios de convergencia (2.5) y (2.6).

Ventajas del método de bisección.

1. Encuentra una raíz aún cuando la función no sea analítica.
2. Siempre encuentra una raíz de la función si se sabe que la raíz existe en un intervalo dado; es decir siempre converge.
3. Sirve como punto inicial para la aplicación de otro método.
4. Por el teorema 2.1.2 se sabe cuantas iteraciones se necesitan realizar con un error determinado Tol .
5. La longitud del subintervalo que contiene a la raíz tiende a cero, por lo que ésta puede tomarse como un criterio de aproximación a la raíz.

Desventajas del método de bisección.

1. Su convergencia es muy lenta comparada con la convergencia de otros métodos, por lo que se sugiere escoger un intervalo inicial $[a,b]$, tan pequeño como sea posible.
2. No toma en cuenta la magnitud de los valores de la función en las aproximaciones calculadas x_n, sólo toma en cuenta el signo de $f(x_n)$, lo que puede resultar que una aproximación intermedia, mejor que la respuesta final, pase desapercibida.
3. Puede atrapar una singularidad como si fuera una raíz, debido a que el método no distingue las raíces de las singularidades.
4. Una tarea importante que se debe de realizar antes de aplicar el método para garantizar la convergencia, es encontrar un intervalo que contenga la raíz buscada, así como verificar que la función sea continua en él. Lo anterior no significa que si no se cumple cualquiera de estos requisitos el método vaya a diverger.

2.2 Método de falsa posición

El *método de la falsa posición* pretende conjugar la seguridad del método de bisección para converger, con la rapidez del método de la secante (ver sección 2.4); también es denominado como: Regla falsa, Posición falsa o Interpolación Lineal. Comienza con un

intervalo $[a,b]$ que encierra a la raíz, es decir $f(a)$ y $f(b)$ son de signos opuestos, (teorema 2.1.1).
Es similar al método de bisección ya que consiste en generar subintervalos que encierren a la raíz; pero la aproximación de la raíz x_n no se obtiene con el punto medio, sino con la intersección de la recta secante a la curva que une a los puntos $(a, f(a))$ y $(b, f(b))$ con el eje x; proporcionando una mejor estimación de la raíz. El reemplazamiento de la curva por una línea recta da una "posición falsa" de la raíz, de aquí el nombre del método.
En la figura 2.5 se muestra la representación gráfica del método de la falsa posición; donde se puede observar que una vez iniciado el proceso iterativo, uno de los extremos del intervalo puede quedarse fijo.

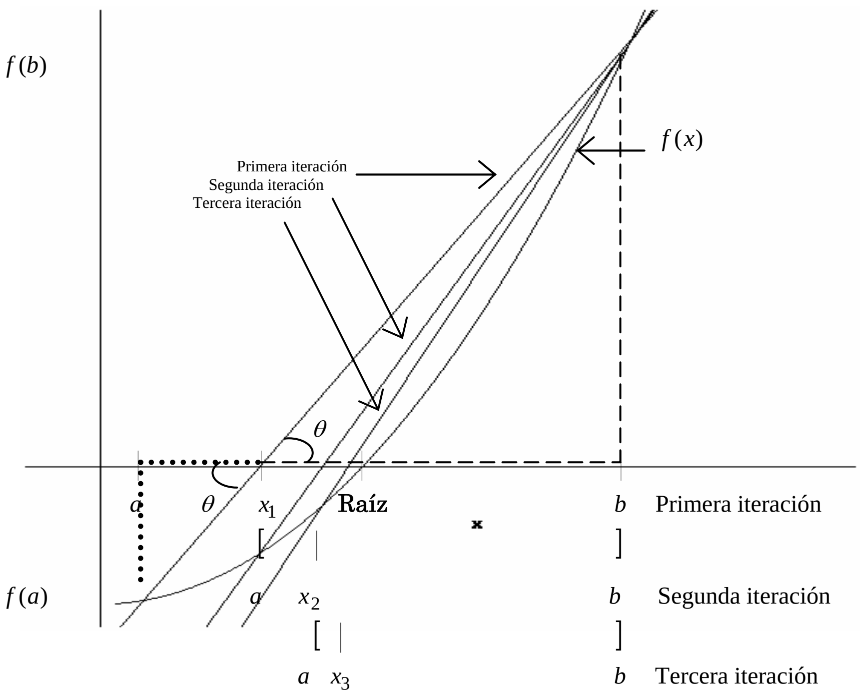


Figura 2.5 Representación gráfica del método de falsa posición

Con el uso de triángulos semejantes, la intersección de la recta secante con el eje x se puede calcular de la siguiente manera (Ver figura 2.5):

$$\cot\theta = \frac{\text{catetoadyacente}}{\text{catetoopuesto}}$$

$$\cot\theta = \frac{x_1 - a}{-f(a)} \qquad \cot\theta = \frac{b - x_1}{f(b)}$$

$$\Rightarrow \frac{x_1 - a}{-f(a)} = \frac{b-x_1}{f(b)}$$

$$f(b)(x_1 - a) = -f(a)(b - x_1)$$

$$x_1 f(b) - af(b) = -bf(a) + x_1 f(a)$$

$$x_1 f(b) - x_1 f(a) = -\text{bf(a)+af(b)}$$
$$x_1(f(b) - f(a)) = \text{af(b)}-\text{bf(a)}$$
$$x_1 = \frac{af(b)-bf(a)}{f(b)-f(a)}$$

El primer paso para aplicar el método de la falsa posición y garantizar que converja a la solución, es comprobar que f es continua en el intervalo $[a,b]$ y que en éste intervalo se encuentre una raíz $(f(a) * f(b) < 0)$, si se cumplen estos dos requisitos se garantiza la convergencia del método; pero no significa que si no se cumple cualquiera de estos el método vaya a diverger.
A continuación, se calcula la intersección con el eje x de la recta secante a la curva que pasa por los puntos $(a, f(a))$ y $(b, f(b))$, que será la primera aproximación de la raíz. Este valor esta dado por:

$$x_1 = \frac{af(b) - bf(a)}{f(b) - f(a)}$$

Para obtener la n-ésima intersección con el eje x de la recta secante a la curva, la ecuación sería:

$$x_n = \frac{af(b) - bf(a)}{f(b) - f(a)} \qquad \textit{fórmuladeiteraciónparaelmétododefalsaposición} \qquad (2.11)$$

A continuación se verifican los signos de $f(a) * f(x_n)$ y $f(x_n) * f(b)$, para saber en que subintervalo se encuentra la raíz (como en el método de bisección):

1. Si $f(a) * f(x_n) < 0$, entonces en el subintervalo $[a, x_n]$ se encuentra la raíz, renombrando $a = a$ y $b = x_n$.
2. Si $f(a) * f(x_n) > 0$, entonces en el subintervalo $[x_n, b]$ se encuentra la raíz, renombrando $a = x_n$ y $b = b$.
3. Si $f(a) * f(x_n) = 0$, entonces x_n es la raíz.

Si ocurre alguno de los dos primeros casos se calcula la intersección con el eje x de la recta secante a la curva con la ecuación (2.11), siendo ésta la aproximación más actualizada de la raíz.
Al igual que en el método de bisección se repite el proceso hasta que se cumplan los criterios de convergencia de las ecuaciones (2.5), (2.6) y (2.7).

ALGORITMO 2.2.1 Método de falsa posición. Para encontrar una aproximación de la raíz, de una ecuación $f(x) = 0$ donde f es una función continua en $[a,b]$ y $f(a) * f(b) < 0$. (Para garantizar la convergencia del método).
Datos: $f(x)$, el intervalo $[a,b]$, una tolerancia Tol y un número máximo de iteraciones N.
Resultados: Una raíz aproximada o un mensaje de falla.

- **Paso 1:** Hacer $n = 1$.
- **Paso 2:** Mientras $n \leq N$, repetir los pasos 3 a 14.

- **Paso 3:** Calcular $f(a)$ y $f(b)$.
- **Paso 4:** Si $n=1$ entonces realizar las siguientes comparaciones. De lo contrario CONTINUAR.
    * **Paso 5:** Si $f(a)*f(b)=0$ entonces IMPRIMIR “Una raíz de la ecuación dada es uno de los extremos del intervalo” y TERMINAR. De lo contrario CONTINUAR.
    * **Paso 6:** Si $f(a)*f(b)>0$, entonces IMPRIMIR “Intervalo incorrecto, no hay raíz o hay varias raíces” y TERMINAR. De lo contrario CONTINUAR.
    * **Paso 7:** Si $f(a)*f(b)<0$, entonces IMPRIMIR “Intervalo correcto” y CONTINUAR. De lo contrario TERMINAR.
- **Paso 8:** Calcular $x_n = \frac{af(b)-bf(a)}{f(b)-f(a)}$.
- **Paso 9:** Calcular $f(x_n)$.
- **Paso 10:** Si $f(a)*f(x_n)=0$ entonces IMPRIMIR “Una raíz de la ecuación dada es x_n” y TERMINAR. De lo contrario CONTINUAR.
- **Paso 11:** Si $f(a)*f(x_n)<0$, entonces hacer $a=a$ y $b=x_n$, de lo contrario hacer $a=x_n$ y $b=b$.
- **Paso 12:** Si $n>1$ entonces calcular $\varepsilon = \left|\frac{x_n - x_{n-1}}{x_n}\right|$ y realizar la siguiente comparación. De lo contrario CONTINUAR.
    * **Paso 13:** Si $\varepsilon \le Tol$ y $|f(x_n)| \le Tol$ entonces IMPRIMIR “Una raíz aproximada de la ecuación dada es x_n” y TERMINAR. De lo contrario CONTINUAR.
- **Paso 14:** Hacer $n=n+1$.

- **Paso 15:** IMPRIMIR mensaje de falla “El método no converge a una raíz” y TERMINAR.

Se puede verificar que el algoritmo del método de falsa posición es idéntico al de el método de bisección (algoritmo 2.1.1), con la excepción del **Paso 1** y de que en el primero se utiliza la ecuación (2.11) y en el segundo la (2.4).

Ejemplo 2.2.1 Usar el método de falsa posición para encontrar una aproximación de la raíz de la ecuación (2.10), ($f(x)=e^x - x^2 + 3x - 2$); en [0,1], con una tolerancia de 10^{-5}; utilizar 6 decimales.

Antes de aplicar el método de falsa posición se grafica la función $f(x)$, obteniendo la figura 2.2; donde se puede observar que la función es continua en el intervalo $[0,1]$ y que en éste intervalo se encuentra exactamente una raíz; por lo que se empieza a aplicar el método de falsa posición, ya que se puede garantizar que va a converger.

Aplicando el algoritmo 2.2.1 del método de falsa posición, se empieza calculando $f(0)$ y $f(1)$, para verificar que en el intervalo $[0,1]$ se encuentra una raíz, obteniéndose:

$$f(0) = e^{(0)} - (0)^2 + 3(0) - 2 = -1.000000$$

$$f(1) = e^{(1)} - (1)^2 + 3(1) - 2 = 2.718282$$

Como $f(0)*f(1)<0$, el intervalo es correcto ya que encierra una raíz (teorema 2.1.1); a continuación se calcula x_1 con la ecuación (2):

$$x_1 = \frac{(0)(2.718282)-(1)(-1.000000)}{(2.718282)-(-1.000000)} = 0.268941$$

después se calcula $f(x_1)$:

$$f(0.268941) = e^{(0.268941)} - (0.268941)^2 + 3(0.268941) - 2 = 0.043072$$

y se multiplica $f(0)*f(0.268941) = -0.043072$; como es <0, el intervalo para la segunda iteración sería: $a=0$ y $b=0.268941$.
Luego se calcula x_2 con la ecuación (2.11):

$$x_2 = \frac{(0)(0.043072)-(0.268941)(-1.000000)}{(0.043072)-(-1.000000)} = 0.257836$$

después se calcula $f(x_2)$:

$$f(0.257836) = e^{(0.257836)} - (0.257836)^2 + 3(0.257836) - 2 = 0.001155$$

y se multiplica $f(0)*f(0.257836) = -0.001155$; como es <0, el intervalo para la tercera iteración sería: $a=0$ y $b=0.257836$.
Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_n - x_{n-1}}{x_n}\right| \le 0.00001$ y que $\left|f(x_n)\right| \le 0.00001$; esto se cumple en la iteración 5: $0.000004 \le 0.00001$ y $0.000001 \le 0.00001$.
Para obtener x_3, x_4 y x_5; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.4.

n	a	b	$f(a)$	$f(b)$	x_n	$f(x_n)$	$f(a)*f(x_n)$	ε
1	0.000000	1.000000	-1.000000	2.718282	0.268941	0.043072	$-$	----
2	0.000000	0.268941	-1.000000	0.043072	0.257836	0.001155	$-$	0.043070
3	0.000000	0.257836	-1.000000	0.001155	0.257539	0.000033	$-$	0.001153
4	0.000000	0.257539	-1.000000	0.000033	0.257531	0.000003	$-$	0.000031
5	0.000000	0.257531	-1.000000	0.000003	**0.257530**	-0.000001	$+$	0.000004

Cuadro 2.4: Ejemplo del método de falsa posición cuando converge

A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.

Para la primera iteración el error relativo no se calcula, porque no se cuenta con una aproximación inicial.
Para la segunda iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{0.257836 - 0.268941}{0.257836}\right| = 0.043070$$

Para la tercera iteración el error relativo se calcula:

$$\left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{0.257539 - 0.257836}{0.257539}\right| = 0.001153$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.4.

$x \approx 0.257530$ Con una tolerancia $\varepsilon = 0.00001$

Comprobación :

$$f(0.257530) = e^{(0.257530)} - (0.257530)^2 + 3(0.257530) - 2 = -0.000001 \approx 0$$

En éste ejemplo se puede observar que el aplicar el método de falsa posición uno de los extremos del intervalo se puede quedar fijo; (ver figura 2.5).
Observando los ejemplos 2.1.1 y 2.2.1, se puede verificar que para la ecuación (2.10) el método de falsa posición converge más rápido que el método de bisección; aunque hay casos en donde éste último converge más rápido.

Ejemplo 2.2.2 Usar el método de falsa posición para encontrar una aproximación de la raíz de la siguiente ecuación; en $[0,1.5]$, con una tolerancia de 10^{-5}; utilizar 7 decimales.

$$f(x) = x^{10} - 1 \tag{2.12}$$

Antes de aplicar el método de falsa posición se grafica la función $f(x)$, obteniendo la figura 2.6; donde se puede observar que la función es continua en el intervalo $[0,1.5]$ y que en éste intervalo se encuentra exactamente una raíz; por lo que se empieza a aplicar el método de falsa posición, ya que se puede garantizar que va a converger.

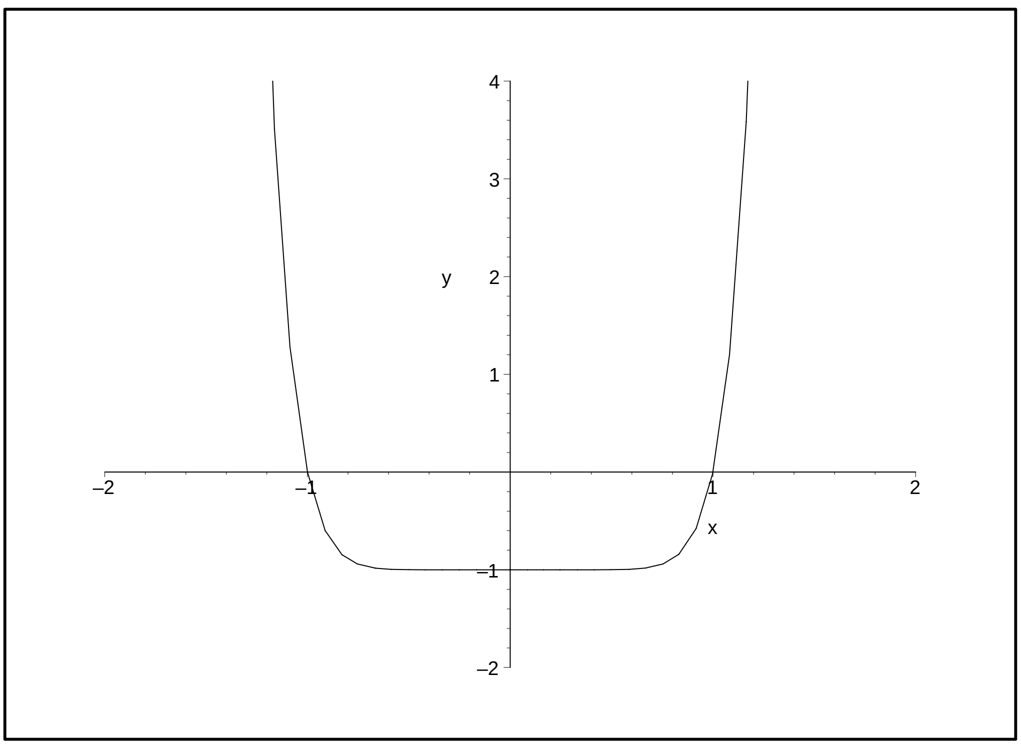


Figura 2.6 Gráfica de la ecuación $f(x) = x^{10} - 1$

Aplicando el algoritmo 2.1.1 del método de falsa posición, se empieza calculando $f(0)$ y $f(1.5)$, para verificar que en el intervalo $[0,1.5]$ se encuentra una raíz, obteniéndose:

$$f(0) = 0^{10} - 1 = -1.0000000$$

$$f(1.5) = 1.5^{10} - 1 = 56.6650391$$

Como $f(0) * f(1.5) < 0$, el intervalo es correcto ya que encierra una raíz (teorema 2.1.1); a continuación se calcula x_1 con la ecuación (2.11):

$$x_1 = \frac{(0)(56.6650391) - (1.5)(-1.0000000)}{(56.6650391) - (-1.0000000)} = 0.0260123$$

después se calcula $f(x_1)$:

$$f(0.0260123) = 0.0260123^{10} - 1 = -1.0000000$$

y se multiplica $f(0) * f(0.0260123) = 1.0000000$; como es > 0, el intervalo para la segunda iteración sería: $a = 0.0260123$ y $b = 1.5000000$.
Luego se calcula x_2 con la ecuación (2):

$$x_2 = \frac{(0.0260123)(56.6650391) - (1.5000000)(-1.0000000)}{(56.6650391) - (-1.0000000)} = 0.0515735$$

después se calcula $f(x_2)$:

$$f(0.0515735) = 0.0515735^{10} - 1 = -1.0000000$$

y se multiplica $f(0.0260123) * f(0.0515735) = 1.0000000$; como es > 0, el intervalo para la tercera iteración sería: $a = 0.0260123$ y $b = 0.0515735$.
Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq 0.00001$ y que $\left|f(x_n)\right| \leq 0.00001$; esto se cumple en la iteración 184: $0.0000009 \leq 0.00001$ y $0.0000001 \leq 0.00001$.
Para obtener $x_3, x_4, x_5, x_6, \ldots, x_{184}$; se repite el proceso anterior. Algunas iteraciones se muestran en el cuadro 2.5.
A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.
Para la primera iteración el error relativo no se calcula, porque no se cuenta con una aproximación inicial.
Para la segunda iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{0.0515735 - 0.0260123}{0.0515735}\right| = 0.4956266$$

Para la tercera iteración el error relativo se calcula:

$$\left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{0.0766914 - 0.0515735}{0.0766914}\right| = 0.3275191$$

n	a	b	$f(a)$	$f(b)$	x_n	$f(x_n)$	$f(a)*f(x_n)$	ε
1	0.0000000	1.5000000	-1.0000000	56.6650391	0.0260123	-1.0000000	+	——
2	0.0260123	1.5000000	-1.0000000	56.6650391	0.0515735	-1.0000000	+	0.4956266
3	0.0515735	1.5000000	-1.0000000	56.6650391	0.0766914	-1.0000000	+	0.3275191
4	0.0766914	1.5000000	-1.0000000	56.6650391	0.1013737	-1.0000000	+	0.2434783
5	0.1013737	1.5000000	-1.0000000	56.6650391	0.1256280	-1.0000000	+	0.1930644
6	0.1256280	1.5000000	-1.0000000	56.6650391	0.1494617	-1.0000000	+	0.1594636
7	0.1494617	1.5000000	-1.0000000	56.6650391	0.1728821	-1.0000000	+	0.1354704
8	0.1728821	1.5000000	-1.0000000	56.6650391	0.1958964	-0.9999999	+	0.1174820
9	0.1958964	1.5000000	-0.9999999	56.6650391	0.2185115	-0.9999998	+	0.1034962
10	0.2185115	1.5000000	-0.9999998	56.6650391	0.2407345	-0.9999993	+	0.0923133
11	0.2407345	1.5000000	-0.9999993	56.6650391	0.2625721	-0.9999984	+	0.0831680
12	0.2625721	1.5000000	-0.9999984	56.6650391	0.2840310	-0.9999966	+	0.0755513
13	0.2840310	1.5000000	-0.9999966	56.6650391	0.3051177	-0.9999930	+	0.0691101
14	0.3051177	1.5000000	-0.9999930	56.6650391	0.3258386	-0.9999865	+	0.0635925
15	0.3258386	1.5000000	-0.9999865	56.6650391	0.3462001	-0.9999753	+	0.0588143
16	0.3462001	1.5000000	-0.9999753	56.6650391	0.3662083	-0.9999566	+	0.0546361
17	0.3662083	1.5000000	-0.9999566	56.6650391	0.3858691	-0.9999268	+	0.0509520
18	0.3858691	1.5000000	-0.9999268	56.6650391	0.4051884	-0.9998807	+	0.0476798
19	0.4051884	1.5000000	-0.9998807	56.6650391	0.4241719	-0.9998115	+	0.0447543
20	0.4241719	1.5000000	-0.9998115	56.6650391	0.4428249	-0.9997101	+	0.0421227
21	0.4428249	1.5000000	-0.9997101	56.6650391	0.4611527	-0.9995650	+	0.0397435
22	0.4611527	1.5000000	-0.9995650	56.6650391	0.4791602	-0.9993620	+	0.0375814
23	0.4791602	1.5000000	-0.9993620	56.6650391	0.4968520	-0.9990832	+	0.0356078
24	0.4968520	1.5000000	-0.9990832	56.6650391	0.5142324	-0.9987070	+	0.0337987
25	0.5142324	1.5000000	-0.9987070	56.6650391	0.5313054	-0.9982076	+	0.0321341
26	0.5313054	1.5000000	-0.9982076	56.6650391	0.5480745	-0.9975543	+	0.0305964
27	0.5480745	1.5000000	-0.9975543	56.6650391	0.5645427	-0.9967117	+	0.0291709
28	0.5645427	1.5000000	-0.9967117	56.6650391	0.5807125	-0.9956387	+	0.0278448
29	0.5807125	1.5000000	-0.9956387	56.6650391	0.5965860	-0.9942888	+	0.0266072
30	0.5965860	1.5000000	-0.9942888	56.6650391	0.6121646	-0.9926094	+	0.0254484
				⋮				
184	0.9999990	1.5000000	-0.0000100	56.6650391	0.9999991	-0.0000090	+	0.0000001

Cuadro 2.5: Ejemplo del método de falsa posición cuando converge muy lentamente

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.5.

$$x \approx 0.9999991 \qquad \text{Con un } \varepsilon = 0.00001$$

Comprobación:

$$f(0.9999991) = 0.9999991^{10} - 1 = -0.0000090 \approx 0$$

Con el ejemplo 2.2.2 se puede comprobar que el aplicar el método de falsa posición uno de los extremos del intervalo se puede quedar fijo; (ver figura 1); así mismo se puede observar que para la ecuación (2.12) el método de falsa posición funciona de manera deficiente, ya que converge muy lentamente. El método de bisección ofrece mejores resultados para esta ecuación, (ver ejercicio 2.1); ya que converge en 20 iteraciones.

Ejemplo 2.2.3 Usar el método de falsa posición para encontrar una aproximación de la primera raíz negativa de la siguiente ecuación; en [-4,4], con una tolerancia de 10^{-5}; utilizar 6 decimales.

$$f(x) = \tan(x)$$

Antes de aplicar el método de falsa posición se grafica la función $f(x)$, obteniendo la figura 2.7; donde se puede observar que la función no es continua en el intervalo $[-4,4]$, y que en éste intervalo se encuentran tres raíces; por lo que se tiene que tomar otro intervalo que contenga exactamente una raíz y que la función sea continua en él. Como se busca la primera raíz negativa se toma el intervalo $[-3.5,-2.5]$, empezando a aplicar el método de la falsa posición, ya que se puede garantizar que va ha converger, a la solución buscada.

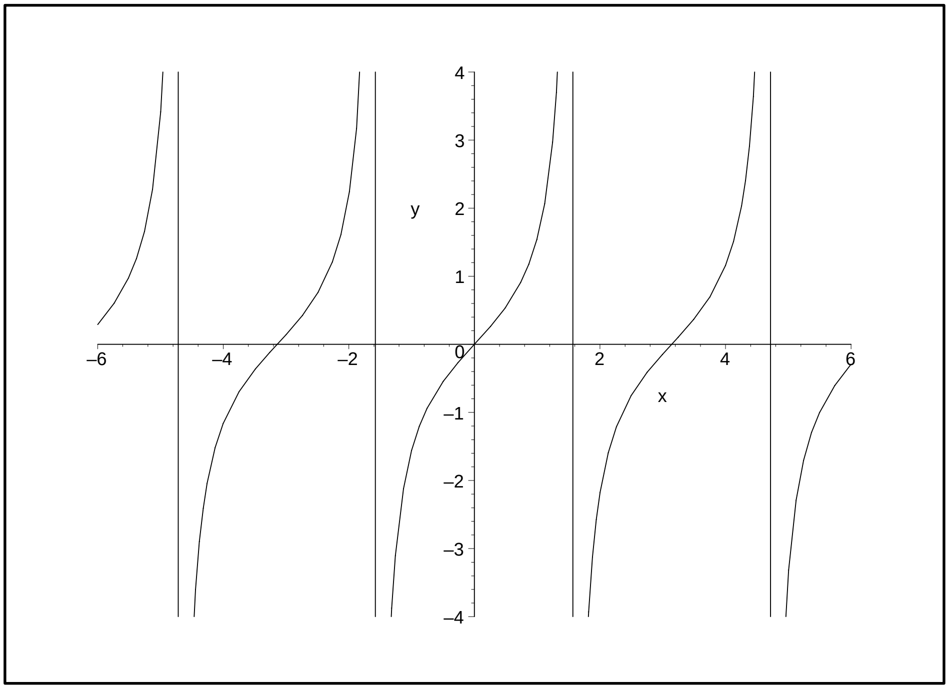


Figure 2.7 Gráfica de la ecuación $f(x)=\tan(x)$

Aplicando el algoritmo 2.2.1 del método de falsa posición, se empieza calculando $f(-3.5)$ y $f(-2.5)$, para verificar que en el intervalo $[-3.5,-2.5]$ se encuentra una raíz, obteniéndose:

$$f(-3.5)=\tan(-3.5)=-0.374586$$
$$f(-2.5)=\tan(-2.5)=0.747022$$

Como $f(-3.5)*f(-2.5)<0$, el intervalo es correcto ya que encierra una raíz (teorema 2.2.1); a continuación se calcula x_1 con la ecuación (2.11):

$$x_1=\frac{(-3.500000)(0.747022)-(-2.500000)(-0.374586)}{(0.747022)-(-0.374586)}=-3.166028$$

después se calcula $f(x_1)$:

$$f(-3.166028)=\tan(-3.166028)=-0.024440$$

y se multiplica $f(-3.500000)*f(-3.166028)=0.009155$; como es >0, el intervalo para la segunda iteración sería: $a=-3.166028$ y $b=-2.500000$.
Luego se calcula x_2 con la ecuación (2.11):

$$x_2 = \frac{(-3.166028)(0.747022)-(-2.500000)(-0.024440)}{(0.747022)-(-0.024440)} = -3.144928$$

después se calcula $f(x_2)$:

$$f(-3.144928) = \tan(-3.144928) = -0.003335$$

y se multiplica $f(-3.166028) * f(-3.144928) = 0.000082$; como es >0, el intervalo para la tercera iteración sería: $a = -3.144928$ y $b = -2.500000$.

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq 0.00001$ y que $|f(x_n)| \leq 0.00001$; esto se cumple en la iteración 6: $0.000003 \leq 0.00001$ y $0.000001 \leq 0.00001$.

Para obtener x_3, x_4, x_5, y x_6; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.6.

A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.

Para la primera iteración el error relativo no se calcula, porque no se cuenta con una aproximación inicial.

Para la segunda iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{-3.144928-(-3.166028)}{-3.144928}\right| = 0.006709$$

Para la tercera iteración el error relativo se calcula:

$$\left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{-3.142062-(-3.144928)}{-3.142062}\right| = 0.000912$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.6.

$$x \approx -3.141594 \qquad \text{Conun } \varepsilon = 0.00001$$

Comprobación:

$$\text{f(-3.141594)} = \tan(-3.141594) = -0.000001 \approx 0$$

Ventajas del método de falsa posición.

1. Siempre encuentra una raíz de la función si se sabe que la raíz existe en un intervalo dado; es decir siempre converge.
2. Casi siempre converge más rápido que el método de bisección.

Desventajas del método de falsa posición.

1. La longitud del subintervalo que contiene a la raíz en general no tiende a cero, debido a que uno de los extremos de los subintervalos se aproxima a la raíz, mientras el otro puede permanecer fijo; es decir la longitud del subintervalo $[a,b]$ no puede tomarse como un criterio de aproximación a la raíz.
2. Una tarea importante que se debe de realizar antes de aplicar el método es encontrar un intervalo que contenga la raíz buscada, así como verificar que la función sea continua en él. Lo anterior no significa que si no se cumple cualquiera de estos requisitos el método vaya a diverger.

2.3 Método de Newton

El *método de Newton* es una de las técnicas numéricas más poderosas y conocidas, para determinar las raíces de una ecuación. Comienza con una aproximación inicial [9] x_0 a la raíz, a continuación se traza una recta tangente a la curva en el punto $(x_0, f(x_0))$. La intersección de la recta tangente con el eje x, se denota como x_1 y se considera una mejor aproximación de la raíz; después se traza otra recta tangente a la curva en el punto $(x_1, f(x_1))$. La nueva intersección de la recta tangente con el eje x, se denota como x_2 y se considera una mejor aproximación de la raíz. El proceso se repite hasta que se cumplan los criterios de convergencia de las ecuaciones (2.5), (2.6) y (2.7).

El método de Newton, se puede aplicar para hallar raíces complejas, siempre y cuando el valor inicial x_0 sea un número complejo.

En la figura 2.8 se muestra la representación gráfica del método de Newton.

[9]El método casi siempre converge muy rápidamente si x_0 está "suficientemente cerca" de la raíz. En la práctica es un tema difícil determinar lo que es "suficientemente cerca", por lo que muchas veces se utiliza el método de bisección para encontrar una aproximación inicial x_0. Así mismo si x_0 "no esta lo suficientemente cerca" de la raíz, el método quizá no converja a la raíz; pero esto no siempre es así.

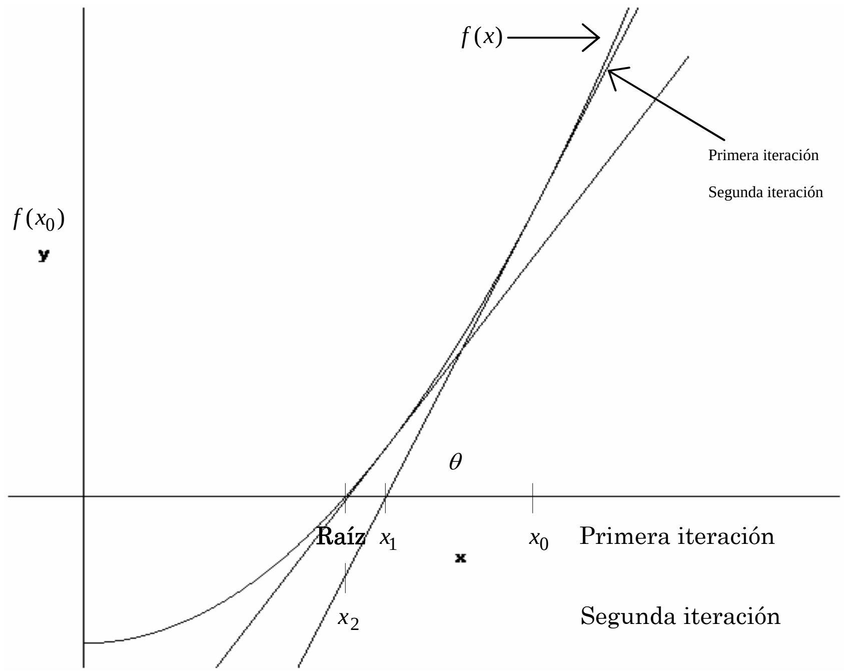


Figura 2.8 Representación gráfica del método de Newton

La intersección de la recta tangente con el eje x se puede calcular de la siguiente manera (Ver figura 2.8):

$$\tan\theta = \frac{\text{catetoopuesto}}{\text{catetoadyacente}}$$

$$\tan\theta = \frac{f(x_0)}{x_0 - x_1} = f'(x_0)$$

Se iguala a la primera derivada de x_0, ya que ésta es equivalente a la pendiente:

$$f'(x_0) = \frac{f(x_0) - 0}{x_0 - x_1}$$

$$\Rightarrow \frac{f(x_0)}{x_0 - x_1} = f'(x_0)$$

$$(x_0 - x_1) f'(x_0) = f(x_0)$$

$$f(x_0) = (x_0 - x_1) f'(x_0)$$

$$f(x_0) = x_0 f'(x_0) - x_1 f'(x_0)$$

$$x_1 f'(x_0) = x_0 f'(x_0) - f(x_0)$$

$$x_1 = \frac{x_0 f'(x_0) - f(x_0)}{f'(x_0)}$$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$

Para obtener la n-ésima intersección con el eje x de la recta tangente a la curva, la ecuación sería:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \qquad fórmula de iteración para el método de Newton \tag{2.13}$$

Un método alternativo para deducir el método de Newton, es a partir de la serie de Taylor.
Al utilizar la serie de Taylor de $f(x_{n+1})$ en torno a una aproximación x_n de la raíz, la ecuación se puede escribir como:

$$f(x_{n+1}) = f(x_n) + f'(x_n)(x_{n+1} - x_n) + \frac{f''(x_n)(x_{n+1} - x_n)^2}{2!} + \ldots \tag{2.14}$$

Truncando la serie de Taylor después del término de la primera derivada, se obtiene:

$$f(x_{n+1}) \approx f(x_n) + f'(x_n)(x_{n+1} - x_n)$$

En la intersección con el eje x, $f(x_{n+1})$ debe ser igual o muy próximo a cero, entonces:

$$0 \approx f(x_n) + f'(x_n)(x_{n+1} - x_n)$$

de donde se despeja a x_{n+1}:

$$0 \approx f(x_n) + x_{n+1} f'(x_n) - x_n f'(x_n)$$

$$x_{n+1} f'(x_n) \approx x_n f'(x_n) - f(x_n)$$

$$x_{n+1} \approx \tfrac{x_n f'(x_n) - f(x_n)}{f'(x_n)}$$

$$x_{n+1} \approx x_n - \tfrac{f(x_n)}{f'(x_n)}$$

Obteniendo la ecuación (2.13).

Algoritmo 2.3.1 Método de Newton. Para encontrar una aproximación de la raíz, de una ecuación $f(x) = 0$.

Datos: $f(x)$, $f'(x)$, una aproximación inicial x_0, una tolerancia Tol y un número máximo de iteraciones N.

Resultados: Una raíz aproximada o un mensaje de falla.

- **Paso 1:** Hacer $n = 0$.
- **Paso 2:** Mientras $n \le N$, repetir los pasos 3 a 11.
  - **Paso 3:** Si $n = 0$ entonces hacer lo siguiente. De lo contrario CONTINUAR.
    - **Paso 4:** Calcular $f(x_n)$
    - **Paso 5:** Si $f(x_n) = 0$, entonces IMPRIMIR "Una raíz de la ecuación dada es x_n" y TERMINAR. De lo contrario CONTINUAR.
  - **Paso 6:** Calcular $x_{n+1} = x_n - \tfrac{f(x_n)}{f'(x_n)}$
  - **Paso 7:** Calcular $f(x_{n+1})$

- **Paso 8:** Si $f(x_{n+1})=0$ entonces IMPRIMIR "Una raíz de la ecuación dada es x_{n+1}" y TERMINAR. De lo contrario CONTINUAR.
- **Paso 9:** Si $n>0$ entonces calcular $\varepsilon=\left|\frac{x_n-x_{n-1}}{x_n}\right|$ y realizar la siguiente comparación. De lo contrario CONTINUAR.
  * **Paso 10:** Si $\varepsilon\leq Tol$ y $\left|f(x_n)\right|\leq Tol$ entonces IMPRIMIR "Una raíz aproximada de la ecuación dada es x_n" y TERMINAR. De lo contrario CONTINUAR.
- **Paso 11:** Hacer $n=n+1$.

- **Paso 12:** IMPRIMIR mensaje de falla "El método no converge a una raíz" y TERMINAR.

En algunos casos lo más difícil al aplicar el método de Newton consiste en el cálculo de la derivada de la función $f(x)$ y la obtención del valor inicial x_0.

Se tiene que tener en cuenta que la naturaleza de la función puede originar dificultades, llegando incluso hacer que el método no converja; algunos casos son:

1. **Caso 1:** Si la aproximación inicial x_0, es un punto de inflexión (es decir $f''(x)=0$), las iteraciones divergen progresivamente de la raíz. (Ver figura 2.9).
2. **Caso 2:** Si el método oscila en los alrededores de un máximo o un mínimo local. Tales oscilaciones pueden alcanzar una pendiente cercana a cero, en cuyo caso la solución se aleja del área de interés. (Ver figura 2.10).
3. **Caso 3:** Si el valor inicial cercano a la raíz, salta a otra raíz muy distante de la anterior. Ésta tendencia a alejarse del área de interés se debe a que se encuentran pendientes cercanas a cero. (Ver figura 2.11).
4. **Caso 4:** Una pendiente nula provoca una división entre cero (lo cual geométricamente representa a una tangente horizontal que jamás corta al eje x). (Ver figura 2.12).

Análisis de convergencia para el método de Newton

Para analizar la convergencia del método de Newton, se utiliza la serie de Taylor, tomando en cuenta que el error en la n-ésima iteración sería: $\varepsilon_n=x-x_n$.

$$\Rightarrow \varepsilon_{n+1}=x-x_{n+1}$$

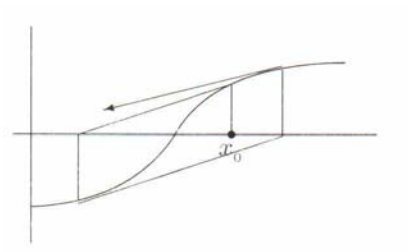
$$\varepsilon_{n-1}=x-x_{n-1}$$

Se sustituye la ecuación (2.13), en ε_{n+1}, obteniéndose:

$$\varepsilon_{n+1}=x-\left[x_n-\frac{f(x_n)}{f'(x_n)}\right]$$

$$\varepsilon_{n+1}=x-x_n+\frac{f(x_n)}{f'(x_n)}$$

como $\varepsilon_n=x-x_n$, entonces:



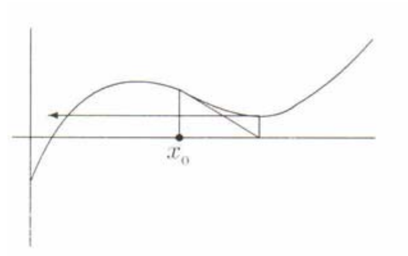
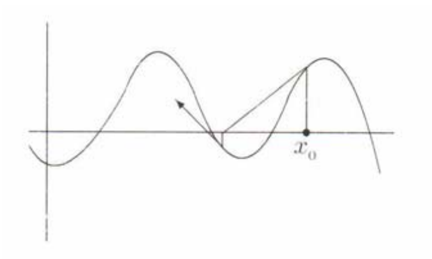


Figura 9. Caso 1 **Figura 10. Caso 2**



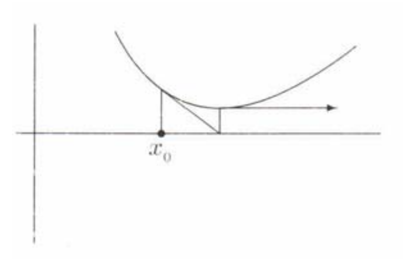


Figura 11. Casos 3 **Figura 12. Caso 4**

$$\varepsilon_{n+1} = \varepsilon_n + \frac{f(x_n)}{f'(x_n)}$$

$$\varepsilon_{n+1} = \frac{\varepsilon_n f'(x_n) + f(x_n)}{f'(x_n)} \tag{2.15}$$

Al utilizar la serie de Taylor de $f(x)$ en torno a una aproximación x_n de la raíz, la ecuación se puede escribir como:

$$f(x) = f(x_n) + f'(x_n)(x - x_n) + \frac{f''(x_n)(x - x_n)^2}{2!} + \ldots$$

Truncando la serie de Taylor después del término de la segunda derivada, se obtiene:

$$f(x) \approx f(x_n) + f'(x_n)(x - x_n) + \frac{f''(x_n)(x - x_n)^2}{2}$$

En la intersección con el eje x, $f(x)$ debe ser igual a cero, entonces:

$$0 \approx f(x_n) + f'(x_n)(x - x_n) + \frac{f''(x_n)(x - x_n)^2}{2}$$

como $\varepsilon_n = x - x_n$, entonces:

$$0 \approx f(x_n) + f'(x_n)(\varepsilon_n) + \frac{f''(x_n)}{2}(\varepsilon_n)^2$$

$$f(x_n) + f'(x_n)(\varepsilon_n) \approx -\frac{f''(x_n)}{2}(\varepsilon_n)^2$$

Se multiplica la ecuación anterior por $\frac{1}{f'(x_n)}$:

$$\Rightarrow \frac{f(x_n)+f'(x_n)(\varepsilon_n)}{f'(x_n)} \approx -\frac{f''(x_n)}{2f'(x_n)}(\varepsilon_n)^2$$

obteniendo del lado derecho de la ecuación anterior ε_{n+1}, (ver ecuación (2.15)):

$$\varepsilon_{n+1} \approx -\frac{f''(x_n)}{2f'(x_n)}(\varepsilon_n)^2$$

es decir el error en una determinada iteración es proporcional al cuadrado del error anterior. Esto significa que si en una determinada iteración se tienen n cifras decimales exactas, en la siguiente iteración se tendrán $2n$ cifras decimales exactas. A este comportamiento se le llama convergencia cuadrática.

Teorema 2.3.1 Sea f una función continua y dos veces derivable en el intervalo $[a,b]$. Si $f(a)*f(b)<0$ y sus dos primeras derivadas $f'(x)$ y $f''(x)$ no se anulan en $[a,b]$ existe una única raíz de la ecuación $f(x)=0$ en dicho intervalo y se puede garantizar la convergencia del método de Newton tomando como valor inicial x_0 el extremo del intervalo en el que la función y su derivada tienen el mismo signo.

Aplicando el teorema 2.3.1 se tiene que: el método de Newton converge a la única raíz x de la ecuación $f(x)$, si f es continua y dos veces derivable en el intervalo $[a,b]$, y si $f(a)*f(b)<0$ y sus dos primeras derivadas $f'(x)$ y $f''(x)$ no se anulan en $[a,b]$, tomando como valor inicial:

$$x_0 = \begin{cases} a & \text{si} f(a)f''(a)>0 \\ b & \text{si} f(b)f''(b)>0 \end{cases} \tag{2.16}$$

es decir el extremo en que la función tiene el mismo signo que su segunda derivada. (Ver figura 2.13)

Ejemplo 2.3.1 Usar el método de Newton para encontrar una aproximación de la raíz de la ecuación (2.10), ($f(x)=e^x-x^2+3x-2$); con $x_0=0.5$, con una tolerancia de 10^{-5}; utilizar 6 decimales.

Antes de aplicar el método de Newton se grafica la función $f(x)$, obteniendo la figura ?; donde se puede observar que la función es continua y que el punto inicial se encuentra cerca de la raíz; por lo que se empieza a aplicar el método.

Se calcula la primera derivada $f'(x)$, (obtenida con Maple):

$$f'(x)=e^x-2x+3$$

Utilizando el algoritmo 2.3.1, se empieza calculando $f(0.5)$ y $f'(0.5)$ obteniéndose:

$$f(0.5) = e^{0.5} - 0.5^2 + 3(0.5) - 2 = 0.898721$$
$$f'(0.5) = e^{0.5} - 2(0.5) + 3 = 3.648721$$

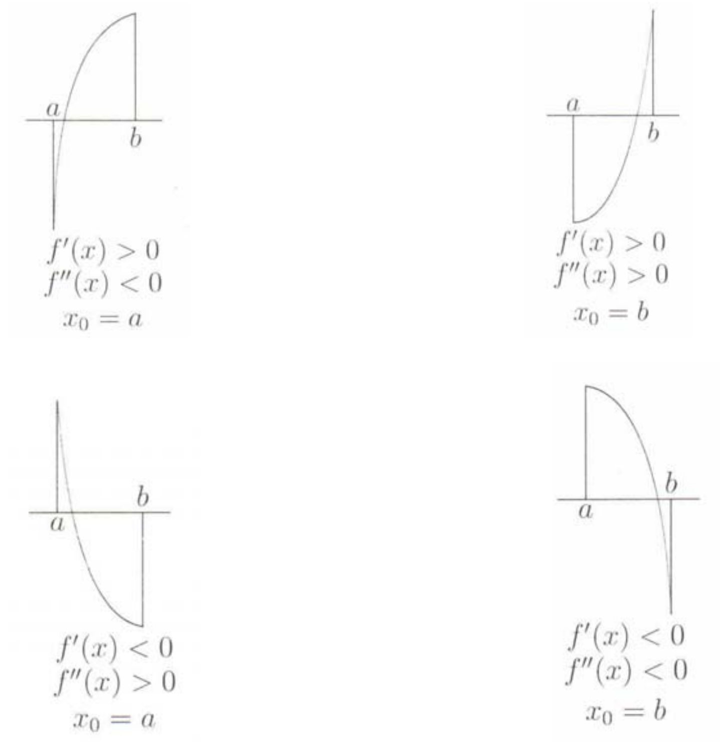


Figure 7. Casos posibles para escoger x_0, cuando se tiene un intervalo $[a,b]$

A continuación se calcula x_1 con la ecuación (2.13):

$$x_1 = 0.5 - \frac{0.898721}{3.648721} = 0.253689$$

después se calcula $f(x_1)$ y $f'(x_1)$:

$$f(0.253689) = e^{0.253689} - 0.253689^2 + 3(0.253689) - 2 = -0.014520$$
$$f'(0.253689) = e^{0.253689} - 2(0.253689) + 3 = 3.781393$$

Luego se calcula x_2 con la ecuación (2.13):

$$x_2 = 0.253689 - \frac{-0.014520}{3.781393} = 0.257529$$

después se calcula $f(x_2)$ y $f'(x_2)$:

$$f(0.257529) = e^{0.257529} - 0.257529^2 + 3(0.257529) - 2 = -0.000005$$
$$f'(0.257529) = e^{0.257529} - 2(0.257529) + 3 = 3.778671$$

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_n - x_{n-1}}{x_n}\right| \le 0.00001$ y que $\left|f(x_n)\right| \le 0.00001$; esto se cumple en la iteración 3: $0.000004 \le 0.00001$ y $m0.000001 \le 0.00001$.

Para obtener x_3; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.7.

n	x_n	$f(x_n)$	$f'(x_n)$	ε
0	0.500000	0.898721	3.648721	——
1	0.253689	-0.014520	3.781393	0.970917
2	0.257529	-0.000005	3.778671	0.014911
3	**0.257530**	-0.000001	3.778671	0.000004

Cuadro 2.7: Ejemplo del método de Newton cuando converge

A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.

Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_1 - x_0}{x_1}\right| = \left|\frac{0.253689 - 0.500000}{0.253689}\right| = 0.970917$$

Para la segunda iteración el error relativo se calcula:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{0.257529 - 0.253689}{0.257529}\right| = 0.014911$$

Para obtener el error relativo de la última iteración se repite el proceso anterior. Los errores se muestran en el cuadro 2.7.

$$x \approx 0.257530 \qquad \text{Con una tolerancia } \text{Tol}=0.00001$$

Comprobación:

$$\text{f}(0.257530)=\text{e}^{(0.257530)} - (0.257530)^2 + 3(0.257530) - 2 = -0.000001 \approx 0$$

Observando los ejemplos 2.1.1, 2.2.1 y 2.3.1, se puede verificar que para la ecuación (2.10) el método de Newton converge más rápido que los métodos de falsa posición y bisección; aunque hay casos en donde estos últimos convergen más rápido.

Ejemplo 2.3.2 Usar el método de Newton para encontrar una aproximación de la raíz positiva más pequeña de la siguiente ecuación; con $x_0 = 3.6$, con una tolerancia de 10^{-6}; utilizar 7 decimales.

$$f(x) = \tan(x) - 0.5x$$

Antes de aplicar el método de Newton se grafica la función $f(x)$, obteniendo la figura 2.14; donde se puede observar que la función no es continua y que el punto inicial no se encuentra tan cerca de la raíz.

Supóngase que no se grafica la función y se empieza aplicar el método.

Se calcula la primera derivada $f'(x)$, (obtenida con Maple):

$$f'(x) = 0.5 + \tan(x)^2$$

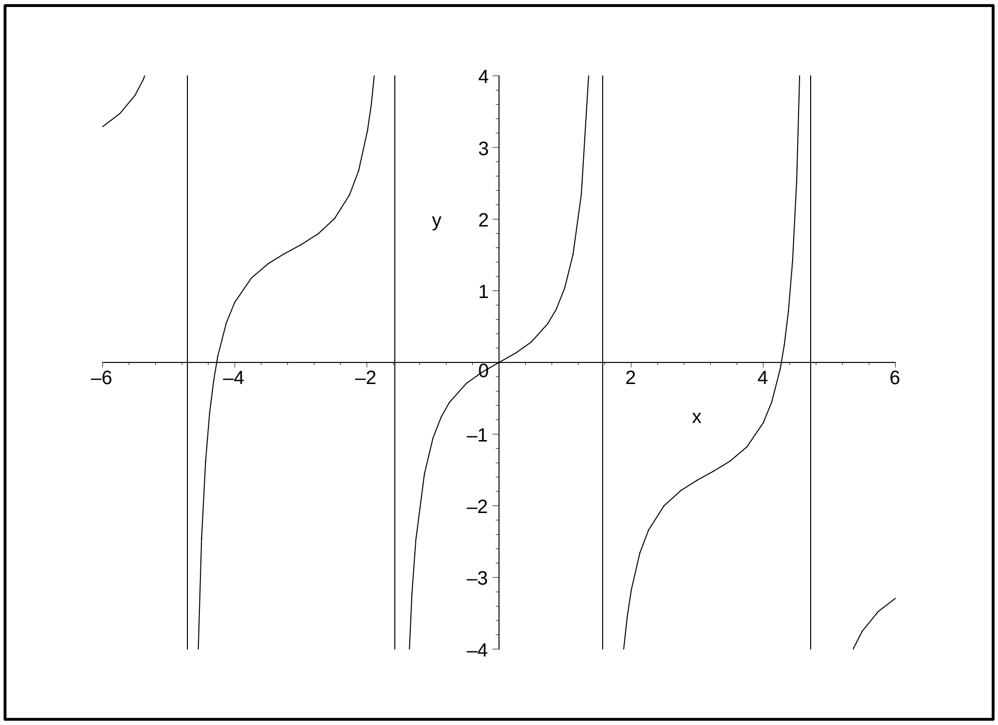


Figura 2.14. Gráfica de la ecuación $f(x)=\tan(x)-0.5x$

Utilizando el algoritmo 2.3.1, se empieza calculando $f(3.6)$ y $f'(3.6)$ obteniéndose:

$$f(3.6)=\tan(3.6)-0.5(3.6)=-1.3065333$$

$$f'(3.6)=0.5+\tan(3.6)^2=0.7435094$$

A continuación se calcula x_1 con la ecuación (2.13):

$$x_1=3.6-\frac{-1.3065333}{0.7435094}=5.3572519$$

después se calcula $f(x_1)$ y $f'(x_1)$:

$$f(5.3572519)=\tan(5.3572519)-0.5(5.3572519)=-4.0081833$$

$$f'(5.3572519)=0.5+\tan(5.3572519)^2=2.2677228$$

Luego se calcula x_2 con la ecuación (2.13):

$$x_2=5.3572519-\frac{-4.0081833}{2.2677228}=7.124745$$

después se calcula $f(x_2)$ y $f'(x_2)$:

$$f(7.1247445)=\tan(7.1247445)-0.5(7.1247445)=-2.4432340$$

$$f'(7.1247445)=0.5+\tan(7.1247445)^2=1.7524705$$

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_n - x_{n-1}}{x_n}\right| \leq 0.00001$ y que $|f(x_n)| \leq 0.00001$; esto nunca se cumple, ya que el método diverge.

Para obtener $x_3, x_4, x_5, x_6, \ldots$; se repite el proceso anterior. Algunas iteraciones se muestran en el cuadro 2.8.

n	x_n	$f(x_n)$	$f'(x_n)$	ε
0	3.6000000	-1.3065333	0.7435094	——
1	5.3572519	-4.0081833	2.2677228	0.3280137
2	7.1247445	-2.4432340	1.7524705	0.2480780
3	8.5189101	-5.5349126	2.1267919	0.1636554
4	11.1213803	-13.4674562	63.0169493	0.2340060
5	11.3350919	-8.4988493	8.5162789	0.0188540
6	12.3330453	-6.4041764	0.5564793	0.0809170
7	23.8414274	-15.4051057	12.6409875	0.4827052
8	25.0600906	-12.6028240	0.5052967	0.0486296
9	50.0015240	-25.2710266	0.5730430	0.4988135
10	94.1012279	-47.1982239	0.5217887	0.4686411
11	184.5558964	-93.3036189	1.5520003	0.4901207
12	244.6741959	-122.7249969	0.6504656	0.2457076
13	433.3467358	-216.8688528	0.5382144	0.4353847
14	836.2881153	-417.4233796	1.0193768	0.4818212
15	1245.7769124	-630.2274064	54.3601902	0.3287015
		$\vdots$		

Cuadro 2.8: Ejemplo del método de Newton cuando no converge

A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.

Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_1 - x_0}{x_1}\right| = \left|\frac{5.3572519 - 3.6000000}{5.3572518}\right| = 0.3280137$$

Para la segunda iteración el error relativo se calcula:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{7.1247445 - 5.3572519}{7.1247445}\right| = 0.2480780$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.8.

El método de Newton no converge a una raíz.

Observando el ejemplo 2.3.2, se puede verificar que quizá el método de Newton no converja a la raíz debido a que la aproximación inicial x_0, “no esta lo suficientemente cerca” de la raíz. Tal vez, sería mejor empezar con $x_0 = 4$, (ver ejercicio 2.5). Pero esto no siempre es así; es decir habrá casos en donde el método converja aunque la aproximación inicial “no este lo suficientemente cerca” de la raíz, (ver ejercicio 2.6).

Ejemplo 2.3.3 Usar el método de Newton para encontrar una aproximación de la raíz de la siguiente ecuación; en $[-0.65, 0.65]$, con una tolerancia de 10^{-5}; utilizar 8 decimales.

$$\frac{3e^x}{1+e^x}=2$$

Se iguala la ecuación con cero, para poder utilizar el método, es decir:

$$\frac{3e^x}{1+e^x}-2=0$$

$$\Rightarrow f(x)=\frac{3e^x}{1+e^x}-2$$

Antes de aplicar el método de Newton se grafica la función $f(x)$, obteniendo la figura 2.15; donde se puede observar que la función es continua en el intervalo $[-0.65, 0.65]$ y que en éste intervalo se encuentra exactamente una raíz.

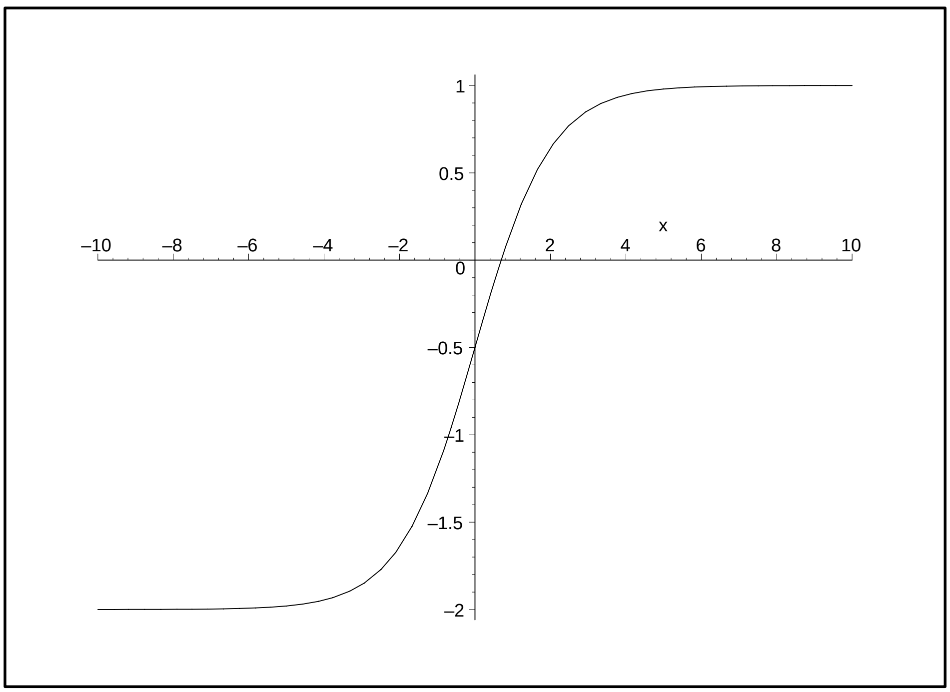


Figura 2.15 Gráfica de la ecuación $f(x)=\frac{3e^x}{1+e^x}-2$

Se calcula la primera y la segunda derivada $f'(x)$ y $f''(x)$, (obtenidas con Maple):

$$f'(x)=\frac{3e^x}{1+e^x}-\frac{3(e^x)^2}{(1+e^x)^2}$$

$$f''(x)=\frac{3e^x}{1+e^x}-\frac{9(e^x)^2}{(1+e^x)^2}+\frac{6(e^x)^3}{(1+e^x)^3}$$

Para saber que aproximación inicial tomar, se verifica en qué extremo del intervalo la función y su derivada tienen el mismo signo, (teorema 2.3.1); obteniéndose:

$$f(-0.65) = \tfrac{3e^{-0.65}}{1+e^{-0.65}} - 2 = -0.97103139$$
$$f''(-0.65) = \tfrac{3e^{-0.65}}{1+e^{-0.65}} - \tfrac{9(e^{-0.65})^2}{(1+e^{-0.65})^2} + \tfrac{6(e^{-0.65})^3}{(1+e^{-0.65})^3} = 0.21229169$$
$$f(0.65) = \tfrac{3e^{0.65}}{1+e^{0.65}} - 2 = -0.02896861$$
$$f''(0.65) = \tfrac{3e^{0.65}}{1+e^{0.65}} - \tfrac{9(e^{0.65})^2}{(1+e^{0.65})^2} + \tfrac{6(e^{0.65})^3}{(1+e^{0.65})^3} = -0.21229169$$

Tomando como valor inicial $x_0 = 0.65$, ya que en este extremo del intervalo la función y su derivada tienen el mismo signo. Por lo que se empieza a aplicar el método de Newton, ya que se puede garantizar que va a converger.
Utilizando el algoritmo 2.3.1, se empieza calculando $f(0.65)$ y $f'(0.65)$ obteniéndose:

$$f(0.65) = \tfrac{3e^{0.65}}{1+e^{0.65}} - 2 = -0.02896861$$
$$f'(0.65) = \tfrac{3e^{0.65}}{1+e^{0.65}} - \tfrac{3(e^{0.65})^2}{(1+e^{0.65})^2} = 0.67604314$$

A continuación se calcula x_1 con la ecuación (2.13):

$$x_1 = 0.65 - \frac{0.02896861}{0.67604314} = 0.69285024$$

después se calcula $f(x_1)$ y $f'(x_1)$:

$$f(0.69285024) = \tfrac{3e^{0.69285024}}{1+e^{0.69285024}} - 2 = -0.00019797$$
$$f'(0.69285024) = \tfrac{3e^{0.69285024}}{1+e^{0.69285024}} - \tfrac{3(e^{0.69285024})^2}{(1+e^{0.69285024})^2} = 0.66673264$$

Luego se calcula x_2 con la ecuación (2.13):

$$x_2 = 0.69285024 - \frac{-0.00019797}{0.66673264} = 0.69314717$$

después se calcula $f(x_2)$ y $f'(x_2)$:

$$f(0.69314717) = \tfrac{3e^{0.69314717}}{1+e^{0.69314717}} - 2 = -0.00000001$$
$$f'(0.69314717) = \tfrac{3e^{0.69314717}}{1+e^{0.69314717}} - \tfrac{3(e^{0.69314717})^2}{(1+e^{0.69314717})^2} = 0.66666667$$

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\tfrac{x_n - x_{n-1}}{x_n}\right| \le 0.00001$ y que $|f(x_n)| \le 0.00001$; esto se cumple en la iteración 3: $0.00000003 \le 0.00001$ y $trm 0.00000001 \le 0.00001$.
Para obtener x_3; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.9.
A continuación se indica como calcular el error relativo $\left|\tfrac{x_n - x_{n-1}}{x_n}\right|$.
Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_1 - x_0}{x_1}\right| = \left|\frac{0.69285024 - 0.65000000}{0.69285024}\right| = 0.06184632$$

n	x_n	$f(x_n)$	$f'(x_n)$	ε
0	0.65000000	-0.02896861	0.67604314	——
1	0.69285024	-0.00019797	0.66673264	0.06184632
2	0.69314717	-0.00000001	0.66666667	0.00042838
3	**0.69314719**	0.00000001	0.66666666	0.00000003

Cuadro 2.9: Ejemplo del método de Newton cuando converge

Para la segunda iteración el error relativo se calcula:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{0.69314717 - 0.69285024}{0.69314717}\right| = 0.00042838$$

Para obtener el error relativo de la última iteración se repite el proceso anterior. Los errores se muestran en el cuadro 2.9.

$$x \approx 0.69314719 \qquad \text{Con una tolerancia Tol=0.00001}$$

Comprobación:

$$f(0.69314719) = \frac{3e^{0.69314719}}{1+e^{0.69314719}} - 2 = 0.00000001 \approx 0$$

Ejemplo 2.3.4 Usar el método de Newton para encontrar una aproximación de la raíz compleja de la siguiente ecuación; con $x_0 = 1+i$, con una tolerancia de 10^{-5}; utilizar 6 decimales.

$$f(x) = x^3 + 2x^2 - x + 5$$

Antes de aplicar el método de Newton se grafica la función $f(x)$, obteniendo la figura 2.16; donde se puede observar que la función sólo tiene una raíz real, por lo que las otras dos raíces son complejas.

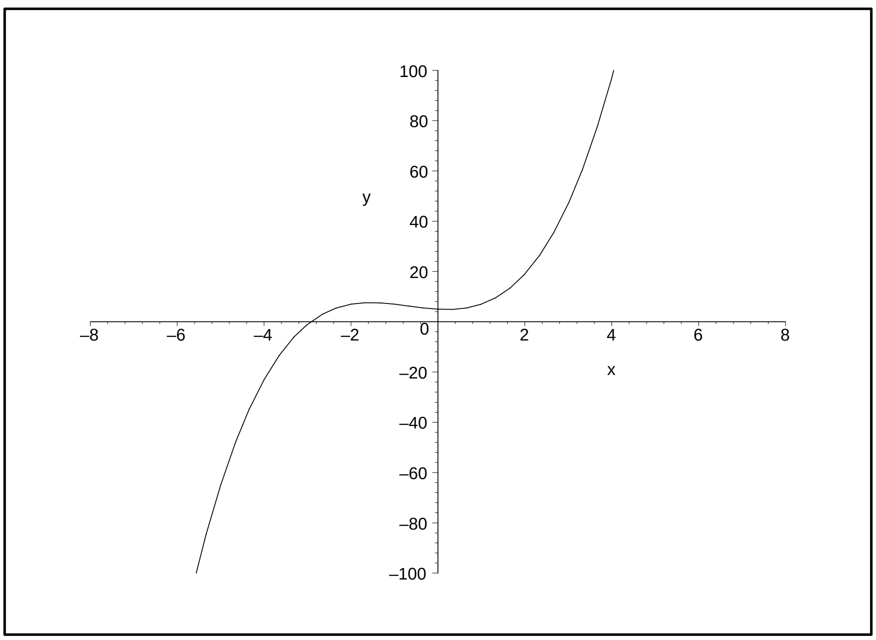


Figura 2.16 Gráfica de la ecuación $f(x) = x^3 + 2x^2 - x + 5$

Se calcula la primera derivada $f'(x)$, (obtenida con Maple):

$$f'(x) = 3x^2 + 4x - 1$$

Utilizando el algoritmo 2.3.1, se empieza calculando $f(1+i)$ y $f'(1+i)$ obteniéndose:

$$f(1+i) = (1+i)^3 + 2(1+i)^2 - (1+i) + 5 = 2+5i$$

$$f'(1+i) = 3(1+i)^2 + 4(1+i) - 1 = 3+10i$$

A continuación se calcula x_1 con la ecuación (2.13):

$$x_1 = (1+i) - \frac{2+5i}{3+10i} = 0.486239 + 1.045872\,i$$

después se calcula $f(x_1)$ y $f'(x_1)$:

$$\begin{aligned} f(0.486239{+}1.045872\,i) &= (0.486239{+}1.045872\,i)^3 + 2(0.486239{+}1.045872\,i)^2 - \\ &\quad (0.486239{+}1.045872\,i) + 5 \\ &= 1.318268{+}0.586095\,i \\ f'(0.486239{+}1.045872\,i) &= 3(0.486239{+}1.045872\,i)^2 + 4(0.486239{+}1.045872\,i) - 1 \\ &= -1.627307{+}7.234746\,i \end{aligned}$$

Luego se calcula x_2 con la ecuación (2.13):

$$x_2 = (0.486239 + 1.045872\,i) - \frac{1.318268 + 0.586095\,i}{-1.627307 + 7.234746\,i} = 0.448140 + 1.236655\,i$$

después se calcula $f(x_2)$ y $f'(x_2)$:

$$\begin{aligned} f(0.448140{+}1.236655\,i) &= (0.448140{+}1.236655\,i)^3 + 2(0.448140{+}1.236655\,i)^2 - \\ &\quad (0.448140{+}1.236655\,i) + 5 \\ &= -0.071155\text{-}0.166043\,i \\ f'(0.448140{+}1.236655\,i) &= 3(0.448140{+}1.236655\,i)^2 + 4(0.448140{+}1.236655\,i) - 1 \\ &= -3.192898{+}8.271787\,i \end{aligned}$$

Para saber cuando dejar de iterar, se debe se cumplir que $\left|f(x_n)\right| \le 0.00001$; esto se cumple en la iteración 5: $0.00000000 \le 0.00001$.

Para obtener x_3, x_4 y x_5; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.10.

A continuación se indica como calcular el error relativo $\left|\frac{x_n - x_{n-1}}{x_n}\right|$.

Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_1 - x_0}{x_1}\right| = \left|\frac{(0.486239 + 1.045872\,i) - (1.000000 + 1.000000\,i)}{0.486239 + 1.045872\,i}\right| = 0.447216$$

Para la segunda iteración el error relativo se calcula:

$$\left|\frac{x_2 - x_1}{x_2}\right| = \left|\frac{(0.448140 + 1.236655\,i) - (0.486239 + 1.045872)\,i}{0.448140 + 1.236655\,i}\right| = 0.147907$$

n	x_n	$f(x_n)$	$f'(x_n)$	ε
0	1.000000+1.000000 i	2.000000+5.000000 i	3.000000+10.000000 i	——
1	0.486239+1.045872 i	1.318268+0.586095 i	−1.627307+7.234746 i	0.447216
2	0.448140+1.236655 i	−0.071155 − 0.166043 i	−3.192898+8.271787 i	0.147907
3	0.462721+1.222425 i	0.001564 − 0.001357 i	−2.989757+8.283546 i	0.015588
4	0.462929+1.222540 i	−0.000001+0.000001 i	−2.989209+8.285833 i	0.000180
5	**0.462926+1.222540** i	−0.000000 − 0.000002 i	−2.989209+8.285830 i	0.000000

Cuadro 2.10: Ejemplo del método de Newton cuando converge

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.10.
Donde el valor absoluto de los números complejos esta definido de la siguiente manera:

$$|a+bi| = \sqrt{a^2+b^2}$$

$$x \approx 0.462926 + 1.222540\,i \qquad \text{Con una tolerancia } \text{Tol=0.00001}$$

Comprobación:

$$\begin{aligned} f(0.462926+1.222540\,i) &= (0.462926+1.222540\,i)^3 + 2(0.462926+1.222540\,i)^2 \\ &\quad -(0.462926+1.222540\,i) + 5 \\ &= -0.000000 - 0.000002\,i \approx 0 + 0i \end{aligned}$$

Ventajas del método de Newton.

1. La razón de convergencia iterativa del método de Newton es alta, cuando funciona.
2. Sirve para encontrar raíces complejas, para lo cual requiere que x_0 sea un número complejo $x_0 = a + bi$.

Desventajas del método de Newton.

1. El método requiere una buena estimación inicial. De otro modo, la solución iterativa puede diverger o converger a una solución irrelevante.
2. La derivada $f'(x)$, no siempre es fácil de calcular.

2. 4 Método de la secante

El *método de la secante* es una variante del método de Newton, aproximando $f'(x_n)$ por la pendiente de la recta, obteniéndose: e calcular. En dichos casos, la derivada se puede aproximar mediante la pendiente de la recta secante, obteniéndose:

$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}} \qquad \text{pendientedelarecta}$$

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

$$x_{n+1} = x_n - \frac{f(x_n)}{\frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}}$$

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \quad \textit{fórmuladeiteraciónparaelmétododelasecante} \qquad (2.17)$$

La ventaja es que no hay que calcular la derivada de f; esto es de gran ayuda en un caso en que f' sea difícil de calcular.

En la ecuación (7) se puede observar que para iterar con el método de la secante se requiere conocer dos aproximaciones iniciales x_0 y x_1.

El método de la secante también esta ligado con el método de falsa posición, ya que ambos comienzan con un intervalo y se basan en la fórmula de interpolación lineal, pero el primero utiliza extrapolaciones y no necesariamente el intervalo encierra la raíz buscada (lo que puede provocar divergencia en el método), mientras que el segundo utiliza únicamente interpolaciones y el intervalo debe encerrar la raíz buscada (por lo que el método siempre converge).

Al igual que el método de falsa posición, la aproximación a la raíz x_n, se obtiene con la intersección de la recta secante a la curva que une a los puntos $(x_0, f(x_0))$ y $(x_1, f(x_1))$ con el eje x y se denota como x_2; proporcionando una mejor estimación de la raíz. Sin embargo la diferencia entre ambos métodos radica en la forma en que uno de los valores iniciales se reemplaza por la nueva aproximación, ya que en el método de la secante se usan las dos últimas aproximaciones x_n y x_{n+1} para obtener el nuevo subintervalo en lugar de buscar el subintervalo que encierra a la raíz; (en consecuencia, algunas veces los valores del subintervalo están del mismo lado de la raíz). Entonces el nuevo intervalo sería $[x_1, x_2]$, después se traza otra recta secante a la curva que une a los puntos $(x_1, f(x_1))$ y $(x_2, f(x_2))$. La nueva intersección de la recta secante con el eje x, se denota como x_3 y se considera una mejor aproximación de la raíz. El proceso se repite hasta que se cumplan los criterios de convergencia de las ecuaciones (2.5), (2.6) y (2.7).

En la figura 2.17 se muestra la representación gráfica del método de la secante.

La intersección de la recta secante con el eje x se puede calcular de la siguiente manera (Ver figura 2.17):

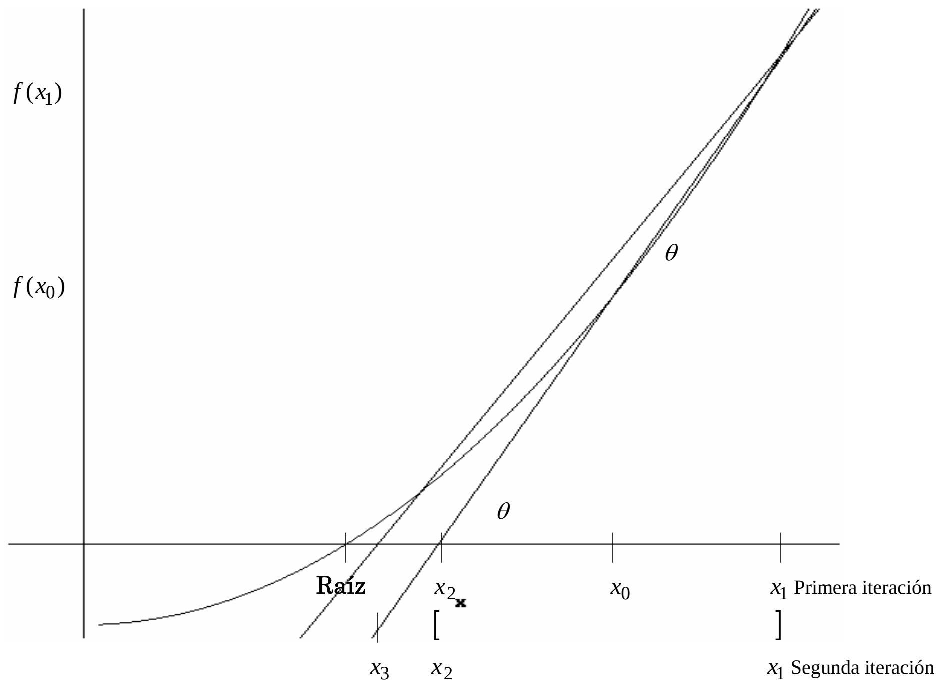


Figura 2.17 Representación gráfica del método de la secante

$$\cot\theta = \frac{\text{catetoadyacente}}{\text{catetoopuesto}}$$

$$\cot\theta = \frac{x_2 - x_1}{f(x_1)} \qquad \cot\theta = \frac{x_1 - x_0}{f(x_0) - f(x_1)}$$

$$\Rightarrow \frac{x_1 - x_2}{f(x_1)} = \frac{x_1 - x_0}{f(x_1) - f(x_0)}$$

$$x_1 - x_2 = f(x_1)\frac{x_1 - x_0}{f(x_1) - f(x_0)}$$

$$-x_2 = -x_1 + f(x_1)\frac{x_1 - x_0}{f(x_1) - f(x_0)}$$

$$x_2 = x_1 - f(x_1)\frac{x_1 - x_0}{f(x_1) - f(x_0)}$$

Para obtener la n-ésima intersección con el eje x de la recta secante a la curva, la ecuación sería:

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

obteniendo la ecuación (2.17).

Algoritmo 2.4.1 Método de la secante Para encontrar una aproximación de la raíz, de una ecuación $f(x)=0$, conocidas dos aproximaciones iniciales x_0 y x_1. **Datos:** $f(x)$, dos aproximaciones iniciales x_0 y x_1, una tolerancia Tol y un número máximo de iteraciones N.
Resultados: Una raíz aproximada o un mensaje de falla.

- **Paso 1:** Hacer $n=1$.
- **Paso 2:** Mientras $n \le N$, repetir los pasos 3 a 11.
  - **Paso 3:** Calcular $f(x_{n-1})$ y $f(x_n)$.
  - **Paso 4:** Si $n=1$ entonces realizar la siguiente comparación. De lo contrario CONTINUAR.
    - **Paso 5:** Si $f(x_{n-1}) * f(x_n) = 0$ entonces IMPRIMIR "Una raíz de la ecuación dada es uno de los extremos del intervalo" y TERMINAR. De lo contrario CONTINUAR.
  - **Paso 6:** Calcular $x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n)-f(x_{n-1})}$
  - **Paso 7:** Calcular $f(x_{n+1})$
  - **Paso 8:** Si $f(x_{n+1})=0$ entonces IMPRIMIR "Una raíz de la ecuación dada es x_{n+1}" y TERMINAR. De lo contrario CONTINUAR.
  - **Paso 9:** Si $n>1$ entonces calcular $\varepsilon = \left|\frac{x_{n+1}-x_n}{x_{n+1}}\right|$ y realizar la siguiente comparación. De lo contrario CONTINUAR.
    - **Paso 10:** Si $\varepsilon \le Tol$ y $\left|f(x_{n+1})\right| \le Tol$ entonces IMPRIMIR "Una raíz aproximada de la ecuación dada es x_{n+1}" y TERMINAR. De lo contrario CONTINUAR.
  - **Paso 11:** Hacer $n=n+1$.
- **Paso 12:** IMPRIMIR mensaje de falla "El método no converge a una raíz" y TERMINAR.

Análisis de convergencia para el método de la secante

Para analizar de la convergencia del método de la secante, se utiliza la serie de Taylor, tomando en cuenta que el error en la n-ésima iteración sería: $\varepsilon_n = x - x_n$.

$$\Rightarrow \varepsilon_{n+1} = x - x_{n+1}$$

$$\varepsilon_{n-1} = x - x_{n-1}$$

Se sustituye la ecuación (2.17), en ε_{n+1}, obteniéndose:

$$\varepsilon_{n+1} = x - \left[x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n)-f(x_{n-1})}\right]$$

$$\varepsilon_{n+1} = x - \left[\frac{x_n f(x_n) - x_n f(x_{n-1}) - x_n f(x_n) + x_{n-1} f(x_n)}{f(x_n)-f(x_{n-1})}\right]$$

$$\varepsilon_{n+1} = x - \left[\frac{-x_n f(x_{n-1}) + x_{n-1} f(x_n)}{f(x_n)-f(x_{n-1})}\right]$$

$$\varepsilon_{n+1} = \frac{xf(x_n) - xf(x_{n-1}) + x_n f(x_{n-1}) - x_{n-1} f(x_n)}{f(x_n)-f(x_{n-1})}$$

$$\varepsilon_{n+1} = \frac{f(x_n)(x - x_{n-1}) - f(x_{n-1})(x - x_n)}{f(x_n) - f(x_{n-1})}$$

como $\varepsilon_n = x - x_n$ y $\varepsilon_{n-1} = x - x_{n-1}$, entonces:

$$\varepsilon_{n+1} = \frac{f(x_n)\varepsilon_{n-1} - f(x_{n-1})\varepsilon_n}{f(x_n) - f(x_{n-1})} \tag{2.18}$$

Se multiplica la ecuación (2.18) por $\frac{x_n - x_{n-1}}{x_n - x_{n-1}}$, obteniéndose:

$$\varepsilon_{n+1} = \left(\frac{x_n - x_{n-1}}{x_n - x_{n-1}}\right)\left(\frac{f(x_n)\varepsilon_{n-1} - f(x_{n-1})\varepsilon_n}{f(x_n) - f(x_{n-1})}\right)$$

$$\varepsilon_{n+1} = \left(\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}\right)\left(\frac{f(x_n)\varepsilon_{n-1} - f(x_{n-1})\varepsilon_n}{x_n - x_{n-1}}\right)$$

Se factoriza en términos de ε_n y ε_{n-1}, obteniéndose:

$$\varepsilon_{n+1} = \left(\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}\right)\left(\frac{\frac{f(x_n)}{\varepsilon_n} - \frac{f(x_{n-1})}{\varepsilon_{n-1}}}{x_n - x_{n-1}}\right)\varepsilon_n\varepsilon_{n-1} \tag{2.19}$$

Al utilizar la serie de Taylor de $f(x_n)$ alrededor de x, se tiene:

$$f(x_n) = f(x) + f'(x)(x_n - x) + \frac{f''(x)(x_n - x)^2}{2!} + \ldots$$

Truncando la serie de Taylor después del término de la segunda derivada, se obtiene:

$$f(x_n) \approx f(x) + f'(x)(x_n - x) + \frac{f''(x)(x_n - x)^2}{2}$$

En la intersección con el eje x, $f(x)$ debe ser igual a cero, entonces:

$$f(x_n) \approx 0 + f'(x)(x_n - x) + \frac{f''(x)(x_n - x)^2}{2}$$

como $-\varepsilon_n = x_n - x$, entonces:

$$f(x_n) \approx f'(x)(-\varepsilon_n) + \frac{f''(x)(-\varepsilon_n)^2}{2}$$

$$f(x_n) \approx \left[-f'(x) + \frac{f''(x)}{2}\varepsilon_n\right]\varepsilon_n$$

$$\frac{f(x_n)}{\varepsilon_n} \approx -f'(x) + \frac{f''(x)}{2}\varepsilon_n \tag{2.20}$$

Al utilizar la serie de Taylor de $f(x_{n-1})$ alrededor de x, se tiene:

$$f(x_{n-1}) = f(x) + f'(x)(x_{n-1} - x) + \frac{f''(x)(x_{n-1} - x)^2}{2!} + \ldots$$

Truncando la serie de Taylor después del término de la segunda derivada, se obtiene:

$$f(x_{n-1}) \approx f(x) + f'(x)(x_{n-1} - x) + \frac{f''(x)(x_{n-1} - x)^2}{2}$$

En la intersección con el eje x, $f(x)$ debe ser igual a cero, entonces:

$$f(x_{n-1}) \approx 0 + f'(x)(x_{n-1} - x) + \frac{f''(x)(x_{n-1} - x)^2}{2}$$

como $-\varepsilon_{n-1} = x_{n-1} - x$, entonces:

$$f(x_{n-1}) \approx f'(x)(-\varepsilon_{n-1}) + \frac{f''(x)(-\varepsilon_{n-1})^2}{2}$$

$$f(x_{n-1}) \approx \left[-f'(x) + \frac{f''(x)}{2}\varepsilon_{n-1} \right] \varepsilon_{n-1}$$

$$\frac{f(x_{n-1})}{\varepsilon_{n-1}} \approx -f'(x) + \frac{f''(x)}{2}\varepsilon_{n-1} \tag{2.21}$$

Se restan las ecuaciones (2.20) y (2.21); obteniéndose:

$$\begin{aligned} \frac{f(x_n)}{\varepsilon_n} - \frac{f(x_{n-1})}{\varepsilon_{n-1}} &\approx -f'(x) + \tfrac{f''(x)}{2}\varepsilon_n - \left[-f'(x) + \tfrac{f''(x)}{2}\varepsilon_{n-1} \right] \\ &\approx -f'(x) + \tfrac{f''(x)}{2}\varepsilon_n + f'(x) - \tfrac{f''(x)}{2}\varepsilon_{n-1} \\ &\approx \tfrac{f''(x)}{2}\varepsilon_n - \tfrac{f''(x)}{2}\varepsilon_{n-1} \\ &\approx \tfrac{f''(x)}{2}(\varepsilon_n - \varepsilon_{n-1}) \end{aligned}$$

como $\varepsilon_n = x - x_n$ y $\varepsilon_{n-1} = x - x_{n-1}$, entonces:

$$\varepsilon_n - \varepsilon_{n-1} = x - x_n - (x - x_{n-1}) = x - x_n - x + x_{n-1} = x_{n-1} - x_n$$

$$\Rightarrow \frac{f(x_n)}{\varepsilon_n} - \frac{f(x_{n-1})}{\varepsilon_{n-1}} \approx \tfrac{f''(x)}{2}(x_{n-1} - x_n)$$

$$\frac{f(x_n)}{\varepsilon_n} - \frac{f(x_{n-1})}{\varepsilon_{n-1}} \approx -\tfrac{f''(x)}{2}(x_n - x_{n-1})$$

$$\frac{\frac{f(x_n)}{\varepsilon_n} - \frac{f(x_{n-1})}{\varepsilon_{n-1}}}{x_n - x_{n-1}} \approx -\tfrac{f''(x)}{2}$$

obteniendo la segunda expresión del lado derecho entre paréntesis de la ecuación (2.19). La primera expresión entre paréntesis de la ecuación (2.19), se puede escribir como:

$$\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} \approx \frac{1}{f'(x)}$$

ya que $f'(x_n)$ se aproxima a la pendiente de la recta.

$$\varepsilon_{n+1} \approx \frac{1}{f'(x)}\left(-\frac{f''(x)}{2}\right)\varepsilon_n\varepsilon_{n-1} \approx -\frac{f''(x)}{2f'(x)}\varepsilon_n\varepsilon_{n-1}$$

es decir el error en una determinada iteración es proporcional al producto de los errores de las dos iteraciones previas. A este comportamiento se le llama convergencia superlineal.

Por lo tanto el método de la secante comparado con el método de Newton, tiene una convergencia ligeramente menor, pero con la ventaja que no hay que derivar $f(x)$.

Ejemplo 2.4.1 Usar el método de la secante para encontrar una aproximación de la raíz de la ecuación (2.10), ($f(x) = e^x - x^2 + 3x - 2$); en [0,1], con una tolerancia de 10^{-5}; utilizar 6 decimales.

Antes de aplicar el método de la secante se grafica la función $f(x)$, obteniendo la figura 2.2; donde se puede observar que la función es continua y que en este caso el intervalo encierra la raíz; por lo que se empieza a aplicar el método.

Utilizando el algoritmo 2.4.1, se empieza calculando $f(0)$ y $f(1)$ obteniéndose:

$$f(0) = e^0 - 0^2 + 3(0) - 2 = -1.000000$$
$$f(1) = e^1 - 1^2 + 3(1) - 2 = 2.718282$$

A continuación se calcula x_2 con la ecuación (2.17):

$$x_2 = 1 - 2.718282\left(\frac{1-0}{2.718282-(-1.000000)}\right) = 0.268941$$

después se calcula $f(x_2)$:

$$f(0.268941) = e^{0.268941} - 0.268941^2 + 3(0.268941) - 2 = 0.043073$$

Luego utilizando las últimas dos iteraciones $x_1 = 1$ y $x_2 = 0.268941$ se calcula x_3 con la ecuación (2.17):

$$x_3 = 0.268941 - 0.043073\left(\frac{0.268941-1.000000}{0.043073-2.718282}\right) = 0.257171$$

después se calcula $f(x_3)$:

$$f(0.257171) = e^{0.257171} - 0.257171^2 + 3(0.257171) - 2 = -0.001359$$

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_{n+1}-x_n}{x_{n+1}}\right| \leq 0.00001$ y que $|f(x_n)| \leq 0.00001$; esto se cumple en la iteración 4: $0.000001 \leq 0.00001$ y $0.000000 \leq 0.00001$.

Para obtener x_4; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.11.

n	x_{n-1}	x_n	$f(x_{n-1})$	$f(x_n)$	x_{n+1}	$f(x_{n+1})$	ε
1	0.000000	1.000000	-1.000000	2.718282	0.268941	0.043073	——
2	1.000000	0.268941	2.718282	0.043073	0.257171	-0.001359	0.045770
3	0.268941	0.257171	0.043073	-0.001359	0.257531	0.000001	0.001398
4	0.257171	0.257531	-0.001359	0.000001	**0.257530**	0.000000	0.000001

Cuadro 2.11: Ejemplo del método de la secante cuando converge

A continuación se indica como calcular el error relativo $\left|\frac{x_{n+1}-x_n}{x_{n+1}}\right|$.

Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_3-x_2}{x_3}\right| = \left|\frac{0.257171-0.268941}{0.257171}\right| = 0.0457707$$

Para la segunda iteración el error relativo se calcula:

$$\left|\frac{x_4-x_3}{x_4}\right| = \left|\frac{0.257531-0.257171}{0.257531}\right| = 0.001398$$

Para obtener el error relativo de la última iteración se repite el proceso anterior. Los errores se muestran en el cuadro 2.11.

$$x \approx 0.257530 \qquad \text{Conunatolerancia Tol=0.00001}$$

Comprobación:

$$\mathrm{f}(0.257530)=\mathrm{e}^{(0.257530)} - (0.257530)^2 + 3(0.257530) - 2 = 0.000000\ldots \approx 0$$

Observando los ejemplos 2.1.1, 2.2.1, 2.3.1 y 2.4.1 se puede verificar que para la ecuación (2.10) el método de Newton converge más rápido que los métodos de falsa posición, bisección y secante; aunque hay casos en donde estos últimos convergen más rápido.

Ejemplo 2.4.2 Usar el método de la secante para encontrar una aproximación de la raíz de la siguiente ecuación; en $[5,6]$ y en $[15,16]$, con una tolerancia de 10^{-5}; utilizar 6 decimales.

$$f(x) = x\log x - 10$$

Antes de aplicar el método de la secante se grafica la función $f(x)$, obteniendo la figura 2.18; donde se puede observar que la función es continua y que en este caso solo el intervalo $[5,6]$ encierra la raíz; por lo que se empieza a aplicar el método.

Para el intervalo $[5,6]$***:***

Utilizando el algoritmo 2.4.1, se empieza calculando $f(5)$ y $f(6)$ obteniéndose:

$$f(5) = 5\log 5 - 10 = -1.952810$$

$$f(6) = 6\log 6 - 10 = 0.750557$$

A continuación se calcula x_2 con la ecuación (2.17):

$$x_2 = 6 - 0.750557\left(\frac{6-5}{0.750557-(-1.952810)}\right) = 5.722362$$

después se calcula $f(x_2)$:

$$f(5.722362) = 5.722362 \log 5.722362 - 10 = -0.018016$$

Luego utilizando las últimas dos iteraciones $x_1 = 6$ y $x_2 = 5.722362$ se calcula x_3 con la ecuación (2.17):

$$x_3 = 5.722362 - \left(-0.018016\right)\left(\frac{5.722362 - 6.000000}{-0.018016 - 0.750557}\right) = 5.728870$$

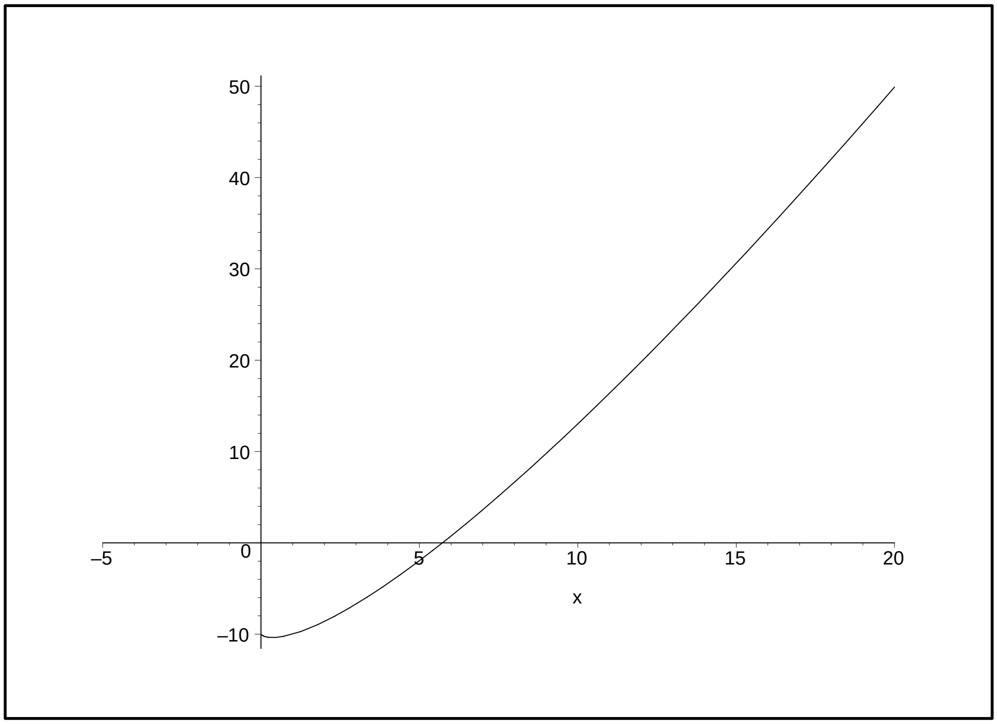


Figura 2.18 Gráfica de la ecuación $f(x) = x\log x - 10$

después se calcula $f(x_3)$:

$$f(5.728870) = 5.728870 \log 5.728870 - 10 = -0.000152$$

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_{n+1}-x_n}{x_{n+1}}\right| \leq 0.00001$ y que $\left|f(x_n)\right| \leq 0.00001$; esto se cumple en la iteración 4: $0.000000 \leq 0.00001$ y $0.000000 \leq 0.00001$.

Para obtener x_4; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.12.

n	x_{n-1}	x_n	$f(x_{n-1})$	$f(x_n)$	x_{n+1}	$f(x_{n+1})$	ε
1	5.000000	6.000000	-1.952810	0.750557	5.722362	-0.018016	——
2	6.000000	5.722362	0.750557	-0.018016	5.728870	-0.000152	0.001136
3	5.722362	5.728870	-0.018016	-0.000152	5.728926	0.000000	0.000010
4	5.728870	5.728926	-0.000152	0.000000	**5.728926**	0.000000	0.000000

Cuadro 2.12: Ejemplo del método de la secante cuando converge

A continuación se indica como calcular el error relativo $\left|\frac{x_{n+1}-x_n}{x_{n+1}}\right|$.

Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_3-x_2}{x_3}\right| = \left|\frac{5.728870-5.722362}{5.728870}\right| = 0.001136$$

Para la segunda iteración el error relativo se calcula:

$$\left|\frac{x_4-x_3}{x_4}\right| = \left|\frac{5.728926-5.728870}{5.728926}\right| = 0.000010$$

Para obtener el error relativo de la última iteración se repite el proceso anterior. Los errores se muestran en el cuadro 2.12.

$$x \approx 5.728926 \qquad \text{Conunatolerancia } \text{Tol}=0.00001$$

Comprobación:

$$\text{f}(5.728926)=5.728926\ \log 5.728926 - 10 = 0.000000\ldots \approx 0$$

Paraelintervalo [15,16]:

Utilizando el algoritmo 2.4.1, se empieza calculando $f(15)$ y $f(16)$ obteniéndose:

$$f(15) = 15\log 15 - 10 = -30.620753$$
$$f(16) = 16\log 16 - 10 = 34.361420$$

A continuación se calcula x_2 con la ecuación (2.17):

$$x_2 = 16 - 34.361420\left(\frac{16-15}{34.361420-(30.620753)}\right) = 6.814092$$

después se calcula $f(x_2)$:

$$f(6.814092) = 6.814092\log 6.814092 - 10 = -3.076193$$

Luego utilizando las últimas dos iteraciones $x_1 = 16$ y $x_2 = 6.814092$ se calcula x_3 con la ecuación (2.17):

$$x_3 = 6.814092 - 3.076193\left(\frac{6.814092-16.000000}{3.076193-34.361420}\right) = 5.910866$$

después se calcula $f(x_3)$:

$$f(5.910866) = 5.910866\log 5.910866 - 10 = 0.502381$$

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_{n+1}-x_n}{x_{n+1}}\right| \leq 0.00001$ y que $|f(x_n)| \leq 0.00001$; esto se cumple en la iteración 5: $0.000006 \leq 0.00001$ y $0.000000 \leq 0.00001$.

Para obtener x_4 y x_5; se repite el proceso anterior. Las iteraciones se muestran en el cuadro 2.13.

n	x_{n-1}	x_n	$f(x_{n-1})$	$f(x_n)$	x_{n+1}	$f(x_{n+1})$	ε
1	15.000000	16.000000	30.620753	34.361420	6.814092	3.076193	——
2	16.000000	6.814092	34.361420	3.076193	5.910866	0.502381	0.152808
3	6.814092	5.910866	3.076193	0.502381	5.734566	0.015488	0.030743
4	5.910866	5.734566	0.502381	0.015488	5.728958	0.000088	0.000979
5	5.734566	5.728958	0.015488	0.000088	**5.728926**	0.000000	0.000006

Cuadro 2.13: Ejemplo del método de la secante cuando converge

A continuación se indica como calcular el error relativo $\left|\frac{x_{n+1}-x_n}{x_{n+1}}\right|$.

Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{5.910866 - 6.814092}{5.910866}\right| = 0.152808$$

Para la segunda iteración el error relativo se calcula:

$$\left|\frac{x_4 - x_3}{x_4}\right| = \left|\frac{5.734566 - 5.910866}{5.734566}\right| = 0.030743$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.13.

$$x \approx 5.728926 \qquad \text{Conunatolerancia Tol=0.00001}$$

Comprobación:

$$\text{f(5.728926)=5.728926} \log 5.728926 - 10 = 0.000000\ldots \approx 0$$

Observando el ejemplo 2.4.2, se puede verificar que no necesariamente para aplicar el método de la secante el intervalo debe encerrar la raíz, así mismo en la mayoría de los casos el método converge más rápido cuando el intervalo se encuentra cercano a ésta.

Ejemplo 2.4.3 Usar el método de la secante para encontrar una aproximación de la raíz de la siguiente ecuación; en $[0, 0.48]$, con una tolerancia de 10^{-5}; utilizar 6 decimales.

$$f(x) = \tan(\pi x) - 6$$

Antes de aplicar el método de la secante se grafica la función $f(x)$, obteniendo la figura 2.19; donde se puede observar que la función no es continua y que en este caso el intervalo encierra la raíz; por lo que se empieza a aplicar el método.

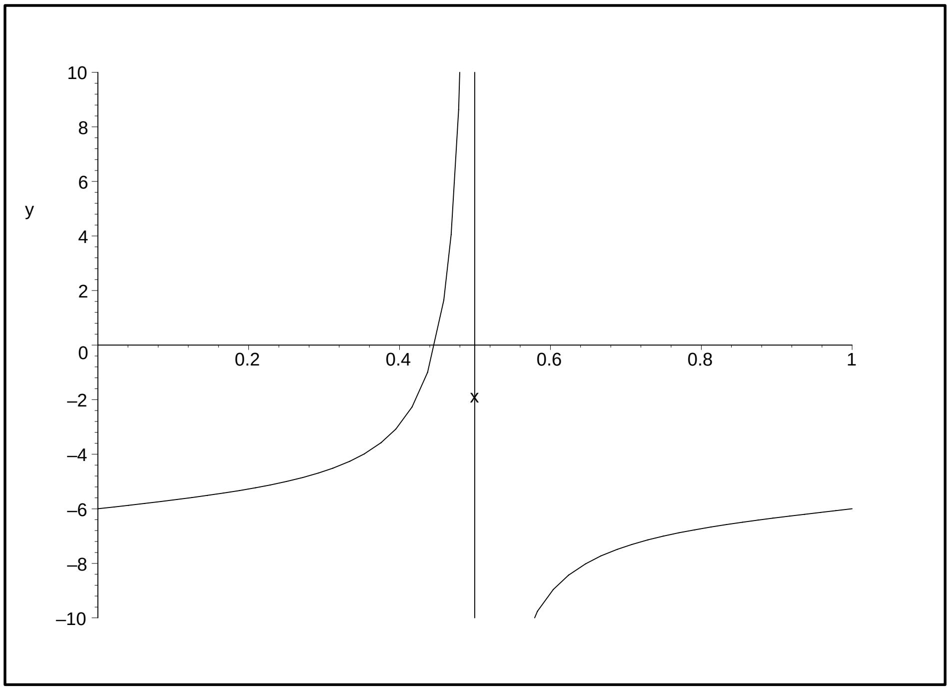


Figura 2.19 Gráfica de la ecuación $f(x) = \tan(\pi x) - 6$

Utilizando el algoritmo 2.4.1, se empieza calculando $f(0)$ y $f(0.48)$ obteniéndose:

$$f(0) = \tan(\pi 0) - 6 = -6.000000$$

$$f(0.48) = \tan(\pi 0.48) - 6 = 9.894545$$

A continuación se calcula x_2 con la ecuación (2.17):

$$x_2 = 0.48 - 9.894545\left(\frac{0.48 - 0}{9.894545 - (-6.000000)}\right) = 0.181194$$

después se calcula $f(x_2)$:

$$f(0.181194) = \tan(\pi 0.181194) - 6 = -5.360105$$

Luego utilizando las últimas dos iteraciones $x_1 = 0.48$ y $x_2 = 0.181194$ se calcula x_3 con la ecuación (2.17):

$$x_3 = 0.181194 - (-5.360105)\left(\frac{0.181194 - 0.480000}{-5.360105 - 9.894545}\right) = 0.286187$$

después se calcula $f(x_3)$:

$$f(0.286187) = \tan(\pi 0.286187) - 6 = -4.742211$$

Para saber cuando dejar de iterar, se debe se cumplir que $\left|\frac{x_{n+1} - x_n}{x_{n+1}}\right| \leq 0.00001$ y que $|f(x_n)| \leq 0.00001$; lo cual no se cumple, (ver cuadro 2.14).

Para obtener $x_3, x_4, x_5, x_6, \ldots$; se repite el proceso anterior. Algunas iteraciones se muestran en el cuadro 2.14.

n	x_{n-1}	x_n	$f(x_{n-1})$	$f(x_n)$	x_{n+1}	$f(x_{n+1})$	ε
1	0.000000	0.480000	-6.000000	9.894545	0.181194	-5.360105	——
2	0.480000	0.181194	9.894545	-5.360105	0.286187	-4.742211	0.366868
3	0.181194	0.286187	-5.360105	-4.742211	1.091986	-5.702695	0.737921
4	0.286187	1.091986	-4.742211	-5.702695	-3.692297	-4.551143	1.295747
5	1.091986	-3.692297	-5.702695	-4.551143	-22.600650	-2.943563	0.836629
6	-3.692297	-22.600650	-4.551143	-2.943563	-57.222832	-6.842372	0.605041
7	-22.600650	-57.222832	-2.943563	-6.842372	3.538758	-14.172096	17.170315
8	-57.222832	3.538758	-6.842372	-14.172096	-113.944398	-5.823523	1.031057
9	3.538758	-113.944398	-14.172096	-5.823523	-195.894431	-5.655624	0.418338
10	-113.944398	-195.894431	-5.823523	-5.655624	-2956.366771	-8.248014	0.933738
11	-195.894431	-2956.366771	-5.655624	-8.248014	5826.423063	-1.943618	1.507407
12	-2956.366771	5826.423063	-8.248014	-1.943618	8534.119765	-5.604926	0.317279
				⋮			

Cuadro 2.14: Ejemplo del método de la secante cuando no converge

A continuación se indica como calcular el error relativo $\left|\frac{x_{n+1}-x_n}{x_{n+1}}\right|$.

Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\left|\frac{x_3 - x_2}{x_3}\right| = \left|\frac{0.286187 - 0.181194}{0.286187}\right| = 0.366868$$

Para la segunda iteración el error relativo se calcula:

$$\left|\frac{x_4 - x_3}{x_4}\right| = \left|\frac{1.091986 - 0.286187}{1.091986}\right| = 0.737921$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 2.14.
El método de la secante no converge a una raíz.

Observando la gráfica se toma otro intervalo que se encuentre más cercano a la raíz.

Para el intervalo $[0.40, 0.48]$:
Utilizando el algoritmo 2.4.1, se empieza calculando $f(0.40)$ y $f(0.48)$ obteniéndose:

$$f(0.40) = \tan(\pi\, 0.40) - 6 = -2.922316$$
$$f(0.48) = \tan(\pi\, 0.48) - 6 = 9.8945448$$

A continuación se calcula x_2 con la ecuación (2.17):

$$x_2 = 0.48 - 9.894545\left(\frac{0.48 - 0.40}{9.894545 - (-2.922316)}\right) = 0.418240$$

después se calcula $f(x_2)$:

$$f(0.418240) = \tan(\pi\, 0.418240) - 6 = -2.192753$$

Capítulo 3

Solución de sistemas de ecuaciones lineales

A diferencia del **capítulo 2** en el que se trataron métodos para determinar el valor de x que satisface a una única ecuación, $f(x)=0$, en éste capítulo el objetivo principal será aplicar algunos métodos numéricos para hallar los valores $x_1, x_2, \ldots, x_n$ que en forma simultánea satisfacen a un sistema de ecuaciones lineales.
Un **sistema de ecuaciones lineales** es un conjunto de n ecuaciones (con coeficientes reales) con n incógnitas $x_1, x_2, x_3, \ldots x_n$, que deben resolverse simultáneamente y que pueden escribirse de forma tradicional de la siguiente manera:

$$\begin{array}{ccccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & + & \ldots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & + & \ldots & + & a_{2n}x_n & = & b_2 \\ & & & & \vdots & & & & \ddots & = & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & a_{n3}x_3 & + & \ldots & + & a_{nn}x_n & = & b_n \end{array} \tag{3.1}$$

El sistema de ecuaciones lineales (3.1) en notación matricial tiene la forma:

$$Ax = b \tag{3.2}$$

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nn} \end{pmatrix}}_{\text{Matriz de coeficientes}} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}}_{\text{Matriz de incógnitas}} = \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}}_{\text{Matriz de términos independientes}} \tag{3.3}$$

Donde:

1. A es la matriz de dimensión $n \times n$ formada por los coeficientes del sistema.
2. x es el vector columna formado por las incógnitas, también llamado vector solución.
3. b es el vector columna formado por los términos independientes.

Los sistemas de ecuaciones lineales que se verán en éste capítulo son del tipo no homogéneo. [1]

La matriz formada por A, a la que se le agrega el véctor de términos independientes, como última columna, se llama la **matriz aumentada** del sistema, y se representa de la siguiente manera:

$$[\mathbf{A}\,|\,\mathbf{b}] = \left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} & b_n \end{array}\right)$$

donde con la línea se separan los coeficientes de las incógnitas, de los términos independientes. Los métodos numéricos que se analizarán para la solución de sistemas de ecuaciones lineales son:

1. **Métodos exactos ó directos**: son aquellos que proporcionan una solución en un número determinado de pasos; la cual sería exacta si no fuera por los errores de redondeo.
    a. Método de Gauss.
    b. Método de Gauss-Jordan.
    c. Inversión de matrices.
    d. Inversión de matrices particionadas.
    e. Gauss-Jordan particionado.
    f. Método de intercambio.
2. **Métodos iterativos ó indirectos**: son aquellos que parten de una aproximación inicial a la solución del sistema dado, y aplicando cierto algoritmo se genera, a partir de dicha aproximación, una sucesión de vectores que si converge lo hace a la solución del sistema.
    a. Método de Jacobi.
    b. Método de Gauss-Seidel.
    c. Método de relajación.

Las matrices asociadas con los sistemas de ecuaciones lineales se clasifican en **densas** y **esparcidas**. Las matrices densas tienen pocos elementos nulos y su orden es relativamente pequeño $(n \leq 50)$; para sistemas con matrices densas se recomienda usar métodos exactos. Las matrices esparcidas tienen pocos elementos no nulos y su orden puede ser muy grande. Los métodos iterativos son recomendados para resolver sistemas con matrices esparcidas.

3.1 Condiciones necesarias y suficientes para la existencia de la solución de sistemas de ecuaciones lineales

No siempre es posible resolver un conjunto de ecuaciones lineales en forma numérica, por ejemplo:

1. Sistemas que representan líneas paralelas, donde no existe solución, ya que las dos líneas jamás se cruzan. Ejemplo:

[1] Cuando al menos uno de los términos independientes de la ecuación (3.1) es distinto de cero, se dice que el conjunto de ecuaciones lineales es no homogéneo.

$$
\begin{array}{rcrcl}
-\frac{1}{2}x_1 & + & x_2 & = & 1 \\
-\frac{1}{2}x_1 & + & x_2 & = & \frac{1}{2}
\end{array}
$$

También llamados *sistemas inconsistentes* [2]. Figura 3.1.

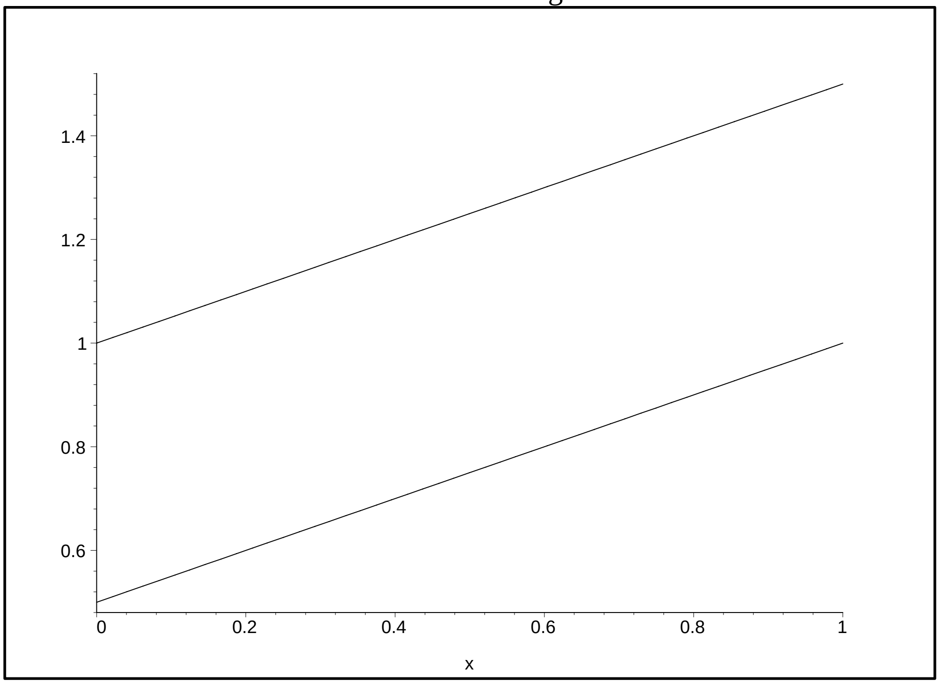


Figura 3.1 Gráfica de sistemas de ecuaciones sin solución.

2. Sistemas que representan dos líneas que coinciden, donde existe un número infinito de soluciones. Ejemplo:

$$
\begin{array}{rcrcl}
-\frac{1}{2}x_1 & + & x_2 & = & 1 \\
-x_1 & + & 2x_2 & = & 2
\end{array}
$$

Se puede observar que si multiplicamos la segunda ecuación por $\frac{1}{2}$, se obtiene la primera, por lo que matemáticamente son idénticas. Cualquier punto (x, y) que satisfaga una de las ecuaciones, también es solución de la otra. Por ejemplo $(-0.4, 0.8)$, $(0.2, 1.1)$, $(0.4, 1.2)$; es decir son *linealmente dependientes* [3]. Figura 3.2

[2]Un conjunto de ecuaciones es inconsistente si el lado izquierdo de al menos una de las ecuaciones se puede eliminar totalmente (sumando o restando otras ecuaciones), mientras que el lado derecho permanece distinto de cero. En éste tipo de sistemas el det $A = 0$.

[3] Si una ecuación es múltiplo de otra, o se puede obtener sumando o restando otras ecuaciones, se dice que esa ecuación es linealmente dependiente de otras. En éste tipo de sistemas el det $A = 0$, por las propiedades (6) y (7) de los determinantes, (sección 3.1.2).

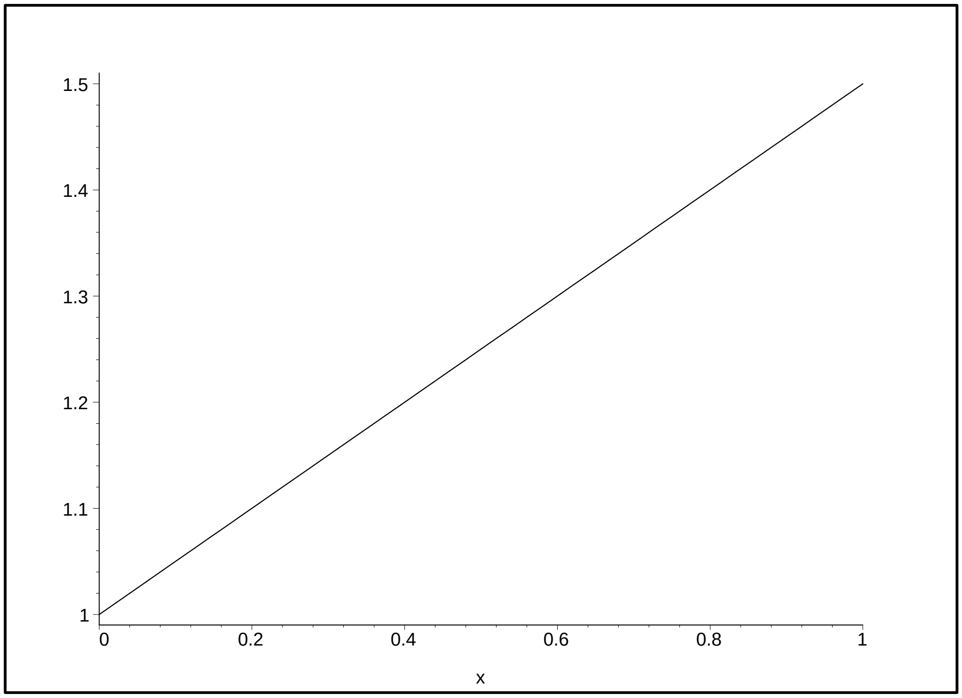


Figura 3.2 Gráfica de sistemas de ecuaciones con número infinito de soluciones.

3. Sistemas muy próximos a ser singulares. Esto se puede observar gráficamente cuando es difícil identificar el punto exacto donde las líneas se intersectan. Ejemplo:

$$\begin{aligned} -\frac{2.3}{5}x_1 + x_2 &= 1.1 \\ -\frac{1}{2}x_1 + x_2 &= 1 \end{aligned}$$

También llamados *sistemas mal condicionados* [4]. Figura 3.3.

[4]Los sistemas mal condicionados son aquellos donde pequeños cambios en los coeficientes generan grandes cambios en la solución; debido a que son extremadamente sensibles a los errores de redondeo. Con el uso de las estrategias de pivoteo (sección 3.3.7) y aumentando la precisión del cálculo (utilizando más cifras significativas), se puede solucionar este problema. En éste tipo de sistemas el det $A \approx 0$.

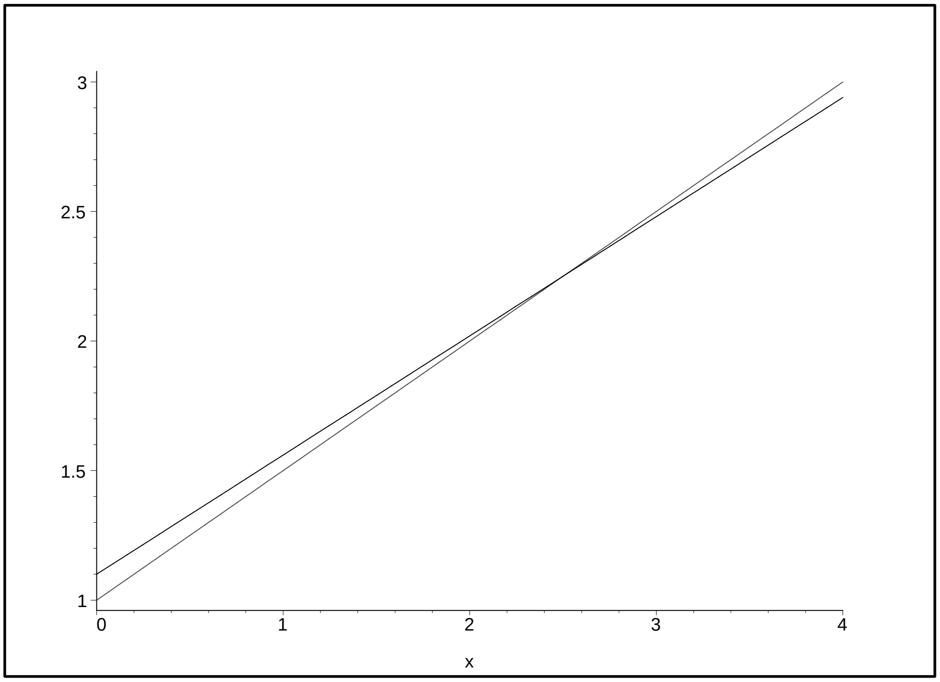


Figura 3.3. Gráfica de sistemas de ecuaciones muy próximos a ser singulares.

4. Sistemas con mayor número de ecuaciones que de incógnitas. Ejemplo:

$$\begin{array}{rcrcr} -x_1 & + & x_2 & = & 1 \\ x_1 & + & 2x_2 & = & -2 \\ 2x_1 & - & x_2 & = & 0 \end{array}$$

5. Como se puede ver en la figura 3.4, las tres ecuaciones no se pueden satisfacer simultáneamente.

Si el número de ecuaciones es mayor que dos, la carencia de independencia lineal o de inconsistencia son menos obvias.

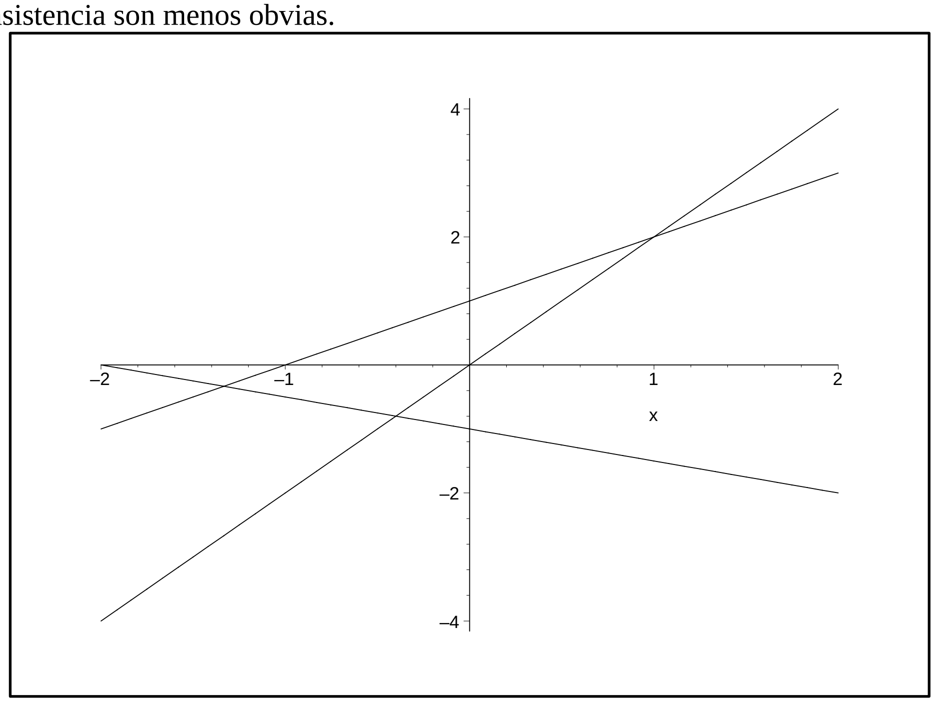


Figura 3.4 Gráfica de sistemas de ecuaciones con mayor número de ecuaciones que de incógnitas.

Con lo anterior se puede concluir que las condiciones necesarias para la existencia de una solución única son:

- El número de ecuaciones debe ser igual al número de incógnitas.
- Cada ecuación debe de ser linealmente independiente; es decir ninguna ecuación se puede eliminar sumando o restando otras ecuaciones.
- $\det A = |A| \neq 0$.

Para poder aplicar alguno de los métodos numéricos que se analizarán en éste capítulo es necesario comprobar que el det $A \neq 0$, por lo que se revisará la definición y algunas propiedades de los determinantes.

3.1.1 Definición de los determinantes

Definición 3.1.1 Determinante. El determinante de una matriz A determina la existencia de una solución única en un sistema de ecuaciones lineales; y se representa por $\det \mathbf{A}$ ó $|\mathbf{A}|$.[5]

1. Si $A = (a)$ es una matriz de 1×1, entonces $\det A = a$.
2. Si A es una matriz de 2×2, entonces $\det A = a_{11}a_{12} - a_{12}a_{21}$.
3. Si A es una matriz de 3×3, entonces:
$$\det A = a_{11}\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12}\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13}\begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}$$
4. Si A es una matriz de $n \times n$, cuando $n > 1$, entonces el determinante esta dado por:
$$\det A = \sum_{j=1}^{n} a_{ij}A_{ij} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij}M_{ij} \qquad \text{para cada} \qquad i = 1, 2, \ldots, n \tag{3.4}$$
o bien por
$$\det A = \sum_{i=1}^{n} a_{ij}A_{ij} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij}M_{ij} \qquad \text{para cada} \qquad j = 1, 2, \ldots, n \tag{3.5}$$
donde A_{ij} es el cofactor ij de A y M_{ij} es el menor ij de A.

Definición 3.1.2 Menor. Sea A una matriz de $n \times n$ y sea M_{ij} la matriz de $(n-1)\times(n-1)$ obtenida de A eliminando el renglón i y la columna j. M_{ij} se llama el menor ij de A.

Definición 3.1.3 Cofactor. Sea A una matriz de $n \times n$. El cofactor ij de A, denotado por A_{ij}, está dado por:
$$A_{ij} = (-1)^{i+j} |M_{ij}|$$
Esto es; el cofactor ij de A se obtiene tomando el determinante del menor ij y multiplicándolo por $(-1)^{i+j}$. Se puede observar que
$$(-1)^{i+j} = \begin{cases} 1 \text{ si } i+j \text{ es par} \\ -1 \text{ si } i+j \text{ es impar} \end{cases}$$

[5]No se debe de confundir esta notación con la del valor absoluto.

Ejemplo 3.1.1 Calcular el determinante de la siguiente matriz, usando la definición 3.1.1.

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

Para calcular el det A, es más adecuado escoger el renglón o la columna que tengan más ceros, para realizar menos cálculos. En este caso es escoge el segundo renglón $(i=2)$ por lo que se ocupará la ecuación (3.4); desarrollándola se obtiene:

$$\det A = a_{21}A_{21}+a_{22}A_{22}+a_{23}A_{23}+a_{24}A_{24} \tag{3.6}$$

Se calculan los menores M_{21}, M_{22}, M_{23} y M_{24} por la definición 3.1.2. Entonces para obtener el menor M_{21}, se elimina el segundo renglón y la primera columna de A, obteniéndose:

$$M_{21}=\begin{pmatrix}0 & 1 & 2\\ -1 & 3 & 1\\ -1 & 4 & 3\end{pmatrix}$$

Para calcular el menor M_{22}, se elimina el segundo renglón y la segunda columna de A, obteniéndose:

$$M_{22}=\begin{pmatrix}2 & 1 & 2\\ 2 & 3 & 1\\ 3 & 4 & 3\end{pmatrix}$$

Se sigue el mismo procedimiento para calcular los menores restantes, obteniéndose:

$$M_{23}=\begin{pmatrix}2 & 0 & 2\\ 2 & -1 & 1\\ 3 & -1 & 3\end{pmatrix}\qquad M_{24}=\begin{pmatrix}2 & 0 & 1\\ 2 & -1 & 3\\ 3 & -1 & 4\end{pmatrix}$$

Después se obtienen los cofactores A_{21}, A_{22}, A_{23} y A_{24} por las definiciones 3.1.1 y 3.1.3:

$$A_{21}=(-1)^3\begin{vmatrix}0 & 1 & 2\\ -1 & 3 & 1\\ -1 & 4 & 3\end{vmatrix}=(-1)^3\left(0\begin{vmatrix}3 & 1\\ 4 & 3\end{vmatrix}-1\begin{vmatrix}-1 & 1\\ -1 & 3\end{vmatrix}+2\begin{vmatrix}-1 & 3\\ -1 & 4\end{vmatrix}\right)=$$

$$(-1[0((3\times 3)-(1\times 4))+(-1)((-1\times 3)-(1\times -1))+2((-1\times 4)-(3\times -1))])=$$

$$(-1[0(9-4)-1(-3+1)+2(-4+3)])=(-1)(0+2-2)=0$$

$$\Rightarrow A_{21}=0$$

$$A_{22}=(-1)^4\begin{vmatrix}2 & 1 & 2\\ 2 & 3 & 1\\ 3 & 4 & 3\end{vmatrix}=(-1)^4\left(2\begin{vmatrix}3 & 1\\ 4 & 3\end{vmatrix}-1\begin{vmatrix}2 & 1\\ 3 & 3\end{vmatrix}+2\begin{vmatrix}2 & 3\\ 3 & 4\end{vmatrix}\right)=$$

$$(1[2((3\times 3)-(1\times 4))+(-1)((2\times 3)-(1\times 3))+2((2\times 4)-(3\times 3))])=$$

$$(1)(10-3-2)=5$$
$$\Rightarrow A_{22}=5$$

$$A_{23}=(-1)^5\begin{vmatrix}2 & 0 & 2\\ 2 & -1 & 1\\ 3 & -1 & 3\end{vmatrix}=(-1)^5\left(2\begin{vmatrix}-1 & 1\\ -1 & 3\end{vmatrix}-0\begin{vmatrix}2 & 1\\ 3 & 3\end{vmatrix}+2\begin{vmatrix}2 & -1\\ 3 & -1\end{vmatrix}\right)=$$

$$(-1[2((-1\times 3)-(1\times -1))+(-0)((2\times 3)-(1\times 3))+2((2\times -1)-(-1\times 3))])=$$
$$(-1)(-4-0+2)=2$$
$$\Rightarrow A_{23}=2$$

$$A_{24}=(-1)^6\begin{vmatrix}2 & 0 & 1\\ 2 & -1 & 3\\ 3 & -1 & 4\end{vmatrix}=(-1)^6\left(2\begin{vmatrix}-1 & 3\\ -1 & 4\end{vmatrix}-0\begin{vmatrix}2 & 3\\ 3 & 4\end{vmatrix}+1\begin{vmatrix}2 & -1\\ 3 & -1\end{vmatrix}\right)=$$

$$(1[2((-1\times 4)-(3\times -1))+(-0)((2\times 4)-(3\times 3))+1((2\times -1)-(-1\times 3))])=$$
$$(1)(-2+0+1)=-1$$
$$\Rightarrow A_{24}=-1$$

Finalmente se sustituyen los cofactores A_{21}, A_{22}, A_{23} y A_{24} en la ecuación (3.6), obteniéndose:

$$\det A=(1\times 0)+(1\times 5)+(0\times 2)+(2\times -1)=3$$
$$\det A=3$$

3.1.2 Propiedades de los determinantes

1. Si A es triangular, esto es A sólo tiene ceros por encima o por debajo de la diagonal principal, entonces $|A|$ es igual al producto de los elementos de la diagonal.

2. Si cualquier renglón o columna de A es un vector cero, entonces $|A|=0$.

3. Si el renglón i o la columna j de A se multiplica por un escalar c, para obtener una matriz triangular superior (inferior); ó una matriz diagonal, entonces $|A|$ se multiplica por el recíproco de c.

4. Si A, B y C son idénticas excepto por el renglón i, y el renglón i de C es la suma de los *i-ésimos* renglones de A y B. Entonces $|C|=|A|+|B|$. La misma afirmación es cierta para columnas.

5. Si B se obtiene al intercambiar dos renglones o columnas de A, entonces $|B|=-|A|$.

6. Si A tiene dos renglones o columnas iguales, entonces $|A|=0$.

7. Si un renglón (columna) de A es múltiplo escalar de otro renglón (columna), entonces $|A|=0$.

8. Si se suma un múltiplo escalar de un renglón (columna) de A a otro renglón (columna) de A, entonces el determinante no cambia.

3.2 Planteamiento de problemas de sistemas lineales

Para resolver problemas mediante el planteamiento de un sistema de ecuaciones lineales, se deben seguir varios pasos:

1. Plantear el problema, entendiendo su enunciado y convirtiéndolo en ecuaciones con coeficientes, constantes y variables o incógnitas.

2. Analizar el tipo de sistema que se obtiene.

3. Elegir un método de resolución (algebraico o gráfico) y aplicarlo.

4. Estudiar si las soluciones obtenidas son pertinentes en el contexto del problema.

5. Comprobar las soluciones en las ecuaciones planteadas.

3.3 Métodos exactos

Cuando la matriz asociada al sistema se clasifica como densa, los métodos exactos son los más usuales y recomendables, porque dan la solución analítica del problema. Salvo algunos casos todos éstos métodos emplean las operaciones elementales de una matriz, para hallar la solución. Usando estas operaciones se simplifica el sistema a tal grado que la solución sea fácil de determinar. Existen varios métodos en esta categoría.

3.3.1 Método de Gauss

Uno de los métodos más antiguos para resolver sistemas de ecuaciones lineales es el *método de Gauss*, el cual tiene mucha importancia debido a que la mayoría de los métodos directos son variantes de éste.

Consiste básicamente en realizar los siguientes pasos:

1. Trasformar la matriz aumentada columna por columna a un sistema triangular superior, aplicando operaciones elementales;

$$\left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & b_n \end{array}\right) \longrightarrow \left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ 0 & a_{22}' & a_{23}' & \cdots & a_{2n}' & b_2' \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn}' & b_n' \end{array}\right)$$

es decir, dado un sistema de “n” ecuaciones con “n” incógnitas se trata de obtener un sistema equivalente cuya primera ecuación tenga n incógnitas, la segunda $n-1$, la tercera $n-2$, y así sucesivamente hasta llegar a la última ecuación, que tendrá una sola incógnita, **(eliminación hacia adelante)**.

2. Resolver la última ecuación, a continuación la penúltima, y así hasta llegar a la primera, **(sustitución hacia atrás ó sustitución regresiva)**.

Operaciones elementales que se pueden aplicar a una matriz aumentada:

1. Multiplicar (o dividir) un renglón por (entre) un número distinto de cero.

2. Sumar el múltiplo de un renglón a otro renglón.

3. Intercambiar dos renglones.

Ejemplo 3.3.1 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss.

$$\begin{array}{rcrcrcr} x_1 & + & x_2 & + & x_3 & = & 11 \\ 2x_1 & - & x_2 & + & x_3 & = & 5 \\ 3x_1 & + & 2x_2 & + & x_3 & = & 24 \end{array}$$

Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A}=\begin{pmatrix}1&1&1\\2&-1&1\\3&2&1\end{pmatrix}\qquad \mathbf{b}=\begin{pmatrix}11\\5\\24\end{pmatrix}$$

Primero se calcula el det A por la definición 3.3.1, para garantizar la existencia de una solución única.

$$\det A=1\begin{vmatrix}-1&1\\2&1\end{vmatrix}-1\begin{vmatrix}2&1\\3&1\end{vmatrix}+1\begin{vmatrix}2&-1\\3&2\end{vmatrix}=$$

$$1((-1\times1)-(1\times2))-1((2\times1)-(1\times3))+1((2\times2)-(-1\times3))=$$
$$1(-3)-1(-1)+1(7)=5$$

Como det $A\neq 0$ se empieza a aplicar el método; combinando las matrices A y b para obtener su forma aumentada, y se empiezan a realizar operaciones elementales para obtener una matriz triangular superior.

$$\left(\begin{array}{ccc|c}1&1&1&11\\2&-1&1&5\\3&2&1&24\end{array}\right)\xrightarrow[R_3\to-3R_1+R_3]{R_2\to-2R_1+R_2}\left(\begin{array}{ccc|c}1&1&1&11\\0&-3&-1&-17\\0&-1&-2&-9\end{array}\right)\xrightarrow{R_3\to-\frac{1}{3}R_2+R_3}$$

$$\left(\begin{array}{ccc|c}1&1&1&11\\0&-3&-1&-17\\0&0&-\frac{5}{3}&-\frac{10}{3}\end{array}\right)$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas.

$$\Rightarrow\begin{array}{rcrcrcr}x_1&+&x_2&+&x_3&=&11\\&-&3x_2&-&x_3&=&17\\&&&-&\frac{5}{3}x_3&=&-\frac{10}{3}\end{array}\qquad\qquad\begin{array}{rcl}x_1&=&4\\x_2&=&5\\x_3&=&2\end{array}$$

Comprobación: [6]

$$\begin{array}{rcrcrcr}(4)&+&(5)&+&(2)&=&11\\2(4)&-&(5)&+&(2)&=&5\\3(4)&+&2(5)&+&(2)&=&24\end{array}$$

Ejemplo 3.3.2 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss.

$$\begin{array}{rcrcrcr}4x_1&-&x_2&+&x_3&=&8\\2x_1&+&5x_2&+&2x_3&=&3\\x_1&+&2x_2&+&4x_3&=&11\end{array}$$

Escribiendo el sistema de ecuaciones en notación matricial se tiene:

---

[6] Se puede comprobar que el resultado obtenido es correcto, si al sustituir cada una de la incógnitas $(x_1,x_2,\ldots,x_n)$ en el sistema original, se cumplen las igualdades.

$$\mathbf{A} = \begin{pmatrix} 4 & -1 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 8 \\ 3 \\ 11 \end{pmatrix}$$

Primero se calcula el det A por la definición 3.3.1, para garantizar la existencia de una solución única.

$$\det A = 4\begin{vmatrix} 5 & 2 \\ 2 & 4 \end{vmatrix} - (-1)\begin{vmatrix} 2 & 2 \\ 1 & 4 \end{vmatrix} + 1\begin{vmatrix} 2 & 5 \\ 1 & 2 \end{vmatrix} =$$

$$4((5\times 4)-(2\times 2))-(-1)((2\times 4)-(2\times 1))+1((2\times 2)-(5\times 1)) =$$
$$4(16)+1(6)+1(-1)=69$$

Como det $A \neq 0$ se empieza a aplicar el método; combinando las matrices A y b para obtener su forma aumentada, y se empiezan a realizar operaciones elementales para obtener una matriz triangular superior.

$$\left(\begin{array}{ccc|c} 4 & -1 & 1 & 8 \\ 2 & 5 & 2 & 3 \\ 1 & 2 & 4 & 11 \end{array}\right) \xrightarrow{R_1 \leftrightarrow R_3} \left(\begin{array}{ccc|c} 1 & 2 & 4 & 11 \\ 2 & 5 & 2 & 3 \\ 4 & -1 & 1 & 8 \end{array}\right) \xrightarrow[R_3 \to -4R_1 + R_3]{R_2 \to -2R_1 + R_2}$$

$$\left(\begin{array}{ccc|c} 1 & 2 & 4 & 11 \\ 0 & 1 & -6 & -19 \\ 0 & -9 & -15 & -36 \end{array}\right) \xrightarrow{R_3 \to 9R_2 + R_3} \left(\begin{array}{ccc|c} 1 & 2 & 4 & 11 \\ 0 & 1 & -6 & -19 \\ 0 & 0 & -69 & -207 \end{array}\right)..$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas.

$$\Rightarrow \begin{array}{rcrcrcr} x_1 & + & 2x_2 & + & 4x_3 & = & 11 \\ & & x_2 & - & 6x_3 & = & -19 \\ & & & - & 69x_3 & = & -207 \end{array} \qquad \begin{array}{rcr} x_1 & = & 1 \\ x_2 & = & -1 \\ x_3 & = & 3 \end{array}$$

Comprobación:

$$\begin{array}{rcrcrcr} 4(1) & - & (-1) & + & (3) & = & 8 \\ 2(1) & + & 5(-1) & + & 2(3) & = & 3 \\ (1) & + & 2(-1) & + & 4(3) & = & 11 \end{array}$$

Con el método de Gauss, se puede:

1. **Resolver el sistema de ecuaciones.**
2. **Obtener el determinante.**
3. **Factorizar la matriz $\mathbf{A}$ en la forma $\mathbf{LU}$.**

Si el método de Gauss puede aplicarse al sistema $Ax=b$, sin intercambio de renglones, entonces la matriz A puede factorizarse como el producto $A = LU$.

Definición 3.3.1 Una matriz invertible A admite una factorización triangular o factorización LU si puede expresarse como el producto de una matriz triangular inferior L, cuyos elementos diagonales son todos iguales a 1, por una matriz triangular superior U:

$$A = LU$$
o, escrito de manera desarrollada,

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}}_{\text{A}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & 1 \end{pmatrix}}_{\text{L}} \underbrace{\begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ 0 & u_{22} & u_{23} & \dots & u_{2n} \\ 0 & 0 & u_{33} & \dots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & u_{nn} \end{pmatrix}}_{\text{U}}$$

La condición de que A sea invertible implica que $u_{kk} \neq 0$ para todo k.

Donde U es la matriz final que se obtiene al aplicar el método de Gauss y L es la matriz formada por los multiplicadores de cada uno de los renglones (con el signo contrario) para obtener U.

Nota: Ésta definición sólo se aplica para el método de Doolittle (Ver capítulo 4: sección 4.1).

4. Obtener la inversa.

Ejemplo 3.3.3 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss; (det $A = -1$, obtenido con Maple).

$$\begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & & & - & 2x_4 & = & -1 \\ 3x_1 & - & 12x_2 & - & 2x_3 & - & 6x_4 & = & -7 \\ -\ 2x_1 & + & 10x_2 & + & 2x_3 & + & 5x_4 & = & 6 \\ -\ x_1 & + & 6x_2 & + & x_3 & + & 3x_4 & = & 4 \end{array} \tag{3.8}$$

Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 1 & -3 & 0 & -2 \\ 3 & -12 & -2 & -6 \\ -2 & 10 & 2 & 5 \\ -1 & 6 & 1 & 3 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} -1 \\ -7 \\ 6 \\ 4 \end{pmatrix}$$

Combinando estas matrices para obtener su forma aumentada, se empiezan a realizar operaciones elementales para obtener una matriz triangular superior.

$$\left(\begin{array}{rrrr|r} 1 & -3 & 0 & -2 & -1 \\ 3 & -12 & -2 & -6 & -7 \\ -2 & 10 & 2 & 5 & 6 \\ -1 & 6 & 1 & 3 & 4 \end{array}\right) \xrightarrow[\substack{R_3 \to 2R_1 + R_3 \\ R_4 \to R_1 + R_4}]{R_2 \to -3R_1 + R_2} \left(\begin{array}{rrrr|r} 1 & -3 & 0 & -2 & -1 \\ 0 & -3 & -2 & 0 & -4 \\ 0 & 4 & 2 & 1 & 4 \\ 0 & 3 & 1 & 1 & 3 \end{array}\right) \xrightarrow[R_4 \to R_2 + R_4]{R_3 \to \frac{4}{3}R_2 + R_3}$$

$$\left(\begin{array}{rrrr|r} 1 & -3 & 0 & -2 & -1 \\ 0 & -3 & -2 & 0 & -4 \\ 0 & 0 & -\frac{2}{3} & 1 & -\frac{4}{3} \\ 0 & 0 & -1 & 1 & -1 \end{array}\right) \xrightarrow{R_4 \to -\frac{3}{2}R_3 + R_4} \left(\begin{array}{rrrr|r} 1 & -3 & 0 & -2 & -1 \\ 0 & -3 & -2 & 0 & -4 \\ 0 & 0 & -\frac{2}{3} & 1 & -\frac{4}{3} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{array}\right) ..$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas.

$$\Rightarrow \quad \begin{array}{rcrcrcrcl} x_1 & - & 3x_2 & & & - & 2x_4 & = & -1 \\ & - & 3x_2 & - & 2x_3 & & & = & -4 \\ & & & - & \frac{2}{3}x_3 & + & x_4 & = & -\frac{4}{3} \\ & & & & & - & \frac{1}{2}x_4 & = & 1 \end{array} \qquad \begin{array}{rcr} x_1 & = & 1 \\ x_2 & = & 2 \\ x_3 & = & -1 \\ x_4 & = & -2 \end{array}$$

Comprobación:

$$\begin{array}{rrrrrrrrrr} & (1) & - & 3(2) & & & - & 2(-2) & = & -1 \\ & 3(1) & - & 12(2) & - & 2(-1) & - & 6(-2) & = & -7 \\ - & 2(1) & + & 10(2) & + & 2(-1) & + & 5(-2) & = & 6 \\ - & (1) & + & 6(2) & + & (-1) & + & 3(-2) & = & 4 \end{array}$$

Ejemplo 3.3.4 Obtener el determinante del sistema de ecuaciones lineales (3.8), por el método de Gauss.
Por las propiedades (1) y (8) de los determinantes.

$$\left(\begin{array}{cccc|c} 1 & -3 & 0 & -2 & -1 \\ 0 & -3 & -2 & 0 & -4 \\ 0 & 0 & -\frac{2}{3} & 1 & -\frac{4}{3} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{array}\right) \qquad |A| = (1)(-3)\left(-\frac{2}{3}\right)\left(-\frac{1}{2}\right) = -1 \qquad |A| = -1$$

Ejemplo 3.3.5 Factorizar la matriz A en la forma LU, formada por el sistema de ecuaciones lineales (3.8), por el método de Gauss.
La matriz U, es la matriz triangular superior que resulta al aplicar el método de Gauss; obtenida en el ejemplo 3.3.3:

$$\underbrace{\begin{pmatrix} 1 & -3 & 0 & -2 \\ 0 & -3 & -2 & 0 \\ 0 & 0 & -\frac{2}{3} & 1 \\ 0 & 0 & 0 & -\frac{1}{2} \end{pmatrix}}_{\text{U}}$$

Para obtener L se forma la matriz triangular inferior con unos en la diagonal y debajo de esta se colocan los escalares por los que multiplicamos cada uno de los renglones para obtener la matriz triangular superior, con el signo contrario, obteniéndose:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -2 & -\frac{4}{3} & 1 & 0 \\ -1 & -1 & \frac{3}{2} & 1 \end{pmatrix}}_{\text{L}}$$

Finalmente, se verifican los resultados anteriores: LU=A

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \\ -2 & -\frac{4}{3} & 1 & 0 \\ -1 & -1 & \frac{3}{2} & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 1 & -3 & 0 & -2 \\ 0 & -3 & -2 & 0 \\ 0 & 0 & -\frac{2}{3} & 1 \\ 0 & 0 & 0 & -\frac{1}{2} \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} 1 & -3 & 0 & -2 \\ 3 & -12 & -2 & -6 \\ -2 & 10 & 2 & 5 \\ -1 & 6 & 1 & 3 \end{pmatrix}}_{A}$$

Ejemplo 3.3.5 Obtener la inversa del sistema de ecuaciones lineales (3.8), por el método de Gauss.

Se empieza combinando la matriz A con la identidad I para obtener su forma aumentada $[A|I]$, después se realizan operaciones elementales sobre los renglones para convertir a A en una matriz triangular superior U. (Se realizan las mismas operaciones en I).

$$\left(\begin{array}{cccc|cccc} 1 & -3 & 0 & -2 & 1 & 0 & 0 & 0 \\ 3 & -12 & -2 & -6 & 0 & 1 & 0 & 0 \\ -2 & 10 & 2 & 5 & 0 & 0 & 1 & 0 \\ -1 & 6 & 1 & 3 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow[\substack{R_3 \to 2R_1+R_3 \\ R_4 \to R_1+R_4}]{R_2 \to -3R_1+R_2} \left(\begin{array}{cccc|cccc} 1 & -3 & 0 & -2 & 1 & 0 & 0 & 0 \\ 0 & -3 & -2 & 0 & -3 & 1 & 0 & 0 \\ 0 & 4 & 2 & 1 & 2 & 0 & 1 & 0 \\ 0 & 3 & 1 & 1 & 1 & 0 & 0 & 1 \end{array}\right)$$

$$\xrightarrow[R_4 \to R_2+R_4]{R_3 \to \frac{4}{3}R_2+R_3} \left(\begin{array}{cccc|cccc} 1 & -3 & 0 & -2 & 1 & 0 & 0 & 0 \\ 0 & -3 & -2 & 0 & -3 & 1 & 0 & 0 \\ 0 & 0 & -\frac{2}{3} & 1 & -2 & \frac{4}{3} & 1 & 0 \\ 0 & 0 & -1 & 1 & -2 & 1 & 0 & 1 \end{array}\right) \xrightarrow{R_4 \to -\frac{3}{2}R_3+R_4}$$

$$\left(\underbrace{\begin{array}{cccc} 1 & -3 & 0 & -2 \\ 0 & -3 & -2 & 0 \\ 0 & 0 & -\frac{2}{3} & 1 \\ 0 & 0 & 0 & -\frac{1}{2} \end{array}}_{U} \left| \begin{array}{cccc} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -2 & \frac{4}{3} & 1 & 0 \\ 1 & -1 & -\frac{3}{2} & 1 \end{array}\right.\right)$$

Para obtener la inversa de A, se forman los siguientes sistemas de ecuaciones lineales. Donde la matriz de coeficientes esta formada por U y la matriz de términos independientes (b) esta formada por cada uno de los vectores que forman la matriz resultante de aplicar la operaciones elementales a I.

$$\underbrace{\left(\begin{array}{cccc|c} 1 & -3 & 0 & -2 & 1 \\ 0 & -3 & -2 & 0 & -3 \\ 0 & 0 & -\frac{2}{3} & 1 & -2 \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{array}\right)}_{\text{U}} \Rightarrow \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & & & - & 2x_4 & = & 1 \\ & - & 3x_2 & - & 2x_3 & & & = & -3 \\ & & & - & \frac{2}{3}x_3 & + & x_4 & = & -2 \\ & & & & & - & \frac{1}{2}x_4 & = & 1 \end{array} \quad \begin{array}{rcr} x_1 & = & 0 \\ x_2 & = & 1 \\ x_3 & = & 0 \\ x_4 & = & -2 \end{array}$$

$$\underbrace{\left(\begin{array}{cccc|c} 1 & -3 & 0 & -2 & 0 \\ 0 & -3 & -2 & 0 & 1 \\ 0 & 0 & -\frac{2}{3} & 1 & \frac{4}{3} \\ 0 & 0 & 0 & -\frac{1}{2} & -1 \end{array}\right)}_{\text{U}} \Rightarrow \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & & & - & 2x_4 & = & 0 \\ & - & 3x_2 & - & 2x_3 & & & = & 1 \\ & & & - & \frac{2}{3}x_3 & + & x_4 & = & \frac{4}{3} \\ & & & & & - & \frac{1}{2}x_4 & = & -1 \end{array} \quad \begin{array}{rcr} x_1 & = & 1 \\ x_2 & = & -1 \\ x_3 & = & 1 \\ x_4 & = & 2 \end{array}$$

$$\underbrace{\left(\begin{array}{cccc|c} 1 & -3 & 0 & -2 & 0 \\ 0 & -3 & -2 & 0 & 0 \\ 0 & 0 & -\frac{2}{3} & 1 & 1 \\ 0 & 0 & 0 & -\frac{1}{2} & -\frac{3}{2} \end{array}\right)}_{\text{U}} \Rightarrow \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & & & - & 2x_4 & = & 0 \\ & - & 3x_2 & - & 2x_3 & & & = & 0 \\ & & & - & \frac{2}{3}x_3 & + & x_4 & = & 1 \\ & & & & & - & \frac{1}{2}x_4 & = & -\frac{3}{2} \end{array} \quad \begin{array}{rcr} x_1 & = & 0 \\ x_2 & = & -2 \\ x_3 & = & 3 \\ x_4 & = & 3 \end{array}$$

$$\underbrace{\left(\begin{array}{cccc|c} 1 & -3 & 0 & -2 & 0 \\ 0 & -3 & -2 & 0 & 0 \\ 0 & 0 & -\frac{2}{3} & 1 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \end{array}\right)}_{\text{U}} \Rightarrow \begin{array}{rcrcrcrcr} x_1 & - & 3x_2 & & & - & 2x_4 & = & 0 \\ & - & 3x_2 & - & 2x_3 & & & = & 0 \\ & & & - & \frac{2}{3}x_3 & + & x_4 & = & 0 \\ & & & & & - & \frac{1}{2}x_4 & = & 1 \end{array} \quad \begin{array}{rcr} x_1 & = & 2 \\ x_2 & = & 2 \\ x_3 & = & -3 \\ x_4 & = & -2 \end{array}$$

Finalmente, la matriz formada por los vectores resultantes de cada uno de los sistemas anteriores, es la inversa de A.

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

Comprobación: Por la definición 3.3.2, (sección 3.3.3).

$$\mathbf{I} = \mathbf{AA}^{-1} = \begin{pmatrix} 1 & -3 & 0 & -2 \\ 3 & -12 & -2 & -6 \\ -2 & 10 & 2 & 5 \\ -1 & 6 & 1 & 3 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 2 \\ 1 & -1 & -2 & 2 \\ 0 & 1 & 3 & -3 \\ -2 & 2 & 3 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

3.3.2 Método de Gauss-Jordan

El *método de Gauss-Jordan* es una variación del método de Gauss. La principal diferencia consiste en que no sólo elimina los términos debajo de la diagonal principal sino también los que están sobre de ella.
Consiste básicamente en realizar el siguiente paso:

1. Transformar el lado derecho de la matriz aumentada columna por columna en la matriz identidad , aplicando operaciones elementales **(eliminación hacia adelante y hacia atrás)**;

$$\left(\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} & b_n \end{array}\right) \longrightarrow \left(\begin{array}{ccccc|c} 1 & 0 & 0 & \cdots & 0 & b_1' \\ 0 & 1 & 0 & \cdots & 0 & b_2' \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & b_n' \end{array}\right)$$

   quedando en el vector de términos independientes el resultado del sistema. En consecuencia no es necesario usar la sustitución hacia atrás para obtener la solución.

Con el método de Gauss-Jordan, se puede:

1. **Resolver el sistema de ecuaciones.**
2. **Obtener el determinante.**
3. **Obtener la inversa.**

Para obtener la inversa de A, por el método de Gauss-Jordan, se empieza combinando la matriz A con la identidad I para obtener su forma aumentada $[A\,|\,I]$, después se realizan operaciones elementales sobre los renglones para convertir a A en I. (Se realizan las mismas operaciones en I). Cuando la mitad izquierda de la matriz aumentada se reduce a la matriz identidad, la mitad derecha se convierte en A^{-1}.

Ejemplo 3.3.7 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss-Jordan; (det $A = -15$, obtenido con Maple).

$$\begin{array}{ccccccc} 4x_1 & + & x_2 & + & 2x_3 & = & 9(9) \\ 2x_1 & + & 4x_2 & - & x_3 & = & -5(10) \\ x_1 & + & x_2 & - & x_3 & = & -9 \end{array} \tag{3.9}$$

Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial tenemos:

$$\mathbf{A} = \begin{pmatrix} 4 & 1 & 2 \\ 2 & 4 & -1 \\ 1 & 1 & -1 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 9 \\ -5 \\ -9 \end{pmatrix}$$

Combinando estas matrices para obtener su forma aumentada, empezamos a realizar operaciones elementales para obtener la matriz identidad.

$$\left(\begin{array}{ccc|c} 4 & 1 & 2 & 9 \\ 2 & 4 & -1 & -5 \\ 1 & 1 & -1 & -9 \end{array}\right) \xrightarrow{R_1 \leftrightarrow R_3} \left(\begin{array}{ccc|c} 1 & 1 & -1 & -9 \\ 2 & 4 & -1 & -5 \\ 4 & 1 & 2 & 9 \end{array}\right) \xrightarrow[R_3 \to -4R_1 + R_3]{R_2 \to -2R_1 + R_2}$$

$$\left(\begin{array}{ccc|c} 1 & 1 & -1 & -9 \\ 0 & 2 & 1 & 13 \\ 0 & -3 & 6 & 45 \end{array}\right) \xrightarrow{\frac{1}{2}R_2} \left(\begin{array}{ccc|c} 1 & 1 & -1 & -9 \\ 0 & 1 & \frac{1}{2} & \frac{13}{2} \\ 0 & -3 & 6 & 45 \end{array}\right) \xrightarrow[R_3 \to 3R_2 + R_3]{R_1 \to -R_2 + R_1} \left(\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & -\frac{31}{2} \\ 0 & 1 & \frac{1}{2} & \frac{13}{2} \\ 0 & 0 & \frac{15}{2} & \frac{129}{2} \end{array}\right) \xrightarrow{\frac{2}{15}R_3}$$

$$\left(\begin{array}{ccc|c} 1 & 0 & -\frac{3}{2} & -\frac{31}{2} \\ 0 & 1 & \frac{1}{2} & \frac{13}{2} \\ 0 & 0 & 1 & \frac{43}{5} \end{array}\right) \xrightarrow[R_2 \to -\frac{1}{2}R_3 + R_2]{R_1 \to \frac{3}{2}R_3 + R_1} \left(\begin{array}{ccc|c} 1 & 0 & 0 & -\frac{13}{5} \\ 0 & 1 & 0 & \frac{11}{5} \\ 0 & 0 & 1 & \frac{43}{5} \end{array}\right) \qquad \begin{array}{rcl} x_1 & = & -\frac{13}{5} \\ x_2 & = & \frac{11}{5} \\ x_3 & = & \frac{43}{5} \end{array}$$

Comprobación:

$$\begin{array}{rcrcrcr} 4(-\dfrac{13}{5}) & + & (\dfrac{11}{5}) & + & 2(\dfrac{43}{5}) & = & 9 \\ 2(-\dfrac{13}{5}) & + & 4(\dfrac{11}{5}) & - & (\dfrac{43}{5}) & = & -5 \\ (-\dfrac{13}{5}) & + & (\dfrac{11}{5}) & - & (\dfrac{43}{5}) & = & -9 \end{array}$$

Ejemplo 3.3.8 Obtener el determinante del sistema de ecuaciones lineales (3.9), por el método de Gauss-Jordan.
Por las propiedades (1), (3), (5) y (8) de los determinantes.

$$\left(\begin{array}{ccc|c} 1 & 0 & 0 & -\frac{13}{5} \\ 0 & 1 & 0 & \frac{11}{5} \\ 0 & 0 & 1 & \frac{43}{5} \end{array}\right) \qquad |A| = (1)\,(1)\,(1)\,(-1)\left(\frac{2}{1}\right)\left(\frac{15}{2}\right) = -15 \qquad \therefore |A| = -15$$

Ejemplo 3.3.9 Obtener la inversa de la siguiente matriz por el método de Gauss-Jordan; (det $A = -9$, obtenido con Maple).

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

Como det $A \neq 0$ se empieza a aplicar el método, ya que se puede garantizar que la matriz A es invertible, por el teorema 3.3.3, (sección 3.3.3).
Se obtiene la matriz aumentada $[A|I]$ y se empiezan a realizar operaciones elementales para transformar A en I.

$$\left(\begin{array}{rrr|rrr} 1 & 2 & -1 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 1 & 0 \\ -1 & 1 & 2 & 0 & 0 & 1 \end{array}\right) \xrightarrow[R_3 \to R_1 + R_3]{R_2 \to -2R_1 + R_2} \left(\begin{array}{rrr|rrr} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & 3 & 3 & 1 & 0 & 1 \end{array}\right) \xrightarrow{-\frac{1}{3}R_2}$$

$$\left(\begin{array}{rrr|rrr} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 3 & 1 & 1 & 0 & 1 \end{array}\right) \xrightarrow[R_3 \to -3R_2 + R_3]{R_1 \to -2R_2 + R_1} \left(\begin{array}{rrr|rrr} 1 & 0 & \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 3 & -1 & 1 & 1 \end{array}\right) \xrightarrow{\frac{1}{3}R_3}$$

$$\left(\begin{array}{rrr|rrr} 1 & 0 & \frac{1}{3} & -\frac{1}{3} & \frac{2}{3} & 0 \\ 0 & 1 & -\frac{2}{3} & \frac{2}{3} & -\frac{1}{3} & 0 \\ 0 & 0 & 1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right) \xrightarrow[R_3 \to \frac{2}{3}R_3 + R_2]{R_1 \to -\frac{1}{3}R_3 + R_1} \left(\begin{array}{rrr|rrr} 1 & 0 & 0 & -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ 0 & 1 & 0 & \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ 0 & 0 & 1 & -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{array}\right)$$

Comprobación: Por la definición 3.3.2,(sección 3.3.3).

$$\mathbf{I} = \mathbf{AA}^{-1} = \begin{pmatrix} 1 & 2 & -1 \\ 2 & 1 & 0 \\ -1 & 1 & 2 \end{pmatrix} \begin{pmatrix} -\frac{2}{9} & \frac{5}{9} & -\frac{1}{9} \\ \frac{4}{9} & -\frac{1}{9} & \frac{2}{9} \\ -\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

3.3.3 Inversión de matrices

Definición 3.3.2 Se dice que A^{-1} es la inversa de la matriz A si y sólo si:

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

siendo I, la matriz identidad.

No todas las matrices tienen inversa. Cuando una matriz no tiene inversa se dice que es una matriz "singular". Cuando la inversa de una matriz existe, se dice que la matriz es invertible ó "no singular".
Una condición que debe de cumplir la matriz para que sea invertible es que sea cuadrada; ésto es evidente, pues la inversa debe verificar que el producto por la izquierda y por la derecha debe dar siempre la matriz identidad, y ésto sólo sucede cuando la matriz es cuadrada. Sin embargo, ésta es una condición necesaria pero no suficiente; es decir, no toda matriz que sea cuadrada es invertible.

Propiedades de la matrices invertibles

1. Si una matriz A es invertible, entonces su inversa es única.
2. Si una matriz A es invertible, entonces se cumple: $(A^{-1})^{-1} = A$.
3. Si A y B son dos matrices invertibles, entonces se cumple: $(AB)^{-1} = B^{-1}A^{-1}$.
4. Si una matriz A es invertible, entonces se cumple que la inversa de la transpuesta es igual a la transpuesta de la inversa: $(A^t)^{-1} = (A^{-1})^t$.

Para poder resolver un sistema de ecuaciones lineales por medio de la inversa lo que se busca es despejar el vector x del sistema $Ax = b$; aplicando la ley asociativa y las propiedades de la matriz identidad, se obtiene:

$$\begin{aligned} A^{-1}Ax &= A^{-1}b \\ (A^{-1}A)x &= A^{-1}b \\ Ix &= A^{-1}b \\ x &= A^{-1}b \end{aligned} \tag{3.10}$$

entonces al realizar la multiplicación de la inversa de la matriz de coeficientes por el vector de términos independientes, se obtendrá el vector solución.

Cálculo de la inversa para una matriz de 1×1

Teorema 3.3.1 Una matriz A de 1×1 es invertible si y sólo si su determinante es diferente de cero $(\det\ A \neq 0)$. La forma de calcular su inversa es:

$$A^{-1} = \frac{1}{A}$$

Cálculo de la inversa para una matriz de 2×2

Teorema 3.3.2 Una matriz A de 2×2 es invertible si y sólo si su determinante es diferente de cero $(\det A \neq 0)$. La forma de calcular su inversa es:

$$A^{-1} = \frac{1}{\det A}\begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix}$$

es decir los elementos que están en la diagonal principal se intercambian y los que están en la diagonal secundaria se les cambia el signo.

Cálculo de la inversa para una matriz de $\mathbf{n \times n}$

Teorema 3.3.3 Una matriz A de $n \times n$ es invertible si y sólo si su determinante es diferente de cero $(\det A \neq 0)$. La forma de calcular su inversa es:

$$A^{-1} = \frac{1}{\det A}\operatorname{adj}A$$

donde adj A es la matriz adjunta de A.
Para poder calcular la inversa de A por medio del teorema 3.3.3, es necesario definir una matriz adjunta:

Definición 3.3.3 Adjunta. Sea A una matriz de $n \times n$. La matriz adjunta de A, es la transpuesta de la matriz de cofactores de A:

$$adjA = (B)^t$$

donde B, es la matriz de cofactores de A.

Ejemplo 3.3.10 Obtener la inversa de la siguiente matriz por el teorema 3.3.3.

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

Primero se calcula det A por la definición 3.1.1, para garantizar que la matriz es invertible.

$$\det A = 1\begin{vmatrix} 2 & 4 \\ 3 & -3 \end{vmatrix} - 2\begin{vmatrix} 2 & 4 \\ 1 & -3 \end{vmatrix} + (-1)\begin{vmatrix} 2 & 2 \\ 1 & 3 \end{vmatrix}$$

$$1((2\times -3)-(4\times 3)) - 2((2\times -3)-(4\times 1)) + (-1)((2\times 3)-(2\times 1)) =$$
$$1(-18) - 2(-10) - 1(4) = -2$$

Como det $A \neq 0$, se calculan los menores por la definición 3.1.2; los menores serán matrices de 2×2, entonces para obtener el menor M_{11}, se elimina el primer renglón y la primera columna de A, obteniéndose:

$$M_{11} = \begin{pmatrix} 2 & 4 \\ 3 & -3 \end{pmatrix}$$

Para calcular el menor M_{12}, se elimina el primer renglón y la segunda columna de A, obteniéndose:

$$M_{12} = \begin{pmatrix} 2 & 4 \\ 1 & -3 \end{pmatrix}$$

Para calcular el menor M_{13}, se elimina el primer renglón y la tercera columna de A, obteniéndose:

$$M_{13} = \begin{pmatrix} 2 & 2 \\ 1 & 3 \end{pmatrix}$$

Se sigue el mismo procedimiento para calcular los menores restantes, obteniéndose:

$$M_{21} = \begin{pmatrix} 2 & -1 \\ 3 & -3 \end{pmatrix} \qquad M_{22} = \begin{pmatrix} 1 & -1 \\ 1 & -3 \end{pmatrix} \qquad M_{23} = \begin{pmatrix} 1 & 2 \\ 1 & 3 \end{pmatrix}$$

$$M_{31} = \begin{pmatrix} 2 & -1 \\ 2 & 4 \end{pmatrix} \qquad M_{32} = \begin{pmatrix} 1 & -1 \\ 2 & 4 \end{pmatrix} \qquad M_{33} = \begin{pmatrix} 1 & 2 \\ 2 & 2 \end{pmatrix}$$

Después se obtienen los cofactores por la definición 3.1.3:

$$A_{11} = (-1)^2 \begin{vmatrix} 2 & 4 \\ 3 & -3 \end{vmatrix} = -18 \qquad A_{12} = (-1)^3 \begin{vmatrix} 2 & 4 \\ 1 & -3 \end{vmatrix} = 10 \qquad A_{13} = (-1)^4 \begin{vmatrix} 2 & 2 \\ 1 & 3 \end{vmatrix} = 4$$

$$A_{21} = (-1)^3 \begin{vmatrix} 2 & -1 \\ 3 & -3 \end{vmatrix} = 3 \qquad A_{22} = (-1)^4 \begin{vmatrix} 1 & -1 \\ 1 & -3 \end{vmatrix} = -2 \qquad A_{23} = (-1)^5 \begin{vmatrix} 1 & 2 \\ 1 & 3 \end{vmatrix} = -1$$

$$A_{31} = (-1)^4 \begin{vmatrix} 2 & -1 \\ 2 & 4 \end{vmatrix} = 10 \qquad A_{32} = (-1)^5 \begin{vmatrix} 1 & -1 \\ 2 & 4 \end{vmatrix} = -6 \qquad A_{33} = (-1)^6 \begin{vmatrix} 1 & 2 \\ 2 & 2 \end{vmatrix} = -2$$

$$B = \begin{pmatrix} -18 & 10 & 4 \\ 3 & -2 & -1 \\ 10 & -6 & -2 \end{pmatrix} \qquad \text{Matrizdecofactores}$$

Luego se obtiene adj A, por la definición 3.3.3:

$$adjA = \begin{pmatrix} -18 & 10 & 4 \\ 3 & -2 & -1 \\ 10 & -6 & -2 \end{pmatrix}^t = \begin{pmatrix} -18 & 3 & 10 \\ 10 & -2 & -6 \\ 4 & -1 & -2 \end{pmatrix}$$

Finalmente se obtiene A^{-1}, por el teorema 3.3.3:

$$A^{-1} = \frac{1}{\det A}\operatorname{adj}A = -\frac{1}{2}\begin{pmatrix} -18 & 3 & 10 \\ 10 & -2 & -6 \\ 4 & -1 & -2 \end{pmatrix} = \begin{pmatrix} 9 & -\frac{3}{2} & -5 \\ -5 & 1 & 3 \\ -2 & \frac{1}{2} & 1 \end{pmatrix}$$

$$\mathbf{A}^{-1} = \begin{pmatrix} 9 & -\frac{3}{2} & -5 \\ -5 & 1 & 3 \\ -2 & \frac{1}{2} & 1 \end{pmatrix}$$

Comprobación: Por la definición 3.3.2.

$$\mathbf{I} = \mathbf{A}\mathbf{A}^{-1} = \begin{pmatrix} 1 & 2 & 1 \\ 2 & 2 & 4 \\ 1 & 3 & -3 \end{pmatrix}\begin{pmatrix} 9 & -\frac{3}{2} & -5 \\ -5 & 1 & 3 \\ -2 & \frac{1}{2} & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Con el ejemplo anterior se puede observar, que si $n > 3$, por lo general es más fácil calcular A^{-1} con otros métodos que usando adj A, ya que por ejemplo, para el caso de una matriz de 4×4 es necesario calcular 17 determinantes (16 para la adj A más det A). Sin embargo el teorema 3.3.3, es muy importante ya que, antes de utilizar cualquier método para calcular la inversa, el cálculo del det A, dice si A^{-1} existe o no.

3.3.4. Inversión de matrices particionadas

Cuando la matriz A es de orden tan grande que no se puede invertir con los métodos antes vistos debido a la capacidad limitada de almacenamiento de la computadora que se dispone, es posible dividirla en submatrices de orden menor.
El método que se verá a continuación, sólo se puede aplicar cuando la partición de una matriz A se puede realizar en cuatro submatrices; la única restricción para realizar la partición, es que las submatrices de la diagonal principal sean cuadradas para asegurar en cierta forma la existencia de las inversas que se requieran y puedan efectuarse los productos matriciales necesarios. Esto se ejemplifica en el cuadro 3.1. etc.

Inversión de matrices realizando 4 particiones

Si un sistema de ecuaciones lineales de la forma (3.2), se puede expresar en forma matricial como:

$$\left(\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right)\left(\begin{array}{c} x_1 \\ \hline x_2 \end{array}\right) = \left(\begin{array}{c} b_1 \\ \hline b_2 \end{array}\right) \tag{3.11}$$

en donde fijada la posición de la submatriz A_{22} en A quedan obligadas las otras submatrices que aparecen en (2). De esta manera, si A_{22} es de orden $m \times m$, los vectores x_2 y b_2 estarán formados por últimos m elementos de x y de b.

Orden de de la matriz	Ejemplos de como realizar 4 particiones
1	no es necesario particionar
2	no es necesario particionar
3	no es necesario particionar
4	$\left(\begin{array}{c\|ccc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right)$ $\left(\begin{array}{cc\|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right)$ $\left(\begin{array}{ccc\|c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right)$
5	$\left(\begin{array}{c\|cccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right)$ $\left(\begin{array}{cc\|ccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right)$ $\left(\begin{array}{ccc\|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right)$ $\left(\begin{array}{cccc\|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right)$
6	$\left(\begin{array}{c\|ccccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$ $\left(\begin{array}{cc\|cccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$ $\left(\begin{array}{ccc\|ccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$ $\left(\begin{array}{cccc\|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$ $\left(\begin{array}{ccccc\|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ \hline a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$
$\vdots$	$\vdots$

Cuadro 3.1: Ejemplos de como realizar 4 particiones.

Para facilitar los cálculos se sugiere que si el orden de la matriz A es múltiplo de 2 (4×4, 6×6, 8×8, etc.), se particione en submatrices del mismo tamaño; pero si el orden de la matriz A no es múltiplo de 2 (5×5, 7×7, 9×9, etc.), se tome la partición en donde las submatrices que forman la diagonal principal sean de orden similar; por ejemplo si el orden de la matriz es 5, se deben tomar la segunda o la tercera partición ejemplificadas en este cuadro, ya que es más fácil calcular la inversa de una matriz de 2×2 y de 3×3; que la de una matriz de 4×4, primera y cuarta partición ejemplificadas en este cuadro.

Para resolver el sistema matricial (3.11), se realizarán las operaciones matriciales indicadas en él, tomando en cuenta que los elementos de dicho sistema son a su vez matrices:

$$\begin{aligned} A_{11}x_1 &+ A_{12}x_2 = b_1 \\ A_{21}x_1 &+ A_{22}x_2 = b_2 \end{aligned} \tag{3.12}$$

Se despeja el vector x_2 de la segunda ecuación de (3.12), para resolver el sistema de ecuaciones lineales; obteniéndose:

$$\begin{gathered} A_{22}x_2 = b_2 - A_{21}x_1 \\ x_2 = A_{22}^{-1}(b_2 - A_{21}x_1) \\ \Rightarrow x_2 = A_{22}^{-1}b_2 - A_{22}^{-1}A_{21}x_1 \end{gathered}$$

Sustituyendo x_2, en la primera ecuación de (3.12) y despejando de ésta el vector x_1, se obtiene:

$$\begin{gathered} A_{11}x_1 + A_{12}(A_{22}^{-1}b_2 - A_{22}^{-1}A_{21}x_1) = b_1 \\ A_{11}x_1 + A_{12}A_{22}^{-1}b_2 - A_{12}A_{22}^{-1}A_{21}x_1 = b_1 \end{gathered}$$

Se factoriza el vector x_1, para poderlo despejar:

$$\begin{gathered} (A_{11} - A_{12}A_{22}^{-1}A_{21})x_1 = b_1 - A_{12}A_{22}^{-1}b_2 \\ \Rightarrow x_1 = (A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}(b_1 - A_{12}A_{22}^{-1}b_2) \end{gathered}$$

Para facilitar la notación en la última expresión se renombran variables de la siguiente manera:

$$C = A_{11} - A_{12}A_{22}^{-1}A_{21} \tag{3.13}$$

$$D = A_{12}A_{22}^{-1} \tag{3.14}$$

$$x_1 = C^{-1}(b_1 - Db_2)$$

$$x_1 = C^{-1}b_1 - C^{-1}Db_2 \tag{3.15}$$

Sustituyendo x_1 en x_2, se obtiene:

$$\begin{gathered} x_2 = A_{22}^{-1}b_2 - A_{22}^{-1}A_{21}(C^{-1}b_1 - C^{-1}Db_2) \\ x_2 = A_{22}^{-1}b_2 - A_{22}^{-1}A_{21}C^{-1}b_1 + A_{22}^{-1}A_{21}C^{-1}Db_2 \\ x_2 = -A_{22}^{-1}A_{21}C^{-1}b_1 + (A_{22}^{-1} + A_{22}^{-1}A_{21}C^{-1}D)b_2 \end{gathered}$$

Nuevamente para facilitar la notación en la última expresión se renombran variables:

$$E = A_{22}^{-1}A_{21} \tag{3.16}$$

$$F = A_{22}^{-1} + EC^{-1}D \tag{3.17}$$

$$x_2 = -EC^{-1}b_1 + Fb_2 \tag{3.18}$$

El sistema de ecuaciones (3.15) y (3.18), se puede escribir de la forma matricial como:

$$\left(\begin{array}{c} x_1 \\ \hline x_2 \end{array}\right) = \left(\begin{array}{c|c} C^{-1} & -C^{-1}D \\ \hline -EC^{-1} & F \end{array}\right) \left(\begin{array}{c} b_1 \\ \hline b_2 \end{array}\right) \tag{3.19}$$

Es decir $x = A^{-1}b$, por lo que la inversa de la matriz A es:

$$\mathbf{A}^{-1} = \left(\begin{array}{c|c} C^{-1} & -C^{-1}D \\ \hline -EC^{-1} & F \end{array}\right) \tag{3.20}$$

en donde la matrices $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$ y $\mathbf{F}$ están definidas por las ecuaciones (3.13), (3.14), (3.16) y (3.17), respectivamente. Se puede observar que para obtener las matrices $\mathbf{C}$, $\mathbf{D}$, $\mathbf{E}$ y $\mathbf{F}$, la submatriz A_{22}, tiene que ser no singular.

Ejemplo 3.3.11 Resolver el siguiente sistema de ecuaciones lineales por el método de inversión de matrices particionadas; (det $A = 2$, obtenido con Maple).

$$\begin{array}{rcrcrcrcl} & & \sqrt{2}x_1 & & & + & 2\sqrt{2}x_3 & & = & \sqrt{2} \\ - & & 4\sqrt{2}x_1 & + & \sqrt{2}x_2 & & & & = & -4\sqrt{2} \\ & & & & & & x_3 & & = & 1 \\ & & & & & & 3x_3 & + \; x_4 & = & 0 \end{array}$$

Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} \sqrt{2} & 0 & 2\sqrt{2} & 0 \\ -4\sqrt{2} & \sqrt{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 1 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} \sqrt{2} \\ -4\sqrt{2} \\ 1 \\ 0 \end{pmatrix}$$

Primero se divide la matriz A en submatrices. Se puede observar en el cuadro (3.1) que la partición de la matriz de orden 4 se puede hacer de tres formas; en este caso se escogió dividir la matriz en submatrices de orden 2, para facilitar los cálculos; obteniéndose:

$$\mathbf{A} = \left(\begin{array}{cc|cc} \sqrt{2} & 0 & 2\sqrt{2} & 0 \\ -4\sqrt{2} & \sqrt{2} & 0 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 1 \end{array}\right)$$

Para garantizar que el método funcione se tiene que verificar que la submatriz A_{22} sea no singular:

$$\mathbf{A}_{22} = \begin{pmatrix} 1 & 0 \\ 3 & 1 \end{pmatrix} \qquad \text{donde} \qquad \det A = 1$$

Por el teorema ?:

$$\mathbf{A}_{22}^{-1} = 1\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}$$

Como A_{22} es no singular, se empieza obteniendo el valor $\mathbf{C}$ con la ecuación (?):

$$\mathbf{C} = \underbrace{\begin{pmatrix} \sqrt{2} & 0 \\ -4\sqrt{2} & \sqrt{2} \end{pmatrix}}_{\mathbf{A_{11}}} - \underbrace{\begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}}_{\mathbf{A_{12}}}\underbrace{\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}}_{\mathbf{A_{22}^{-1}}}\underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}}_{\mathbf{A_{21}}} = \begin{pmatrix} \sqrt{2} & 0 \\ -4\sqrt{2} & \sqrt{2} \end{pmatrix}$$

$$\Rightarrow \mathbf{C}^{-1} = \frac{1}{2}\begin{pmatrix} \sqrt{2} & 0 \\ 4\sqrt{2} & \sqrt{2} \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 \\ 2\sqrt{2} & \frac{\sqrt{2}}{2} \end{pmatrix}$$

Después se obtiene el valor de $\mathbf{D}$ con la ecuación (3.14):

$$\mathbf{D} = \underbrace{\begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}}_{\mathbf{A_{12}}}\underbrace{\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}}_{\mathbf{A_{22}^{-1}}} = \begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Rightarrow -\mathbf{C}^{-1}\mathbf{D} = -\begin{pmatrix} \frac{\sqrt{2}}{2} & 0 \\ 2\sqrt{2} & \frac{\sqrt{2}}{2} \end{pmatrix}\begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} -2 & 0 \\ -8 & 0 \end{pmatrix}$$

Luego se obtiene el valor de $\mathbf{E}$ con la ecuación (3.16):

$$\mathbf{E} = \underbrace{\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}}_{\mathbf{A_{22}^{-1}}}\underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}}_{\mathbf{A_{21}}} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Rightarrow -\mathbf{E}\mathbf{C}^{-1} = -\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} \frac{\sqrt{2}}{2} & 0 \\ 2\sqrt{2} & \frac{\sqrt{2}}{2} \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Después se obtiene el valor de $\mathbf{F}$ con la ecuación (3.17):

$$\Rightarrow \mathbf{F} = \underbrace{\begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}}_{\mathbf{A_{22}^{-1}}} + \underbrace{\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}}_{\mathbf{EC^{-1}}}\underbrace{\begin{pmatrix} 2\sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}}_{\mathbf{D}} = \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix}$$

Sustituyendo $\mathbf{C}^{-1}$, $-\mathbf{C}^{-1}\mathbf{D}$, $-\mathbf{E}\mathbf{C}^{-1}$ y $\mathbf{F}$ en (3.20), se obtiene:

$$\mathbf{A}^{-1} = \left(\begin{array}{cc|cc} \frac{\sqrt{2}}{2} & 0 & -2 & 0 \\ 2\sqrt{2} & \frac{\sqrt{2}}{2} & -8 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & -3 & 1 \end{array}\right)$$

Comprobación: Por la definición 3.3.2.

$$\mathbf{I} = \mathbf{AA}^{-1} = \begin{pmatrix} \sqrt{2} & 0 & 2\sqrt{2} & 0 \\ -4\sqrt{2} & \sqrt{2} & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 3 & 1 \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{2} & 0 & -2 & 0 \\ 2\sqrt{2} & \frac{\sqrt{2}}{2} & -8 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -3 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Finalmente, por la ecuación (3.10) se multiplica A^{-1} por b, a fin de obtener el valor de las incógnitas.

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \left(\begin{array}{cc|cc} \frac{\sqrt{2}}{2} & 0 & -2 & 0 \\ 2\sqrt{2} & \frac{\sqrt{2}}{2} & -8 & 0 \\ \hline 0 & 0 & 1 & 0 \\ 0 & 0 & -3 & 1 \end{array}\right) \begin{pmatrix} \sqrt{2} \\ -4\sqrt{2} \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -8 \\ 1 \\ -3 \end{pmatrix} \qquad \therefore \begin{array}{l} x_1 = -1 \\ x_2 = -8 \\ x_3 = 1 \\ x_4 = -3 \end{array}$$

Comprobación:

$$\begin{array}{rcrcrcrcl} & & \sqrt{2}(-1) & & & + & 2\sqrt{2}(1) & & & = & \sqrt{2} \\ - & & 4\sqrt{2}(-1) & + & \sqrt{2}(-8) & & & & & = & -4\sqrt{2} \\ & & & & & & (1) & & & = & 1 \\ & & & & & & 3(1) & + & 1(-3) & = & 0 \end{array}$$

3.3.5 Gauss-Jordan particionado

El sistema de ecuaciones lineales de la forma (3.2), se puede particionar en submatrices, para resolverse siguiendo el procedimiento de eliminación de Gauss-Jordan. A continuación se verá el procedimiento realizando 4 y 9 particiones. Se pueden obtener más particiones siguiendo el mismo procedimiento.

La única restricción para realizar la partición, al igual que en el tema de inversión de matrices particionadas; es que las submatrices de la diagonal principal sean cuadradas para asegurar en cierta forma la existencia de las inversas que se requieran y puedan efectuarse los productos matriciales necesarios. Ver cuadro 3.1 y 3.2.

Gauss-Jordan particionado realizando 4 particiones

Para 4 particiones se utilizará la siguiente matriz ampliada:

$$\left(\begin{array}{cc|c} A_{11} & A_{12} & b_1 \\ A_{21} & A_{22} & b_2 \end{array}\right) \tag{3.21}$$

Una vez fijada la posición de la submatriz A_{22} en A, quedan obligadas las otras submatrices que aparecen en (1); es decir, si A_{22} es de orden $m \times m$, el vector b_2 estará formado por últimos m elementos de b.

La transformación de la matriz A en la matriz identidad se logra realizando los siguientes pasos, tomando en cuenta que los elementos de dicho sistema son a su vez matrices:

1. Premultiplicación de todos los elementos del primer renglón de (3.21) por A_{11}^{-1}, la inversa del pivote; obteniéndose:

$$\left(\begin{array}{cc|c} I & A'_{12} & b'_1 \\ A_{21} & A_{22} & b_2 \end{array} \right) \tag{3.22}$$

en donde:

$$I = A_{11}^{-1}A_{11}$$
$$A_{12}' = A_{11}^{-1}A_{12}$$
$$b_1' = A_{11}^{-1}b_1$$
$$|A_{11}| \neq 0$$

2. Restar a los elementos matriciales del segundo renglón los correspondientes elementos del primer renglón premultiplicados por A_{21}, el elemento que se trata de eliminar; obteniéndose:

$$\left(\begin{array}{cc|c} I & A'_{12} & b'_1 \\ 0 & A'_{22} & b'_2 \end{array} \right) \tag{3.23}$$

en donde:

$$0 = A_{21} - A_{21}I$$
$$A_{22}' = A_{22} - A_{21}A_{12}'$$
$$b_2' = b_2 - A_{21}b_1'$$

3. Premultiplicación de los elementos del segundo renglón por $(A_{22}')^{-1}$, la inversa del nuevo pivote; obteniéndose:

$$\left(\begin{array}{cc|c} I & A'_{12} & b'_1 \\ 0 & I & c_2 \end{array} \right) \tag{3.24}$$

en donde:

$$I = (A_{22}')^{-1}A_{22}'$$
$$c_2 = (A_{22}')^{-1}b_2'$$
$$|A_{22}'| \neq 0$$

4. Finalmente, restar de los elementos del primer renglón los correspondientes elementos del segundo, premultiplicados previamente por A'_{12}, el elemento matricial que ahora se trata de eliminar; obteniéndose:

$$\left(\begin{array}{cc|c} I & 0 & c_1 \\ 0 & I & c_2 \end{array}\right) \tag{3.25}$$

en donde:

$$0 = A'_{12} - A'_{12}I$$
$$c_1 = b'_1 - A'_{12}c_2$$

La solución del sistema es:

$$x = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}$$

Ejemplo 3.3.12 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss-Jordan particionado, realizando 4 particiones; (det $A = -8$, obtenido con Maple).

$$\begin{array}{ccccccccc} x_1 & - & x_2 & + & 2x_3 & - & x_4 & = & 6 \\ x_1 & & & - & x_3 & + & x_4 & = & 4 \\ 2x_1 & + & x_2 & + & 3x_3 & - & 4x_4 & = & -2 \\ & - & x_2 & + & x_3 & - & x_4 & = & 5 \end{array}$$

Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 2 & -1 \\ 1 & 0 & -1 & 1 \\ 2 & 1 & 3 & -4 \\ 0 & -1 & 1 & -1 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 6 \\ 4 \\ -2 \\ 5 \end{pmatrix}$$

Combinando estas matrices para obtener su forma aumentada y dividiendo la matriz A en submatrices de orden 2 (ver cuadro 3.1); se obtiene:

$$[\mathbf{A}|\mathbf{b}] = \left(\begin{array}{cc|cc|c} 1 & -1 & 2 & -1 & 6 \\ 1 & 0 & -1 & 1 & 4 \\ \hline 2 & 1 & 3 & -4 & -2 \\ 0 & -1 & 1 & -1 & 5 \end{array}\right)$$

Se empieza tomando como pivote a la submatriz A_{11} y se obtiene su inversa:

$$\mathbf{A}_{11} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \qquad \text{donde} \qquad \det A = 1$$

Por el teorema 3.3.2:

$$\mathbf{A}_{11}^{-1} = 1\begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}$$

Como A_{11} es no singular, se premultiplican todos los elementos del primer renglón por A_{11}^{-1}, la inversa del pivote; obteniéndose:

$$\text{Matriz llevada a la forma (3.22)} \qquad \left(\begin{array}{cc|cc|c} 1 & 0 & -1 & 1 & 4 \\ 0 & 1 & -3 & 2 & -2 \\ \hline 2 & 1 & 3 & -4 & -2 \\ 0 & -1 & 1 & -1 & 5 \end{array}\right)$$

Operaciones:

$$I = A_{11}^{-1}A_{11} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A_{12}' = A_{11}^{-1}A_{12} = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ -3 & 2 \end{pmatrix}$$

$$b_1' = A_{11}^{-1}b_1 = \begin{pmatrix} 0 & 1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 6 \\ 4 \end{pmatrix} = \begin{pmatrix} 4 \\ -2 \end{pmatrix}$$

Después se restan a los elementos matriciales del segundo renglón los correspondientes elementos del primer renglón premultiplicados por A_{21}, el elemento que se trata de eliminar; obteniéndose:

$$\text{Matriz llevada a la forma (3.23)} \qquad \left(\begin{array}{cc|cc|c} 1 & 0 & -1 & 1 & 4 \\ 0 & 1 & -3 & 2 & -2 \\ \hline 0 & 0 & 8 & -8 & -8 \\ 0 & 0 & -2 & 1 & 3 \end{array}\right)$$

Operaciones:

$$0 = A_{21} - A_{21}I = \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix} - \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$A_{22}' = A_{22} - A_{21}A_{12}' = \begin{pmatrix} 3 & -4 \\ 1 & -1 \end{pmatrix} - \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} -1 & 1 \\ -3 & 2 \end{pmatrix} = \begin{pmatrix} 8 & -8 \\ -2 & 1 \end{pmatrix}$$

$$b_2' = b_2 - A_{21}b_1' = \begin{pmatrix} -2 \\ 5 \end{pmatrix} - \begin{pmatrix} 2 & 1 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 4 \\ -2 \end{pmatrix} = \begin{pmatrix} -8 \\ 3 \end{pmatrix}$$

Luego se toma como pivote a la submatriz A_{22}' y se obtiene su inversa:

$$\mathbf{A}_{22}' = \begin{pmatrix} 8 & -8 \\ -2 & 1 \end{pmatrix} \qquad \text{donde} \qquad \det A = -8$$

Por el teorema 3.3.2:

$$(\mathbf{A}_{22}')^{-1} = -\frac{1}{8}\begin{pmatrix} 1 & 8 \\ 2 & 8 \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{8} & -1 \\ -\dfrac{1}{4} & -1 \end{pmatrix}$$

Como $(A_{22}')^{-1}$ es no singular, se premultiplican todos los elementos del segundo renglón por $(A_{22}')^{-1}$, la inversa del nuevo pivote; obteniéndose:

$$\text{Matriz llevada a la forma (3.24)} \qquad \left(\begin{array}{cc|cc|c} 1 & 0 & -1 & 1 & 4 \\ 0 & 1 & -3 & 2 & -2 \\ \hline 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

Operaciones:

$$I = (A_{22}')^{-1} A_{22}' = \begin{pmatrix} -\frac{1}{8} & -1 \\ -\frac{1}{4} & -1 \end{pmatrix}\begin{pmatrix} 8 & -8 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$c_2 = (A_{22}')^{-1} b_2' = \begin{pmatrix} -\frac{1}{8} & -1 \\ -\frac{1}{4} & -1 \end{pmatrix}\begin{pmatrix} -8 \\ 3 \end{pmatrix} = \begin{pmatrix} -2 \\ -1 \end{pmatrix} \qquad \Rightarrow \begin{array}{rcr} x_3 & = & -2 \\ x_4 & = & -1 \end{array}$$

Finalmente se restan a los elementos matriciales del primer renglón los correspondientes elementos del segundo, premultiplicados previamente por A_{12}', el elemento matricial que ahora se trata de eliminar; obteniéndose:

$$\text{Matriz llevada a la forma (3.25)} \qquad \left(\begin{array}{cc|cc|c} 1 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & -6 \\ \hline 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 1 & -1 \end{array}\right)$$

Operaciones:

$$0 = A_{12}' - A_{12}' I = \begin{pmatrix} -1 & 1 \\ -3 & 2 \end{pmatrix} - \begin{pmatrix} -1 & 1 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$c_1 = b_1' - A_{12}' c_2 = \begin{pmatrix} 4 \\ -2 \end{pmatrix} - \begin{pmatrix} -1 & 1 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} -2 \\ -1 \end{pmatrix} = \begin{pmatrix} 3 \\ -6 \end{pmatrix} \qquad \Rightarrow \begin{array}{rcr} x_1 & = & 3 \\ x_2 & = & -6 \end{array}$$

Obteniendo así la solución del sistema:

$$\begin{array}{rcr} x_1 & = & 3 \\ x_2 & = & -6 \\ x_3 & = & -2 \\ x_4 & = & -1 \end{array}$$

Comprobación:

$$\begin{array}{ccccccccc} (3) & - & (-6) & + & 2(-2) & - & (-1) & = & 6 \\ (3) & & & - & (-2) & + & (-1) & = & 4 \\ 2(3) & + & (-6) & + & 3(-2) & - & 4(-1) & = & -2 \\ & - & (-6) & + & (-2) & - & (-1) & = & 5 \end{array}$$

Ejemplo 3.3.13 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss-Jordan particionado, realizando 4 particiones; (det $A = -56$, obtenido con Maple).

$$\begin{array}{ccccccccccc} x_1 & + & 2x_2 & + & 3x_3 & + & 4x_4 & + & 5x_5 & = & 33 \\ & - & x_2 & + & 2x_3 & - & x_4 & + & 2x_5 & = & 27 \\ x_1 & & & & & + & 2x_4 & - & x_5 & = & -10 \\ x_1 & + & x_2 & - & x_3 & + & x_4 & + & x_5 & = & 4 \\ & & & & & & & & 4x_5 & = & 36 \end{array} \tag{3.26}$$

Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & -1 & 2 & -1 & 2 \\ 1 & 0 & 0 & 2 & -1 \\ 1 & 1 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 33 \\ 27 \\ -10 \\ 4 \\ 36 \end{pmatrix}$$

Combinando estas matrices para obtener su forma aumentada y dividiendo la matriz A en 4 submatrices (ver cuadro 3.1); se obtiene:

$$[\mathbf{A}|\mathbf{b}] = \left(\begin{array}{ccc|cc|c} 1 & 2 & 3 & 4 & 5 & 33 \\ 0 & -1 & 2 & -1 & 2 & 27 \\ 1 & 0 & 0 & 2 & -1 & -10 \\ \hline 1 & 1 & -1 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Se empieza tomando como pivote a la submatriz A_{11} y se obtiene su inversa:

$$\mathbf{A}_{11} = \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & 2 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{donde} \qquad \det A_{11} = 7$$

Operaciones:

$$\det A_{11} = 1\begin{vmatrix} -1 & 2 \\ 0 & 0 \end{vmatrix} - 2\begin{vmatrix} 0 & 2 \\ 1 & 0 \end{vmatrix} + 3\begin{vmatrix} 0 & -1 \\ 1 & 0 \end{vmatrix} =$$

$$1((-1\times 0)-(2\times 0))-2((0\times 0)-(2\times 1))+3((0\times 0)-(-1\times 1)) =$$
$$1(0)-2(-2)+3(1) = 7$$

Para obtener A_{11}^{-1}, se utiliza la matriz aumentada $[A_{11} \mid I]$ y se empiezan a realizar operaciones elementales para transformar A en I. (Método de Gauss-Jordan).

$$\left(\begin{array}{ccc|ccc} 1 & 2 & 3 & 1 & 0 & 0 \\ 0 & -1 & 2 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{array}\right) \xrightarrow{R_1 \leftrightarrow R_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 3 & 1 & 0 & 0 \end{array}\right) \xrightarrow{R_3 \to -R_1 + R_3}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & -1 & 2 & 0 & 1 & 0 \\ 0 & 2 & 3 & 1 & 0 & -1 \end{array}\right) \xrightarrow{R_2 \to -1R_2} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -2 & 0 & -1 & 0 \\ 0 & 2 & 3 & 1 & 0 & -1 \end{array}\right) \xrightarrow{R_3 \to -2R_2 + R_3}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -2 & 0 & -1 & 0 \\ 0 & 0 & 7 & 1 & 2 & -1 \end{array}\right) \xrightarrow{R_3 \to \frac{1}{7}R_3} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & -2 & 0 & -1 & 0 \\ 0 & 0 & 1 & \frac{1}{7} & \frac{2}{7} & -\frac{1}{7} \end{array}\right) \xrightarrow{R_2 \to 2R_3 + R_2}$$

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & \frac{2}{7} & -\frac{3}{7} & -\frac{2}{7} \\ 0 & 0 & 1 & \frac{1}{7} & \frac{2}{7} & -\frac{1}{7} \end{array}\right)$$

La matriz del lado derecho, es la inversa de A.

$$\mathbf{A}_{11}^{-1} = \begin{pmatrix} 0 & 0 & 1 \\ \frac{2}{7} & -\frac{3}{7} & -\frac{2}{7} \\ \frac{1}{7} & \frac{2}{7} & -\frac{1}{7} \end{pmatrix}$$

Como A_{11} es no singular, se premultiplican todos los elementos del primer renglón por A_{11}^{-1}, la inversa del pivote; obteniéndose:

$$\text{Matriz llevada a la forma (3.22)} \qquad \left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & 2 & -1 & -10 \\ 0 & 1 & 0 & 1 & \frac{6}{7} & \frac{5}{7} \\ 0 & 0 & 1 & 0 & \frac{10}{7} & \frac{97}{7} \\ \hline 1 & 1 & -1 & 1 & 1 & 4 \\ 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Operaciones:

$$I = A_{11}^{-1}A_{11} = \begin{pmatrix} 0 & 0 & 1 \\ \frac{2}{7} & -\frac{3}{7} & -\frac{2}{7}(6) \\ \frac{1}{7} & \frac{2}{7} & -\frac{1}{7} \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 0 & -1 & 2(7) \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0(8) \\ 0 & 0 & 1 \end{pmatrix}$$

$$A_{12}^{'} = A_{11}^{-1}A_{12} = \begin{pmatrix} 0 & 0 & 1 \\ \frac{2}{7} & -\frac{3}{7} & -\frac{2}{7}(9) \\ \frac{1}{7} & \frac{2}{7} & -\frac{1}{7} \end{pmatrix} \begin{pmatrix} 4 & 5 \\ -1 & 2(10) \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & \frac{6}{7}(11) \\ 0 & \frac{10}{7} \end{pmatrix}$$

$$b_1^{'} = A_{11}^{-1}b_1 = \begin{pmatrix} 0 & 0 & 1 \\ \frac{2}{7} & -\frac{3}{7} & -\frac{2}{7}(12) \\ \frac{1}{7} & \frac{2}{7} & -\frac{1}{7} \end{pmatrix} \begin{pmatrix} 33 \\ 27(13) \\ -10 \end{pmatrix} = \begin{pmatrix} -10 \\ \frac{5}{7}(14) \\ \frac{97}{7} \end{pmatrix}$$

Después se restan a los elementos matriciales del segundo renglón los correspondientes elementos del primer renglón premultiplicados por A_{21}, el elemento que se trata de eliminar; obteniéndose:

$$\text{Matriz llevada a la forma (3.23)} \qquad \left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & 2 & -1 & -10 \\ 0 & 1 & 0 & 1 & \frac{6}{7} & \frac{5}{7} \\ 0 & 0 & 1 & 0 & \frac{10}{7} & \frac{97}{7} \\ \hline 0 & 0 & 0 & -2 & \frac{18}{7} & \frac{190}{7} \\ 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Operaciones:

$$0 = A_{21} - A_{21}I = \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} - \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0(15) \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

$$A_{22}' = A_{22} - A_{21}A_{12}' = \begin{pmatrix} 1 & 1 \\ 0 & 4 \end{pmatrix} - \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 2 & -1 \\ 1 & \frac{6}{7}(16) \\ 0 & \frac{10}{7} \end{pmatrix} = \begin{pmatrix} -2 & \frac{18}{7} \\ 0 & 4 \end{pmatrix}$$

$$b_2' = b_2 - A_{21}b_1' = \begin{pmatrix} 4 \\ 36 \end{pmatrix} - \begin{pmatrix} 1 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} -10 \\ \frac{5}{7}(17) \\ \frac{97}{7} \end{pmatrix} = \begin{pmatrix} \frac{190}{7} \\ 36 \end{pmatrix}$$

Luego se toma como pivote a la submatriz A_{22}' y se obtiene su inversa:

$$\mathbf{A}_{22}' = \begin{pmatrix} -2 & \dfrac{18}{7} \\ 0 & 4 \end{pmatrix} \qquad \text{donde} \qquad \det A = -8$$

Por el teorema 3.3.2:

$$(\mathbf{A}_{22}')^{-1} = -\frac{1}{8}\begin{pmatrix} 4 & -\dfrac{18}{7} \\ 0 & -2 \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{2} & \dfrac{9}{28} \\ 0 & \dfrac{1}{4} \end{pmatrix}$$

Como $(A_{22}')^{-1}$ es no singular, se premultiplican todos los elementos del segundo renglón por $(A_{22}')^{-1}$, la inversa del nuevo pivote; obteniéndose:

$$\text{Matriz llevada a la forma (3.24)} \qquad \left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & 2 & -1 & -10 \\ 0 & 1 & 0 & 1 & \frac{6}{7} & \frac{5}{7} \\ 0 & 0 & 1 & 0 & \frac{10}{7} & \frac{97}{7} \\ \hline 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 9 \end{array}\right)$$

Operaciones:

$$I = (A_{22}^{'})^{-1} A_{22}^{'} = \begin{pmatrix} -\frac{1}{2} & \frac{9}{28} \\ 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} -2 & \frac{18}{7} \\ 0 & 4 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$c_2 = (A_{22}^{'})^{-1} b_2^{'} = \begin{pmatrix} -\frac{1}{2} & \frac{9}{28} \\ 0 & \frac{1}{4} \end{pmatrix} \begin{pmatrix} \frac{190}{7} \\ 36 \end{pmatrix} = \begin{pmatrix} -2 \\ 9 \end{pmatrix} \qquad \Rightarrow \begin{matrix} x_3 & = & -2 \\ x_4 & = & 9 \end{matrix}$$

Finalmente se restan a los elementos matriciales del primer renglón los correspondientes elementos del segundo, premultiplicados previamente por $A_{12}^{'}$, el elemento matricial que ahora se trata de eliminar; obteniéndose:

Matriz llevada a la forma (3.25) $\left(\begin{array}{ccc|cc|c} 1 & 0 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 0 & -5 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ \hline 0 & 0 & 0 & 1 & 0 & -2 \\ 0 & 0 & 0 & 0 & 1 & 9 \end{array}\right)$

Operaciones:

$$0 = A_{12}^{'} - A_{12}^{'} I = \begin{pmatrix} 2 & -1 \\ 1 & \frac{6}{7}(18) \\ 0 & \frac{10}{7} \end{pmatrix} - \begin{pmatrix} 2 & -1 \\ 1 & \frac{6}{7}(19) \\ 0 & \frac{10}{7} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0(20) \\ 0 & 0 \end{pmatrix}$$

$$c_1 = b_1^{'} - A_{12}^{'} c_2 = \begin{pmatrix} -10 \\ \frac{5}{7}(21) \\ \frac{97}{7} \end{pmatrix} - \begin{pmatrix} 2 & -1 \\ 1 & \frac{6}{7}(22) \\ 0 & \frac{10}{7} \end{pmatrix} \begin{pmatrix} -2 \\ 9 \end{pmatrix} = \begin{pmatrix} 3 \\ -5(23) \\ 1 \end{pmatrix} \qquad \Rightarrow \begin{matrix} x_1 & = & 3 \\ x_2 & = & -5(24) \\ x_3 & = & 1 \end{matrix}$$

Obteniendo así la solución del sistema:

$$\begin{matrix} x_1 & = & 3 \\ x_2 & = & -5 \\ x_3 & = & 1 \\ x_4 & = & -2 \\ x_5 & = & 9 \end{matrix}$$

Comprobación:

$$\begin{matrix} (3) & + & 2(-5) & + & 3(1) & + & 4(-2) & + & 5(9) & = & 33 \\ & - & (-5) & + & 2(1) & - & (-2) & + & 2(9) & = & 27 \\ (3) & & & & & + & 2(-2) & - & (9) & = & -10 \\ (3) & + & (-5) & - & (1) & + & (-2) & + & (9) & = & 4 \\ & & & & & & & & 4(9) & = & 36 \end{matrix}$$

Gauss-Jordan particionado realizando 9 particiones

Para 9 particiones se utilizará la siguiente matriz ampliada:

$$\left(\begin{array}{ccc|c} A_{11} & A_{12} & A_{13} & b_1 \\ A_{21} & A_{22} & A_{23} & b_2 \\ A_{31} & A_{32} & A_{33} & b_3 \end{array}\right) \tag{3.27}$$

Una vez fijada la posición de la submatriz A_{33} en A, quedan obligadas las otras submatrices que aparecen en (1); es decir, si A_{33} es de orden $m \times m$, el vector b_3 estará formado por últimos m elementos de b.
La transformación de la matriz A en la matriz identidad se logra realizando los siguientes pasos:

1. Premultiplicación de todos los elementos del primer renglón de (?) por A_{11}^{-1}, la inversa del pivote; obteniéndose:

$$\left(\begin{array}{ccc|c} I & A'_{12} & A'_{13} & b'_1 \\ A_{21} & A_{22} & A_{23} & b_2 \\ A_{31} & A_{32} & A_{33} & b_3 \end{array}\right) \tag{3.28}$$

en donde:

$$I = A_{11}^{-1} A_{11}$$
$$A'_{12} = A_{11}^{-1} A_{12}$$
$$A'_{13} = A_{11}^{-1} A_{13}$$
$$b'_1 = A_{11}^{-1} b_1$$
$$| A_{11} | \neq 0$$

2. Restar a los elementos matriciales del segundo renglón los correspondientes elementos del primer renglón premultiplicados por A_{21}, el elemento que se trata de eliminar; obteniéndose:

$$\left(\begin{array}{ccc|c} I & A'_{12} & A'_{13} & b'_1 \\ 0 & A'_{22} & A'_{23} & b'_2 \\ A_{31} & A_{32} & A_{33} & b_3 \end{array}\right) \tag{3.29}$$

en donde:

$$0 = A_{21} - A_{21} I$$
$$A'_{22} = A_{22} - A_{21} A'_{12}$$
$$A'_{23} = A_{23} - A_{21} A'_{13}$$
$$b'_2 = b_2 - A_{21} b'_1$$

<table>
<tr><th>Orden de
de la matriz</th><th>Ejemplos de como realizar
9 particiones</th></tr>
<tr><td>1</td><td>no es necesario particionar</td></tr>
<tr><td>2</td><td>no es necesario particionar</td></tr>
<tr><td>3</td><td>no es necesario particionar</td></tr>
<tr><td>4</td><td>

$$\left(\begin{array}{c|c|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right) \left(\begin{array}{cc|c|c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right) \left(\begin{array}{c|cc|c} a_{11} & a_{12} & a_{13} & a_{14} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right)$$

</td></tr>
<tr><td>5</td><td>

$$\left(\begin{array}{c|c|ccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right) \left(\begin{array}{c|cc|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right) \left(\begin{array}{c|ccc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right)$$

$$\left(\begin{array}{cc|c|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right) \left(\begin{array}{cc|cc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right) \left(\begin{array}{ccc|c|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{array}\right)$$

</td></tr>
<tr><td>6</td><td>

$$\left(\begin{array}{c|c|cccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right) \left(\begin{array}{c|cc|ccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right) \left(\begin{array}{c|ccc|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$$

$$\left(\begin{array}{c|cccc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ \hline a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ \hline a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right) \left(\begin{array}{cc|c|ccc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right) \left(\begin{array}{cc|cc|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$$

$$\left(\begin{array}{cc|ccc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ \hline a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right) \left(\begin{array}{ccc|c|cc} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right) \left(\begin{array}{ccc|cc|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ \hline a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$$

$$\left(\begin{array}{cccc|c|c} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} & a_{36} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} & a_{46} \\ \hline a_{51} & a_{52} & a_{53} & a_{54} & a_{55} & a_{56} \\ \hline a_{61} & a_{62} & a_{63} & a_{64} & a_{65} & a_{66} \end{array}\right)$$

</td></tr>
<tr><td>$\vdots$</td><td>$\vdots$</td></tr>
</table>

Cuadro 3.2: Ejemplos de como realizar 9 particiones.

Para facilitar los cálculos se sugiere que si el orden de la matriz A es múltiplo de 3 (6×6, 9×9, 12×12, etc.), se particione en submatrices del mismo tamaño; pero si el orden de la matriz A no es múltiplo de 3 (5×5, 7×7, 8×8, etc.), se tome la partición en donde las submatrices que forman la diagonal principal sean de orden similar; por ejemplo si el orden de la matriz es 5, se deben tomar la seguna, cuarta o quinta partición ejemplificadas en este cuadro, ya que es más fácil calcular la inversa de una matriz de 2×2 y de 1×1; que la de una matriz de 3×3, primera, tercera y sexta partición ejemplificadas en este cuadro.

3. Restar a los elementos matriciales del tercer renglón los correspondientes elementos del primer renglón premultiplicados por A_{31}, el elemento que se trata de eliminar; obteniéndose:

$$\left(\begin{array}{ccc|c} I & A'_{12} & A'_{13} & b'_1 \\ 0 & A'_{22} & A'_{23} & b'_2 \\ 0 & A'_{32} & A'_{33} & b'_3 \end{array}\right) \tag{3.30}$$

en donde:

$$0 = A_{31} - A_{31}I$$
$$A'_{32} = A_{32} - A_{31}A'_{12}$$
$$A'_{33} = A_{33} - A_{31}A'_{13}$$
$$b'_3 = b_3 - A_{31}b'_1$$

4. Premultiplicación de los elementos del segundo renglón por $(A'_{22})^{-1}$, la inversa del nuevo pivote; obteniéndose:

$$\left(\begin{array}{ccc|c} I & A'_{12} & A'_{13} & b'_1 \\ 0 & I & A''_{23} & b''_2 \\ 0 & A'_{32} & A'_{33} & b'_3 \end{array}\right) \tag{3.31}$$

en donde:

$$I = (A'_{22})^{-1}A'_{22}$$
$$A''_{23} = (A'_{22})^{-1}A'_{23}$$
$$b''_2 = (A'_{22})^{-1}b'_2$$
$$|A'_{22}| \neq 0$$

5. Restar de los elementos del primer renglón los correspondientes elementos del segundo, premultiplicados previamente por A'_{12}, el elemento matricial que ahora se trata de eliminar; obteniéndose:

$$\left(\begin{array}{ccc|c} I & 0 & A''_{13} & b''_1 \\ 0 & I & A''_{23} & b''_2 \\ 0 & A'_{32} & A'_{33} & b'_3 \end{array}\right) \tag{3.32}$$

en donde:

$$0 = A'_{12} - A'_{12}I$$
$$A''_{13} = A'_{13} - A'_{12}A''_{23}$$
$$b''_1 = b'_1 - A'_{12}b''_2$$

6. Restar de los elementos del tercer renglón los correspondientes elementos del segundo, premultiplicados previamente por A'_{32}, el elemento matricial que ahora se trata de eliminar; obteniéndose:

$$\left(\begin{array}{ccc|c} I & 0 & A''_{13} & b''_1 \\ 0 & I & A''_{23} & b''_2 \\ 0 & 0 & A''_{33} & b''_3 \end{array}\right) \tag{3.33}$$

en donde:

$$0 = A'_{32} - A'_{32}I$$
$$A''_{33} = A'_{33} - A'_{32}A''_{23}$$
$$b''_3 = b'_3 - A'_{32}b''_2$$

7. Premultiplicación de los elementos del tercer renglón por $(A''_{33})^{-1}$, la inversa del nuevo pivote; obteniéndose:

$$\left(\begin{array}{ccc|c} I & 0 & A''_{13} & b''_1 \\ 0 & I & A''_{23} & b''_2 \\ 0 & 0 & I & c_3 \end{array}\right) \tag{3.34}$$

en donde:

$$I = (A''_{33})^{-1}A''_{33}$$
$$c_3 = (A''_{33})^{-1}b''_3$$
$$|A''_{33}| \neq 0$$

8. Restar de los elementos del primer renglón los correspondientes elementos del tercero, premultiplicados previamente por A''_{13}, el elemento matricial que ahora se trata de eliminar; obteniéndose:

$$\left(\begin{array}{ccc|c} I & 0 & 0 & c_1 \\ 0 & I & A''_{23} & b''_2 \\ 0 & 0 & I & c_3 \end{array}\right) \tag{3.35}$$

en donde:

$$0 = A''_{13} - A''_{13}I$$
$$c_1 = b''_1 - A''_{13}c_3$$

9. Finalmente, restar de los elementos del segundo renglón los correspondientes elementos del tercero, premultiplicados previamente por A''_{23}, el elemento matricial que ahora se trata de eliminar; obteniéndose:

$$\left(\begin{array}{ccc|c} I & 0 & 0 & c_1 \\ 0 & I & 0 & c_2 \\ 0 & 0 & I & c_3 \end{array}\right) \tag{3.36}$$

en donde:

$$0 = A''_{23} - A''_{23}I$$

$$c_2 = b_2'' - A_{23}''c_3$$

La solución del sistema es:

$$x = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

Ejemplo 3.3.14 Resolver el sistema de ecuaciones lineales (.3.26) por el método de Gauss-Jordan particionado, realizando 9 particiones.
Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones (3.26) en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 0 & -1 & 2 & -1 & 2 \\ 1 & 0 & 0 & 2 & -1 \\ 1 & 1 & -1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 4 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 33 \\ 27 \\ -10 \\ 4 \\ 36 \end{pmatrix}$$

Combinando estas matrices para obtener su forma aumentada y dividiendo la matriz A en 9 submatrices (ver cuadro 3.2); se obtiene:

$$[\mathbf{A}|\mathbf{b}] = \left(\begin{array}{cc|cc|c|c} 1 & 2 & 3 & 4 & 5 & 33 \\ 0 & -1 & 2 & -1 & 2 & 27 \\ \hline 1 & 0 & 0 & 2 & -1 & -10 \\ 1 & 1 & -1 & 1 & 1 & 4 \\ \hline 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Se empieza tomando como pivote a la submatriz A_{11} y se obtiene su inversa:

$$\mathbf{A}_{11} = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} \qquad \text{donde} \qquad \det A = -1$$

Por el teorema 3.3.2:

$$\mathbf{A}_{11}^{-1} = -1\begin{pmatrix} -1 & -2 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}$$

Como A_{11} es no singular, se premultiplican todos los elementos del primer renglón por A_{11}^{-1}, la inversa del pivote; obteniéndose:

$$\text{Matriz llevada a la forma (3.28)} \qquad \left(\begin{array}{cc|cc|c|c} 1 & 0 & 7 & 2 & 9 & 87 \\ 0 & 1 & -2 & 1 & -2 & -27 \\ \hline 1 & 0 & 0 & 2 & -1 & -10 \\ 1 & 1 & -1 & 1 & 1 & 4 \\ \hline 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Operaciones:

$$I = A_{11}^{-1}A_{11} = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A_{12}^{'} = A_{11}^{-1}A_{12}\begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 3 & 4 \\ 2 & -1 \end{pmatrix} = \begin{pmatrix} 7 & 2 \\ -2 & 1 \end{pmatrix}$$

$$A_{13}^{'} = A_{11}^{-1}A_{13} = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 5 \\ 2 \end{pmatrix} = \begin{pmatrix} 9 \\ -2 \end{pmatrix}$$

$$b_1^{'} = A_{11}^{-1}b_1 = \begin{pmatrix} 1 & 2 \\ 0 & -1 \end{pmatrix}\begin{pmatrix} 33 \\ 27 \end{pmatrix} = \begin{pmatrix} 87 \\ -27 \end{pmatrix}$$

Después se restan a los elementos matriciales del segundo renglón los correspondientes elementos del primer renglón premultiplicados por A_{21}, el elemento que se trata de eliminar; obteniéndose:

$$\text{Matriz llevada a la forma (3.29)} \qquad \left(\begin{array}{cc|cc|c|c} 1 & 0 & 7 & 2 & 9 & 87 \\ 0 & 1 & -2 & 1 & -2 & -27 \\ \hline 0 & 0 & -7 & 0 & -10 & -97 \\ 0 & 0 & -6 & -2 & -6 & -56 \\ \hline 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Operaciones:

$$0 = A_{21} - A_{21}I = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$A_{22}^{'} = A_{22} - A_{21}A_{12}^{'} = \begin{pmatrix} 0 & 2 \\ -1 & 1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 7 & 2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} -7 & 0 \\ -6 & -2 \end{pmatrix}$$

$$A_{23}^{'} = A_{23} - A_{21}A_{13}^{'} = \begin{pmatrix} 1 \\ -1 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 9 \\ -2 \end{pmatrix} = \begin{pmatrix} -10 \\ -6 \end{pmatrix}$$

$$b_2^{'} = b_2 - A_{21}b_1^{'} = \begin{pmatrix} -10 \\ 4 \end{pmatrix} - \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 87 \\ -27 \end{pmatrix} = \begin{pmatrix} -97 \\ -56 \end{pmatrix}$$

Luego se restar a los elementos matriciales del tercer renglón los correspondientes elementos del primer renglón premultiplicados por A_{31}, el elemento que se trata de eliminar; obteniéndose:

$$\text{Matriz llevada a la forma (3.30)} \qquad \left(\begin{array}{cc|cc|c|c} 1 & 0 & 7 & 2 & 9 & 87 \\ 0 & 1 & -2 & 1 & -2 & -27 \\ \hline 0 & 0 & -7 & 0 & -10 & -97 \\ 0 & 0 & -6 & -2 & -6 & -56 \\ \hline 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Operaciones:

$$0 = A_{31} - A_{31}I = \begin{pmatrix} 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

$$A_{32}' = A_{32} - A_{31}A_{12}' = \begin{pmatrix} 0 & 0 \end{pmatrix} - \begin{pmatrix} 0 & 0 \end{pmatrix}\begin{pmatrix} 7 & 2 \\ -2 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix}$$

$$A_{33}' = A_{33} - A_{31}A_{13}' = \begin{pmatrix} 4 \end{pmatrix} - \begin{pmatrix} 0 & 0 \end{pmatrix}\begin{pmatrix} 9 \\ -2 \end{pmatrix} = \begin{pmatrix} 4 \end{pmatrix}$$

$$b_3' = b_3 - A_{31}b_1' = \begin{pmatrix} 36 \end{pmatrix} - \begin{pmatrix} 0 & 0 \end{pmatrix}\begin{pmatrix} 87 \\ -27 \end{pmatrix} = \begin{pmatrix} 36 \end{pmatrix}$$

Nota: El paso anterior se puede omitir, ya que para este ejemplo el elemento $A_{31} = 0$, por lo que ya no es necesario eliminarlo.
Luego se toma como pivote a la submatriz A_{22}' y se obtiene su inversa:

$$\mathbf{A}_{22}' = \begin{pmatrix} -7 & 0 \\ -6 & -2 \end{pmatrix} \qquad \text{donde} \qquad \det A = 14$$

Por el teorema 3.3.2:

$$(\mathbf{A}_{22}')^{-1} = \frac{1}{14}\begin{pmatrix} -2 & 0 \\ 6 & -7 \end{pmatrix} = \begin{pmatrix} -\frac{1}{7} & 0 \\ \frac{3}{7} & -\frac{1}{2} \end{pmatrix}$$

Como $(A_{22}')^{-1}$ es no singular, se premultiplican todos los elementos del segundo renglón por $(A_{22}')^{-1}$, la inversa del nuevo pivote; obteniéndose:

$$\text{Matriz llevada a la forma (3.31)} \qquad \left(\begin{array}{cc|cc|c|c} 1 & 0 & 7 & 2 & 9 & 87 \\ 0 & 1 & -2 & 1 & -2 & -27 \\ \hline 0 & 0 & 1 & 0 & \frac{10}{7} & \frac{97}{7} \\ 0 & 0 & 0 & 1 & -\frac{9}{7} & -\frac{95}{7} \\ \hline 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Operaciones:

$$I = (A_{22}')^{-1}A_{22}' = \begin{pmatrix} -\frac{1}{7} & 0 \\ \frac{3}{7} & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} -7 & 0 \\ -6 & -2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$A_{23}'' = (A_{22}')^{-1}A_{23}' = \begin{pmatrix} -\frac{1}{7} & 0 \\ \frac{3}{7} & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} -10 \\ -6 \end{pmatrix} = \begin{pmatrix} \frac{10}{7} \\ -\frac{9}{7} \end{pmatrix}$$

$$b_2'' = (A_{22}')^{-1}b_2' = \begin{pmatrix} -\frac{1}{7} & 0 \\ \frac{3}{7} & -\frac{1}{2} \end{pmatrix}\begin{pmatrix} -97 \\ -56 \end{pmatrix} = \begin{pmatrix} \frac{97}{7} \\ -\frac{95}{7} \end{pmatrix}$$

Después se restan de los elementos del primer renglón los correspondientes elementos del segundo, premultiplicados previamente por A_{12}', el elemento matricial que ahora se trata de eliminar; obteniéndose:

Matriz llevada a la forma (3.32) $$\left(\begin{array}{cc|cc|c|c} 1 & 0 & 0 & 0 & \frac{11}{7} & \frac{120}{7} \\ 0 & 1 & 0 & 0 & \frac{15}{7} & \frac{100}{7} \\ \hline 0 & 0 & 1 & 0 & \frac{10}{7} & \frac{97}{7} \\ 0 & 0 & 0 & 1 & -\frac{9}{7} & -\frac{95}{7} \\ \hline 0 & 0 & 0 & 0 & 4 & 36 \end{array}\right)$$

Operaciones:

$$0 = A_{12}^{'} - A_{12}^{'} I = \begin{pmatrix} 7 & 2 \\ -2 & 1 \end{pmatrix} - \begin{pmatrix} 7 & 2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

$$A_{13}^{''} = A_{13}^{'} - A_{12}^{'} A_{23}^{''} = \begin{pmatrix} 9 \\ -2 \end{pmatrix} - \begin{pmatrix} 7 & 2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} \frac{10}{7} \\ -\frac{9}{7} \end{pmatrix} = \begin{pmatrix} \frac{11}{7} \\ \frac{15}{7} \end{pmatrix}$$

$$b_1^{''} = b_1^{'} - A_{12}^{'} b_2^{''} = \begin{pmatrix} 87 \\ -27 \end{pmatrix} - \begin{pmatrix} 7 & 2 \\ -2 & 1 \end{pmatrix}\begin{pmatrix} \frac{97}{7} \\ -\frac{95}{7} \end{pmatrix} = \begin{pmatrix} \frac{120}{7} \\ \frac{100}{7} \end{pmatrix}$$

Como el elemento $A_{32}^{'}$ es cero, ya no se elimina;

$$\Rightarrow A_{33}^{''} = (4)$$
$$\Rightarrow b_3^{''} = (36)$$

Luego se toma como pivote a la submatriz $(A_{33}^{''})^{-1}$ y se obtiene su inversa:
Por el teorema 3.3.1:

$$(\mathbf{A}_{33}^{''})^{-1} = \frac{1}{4}$$

Como $(A_{33}^{''})^{-1}$ es no singular, se premultiplican todos los elementos del tercer renglón por $(A_{33}^{''})^{-1}$, la inversa del nuevo pivote; obteniéndose:

Matriz llevada a la forma (3.34) $$\left(\begin{array}{cc|cc|c|c} 1 & 0 & 0 & 0 & \frac{11}{7} & \frac{120}{7} \\ 0 & 1 & 0 & 0 & \frac{15}{7} & \frac{100}{7} \\ \hline 0 & 0 & 1 & 0 & \frac{10}{7} & \frac{97}{7} \\ 0 & 0 & 0 & 1 & -\frac{9}{7} & -\frac{95}{7} \\ \hline 0 & 0 & 0 & 0 & 1 & 9 \end{array}\right)$$

Operaciones:

$$I = (A_{33}^{''})^{-1} A_{33}^{''} = \left(\tfrac{1}{4}\right)(4) = (1)$$

$$c_3 = (A_{33}^{''})^{-1} b_3^{''} = \left(\tfrac{1}{4}\right)(36) = (9) \qquad \Rightarrow x_5 \;=\; 9$$

Después se restan de los elementos del primer renglón los correspondientes elementos del tercero, premultiplicados previamente por $A_{13}^{''}$, el elemento matricial que ahora se trata de eliminar; obteniéndose:

$$\text{Matriz llevada a la forma (3.35)} \qquad \left(\begin{array}{cc|cc|c|c} 1 & 0 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 0 & -5 \\ \hline 0 & 0 & 1 & 0 & \frac{10}{7} & \frac{97}{7} \\ 0 & 0 & 0 & 1 & -\frac{9}{7} & -\frac{95}{7} \\ \hline 0 & 0 & 0 & 0 & 1 & 9 \end{array}\right)$$

Operaciones:

$$0 = A_{13}'' - A_{13}''I = \begin{pmatrix} \frac{11}{7} \\ \frac{15}{7} \end{pmatrix} - \begin{pmatrix} \frac{11}{7} \\ \frac{15}{7} \end{pmatrix}(1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$c_1 = b_1'' - A_{13}''c_3 = \begin{pmatrix} \frac{120}{7} \\ \frac{100}{7} \end{pmatrix} - \begin{pmatrix} \frac{11}{7} \\ \frac{15}{7} \end{pmatrix}(9) = \begin{pmatrix} 3 \\ -5 \end{pmatrix} \qquad \Rightarrow \begin{array}{ccc} x_1 & = & 3 \\ x_2 & = & -5 \end{array}$$

Finalmente se restan a los elementos matriciales del segundo renglón los correspondientes elementos del tercero, premultiplicados previamente por A_{23}'', el elemento matricial que ahora se trata de eliminar; obteniéndose:

$$\text{Matriz llevada a la forma (3.36)} \qquad \left(\begin{array}{cc|cc|c|c} 1 & 0 & 0 & 0 & 0 & 3 \\ 0 & 1 & 0 & 0 & 0 & -5 \\ \hline 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & 2 \\ \hline 0 & 0 & 0 & 0 & 1 & 9 \end{array}\right)$$

Operaciones:

$$0 = A_{23}'' - A_{23}''I = \begin{pmatrix} \frac{10}{7} \\ -\frac{9}{7} \end{pmatrix} - \begin{pmatrix} \frac{10}{7} \\ -\frac{9}{7} \end{pmatrix}(1) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$c_2 = b_2'' - A_{23}''c_3 = \begin{pmatrix} \frac{97}{7} \\ -\frac{95}{7} \end{pmatrix} - \begin{pmatrix} \frac{10}{7} \\ -\frac{9}{7} \end{pmatrix}(9) = \begin{pmatrix} -1 \\ 2 \end{pmatrix} \qquad \Rightarrow \begin{array}{ccc} x_3 & = & -1 \\ x_4 & = & 2 \end{array}$$

Obteniendo así la solución del sistema:

$$\begin{array}{ccc} x_1 & = & 3 \\ x_2 & = & -5 \\ x_3 & = & 1 \\ x_4 & = & -2 \\ x_5 & = & 9 \end{array}$$

Comprobación:

$$\begin{array}{ccccccccccc} (3) & + & 2(-5) & + & 3(1) & + & 4(-2) & + & 5(9) & = & 33 \\ & - & (-5) & + & 2(1) & - & (-2) & + & 2(9) & = & 27 \\ (3) & & & & & + & 2(-2) & - & (9) & = & -10 \\ (3) & + & (-5) & - & (1) & + & (-2) & + & (9) & = & 4 \\ & & & & & & & & 4(9) & = & 36 \end{array}$$

3.3.6 Método de intercambio

El *método de intercambio* es un método exacto para invertir matrices parecido al de Gauss-Jordan. Con este método se invierte la matriz en el mismo sitio en donde se tiene la matriz original; por lo que ya no es necesario separar espacio adicional para almacenar la inversa.

Básicamente el método consiste en despejar una incógnita de una ecuación y sustituirla en todas las otras, tomar otra incógnita de cualquier otra ecuación y sustituirla en todas las demás, etc.; al repetir n veces el proceso se tendrán despejadas todas las incógnitas y resuelto el sistema.

Lo importante de este método es la forma en que se puede sistematizar, utilizándose reglas fijas semejantes a las del proceso de Gauss-Jordan. Para obtener éstas se va a particularizar, considerando un sistema de ecuaciones con tres incógnitas $(n=3)$, por lo que el sistema tiene la forma:

$$\begin{array}{ccccccc} a_{11}x_1 & + & a_{12}x_2 & + & a_{13}x_3 & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & a_{23}x_3 & = & b_2 \\ a_{31}x_1 & + & a_{32}x_2 & + & a_{33}x_3 & = & b_3 \end{array} \tag{3.37}$$

que se puede representar por medio de la tabla

$$\begin{array}{c|ccc} & x_1 & x_2 & x_3 \\ \hline b_1 & a_{11} & a_{12} & a_{13} \\ b_2 & a_{21} & a_{22} & a_{23} \\ b_3 & a_{31} & a_{32} & a_{33} \end{array}$$

Aplicando el método se puede despejar, por ejemplo x_3, de la segunda ecuación de (3.37):

$$a_{23}x_3 = b_2 - a_{21}x_1 - a_{22}x_2$$

$$\Rightarrow x_3 = \frac{b_2}{a_{23}} - \frac{a_{21}}{a_{23}}x_1 - \frac{a_{22}}{a_{23}}x_2 \qquad \text{donde} \qquad a_{23} \neq 0 \tag{3.38}$$

Sustituyendo x_3 en la primera ecuación de (3.37), se obtiene:

$$b_1 = a_{11}x_1 + a_{12}x_2 + a_{13}\left[\frac{b_2}{a_{23}} - \frac{a_{21}}{a_{23}}x_1 - \frac{a_{22}}{a_{23}}x_2\right]$$

Se factoriza b_1 en términos de x_1 y x_2:

$$b_1 = a_{11}x_1 + a_{12}x_2 + \frac{a_{13}}{a_{23}}b_2 - \frac{a_{21}}{a_{23}}a_{13}x_1 - \frac{a_{22}}{a_{23}}a_{13}x_2$$

$$\Rightarrow b_1 = \left(a_{11} - \frac{a_{21}}{a_{23}}a_{13}\right)x_1 + \left(a_{12} - \frac{a_{22}}{a_{23}}a_{13}\right)x_2 + \frac{a_{13}}{a_{23}}b_2 \tag{3.39}$$

Sustituyendo x_3 en la tercera ecuación de (1), se obtiene:

$$b_3 = a_{31}x_1 + a_{32}x_2 + a_{33}\left[\frac{b_2}{a_{23}} - \frac{a_{21}}{a_{23}}x_1 - \frac{a_{22}}{a_{23}}x_2\right]$$

Se factoriza b_3 en términos de x_1 y x_2:

$$b_3 = a_{31}x_1 + a_{32}x_2 + \frac{a_{33}}{a_{23}}b_2 - \frac{a_{21}}{a_{23}}a_{33}x_1 - \frac{a_{22}}{a_{23}}a_{33}x_2$$

$$\Rightarrow b_3 = \left(a_{31} - \frac{a_{21}}{a_{23}}a_{33}\right)x_1 + \left(a_{32} - \frac{a_{22}}{a_{23}}a_{33}\right)x_2 + \frac{a_{33}}{a_{23}}b_2 \tag{3.40}$$

Obteniéndose un nuevo sistema con (3.38), (3.39) y (3.40); el cual puede escribirse en forma de tabla, como se muestra a continuación:

	x_1	x_2	b_2
b_1	$a_{11}-\frac{a_{21}}{a_{23}}a_{13}$	$a_{12}-\frac{a_{22}}{a_{23}}a_{13}$	$\frac{a_{13}}{a_{23}}$
x_3	$-\frac{a_{21}}{a_{23}}$	$-\frac{a_{22}}{a_{23}}$	$\frac{1}{a_{23}}$
b_3	$a_{31}-\frac{a_{21}}{a_{23}}a_{33}$	$a_{32}-\frac{a_{22}}{a_{23}}a_{33}$	$\frac{a_{33}}{a_{23}}$

Comparando las tablas presentadas se observa que al despejar x_3 de la segunda ecuación y sustituirla en las otras se intercambiaron las variables x_3 y b_2, que aparecen fuera del cuadro, situación que da nombre al método. Con respecto a los valores en el interior de la tabla, y llamando pivote al elemento a_{23} [7], se ve que para pasar de una tabla a otra se debe:

1. Dividir todos los elementos del renglón del pivote entre el propio pivote y cambiarles de signo, (menos el pivote).

2. Sumar a los elementos de todos los renglones diferentes al del pivote, con excepción de los que estén en la columna del pivote, los correspondientes elementos del renglón del pivote ya corregido en (1), multiplicados por el elemento correspondiente en la columna pivote.

3. Dividir todos los elementos de la columna del pivote entre el propio pivote,(excepto el elemento pivote).

4. Reemplazar el elemento pivote por su recíproco.

Al repetir tres veces los pasos anteriores sobre la matriz considerada, de orden tres, buscando intercambiar términos independientes por incógnitas [8], se llegará a un cuadro como el siguiente:

	b_1	b_2	b_3
x_1	c_{11}	c_{12}	c_{13}
x_2	c_{21}	c_{22}	c_{23}
x_3	c_{31}	c_{32}	c_{33}

(3.41)

de donde

[7] Se selecciona como pivote al elemento mayor en valor absoluto entre los posibles pivotes para reducir así al máximo los errores de redondeo.

[8] Si al realizar el intercambio de términos independientes por incógnitas, los encabezados no quedan de forma ordenada, será necesario intercambiar renglones y columnas a manera de que esto suceda.

$$
\begin{array}{ccccccc}
c_{11}b_1 & + & c_{12}b_2 & + & c_{13}b_3 & = & x_1 \\
c_{21}b_1 & + & c_{22}b_2 & + & c_{23}b_3 & = & x_2 \\
c_{31}b_1 & + & c_{32}b_2 & + & c_{33}b_3 & = & x_3
\end{array} \tag{3.42}
$$

que se puede representar matricialmente de la siguiente forma:

$$
\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \underbrace{\begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}}_{C} \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \tag{3.43}
$$

Al comparar esta última expresión con la ecuación (3.10), $C = A^{-1}$, es decir que los elementos de la tabla (3.41) forman la inversa de la matriz A.

Ejemplo 3.3.15 Resolver el siguiente sistema de ecuaciones lineales por el método de intercambio; (det $A = -12$, obtenido con Maple).

$$
\begin{array}{ccccccccc}
- & x_1 & + & x_2 & & & - & 3x_4 & = & 4 \\
 & x_1 & & & + & 3x_3 & + & x_4 & = & 0 \\
 & & + & x_2 & - & x_3 & - & x_4 & = & 3 \\
 & 3x_1 & & & + & x_3 & + & 2x_4 & = & 1
\end{array} \tag{3.44}
$$

Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

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

Se representa el sistema (3.34) por medio de la siguiente tabla:

	x_1	x_2	x_3	x_4
b_1	-1	1	0	-3
b_2	1	0	3	1
b_3	0	1	-1	-1
b_4	3	0	1	2

Se escoge como pivote al elemento mayor en valor absoluto, en este ejemplo los elementos que cumplen con esta característica son: a_{14}, a_{23} y a_{41}; por lo que se puede escoger como elemento pivote a cualquiera de ellos. En este caso se escoge como elemento pivote a a_{23}, intercambiando x_3 y b_2. Se sugiere identificar en la tabla el renglón y la columna del pivote, realizando una división, para facilitar los cálculos al aplicar el método.
Los pasos del método de intercambio se tienen que realizar 4 veces, ya que el orden de la matriz A es 4, para lograr intercambiar todos los términos independientes por las incógnitas.

Pasos del método de intercambio (primera vez)

Siguiendo los pasos del método, se divide todos los elementos del renglón del pivote entre el propio pivote y se les cambia de signo, (menos el pivote); obteniéndose:

	x_1	x_2	b_2	x_4
b_1	-1	1	0	-3
x_3	$-(\frac{1}{3})$	$-(\frac{0}{3})$	3	$-(\frac{1}{3})$
b_3	0	1	-1	-1
b_4	3	0	1	2

Después se suman a los elementos de todos los renglones diferentes al del pivote, con excepción de los que estén en la columna del pivote, los correspondientes elementos del renglón del pivote ya corregido en el paso anterior, multiplicados por el elemento correspondiente en la columna pivote; obteniéndose:

	x_1	x_2	b_2	x_4
b_1	$-1+(-\frac{1}{3})(0)$	$1+(0)(0)$	0	$-3+(-\frac{1}{3})(0)$
x_3	$-\frac{1}{3}$	0	3	$-\frac{1}{3}$
b_3	$0+(-\frac{1}{3})(-1)$	$1+(0)(-1)$	-1	$-1+(-\frac{1}{3})(-1)$
b_4	$3+(-\frac{1}{3})(1)$	$0+(0)(1)$	1	$2+(-\frac{1}{3})(1)$

Realizando las operaciones:

	x_1	x_2	b_2	x_4
b_1	-1	1	0	-3
x_3	$-\frac{1}{3}$	0	3	$-\frac{1}{3}$
b_3	$\frac{1}{3}$	1	-1	$-\frac{2}{3}$
b_4	$\frac{8}{3}$	0	1	$\frac{5}{3}$

Luego se dividen todos los elementos de la columna del pivote entre el propio pivote,(excepto el elemento pivote); obteniéndose:

	x_1	x_2	b_2	x_4
b_1	-1	1	$\frac{0}{3}$	-3
x_3	$-\frac{1}{3}$	0	3	$-\frac{1}{3}$
b_3	$\frac{1}{3}$	1	$-\frac{1}{3}$	$-\frac{2}{3}$
b_4	$\frac{8}{3}$	0	$\frac{1}{3}$	$\frac{5}{3}$

Después se reemplaza el elemento pivote por su recíproco; obteniéndose:

	x_1	x_2	b_2	x_4
b_1	-1	1	0	-3
x_3	$-\frac{1}{3}$	0	$\frac{1}{3}$	$-\frac{1}{3}$
b_3	$\frac{1}{3}$	1	$-\frac{1}{3}$	$-\frac{2}{3}$
b_4	$\frac{8}{3}$	0	$\frac{1}{3}$	$\frac{5}{3}$

Pasos del método de intercambio (segunda vez)

Ahora se escoge como nuevo pivote a –3 (elemento mayor en valor absoluto), intercambiando x_4 y b_1. Y se repiten los pasos anteriores.

Se divide todos los elementos del renglón del pivote entre el propio pivote y se les cambia de signo, (menos el pivote); obteniéndose:

	x_1	x_2	b_2	b_1
x_4	$-(\frac{-1}{-3})$	$-(\frac{1}{-3})$	$-(\frac{0}{-3})$	-3
x_3	$-\frac{1}{3}$	0	$\frac{1}{3}$	$-\frac{1}{3}$
b_3	$\frac{1}{3}$	1	$-\frac{1}{3}$	$-\frac{2}{3}$
b_4	$\frac{8}{3}$	0	$\frac{1}{3}$	$\frac{5}{3}$

Después se suman a los elementos de todos los renglones diferentes al del pivote, con excepción de los que estén en la columna del pivote, los correspondientes elementos del renglón del pivote ya corregido en el paso anterior, multiplicados por el elemento correspondiente en la columna pivote; obteniéndose:

	x_1	x_2	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{1}{3}$	0	-3
x_3	$-\frac{1}{3}+(-\frac{1}{3})(-\frac{1}{3})$	$0+(\frac{1}{3})(-\frac{1}{3})$	$\frac{1}{3}+(0)(-\frac{1}{3})$	$-\frac{1}{3}$
b_3	$\frac{1}{3}+(-\frac{1}{3})(-\frac{2}{3})$	$1+(\frac{1}{3})(-\frac{2}{3})$	$-\frac{1}{3}+(0)(-\frac{2}{3})$	$-\frac{2}{3}$
b_4	$\frac{8}{3}+(-\frac{1}{3})(\frac{5}{3})$	$0+(\frac{1}{3})(\frac{5}{3})$	$\frac{1}{3}+(0)(\frac{5}{3})$	$\frac{5}{3}$

Realizando las operaciones:

	x_1	x_2	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{1}{3}$	0	-3
x_3	$-\frac{2}{9}$	$-\frac{1}{9}$	$\frac{1}{3}$	$-\frac{1}{3}$
b_3	$\frac{5}{9}$	$\frac{7}{9}$	$-\frac{1}{3}$	$-\frac{2}{3}$
b_4	$\frac{19}{9}$	$\frac{5}{9}$	$\frac{1}{3}$	$\frac{5}{3}$

Luego se dividen todos los elementos de la columna del pivote entre el propio pivote,(excepto el elemento pivote); obteniéndose:

	x_1	x_2	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{1}{3}$	0	-3
x_3	$-\frac{2}{9}$	$-\frac{1}{9}$	$\frac{1}{3}$	$-\frac{1}{3}/(-3)=\frac{1}{9}$
b_3	$\frac{5}{9}$	$\frac{7}{9}$	$-\frac{1}{3}$	$-\frac{2}{3}/(-3)=\frac{2}{9}$
b_4	$\frac{19}{9}$	$\frac{5}{9}$	$\frac{1}{3}$	$\frac{5}{3}/(-3)=-\frac{5}{9}$

Después se reemplaza el elemento pivote por su recíproco; obteniéndose:

	x_1	x_2	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{1}{3}$	0	$-\frac{1}{3}$
x_3	$-\frac{2}{9}$	$-\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$
b_3	$\frac{5}{9}$	$\frac{7}{9}$	$-\frac{1}{3}$	$\frac{2}{9}$
b_4	$\frac{19}{9}$	$\frac{5}{9}$	$\frac{1}{3}$	$-\frac{5}{9}$

Pasos del método de intercambio (tercera vez)

Ahora se escoge como nuevo pivote a $\frac{19}{9}$ (elemento mayor en valor absoluto), intercambiando x_1 y b_4. Y se repiten los pasos anteriores.

Se divide todos los elementos del renglón del pivote entre el propio pivote y se les cambia de signo, (menos el pivote); obteniéndose:

	b_4	x_2	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{1}{3}$	0	$-\frac{1}{3}$
x_3	$-\frac{2}{9}$	$-\frac{1}{9}$	$\frac{1}{3}$	$\frac{1}{9}$
b_3	$\frac{5}{9}$	$\frac{7}{9}$	$-\frac{1}{3}$	$\frac{2}{9}$
x_1	$\frac{19}{9}$	$-(\frac{5}{9}/\frac{19}{9}) = -\frac{5}{19}$	$-(\frac{1}{3}/\frac{19}{9}) = -\frac{3}{19}$	$-(-\frac{5}{9}/\frac{19}{9}) = \frac{5}{19}$

Después se suman a los elementos de todos los renglones diferentes al del pivote, con excepción de los que estén en la columna del pivote, los correspondientes elementos del renglón del pivote ya corregido en el paso anterior, multiplicados por el elemento correspondiente en la columna pivote; obteniéndose:

	b_4	x_2	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{1}{3} + (-\frac{5}{19})(-\frac{1}{3})$	$0 + (-\frac{3}{19})(-\frac{1}{3})$	$-\frac{1}{3} + (\frac{5}{19})(-\frac{1}{3})$
x_3	$-\frac{2}{9}$	$-\frac{1}{9} + (-\frac{5}{19})(-\frac{2}{9})$	$\frac{1}{3} + (-\frac{3}{19})(-\frac{2}{9})$	$\frac{1}{9} + (\frac{5}{19})(-\frac{2}{9})$
b_3	$\frac{5}{9}$	$\frac{7}{9} + (-\frac{5}{19})(\frac{5}{9})$	$-\frac{1}{3} + (-\frac{3}{19})(\frac{5}{9})$	$\frac{2}{9} + (\frac{5}{19})(\frac{5}{9})$
x_1	$\frac{19}{9}$	$-\frac{5}{19}$	$-\frac{3}{19}$	$\frac{5}{19}$

Realizando las operaciones:

	b_4	x_2	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{8}{19}$	$\frac{1}{19}$	$-\frac{8}{19}$
x_3	$-\frac{2}{9}$	$-\frac{1}{19}$	$\frac{7}{19}$	$\frac{1}{19}$
b_3	$\frac{5}{9}$	$\frac{12}{19}$	$-\frac{8}{19}$	$\frac{7}{19}$
x_1	$\frac{19}{9}$	$-\frac{5}{19}$	$-\frac{3}{19}$	$\frac{5}{19}$

Luego se dividen todos los elementos de la columna del pivote entre el propio pivote,(excepto el elemento pivote); obteniéndose:

	b_4	x_2	b_2	b_1
x_4	$-\frac{1}{3}/\frac{19}{9} = -\frac{3}{19}$	$\frac{8}{19}$	$\frac{1}{19}$	$-\frac{8}{19}$
x_3	$-\frac{2}{9}/\frac{19}{9} = -\frac{2}{19}$	$-\frac{1}{19}$	$\frac{7}{19}$	$\frac{1}{19}$
b_3	$\frac{5}{9}/\frac{19}{9} = \frac{5}{19}$	$\frac{12}{19}$	$-\frac{8}{19}$	$\frac{7}{19}$
x_1	$\frac{19}{9}$	$-\frac{5}{19}$	$-\frac{3}{19}$	$\frac{5}{19}$

Después se reemplaza el elemento pivote por su recíproco; obteniéndose:

	b_4	x_2	b_2	b_1
x_4	$-\frac{3}{19}$	$\frac{8}{19}$	$\frac{1}{19}$	$-\frac{8}{19}$
x_3	$-\frac{2}{19}$	$-\frac{1}{19}$	$\frac{7}{19}$	$\frac{1}{19}$
b_3	$\frac{5}{19}$	$\frac{12}{19}$	$-\frac{8}{19}$	$\frac{7}{19}$
x_1	$\frac{9}{19}$	$-\frac{5}{19}$	$-\frac{3}{19}$	$\frac{5}{19}$

Pasos del método de intercambio (cuarta vez)

Ahora se escoge como nuevo pivote a $\frac{12}{19}$ (elemento mayor en valor absoluto), intercambiando x_2 y b_3. Y se repiten los pasos anteriores.

Se dividen todos los elementos del renglón del pivote entre el propio pivote y se les cambia de signo, (menos el pivote); obteniéndose:

	b_4	b_3	b_2	b_1
x_4	$-\frac{3}{19}$	$\frac{8}{19}$	$\frac{1}{19}$	$-\frac{8}{19}$
x_3	$-\frac{2}{19}$	$-\frac{1}{19}$	$\frac{7}{19}$	$\frac{1}{19}$
x_2	$-(\frac{5}{19}/\frac{12}{19}) = -\frac{5}{12}$	$\frac{12}{19}$	$-(-\frac{8}{19}/\frac{12}{19}) = \frac{2}{3}$	$-(\frac{7}{19}/\frac{12}{19}) = -\frac{7}{12}$
x_1	$\frac{9}{19}$	$-\frac{5}{19}$	$-\frac{3}{19}$	$\frac{5}{19}$

Después se suman a los elementos de todos los renglones diferentes al del pivote, con excepción de los que estén en la columna del pivote, los correspondientes elementos del renglón del pivote ya corregido en el paso anterior, multiplicados por el elemento correspondiente en la columna pivote; obteniéndose:

	b_4	b_3	b_2	b_1
x_4	$-\frac{3}{19} + (-\frac{5}{12})(\frac{8}{19})$	$\frac{8}{19}$	$\frac{1}{19} + (\frac{2}{3})(\frac{8}{19})$	$-\frac{8}{19} + (-\frac{7}{12})(\frac{8}{19})$
x_3	$-\frac{2}{19} + (-\frac{5}{12})(-\frac{1}{19})$	$-\frac{1}{19}$	$\frac{7}{19} + (\frac{2}{3})(-\frac{1}{19})$	$\frac{1}{19} + (-\frac{7}{12})(-\frac{1}{19})$
x_2	$-\frac{5}{12}$	$\frac{12}{19}$	$\frac{2}{3}$	$-\frac{7}{12}$
x_1	$\frac{9}{19} + (-\frac{5}{12})(-\frac{5}{19})$	$-\frac{5}{19}$	$-\frac{3}{19} + (\frac{2}{3})(-\frac{5}{19})$	$\frac{5}{19} + (-\frac{7}{12})(-\frac{5}{19})$

Realizando las operaciones:

	b_4	b_3	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{8}{19}$	$\frac{1}{3}$	$-\frac{2}{3}$
x_3	$-\frac{1}{12}$	$-\frac{1}{19}$	$\frac{1}{3}$	$\frac{1}{12}$
x_2	$-\frac{5}{12}$	$\frac{12}{19}$	$\frac{2}{3}$	$-\frac{7}{12}$
x_1	$\frac{7}{12}$	$-\frac{5}{19}$	$-\frac{1}{3}$	$\frac{5}{12}$

Luego se dividen todos los elementos de la columna del pivote entre el propio pivote,(excepto el elemento pivote); obteniéndose:

	b_4	b_3	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{8}{19}/\frac{12}{19}=\frac{2}{3}$	$\frac{1}{3}$	$-\frac{2}{3}$
x_3	$-\frac{1}{12}$	$-\frac{1}{19}/\frac{12}{19}=-\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{12}$
x_2	$-\frac{5}{12}$	$\frac{12}{19}$	$\frac{2}{3}$	$-\frac{7}{12}$
x_1	$\frac{7}{12}$	$-\frac{5}{19}/\frac{12}{19}=-\frac{5}{12}$	$-\frac{1}{3}$	$\frac{5}{12}$

Después se reemplaza el elemento pivote por su recíproco; obteniéndose:

	b_4	b_3	b_2	b_1
x_4	$-\frac{1}{3}$	$\frac{2}{3}$	$\frac{1}{3}$	$-\frac{2}{3}$
x_3	$-\frac{1}{12}$	$-\frac{1}{12}$	$\frac{1}{3}$	$\frac{1}{12}$
x_2	$-\frac{5}{12}$	$\frac{19}{12}$	$\frac{2}{3}$	$-\frac{7}{12}$
x_1	$\frac{7}{12}$	$-\frac{5}{12}$	$-\frac{1}{3}$	$\frac{5}{12}$

Como al realizar el intercambio de términos independientes por incógnitas, los encabezados no quedan de forma ordenada, será necesario intercambiar renglones y columnas a manera de que esto suceda; obteniéndose:

	b_1	b_2	b_3	b_4
x_1	$\frac{5}{12}$	$-\frac{1}{3}$	$-\frac{5}{12}$	$\frac{7}{12}$
x_2	$-\frac{7}{12}$	$\frac{2}{3}$	$\frac{19}{12}$	$-\frac{5}{12}$
x_3	$\frac{1}{12}$	$\frac{1}{3}$	$-\frac{1}{12}$	$-\frac{1}{12}$
x_4	$-\frac{2}{3}$	$\frac{1}{3}$	$\frac{2}{3}$	$-\frac{1}{3}$

Los elementos de la tabla anterior forman la inversa de la matriz A:

$$\mathbf{A}^{-1}=\begin{pmatrix} \frac{5}{12} & -\frac{1}{3} & -\frac{5}{12} & \frac{7}{12} \\ -\frac{7}{12} & \frac{2}{3} & \frac{19}{12} & -\frac{5}{12} \\ \frac{1}{12} & \frac{1}{3} & -\frac{1}{12} & -\frac{1}{12} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{pmatrix}$$

Comprobación: Por la definición 3.3.2, (sección 3.3.3).

$$\mathbf{I}=\mathbf{A}\mathbf{A}^{-1}=\begin{pmatrix} -1 & 1 & 0 & -3 \\ 1 & 0 & 3 & 1 \\ 0 & 1 & -1 & -1 \\ 3 & 0 & 1 & 2 \end{pmatrix}\begin{pmatrix} \frac{5}{12} & -\frac{1}{3} & -\frac{5}{12} & \frac{7}{12} \\ -\frac{7}{12} & \frac{2}{3} & \frac{19}{12} & -\frac{5}{12} \\ \frac{1}{12} & \frac{1}{3} & -\frac{1}{12} & -\frac{1}{12} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{pmatrix}=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Finalmente, por la ecuación (3.10) se multiplica A^{-1} por b, a fin de obtener el valor de las incógnitas.

$$\mathbf{x}=\mathbf{A}^{-1}\mathbf{b}=\begin{pmatrix} \frac{5}{12} & -\frac{1}{3} & -\frac{5}{12} & \frac{7}{12} \\ -\frac{7}{12} & \frac{2}{3} & \frac{19}{12} & -\frac{5}{12} \\ \frac{1}{12} & \frac{1}{3} & -\frac{1}{12} & -\frac{1}{12} \\ -\frac{2}{3} & \frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \end{pmatrix}\begin{pmatrix}4\\0\\3\\1\end{pmatrix}=\begin{pmatrix}1\\2\\0\\-1\end{pmatrix} \qquad \begin{aligned} x_1 &= 1 \\ x_2 &= 2 \\ x_3 &= 0 \\ x_4 &= -1 \end{aligned}$$

Comprobación:

$$\begin{array}{rrrrrrrrrr} - & (1) & + & (2) & & & - & 3(-1) & = & 4 \\ & (1) & & & + & 3(0) & + & (-1) & = & 0 \\ & & + & (2) & - & (0) & - & (-1) & = & 3 \\ & 3(1) & & & + & (0) & + & 2(-1) & = & 1 \end{array}$$

3.3.7 Estrategias de pivoteo

El pivoteo tiene dos finalidades: 1) superar la dificultad que presentan cuando un elemento pivote es cero, ya que origina una división entre cero; y 2) disminuir los errores de redondeo; debido a que si la magnitud del pivote es pequeña comparada con la de los otros elementos, entonces se pueden introducir errores de redondeo.

La única condición para que un elemento pueda ser pivote es que sea distinto de 0.

Existen tres tipos de pivoteo:

1. **Pivoteo máximo por columna o pivoteo parcial:** Consiste en elegir un elemento no nulo (denominado "pivote") para cada columna, a partir de la primera, que es el coeficiente de la columna cuyo valor absoluto es el mayor; para lograr esto, se intercambia en cada caso el renglón i por el renglón k cuyo elemento a_{ki} satisfaga la condición anterior; ya que se eligió el pivote se procede de modo habitual aplicando el método deseado.
2. **Pivoteo total:** Consiste en producir dominancia diagonal en la matriz de coeficientes, para ello se busca en toda la matriz el elemento de mayor valor absoluto, realizando intercambio de renglones y columnas; es decir consiste en sustituir el pivote a_{ii}, en cada caso, por aquel elemento a_{kj}, cuyo valor absoluto sea el mayor; para lograr esto, se intercambia el renglón i por el k y la columna i por la j, y se procede de modo habitual aplicando el método deseado; el proceso se repite $(n-1)$ veces. Este tipo de pivoteo se usa en muy raras ocasiones debido que al intercambiar columnas se cambia el orden de las x y, en consecuencia, se agrega complejidad significativa al realizar el programa en la computadora.
3. **Pivoteo escalado de fila o escalamiento:** Sólo se usa para matrices con elementos muy diferentes en magnitud y cuando un pivote es mucho más "pequeño" que alguno de los coeficientes de la ecuación que él encabeza. Consiste en escalar una vez antes del proceso de eliminación; es decir se divide el renglón o renglones entre el elemento de mayor magnitud.

Para escoger el pivote a_{jj}, para $j=1,2,\ldots,n-1$, se realizan los siguientes pasos:

1. Para $i = j, j+1, \ldots, n$, se calcula

$$s_i = \max_{1 \le l \le n} |a_{il}| \qquad \text{factordeescala}$$

2. Para $i = j, j+1, \ldots, n$, se calcula

$$\frac{|a_{ij}|}{s_i}$$

3. Se encuentra el menor entero k tal que:

$$\max_{j \le i \le n} \frac{|a_{ij}|}{s_i} = \frac{|a_{kj}|}{s_k}$$

Si tal k existe y $k \ne j$, entonces se cambia el renglón j por el renglón k, y se procede de modo habitual aplicando el método deseado.

4. Se repite el paso $b)$ y $c)$ hasta que $j = n-1$.

Los factores de cambio de escala $s_1, s_2, \ldots s_n$ se calculan sólo una vez, al inicio del procedimiento y también deben intercambiarse al realizar intercambio de renglones.

Ejemplo 3.3.16 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss, sin utilizar alguna estrategia de pivoteo, con 4 cifras significativas; (det $A = 0.12457$, obtenido con Maple).

$$\begin{array}{ccccccc} 0.0002x_1 & - & 0.00031x_2 & + & 0.0017x_3 & = & 0.00609 \\ 5x_1 & - & 7x_2 & + & 6x_3 & = & 7 \\ 8x_1 & + & 6x_2 & + & 3x_3 & = & 2 \end{array} \tag{3.45}$$

Como det $A \ne 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 0.0002 & -0.00031 & 0.0017 \\ 5 & -7 & 6 \\ 8 & 6 & 3 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 0.00609 \\ 7 \\ 2 \end{pmatrix}$$

Combinando las matrices A y b para obtener su forma aumentada, se empiezan a realizar operaciones elementales para obtener una matriz triangular superior.

$$\left(\begin{array}{rrr|r} 0.0002 & -0.00031 & 0.0017 & 0.00609 \\ 5 & -7 & 6 & 7 \\ 8 & 6 & 3 & 2 \end{array}\right) \xrightarrow[R_3 \to -\frac{8}{0.0002}R_1 + R_3]{R_2 \to -\frac{5}{0.0002}R_1 + R_2}$$

$$\left(\begin{array}{rrr|r} 0.0002 & -0.00031 & 0.0017 & 0.00609 \\ 0 & 0.750 & -36.50 & -145.2 \\ 0 & 18.40 & -65.00 & -241.6 \end{array}\right) R_3 \to -\frac{18.40}{0.750}R_2 + R_3$$

$$\left(\begin{array}{rrr|r} 0.0002 & -0.00031 & 0.0017 & 0.00609 \\ 0 & 0.750 & -36.50 & -145.2 \\ 0 & 0 & 830.3 & 3320 \end{array}\right)$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas.

$$\Rightarrow \begin{array}{rcrcrcr} 0.0002x_1 & - & 0.00031x_2 & + & 0.0017x_3 & = & 0.00609 \\ & & 0.750x_2 & - & 36.50x_3 & = & -145.2 \\ & & & & 830.3x_3 & = & 3320 \end{array} \qquad \begin{array}{rcr} x_1 & \approx & 3.999 \\ x_2 & \approx & 1.067 \\ x_3 & \approx & -1.885 \end{array}$$

Ejemplo 3.3.17 Resolver el sistema de ecuaciones lineales (3.45) por el método de Gauss, utilizando pivoteo parcial, con 4 cifras significativas.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 0.0002 & -0.00031 & 0.0017 \\ 5 & -7 & 6 \\ 8 & 6 & 3 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 0.00609 \\ 7 \\ 2 \end{pmatrix}$$

Se combinan las matrices A y b para obtener su forma aumentada. Empezando a aplicar pivoteo parcial; para escoger el pivote a_{11}, de la primera columna con elementos diferentes de cero (llamada columna pivote), se selecciona el elemento de mayor valor absoluto, en este caso es el elemento $a_{13} = 8$; a este elemento se le llama pivote:

$$\left(\begin{array}{rrr|r} 0.0002 & -0.00031 & 0.0017 & 0.00609 \\ 5 & -7 & 6 & 7 \\ 8 & 6 & 3 & 2 \end{array}\right)$$

Se intercambia el renglón 1 por el renglón 3, para llevar este elemento de la columna a la posición diagonal.

$$\left(\begin{array}{rrr|r} 8 & 6 & 3 & 2 \\ 5 & -7 & 6 & 7 \\ 0.0002 & -0.00031 & 0.0017 & 0.00609 \end{array}\right)$$

Se empiezan a realizar operaciones elementales para hacer cero los elementos que se encuentran debajo del pivote.

$$\left(\begin{array}{ccc|c} 8 & 6 & 3 & 2 \\ 5 & -7 & 6 & 7 \\ 0.0002 & -0.00031 & 0.0017 & 0.00609 \end{array}\right) \xrightarrow[R_3 \to -\frac{0.0002}{8}R_1 + R_3]{R_2 \to -\frac{5}{8}R_1 + R_2}$$

$$\left(\begin{array}{ccc|c} 8 & 6 & 3 & 2 \\ 0 & -10.75 & 4.125 & 5.750 \\ 0 & -0.0004600 & 0.001625 & 0.006040 \end{array}\right)$$

De la segunda columna, se selecciona el nuevo pivote a_{22}, que es el elemento de mayor valor absoluto de la diagonal principal para abajo, en este caso es el número -10.75; por lo que no se necesita realizar intercambio de renglones, ya que este elemento se encuentra en la posición diagonal.

$$\left(\begin{array}{ccc|c} 8 & 6 & 3 & 2 \\ 0 & -10.75 & 4.125 & 5.750 \\ 0 & -0.0004600 & 0.001625 & 0.006040 \end{array}\right)$$

Se empiezan a realizar operaciones elementales para hacer cero los elementos que se encuentran debajo del pivote.

$$\left(\begin{array}{ccc|c} 8 & 6 & 3 & 2 \\ 0 & -10.75 & 4.125 & 5.750 \\ 0 & -0.0004600 & 0.001625 & 0.006040 \end{array}\right) R_3 \to -\frac{18.40}{0.750}R_2 + R_3$$

$$\left(\begin{array}{ccc|c} 8 & 6 & 3 & 2 \\ 0 & -10.75 & 4.125 & 5.750 \\ 0 & 0 & 0.001448 & 0.005794 \end{array}\right)$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas.

$$\begin{array}{rrrrrrrrl} & 8x_1 & + & 6x_2 & + & 3x_3 & = & 2 & \qquad x_1 \approx 4.001 \\ \Rightarrow & & - & 10.75x_2 & + & 4.125x_3 & = & 5.750 & \qquad x_2 \approx 1.000 \\ & & & & & 0.001448x_3 & = & 0.005794 & \qquad x_3 \approx -2.000 \end{array}$$

La solución exacta del sistema (3.45) es $x_1 = 4$, $x_2 = 1$ y $x_3 = -2$. Comparando los resultados obtenidos de los ejemplos 3.16 y 3.17, se puede observar que si no se aplica pivoteo parcial los errores son significativos:

$$\varepsilon_{r,x_1} = \left|\tfrac{4-3.999}{4}\right| \times 100 = 0.025\%$$

$$\varepsilon_{r,x_2} = \left|\tfrac{1-1.067}{1}\right| \times 100 = 6.7\%$$

$$\varepsilon_{r,x_3} = \left|\tfrac{-2-(-1.885)}{4}\right| \times 100 = 5.75\%$$

aplicando pivoteo parcial y un redondeo a 4 cifras significativas, x_2 y x_3 se obtienen de manera exacta y x_1 se obtiene con un error relativo de $(|4-4.001|/4) \times 100 = 0.025\%$.

Ejemplo 3.3.18 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss, sin utilizar alguna estrategia de pivoteo, con 4 cifras significativas; (det $A = -65.679715$, obtenido con Maple).

$$\begin{array}{rcrcrcr} 1.78x_1 & + & 3.01x_2 & - & 4.88x_3 & = & -7.70 \\ 4.63x_1 & - & 1.06x_2 & - & 2.27x_3 & = & -6.36 \\ -\ 3.39x_1 & + & 9.81x_2 & - & 4.78x_3 & = & 3.95 \end{array} \tag{3.46}$$

Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 1.78 & 3.01 & -4.88 \\ 4.63 & -1.06 & -2.27 \\ -3.39 & 9.81 & -4.78 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} -7.70 \\ -6.36 \\ 3.95 \end{pmatrix}$$

Combinando las matrices A y b para obtener su forma aumentada, se empiezan a realizar operaciones elementales para obtener una matriz triangular superior.

$$\left(\begin{array}{rrr|r} 1.78 & 3.01 & -4.88 & -7.70 \\ 4.63 & -1.06 & -2.27 & -6.36 \\ -3.39 & 9.81 & -4.78 & 3.95 \end{array}\right) \xrightarrow[R_3 \to -(-\frac{3.39}{1.78})R_1 + R_3]{R_2 \to -\frac{4.63}{1.78}R_1 + R_2}$$

$$\left(\begin{array}{rrr|r} 1.78 & 3.01 & -4.88 & -7.70 \\ 0 & -8.889 & 10.42 & 13.67 \\ 0 & 15.54 & -14.07 & -10.71 \end{array}\right) R_3 \to -\frac{15.54}{-8.889}R_2 + R_3$$

$$\left(\begin{array}{rrr|r} 1.78 & 3.01 & -4.88 & -7.70 \\ 0 & -8.889 & 10.42 & 13.67 \\ 0 & 0 & 4.14 & 13.19 \end{array}\right)$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas.

$$\Rightarrow \begin{array}{rcrcrcr} 1.78x_1 & + & 3.01x_2 & - & 4.88x_3 & = & -7.70 \\ & - & 8.889x_2 & + & 10.42x_3 & = & 13.67 \\ & & & & 4.14x_3 & = & 13.19 \end{array} \qquad \begin{array}{rcl} x_1 & \approx & 0.6966 \\ x_2 & \approx & 2.197 \\ x_3 & \approx & 3.186 \end{array}$$

Ejemplo 3.3.19 Resolver el sistema de ecuaciones lineales (3.46) por el método de Gauss utilizando pivoteo total, con 4 cifras significativas.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 1.78 & 3.01 & -4.88 \\ 4.63 & -1.06 & -2.27 \\ -3.39 & 9.81 & -4.78 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} -7.70 \\ -6.36 \\ 3.95 \end{pmatrix}$$

Se combinan las matrices A y b para obtener su forma aumentada. Empezando a aplicar pivoteo completo; para escoger el pivote a_{11} se busca en toda la matriz A el elemento de mayor valor absoluto, en este caso es el elemento $a_{3,2} = 9.81$; a este elemento se le llama pivote:

$$\left(\begin{array}{rrr|r} 1.78 & 3.01 & -4.88 & -7.70 \\ 4.63 & -1.06 & -2.27 & -6.36 \\ -3.39 & 9.81 & -4.78 & 3.95 \end{array}\right)$$

Se intercambia el renglón 1 por el renglón 3:

$$\left(\begin{array}{rrr|r} -3.39 & 9.81 & -4.78 & 3.95 \\ 4.63 & -1.06 & -2.27 & -6.36 \\ 1.78 & 3.01 & -4.88 & -7.70 \end{array}\right)$$

Después se intercambian la columna 1 por la columna 2:

$$\left(\begin{array}{rrr|r} 9.81 & -3.39 & -4.78 & 3.95 \\ -1.06 & 4.63 & -2.27 & -6.36 \\ 3.01 & 1.78 & -4.88 & -7.70 \end{array}\right)$$

Se empiezan a realizar operaciones elementales para hacer cero los elementos que se encuentran debajo del pivote.

$$\left(\begin{array}{rrr|r} 9.81 & -3.39 & -4.78 & 3.95 \\ -1.06 & 4.63 & -2.27 & -6.36 \\ 3.01 & 1.78 & -4.88 & -7.70 \end{array}\right) \xrightarrow[R_3 \to -3.01R_1+R_3]{R_2 \to 1.06R_1+R_2}$$

$$\left(\begin{array}{rrr|r} 9.81 & -3.39 & -4.78 & 3.95 \\ 0 & 4.264 & -2.787 & -5.933 \\ 0 & 2.820 & -3.413 & -8.912 \end{array}\right)$$

Para escoger el pivote a_{22} se busca en toda la matriz a excepción de la columna ya afectada, el elemento de mayor valor absoluto, en este caso es $a_{22} = 4.264$; por lo que no se necesita realizar intercambio de renglones y columnas, ya que este elemento se encuentra en la posición diagonal.

$$\left(\begin{array}{rrr|r} 9.81 & -3.39 & -4.78 & 3.95 \\ 0 & 4.264 & -2.787 & -5.933 \\ 0 & 2.820 & -3.413 & -8.912 \end{array}\right)$$

Se empiezan a realizar operaciones elementales para hacer cero los elementos que se encuentran debajo del pivote.

$$\left(\begin{array}{rrr|r} 9.81 & -3.39 & -4.78 & 3.95 \\ 0 & 4.264 & -2.787 & -5.933 \\ 0 & 2.820 & -3.413 & -8.912 \end{array}\right) R_3 \to -\frac{2.820}{4.264}R_2+R_3$$

$$\left(\begin{array}{rrr|r} 9.81 & -3.39 & -4.78 & 3.95 \\ 0 & 4.264 & -2.787 & -5.933 \\ 0 & 0 & -1.570 & -4.988 \end{array}\right)$$

Como se intercambiaron la primera columna por la segunda, se cambia x_2 por x_1 en el sistema obtenido. Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas.

$$\Rightarrow \begin{array}{rcrcrcr} 9.81x_2 & - & 3.39x_1 & - & 4.78x_3 & = & 3.95 \\ & & 4.262x_1 & - & 2.787x_3 & = & -5.933 \\ & & & & -1.570x_3 & = & -4.988 \end{array} \qquad \begin{array}{rcl} x_1 & \approx & 0.6850 \\ x_2 & \approx & 2.188 \\ x_3 & \approx & 3.177 \end{array}$$

Una buena aproximación al sistema (3.46) es $x_1 = 0.6842$, $x_2 = 2.187$ y $x_3 = 3.176$. Comparando los resultados obtenidos de los ejemplos 3.3.18 y 3.3.19, se puede observar que si se aplica pivoteo total los errores relativos son menores. Errores sin aplicar pivoteo total:

$$\varepsilon_{r,x_1} = \left|\tfrac{0.6842-0.6966}{0.6842}\right| \times 100 = 1.81\%$$
$$\varepsilon_{r,x_2} = \left|\tfrac{2.187-2.197}{2.187}\right| \times 100 = 0.457\%$$
$$\varepsilon_{r,x_3} = \left|\tfrac{3.176-3.186}{3.176}\right| \times 100 = 0.315\%$$

Errores aplicando pivoteo total:

$$\varepsilon_{r,x_1} = \left|\tfrac{0.6842-0.6850}{0.6842}\right| \times 100 = 0.117\%$$
$$\varepsilon_{r,x_2} = \left|\tfrac{2.187-2.188}{2.187}\right| \times 100 = 0.0457\%$$
$$\varepsilon_{r,x_3} = \left|\tfrac{3.176-3.177}{3.176}\right| \times 100 = 0.0315\%$$

Al aplicar pivoteo total, se reducen los errores de redondeo y se obtiene una mejor aproximación de la solución.
Nota: En la mayor parte de los problemas el pivoteo completo no es mucho más exacto que el pivoteo parcial; al menos no el suficiente para justificar el trabajo adicional que implica; por esta razón el pivoteo parcial se usa más.

Ejemplo 3.3.20 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss, utilizando pivoteo escalado, con 4 cifras significativas; (det $A = 28696.6707$, obtenido con Maple).

$$\begin{array}{rcrcrcrcr} 1.19x_1 & + & 2.11x_2 & - & 100x_3 & + & x_4 & = & 1.12 \\ 14.2x_1 & - & 0.122x_2 & + & 12.2x_3 & + & x_4 & = & 3.44 \\ & & 100x_2 & - & 99.9x_3 & + & x_4 & = & 2.15 \\ 15.3x_1 & + & 0.110x_2 & - & 13.1x_3 & - & x_4 & = & 4.16 \end{array}$$

Como det $A \neq 0$ se empieza a aplicar el método.
Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 1.19 & 2.11 & -100 & 1 \\ 14.2 & -0.122 & 12.2 & -1 \\ 0 & 100 & -99.9 & 1 \\ 15.3 & 0.110 & -13.1 & -1 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 1.12 \\ 3.44 \\ 2.15 \\ 4.16 \end{pmatrix}$$

Se combinan las matrices A y b para obtener su forma aumentada:

$$\left(\begin{array}{cccc|c} 1.19 & 2.11 & -100 & 1 & 1.12 \\ 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 15.3 & 0.110 & -13.1 & -1 & 4.16 \end{array}\right)$$

Empezando a aplicar pivoteo escalado, se calculan los factores de escala s_1, s_2, s_3 y s_4:

$$s_1 = \max_{1 \le l \le 4}(3)\{|1.19|,|2.11|,|-100|,|1|\} = 100$$

$$s_2 = \max_{1 \le l \le 4}(4)\{|14.2|,|-0.122|,|12.2|,|-1|\} = 14.2$$

$$s_3 = \max_{1 \le l \le 4}(5)\{|0|,|100|,|-99.9|,|1|\} = 100$$

$$s_4 = \max_{1 \le l \le 4}(6)\{|15.3|,|0.110|,|-13.1|,|-1|\} = 15.3$$

Para escoger el pivote a_{11}; se calculan los siguientes elementos:

$$\frac{|a_{11}|}{s_1} = \tfrac{|1.19|}{100} = -0.01190$$

$$\frac{|a_{21}|}{s_2} = \tfrac{|14.2|}{14.2} = 1.000$$

$$\frac{|a_{31}|}{s_3} = \tfrac{|0|}{100} = 0$$

$$\frac{|a_{41}|}{s_4} = \tfrac{|15.3|}{15.3} = 1.000$$

Después se encuentra el máximo de los elementos anteriores:

$$\max\{-0.01190, 1.000, 0, 1.000\} = \frac{|a_{21}|}{s_1} = \frac{|a_{41}|}{s_4} = 1$$

En este caso existen dos máximos, por lo que se puede tomar cualquiera de ellos, tomando a $\tfrac{|a_{21}|}{s_1}$, entonces se intercambia el renglón 1 por el renglón 2.

$$\left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 1.19 & 2.11 & -100 & 1 & 1.12 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 15.3 & 0.110 & -13.1 & -1 & 4.16 \end{array}\right)$$

Se empiezan a realizar operaciones elementales para hacer cero los elementos que se encuentran debajo del pivote.

$$\left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 1.19 & 2.11 & -100 & 1 & 1.12 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 15.3 & 0.110 & -13.1 & -1 & 4.16 \end{array}\right) \xrightarrow[R_4 \to -\frac{15.3}{14.2}R_1 + R_4]{R_2 \to -\frac{1.19}{14.2}R_1 + R_2} \left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 0 & 2.120 & -101.0 & 1.084 & 0.8317 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 0 & 0.2414 & -26.24 & 0.077 & 0.455 \end{array}\right)$$

Para escoger el pivote a_{22}; se calculan los siguientes elementos:

$$\frac{|a_{22}|}{s_1} = \frac{|2.120|}{100} = 0.02120$$

$$\frac{|a_{32}|}{s_3} = \frac{|100|}{100} = 1.000$$

$$\frac{|a_{42}|}{s_4} = \frac{|0.2414|}{15.3} = 0.01578$$

Después se encuentra el máximo de los elementos anteriores:

$$\max\{0.02120, 1.000, 0.01578\} = \frac{|a_{32}|}{s_3} = 1$$

Tomando a $\frac{|a_{32}|}{s_3}$, entonces se cambia el renglón 2 por el renglón 3.

$$\left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 0 & 2.120 & -101.0 & 1.084 & 0.8317 \\ 0 & 0.2414 & -26.24 & 0.077 & 0.455 \end{array}\right)$$

Se empiezan a realizar operaciones elementales para hacer cero los elementos que se encuentran debajo del pivote.

$$\left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 0 & 2.120 & -101.0 & 1.084 & 0.8317 \\ 0 & 0.2414 & -26.24 & 0.077 & 0.455 \end{array}\right)$$

$$\xrightarrow[R_4 \to -\frac{0.2414}{100}R_2 + R_4]{R_3 \to -\frac{2.120}{100}R_2 + R_3} \left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 0 & 0 & -98.88 & 1.063 & 0.7861 \\ 0 & 0 & -26.00 & 0.07459 & 0.4498 \end{array}\right)$$

Para escoger el pivote a_{33}; se calculan los siguientes elementos:

$$\frac{|a_{33}|}{s_1} = \frac{|-98.88|}{100} = 0.9888$$

$$\frac{|a_{43}|}{s_4} = \frac{|-26.00|}{15.3} = 1.699$$

Después se encuentra el máximo de los elementos anteriores:

$$\max\{0.9888, 1.699\} = \frac{|a_{43}|}{s_4} = 1.699$$

Tomando a $\frac{|a_{43}|}{s_4}$, entonces se cambia el renglón 3 por el renglón 4.

$$\left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 0 & 0 & -26.00 & 0.07459 & 0.4498 \\ 0 & 0 & -98.88 & 1.063 & 0.7861 \end{array}\right)$$

Se empiezan a realizar operaciones elementales para hacer cero los elementos que se encuentran debajo del pivote.

$$\left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 0 & 0 & -26.00 & 0.07459 & 0.4498 \\ 0 & 0 & -98.88 & 1.063 & 0.7861 \end{array}\right)$$

$$\xrightarrow{R_4 \to -\left(\frac{-98.88}{-26.00}\right)R_3 + R_4} \left(\begin{array}{cccc|c} 14.2 & -0.122 & 12.2 & -1 & 3.44 \\ 0 & 100 & -99.9 & 1 & 2.15 \\ 0 & 0 & -26.00 & 0.07459 & 0.4498 \\ 0 & 0 & 0 & 0.7793 & -0.9249 \end{array}\right)$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas.

$$\Rightarrow \begin{array}{rcrcrcrcr} 14.2x_1 & - & 0.122x_2 & + & 12.2x_3 & - & x_4 & = & 3.44 \\ & - & 100x_2 & - & 99.9x_3 & + & x_4 & = & 2.15 \\ & & & - & 26.00x_3 & + & 0.07459x_4 & = & 2.15 \\ & & & & & + & 0.7793x_4 & = & -0.9249 \end{array} \qquad \begin{array}{rcr} x_1 & \approx & 0.1765 \\ x_2 & \approx & 0.01269 \\ x_3 & \approx & -0.02070 \\ x_4 & \approx & -1.187 \end{array}$$

3.4 Métodos iterativos

Éstos se emplean rara vez para resolver sistemas de ecuaciones lineales de dimensión pequeña, ya que el tiempo requerido para lograr un precisión suficiente excede al de los métodos exactos. Sin embargo, para sistemas grandes son eficientes en términos de velocidad y requerimientos de la memoria de la computadora.

Los sistemas de ecuaciones lineales grandes se presentan en la solución numérica de ecuaciones diferenciales parciales, en la solución de modelos resultantes en la simulación de columnas de destilación, en análisis de circuitos, en problemas con valores en la frontera, etc.; (los cuales en general forman sistemas con matrices esparcidas).

3.4.1 Mejoramiento iterativo de la solución

Los métodos iterativos para resolver un sistema de ecuaciones lineales $Ax = b$ comienzan con una aproximación inicial [9] $x^{(0)}$ a la solución x y genera una sucesión de vectores $\{x^{(k)}\}_{k=0}^{\infty}$ que si convergen, lo hacen a x. La mayoría de éstos métodos involucran un proceso que convierte el sistema $Ax = b$ en un sistema equivalente de la forma $x = Tx + c$ para alguna matriz T de $n \times n$ y un vector c. Se construye entonces la sucesión de vectores a partir de la fórmula de iteración:

$$x^{(k)} = Tx^{(k-1)} + c \qquad \text{para} k = 1, 2, 3\ldots \tag{3.47}$$

Antes de empezar el estudio de los métodos iterativos con que se resuelven los sistemas lineales de ecuaciones, se necesita medir la distancia entre dos vectores columna, para determinar si una serie de esos vectores converge a la solución del sistema.

Las siguientes son algunas normas vectoriales en R^n. Si $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in R^n$ entonces,

1. **La norma euclidiana (o norma 2) definida por**:

$$\|X\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}} \tag{3.48}$$

2. **La norma suma (o norma 1) definida por**:

$$\|X\|_1 = \sum_{i=1}^{n} |x_i| \tag{3.49}$$

3. **La norma del máximo (o norma ∞) definida por**:

$$\|X\|_\infty = \max_{1 \le i \le n} |x_i| \tag{3.50}$$

Estas normas en R^n, inducen las siguientes nociones de distancia entre dos vectores X, $Y \in R^n$:

1. **La norma euclidiana (o norma 2) definida por**:

$$\|X - Y\|_2 = \left(\sum_{i=1}^{n} (x_i - y_i)^2 \right)^{\frac{1}{2}} \tag{3.51}$$

2. **La norma suma (o norma 1) definida por**:

$$\|X - Y\|_1 = \sum_{i=1}^{n} |x_i - y_i| \tag{3.52}$$

---

[9] El vector inicial $x^{(0)}$ puede ser arbitrario. Sin embargo, si se conoce un vector como buena aproximación a la solución, éste debe utilizarse como $x^{(0)}$; de lo contrario se puede igualar simplemente a cero.

3. **La norma del máximo (o norma ∞) definida por**:

$$\|X - Y\|_\infty = \max_{1 \le i \le n} |x_i - y_i| \tag{3.53}$$

Definición 3.4.1 Una matriz es estrictamente dominante diagonalmente (E.D.D.) si cada elemento de la diagonal principal es mayor en (valor absoluto) que la suma de los valores absolutos de todos los demás elementos de la misma fila o columna; es decir:

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{ij}| \qquad 1 \le i \le n \qquad \text{E.D.Dporfilas}$$

$$|a_{ii}| > \sum_{\substack{j=1 \\ j \ne i}}^{n} |a_{ji}| \qquad 1 \le i \le n \qquad \text{E.D.Dporcolumnas} \tag{3.54}$$

Teorema 3.4.1 Si A es estrictamente dominante diagonalmente, entonces la sucesión de vectores que resulta al aplicar los métodos iterativos converge a la solución de Ax=b para cualquier vector inicial.

El teorema 3.4.1 es una implicación de un solo sentido. El hecho de que un sistema no sea estrictamente dominante diagonalmente, no significa que los métodos iterativos vayan a ser divergentes. Habrá casos en donde uno de los métodos converge y el otro diverge. Sin embargo, si cualquiera de éstos métodos converge, entonces deben converger hacia la solución, no pueden converger hacia algún otro punto.
Cuando la matriz A, no es $E.D.D$, se recomienda reordenar las ecuaciones del sistema dado, realizando intercambios de renglones o columnas, de modo que la matriz de coeficientes del sistema reordenado se convierta en una matriz $E.D.D$, ó sea lo más cercana posible a una matriz $E.D.D$ (colocando primero las ecuaciones que tienen un coeficiente dominante y luego las restantes).
Teorema 3.4.2 Para cualquier $x^{(0)}$ en R^n, la sucesión de vectores $\{x^{(k)}\}_{k=0}^{\infty}$ definida por la fórmula de iteración $x^{(k)} = Tx^{(k-1)} + c$ para $k = 1, 2, 3, \ldots$, converge a la única solución x del sistema $x = Tx + c$ si y sólo si $\rho(T) < 1$.

Para aplicar el teorema 3.4.2 es necesario definir $\rho(T)$ [10](radio espectral).
Definición 3.4.2 El radio espectral $\rho(A)$ de una matriz A, está definido por:

$$\rho(A) = \max |\lambda| \qquad \text{donde}$$

λ es un valor característico de A
(Cuando λ es un número complejo, $\lambda = \alpha + \beta i$, se tiene $|\lambda| = (\alpha^2 + \beta^2)^{\frac{1}{2}}$.)

Para calcular los valores de λ, se utiliza la siguiente definición:

Definición 3.4.3 Si A es una matriz cuadrada, el polinomio definido por

$$\rho(\lambda) = \det(A - \lambda I)$$

[10]Mientras $\rho(T)$ se aproxime más a cero más rápida será la convergencia.

recibe el nombre de polinomio característico de A, con este polinomio se obtienen los valores de λ.

En resumen, con el polinomio característico de A se obtienen los valores de λ, y si el máximo de éstos en valor absoluto es menor que 1, el método converge a la única solución x del sistema; de lo contrario el método no converge.

3.4.2 Método de Jacobi

Dado un sistema de ecuaciones de la forma (3.1) (donde $a_{ii} \neq 0$ para todo $i = 1, 2, \ldots, n$), el *método de Jacobi* consiste básicamente en realizar los siguientes pasos:

1. Despejar de cada ecuación la variable sobre la diagonal principal, es decir despejar x_1 de la primera ecuación, x_2 de la segunda, x_3 de la tercera, etc., obteniéndose:

$$\begin{array}{rlllll}
x_1 = & & -\frac{a_{12}}{a_{11}}x_2 & -\frac{a_{13}}{a_{11}}x_3 & -\ldots -\frac{a_{1n}}{a_{11}}x_n & +\frac{b_1}{a_{11}} \\
x_2 = & -\frac{a_{21}}{a_{22}}x_1 & & -\frac{a_{23}}{a_{22}}x_3 & -\ldots -\frac{a_{2n}}{a_{22}}x_n & +\frac{b_2}{a_{22}} \\
x_3 = & -\frac{a_{31}}{a_{33}}x_1 & -\frac{a_{32}}{a_{33}}x_2 & & -\ldots -\frac{a_{3n}}{a_{33}}x_n & +\frac{b_3}{a_{33}} \\
\vdots & \vdots & & & \ddots & \vdots \\
x_n = & -\frac{a_{n1}}{a_{nn}}x_1 & -\frac{a_{n2}}{a_{nn}}x_2 & -\frac{a_{n3}}{a_{nn}}x_3 & -\ldots & +\frac{b_n}{a_{nn}}
\end{array} \tag{3.55}$$

Generalizando es sistema de ecuaciones (3.55), se obtiene:

$$x_i = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{a_{ij}x_j}{a_{ii}} \qquad \text{para } i = 1, 2, 3 \ldots n \tag{3.56}$$

Así mismo el sistema (3.55) se puede representar matricialmente de la siguiente forma:

$$\underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 0 & -\frac{a_{12}}{a_{11}} & -\frac{a_{13}}{a_{11}} & \ldots & -\frac{a_{1n}}{a_{11}} \\ -\frac{a_{21}}{a_{22}} & 0 & -\frac{a_{23}}{a_{22}} & \ldots & -\frac{a_{2n}}{a_{22}} \\ -\frac{a_{31}}{a_{33}} & -\frac{a_{32}}{a_{33}} & 0 & \ldots & -\frac{a_{3n}}{a_{11}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\frac{a_{n1}}{a_{nn}} & -\frac{a_{n2}}{a_{nn}} & -\frac{a_{n3}}{a_{nn}} & \ldots & 0 \end{pmatrix}}_{T} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{pmatrix}}_{x} + \underbrace{\begin{pmatrix} \frac{b_1}{a_{11}} \\ \frac{b_2}{a_{22}} \\ \frac{b_3}{a_{33}} \\ \vdots \\ \frac{b_n}{a_{nn}} \end{pmatrix}}_{c}$$

2. Asignar un vector inicial $x^{(0)} = (x_1, x_2, x_3, \ldots, x_n)^t$. (El superíndice indica el número de la sustitución efectuada).

3. Sustituir el vector inicial $x^{(0)}$ en el sistema (3.55) para obtener $x^{(1)}$.

4. Sustituir el vector $x^{(1)}$ en el sistema (3.55) para obtener $x^{(2)}$; se repite el proceso k veces para llegar a la aproximación; como se muestra en el siguiente sistema:

$$
\begin{array}{llllllll}
x_1^k = & & -\frac{a_{12}}{a_{11}}x_2^{k-1} & -\frac{a_{13}}{a_{11}}x_3^{k-1} & - & \ldots & -\frac{a_{1n}}{a_{11}}x_n^{k-1} & +\frac{b_1}{a_{11}} \\
x_2^k = & -\frac{a_{21}}{a_{22}}x_1^{k-1} & & -\frac{a_{23}}{a_{22}}x_3^{k-1} & - & \ldots & -\frac{a_{2n}}{a_{22}}x_n^{k-1} & +\frac{b_2}{a_{22}} \\
x_3^k = & -\frac{a_{31}}{a_{33}}x_1^{k-1} & -\frac{a_{32}}{a_{33}}x_2^{k-1} & & - & \ldots & -\frac{a_{3n}}{a_{33}}x_n^{k-1} & +\frac{b_3}{a_{33}} \\
\vdots & & \vdots & & & \ddots & & \vdots \\
x_n^k = & -\frac{a_{n1}}{a_{nn}}x_1^{k-1} & -\frac{a_{n2}}{a_{nn}}x_2^{k-1} & -\frac{a_{n3}}{a_{nn}}x_3^{k-1} & - & \ldots & & +\frac{b_n}{a_{nn}}
\end{array} \tag{3.57}
$$

Generalizando el sistema de ecuaciones (1), se obtiene el **método iterativo de Jacobi**:

$$x_i^{(k)} = \frac{b_i}{a_{ii}} - \sum_{\substack{j=1 \\ j\neq i}}^{n} \frac{a_{ij}x_j^{(k-1)}}{a_{ii}} \qquad \text{para } i = 1, 2, 3, \ldots n \tag{3.58}$$

5. Escogida alguna norma vectorial $(\|.\|_1,\ \|.\|_2 \text{ ó } \|.\|_\infty)$, alguna tolerancia $\varepsilon > 0$ y un número máximo de iteraciones N; continuar iterando hasta que se satisfaga alguno de los siguientes criterios de convergencia:

$$\|X^{(k)} - X^{(k-1)}\| \leq \varepsilon \tag{3.59}$$

$$\frac{\|X^{(k)} - X^{(k-1)}\|}{\|X^{(k)}\|} \leq \varepsilon \tag{3.60}$$

$$k > N \tag{3.61}$$

Para obtener el primer criterio de convergencia $\|X^{(k)} - X^{(k-1)}\|$ (*error absoluto*), se utiliza cualquiera de las tres normas definidas en las ecuaciones (3.51),(3.52) y (3.53).

Para obtener el segundo criterio de convergencia $\frac{\|X^{(k)}-X^{(k-1)}\|}{\|X^{(k)}\|}$ (*error relativo*), se utiliza cualquiera de las tres normas definidas en las ecuaciones (3.51),(3.52) y (3.53) para el numerador; y para el denominador se utiliza cualquiera de las tres normas definidas en las ecuaciones (3.48),(3.49) y (3.50). Se tiene que utilizar la misma norma en el numerador que en el denominador.

De las tres normas la que se utiliza con frecuencia es la $\|.\|_\infty$.

Análisis de convergencia para el método de Jacobi

Para analizar de la convergencia del método de Jacobi, se empieza descomponiendo la matriz A en la forma

$$A = D - L - U$$

donde D es la matriz diagonal cuya diagonal es la diagonal principal de A, $-L$ es la matriz triangular inferior obtenida de la parte triangular estrictamente inferior de A, y $-U$ es la matriz triangular superior obtenida de la parte triangular estrictamente superior de A. Con esta notación

$$\mathbf{A}=\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}$$

se divide en

$$\mathbf{A}=\underbrace{\begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}}_{D}-\underbrace{\begin{pmatrix} 0 & 0 & 0 & \dots & 0 \\ -a_{21} & 0 & 0 & \dots & 0 \\ -a_{31} & -a_{32} & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & -a_{n3} & \dots & 0 \end{pmatrix}}_{L}-\underbrace{\begin{pmatrix} 0 & -a12 & -a_{13} & \dots & -a_{1n} \\ 0 & 0 & -a_{23} & \dots & -a_{2n} \\ 0 & 0 & 0 & \dots & -a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}}_{U}$$

$=D-L-U$

$$\begin{aligned} Ax=b &\Rightarrow \text{(D−L−U)x=b} \\ &\Rightarrow \text{Dx−(L+U)x=b} \\ &\Rightarrow \text{Dx=(L+U)x+b} \\ &\Rightarrow x=D^{-1}(L+U)x+D^{-1}b \end{aligned}$$

lo que conduce a la fórmula vectorial de iteración del método de Jacobi:

$$x^{(k)}=D^{-1}(L+U)x^{(k-1)}+D^{-1}b \qquad \text{para} k=1,2,3,\dots$$

que se usa para efectos teóricos, mientras que la ecuación (3.58) se usa para los cálculos numéricos.

Donde

$$T_j=D^{-1}(L+U) \qquad matrizdeiteracióndelmétododeJacobi \tag{3.62}$$

Aplicando el teorema 3.4.2 se tiene que: el método iterativo de Jacobi converge a la única solución x del sistema $x=T_j\,x+c$, si y sólo si $\rho(T_j)<1$.

Aplicando la definición 3.4.2 se tiene que: $\rho(T_j)=\max|\lambda|$, donde λ es un valor característico de T_j.

Ejemplo 3.4.1 Resolver el siguiente sistema de ecuaciones lineales por el método de Jacobi, con 6 decimales y una tolerancia de $\varepsilon=0.0001$; (det $A=2125$, obtenido con Maple).

$$\begin{array}{rcrcrcrcr} 10x_1 & + & 5x_2 & & & & & = & 6 \\ 5x_1 & + & 10x_2 & - & 4x_3 & & & = & 25 \\ & - & 4x_2 & + & 8x_3 & - & x_4 & = & -11 \\ & & & - & x_3 & + & 5x_4 & = & -11 \end{array} \tag{3.63}$$

Como det $A\neq 0$ y el sistema es estrictamente dominante diagonalmente, se empieza a aplicar el método; ya que se puede garantizar que el método va ha converger, (teorema 3.4.1).

Se despeja de cada ecuación la variable sobre la diagonal principal, es decir se despeja x_1 de la primera ecuación, x_2 de la segunda, x_3 de la tercera y x_4 de la cuarta, obteniéndose:

$$
\begin{aligned}
x_1 &= & &-\frac{1}{2}x_2 & & & &+\frac{3}{5}\\
x_2 &= -\frac{1}{2}x_1 & & & &+\frac{2}{5}x_3 & &+\frac{5}{2}\\
x_3 &= & &\frac{1}{2}x_2 & & &+\frac{1}{8}x_4 &-\frac{11}{8}\\
x_4 &= & & & &\frac{1}{5}x_3 & &-\frac{11}{5}
\end{aligned}
\tag{3.64}
$$

Escribiendo el sistema (3.64) en la forma $x = Tx + c$, se obtiene:

$$
\mathbf{T} = \underbrace{\begin{pmatrix} 0 & -\frac{1}{2} & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{2}{5} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{8} \\ 0 & 0 & \frac{1}{5} & 0 \end{pmatrix}}_{T} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}}_{x} + \underbrace{\begin{pmatrix} \frac{3}{5} \\ \frac{5}{2} \\ -\frac{11}{8} \\ -\frac{11}{5} \end{pmatrix}}_{c}
$$

Como no se conoce una buena aproximación inicial, se iguala el vector inicial con ceros;

$$
x^{(0)} = \begin{pmatrix} 0.000000 \\ 0.000000 \\ 0.000000 \\ 0.000000 \end{pmatrix} = \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \\ x_4^{(0)} \end{pmatrix}
$$

Después se sustituye el vector inicial $x^{(0)}$ en el sistema (3.64) para obtener $x^{(1)}$:

$$
\begin{aligned}
x_1^{(1)} &= -\tfrac{1}{2}(0.000000) + \tfrac{3}{5}\\
x_2^{(1)} &= -\tfrac{1}{2}(0.000000) + \tfrac{2}{5}(0.000000) + \tfrac{5}{2}\\
x_3^{(1)} &= \tfrac{1}{2}(0.000000) + \tfrac{1}{8}(0.000000) - \tfrac{11}{8}\\
x_4^{(1)} &= \tfrac{1}{5}(0.000000) - \tfrac{11}{5}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
x_1^{(1)} &= 0.600000\\
x_2^{(1)} &= 2.500000\\
x_3^{(1)} &= -1.375000\\
x_4^{(1)} &= -2.200000
\end{aligned}
$$

Luego se sustituye $x^{(1)}$ en el sistema (3.64) para obtener $x^{(2)}$:

$$
\begin{aligned}
x_1^{(2)} &= -\tfrac{1}{2}(2.500000) + \tfrac{3}{5}\\
x_2^{(2)} &= -\tfrac{1}{2}(0.600000) + \tfrac{2}{5}(-1.375000) + \tfrac{5}{2}\\
x_3^{(2)} &= \tfrac{1}{2}(2.500000) + \tfrac{1}{8}(-2.200000) - \tfrac{11}{8}\\
x_4^{(2)} &= \tfrac{1}{5}(-1.375000) - \tfrac{11}{5}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
x_1^{(2)} &= -0.650000\\
x_2^{(2)} &= 1.650000\\
x_3^{(2)} &= -0.400000\\
x_4^{(2)} &= -2.475000
\end{aligned}
$$

Después se sustituye $x^{(2)}$ en el sistema (3.64) para obtener $x^{(3)}$:

$$
\begin{aligned}
x_1^{(3)} &= -\tfrac{1}{2}(1.650000) + \tfrac{3}{5}\\
x_2^{(3)} &= -\tfrac{1}{2}(-0.650000) + \tfrac{2}{5}(-0.400000) + \tfrac{5}{2}\\
x_3^{(3)} &= \tfrac{1}{2}(1.650000) + \tfrac{1}{8}(-2.475000) - \tfrac{11}{8}\\
x_4^{(3)} &= \tfrac{1}{5}(-0.400000) - \tfrac{11}{5}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
x_1^{(3)} &= -0.225000\\
x_2^{(3)} &= 2.665000\\
x_3^{(3)} &= -0.859375\\
x_4^{(3)} &= -2.280000
\end{aligned}
$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$x_4^{(k)}$	ε
0	0.000000	0.000000	0.000000	0.000000	——
1	0.600000	2.500000	-1.375000	-2.200000	1.000000
2	-0.650000	1.650000	-0.400000	-2.475000	0.505051
3	-0.225000	2.665000	-0.859375	-2.280000	0.380863
4	-0.732500	2.268750	-0.327500	-2.371875	0.224242
5	-0.534375	2.735250	-0.537109	-2.265500	0.170551
6	-0.767625	2.552344	-0.290563	-2.307422	0.096596
7	-0.676172	2.767588	-0.387256	-2.258113	0.077773
8	-0.783794	2.683184	-0.273470	-2.277451	0.042407
9	-0.741592	2.782509	-0.318090	-2.254694	0.035696
10	-0.791254	2.743560	-0.265582	-2.263618	0.019138
11	-0.771780	2.789394	-0.286172	-2.253116	0.016432
12	-0.794697	2.771421	-0.261942	-2.257234	0.008743
13	-0.785711	2.792572	-0.271444	-2.252388	0.007574
14	-0.796286	2.784278	-0.260263	-2.254289	0.004016
15	-0.792139	2.794038	-0.264647	-2.252053	0.003493
16	-0.797019	2.790211	-0.259488	-2.252929	0.001849
17	-0.795105	2.794714	-0.261511	-2.251898	0.001612
18	-0.797357	2.792948	-0.259130	-2.252302	0.000852
19	-0.796474	2.795027	-0.260064	-2.251826	0.000744
20	-0.797513	2.794212	-0.258965	-2.252013	0.000393
21	-0.797106	2.795171	-0.259396	-2.251793	0.000343
22	-0.797585	2.794795	-0.258889	-2.251879	0.000181
23	-0.797397	2.795237	-0.259088	-2.251778	0.000158
24	-0.797619	2.795064	-0.258854	-2.251818	0.000084

Cuadro 3.3: Ejemplo del método de Jacobi

Escogiendo la ecuación (3.60) (*error relativo*) para saber cuando dejar de iterar, se debe se cumplir que $\frac{\|X^{(k)}-X^{(k-1)}\|}{\|X^{(k)}\|} \leq 0.0001;$ esto se cumple en la iteración 24, $(0.000084 \leq 0.0001)$.

Para obtener $x^{(4)}, x^{(5)}, x^{(6)}, x^{(7)}, \cdots, x^{(24)}$; se repite el proceso anterior. Las iteraciones se muestran el cuadro 3.3.

A continuación se indica como calcular la ecuación (3.60) con $\|.\|_\infty$.
Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\frac{\left\|X^{(1)}-X^{(0)}\right\|_\infty}{\left\|X^{(1)}\right\|_\infty} = \frac{2.500000}{2.500000} = 1.000000$$

Operaciones:

$$\left\|X^{(1)}-X^{(0)}\right\|_\infty = \max\left\{\begin{matrix} |\,0.600000-0.000000\,|, \\ |\,2.500000-0.000000\,|, \\ |-1.375000-0.000000\,|, \\ |-2.200000-0.000000\,| \end{matrix}\right\} = \left\{\begin{matrix} 0.600000, \\ 2.500000, \\ 1.375000, \\ 2.200000 \end{matrix}\right\} = 2.500000$$

$$\left\|X^{(1)}\right\|_{\infty}=\max\begin{Bmatrix}|0.600000|,\\|2.500000|,\\|-1.375000|,\\|-2.200000|\end{Bmatrix}=\begin{Bmatrix}0.600000,\\2.500000,\\1.375000,\\2.200000\end{Bmatrix}=2.500000$$

Para la segunda iteración el error relativo se calcula:

$$\frac{\left\|X^{(2)}-X^{(1)}\right\|_{\infty}}{\left\|X^{(2)}\right\|_{\infty}}=\frac{1.250000}{2.475000}=0.505051$$

Operaciones:

$$\left\|X^{(2)}-X^{(1)}\right\|_{\infty}=\max\begin{Bmatrix}|-0.650000-0.600000|,\\|1.650000-2.500000|,\\|-0.400000-(-1.375000)|,\\|-2.475000-(-2.200000)|\end{Bmatrix}=\begin{Bmatrix}1.250000,\\0.850000,\\0.975000,\\0.275000\end{Bmatrix}=1.250000$$

$$\left\|X\right\|^{(2)}_{\infty}=\max\begin{Bmatrix}|-0.650000|,\\|1.650000|,\\|-0.400000|,\\|-2.475000|\end{Bmatrix}=\begin{Bmatrix}0.650000,\\1.650000,\\0.400000,\\2.475000\end{Bmatrix}=2.475000$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 3.3.

$$\begin{array}{lcr} x_1 & \approx & -0.797619 \\ x_2 & \approx & 2.795064 \\ x_3 & \approx & -0.258854 \\ x_4 & \approx & -2.251818 \end{array}\qquad \text{Conun } \varepsilon=0.0001$$

Comprobación:

$$\begin{array}{rcl}
10(-0.797619)+5(2.795064) & =5.999130 & \approx 6\\
5(-0.797619)+10(2.795064)-4(-0.258854) & =24.997961 & \approx 25\\
-4(2.795064)+8(-0.258854)-(-2.251818) & =-10.999270 & \approx -11\\
-(-0.258854)+5(-2.251818) & =-11.000236 & \approx -11
\end{array}$$

Ejemplo 3.4.2 Resolver el siguiente sistema de ecuaciones lineales por el método de Jacobi con una tolerancia de $\varepsilon=0.0001$; (det $A=-12$, obtenido con Maple).

$$\begin{array}{rcrcrcr}
2x_1 & - & x_2 & + & x_3 & = & -1\\
x_1 & + & x_2 & + & 3x_3 & = & 0\\
3x_1 & + & 3x_2 & + & 5x_3 & = & 4
\end{array}\tag{3.65}$$

Aunque det $A\neq 0$, el sistema no es estrictamente dominante diagonalmente; por lo que se reordenan las ecuaciones del sistema dado, de modo que la matriz de coeficientes se convierta en una matriz $E.D.D$ ó sea lo más cercana posible a una matriz $E.D.D$. En el caso, la reordenación más conveniente es intercambiar el segundo renglón por el tercero, obteniéndose:

$$
\begin{array}{rcrcrcr}
2x_1 & - & x_2 & + & x_3 & = & -1 \\
3x_1 & + & 3x_2 & + & 5x_3 & = & 4 \\
x_1 & + & x_2 & + & 3x_3 & = & 0
\end{array}
$$

Se puede observar que la matriz de coeficientes de este sistema reordenado tampoco es $E.D.D.$, por lo que antes de empezar a aplicar el método se debe verificar que $\rho(T_j) < 1$.

Para verificar esto se empieza obteniendo la matriz T_j; una forma de obtenerla, es despejando de cada ecuación la variable sobre la diagonal principal, es decir se despeja x_1 de la primera ecuación, x_2 de la segunda y x_3 de la tercera[11], obteniéndose:

$$
\begin{array}{rcrrcrcrcr}
x_1 & = & & & & \frac{1}{2}x_2 & - & \frac{1}{2}x_3 & - & \frac{1}{2} \\
x_2 & = & - & x_1 & & & - & \frac{5}{3}x_3 & + & \frac{4}{3} \\
x_3 & = & - & \frac{1}{3}x_1 & - & \frac{1}{3}x_2 & + & & + & 0
\end{array}
\tag{3.66}
$$

Escribiendo el sistema (5) en la forma $x = Tx + c$, se obtiene:

$$
\mathbf{x} = \underbrace{\begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & -\frac{5}{3} \\ -\frac{1}{3} & -\frac{1}{3} & 0 \end{pmatrix}}_{T_j} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{x} + \underbrace{\begin{pmatrix} -\frac{1}{2} \\ \frac{4}{3} \\ 0 \end{pmatrix}}_{c}
$$

$$
\Rightarrow \mathbf{T_j} = \begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & -\frac{5}{3} \\ -\frac{1}{3} & -\frac{1}{3} & 0 \end{pmatrix}
$$

Después se obtiene el polinomio característico $\rho(\lambda)$ de T_j con la definición 3.4.3:

$$
\rho(\lambda) = \det(T_j - \lambda I)
$$

$$
\det\left(\begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & -\frac{5}{3} \\ -\frac{1}{3} & -\frac{1}{3} & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = \det\begin{pmatrix} -\lambda & \frac{1}{2} & -\frac{1}{2} \\ -1 & -\lambda & -\frac{5}{3} \\ -\frac{1}{3} & -\frac{1}{3} & -\lambda \end{pmatrix}
$$

[11] La otra forma de obtener T_j, es con la ecuación (3.62).

Por la definición 3.1.1, det $\begin{pmatrix} -\lambda & \frac{1}{2} & -\frac{1}{2} \\ -1 & -\lambda & -\frac{5}{3} \\ -\frac{1}{3} & -\frac{1}{3} & -\lambda \end{pmatrix}$ es:

$$-\lambda \begin{vmatrix} -\lambda & -\frac{5}{3} \\ -\frac{1}{3} & -\lambda \end{vmatrix} - \frac{1}{2} \begin{vmatrix} -1 & -\frac{5}{3} \\ -\frac{1}{3} & -\lambda \end{vmatrix} + (-\frac{1}{2}) \begin{vmatrix} -1 & -\lambda \\ -\frac{1}{3} & -\frac{1}{3} \end{vmatrix} =$$

$$\begin{aligned} -\lambda((-\lambda \times -\lambda) - (-\frac{5}{3} \times -\frac{1}{3})) - \frac{1}{2}((-1 \times -\lambda) - (-\frac{5}{3} \times -\frac{1}{3})) & \\ +(-\frac{1}{2})((-1 \times -\frac{1}{3}) - (-\lambda \times -\frac{1}{3})) & = \\ -\lambda(\lambda^2 - \frac{5}{9}) - \frac{1}{2}(\lambda - \frac{5}{9}) - \frac{1}{2}(\frac{1}{3} - \frac{1}{3}\lambda) & = \\ -\lambda^3 + \frac{5}{9}\lambda - \frac{1}{2}\lambda + \frac{5}{18} - \frac{1}{6} + \frac{1}{6}\lambda & = -\lambda^3 + \frac{2}{9}\lambda + \frac{1}{9} \end{aligned}$$

$$\Rightarrow \rho(\lambda) = -\lambda^3 + \frac{2}{9}\lambda + \frac{1}{9}$$

Las raíces del polinomio característico $\rho(\lambda)$ son, (estas se obtuvieron con Maple):

$$\begin{aligned} \lambda_1 &\approx 0.631096 \\ \lambda_2 &\approx 0.315548 + 0.276567i \\ \lambda_3 &\approx 0.315548 - 0.276567i \end{aligned}$$

Por la definición 3.4.2 se tiene que:

$$\rho(T_j) \approx \max\{|0.631096|, |0.315548 + 0.276567i|, |0.315548 - 0.276567i|\} =$$

Operaciones:

$$\begin{aligned} |0.315548 + 0.276567i| &= |(0.315548^2 + 0.276567^2)^{\frac{1}{2}}| = 0.419595 \\ |0.315548 - 0.276567i| &= |(0.315548^2 + (-0.276567)^2)^{\frac{1}{2}}| = 0.419595 \end{aligned}$$

$$\max\{|0.631096|, |0.419595|\} = 0.631096$$

$$\Rightarrow \rho(T_j) \approx 0.631096 < 1$$

Por el teorema 3.4.2 el método iterativo de Jacobi converge a la única solución x del sistema, ya que $\rho(T_j) < 1$, por lo que se empieza a aplicar el método.

Como no se conoce una buena aproximación inicial, se iguala el vector inicial con ceros;

$$x^{(0)} = \begin{pmatrix} 0.000000 \\ 0.000000 \\ 0.000000 \end{pmatrix} = \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{pmatrix}$$

Después se sustituye el vector inicial $x^{(0)}$ en el sistema (3.66) para obtener $x^{(1)}$:

$$
\begin{array}{llll}
x_1^{(1)} = & \frac{1}{2}(0.000000) - \frac{1}{2}(0.000000) - \frac{1}{2} & & x_1^{(1)} = -0.500000 \\
x_2^{(1)} = -\ 1(0.000000) & -\frac{5}{3}(0.000000) + \frac{4}{3} & \Rightarrow & x_2^{(1)} = 1.333333 \\
x_3^{(1)} = -\frac{1}{3}(0.000000) - \frac{1}{3}(0.000000) + & +\ 0 & & x_3^{(1)} = 0.000000
\end{array}
$$

Luego se sustituye $x^{(1)}$ en el sistema (3.66) para obtener $x^{(2)}$:

$$
\begin{array}{llll}
x_1^{(2)} = & \frac{1}{2}(1.333333) - \frac{1}{2}(0.000000) - \frac{1}{2} & & x_1^{(2)} = 0.166667 \\
x_2^{(2)} = -\ 1(-0.500000) & -\frac{5}{3}(0.000000) + \frac{4}{3} & \Rightarrow & x_2^{(2)} = 1.833333 \\
x_3^{(2)} = -\frac{1}{3}(-0.500000) - \frac{1}{3}(1.333333) + & +\ 0 & & x_3^{(2)} = -0.277778
\end{array}
$$

Después se sustituye $x^{(2)}$ en el sistema (3.66) para obtener $x^{(3)}$:

$$
\begin{array}{llll}
x_1^{(3)} = & \frac{1}{2}(1.833333) - \frac{1}{2}(-0.277778) - \frac{1}{2} & & x_1^{(3)} = 0.555556 \\
x_2^{(3)} = -\ 1(0.166667) & -\frac{5}{3}(-0.277778) + \frac{4}{3} & \Rightarrow & x_2^{(3)} = 1.629630 \\
x_3^{(3)} = -\frac{1}{3}(0.166667) - \frac{1}{3}(1.833333) + & +\ 0 & & x_3^{(3)} = -0.666667
\end{array}
$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	ε
0	0.000000	0.000000	0.000000	——
1	-0.500000	1.333333	0.000000	1.000000
2	0.166667	1.833333	-0.277778	0.363636
3	0.555556	1.629630	-0.666667	0.238636
4	0.648148	1.888889	-0.728395	0.137255
5	0.808642	1.899177	-0.845679	0.084507
6	0.872428	1.934156	-0.902606	0.032979
7	0.918381	1.965249	-0.935528	0.023383
8	0.950389	1.974166	-0.961210	0.016213
9	0.967688	1.984962	-0.974851	0.008715
10	0.979907	1.990398	-0.984216	0.006139
11	0.987307	1.993788	-0.990101	0.003712
12	0.991945	1.996195	-0.993698	0.002323
13	0.994947	1.997553	-0.996047	0.001503
14	0.996800	1.998464	-0.997500	0.000927
15	0.997982	1.999033	-0.998421	0.000591
16	0.998727	1.999387	-0.999005	0.000373
17	0.999196	1.999615	-0.999371	0.000234
18	0.999493	1.999756	-0.999604	0.000149
19	0.999680	1.999846	-0.999750	0.000093

Cuadro 3.4: Ejemplo del método de Jacobi

Escogiendo la ecuación (3.60) (*error relativo*) para saber cuando dejar de iterar, se debe se cumplir que $\frac{\|X^{(k)} - X^{(k-1)}\|}{\|X^{(k)}\|} \leq 0.0001$; esto se cumple en la iteración 19, $(0.000093 \leq 0.0001)$.

Para obtener $x^{(4)}, x^{(5)}, x^{(6)}, x^{(7)}, \cdots, x^{(19)}$; se repite el proceso anterior. Las iteraciones se muestran el cuadro 3.4.

A continuación se indica como calcular la ecuación (3.60) con $\|\cdot\|_\infty$.
Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\frac{\left\|X^{(1)}-X^{(0)}\right\|_\infty}{\left\|X^{(1)}\right\|_\infty}=\frac{1.333333}{1.333333}=1.000000$$

Operaciones:

$$\left\|X^{(1)}-X^{(0)}\right\|_\infty=\max\left\{\begin{matrix}|-0.500000-0.000000|,\\ |1.333333-0.000000|,\\ |0.000000-0.000000|\end{matrix}\right\}=\left\{\begin{matrix}0.500000,\\ 1.333333,\\ 0.000000\end{matrix}\right\}=1.333333$$

$$\left\|X^{(1)}\right\|_\infty=\max\left\{\begin{matrix}|-0.500000|,\\ |1.333333|,\\ |0.000000|\end{matrix}\right\}=\left\{\begin{matrix}0.500000,\\ 1.333333,\\ 0.000000\end{matrix}\right\}=1.333333$$

Para la segunda iteración el error relativo se calcula:

$$\frac{\left\|X^{(2)}-X^{(1)}\right\|_\infty}{\left\|X^{(2)}\right\|_\infty}=\frac{0.666667}{1.833333}=0.363636$$

Operaciones:

$$\left\|X^{(2)}-X^{(1)}\right\|_\infty=\max\left\{\begin{matrix}|0.166667-(-0.500000)|,\\ |1.833333-1.333333|,\\ |-0.277778-0.000000|\end{matrix}\right\}=\left\{\begin{matrix}0.666667,\\ 0.500000,\\ 0.277778\end{matrix}\right\}=0.666667$$

$$\left\|X^{(2)}\right\|_\infty=\max\left\{\begin{matrix}|0.166667|,\\ |1.833333|,\\ |-0.277778|\end{matrix}\right\}=\left\{\begin{matrix}0.166667,\\ 1.833333,\\ 0.277778\end{matrix}\right\}=1.833333$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 3.4.

$$\begin{matrix}x_1 & \approx & 0.999680 & \\ x_2 & \approx & 1.999846 & \text{Conun } \varepsilon=0.0001\\ x_3 & \approx & -0.999750 & \end{matrix}$$

Comprobación:

$$\begin{matrix}2(0.999680) & - & (1.999846) & + & (-0.999750) & = & -1.000236 & \approx & -1\\ (0.999680) & + & (1.999846) & + & 3(-0.999750) & = & 0.000277 & \approx & 0\\ 3(0.999680) & + & 3(1.999846) & + & 5(-0.999750) & = & 3.999830 & \approx & 4\end{matrix}$$

Ejemplo 3.4.3 Resolver el sistema de ecuaciones lineales (3.65) por el método de Jacobi, con 6 decimales y con una tolerancia de $\varepsilon = 0.0001$. Sin reordenar el sistema.

El sistema no es estrictamente dominante diagonalmente; por lo que no se puede garantizar que el método vaya ha converger.
Antes de empezar a aplicar el método se debe verificar que $\rho(T_j) < 1$.

Para verificar esto se empieza obteniendo la matriz T_j; una forma de obtenerla, es despejando de cada ecuación del sistema (1) la variable sobre la diagonal principal, es decir se despeja x_1 de la primera ecuación, x_2 de la segunda y x_3 de la tercera, obteniéndose:

$$\begin{aligned}
x_1 &= & & \tfrac{1}{2}x_2 & - & \tfrac{1}{2}x_3 & - & \tfrac{1}{2} \\
x_2 &= -\ x_1 & & & - & 3x_3 & & \\
x_3 &= -\ \tfrac{3}{5}x_1 & - & \tfrac{3}{5}x_2 & + & & + & \tfrac{4}{5}
\end{aligned} \tag{3.67}$$

Escribiendo el sistema (3.67) en la forma $x = Tx + c$, se obtiene:

$$\mathbf{x} = \underbrace{\begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & -3 \\ -\frac{3}{5} & -\frac{3}{5} & 0 \end{pmatrix}}_{T_j} \underbrace{\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}}_{x} + \underbrace{\begin{pmatrix} -\frac{1}{2} \\ 0 \\ \frac{4}{5} \end{pmatrix}}_{c}$$

$$\Rightarrow \mathbf{T_j} = \begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & -3 \\ -\frac{3}{5} & -\frac{3}{5} & 0 \end{pmatrix}$$

Después se obtiene el polinomio característico $\rho(\lambda)$ de T_j con la definición 3.4.3:

$$\rho(\lambda) = \det(T_j - \lambda I)$$

$$\det\left(\begin{pmatrix} 0 & \frac{1}{2} & -\frac{1}{2} \\ -1 & 0 & -3 \\ -\frac{3}{5} & -\frac{3}{5} & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) = \det \begin{pmatrix} -\lambda & \frac{1}{2} & -\frac{1}{2} \\ -1 & -\lambda & -3 \\ -\frac{3}{5} & -\frac{3}{5} & -\lambda \end{pmatrix}$$

Por la definición 3.1.1, $\det \begin{pmatrix} -\lambda & \frac{1}{2} & -\frac{1}{2} \\ -1 & -\lambda & -3 \\ -\frac{3}{5} & -\frac{3}{5} & -\lambda \end{pmatrix}$ es:

$$-\lambda \begin{vmatrix} -\lambda & -3 \\ -\frac{3}{5} & -\lambda \end{vmatrix} - \frac{1}{2} \begin{vmatrix} -1 & -3 \\ -\frac{3}{5} & -\lambda \end{vmatrix} + (-\frac{1}{2}) \begin{vmatrix} -1 & -\lambda \\ -\frac{3}{5} & -\frac{3}{5} \end{vmatrix} =$$

$$
\begin{aligned}
&-\lambda((-\lambda\times-\lambda)-(-3\times-\tfrac{3}{5}))-\tfrac{1}{2}((-1\times-\lambda)-(-3\times-\tfrac{3}{5})) \\
&\qquad +(-\tfrac{1}{2})((-1\times-\tfrac{3}{5})-(-\lambda\times-\tfrac{3}{5})) && = \\
&-\lambda(\lambda^2-\tfrac{9}{5})-\tfrac{1}{2}(\lambda-\tfrac{9}{5})-\tfrac{1}{2}(\tfrac{3}{5}-\tfrac{3}{5}\lambda) && = \\
&-\lambda^3+\tfrac{9}{5}\lambda-\tfrac{1}{2}\lambda+\tfrac{9}{10}-\tfrac{3}{10}+\tfrac{3}{10}\lambda && = \quad -\lambda^3+\tfrac{8}{5}\lambda+\tfrac{3}{5}
\end{aligned}
$$

$$\Rightarrow \rho(\lambda)=-\lambda^3+\frac{8}{5}\lambda+\frac{3}{5}$$

Las raíces del polinomio característico $\rho(\lambda)$ son, (estas se obtuvieron con Maple):

$$
\begin{aligned}
\lambda_1 &\approx -1.000000 \\
\lambda_2 &\approx -0.421954 \\
\lambda_3 &\approx 1.421954
\end{aligned}
$$

Por la definición 3.4.2 se tiene que:

$$\rho(T_j)\approx \max\{|-1.000000|,|-0.421954|,|1.421954|\}=1.421954$$

$$\Rightarrow \rho(T_j)\approx 1.421954>1$$

Por el teorema 3.4.2 el método iterativo de Jacobi no converge a la única solución x del sistema, ya que $\rho(T_j)>1$.

Comparando los ejemplos 3.4.2 y 3.4.3, se puede observar que cuando la matriz A, no es $E.D.D$, es muy importante reordenar las ecuaciones del sistema dado, de modo que la matriz de coeficientes del sistema reordenado se convierta en una matriz $E.D.D$, ó sea lo más cercana posible a una matriz $E.D.D$; ya que esto puede ser la diferencia para que el método converja a la única solución x del sistema.

Ejemplo 3.4.4 Verificar si para el siguiente sistema de ecuaciones lineales el método de Jacobi converge; sin reordenar el sistema; (det $A=1$, obtenido con Maple).

$$
\begin{array}{rcrcrcl}
x_1 & + & 2x_2 & - & 2x_3 & = & 7 \\
x_1 & + & x_2 & + & x_3 & = & 2 \\
2x_1 & + & 2x_2 & + & x_3 & = & 5
\end{array}
\tag{3.68}
$$

Aunque det $A\neq 0$, el sistema no es estrictamente dominante diagonalmente; por lo que no se puede garantizar que el método vaya ha converger.

Antes de empezar a aplicar el método se debe verificar que $\rho(T_j)<1$.

Para verificar esto se empieza obteniendo la matriz T_j; con la ecuación (3.62):

$$\mathbf{T_j} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}}_{D} \left(\underbrace{\begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -2 & -2 & 0 \end{pmatrix}}_{L} + \underbrace{\begin{pmatrix} 0 & -2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}}_{U} \right)$$

$$\mathbf{T_j} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix}$$

$$\Rightarrow \mathbf{T_j} = \begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix}$$

Después se obtiene el polinomio característico $\rho(\lambda)$ de T_j con la definición 3.4.3:

$$\rho(\lambda) = \det(T_j - \lambda I)$$

$$\det\left(\begin{pmatrix} 0 & -2 & 2 \\ -1 & 0 & -1 \\ -2 & -2 & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \right) = \det \begin{pmatrix} -\lambda & -2 & 2 \\ -1 & -\lambda & -1 \\ -2 & -2 & -\lambda \end{pmatrix}$$

Por la definición 3.1.1, $\det \begin{pmatrix} -\lambda & -2 & 2 \\ -1 & -\lambda & -1 \\ -2 & -2 & -\lambda \end{pmatrix}$ es:

$$-\lambda \begin{vmatrix} -\lambda & -1 \\ -2 & -\lambda \end{vmatrix} - (-2) \begin{vmatrix} -1 & -1 \\ -2 & -\lambda \end{vmatrix} + 2 \begin{vmatrix} -1 & -\lambda \\ -2 & -2 \end{vmatrix} =$$

$$\begin{aligned} -\lambda((-\lambda \times -\lambda) - (-1 \times -2)) + 2((-1 \times -\lambda) - (-1 \times -2)) & \\ +2((-1 \times -2) - (-\lambda \times -2)) & = \\ -\lambda(\lambda^2 \times -2) + 2(\lambda \times -2)) + 2(2 - 2\lambda) & = \\ -\lambda^3 + 2\lambda + 2\lambda - 4 + 4 - 4\lambda) & = -\lambda^3 \end{aligned}$$

$$\Rightarrow \rho(\lambda) = -\lambda^3$$

Las raíces del polinomio característico $\rho(\lambda)$ son, (estas se obtuvieron con Maple):

$$\begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= 0 \\ \lambda_3 &= 0 \end{aligned}$$

Por la definición 3.4.2 se tiene que:

$$\rho(T_j) = \max\{|\,0\,|\} = 0$$

$$\Rightarrow \rho(T_j) = 0 < 1$$

Por el teorema 3.4.2 el método iterativo de Jacobi converge a la única solución x del sistema, ya que $\rho(T_j) < 1$.

3.4.3 Método de Gauss-Seidel

El *método de Gauss-Seidel* es una mejora del método de Jacobi.
Dado un sistema de ecuaciones de la forma (3.1), (donde $a_{ii} \neq 0$ para todo $i = 1, 2, \ldots, n$), consiste básicamente en realizar los siguientes pasos:

1. Despejar de cada ecuación la variable sobre la diagonal principal, es decir despejar x_1 de la primera ecuación, x_2 de la segunda, x_3 de la tercera, etc., obteniéndose el sistema de ecuaciones (3.55).

2. Asignar un vector inicial $x^{(0)} = (x_1, x_2, x_3, \ldots, x_n)^t$.

3. Sustituir el vector inicial $x^{(0)}$ en la primera ecuación del sistema (3.55) para obtener un nuevo valor para la primera incógnita.

4. Pasar a la segunda ecuación del sistema (3.55) y utilizar el valor calculado para la incógnita del paso 3 y los valores obtenidos en la iteración anterior para las incógnitas restantes. Este procedimiento se repite hasta obtener los nuevos valores de todas las incógnitas despejadas. Es decir siempre se utiliza el valor más reciente de cada incógnita en todos los cálculos; como se muestra en el siguiente sistema:

$$\begin{array}{rcccccccc}
x_1^k = & & -\frac{a_{12}}{a_{11}}x_2^{k-1} & -\frac{a_{13}}{a_{11}}x_3^{k-1} & - & \cdots & -\frac{a_{1n}}{a_{11}}x_n^{k-1} & +\frac{b_1}{a_{11}} \\
x_2^k = & -\frac{a_{21}}{a_{22}}x_1^k & & -\frac{a_{23}}{a_{22}}x_3^{k-1} & - & \cdots & -\frac{a_{2n}}{a_{22}}x_n^{k-1} & +\frac{b_2}{a_{22}} \\
x_3^k = & -\frac{a_{31}}{a_{33}}x_1^k & -\frac{a_{32}}{a_{33}}x_2^k & & - & \cdots & -\frac{a_{3n}}{a_{33}}x_n^{k-1} & +\frac{b_3}{a_{33}} \\
\vdots & & \vdots & & & \ddots & & \vdots \\
x_n^k = & -\frac{a_{n1}}{a_{nn}}x_1^k & -\frac{a_{n2}}{a_{nn}}x_2^k & -\frac{a_{n3}}{a_{nn}}x_3^k & - & \cdots & & +\frac{b_n}{a_{nn}}
\end{array} \tag{3.69}$$

Generalizando el sistema de ecuaciones (1), se obtiene el **método iterativo de Gauss-Seidel**:

$$x_i^{(k)} = \frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1} \frac{a_{ij}x_j^{(k)}}{a_{ii}} - \sum_{j=i+1}^{n} \frac{a_{ij}x_j^{(k-1)}}{a_{ii}} \qquad \text{para} i = 1, 2, 3 \ldots n \tag{3.70}$$

5. Continuar iterando hasta que se satisfaga alguno de los criterios de convergencia definidos en las ecuaciones (3.59), (3.60) y (3.61).

Análisis de convergencia para el método de Gauss-Seidel

Para analizar de la convergencia del método de Gauss-Seidel, se empieza multiplicando ambos lados del sistema de ecuaciones (3.69) por a_{ii}, obteniéndose:

$$
\begin{array}{lllllll}
a_{11}x_1^k = & & -\frac{a_{12}}{a_{11}}a_{11}x_2^{k-1} & -\frac{a_{13}}{a_{11}}a_{11}x_3^{k-1} & - \ \dots & -\frac{a_{1n}}{a_{11}}a_{11}x_n^{k-1} & +\frac{b_1}{a_{11}}a_{11} \\
a_{22}x_2^k = & -\frac{a_{21}}{a_{22}}a_{22}x_1^k & & -\frac{a_{23}}{a_{22}}a_{22}x_3^{k-1} & - \ \dots & -\frac{a_{2n}}{a_{22}}a_{22}x_n^{k-1} & +\frac{b_2}{a_{22}}a_{22} \\
a_{33}x_3^k = & -\frac{a_{31}}{a_{33}}a_{33}x_1^k & -\frac{a_{32}}{a_{33}}a_{33}x_2^k & & - \ \dots & -\frac{a_{3n}}{a_{33}}a_{33}x_n^{k-1} & +\frac{b_3}{a_{33}}a_{33} \\
\vdots & & \vdots & & \ddots & & \vdots \\
a_{nn}x_n^k = & -\frac{a_{n1}}{a_{nn}}a_{nn}x_1^k & -\frac{a_{n2}}{a_{nn}}a_{nn}x_2^k & -\frac{a_{n3}}{a_{nn}}a_{nn}x_3^k & - \ \dots & & +\frac{b_n}{a_{nn}}a_{nn}
\end{array}
$$

Simplificando el sistema anterior y reuniendo todos los k-ésimos términos, se obtiene:

$$
\begin{array}{rllllll}
a_{11}x_1^k = & & -a_{12}x_2^{k-1} & -a_{13}x_3^{k-1} & - \ \dots & -a_{1n}x_n^{k-1} & +b_1 \\
a_{21}x_1^k + a_{22}x_2^k = & & & -a_{23}x_3^{k-1} & - \ \dots & -a_{2n}x_n^{k-1} & +b_2 \\
a_{31}x_1^k + a_{32}x_2^k + a_{33}x_3^k = & & & \dots & & -a_{3n}x_n^{k-1} & +b_3 \\
\vdots & & & & \ddots & & \vdots \\
a_{n1}x_1^k + a_{n2}x_2^k + a_{n3}x_3^k + a_{nn}x_n^k = & & & & \dots & & +b_n
\end{array}
$$

Lo cual se puede expresar matricialmente de la siguiente forma:

$$
\underbrace{\begin{pmatrix} a_{11} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & 0 & \dots & 0 \\ a_{31} & a_{22} & a_{23} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}}_{D-L}
\underbrace{\begin{pmatrix} x_1^k \\ x_2^k \\ x_3^k \\ \vdots \\ x_n^k \end{pmatrix}}_{x^k}
=
\underbrace{\begin{pmatrix} 0 & -a_{12} & -a_{13} & \dots & -a_{1n} \\ 0 & 0 & -a_{23} & \dots & -a_{2n} \\ 0 & 0 & 0 & \dots & -a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}}_{U}
\underbrace{\begin{pmatrix} x_1^{k-1} \\ x_2^{k-1} \\ x_3^{k-1} \\ \vdots \\ x_n^{k-1} \end{pmatrix}}_{x^{k-1}}
+
\underbrace{\begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix}}_{b}
$$

Con las definiciones de D, L y U para el análisis de convergencia del método de Jacobi, se obtiene:

$$(D-L)x^{(k)} = Ux^{(k-1)} + b$$

Despejando $x^{(k)}$ de la ecuación anterior, se deduce que la fórmula vectorial de iteración del método de Gauss-Seidel es:

$$x^{(k)} = (D-L)^{-1}Ux^{(k-1)} + (D-L)^{-1}b \qquad \text{para} k = 1, 2, 3, \dots$$

Donde

$$T_g = (D-L)^{-1}U \qquad matrizdeiteracióndelmétododeGauss - Seidel \tag{3.71}$$

Aplicando el teorema 3.4.2 se tiene que: el método iterativo de Gauss-Seidel converge a la única solución x del sistema $x = T_g\, x + c$, si y sólo si $\rho(T_g) < 1$.

Aplicando la definición 3.4.2 se tiene que: $\rho(T_g) = \max|\lambda|$, donde λ es un valor característico de T_g.

Ejemplo 3.4.5 Resolver el sistema de ecuaciones lineales (3.63) por el método de Gauss-Seidel, con 6 decimales.

Se despeja de cada ecuación la variable sobre la diagonal principal, es decir se despeja x_1 de la primera ecuación, x_2 de la segunda, x_3 de la tercera y x_4 de la cuarta, obteniéndose el sistema (3.64).
Como no se conoce una buena aproximación inicial, se iguala el vector inicial con ceros;

$$x^{(0)} = \begin{pmatrix} 0.000000 \\ 0.000000 \\ 0.000000 \\ 0.000000 \end{pmatrix} = \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \\ x_4^{(0)} \end{pmatrix}$$

Después se sustituye $x^{(0)}$ en la primera ecuación del sistema (3.64) para obtener un nuevo valor $x_1^{(1)}$. Para la segunda ecuación del sistema 3.64) se utilizar el valor calculado $x_1^{(1)}$ y los valores previos para las incógnitas restantes, obteniendo $x_2^{(1)}$. Para la tercera ecuación del sistema (3.64) se utilizan los valores calculados $x_1^{(1)}$ y $x_2^{(1)}$ y los valores previos para las incógnitas restantes, obteniendo $x_3^{(1)}$. Finalmente para la cuarta ecuación del sistema (3.64) se utilizan los valores calculados $x_1^{(1)}$, $x_2^{(1)}$ y $x_3^{(1)}$ y los valores previos para las incógnitas restantes, obteniendo $x_4^{(1)}$. Es decir siempre se utiliza el valor más reciente de cada incógnita en todos los cálculos.

$$\begin{array}{llllll}
x_1^{(1)} = & -\tfrac{1}{2}(0.000000) & & & +\ \tfrac{3}{5} & \\
x_2^{(1)} = -\tfrac{1}{2}(0.600000) & & +\ \tfrac{2}{5}(0.000000) & & +\ \tfrac{5}{2} & \\
x_3^{(1)} = & \tfrac{1}{2}(2.200000) & & +\ \tfrac{1}{8}(0.000000) & -\ \tfrac{11}{8} & \Rightarrow \\
x_4^{(1)} = & & \tfrac{1}{5}(-0.275000) & & -\ \tfrac{11}{5} &
\end{array}
\quad
\begin{array}{l}
x_1^{(1)} = 0.600000 \\
x_2^{(1)} = 2.200000 \\
x_3^{(1)} = -0.275000 \\
x_4^{(1)} = -2.255000
\end{array}$$

Se sigue el procedimiento anterior para obtener $x^{(2)}$:

$$\begin{array}{llllll}
x_1^{(2)} = & -\tfrac{1}{2}(2.200000) & & & +\ \tfrac{3}{5} & \\
x_2^{(2)} = -\tfrac{1}{2}(-0.500000) & & +\ \tfrac{2}{5}(-0.275000) & & +\ \tfrac{5}{2} & \\
x_3^{(2)} = & \tfrac{1}{2}(2.640000) & & +\ \tfrac{1}{8}(-2.255000) & -\ \tfrac{11}{8} & \Rightarrow \\
x_4^{(2)} = & & \tfrac{1}{5}(-0.336875) & & -\ \tfrac{11}{5} &
\end{array}
\quad
\begin{array}{l}
x_1^{(2)} = -0.500000 \\
x_2^{(2)} = 2.640000 \\
x_3^{(2)} = -0.336875 \\
x_4^{(2)} = -2.267375
\end{array}$$

Se sigue el mismo procedimiento para obtener $x^{(3)}$:

$$\begin{array}{llllll}
x_1^{(3)} = & -\tfrac{1}{2}(2.640000) & & & +\ \tfrac{3}{5} & \\
x_2^{(3)} = -\tfrac{1}{2}(-0.720000) & & +\ \tfrac{2}{5}(-0.336875) & & +\ \tfrac{5}{2} & \\
x_3^{(3)} = & \tfrac{1}{2}(2.725250) & & +\ \tfrac{1}{8}(-2.267375) & -\ \tfrac{11}{8} & \Rightarrow \\
x_4^{(3)} = & & \tfrac{1}{5}(-0.295797) & & -\ \tfrac{11}{5} &
\end{array}
\quad
\begin{array}{l}
x_1^{(3)} = -0.720000 \\
x_2^{(3)} = 2.725250 \\
x_3^{(3)} = -0.295797 \\
x_4^{(3)} = -2.259159
\end{array}$$

Escogiendo la ecuación (3.60) (*error relativo*) para saber cuando dejar de iterar, se debe se cumplir que $\frac{\left\| X^{(k)} - X^{(k-1)} \right\|}{\left\| X^{(k)} \right\|} \leq 0.0001$; esto se cumple en la iteración 11, $(0.000065 \leq 0.0001)$.

Para obtener $x^{(4)}, x^{(5)}, x^{(6)}, x^{(7)}, \cdots, x^{(11)}$; se repite el proceso anterior. Las iteraciones se muestran el cuadro 3.5.
A continuación se indica como calcular la ecuación (3.60) con $\left\| . \right\|_\infty$.
Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\frac{\left\| X^{(1)} - X^{(0)} \right\|_\infty}{\left\| X^{(1)} \right\|_\infty} = \frac{2.255000}{2.255000} = 1.000000$$

Operaciones:

$$\left\| X^{(1)} - X^{(0)} \right\|_\infty = \max \begin{Bmatrix} |\,0.600000 - 0.000000\,|, \\ |\,2.200000 - 0.000000\,|, \\ |-0.275000 - 0.000000\,|, \\ |-2.255000 - 0.000000\,| \end{Bmatrix} = \begin{Bmatrix} 0.600000, \\ 2.200000, \\ 0.275000, \\ 2.255000 \end{Bmatrix} = 2.255000$$

$$\left\| X^{(1)} \right\|_\infty = \max \begin{Bmatrix} |\,0.600000\,|, \\ |\,2.200000\,|, \\ |-0.275000\,|, \\ |-2.255000\,| \end{Bmatrix} = \begin{Bmatrix} 0.600000, \\ 2.200000, \\ 0.275000, \\ 2.255000 \end{Bmatrix} = 2.255000$$

Para la segunda iteración el error relativo se calcula:

$$\frac{\left\| X^{(2)} - X^{(1)} \right\|_\infty}{\left\| X^{(2)} \right\|_\infty} = \frac{1.100000}{2.640000} = 0.416667$$

Operaciones:

$$\left\| X^{(2)} - X^{(1)} \right\|_\infty = \max \begin{Bmatrix} |-0.500000 - 0.600000\,|, \\ |\,2.640000 - 2.200000\,|, \\ |-0.336875 - (-0.275000)\,|, \\ |-2.267375 - (-2.255000)\,| \end{Bmatrix} = \begin{Bmatrix} 1.100000, \\ 0.440000, \\ 0.061875, \\ 0.012375 \end{Bmatrix} = 1.100000$$

$$\left\| X^{(2)} \right\|_\infty = \max \begin{Bmatrix} |-0.500000\,|, \\ |\,2.640000\,|, \\ |-0.336875\,|, \\ |-2.267375\,| \end{Bmatrix} = \begin{Bmatrix} 0.500000, \\ 2.640000, \\ 0.336875, \\ 2.267375 \end{Bmatrix} = 2.640000$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 3.5.

$$\begin{aligned} x_1 &\approx -0.797491 \\ x_2 &\approx 2.795150 \\ x_3 &\approx -0.258900 \\ x_4 &\approx -2.251780 \end{aligned} \qquad \text{Conun } \varepsilon = 0.0001$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$x_4^{(k)}$	ε
0	0.000000	0.000000	0.000000	0.000000	——
1	0.600000	2.200000	-0.275000	-2.255000	1.000000
2	-0.500000	2.640000	-0.336875	-2.267375	0.416667
3	-0.720000	2.725250	-0.295797	-2.259159	0.080727
4	-0.762625	2.762994	-0.275898	-2.255180	0.015427
5	-0.781497	2.780389	-0.266703	-2.253341	0.006787
6	-0.790195	2.788416	-0.262459	-2.252492	0.003119
7	-0.794208	2.792120	-0.260501	-2.252100	0.001437
8	-0.796060	2.793830	-0.259598	-2.251920	0.000663
9	-0.796915	2.794618	-0.259181	-2.251836	0.000306
10	-0.797309	2.794982	-0.258988	-2.251798	0.000141
11	-0.797491	2.795150	-0.258900	-2.251780	0.000065

Cuadro 3.5: Ejemplo del método de Gauss-Seidel

Comprobación:

$$\begin{array}{rcrcrcrcrcr}
10(-0.797491) & + & 5(2.795150) & & & & & = & 6.000840 & \approx & 6 \\
5(-0.797491) & + & 10(2.795150) & - & 4(-0.258900) & & & = & 24.999645 & \approx & 25 \\
 & - & 4(2.795150) & + & 8(-0.258900) & - & (-2.251780) & = & -11.000002 & \approx & -11 \\
 & & & - & (-0.258900) & + & 5(-2.251780) & = & -11.000000 & \approx & -11
\end{array}$$

Comparando los ejemplos 3.4.1 y 3.4.5, se puede observar que si un sistema es estrictamente dominante diagonalmente, entonces el método de Gauss-Seidel converge más rápidamente que el de Jacobi. (Para el método de Gauss-Seidel se realizaron 11 iteraciones y para el de Jacobi 24).

Ejemplo 3.4.6 Resolver el sistema de ecuaciones lineales (3.68) por el método de Gauss-Seidel, con 6 decimales y una tolerancia de $\varepsilon = 0.0001$. Sin reordenar el sistema. Aunque det $A \neq 0$, el sistema no es estrictamente dominante diagonalmente; por lo que no se puede garantizar que el método vaya ha converger.
Antes de empezar a aplicar el método se debe verificar que $\rho(T_g) < 1$.

Para verificar esto se empieza obteniendo la matriz T_g; con la ecuación (3.71):

$$\mathbf{T_g} = \left(\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{D} - \underbrace{\begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ -2 & -2 & 0 \end{pmatrix}}_{L} \right)^{-1} \underbrace{\begin{pmatrix} 0 & -2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix}}_{U}$$

$$\mathbf{T_g} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 2 & 2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & -2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 0 & -2 & 2 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -2 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 2 \end{pmatrix}$$

$$\Rightarrow \mathbf{T_g} = \begin{pmatrix} 0 & -2 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 2 \end{pmatrix}$$

Después se obtiene el polinomio característico $\rho(\lambda)$ de T_g con la definición 3.4.3:

$$\rho(\lambda) = \det(T_g - \lambda I)$$

$$\det\left(\begin{pmatrix} 0 & -2 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 2 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = \det\begin{pmatrix} -\lambda & -2 & 2 \\ 0 & 2-\lambda & -3 \\ 0 & 0 & 2-\lambda \end{pmatrix}$$

Por la definición 3.1.1, $\det\begin{pmatrix} -\lambda & -2 & 2 \\ 0 & 2-\lambda & -3 \\ 0 & 0 & 2-\lambda \end{pmatrix}$ es:

$$-\lambda\begin{vmatrix} 2-\lambda & -3 \\ 0 & 2-\lambda \end{vmatrix} - (-2)\begin{vmatrix} 0 & -3 \\ 0 & 2-\lambda \end{vmatrix} + 2\begin{vmatrix} 0 & 2-\lambda \\ 0 & 0 \end{vmatrix} =$$

$$\begin{aligned} -\lambda((2-\lambda\times 2-\lambda)-(-3\times 0))+2((0\times 2-\lambda)-(-3\times 0))+2((0\times 0)-(2-\lambda\times 0)) &= \\ -\lambda(\lambda^2-4\lambda+4)+2(0)-2(0) &= \\ -\lambda^3+4\lambda^2-4 & \end{aligned}$$

$$\Rightarrow \rho(\lambda) = -\lambda^3 + 4\lambda^2 - 4$$

Las raíces del polinomio característico $\rho(\lambda)$ son, (estas se obtuvieron con Maple):

$$\begin{aligned} \lambda_1 &= 0 \\ \lambda_2 &= 2 \\ \lambda_3 &= 2 \end{aligned}$$

Por la definición 3.4.2 se tiene que:

$$\rho(T_g) = \max\{|0|, |2|\} = 2$$

$$\Rightarrow \rho(T_g) = 2 > 1$$

Por el teorema 3.4.2 el método iterativo de Gauss-Seidel no converge a la única solución x del sistema, ya que $\rho(T_g) > 1$.

Comparando los ejemplos 3.4.4 y 3.4.6, se puede observar que el hecho de que un sistema no sea estrictamente dominante diagonalmente, no significa que los métodos iterativos vayan a ser divergentes. Habrá casos en donde uno de los métodos converge y el otro diverge. En este caso el método de Jacobi converge y el método de Gauss-Seidel no.

3.4.4 Método de sobrerrelajación sucesiva (SOR)

La relajación representa una ligera modificación al método de Gauss-Seidel y ésta permite mejorar la convergencia. Después de que se calcula cada nuevo valor de x, por medio de la ecuación (3.55), ese valor se modifica mediante un promedio ponderado de los resultados de las iteraciones anterior y actual:

$$x_i^{(k)} = wx_i^{(k)} + (1-w)x_i^{(k-1)}$$

donde w es un factor de ponderación que tiene un valor entre 0 y 2.
Sustituyendo la ecuación (3.70), en la ecuación anterior, se obtiene el **método iterativo de SOR**:

$$x_i^{(k)} = (1-w)x_i^{(k-1)} + w\left(\frac{b_i}{a_{ii}} - \sum_{j=1}^{i-1}\frac{a_{ij}x_j^{(k)}}{a_{ii}} - \sum_{j=i+1}^{n}\frac{a_{ij}x_j^{(k-1)}}{a_{ii}}\right) \text{para} i = 1,2,3\ldots n \qquad (3.72)$$

Donde:

1. Para $0 < w < 1$ el método se denomina de **sub-relajación** y se puede usar para obtener convergencia en algunos sistemas para los cuales el método de Gauss-Seidel no es convergente.
2. Para $1 < w < 2$ el método se denomina de **sobre-relajación** y se puede usar para acelerar la convergencia en algunos sistemas que son convergentes por el método de Gauss-Seidel.
3. Observe que si $w = 1$, el método se convierte en el método de Gauss-Seidel.

En general el cálculo de w es complicado y sólo en sistemas especiales (matriz definida positiva y tridiagonal) se tiene una ecuación para calcularla.
Si A es una matriz definida positiva y tridiagonal, entonces la selección óptima de w para el método de SOR es:

$$w = \frac{2}{1+\sqrt{1-[\rho(T_j)]^2}} \qquad (3.73)$$

donde $\rho(T_j)$ es el radio espectral de la matriz T_j. Ver definición 3.4.2.
Desarrollando la ecuación (3.72), se obtiene:

$$\begin{array}{llllllll}
x_1^k = & (1-w)x_1^{k-1} & & -\frac{a_{12}}{a_{11}}wx_2^{k-1} & -\frac{a_{13}}{a_{11}}wx_3^{k-1} & - \;\ldots & -\frac{a_{1n}}{a_{11}}wx_n^{k-1} & +\frac{b_1}{a_{11}}w \\
x_2^k = & (1-w)x_2^{k-1} & -\frac{a_{21}}{a_{22}}wx_1^k & & -\frac{a_{23}}{a_{22}}wx_3^{k-1} & - \;\ldots & -\frac{a_{2n}}{a_{22}}wx_n^{k-1} & +\frac{b_2}{a_{22}}w \\
x_3^k = & (1-w)x_3^{k-1} & -\frac{a_{31}}{a_{33}}wx_1^k & -\frac{a_{32}}{a_{33}}wx_2^k & & - \;\ldots & -\frac{a_{3n}}{a_{33}}wx_n^{k-1} & +\frac{b_3}{a_{33}}w \\
\vdots & \vdots & & & & \ddots & \vdots(6) & \\
x_n^k = & (1-w)x_n^{k-1} & -\frac{a_{n1}}{a_{nn}}wx_1^k & -\frac{a_{n2}}{a_{nn}}wx_2^k & -\frac{a_{n3}}{a_{nn}}wx_3^k & - \;\ldots & & +\frac{b_n}{a_{nn}}w
\end{array} \qquad (3.74)$$

Dado un sistema de ecuaciones de la forma (3.1), (donde $a_{ii} \neq 0$ para todo $i = 1,2,\ldots,n$), el método de SOR consiste básicamente en realizar los siguientes pasos:

1. Asignar un vector inicial $x^{(0)} = (x_1, x_2, x_3, \ldots, x_n)^t$ y un valor para w.
2. Sustituir el vector inicial $x^{(0)}$ y el valor de w en la primera ecuación del sistema (3.74) para obtener un nuevo valor para la primera incógnita.
3. Pasar a la segunda ecuación del sistema (3.74) y utilizar el valor calculado para la incógnita del paso 2 y los valores obtenidos en la iteración anterior para las incógnitas restantes. Este procedimiento se repite hasta obtener los nuevos valores de todas las incógnitas despejadas. Es decir siempre se utiliza el valor

más reciente de cada incógnita en todos los cálculos.

4. Continuar iterando hasta que se satisfaga alguno de los criterios de convergencia definidos en las ecuaciones (3.59), (3.60) y (3.61).

Análisis de convergencia para el método de SOR

Para analizar de la convergencia del método de SOR, se empieza multiplicando ambos lados del sistema de ecuaciones (3.74) por a_{ii}, obteniéndose:

$$\begin{array}{llllllll}
a_{11}x_1^k = (1-w)a_{11}x_1^{k-1} & & & -w\frac{a_{12}}{a_{11}}a_{11}x_2^{k-1} & -w\frac{a_{13}}{a_{11}}a_{11}x_3^{k-1} & - & \ldots -w\frac{a_{1n}}{a_{11}}a_{11}x_n^{k-1} & +w\frac{b_1}{a_{11}}a_{11} \\
a_{22}x_2^k = (1-w)a_{22}x_2^{k-1} & -w\frac{a_{21}}{a_{22}}a_{22}x_1^k & & & -w\frac{a_{23}}{a_{22}}a_{22}x_3^{k-1} & - & \ldots -w\frac{a_{2n}}{a_{22}}a_{22}x_n^{k-1} & +w\frac{b_2}{a_{22}}a_{22} \\
a_{33}x_3^k = (1-w)a_{33}x_3^{k-1} & -w\frac{a_{31}}{a_{33}}a_{33}x_1^k & -w\frac{a_{32}}{a_{33}}a_{33}x_2^k & & & - & \ldots -w\frac{a_{3n}}{a_{33}}a_{33}x_n^{k-1} & +w\frac{b_3}{a_{33}}a_{33} \\
\vdots & \vdots & & & & \ddots & \vdots & \\
a_{nn}x_n^k = (1-w)a_{nn}x_n^{k-1} & -w\frac{a_{n1}}{a_{nn}}a_{nn}x_1^k & -w\frac{a_{n2}}{a_{nn}}a_{nn}x_2^k & -w\frac{a_{n3}}{a_{nn}}a_{nn}x_3^k & & - & \ldots & +w\frac{b_n}{a_{nn}}a_{nn}
\end{array}$$

Simplificando el sistema anterior y reuniendo todos los k-ésimos términos, se obtiene:

$$\begin{array}{rllllll}
a_{11}x_1^k = & (1-w)a_{11}x_1^{k-1} & -wa_{12}x_2^{k-1} & -wa_{13}x_3^{k-1} & - & \ldots -wa_{1n}x_n^{k-1} & +wb_1 \\
wa_{21}x_1^k + a_{22}x_2^k = & & (1-w)a_{22}x_2^{k-1} & -wa_{23}x_3^{k-1} & - & \ldots -wa_{2n}x_n^{k-1} & +wb_2 \\
wa_{31}x_1^k + wa_{32}x_2^k + a_{33}x_3^k = & & & (1-w)a_{33}x_3^{k-1} & - & \ldots -wa_{3n}x_n^{k-1} & +wb_3 \\
\vdots & & & & & \ddots & \vdots \\
wa_{n1}x_1^k + wa_{n2}x_2^k + wa_{n3}x_3^k + a_{nn}x_n^k = & & & & & \ldots & +wb_n
\end{array}$$

Lo cual se puede expresar matricialmente de la siguiente forma:

$$\left(\underbrace{\begin{pmatrix} a_{11} & 0 & 0 & \ldots & 0 \\ 0 & a_{22} & 0 & \ldots & 0 \\ 0 & & a_{33} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \ldots & a_{nn} \end{pmatrix}}_{D} - w \underbrace{\begin{pmatrix} 0 & 0 & 0 & \ldots & 0 \\ -a_{21} & 0 & 0 & \ldots & 0 \\ -a_{31} & -a_{32} & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & -a_{n3} & \ldots & 0 \end{pmatrix}}_{L}\right) \underbrace{\begin{pmatrix} x_1^k \\ x_2^k \\ x_3^k \\ \vdots \\ x_n^k \end{pmatrix}}_{x^k} =$$

$$\left((1-w)\underbrace{\begin{pmatrix} a_{11} & 0 & 0 & \ldots & 0 \\ 0 & a_{22} & 0 & \ldots & 0 \\ 0 & & a_{33} & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & \ldots & a_{nn} \end{pmatrix}}_{D} + w \underbrace{\begin{pmatrix} 0 & -a_{12} & -a_{13} & \ldots & -a_{1n} \\ 0 & 0 & -a_{23} & \ldots & -a_{2n} \\ 0 & 0 & 0 & \ldots & -a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}}_{U}\right) \underbrace{\begin{pmatrix} x_1^{k-1} \\ x_2^{k-1} \\ x_3^{k-1} \\ \vdots \\ x_n^{k-1} \end{pmatrix}}_{x^{k-1}} + w \underbrace{\begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_n \end{pmatrix}}_{b}$$

Con la definiciones de D, L y U para el análisis de convergencia del método de Jacobi, se obtiene:

$$(D - wL)x^{(k)} = [(1-w)D + wU]x^{(k-1)} + wb$$

Despejando $x^{(k)}$ de la ecuación anterior, se deduce que la fórmula vectorial de iteración del método de SOR es:

$$x^{(k)} = (D - wL)^{-1}[(1-w)D + wU]x^{(k-1)} + w(D - wL)^{-1}b \qquad \text{para}\, k = 1, 2, 3, \ldots$$

Donde

$$T_w = (D - wL)^{-1}[(1-w)D + wU] \qquad \textit{matriz de iteración del método de SOR} \qquad (3.75)$$

Aplicando el teorema 3.4.2 se tiene que: el método iterativo de SOR converge a la única solución x del sistema $x = T_w\, x + c$, si y sólo si $\rho(T_w) < 1$.

Aplicando la definición 3.4.2 se tiene que: $\rho(T_w) = \max|\lambda|$, donde λ es un valor característico de T_w.

Ejemplo 3.4.7 Resolver el siguiente sistema de ecuaciones lineales por el método de SOR con una tolerancia de $\varepsilon = 0.0001$; (det $A = 950$, obtenido con Maple).

$$\begin{array}{rcrcrcl} 10x_1 & - & x_2 & & & = & 9 \\ -\ x_1 & + & 10x_2 & - & 2x_3 & = & 7 \\ & - & 2x_2 & + & 10x_3 & = & 6 \end{array} \tag{3.76}$$

Como det $A \neq 0$ y el sistema es estrictamente dominante diagonalmente, se empieza a aplicar el método; ya que se puede garantizar que el método va ha converger, (teorema 3.4.1).

Como no se conoce una buena aproximación inicial, se iguala el vector inicial con ceros;

$$x^{(0)} = \begin{pmatrix} 0.000000 \\ 0.000000 \\ 0.000000 \end{pmatrix} = \begin{pmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{pmatrix}$$

Al no tener un valor para w; como la matriz es definida positiva y tridiagonal, para obtenerlo empieza calculando la matriz T_j; con la ecuación (3.62):

$$\mathbf{T_j} = \underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1}}_{D} \left(\underbrace{\begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 2 & 0 \end{pmatrix}}_{L} + \underbrace{\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}}_{U} \right)$$

$$\mathbf{T_j} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}^{-1} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{10} & 0 & 0 \\ 0 & \frac{1}{10} & 0 \\ 0 & 0 & \frac{1}{10} \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 2 \\ 0 & 2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{10} & 0 \\ \frac{1}{10} & 0 & \frac{1}{5} \\ 0 & \frac{1}{5} & 0 \end{pmatrix}$$

$$\Rightarrow \mathbf{T_j} = \begin{pmatrix} 0 & \frac{1}{10} & 0 \\ \frac{1}{10} & 0 & \frac{1}{5} \\ 0 & \frac{1}{5} & 0 \end{pmatrix}$$

Después se obtiene el polinomio característico $\rho(\lambda)$ de T_j con la definición 3.4.3:

$$\rho(\lambda) = \det(T_j - \lambda I)$$

$$\det\left(\begin{pmatrix} 0 & \frac{1}{10} & 0 \\ \frac{1}{10} & 0 & \frac{1}{5} \\ 0 & \frac{1}{5} & 0 \end{pmatrix} - \lambda \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right) = \det\begin{pmatrix} -\lambda & \frac{1}{10} & 0 \\ \frac{1}{10} & -\lambda & \frac{1}{5} \\ 0 & \frac{1}{5} & -\lambda \end{pmatrix}$$

Por la definición 3.1.1, $\det\begin{pmatrix} -\lambda & \frac{1}{10} & 0 \\ \frac{1}{10} & -\lambda & \frac{1}{5} \\ 0 & \frac{1}{5} & -\lambda \end{pmatrix}$ es:

$$-\lambda \begin{vmatrix} -\lambda & \frac{1}{5} \\ \frac{1}{5} & -\lambda \end{vmatrix} - \frac{1}{10}\begin{vmatrix} \frac{1}{10} & \frac{1}{5} \\ 0 & -\lambda \end{vmatrix} + 0 \begin{vmatrix} \frac{1}{10} & -\lambda \\ 0 & \frac{1}{5} \end{vmatrix} =$$

$$\begin{aligned} -\lambda((-\lambda\times-\lambda)-(\frac{1}{5}\times\frac{1}{5}))+\frac{1}{10}((\frac{1}{10}\times-\lambda)-(\frac{1}{5}\times 0))+0 \\ -\lambda(\lambda^2-\frac{1}{10})+\frac{1}{10}(-\frac{1}{10}\lambda-0))+0 &= \\ -\lambda^3+\frac{1}{10}\lambda-\frac{1}{20}\lambda &= -\lambda^3+\frac{1}{20}\lambda \end{aligned}$$

$$\Rightarrow \rho(\lambda) = -\lambda^3 + \frac{1}{20}\lambda$$

Las raíces del polinomio característico $\rho(\lambda)$ son, (estas se obtuvieron con Maple):

$$\begin{aligned} \lambda_1 &\approx 0.000000 \\ \lambda_2 &\approx -0.223607 \\ \lambda_3 &\approx 0.223607 \end{aligned}$$

Por la definición 3.4.2 se tiene que:

$$\rho(T_j) = \max\{|0|,|0.223607|\} = 0.223607$$

Luego se obtiene el valor de w con la ecuación (3.73):

$$\begin{aligned} w &= \frac{2}{1+\sqrt{1-[\rho(T_j)]^2}} = \frac{2}{1+\sqrt{1-[0.223607]^2}} \\ &= \frac{2}{1+\sqrt{0.974679}} = \frac{2}{1.974679} \approx 1.012823 \\ &\Rightarrow w \approx 1.013 \end{aligned}$$

Después se lleva el sistema (3.76) a la forma (3.74), obteniéndose:

$$
\begin{array}{lcl}
x_1^k &=& (1-w)x_1^{k-1} - w(-\frac{1}{10})x_2^{k-1} + w\frac{9}{10}\\
x_2^k &=& (1-w)x_2^{k-1} - w(-\frac{1}{10})x_1^{k} - w(-\frac{2}{10})x_3^{k-1} + w\frac{7}{10}\\
x_3^k &=& (1-w)x_3^{k-1} - w(-\frac{2}{10})x_2^{k} + \quad + w\frac{6}{10}
\end{array}
$$

Sustituyendo el valor de $w = 1.013$ y realizando las operaciones en el sistema anterior, se obtiene:

$$
\begin{array}{lcl}
x_1^k &=& -0.013000x_1^{k-1} + 0.101300x_2^{k-1} + 0.911700\\
x_2^k &=& -0.013000x_2^{k-1} + 0.101300x_1^{k} + 0.202600x_3^{k-1} + 0.709100\\
x_3^k &=& -0.013000x_3^{k-1} + 0.202600x_2^{k} + \quad + 0.607800
\end{array} \tag{3.77}
$$

Después se sustituye $x^{(0)}$ en la primera ecuación del sistema (3.77) para obtener un nuevo valor $x_1^{(1)}$. Para la segunda ecuación del sistema (3.77) se utilizar el valor calculado $x_1^{(1)}$ y los valores previos para las incógnitas restantes, obteniendo $x_2^{(1)}$. Finalmente para la tercera ecuación del sistema (3.77) se utilizan los valores calculados $x_1^{(1)}$ y $x_2^{(1)}$ y los valores previos para las incógnitas restantes, obteniendo $x_3^{(1)}$. Es decir siempre se utiliza el valor más reciente de cada incógnita en todos los cálculos; como en el método de Gauss-Seidel; obteniéndose:

$$
\begin{array}{lcl}
x_1^{(1)} &=& -0.013000(0.000000) + 0.101300(0.000000) + 0.911700\\
x_2^{(1)} &=& -0.013000(0.000000) + 0.101300(0.911700) + 0.202600(0.000000) + 0.709100\\
x_3^{(1)} &=& -0.013000(0.000000) + 0.202600(0.801455) + \quad + 0.607800
\end{array}
$$

$$
\begin{array}{rcl}
x_1^{(1)} &=& 0.911700\\
\Rightarrow x_2^{(1)} &=& 0.801455\\
x_3^{(1)} &=& 0.770200
\end{array}
$$

Se sigue el procedimiento anterior para obtener $x^{(2)}$:

$$
\begin{array}{lcl}
x_1^{(2)} &=& -0.013000(0.911700) + 0.101300(0.801455) + 0.911700\\
x_2^{(2)} &=& -0.013000(0.801455) + 0.101300(0.981035) + 0.202600(0.770200) + 0.709100\\
x_3^{(2)} &=& -0.013000(0.770200) + 0.202600(0.954097) + \quad + 0.607800
\end{array}
$$

$$
\begin{array}{rcl}
x_1^{(2)} &=& 0.981035\\
\Rightarrow x_2^{(2)} &=& 0.954097\\
x_3^{(2)} &=& 0.791100
\end{array}
$$

Escogiendo la ecuación (3.60) (*error relativo*) para saber cuando dejar de iterar, se debe se cumplir que $\frac{\left\|X^{(k)}-X^{(k-1)}\right\|}{\left\|X^{(k)}\right\|} \leq 0.0001$; esto se cumple en la iteración 5, $(0.000005 \leq 0.0001)$.

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	ε
0	0.000000	0.000000	0.000000	——
1	0.911700	0.801455	0.770200	1.000000
2	0.981035	0.954097	0.791100	0.155593
3	0.995597	0.957825	0.791600	0.014626
4	0.995785	0.957894	0.791600	0.000189
5	0.995789	0.957895	0.791600	0.000005

Cuadro 3.6: Ejemplo del método de SOR

Para obtener $x^{(3)}$, $x^{(4)}$ y $x^{(5)}$; se repite el proceso anterior. Las iteraciones se muestran el cuadro 3.6.

A continuación se indica como calcular la ecuación (3.60) con $\|\cdot\|_\infty$.

Para la primera iteración el error relativo se calcula de la siguiente manera:

$$\frac{\left\|X^{(1)}-X^{(0)}\right\|_\infty}{\left\|X^{(1)}\right\|_\infty} = \frac{0.911700}{0.911700} = 1.000000$$

Operaciones:

$$\left\|X^{(1)}-X^{(0)}\right\|_\infty = \max\begin{Bmatrix} |\,0.911700-0.000000\,|, \\ |\,0.801455-0.000000\,|, \\ |\,0.770200-0.000000\,| \end{Bmatrix} = \begin{Bmatrix} 0.911700, \\ 0.801455, \\ 0.770200 \end{Bmatrix} = 0.911700$$

$$\left\|X^{(1)}\right\|_\infty = \max\begin{Bmatrix} |\,0.911700\,|, \\ |\,0.801455\,|, \\ |\,0.770200\,| \end{Bmatrix} = \begin{Bmatrix} 0.500000, \\ 1.333333, \\ 0.000000 \end{Bmatrix} = 0.911700$$

Para la segunda iteración el error relativo se calcula:

$$\frac{\left\|X^{(2)}-X^{(1)}\right\|_\infty}{\left\|X^{(2)}\right\|_\infty} = \frac{0.152642}{0.981035} = 0.155593$$

Operaciones:

$$\left\|X^{(2)}-X^{(1)}\right\|_\infty = \max\begin{Bmatrix} |\,0.981035-0.911700\,|, \\ |\,0.954097-0.801455\,|, \\ |\,0.791100-0.770200\,| \end{Bmatrix} = \begin{Bmatrix} 0.069335, \\ 0.152642, \\ 0.020913 \end{Bmatrix} = 0.152642$$

$$\left\|X^{(2)}\right\|_\infty = \max\begin{Bmatrix} |\,0.981035\,|, \\ |\,0.954097\,|, \\ |\,0.791100\,| \end{Bmatrix} = \begin{Bmatrix} 0.981035, \\ 0.954097, \\ 0.791100 \end{Bmatrix} = 0.981035$$

Para obtener el error relativo de las demás iteraciones se repite el proceso anterior. Los errores se muestran en el cuadro 3.6.

$$\begin{aligned} x_1 &\approx 0.995789 \\ x_2 &\approx 0.957895 \qquad \text{Conun } \varepsilon = 0.0001 \\ x_3 &\approx 0.791600 \end{aligned}$$

Comprobación:

$$\begin{array}{rrrrrrrrr} & 10(0.995789) & - & (0.957895) & & & = & 8.999995 & \approx & 9 \\ - & (0.995789) & + & 10(0.957895) & - & 2(0.791600) & = & 6.999961 & \approx & 7 \\ & & - & 2(0.957895) & + & 10(0.791600) & = & 6.000210 & \approx & 6 \end{array}$$

3.5 Aplicaciones y ejercicios

Ejercicio 3.1 Una bióloga ha colocado tres cepas bacterianas (denotadas como I, II y III) en un tubo de ensayo, donde serán alimentadas con tres diferentes fuentes alimenticias (A, B y C). Cada día 2300 unidades de A, 800 de B y 1500 de C, se colocan en un tubo de ensayo, y cada bacteria consume cierto número de unidades de cada alimento por día, como se muestra en el cuadro 3.7. ¿Cuántas bacterias de cada cepa pueden coexistir en un tubo de ensayo y consumir todo el alimento?. (Resolver por Gauss-Jordan).

	Cepa bacteriana I	Cepa bacteriana II	Cepa bacteriana III
Alimento A	2	2	4
Alimento B	1	2	0
Alimento C	1	3	1

Cuadro 3.7: Cepas bacterianas y alimento que consumen

Solución:

Sean x_1, x_2 y x_3, los números de bacterias de las cepas I, II y III, respectivamente. Puesto que cada una de las bacterias de x_1 de la cepa I consume 2 unidades de A por día, la cepa I consume un total de $2x_1$ unidades por día. De manera similar las cepas II y III consumen un total de $2x_2$ y $4x_3$ unidades de alimento A diariamente. Puesto que se quiere acabar con todas las 2300 unidades de A, se obtiene la siguiente ecuación:

$$2x_1 + 2x_2 + 4x_3 = 2300$$

Del mismo modo de obtienen las ecuaciones correspondientes al consumo de B y C.

$$\begin{array}{ccccccc} x_1 & + & 2x_2 & & & = & 800 \\ x_1 & + & 3x_2 & + & x_3 & = & 1500 \end{array}$$

Así se tiene un sistema de tres ecuaciones lineales con tres variables. (det $A = 6$). Resolviendo por el método de Gauss- Jordan, se obtiene:

$$\left(\begin{array}{ccc|c} 2 & 2 & 4 & 2300 \\ 1 & 2 & 0 & 800 \\ 1 & 3 & 1 & 1500 \end{array}\right) \rightarrow \left(\begin{array}{ccc|c} 1 & 0 & 0 & 100 \\ 0 & 1 & 0 & 350 \\ 0 & 0 & 1 & 350 \end{array}\right)$$

$x_1 = 100$, $x_2 = 350$ y $x_3 = 350$; es decir la bióloga debería colocar 100 bacterias de la cepa I, 350 de cada una de las cepas II y III en el tubo de ensayo si ella desea que todo el alimento sea consumido.

Ejercicio 3.2 Aplicación en criptografía.
Un proceso para encriptar un mensaje secreto es usar cierta matriz cuadrada cuyos elementos son enteros con elementos enteros de la inversa. Se recibe un mensaje, se asigna un número a cada letra (por ejemplo, A=1, B=2, etc., y espacio=27), se arreglan los números en una matriz de izquierda a derecha en cada renglón , donde el número de elementos en el renglón es igual al tamaño de la matriz de código, se multiplica esta matriz por la matriz de código por la derecha, se transcribe el mensaje a una cadena de números (que se lee de izquierda a derecha a lo largo de cada renglón), y se manda el mensaje encriptado.

La persona que debe recibir el mensaje conoce la matriz de código. Arregla el mensaje encriptado en una matriz B de izquierda a derecha en cada renglón, en donde el número de elementos en un renglón coincide con el tamaño de la matriz de código, multiplica la matriz B por la derecha por la inversa de la matriz de código y puede leer el mensaje decodificado (de izquierda a derecha por cada renglón).

Se ha recibido el siguiente mensaje que fue encriptado usando la siguiente matriz de código A. Decodifíquelo usando inversión de matrices particionadas. (Suponga que A=1, B=2, etc., y espacio=27).

Mensaje: 47, 49, -19, 257, 487, 10, -9, 63, 137, 236, 79, 142, -184, 372, 536, 59, 70, -40, 332, 588.

Nota: El primer renglón de la matriz B con el mensaje encriptado que se construye es 47, 49, -19, 257, 487, ya que el número de elementos en un renglón debe de coincidir con el tamaño de la matriz de código $(n=5)$.

Solución:

A=1	E=5	I=9	M=13	Q=17	U=21	Y=25
B=2	F=6	J=10	N=14	R=18	V=22	Z=26
C=3	G=7	K=11	O=15	S=19	W=23	espacio=27
D=4	H=8	L=12	P=16	T=20	X=24	

Cuadro 3.8: Asignación de numeros a cada letra

La matriz B, esta formada por:

$$\mathbf{B}=\begin{pmatrix} 47 & 49 & -19 & 257 & 487 \\ 10 & -9 & 63 & 137 & 236 \\ 79 & 142 & -184 & 372 & 536 \\ 59 & 70 & -40 & 332 & 588 \end{pmatrix}$$

Para poder decodificar el mensaje se multiplica la matriz B por la derecha por la inversa de la matriz de código A.

La inversa de la matriz A que se obtiene de aplicar inversión de matrices particionadas es; (det $A=1$):

$$\mathbf{A}^{-1} = \left(\begin{array}{ccc|cc} 14 & 1 & -2 & 1 & -5 \\ 22 & -7 & -4 & 6 & -19 \\ 13 & -3 & -2 & 3 & -10 \\ \hline -2 & 1 & 1 & -1 & 2 \\ -2 & 0 & 0 & 0 & 1 \end{array}\right)$$

Finalmente la matriz decodificada es:

$$\mathbf{BA}^{-1} = \begin{pmatrix} 47 & 49 & -19 & 257 & 487 \\ 10 & -9 & 63 & 137 & 236 \\ 79 & 142 & -184 & 372 & 536 \\ 59 & 70 & -40 & 332 & 588 \end{pmatrix} \begin{pmatrix} 14 & 1 & -2 & 1 & -5 \\ 22 & -7 & -4 & 6 & -19 \\ 13 & -3 & -2 & 3 & -10 \\ -2 & 1 & 1 & -1 & 2 \\ -2 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{BA}^{-1} = \begin{pmatrix} 1 & 18 & 5 & 27 & 25 \\ 15 & 21 & 27 & 8 & 1 \\ 22 & 9 & 14 & 7 & 27 \\ 6 & 21 & 14 & 27 & 27 \end{pmatrix}$$

Para leer el mensaje decodificado se forma una cadena de números de izquierda a derecha con los elementos de la matriz $\mathbf{BA}^{-1}$.***Mensaje decodificado:*** 1, 18, 5, 27, 25, 15, 21, 27, 8, 1, 22, 9, 14, 7, 27, 6, 21, 14, 27, 27.

Utilizando el cuadro (3.8), el mensaje decodificado dice: *ARE YOU HAVING FUN*.

Ejercicio 3.3 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss-Jordan particionado, realizando 9 particiones.

$$
\begin{array}{rcrcrcrcrcrcrcrcrcrcr}
x_1 & + & 2x_2 & + & 3x_3 & + & 4x_4 & + & 5x_5 & + & 6x_6 & + & 7x_7 & + & 8x_8 & + & 9x_9 & + & 10x_{10} & = & 385 \\
2x_1 & + & x_2 & & & & & & & & & & & & & & & & & = & 4 \\
3x_1 & & & + & x_3 & & & & & & & & & & & & & & & = & 6 \\
4x_1 & & & & & + & x_4 & & & & & & & & & & & & & = & 8 \\
5x_1 & & & & & & & + & x_5 & & & & & & & & & & & = & 10 \\
6x_1 & & & & & & & & & + & x_6 & & & & & & & & & = & 12 \\
7x_1 & & & & & & & & & & & + & x_7 & & & & & & & = & 14 \\
8x_1 & & & & & & & & & & & & & + & x_8 & & & & & = & 16 \\
9x_1 & & & & & & & & & & & & & & & + & x_9 & & & = & 18 \\
10x_1 & & & & & & & & & & & & & & & & & + & x_{10} & = & 20
\end{array}
$$

Solución:
det $A = -383$.

$$
[\mathbf{A}|\mathbf{b}] = \left(\begin{array}{ccc|cccc|ccc|c}
1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 385 \\
2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\
3 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 6 \\
\hline
4 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 8 \\
5 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 10 \\
6 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 12 \\
7 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 14 \\
\hline
8 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 16 \\
9 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 18 \\
10 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 20
\end{array}\right) \rightarrow
$$

$$
\left(\begin{array}{ccc|cccc|ccc|c}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \\
\hline
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 4 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 5 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 6 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 7 \\
\hline
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 8 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 9 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 10
\end{array}\right)
$$

$x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$, $x_5 = 5$, $x_6 = 6$, $x_7 = 7$, $x_8 = 8$, $x_9 = 9$ y $x_{10} = 10$.

Ejercicio 3.4. Resolver el siguiente sistema de ecuaciones lineales por el método de intercambio.

$$\begin{array}{ccccccccc} 5x_1 & + & 7x_2 & + & 6x_3 & + & 5x_4 & = & 23 \\ 7x_1 & + & 10x_2 & + & 8x_3 & + & 7x_4 & = & 32 \\ 6x_1 & + & 8x_2 & + & 10x_3 & + & 9x_4 & = & 33 \\ 5x_1 & + & 7x_2 & + & 9x_3 & + & 10x_4 & = & 31 \end{array}$$

Solución:

det $A = 1$.

$$\begin{array}{c|cccc} & x_1 & x_2 & x_3 & x_4 \\ \hline b_1 & 5 & 7 & 6 & 5 \\ b_2 & 7 & 10 & 8 & 7 \\ b_3 & 6 & 8 & 10 & 9 \\ b_4 & 5 & 7 & 9 & 10 \end{array} \rightarrow \begin{array}{c|cccc} & b_1 & b_2 & b_3 & b_4 \\ \hline x_1 & 68 & -41 & -17 & 10 \\ x_2 & -41 & 25 & 10 & -6 \\ x_3 & -17 & 10 & 5 & -3 \\ x_4 & 10 & -6 & -3 & 2 \end{array}$$

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b} = \begin{pmatrix} 68 & -41 & -17 & 10 \\ -41 & 25 & 10 & -6 \\ -17 & 10 & 5 & -3 \\ 10 & -6 & -3 & 2 \end{pmatrix} \begin{pmatrix} 23 \\ 32 \\ 33 \\ 31 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

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

Ejercicio 3.5 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss, utilizando pivoteo parcial.

$$\begin{array}{rcrcrcr} 0.2641x_1 & + & 0.1735x_2 & + & 0.8642x_3 & = & -0.7521 \\ -\ 0.8641x_1 & - & 0.4243x_2 & + & 0.0711x_3 & = & 0.2501 \\ 0.9411x_1 & + & 0.0175x_2 & + & 0.1463x_3 & = & 0.6310 \end{array}$$

Solución:

det $A \approx 0.348835$.

$$\left(\begin{array}{rrr|r} 0.2641 & 0.1735 & 0.8642 & -0.7521 \\ -0.8641 & -0.4243 & 0.0711 & 0.2501 \\ 0.9411 & 0.0175 & 0.1463 & 0.6310 \end{array}\right) \to \left(\begin{array}{rrr|r} 0.9411 & 0.0175 & 0.1463 & 0.6310 \\ 0 & -0.408232 & 0.205430 & 0.829472 \\ 0 & 0 & 0.907981 & -0.586627 \end{array}\right)$$

Realizando la sustitución hacia atrás; $x_1 = 0.814757$, $x_2 = -2.35698$ y $x_3 = -0.646078$.

Ejercicio 3.6 Supóngase que se calienta cada borde de una placa metálica a una temperatura constante, como se muestra en la figura 3.5

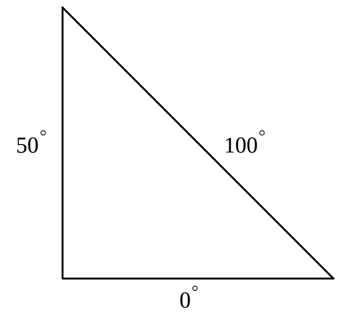


Figura 3.5 Una placa de metal caliente

Finalmente, la temperatura en los puntos interiores alcanzará el equilibrio donde puede mostrarse que siguiente propiedad se mantendrá:

1. La temperatura en cada punto interior P sobre una placa es el promedio de las temperaturas sobre la circunferencia de cualquier círculo centrado en P dentro de la placa (figura 3.6).

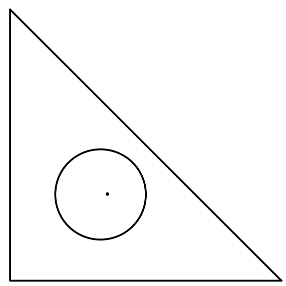


Figure 3.6 Temperatura en cada punto interior P

Para aplicar esta propiedad en un ejemplo real se requiere de técnicas de cálculo. Como una alternativa, se puede aproximar la situación sobreponiendo una cuadrícula o malla sobre la placa, que tenga un número finito de puntos interiores, como se muestra en la figura 3.7.

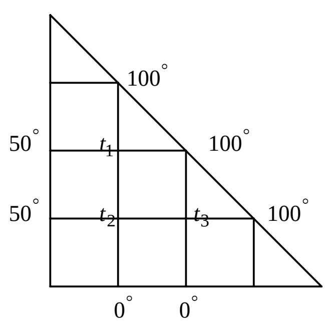


Figure 3. 7 Versión discreta del problema de la placa calentada

El homólogo discreto de la propiedad de promediación que gobierna las temperaturas en equilibrio se establece como sigue: *La temperatura de cada punto interior P es el promedio de las temperaturas en los puntos adyacentes a P*.

Usar el método de Jacobi para obtener las temperaturas de equilibrio del ejemplo que se muestra en la figura 3, donde se tienen tres puntos interiores, cada uno de los cuales es adyacente a otros cuatro puntos; con una tolerancia de 10^{-4}.

Solución:

Sean las temperaturas en equilibrio de los puntos interiores t_1, t_2 y t_3, como se muestra. Entonces, por la propiedad de promediación de temperatura, se tiene que:

$$t_1 = \frac{100+100+t_2+50}{4}$$

$$t_2 = \frac{t_1+t_3+0+50}{4}$$

$$t_3 = \frac{100+100+0+t_2}{4}$$

$$\begin{array}{rcrcrcl} 4t_1 & - & t_2 & & & = & 250 \\ -t_1 & + & 4t_2 & - & t_3 & = & 50 \\ & - & t_2 & + & 4t_3 & = & 200 \end{array}$$

$\det A = 56$, obtenido con Maple, y el sistema es $E.D.D.$

Las temperaturas de equilibrio en los puntos interiores son:

$$\begin{array}{lcl} t_1 & \approx & 74.105072 \\ t_2 & \approx & 46.427155 \qquad \text{Conun } \varepsilon = 0.0001 \\ t_3 & \approx & 61.605072 \end{array}$$

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	ε
0	0.000000	0.000000	0.000000	——
1	62.500000	12.500000	50.000000	1.000000
2	65.625000	40.625000	53.125000	0.428571
3	72.656250	42.187500	60.156250	0.096774
4	73.046875	45.703125	60.546875	0.048128
5	73.925781	45.898438	61.425781	0.011889
6	73.974610	46.337891	61.474610	0.005941
7	74.084473	46.362305	61.584473	0.001483
8	74.090576	46.417237	61.590576	0.000741
9	74.104309	46.420288	61.604309	0.000185
10	74.105072	46.427155	61.605072	0.000093

Cuadro 3.9: Ejercicio del método de Jacobi

Ejercicio 3.7 Determinar si el siguiente sistema de ecuaciones lineales converge al resolverlo con el método de Jacobi.

$$\begin{array}{rcrcrcrcrcr}
 & & 2x_1 & - & x_2 & & & & & = & 3 \\
- & & x_1 & + & 2x_2 & - & x_3 & & & = & -2 \\
 & & & - & x_2 & + & 2x_3 & - & x_4 & = & 2 \\
 & & & & & - & x_3 & + & 2x_4 \; - \; x_5 & = & -2 \\
 & & & & & & & - & x_4 \; + \; 2x_5 & = & 1
\end{array}$$

Solución:

det $A = 6$ y la matriz no es $E.D.D$.

$$\begin{array}{rclcl}
x_1 & = & \frac{1}{2}x_2 & + & \frac{3}{2} \\
x_2 & = & \frac{1}{2}x_1 + \frac{1}{2}x_3 & - & \frac{2}{2} \\
x_3 & = & \frac{1}{2}x_2 + \frac{1}{2}x_4 & + & \frac{2}{2} \\
x_4 & = & \frac{1}{2}x_3 + \frac{1}{2}x_5 & - & \frac{2}{2} \\
x_5 & = & \frac{1}{2}x_4 & + & \frac{1}{2}
\end{array}$$

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

$$\rho(\lambda) = -\lambda^5 + \lambda^3 - \frac{3}{16}\lambda$$

Las raíces del polinomio característico $\rho(\lambda)$ son:

$$\begin{array}{ccc} \lambda_1 & = & 0 \\ \lambda_2 & = & \frac{1}{2} \\ \lambda_3 & = & -\frac{1}{2} \\ \lambda_4 & = & \frac{1}{2}\sqrt{3} \\ \lambda_5 & = & -\frac{1}{2}\sqrt{3} \end{array}$$

$$\rho(T_j) \approx \max\left\{|0|, \left|\frac{1}{2}\right|, \left|-\frac{1}{2}\right|, \left|\frac{1}{2}\sqrt{3}\right|, \left|-\frac{1}{2}\sqrt{3}\right|\right\} = \frac{1}{2}\sqrt{3}$$

$$\Rightarrow \rho(T_j) \approx .866025 < 1$$

Por el teorema 3.4.2 el método iterativo de Jacobi converge a la única solución x del sistema, ya que $\rho(T_j) < 1$.

Ejercicio 3.8 Una delgada tira de papel de 1 unidad de largo se coloca a lo largo de una recta numérica de modo que sus extremos están en 0 y 1. El papel se dobla a la mitad, con su extremo derecho hacia la izquierda, de tal modo que sus extremos se encuentran ahora en 0 y $\frac{1}{2}$. Acto seguido, se dobla nuevamente a la mitad, esta vez con el extremo izquierdo hacia la derecha, de tal suerte que sus extremos se encuentran en $\frac{1}{4}$ y $\frac{1}{2}$. La figura 1 muestra este proceso. Se continua doblando el papel a la mitad, alternando la derecha hacia la izquierda e izquierda hacia la derecha. Si se pudiera continuar indefinidamente, es claro que los extremos del papel convergirían a un punto. Este punto es el que se desea encontrar.

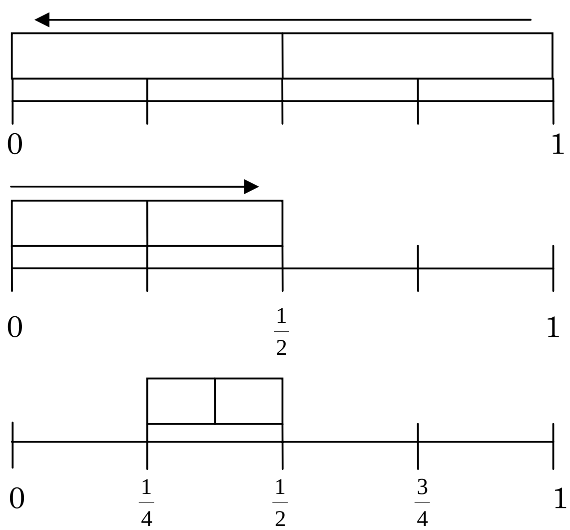


Figure 3.8 Doblado de una tira de papel

1. Si x_1 corresponde al extremo izquierdo del papel y x_2 corresponde al extremo derecho. Realizar una tabla con los primeros seis valores de $[x_1, x_2]$.

2. Encuentre dos ecuaciones lineales de la forma $x_2 = ax_1 + b$ y $x_1 = cx_2 + d$ que determinen los nuevos valores de los extremos de cada iteración.

3. Usar el método de Gauss-Seidel para aproximar el punto en el cual los extremos del papel lleguen a converger; con una tolerancia de 10^{-3}.

4. Resolver el sistema de ecuaciones obtenido en 2 con un método exacto y comparar resultados.

Solución:

1.

n	0	1	2	3	4	5	6
x_1	0	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{5}{16}$	$\frac{5}{16}$	$\frac{21}{64}$
x_2	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{3}{8}$	$\frac{3}{4}$	$\frac{11}{32}$	$\frac{11}{32}$

2. $2x_1 + x_2 = 1$ y $x_1 + 2x_2 = 1$

3. El sistema obtenido en el punto anterior es $E.D.D.$; ver cuadro 3.10 (det $A = 3$, obtenido con Maple).

n	0	1	2	3	4	5	6	7
x_1	0	0.0000	0.2500	0.3125	0.3281	0.3320	0.3330	0.3332
x_2	1	0.5000	0.3750	0.3438	0.3360	0.3340	0.3335	0.3334

Cuadro 3.10: Ejercicio del método de Gauss-Seidel

4. $x_1 = x_2 = \frac{1}{3}$, las iteraciones del punto 3 convergen hacia $x_1 = x_2 = 0.3333$, mismo resultado que se obtiene al usar un método exacto.

Ejercicio 3.9 Resolver el siguiente sistema de ecuaciones lineales por el método de Gauss-Seidel, con 6 decimales y una tolerancia de $\varepsilon = 0.0001$. Sin reordenar el sistema.

$$\begin{array}{rcrcrcr} x_1 & - & 3x_2 & + & 5x_3 & = & 5 \\ 8x_1 & - & x_2 & - & x_3 & = & 8 \\ -\ 2x_1 & + & 4x_2 & + & x_3 & = & 4 \end{array}$$

Solución:

det $A = 171$ y la matriz no es $E.D.D$.

$$\mathbf{T_g} = \begin{pmatrix} 0 & 3 & -5 \\ 0 & 24 & -41 \\ 0 & -90 & 154 \end{pmatrix}$$

$$\rho(\lambda) = -\lambda^3 + 178\lambda^2 - 6\lambda$$

Las raíces del polinomio característico $\rho(\lambda)$ son:

$$\begin{array}{lcr} \lambda_1 & \approx & 0.000000 \\ \lambda_2 & \approx & 0.033714 \\ \lambda_3 & \approx & 177.966286 \end{array}$$

$$\rho(T_g) \approx \max\{|0.000000|, |0.033714|, |177.966286|\} = 177.966286$$

$$\Rightarrow \rho(T_g) \approx 177.966286 > 1$$

Por el teorema 3.4.2 el método iterativo de Gauss-Seidel no converge a la única solución x del sistema, ya que $\rho(T_g) > 1$.

Ejercicio 3.10 Resolver el siguiente sistema de ecuaciones lineales por el método de SOR, con 6 decimales y una tolerancia de $\varepsilon = 0.0001$.

$$\begin{array}{rcrcrcrcr} 10x_1 & - & x_2 & + & 2x_3 & & & = & 6 \\ -x_1 & + & 11x_2 & - & x_3 & + & 3x_4 & = & 25 \\ 2x_1 & - & x_2 & + & 10x_3 & - & x_4 & = & -11 \\ & & 3x_2 & - & x_3 & + & 8x_4 & = & 15 \end{array}$$

Solución:
det $A = 7395$ y la matriz es $E.D.D$.

k	$x_1^{(k)}$	$x_2^{(k)}$	$x_3^{(k)}$	$x_4^{(k)}$	ε
0	0.000000	0.000000	0.000000	0.000000	——
1	0.720000	2.805818	-1.156102	0.813967	1.000000
2	1.190163	1.903436	-1.048330	1.073411	0.474081
3	0.961979	1.985867	-0.974096	0.995563	0.114904
4	0.999691	2.007071	-1.004791	0.996987	0.018790
5	1.002060	1.999274	-0.999985	1.000932	0.003900
6	0.999497	1.999787	-0.999796	0.999940	0.001282
7	1.000026	2.000087	-1.000044	0.999966	0.000264
8	1.000016	1.999991	-1.000000	1.000011	0.000048

Cuadro 3.11: Ejercicio del método de SOR

$x_1 \approx 1.000016$, $x_2 \approx 1.999991$, $x_3 \approx -1.000000$ y $x_4 \approx 1.000011$.

Capítulo 4

Factorización LU y sus aplicaciones

Si es posible factorizar la matriz A de un sistema de ecuaciones lineales como el producto de una matriz triangular inferior L con una matriz triangular superior U: $A = LU$;

$$\underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}}_{A} = \underbrace{\begin{pmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ l_{31} & l_{32} & l_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{nn} \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ 0 & u_{22} & u_{23} & \dots & u_{2n} \\ 0 & 0 & u_{33} & \dots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & u_{nn} \end{pmatrix}}_{U}$$

entonces la ecuación $Ax = b$, se transforma en:

$$(LU)x = \mathrm{b}$$
$$\Rightarrow L(Ux) = \mathrm{b}$$

que puede solucionarse , resolviendo los sistemas

$$Ly = b \tag{4.1}$$
$$Ux = y \tag{4.2}$$

El sistema (4.1) se resuelve realizando una sustitución hacia adelante, obteniendo así el vector y; después se toma este vector como términos independientes para el sistema (4.2) y se resuelve realizando una sustitución hacia atrás para obtener el vector x; que es la solución del sistema original; contando con la ventaja de que la solución de estos sistemas es inmediata.

Cabe mencionar que si se puede factorizar la matriz A en la forma LU, L y U no son únicas; así mismo no todas las matrices pueden factorizarse en la forma LU.

Existen tres métodos para factorizar una matriz en la forma LU.

1. **Doolittle** : requiere que haya unos en la diagonal de L, ($l_{ii} = 1$ para $1 \le i \le n$).
2. **Crout** : requiere que haya unos en la diagonal de U, ($u_{ii} = 1$ para $1 \le i \le n$).
3. **Cholesky** : requiere que $l_{ii} = u_{ii}$, (para $1 \le i \le n$).

Una vez que se determina la factorización LU, se puede resolver en forma simplificada cualquier sistema que contenga la matriz A; es decir cuando se tienen que resolver varios conjuntos de ecuaciones lineales en los que todas las matrices de coeficientes son iguales pero los términos independientes son distintos. La solución de este tipo de ecuaciones lineales utilizando la factorización LU, tiende a ser más eficiente y simplificada.

4.1 Método de Doolittle

Aún cuando las matrices L y U pueden obtenerse por el método de Gauss (ver sección 3.3.1), es deseable encontrar un método más directo para su determinación. Para esto se analizará *el método de Doolittle*, el cual está basado en la factorización de la matriz A como sigue:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ l_{21} & 1 & 0 & \dots & 0 \\ l_{31} & l_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ 0 & u_{22} & u_{23} & \dots & u_{2n} \\ 0 & 0 & u_{33} & \dots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & u_{nn} \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}}_{A}$$

Para ilustrar cómo funciona el método, se analizará la factorización de una matriz A de orden 4, donde $l_{ii} = 1$. Se comienza alternando entre renglones de U y columnas de L, tomando en cuenta el elemento a_{ij} para obtener los renglones de u_{ij} y las columnas de l_{ij}, respectivamente:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ l_{21} & 1 & 0 & 0 \\ l_{31} & l_{32} & 1 & 0 \\ l_{41} & l_{42} & l_{43} & 1 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}}_{A}$$

entonces:

1. Para obtener el **primer renglón de** U, (u_{11}, u_{12}, u_{13} y u_{14}); se toman en cuenta los elementos a_{11}, a_{12}, a_{13} y a_{14}, y se realiza la multiplicación que se necesita para obtenerlos, (por ejemplo para obtener el elemento a_{11}, se multiplica primer renglón por primera columna, para obtener el elemento a_{12}, se multiplica primer renglón por segunda columna, etc.); despejando después los elementos de U, que se quieren encontrar:

$$\begin{array}{rcl} a_{11} & = & 1u_{11} \\ a_{12} & = & 1u_{12} \\ a_{13} & = & 1u_{13} \\ a_{14} & = & 1u_{14} \end{array} \quad \Rightarrow \quad \begin{array}{rcl} u_{11} & = & a_{11} \\ u_{12} & = & a_{12} \\ u_{13} & = & a_{13} \\ u_{14} & = & a_{14} \end{array}$$

2. Para obtener la **primera columna de** L, (1_{21}, l_{31}, y l_{41}); se toman en cuenta los elementos a_{21}, a_{31} y a_{41}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de L que se quieren encontrar:

$$
\begin{aligned}
a_{21} &= l_{21}u_{11} \\
a_{31} &= l_{31}u_{11} \\
a_{41} &= 1_{41}u_{11}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
l_{21} &= \frac{a_{21}}{u_{11}} \\
l_{31} &= \frac{a_{31}}{u_{11}} \\
l_{41} &= \frac{a_{41}}{u_{11}}
\end{aligned}
$$

3. Para obtener el **segundo renglón de** U, (u_{22}, u_{23} y u_{24}); se toman en cuenta los elementos a_{22}, a_{23} y a_{24}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de U, que se quieren encontrar:

$$
\begin{aligned}
a_{22} &= l_{21}u_{12} + u_{22} \\
a_{23} &= l_{21}u_{13} + u_{23} \\
a_{24} &= l_{21}u_{14} + u_{24}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
u_{22} &= a_{22} - l_{21}u_{12} \\
u_{23} &= a_{23} - l_{21}u_{13} \\
u_{24} &= a_{24} - l_{21}u_{14}
\end{aligned}
$$

4. Para obtener la **segunda columna de** L, (l_{32} y l_{34}); se toman en cuenta los elementos a_{32} y a_{42}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de L, que se quieren encontrar:

$$
\begin{aligned}
a_{32} &= l_{31}u_{12} + l_{32}u_{22} \\
a_{42} &= l_{41}u_{12} + l_{42}u_{22}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
l_{32} &= \frac{a_{32} - l_{31}u_{12}}{u_{22}} \\
l_{42} &= \frac{a_{42} - l_{41}u_{12}}{u_{22}}
\end{aligned}
$$

5. Para obtener el **tercer renglón de** U, (u_{33} y u_{34}); se toman en cuenta los elementos a_{33} y a_{32}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de U, que se quieren encontrar:

$$
\begin{aligned}
a_{33} &= l_{31}u_{13} + l_{32}u_{23} + u_{33} \\
a_{34} &= l_{31}u_{14} + l_{32}u_{24} + u_{34}
\end{aligned}
\quad \Rightarrow \quad
\begin{aligned}
u_{33} &= a_{33} - l_{31}u_{13} - l_{32}u_{23} \\
u_{34} &= a_{34} - l_{31}u_{14} - l_{32}u_{24}
\end{aligned}
$$

6. Para obtener la **tercera columna de** L, (l_{43}); se toma en cuenta el elemento a_{43}, y se realiza la multiplicación que se necesita para obtenerlo; despejando después el elemento de L, que se quiere encontrar:

$$
a_{43} = l_{41}u_{13} + l_{42}u_{23} + l_{43}u_{33} \quad \Rightarrow l_{43} = \frac{a_{43} - l_{41}u_{13} - l_{42}u_{23}}{u_{33}}
$$

7. Para obtener el **cuarto renglón de** U, (u_{44}); se toma en cuenta el elemento a_{44}, y se realiza la multiplicación que se necesita para obtenerlo; despejando después el elemento de U, que se quiere encontrar:

$$
a_{44} = l_{41}u_{14} + l_{42}u_{24} + l_{43}u_{34} + u_{44} \quad \Rightarrow u_{44} = a_{44} - l_{41}u_{14} - l_{42}u_{24} - l_{43}u_{34}
$$

Con lo anterior se puede observar que al utilizar el elemento a_{ij} en el cálculo de u_{ij} y l_{ij} según sea el caso, este elemento no vuelve a emplearse como tal, por lo que los elementos de L y U generados pueden guardarse en A y ahorrar memoria de esa manera:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix} \longrightarrow \begin{pmatrix} u_{11} & u_{12} & u_{13} & \dots & u_{1n} \\ l_{21} & u_{22} & u_{23} & \dots & u_{2n} \\ l_{31} & l_{32} & u_{33} & \dots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & u_{nn} \end{pmatrix}$$

sabiendo que $l_{ii} = 1$.

Generalizando la obtención de los renglones de L y las columnas de U; la forma general para obtener los coeficientes de las matrices L y U sería:

$$\begin{gathered} u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik}\, u_{kj} \qquad \text{para} j = i, i+1, \dots, n \\ i = 1, 2, \dots, n \\ l_{ij} = \frac{1}{u_{jj}} \left(a_{ij} - \sum_{k=1}^{j-1} l_{ik}\, u_{kj} \right) \qquad \text{para} i = j+1, \dots, n \\ j = 1, 2, \dots, n-1 \\ l_{ii} = 1 \qquad \text{para} i = 1, 2, 3, \dots, n \end{gathered} \tag{4.9}$$

Con la convención en las sumatorias que $\sum_{k=1}^{0} = 0$.

Teorema 4.1.1 Si los n menores principales de la matriz A de $n \times n$ son no singulares, entonces la matriz A, se puede factorizar en la forma LU; donde $l_{ii} = 1$.

Teorema 4.1.2 Una matriz A se puede factorizar en la forma LU si y sólo si sus primeras submatrices principales tienen determinante distinto de cero.

Aplicando el teorema 4.1.1 se tiene que: con el método de Doolittle se puede factorizar la matriz A, en la forma LU si y sólo si para todo M_{ij}, $|M_{ij}| \neq 0$; donde M_{ij} es el menor ij de A. (Ver definición 3.1.2 para obtenerlo).

Aplicando el teorema 4.1.2 se tiene que: con el método de Doolittle se puede factorizar la matriz A, en la forma LU si y sólo si los determinantes de las submatrices principales de A son distintos de cero; es decir:

$$|a_{11}| \neq 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \qquad \dots, \qquad \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \neq 0$$

Para reducir los errores de redondeo, se debe utilizar alguna de las estrategias de pivoteo (sección 3.3.7), al aplicar el método de Doolittle.

Ejemplo 4.1.1 Resolver el siguiente sistema de ecuaciones lineales por el método de Doolittle; (det $A = 378$, obtenido con Maple).

$$
\begin{array}{rcrcrcrcr}
4x_1 & + & 5x_2 & + & 2x_3 & - & x_4 & = & 3 \\
5x_1 & + & 8x_2 & + & 7x_3 & + & 6x_4 & = & 2 \\
3x_1 & + & 7x_2 & - & 4x_3 & - & 2x_4 & = & 0 \\
-x_1 & + & 6x_2 & - & 2x_3 & + & 5x_4 & = & 1
\end{array}
\tag{4.4}
$$

Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$
\mathbf{A} = \begin{pmatrix} 4 & 5 & 2 & -1 \\ 5 & 8 & 7 & 6 \\ 3 & 7 & -4 & -2 \\ -1 & 6 & -2 & 5 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 3 \\ 2 \\ 0 \\ 1 \end{pmatrix}
$$

Aunque det $A \neq 0$, antes de empezar a aplicar el método se debe verificar que los determinantes de las submatrices principales de A sean también $\neq 0$, (teorema 4.1.2); para que la matriz A, se pueda factorizar en la forma LU.

Para verificar esto se obtiene el determinante de las submatrices principales de A, (estos se obtuvieron con Maple):

$$
|4| = 4 \neq 0, \qquad \begin{vmatrix} 4 & 5 \\ 5 & 8 \end{vmatrix} = 7 \neq 0,
$$

$$
\begin{vmatrix} 4 & 5 & 2 \\ 5 & 8 & 7 \\ 3 & 7 & -4 \end{vmatrix} = -97 \neq 0, \qquad \begin{vmatrix} 4 & 5 & 2 & -1 \\ 5 & 8 & 7 & 6 \\ 3 & 7 & -4 & -2 \\ -1 & 6 & -2 & 5 \end{vmatrix} = 378 \neq 0
$$

Como los determinantes de las submatrices principales de A son $\neq 0$, se empieza aplicar el método, utilizando la ecuación (4.3); obteniéndose:

Primer renglón de U

$u_{11} = 4$

$u_{12} = 5$

$u_{13} = 2$

$u_{14} = -1$

Primera columna de L

$l_{21} = \frac{5}{4}$

$l_{31} = \frac{3}{4}$

$l_{41} = -\frac{1}{4}$

Segundo renglón de U

$u_{22} = 8 - \frac{5}{4}(5) = \frac{7}{4}$

$u_{23} = 7 - \frac{5}{4}(2) = \frac{9}{2}$

$u_{24} = 6 - \frac{5}{4}(-1) = \frac{29}{4}$

Segunda columna de L

$l_{32} = \frac{7 - \frac{3}{4}(5)}{\frac{7}{4}} = \frac{13}{7}$

$l_{42} = \frac{6 - \left(-\frac{1}{4}\right)(5)}{\frac{7}{4}} = \frac{29}{7}$

Tercer renglón de U

$u_{33} = -4 - \frac{3}{4}(2) - \frac{13}{7}\left(\frac{9}{2}\right) = -\frac{97}{7}$

$u_{34} = -2 - \frac{3}{4}(-1) - \frac{13}{7}\left(\frac{29}{4}\right) = -\frac{103}{7}$

Tercera columna de L

$l_{43} = \frac{-2 - \left(\frac{1}{4}\right)(2) - \left(\frac{29}{7}\right)\left(\frac{9}{2}\right)}{-\frac{97}{7}} = \frac{141}{97}$

Cuarto renglón de U

$u_{44} = 5 - \left(-\frac{1}{4}\right)(-1) - \frac{29}{7}\left(\frac{29}{4}\right) - \frac{141}{97}\left(-\frac{103}{7}\right) = -\frac{378}{97}$

Entonces la matriz A, se transforma de la siguiente manera:

$$\Rightarrow \begin{pmatrix} 4 & 5 & 2 & -1 \\ 5 & 8 & 7 & 6 \\ 3 & 7 & -4 & -2 \\ -1 & 6 & -2 & 5 \end{pmatrix} \longrightarrow \begin{pmatrix} 4 & 5 & 2 & -1 \\ \frac{5}{4} & \frac{7}{4} & \frac{9}{2} & \frac{29}{4} \\ \frac{3}{4} & \frac{13}{7} & -\frac{97}{7} & -\frac{103}{7} \\ -\frac{1}{4} & \frac{29}{7} & \frac{141}{97} & -\frac{378}{97} \end{pmatrix}$$

donde:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{5}{4} & 1 & 0 & 0 \\ \frac{3}{4} & \frac{13}{7} & 1 & 0 \\ -\frac{1}{4} & \frac{29}{7} & \frac{141}{97} & 1 \end{pmatrix} \quad \text{y} \quad U = \begin{pmatrix} 4 & 5 & 2 & -1 \\ 0 & \frac{7}{4} & \frac{9}{2} & \frac{29}{4} \\ 0 & 0 & -\frac{97}{7} & -\frac{103}{7} \\ 0 & 0 & 0 & -\frac{378}{97} \end{pmatrix}$$

Comprobación:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{5}{4} & 1 & 0 & 0 \\ \frac{3}{4} & \frac{13}{7} & 1 & 0 \\ -\frac{1}{4} & \frac{29}{7} & \frac{141}{97} & 1 \end{pmatrix}}_{\text{L}} \underbrace{\begin{pmatrix} 4 & 5 & 2 & -1 \\ 0 & \frac{7}{4} & \frac{9}{2} & \frac{29}{4} \\ 0 & 0 & -\frac{97}{7} & -\frac{103}{7} \\ 0 & 0 & 0 & -\frac{378}{97} \end{pmatrix}}_{\text{U}} = \underbrace{\begin{pmatrix} 4 & 5 & 2 & -1 \\ 5 & 8 & 7 & 6 \\ 3 & 7 & -4 & -2 \\ -1 & 6 & -2 & 5 \end{pmatrix}}_{\text{A}}$$

Para obtener la solución del sistema de ecuaciones lineales (4.4) se empieza resolviendo el sistema (4.1); obteniéndose:

$$Ly = b \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{5}{4} & 1 & 0 & 0 \\ \frac{3}{4} & \frac{13}{7} & 1 & 0 \\ -\frac{1}{4} & \frac{29}{7} & \frac{141}{97} & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 3 \\ 2 \\ 0 \\ 1 \end{pmatrix}$$

Se aplica la sustitución hacia adelante, a fin de obtener el valor de las incógnitas del vector y.

$$\Rightarrow \begin{array}{rcrcrcrcl} y_1 & & & & & & & = & 3 \\ \frac{5}{4}y_1 & + & y_2 & & & & & = & 2 \\ \frac{3}{4}y_1 & + & \frac{13}{7}y_2 & + & y_3 & & & = & 0 \\ -\frac{1}{4}y_1 & + & \frac{29}{7}y_2 & + & \frac{141}{97}y_3 & + & y_4 & = & 1 \end{array} \qquad \Rightarrow \begin{array}{rcl} y_1 & = & 3 \\ y_2 & = & -\frac{7}{4} \\ y_3 & = & 1 \\ y_4 & = & \frac{732}{97} \end{array}$$

Después se resuelve el sistema (4.2); obteniéndose:

$$Ux = y \qquad \begin{pmatrix} 4 & 5 & 2 & -1 \\ 0 & \frac{7}{4} & \frac{9}{2} & \frac{29}{4} \\ 0 & 0 & -\frac{97}{7} & -\frac{103}{7} \\ 0 & 0 & 0 & -\frac{378}{97} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 3 \\ -\frac{7}{4} \\ 1 \\ \frac{732}{97} \end{pmatrix}$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas del vector x.

$$\Rightarrow \begin{array}{rcrcrcrcl} 4x_1 & + & 5x_2 & + & 2x_3 & - & x_4 & = & 3 \\ & & \frac{7}{4}x_2 & + & \frac{9}{2}x_3 & + & \frac{29}{4}x_4 & = & -\frac{7}{4} \\ & & & & -\frac{97}{7}x_3 & - & \frac{103}{7}x_4 & = & 1 \\ & & & & & & \frac{378}{97}x_4 & = & \frac{732}{97} \end{array} \qquad \begin{array}{rcl} x_1 & = & -\frac{197}{63} \\ x_2 & = & \frac{121}{63} \\ x_3 & = & \frac{125}{63} \\ x_4 & = & -\frac{122}{63} \end{array}$$

Comprobación:

$$\begin{array}{rcrcrcrcl} 4(-\frac{197}{63}) & + & 5(\frac{121}{63}) & + & 2(\frac{125}{63}) & - & (-\frac{122}{63}) & = & 3 \\ 5(-\frac{197}{63}) & + & 8(\frac{121}{63}) & + & 7(\frac{125}{63}) & + & 6(-\frac{122}{63}) & = & 2 \\ 3(-\frac{197}{63}) & + & 7(\frac{121}{63}) & - & 4(\frac{125}{63}) & - & 2(-\frac{122}{63}) & = & 0 \\ -(-\frac{197}{63}) & + & 6(\frac{121}{63}) & - & 2(\frac{125}{63}) & + & 5(-\frac{122}{63}) & = & 1 \end{array}$$

Ejemplo 4.1.2 Resolver el siguiente sistema de ecuaciones lineales por el método de Doolittle, utilizando pivoteo parcial, con 5 cifras significativas; (det $A = 192639.4330$, obtenido con Maple).

$$\begin{array}{rcrcrcl} 1.9999x_1 & + & 17.01x_2 & + & 9.6x_3 & = & 1 \\ 1.6x_1 & + & 5.2x_2 & + & 1.7x_3 & = & 0 \\ 3.444x_1 & + & 16100x_2 & - & 9.1x_3 & = & 0 \end{array} \tag{4.5}$$

Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A}=\begin{pmatrix}1.9999 & 17.01 & 9.6\\ 1.6 & 5.2 & 1.7\\ 3.4444 & 16100 & -9.1\end{pmatrix}\qquad \mathbf{b}=\begin{pmatrix}1\\0\\0\end{pmatrix}$$

Aunque $\det A\neq 0$, antes de empezar a aplicar el método se debe verificar que los determinantes de las submatrices principales de A sean también $\neq 0$, (teorema 4.2.1); para que la matriz A, se pueda factorizar en la forma LU.
Para verificar esto se obtiene el determinante de las submatrices principales de A, (estos se obtuvieron con Maple):

$$\left|1.9999\right| = 1.9999\neq 0,\qquad \begin{vmatrix}1.9999 & 17.01\\ 1.6 & 5.2\end{vmatrix} = -16.817\neq 0,$$

$$\begin{vmatrix}1.9999 & 17.01 & 9.6\\ 1.6 & 5.2 & 1.7\\ 3.4444 & 16100 & -9.1\end{vmatrix} = 192639.4330\neq 0,$$

Como los determinantes de las submatrices principales de A son $\neq 0$, se aplica el método, utilizando la ecuación (4.3) y pivoteo parcial (sección 3.3.7).
Se combinan las matrices A y b para obtener su forma aumentada. Empezando a aplicar pivoteo parcial; para escoger el pivote a_{11}, de la primera columna con elementos diferentes de cero (llamada columna pivote), se selecciona el elemento de mayor valor absoluto, en este caso es el elemento $a_{13} = 3.4444$; a este elemento se le llama pivote:

$$\left(\begin{array}{ccc|c}1.9999 & 17.01 & 9.6 & 1\\ 1.6 & 5.2 & 1.7 & 0\\ 3.4444 & 16100 & -9.1 & 0\end{array}\right)$$

Se intercambia el renglón 1 por el renglón 3, para llevar este elemento de la columna a la posición diagonal.

$$\left(\begin{array}{ccc|c}3.4444 & 16100 & -9.1 & 0\\ 1.6 & 5.2 & 1.7 & 0\\ 1.9999 & 17.01 & 9.6 & 1\end{array}\right)$$

Se obtiene el primer renglón de U y la primera columna de L y se guardan en los elementos correspondientes de A.

Primer renglón de U	Primera columna de L
$u_{11} = 3.4444$	$l_{21} = \frac{1.6}{3.4444} = 0.46458$
$u_{12} = 16100$	$l_{31} = \frac{1.9999}{3.4444} = 0.58069$
$u_{13} = -9.1$	

$$\left(\begin{array}{ccc|c} 3.4444 & 16100 & -9.1 & 0 \\ 0.46458 & 5.2 & 1.7 & 0 \\ 0.58069 & 17.01 & 9.6 & 1 \end{array}\right)$$

D e la segunda columna, se selecciona el nuevo pivote a_{22}, que es el elemento de mayor valor absoluto de la diagonal principal para abajo, en este caso es el elemento $a_{32} = 17.01$; a este elemento se le llama pivote.

$$\left(\begin{array}{ccc|c} 3.4444 & 16100 & -9.1 & 0 \\ 0.46458 & 5.2 & 1.7 & 0 \\ 0.58069 & 17.01 & 9.6 & 1 \end{array}\right)$$

Se intercambia el renglón 2 por el renglón 3, para llevar este elemento de la columna a la posición diagonal.

$$\left(\begin{array}{ccc|c} 3.4444 & 16100 & -9.1 & 0 \\ 0.58069 & 17.01 & 9.6 & 1 \\ 0.46458 & 5.2 & 1.7 & 0 \end{array}\right)$$

Se obtiene el segundo renglón de U y la segunda columna de L y se guardan en los elementos correspondientes de A.

Segundo renglón de U	Segunda columna de L
$u_{22} = 17.01 - (0.58069)(16100) = -9332.1$	$l_{32} = \frac{5.2-(0.46458)(16100)}{14.884} = 0.80095$
$u_{23} = 9.6 - (0.58069)(-9.1) = 14.884$	

$$\left(\begin{array}{ccc|c} 3.4444 & 16100 & -9.1 & 0 \\ 0.58069 & -9332.1 & 14.884 & 1 \\ 0.46458 & 0.80095 & 1.7 & 0 \end{array}\right)$$

Finalmente se obtiene el tercer renglón de U y se guarda en los elementos correspondientes de A.

Tercer renglón de U

$$u_{33} = 1.7 - (0.46458)(-9.1) - (0.80095)(14.884) = -5.9933$$

$$\left(\begin{array}{ccc|c} 3.4444 & 16100 & -9.1 & 0 \\ 0.58069 & -9332.1 & 14.884 & 1 \\ 0.46458 & 0.80095 & -5.9933 & 0 \end{array}\right)$$

Entonces la matriz A, se transforma de la siguiente manera:

$$\Rightarrow \begin{pmatrix} 3.4444 & 16100 & -9.1 \\ 1.9999 & 17.01 & 9.6 \\ 1.6 & 5.2 & 1.7 \end{pmatrix} \longrightarrow \begin{pmatrix} 3.4444 & 16100 & -9.1 \\ 0.58069 & -9332.1 & 14.884 \\ 0.46458 & 0.80095 & -5.9933 \end{pmatrix}$$

donde:

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 0.58069 & 1 & 0 \\ 0.46458 & 0.80095 & 1 \end{pmatrix} \quad \text{y} \quad U = \begin{pmatrix} 3.4444 & 16100 & -9.1 \\ 0 & -9332.1 & 14.884 \\ 0 & 0 & -5.9933 \end{pmatrix}$$

Comprobación:

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 \\ 0.58069 & 1 & 0 \\ 0.46458 & 0.80095 & 1 \end{pmatrix}}_{\text{L}} \underbrace{\begin{pmatrix} 3.4444 & 16100 & -9.1 \\ 0 & -9332.1 & 14.884 \\ 0 & 0 & -5.9933 \end{pmatrix}}_{\text{U}} = \underbrace{\begin{pmatrix} 3.4444 & 16100 & -9.1 \\ 1.9999 & 17.01 & 9.6 \\ 1.6 & 5.2 & 1.7 \end{pmatrix}}_{\text{A}}$$

Se puede observar que al multiplicar L por U, no se obtiene la matriz A original, se obtiene la matriz afectada por el intercambio de renglones; por lo que al resolver el sistema (4.1), el vector b, también es afectado por el intercambio de renglones.
Para obtener la solución del sistema de ecuaciones lineales (4.5) se empieza resolviendo el sistema (4.1); obteniéndose:

$$Ly = b \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0.58069 & 1 & 0 \\ 0.46458 & 0.80095 & 1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

Se aplica la sustitución hacia adelante, a fin de obtener el valor de las incógnitas del vector y.

$$\Rightarrow \begin{array}{ccccccc} y_1 & & & & & = & 0 \\ (0.58069)y_1 & + & y_2 & & & = & 1 \\ (0.46458)y_1 & + & (0.80095)y_2 & + & y_3 & = & 0 \end{array} \qquad \Rightarrow \begin{array}{ccc} y_1 & = & 0 \\ y_2 & = & 1 \\ y_3 & = & -0.80095 \end{array}$$

Después se resuelve el sistema (4.2); obteniéndose:

$$Ux = y \qquad \begin{pmatrix} 3.4444 & 16100 & -9.1 \\ 0 & -9332.1 & 14.884 \\ 0 & 0 & -5.9933 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ -0.80095 \end{pmatrix}$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas del vector x.

$$\Rightarrow \begin{array}{ccccccc} (3.4444)x_1 & + & (16100)x_2 & + & (-9.1)x_3 & = & 0 \\ & & (-9332.1)x_2 & + & (14.884)x_3 & = & 1 \\ & & & & (-5.9933)x_3 & = & -0.80095 \end{array} \qquad \begin{array}{ccc} x_1 & \approx & -0.14237 \\ x_2 & \approx & 0.00010599 \\ x_3 & \approx & 0.13364 \end{array}$$

Comprobación:

$$\begin{array}{rcrcrcrcl} 1.9999(-0.14237) & + & 17.01(0.00010599) & + & 9.6(0.13364) & = & 0.99997 & \approx & 1 \\ 1.6(-0.14237) & + & 5.2(0.00010599) & + & 1.7(0.13364) & = & -0.00005 & \approx & 0 \\ 3.444(-0.14237) & + & 16100(0.00010599) & - & 9.1(0.13364) & = & 0 & \approx & 0 \end{array}$$

Nota: Una buena aproximación al sistema (4.5) es $x_1 = -0.14232$, $x_2 = 0.00010597$ y $x_3 = 0.13363$. Al resolver el ejemplo 4.2.1, sin utilizar alguna estrategia de pivoteo la solución que se obtiene es: $x_1 \approx -0.14233$, $x_2 \approx 0.00010860$ y $x_3 \approx 0.13362$. Comparando estos resultados con los obtenidos en el ejemplo 4.1.2 , se puede observar que si se aplica pivoteo parcial los errores relativos son menores.

4.2 Método de Crout

El *método de Crout*, está basado en la factorización de la matriz A como sigue:

$$\underbrace{\begin{pmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ l_{31} & l_{32} & l_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{nn} \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 1 & u_{12} & u_{13} & \dots & u_{1n} \\ 0 & 1 & u_{23} & \dots & u_{2n} \\ 0 & 0 & 1 & \dots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}}_{A}$$

Para ilustrar cómo funciona el método, se analizará la factorización de una matriz A de orden 4, donde $u_{ii} = 1$. Se comienza alternando entre columnas de L y renglones de U, tomando en cuenta el elemento a_{ij} para obtener los renglones de u_{ij} y las columnas de l_{ij}, respectivamente:

$$\underbrace{\begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 1 & u_{12} & u_{13} & u_{14} \\ 0 & 1 & u_{23} & u_{24} \\ 0 & 0 & 1 & u_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}}_{A}$$

entonces:

1. Para obtener la **primera columna de** L, (l_{11}, l_{21}, l_{31} y l_{41}); se toman en cuenta los elementos a_{11}, a_{21}, a_{31} y a_{41} , y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de L, que se quieren encontrar:

$$\begin{array}{rcl} a_{11} & = & l_{11} \\ a_{21} & = & l_{21} \\ a_{31} & = & l_{31} \\ a_{41} & = & 1_{41} \end{array} \quad \Rightarrow \quad \begin{array}{rcl} l_{11} & = & a_{11} \\ l_{21} & = & a_{21} \\ l_{31} & = & a_{31} \\ l_{41} & = & a_{41} \end{array}$$

2. Para obtener el **primer renglón de** U, (u_{12}, u_{13} y u_{14}); se toman en cuenta los elementos a_{12}, a_{13} y a_{14}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de U, que se quieren encontrar:

$$\begin{aligned} a_{12} &= l_{11}u_{12} \\ a_{13} &= l_{11}u_{13} \\ a_{14} &= l_{11}u_{14} \end{aligned} \quad \Rightarrow \quad \begin{aligned} u_{12} &= \frac{a_{12}}{l_{11}} \\ u_{13} &= \frac{a_{13}}{l_{11}} \\ u_{14} &= \frac{a_{14}}{l_{11}} \end{aligned}$$

3. Para obtener la **segunda columna de** L, (l_{22}, l_{32} y l_{42}); se toman en cuenta los elementos a_{22}, a_{32} y a_{42}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de L, que se quieren encontrar:

$$\begin{aligned} a_{22} &= l_{21}u_{12} + l_{22} \\ a_{32} &= l_{31}u_{12} + l_{32} \\ a_{42} &= l_{41}u_{12} + l_{42} \end{aligned} \quad \Rightarrow \quad \begin{aligned} l_{22} &= a_{22} - l_{21}u_{12} \\ l_{32} &= a_{32} - l_{31}u_{12} \\ l_{42} &= a_{42} - l_{41}u_{12} \end{aligned}$$

4. Para obtener el **segundo renglón de** U, (u_{23} y u_{24}); se toman en cuenta los elementos a_{23} y a_{24}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de U, que se quieren encontrar:

$$\begin{aligned} a_{23} &= l_{21}u_{13} + l_{22}u_{23} \\ a_{24} &= l_{21}u_{14} + l_{22}u_{24} \end{aligned} \quad \Rightarrow \quad \begin{aligned} u_{23} &= \frac{a_{23} - l_{21}u_{13}}{l_{22}} \\ u_{24} &= \frac{a_{24} - l_{21}u_{14}}{l_{22}} \end{aligned}$$

5. Para obtener la **tercera columna de** L, (l_{33} y l_{43}); se toman en cuenta los elementos a_{33} y a_{43}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de L, que se quieren encontrar:

$$\begin{aligned} a_{33} &= l_{31}u_{13} + l_{32}u_{23} + l_{33} \\ a_{43} &= l_{41}u_{13} + l_{42}u_{23} + l_{43} \end{aligned} \quad \Rightarrow \quad \begin{aligned} l_{33} &= a_{33} - l_{31}u_{13} - l_{32}u_{23} \\ l_{43} &= a_{43} - l_{41}u_{13} - l_{42}u_{23} \end{aligned}$$

6.

7. Para obtener el **tercer renglón de** U, (u_{34}); se toma en cuenta el elementos a_{34}, y se realiza la multiplicación que se necesita para obtenerlo; despejando después el elemento de U, que se quiere encontrar:

$$a_{34} = l_{31}u_{14} + l_{32}u_{24} + l_{33}u_{34} \quad \Rightarrow u_{34} = \frac{a_{34} - l_{31}u_{14} - l_{32}u_{24}}{l_{33}}$$

8. Para obtener la **cuarta columna de** L, (l_{44}); se toma en cuenta el elemento a_{44}, y se realiza la multiplicación que se necesita para obtenerlo; despejando después el elemento de L, que se quiere encontrar:

$$a_{44} = l_{41}u_{14} + l_{42}u_{24} + l_{43}u_{34} + l_{44} \qquad \Rightarrow l_{44} = a_{44} - l_{41}u_{14} - l_{42}u_{24} - l_{43}u_{34}$$

Al igual que en el método de Doolittle, se puede observar que al utilizar el elemento a_{ij} en el cálculo de u_{ij} y l_{ij} según sea el caso, este elemento no vuelve a emplearse como tal, por lo que los elementos de L y U generados pueden guardarse en A y ahorrar memoria de esa manera:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{pmatrix} \longrightarrow \begin{pmatrix} l_{11} & u_{12} & u_{13} & \cdots & u_{1n} \\ l_{21} & l_{22} & u_{23} & \cdots & u_{2n} \\ l_{31} & l_{32} & l_{33} & \cdots & u_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \cdots & l_{nn} \end{pmatrix}$$

sabiendo que $u_{ii} = 1$.

Generalizando la obtención de los renglones de L y las columnas de U; la forma general para obtener los coeficientes de las matrices L y U sería:

$$\begin{aligned} u_{ij} &= \frac{1}{l_{ii}}\left(a_{ij} - \sum_{k=1}^{i-1} l_{ik}\, u_{kj}\right) \qquad \text{para} j = i+1, \ldots, n \\ & \qquad i = 1, 2, \ldots, n-1 \\ l_{ij} &= a_{ij} - \sum_{k=1}^{j-1} l_{ik}\, u_{kj} \qquad \text{para} i = j, j+1, \ldots, n \\ & \qquad j = 1, 2, \ldots, n \\ u_{ii} &= 1 \qquad \text{para} i = 1, 2, 3, \ldots, n \end{aligned} \tag{4.6}$$

Con la convención en las sumatorias que $\sum_{k=1}^{0} = 0$.

Aplicando el teorema 4.1.2 se tiene que: con el método de Crout se puede factorizar la matriz A, en la forma LU si y sólo si los determinantes de las submatrices principales de A son distintos de cero; es decir:

$$|a_{11}| \neq 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \qquad \ldots, \qquad \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} \neq 0$$

Para reducir los errores de redondeo, se debe utilizar alguna de las estrategias de pivoteo (sección 3.3.7), al aplicar el método de Crout.

Ejemplo 4.2.1 Resolver el siguiente sistema de ecuaciones lineales por el método de Crout; (det $A = 191$, obtenido con Maple).

$$\begin{array}{rcrcrcrcr} 6x_1 & + & 2x_2 & + & x_3 & - & x_4 & = & 0 \\ 2x_1 & + & 4x_2 & + & x_3 & & & = & 7 \\ x_1 & + & x_2 & + & 4x_3 & - & x_4 & = & -1 \\ -x_1 & & & - & x_3 & + & 3x_4 & = & -2 \end{array} \tag{4.7}$$

Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A}=\begin{pmatrix}6 & 2 & 1 & -1\\ 2 & 4 & 1 & 0\\ 1 & 1 & 4 & -1\\ -1 & 0 & -1 & 3\end{pmatrix}\qquad \mathbf{b}=\begin{pmatrix}0\\ 7\\ -1\\ -2\end{pmatrix}$$

Aunque det $A \neq 0$, antes de empezar a aplicar el método se debe verificar que los determinantes de las submatrices principales de A sean también $\neq 0$, (teorema 4.1.2); para que la matriz A, se pueda factorizar en la forma LU.

Para verificar esto se obtiene el determinante de las submatrices principales de A, (estos se obtuvieron con Maple):

$$|6| = 6 \neq 0, \qquad \begin{vmatrix}6 & 2\\ 2 & 4\end{vmatrix} = 20 \neq 0,$$

$$\begin{vmatrix}6 & 2 & 1\\ 2 & 4 & 1\\ 1 & 1 & 4\end{vmatrix} = 74 \neq 0, \qquad \begin{vmatrix}6 & 2 & 1 & -1\\ 2 & 4 & 1 & 0\\ 1 & 1 & 4 & -1\\ -1 & 0 & -1 & 3\end{vmatrix} = 191 \neq 0$$

Como los determinantes de las submatrices principales de A son $\neq 0$, se empieza aplicar el método, utilizando la ecuación (4.6); obteniéndose:

Primera columna de L

$l_{11} = 6$
$l_{21} = 2$
$l_{31} = 1$
$l_{41} = -1$

Primer renglón de U

$u_{12} = \frac{1}{3}$
$u_{13} = \frac{1}{6}$
$u_{14} = -\frac{1}{6}$

Segunda columna de L

$l_{22} = 4 - 2\left(\frac{1}{3}\right) = \frac{10}{3}$
$l_{32} = 1 - 1\left(\frac{1}{3}\right) = \frac{2}{3}$
$l_{42} = 0 - (-1)\left(\frac{1}{3}\right) = \frac{1}{3}$

Segundo renglón de U

$u_{23} = \frac{1-2\left(\frac{1}{6}\right)}{\frac{10}{3}} = \frac{1}{5}$
$u_{24} = \frac{0-2\left(-\frac{1}{6}\right)}{\frac{10}{3}} = \frac{1}{10}$

Tercera columna de L

$l_{33} = 4 - 1\left(\frac{1}{6}\right) - \frac{2}{3}\left(\frac{1}{5}\right) = \frac{37}{10}$
$l_{43} = -1 - (-1)\left(\frac{1}{6}\right) - \frac{1}{3}\left(\frac{1}{5}\right) = -\frac{9}{10}$

Tercer renglón de U

$u_{34} = \frac{-1-1\left(-\frac{1}{6}\right)-\frac{2}{3}\left(\frac{1}{10}\right)}{\frac{37}{10}} = -\frac{9}{37}$

Cuarta columna de L

$l_{44} = 3 - (-1)\left(-\frac{1}{6}\right) - \frac{1}{3}\left(\frac{1}{10}\right) - \left(-\frac{9}{10}\right)\left(-\frac{9}{37}\right) = \frac{191}{74}$

Entonces la matriz A, se transforma de la siguiente manera:

$$\Rightarrow \begin{pmatrix} 6 & 2 & 1 & -1 \\ 2 & 4 & 1 & 0 \\ 1 & 1 & 4 & -1 \\ -1 & 0 & -1 & 3 \end{pmatrix} \longrightarrow \begin{pmatrix} 6 & \frac{1}{3} & \frac{1}{6} & -\frac{1}{6} \\ 2 & \frac{10}{3} & \frac{1}{5} & \frac{1}{10} \\ 1 & \frac{2}{3} & \frac{37}{10} & -\frac{9}{37} \\ -1 & \frac{1}{3} & -\frac{9}{10} & \frac{191}{74} \end{pmatrix}$$

donde:

$$L = \begin{pmatrix} 6 & 0 & 0 & 0 \\ 2 & \frac{10}{3} & 0 & 0 \\ 1 & \frac{2}{3} & \frac{37}{10} & 0 \\ -1 & \frac{1}{3} & -\frac{9}{10} & \frac{191}{74} \end{pmatrix} \quad \text{y} \quad U = \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{6} & -\frac{1}{6} \\ 0 & 1 & \frac{1}{5} & \frac{1}{10} \\ 0 & 0 & 1 & -\frac{9}{37} \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Comprobación:

$$\underbrace{\begin{pmatrix} 6 & 0 & 0 & 0 \\ 2 & \frac{10}{3} & 0 & 0 \\ 1 & \frac{2}{3} & \frac{37}{10} & 0 \\ -1 & \frac{1}{3} & -\frac{9}{10} & \frac{191}{74} \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{6} & -\frac{1}{6} \\ 0 & 1 & \frac{1}{5} & \frac{1}{10} \\ 0 & 0 & 1 & -\frac{9}{37} \\ 0 & 0 & 0 & 1 \end{pmatrix}}_{U} = \underbrace{\begin{pmatrix} 6 & 2 & 1 & -1 \\ 2 & 4 & 1 & 0 \\ 1 & 1 & 4 & -1 \\ -1 & 0 & -1 & 3 \end{pmatrix}}_{A}$$

Para obtener la solución del sistema de ecuaciones lineales (4.7) se empieza resolviendo el sistema 4.1); obteniéndose:

$$Ly = b \qquad \begin{pmatrix} 6 & 0 & 0 & 0 \\ 2 & \frac{10}{3} & 0 & 0 \\ 1 & \frac{2}{3} & \frac{37}{10} & 0 \\ -1 & \frac{1}{3} & -\frac{9}{10} & \frac{191}{74} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 7 \\ -1 \\ -2 \end{pmatrix}$$

Se aplica la sustitución hacia adelante, a fin de obtener el valor de las incógnitas del vector y.

$$\Rightarrow \begin{array}{rcrcrcrcl} & & 6y_1 & & & & & = & 0 \\ & & 2y_1 & + & \frac{10}{3}y_2 & & & = & 7 \\ & & y_1 & + & \frac{2}{3}y_2 & + & \frac{37}{10}y_3 & = & -1 \\ - & & y_1 & + & \frac{1}{3}y_2 & - & \frac{9}{10}y_3 + \frac{191}{74}y_4 & = & -2 \end{array} \qquad \Rightarrow \begin{array}{rcl} y_1 & = & 0 \\ y_2 & = & \frac{21}{10} \\ y_3 & = & -\frac{24}{37} \\ y_4 & = & -\frac{243}{191} \end{array}$$

Después se resuelve el sistema (4.2); obteniéndose:

$$Ux = y \qquad \begin{pmatrix} 1 & \frac{1}{3} & \frac{1}{6} & -\frac{1}{6} \\ 0 & 1 & \frac{1}{5} & \frac{1}{10} \\ 0 & 0 & 1 & -\frac{9}{37} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{21}{10} \\ -\frac{24}{37} \\ -\frac{243}{191} \end{pmatrix}$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas del vector x.

$$\begin{array}{rrrrrrrcr} x_1 & + & \frac{1}{3}x_2 & + & \frac{1}{6}x_3 & - & \frac{1}{6}x_4 & = & 0 \\ & & x_2 & + & \frac{1}{5}x_3 & + & \frac{1}{10}x_4 & = & \frac{21}{10} \\ & & & & x_3 & - & \frac{9}{37}x_4 & = & -\frac{24}{37} \\ & & & & & & x_4 & = & -\frac{243}{191} \end{array} \Rightarrow \begin{array}{rcr} x_1 & = & -\frac{164}{191} \\ x_2 & = & \frac{462}{191} \\ x_3 & = & -\frac{183}{191} \\ x_4 & = & -\frac{243}{191} \end{array}$$

Comprobación:

$$\begin{array}{rrrrrrrcr} 6(-\frac{164}{191}) & + & 2(\frac{462}{191}) & + & (-\frac{183}{191}) & - & (-\frac{243}{191}) & = & 0 \\ 2(-\frac{164}{191}) & + & 4(\frac{462}{191}) & + & (-\frac{183}{191}) & & & = & 7 \\ (-\frac{164}{191}) & + & (\frac{462}{191}) & + & 4(-\frac{183}{191}) & - & (-\frac{243}{191}) & = & -1 \\ -\ (-\frac{164}{191}) & & & - & (-\frac{183}{191}) & + & 3(-\frac{243}{191}) & = & -2 \end{array}$$

4.3 Método de Cholesky

El *método de Cholesky*, está basado en la factorización de la matriz A como sigue:

$$\underbrace{\begin{pmatrix} l_{11} & 0 & 0 & \dots & 0 \\ l_{21} & l_{22} & 0 & \dots & 0 \\ l_{31} & l_{32} & l_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{nn} \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} l_{11} & l_{21} & l_{31} & \dots & l_{n1} \\ 0 & l_{22} & l_{32} & \dots & l_{n2} \\ 0 & 0 & l_{33} & \dots & l_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & l_{nn} \end{pmatrix}}_{L^t} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix}}_{A}$$

Para poder aplicar éste método, se requiere que la matriz A tenga coeficientes reales, sea simétrica y definida positiva; y esta basado en la factorización de la matriz A como $A = LL^t$.

Teorema 4.3.1 Si A es una matriz simétrica y definida positiva, entonces tiene una factorización única $A = LL^t$ en donde L es una matriz triangular inferior con diagonal positiva.

Para ilustrar cómo funciona el método, se analizará la factorización de una matriz A de orden 4, donde $l_{ii} = u_{ii}$; tomando en cuenta el elemento a_{ij} para obtener las columnas de l_{ij}:

$$\underbrace{\begin{pmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} l_{11} & l_{21} & l_{31} & l_{41} \\ 0 & l_{22} & l_{32} & l_{42} \\ 0 & 0 & l_{33} & l_{43} \\ 0 & 0 & 0 & l_{44} \end{pmatrix}}_{L^t} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix}}_{A}$$

entonces:

1. Para obtener la **primera columna de** L, (l_{11}, l_{21}, l_{31} y l_{41}); se toman en cuenta los elementos a_{11}, a_{21}, a_{31} y a_{41}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de L, que se quieren encontrar:

$$\begin{aligned} a_{11} &= l_{11}^2 \\ a_{21} &= l_{21}l_{11} \\ a_{31} &= l_{31}l_{11} \\ a_{41} &= l_{41}l_{11} \end{aligned} \quad \Rightarrow \quad \begin{aligned} l_{11} &= \sqrt{a_{11}} \\ l_{21} &= \frac{a_{21}}{l_{11}} \\ l_{31} &= \frac{a_{31}}{l_{11}} \\ l_{41} &= \frac{a_{41}}{l_{11}} \end{aligned}$$

2. Para obtener la **segunda columna de** L, (l_{22}, l_{32} y l_{42}); se toman en cuenta los elementos a_{22}, a_{32} y a_{42}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de L, que se quieren encontrar:

$$\begin{aligned} a_{22} &= l_{21}^2 + l_{22}^2 \\ a_{32} &= l_{31}l_{21} + l_{32}l_{22} \\ a_{42} &= l_{41}l_{21} + l_{42}l_{22} \end{aligned} \quad \Rightarrow \quad \begin{aligned} l_{22} &= \sqrt{a_{22} - l_{21}^2} \\ l_{32} &= \frac{a_{32} - l_{31}l_{21}}{l_{22}} \\ l_{42} &= \frac{a_{42} - l_{41}l_{21}}{l_{22}} \end{aligned}$$

3. Para obtener la **tercera columna de** L, (l_{33} y l_{43}); se toman en cuenta los elementos a_{33} y a_{43}, y se realiza la multiplicación que se necesita para obtenerlos; despejando después los elementos de L, que se quieren encontrar:

$$\begin{aligned} a_{33} &= l_{31}^2 + l_{32}^2 + l_{33}^2 \\ a_{43} &= l_{41}l_{31} + l_{42}l_{32} + l_{43}l_{33} \end{aligned} \quad \Rightarrow \quad \begin{aligned} l_{33} &= \sqrt{a_{33} - l_{31}^2 - l_{32}^2} \\ l_{43} &= \frac{a_{43} - l_{41}l_{31} - l_{42}l_{32}}{l_{33}} \end{aligned}$$

4. Para obtener la **cuarta columna de** L, (l_{44}); se toma en cuenta el elemento a_{44}, y se realiza la multiplicación que se necesita para obtenerlo; despejando después el elemento de L, que se quiere encontrar:

$$a_{44} = l_{41}^2 + l_{42}^2 + l_{43}^2 + l_{44}^2 \quad \Rightarrow l_{44} = \sqrt{a_{44} - l_{41}^2 - l_{42}^2 - l_{43}^2}$$

Al igual que en el método de Doolittle y en el de Crout, se puede observar que al utilizar el elemento a_{ij} en el cálculo l_{ij}, este elemento no vuelve a emplearse como tal, por lo que los elementos de L generados pueden guardarse en A y ahorrar memoria de esa manera:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{pmatrix} \longrightarrow \begin{pmatrix} l_{11} & l_{21} & l_{31} & \dots & l_{n1} \\ l_{21} & l_{22} & l_{32} & \dots & l_{n2} \\ l_{31} & l_{32} & l_{33} & \dots & l_{n3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & l_{n3} & \dots & l_{nn} \end{pmatrix}$$

sabiendo que $l_{ii} = u_{ii}$.

Generalizando la obtención de las columnas de L; la forma general para obtener los coeficientes de la matriz L sería:

$$\begin{gathered} l_{11} = \sqrt{a_{11}} \\ l_{i1} = \frac{a_{i1}}{l_{11}} \qquad \text{para} i = 2, 3, \dots, n \\ l_{ij} = \left(a_{ij} - \sum_{k=1}^{i-1} l_{ik}^2 \right)^{\frac{1}{2}} \qquad \text{para} j = 2, 3, \dots, n \\ l_{ij} = \frac{1}{l_{ij}} \left(a_{ij} - \sum_{k=1}^{i-1} l_{ik} l_{jk} \right) \qquad \text{para} j = 2, 3, \dots, n \\ i = j+1, j+2, \dots, n-1 \\ l_{ij} = 0 \qquad \text{para} i < j \end{gathered} \tag{4.8}$$

Teorema 4.3.2 Una matriz simétrica A es definida positiva si y sólo si sus primeras submatrices principales tienen determinante positivo.

Aplicando el teorema 4.3.2 se tiene que: con el método de Cholesky se puede factorizar la matriz A, en la forma LL^t si y sólo si los determinantes de las submatrices principales de A son positivos; es decir:

$$|a_{11}| > 0, \qquad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} > 0, \qquad \dots, \qquad \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} > 0$$

Teorema 4.3.3 Si se multiplica una matriz cuadrada A de orden n por su transpuesta, entonces el producto de $A^t A$ resulta una matriz simétrica y definida positiva.

Este último teorema permite no restringir el método de Cholesky, aplicándolo a matrices de coeficientes cualesquiera, ya que si la matriz de coeficientes no es simétrica o no es definida positiva, basta multiplicar por la transpuesta para que tales condiciones se cumplan. En efecto, si $Ax = b$, entonces el sistema equivalente cuenta con una matriz de coeficientes $A^t A$ simétrica y definida positiva, y el vector de constantes es $A^t b$; el cual se puede resolver por el método de Cholesky.

Es importante observar que con este método no se puede aplicar alguna estrategia de pivoteo para reducir los errores de redondeo; ya que al realizar algún intercambio de renglón o de columna, la matriz A ya no cumpliría con la propiedad de ser simétrica.

Ejemplo 4.3.2 Resolver el siguiente sistema de ecuaciones lineales por el método de Cholesky, con 5 cifras significativas; (det $A = 808$, obtenido con Maple).

$$\begin{array}{rcrcrcrcl} 5x_1 & + & x_2 & + & 2x_3 & - & x_4 & = & 1 \\ x_1 & + & 7x_2 & & & + & 3x_4 & = & 2 \\ 2x_1 & & & + & 5x_3 & + & x_4 & = & 3 \\ -x_1 & + & 3x_2 & + & x_3 & + & 8x_4 & = & 4 \end{array} \tag{4.9}$$

Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 5 & 1 & 2 & -1 \\ 1 & 7 & 0 & 3 \\ 2 & 0 & 5 & 1 \\ -1 & 3 & 1 & 8 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

Aunque det $A \neq 0$, antes de empezar a aplicar el método se requiere verificar que la matriz A tenga coeficientes reales, sea simétrica y definida positiva.
Los coeficientes son reales. Para verificar si es simétrica se debe de cumplir que $A = A^t$.

$$\mathbf{A} = \begin{pmatrix} 5 & 1 & 2 & -1 \\ 1 & 7 & 0 & 3 \\ 2 & 0 & 5 & 1 \\ -1 & 3 & 1 & 8 \end{pmatrix} = \mathbf{A}^{\mathbf{t}} \begin{pmatrix} 5 & 1 & 2 & -1 \\ 1 & 7 & 0 & 3 \\ 2 & 0 & 5 & 1 \\ -1 & 3 & 1 & 8 \end{pmatrix}$$

Para saber si es definida positiva se deben calcular los determinantes de las submatrices principales de A, los cuales tienen que ser > 0, (teorema 4.3.2); para que la matriz A, se pueda factorizar en la forma LL^t.
Para verificar esto se obtiene el determinante de las submatrices principales de A, (estos se obtuvieron con Maple):

$$|5| = 5 > 0, \qquad \begin{vmatrix} 5 & 1 \\ 1 & 7 \end{vmatrix} = 34 > 0,$$

$$\begin{vmatrix} 5 & 1 & 2 \\ 1 & 7 & 0 \\ 2 & 0 & 5 \end{vmatrix} = 142 > 0, \qquad \begin{vmatrix} 5 & 1 & 2 & -1 \\ 1 & 7 & 0 & 3 \\ 2 & 0 & 5 & 1 \\ -1 & 3 & 1 & 8 \end{vmatrix} = 808 > 0$$

Como los determinantes de las submatrices principales de A son > 0, se empieza aplicar el método, utilizando la ecuación (4.8); obteniéndose:

Primera columna de L

$l_{11} = \sqrt{5} = 2.2361$
$l_{21} = \frac{1}{\sqrt{5}} = 0.44721$
$l_{31} = \frac{2}{\sqrt{5}} = 0.89441$
$l_{41} = -\frac{1}{\sqrt{5}} = -0.44721$

Segunda columna de L

$l_{22} = \sqrt{7 - (0.44721)^2} = 2.6077$
$l_{32} = \frac{0-(0.89441)(0.44721)}{2.6077} = -0.15339$
$l_{42} = \frac{3-(-0.44721)(0.44721)}{2.6077} = 1.2271$

Tercera columna de L

$l_{33} = \sqrt{5 - (0.89441)^2 - (-0.15339)^2} = 2.0436$
$l_{43} = \frac{1-(-0.44721)(0.89441)-(1.2271)(-0.15339)}{2.0436} = 0.77716$

Cuarta columna de L

$l_{44} = \sqrt{8 - (-0.44721)^2 - (1.2271)^2 - (0.77716)^2} = 2.3854$

Entonces la matriz A, se transforma de la siguiente manera:

$$\Rightarrow \begin{pmatrix} 5 & 1 & 2 & -1 \\ 1 & 7 & 0 & 3 \\ 2 & 0 & 5 & 1 \\ -1 & 3 & 1 & 8 \end{pmatrix} \longrightarrow \begin{pmatrix} 2.2361 & 0.44721 & 0.89441 & -0.44721 \\ 0.44721 & 2.6077 & -0.15339 & 1.2271 \\ 0.89441 & -0.15339 & 2.0436 & 0.77716 \\ -0.44721 & 1.2271 & 0.77716 & 2.3854 \end{pmatrix}$$

donde:

$$L = \begin{pmatrix} 2.2361 & 0 & 0 & 0 \\ 0.44721 & 2.6077 & 0 & 0 \\ 0.89441 & -0.15339 & 2.0436 & 0 \\ -0.44721 & 1.2271 & 0.77716 & 2.3854 \end{pmatrix} \quad \text{y}$$

$$L^t = \begin{pmatrix} 2.2361 & 0.44721 & 0.89441 & -0.44721 \\ 0 & 2.6077 & -0.15339 & 1.2271 \\ 0 & 0 & 2.0436 & 0.77716 \\ 0 & 0 & 0 & 2.3854 \end{pmatrix}$$

Comprobación:

$$\underbrace{\begin{pmatrix} 2.2361 & 0 & 0 & 0 \\ 0.44721 & 2.6077 & 0 & 0 \\ 0.89441 & -0.15339 & 2.0436 & 0 \\ -0.44721 & 1.2271 & 0.77716 & 2.3854 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 2.2361 & 0.44721 & 0.89441 & -0.44721 \\ 0 & 2.6077 & -0.15339 & 1.2271 \\ 0 & 0 & 2.0436 & 0.77716 \\ 0 & 0 & 0 & 2.3854 \end{pmatrix}}_{L^t} =$$

$$\underbrace{\begin{pmatrix} 5.0001 & 1.0000 & 2.0000 & -1.0000 \\ 1.0000 & 7.0001 & -0.00001 & 2.9999 \\ 2.0000 & -0.00001 & 4.9998 & 0.99999 \\ -1.0000 & 2.9999 & 0.99999 & 7.9999 \end{pmatrix}}_{\approx A}$$

Para obtener la solución del sistema de ecuaciones lineales (4.9) se empieza resolviendo el sistema (4.1); obteniéndose:

$$Ly = b \qquad \begin{pmatrix} 2.2361 & 0 & 0 & 0 \\ 0.44721 & 2.6077 & 0 & 0 \\ 0.89441 & -0.15339 & 2.0436 & 0 \\ -0.44721 & 1.2271 & 0.77716 & 2.3854 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}$$

Se aplica la sustitución hacia adelante, a fin de obtener el valor de las incógnitas del vector y.

$$\Rightarrow \begin{array}{rcrcrcrcl} & & 2.2361y_1 & & & & & = & 1 \\ & & 0.44721y_1 & + & 2.6077y_2 & & & = & 2 \\ & & 0.89441y_1 & - & 0.15339y_2 & + & 2.0436y_3 & = & 3 \\ - & & 0.44721y_1 & + & 1.2271y_2 & + & 0.77716y_3 + 2.3854y_4 & = & 4 \end{array} \qquad \Rightarrow \begin{array}{rcl} y_1 & = & 0.44721 \\ y_2 & = & 0.69027 \\ y_3 & = & 1.3241 \\ y_4 & = & 0.97424 \end{array}$$

Después se resuelve el sistema (4.2); obteniéndose:

$$L^t x = y \qquad \begin{pmatrix} 2.2361 & 0.44721 & 0.89441 & -0.44721 \\ 0 & 2.6077 & -0.15339 & 1.2271 \\ 0 & 0 & 2.0436 & 0.77716 \\ 0 & 0 & 0 & 2.3854 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 0.44721 \\ 0.69027 \\ 1.3241 \\ 0.97424 \end{pmatrix}$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas del vector x.

$$\Rightarrow \begin{array}{rcrcrcrcl} 2.2361x_1 & + & 0.44721x_2 & + & 0.89441x_3 & - & 0.44721x_4 & = & 0.44721 \\ & & 2.6077x_2 & - & 0.15339x_3 & + & 1.2271x_4 & = & 0.69027 \\ & & & & 2.0436x_3 & + & 0.77716x_4 & = & 1.3241 \\ & & & & & & 2.3854x_4 & = & 0.97424 \end{array} \qquad \begin{array}{rcl} x_1 & \approx & 0.064343 \\ x_2 & \approx & 0.10149 \\ x_3 & \approx & 0.49261 \\ x_4 & \approx & 0.40842 \end{array}$$

Comprobación:

$$\begin{array}{rcrcrcrcrcrcl} & & 5(0.064343) & + & (0.10149) & + & 2(0.49261) & - & (0.40842) & = & 0.99998 & \approx & 1 \\ & & (0.064343) & + & 7(0.10149) & + & & & 3(0.40842) & = & 2.0001 & \approx & 2 \\ & & 2(0.064343) & & & + & 5(0.49261) & + & (0.40842) & = & 3.0001 & \approx & 3 \\ - & & (0.064343) & + & 3(0.10149) & + & (0.49261) & + & 8(0.40842) & = & 4.0001 & \approx & 4 \end{array}$$

Ejemplo 4.3.2 Resolver el siguiente sistema de ecuaciones lineales por el método de Cholesky, con 6 cifras significativas; (det $A = -31$, obtenido con Maple).

$$\begin{array}{ccccccc} 4x_1 & + & 3x_2 & - & x_3 & = & 0 \\ 2x_1 & + & x_2 & & & = & 1 \\ 3x_1 & + & 8x_2 & + & 9x_3 & = & 4 \end{array} \tag{4.10}$$

Escribiendo el sistema de ecuaciones en notación matricial se tiene:

$$\mathbf{A} = \begin{pmatrix} 4 & 3 & -1 \\ 2 & 1 & 0 \\ 3 & 8 & 9 \end{pmatrix} \qquad \mathbf{b} = \begin{pmatrix} 0 \\ 1 \\ 4 \end{pmatrix}$$

Aunque det $A \neq 0$, antes de empezar a aplicar el método se requiere verificar que la matriz A tenga coeficientes reales, sea simétrica y definida positiva.
Los coeficientes son reales. Para verificar si es simétrica se debe de cumplir que $A = A^t$.

$$\mathbf{A} = \begin{pmatrix} 4 & 3 & -1 \\ 2 & 1 & 0 \\ 3 & 8 & 9 \end{pmatrix} \neq \mathbf{A^t} \begin{pmatrix} 4 & 2 & 3 \\ 3 & 1 & 8 \\ -1 & 0 & 9 \end{pmatrix}$$

Para saber si es definida positiva se deben calcular los determinantes de las submatrices principales de A, los cuales tienen que ser > 0, (teorema 4.3.2); para que la matriz A, se pueda factorizar en la forma LL^t.
Para verificar esto se obtiene el determinante de las submatrices principales de A, (estos se obtuvieron con Maple):

$$|4| = 4 > 0, \qquad \begin{vmatrix} 4 & 3 \\ 2 & 1 \end{vmatrix} = -2 < 0,$$

$$\begin{vmatrix} 4 & 3 & -1 \\ 2 & 1 & 0 \\ 3 & 8 & 9 \end{vmatrix} = -31 < 0$$

Como la matriz A, no es simétrica y no es definida positiva, basta multiplicar por la transpuesta para que tales condiciones se cumplan, (teorema ?); obteniéndose un sistema equivalente: $A' = A^t A$ y $b' = A^t b$.

$$\mathbf{A'} = \mathbf{A^t A} = \begin{pmatrix} 4 & 2 & 3 \\ 3 & 1 & 8 \\ -1 & 0 & 9 \end{pmatrix} \begin{pmatrix} 4 & 3 & -1 \\ 2 & 1 & 0 \\ 3 & 8 & 9 \end{pmatrix} = \begin{pmatrix} 29 & 38 & 23 \\ 38 & 74 & 69 \\ 23 & 69 & 82 \end{pmatrix}$$

$$\mathbf{b'} = \mathbf{A^t b} = \begin{pmatrix} 4 & 2 & 3 \\ 3 & 1 & 8 \\ -1 & 0 & 9 \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 4 \end{pmatrix} = \begin{pmatrix} 14 \\ 33 \\ 36 \end{pmatrix}$$

Se verificar si A' es simétrica ($A = A^t$).

$$\mathbf{A'} = \begin{pmatrix} 29 & 38 & 23 \\ 38 & 74 & 69 \\ 23 & 69 & 82 \end{pmatrix} = \mathbf{A'^t} \begin{pmatrix} 29 & 38 & 23 \\ 38 & 74 & 69 \\ 23 & 69 & 82 \end{pmatrix}$$

Se verificar si los determinantes de las submatrices principales de A' son > 0, (estos se obtuvieron con Maple):

$$|29| = 29 > 0, \qquad \begin{vmatrix} 29 & 28 \\ 38 & 74 \end{vmatrix} = 702 > 0,$$

$$\begin{vmatrix} 29 & 28 & 23 \\ 38 & 74 & 69 \\ 23 & 69 & 82 \end{vmatrix} = 961 > 0$$

Como A' es simétrica y sus determinantes de las submatrices principales son > 0 , se empieza aplicar el método, utilizando la ecuación (4.8); obteniéndose:

Primera columna de L

$l_{11} = \sqrt{29} = 5.38516$

$l_{21} = \frac{38}{\sqrt{29}} = 7.05643$

$l_{31} = \frac{23}{\sqrt{29}} = 4.27100$

Segunda columna de L

$l_{22} = \sqrt{74 - (7.05643)^2} = 4.92004$

$l_{32} = \frac{69-(4.27100)(7.05643)}{4.92004} = 7.89872$

Tercera columna de L

$l_{33} = \sqrt{82 - (4.27100)^2 - (7.89872)^2} = 1.16996$

Entonces la matriz A', se transforma de la siguiente manera:

$$\Rightarrow \begin{pmatrix} 29 & 38 & 23 \\ 38 & 74 & 69 \\ 23 & 69 & 82 \end{pmatrix} \longrightarrow \begin{pmatrix} 5.38516 & 7.05643 & 4.27100 \\ 7.05643 & 4.92004 & 7.89872 \\ 4.27100 & 7.89872 & 1.16996 \end{pmatrix}$$

donde:

$$L = \begin{pmatrix} 5.38516 & 0 & 0 \\ 7.05643 & 4.92004 & 0 \\ 4.27100 & 7.89872 & 1.16996 \end{pmatrix} \quad \text{y} \quad L^t = \begin{pmatrix} 5.38516 & 7.05643 & 4.27100 \\ 0 & 4.92004 & 7.89872 \\ 0 & 0 & 1.16996 \end{pmatrix}$$

Comprobación:

$$\underbrace{\begin{pmatrix} 5.38516 & 0 & 0 \\ 7.05643 & 4.92004 & 0 \\ 4.27100 & 7.89872 & 1.16996 \end{pmatrix}}_{L} \underbrace{\begin{pmatrix} 5.38516 & 7.05643 & 4.27100 \\ 0 & 4.92004 & 7.89872 \\ 0 & 0 & 1.16996 \end{pmatrix}}_{L^t} =$$

$$\underbrace{\begin{pmatrix} 28.9999 & 38.0000 & 23.0000 \\ 38.0000 & 74.0000 & 69.0000 \\ 23.0000 & 69.0000 & 82.0000 \end{pmatrix}}_{\approx A'}$$

Para obtener la solución del sistema de ecuaciones lineales (4.10) se empieza resolviendo el sistema (4.1); obteniéndose:

$$Ly = b' \qquad \begin{pmatrix} 5.38516 & 0 & 0 \\ 7.05643 & 4.92004 & 0 \\ 4.27100 & 7.89872 & 1.16996 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} 14 \\ 33 \\ 36 \end{pmatrix}$$

Se aplica la sustitución hacia adelante, a fin de obtener el valor de las incógnitas del vector y.

$$\Rightarrow \begin{array}{lllllll} 5.38516y_1 & & & & & = & 14 \\ 7.05643y_1 & + & 4.92004y_2 & & & = & 33 \\ 4.27100y_1 & + & 7.89872y_2 & + & 1.16996y_3 & = & 36 \end{array} \qquad \Rightarrow \begin{array}{lll} y_1 & = & 2.59974 \\ y_2 & = & 2.97866 \\ y_3 & = & 1.17004 \end{array}$$

Después se resuelve el sistema (4.2); obteniéndose:

$$L^t x = y \qquad \begin{pmatrix} 5.38516 & 7.05643 & 4.27100 \\ 0 & 4.92004 & 7.89872 \\ 0 & 0 & 1.16996 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} 2.59974 \\ 2.97866 \\ 1.17004 \end{pmatrix}$$

Finalmente, se aplica la sustitución hacia atrás, a fin de obtener el valor de las incógnitas del vector x.

$$\Rightarrow \begin{array}{lllllll} 5.38516x_1 & + & 7.05643x_2 & + & 4.27100x_3 & = & 0.44721 \\ & & 4.92004x_2 & + & 7.89872x_3 & = & 0.69027 \\ & & & & 1.16996x_3 & = & 1.3241 \end{array} \qquad \begin{array}{lll} x_1 & \approx & 1.00010 \\ x_2 & \approx & -1.00011 \\ x_3 & \approx & 1.00007 \end{array}$$

Comprobación:

$$\begin{array}{llllllll} 4(1.00010) & + & 3(-1.00011) & - & (1.00007) & = & 0.00000 \approx & 0 \\ 2(1.00010) & + & (-1.00011) & & & = & 1.00009 \approx & 1 \\ 3(1.00010) & + & 8(-1.00011) & + & 9(1.00007) & = & 4.00005 \approx & 4 \end{array}$$

4.4 Aplicaciones y ejercicios

Ejercicio 4.1 Resolver los siguientes sistemas de ecuaciones lineales por el método de Doolittle, con 6 cifras significativas.

a)
$$\begin{array}{rcrcrcr} \mathbf{x}_1 & + & \mathbf{x}_2 & + & \mathbf{x}_3 & = & 1 \\ 2\mathbf{x}_1 & - & \mathbf{x}_2 & + & 3\mathbf{x}_3 & = & 4 \\ 3\mathbf{x}_1 & + & 2\mathbf{x}_2 & - & 2\mathbf{x}_3 & = & -2 \end{array}$$

b)
$$\begin{array}{rcrcrcr} \mathbf{x}_1 & + & \mathbf{x}_2 & + & \mathbf{x}_3 & = & -2 \\ 2\mathbf{x}_1 & - & \mathbf{x}_2 & + & 3\mathbf{x}_3 & = & 5 \\ 3\mathbf{x}_1 & + & 2\mathbf{x}_2 & - & 2\mathbf{x}_3 & = & 1 \end{array}$$

c)
$$\begin{array}{rcrcrcr} \mathbf{x}_1 & + & \mathbf{x}_2 & + & \mathbf{x}_3 & = & 2 \\ 2\mathbf{x}_1 & - & \mathbf{x}_2 & + & 3\mathbf{x}_3 & = & -1 \\ 3\mathbf{x}_1 & + & 2\mathbf{x}_2 & - & 2\mathbf{x}_3 & = & 4 \end{array}$$

Solución:

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 2.00000 & 1 & 0 \\ 3.00000 & 0.333333 & 1 \end{pmatrix} \qquad U = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -3.00000 & 1.00000 \\ 0 & 0 & -5.33333 \end{pmatrix}$$

a)
$$\begin{array}{lcl} \mathbf{x}_1 & \approx & 0.249998 \\ \mathbf{x}_2 & \approx & -0.312500 \\ \mathbf{x}_3 & \approx & 1.06250 \end{array}$$

b)
$$\begin{array}{lcl} \mathbf{x}_1 & \approx & 2.00000 \\ \mathbf{x}_2 & \approx & -3.25000 \\ \mathbf{x}_3 & \approx & -0.750000 \end{array}$$

c)
$$\begin{array}{lcl} \mathbf{x}_1 & \approx & 0.250000 \\ \mathbf{x}_2 & \approx & 1.68750 \\ \mathbf{x}_3 & \approx & 0.0625004 \end{array}$$

Ejercicio 4.2 Comprobar que la siguiente matriz se puede factorizar por el método de Doolittle; y realizarla.

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 1 & 4 & 3 & 0 & 0 \\ 0 & 4 & 9 & 4 & 0 \\ 0 & 0 & 9 & 16 & 5 \\ 0 & 0 & 0 & 16 & 25 \end{pmatrix}$$

Solución:

$$|1| = 1 \neq 0, \qquad \begin{vmatrix} 1 & 2 \\ 1 & 4 \end{vmatrix} = 2 \neq 0, \qquad \begin{vmatrix} 1 & 2 & 0 \\ 1 & 4 & 3 \\ 0 & 4 & 9 \end{vmatrix} = 6 \neq 0$$

$$\begin{vmatrix} 1 & 2 & 0 & 0 \\ 1 & 4 & 3 & 0 \\ 0 & 4 & 9 & 4 \\ 0 & 0 & 9 & 16 \end{vmatrix} = 24 \neq 0, \qquad \begin{vmatrix} 1 & 2 & 0 & 0 & 0 \\ 1 & 4 & 3 & 0 & 0 \\ 0 & 4 & 9 & 4 & 0 \\ 0 & 0 & 9 & 16 & 5 \\ 0 & 0 & 0 & 16 & 25 \end{vmatrix} = 120 \neq 0$$

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 2 & 1 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 4 & 1 \end{pmatrix} \qquad U = \begin{pmatrix} 1 & 2 & 0 & 0 & 0 \\ 0 & 2 & 3 & 0 & 0 \\ 0 & 0 & 3 & 4 & 0 \\ 0 & 0 & 0 & 4 & 5 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

Ejercicio 4.3 Obtener la factorización LU de la siguiente matriz por el método de Doolittle.

$$\mathbf{A} = \begin{pmatrix} 2 & 4 & -1 & 0 \\ 4 & 10 & -1 & -1 \\ 6 & 10 & -7 & 1 \\ 0 & 2 & 1 & -2 \end{pmatrix}$$

Solución:

$$L = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ 3 & -1 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix} \qquad U = \begin{pmatrix} 2 & 4 & -1 & 0 \\ 0 & 2 & 4 & -1 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}$$

Ejercicio 4.4 Resolver el siguiente sistema de ecuaciones lineales por el método de Cholesky.

$$\begin{array}{rcrcrcl} x_1 & + & 5x_2 & + & 5x_3 & = & 3 \\ x_1 & + & 2x_2 & + & x_3 & = & 2 \\ x_1 & + & 2x_2 & + & 3x_3 & = & 1 \end{array}$$

Solución:

$$\mathbf{A}' = \mathbf{A}^{\mathbf{t}}\mathbf{A} = \begin{pmatrix} 1 & 1 & 1 \\ 5 & 2 & 2 \\ 5 & 1 & 3 \end{pmatrix}\begin{pmatrix} 1 & 5 & 5 \\ 1 & 2 & 1 \\ 1 & 2 & 3 \end{pmatrix} = \begin{pmatrix} 3 & 9 & 9 \\ 9 & 33 & 33 \\ 9 & 33 & 35 \end{pmatrix}$$

$$\mathbf{b}' = \mathbf{A}^{\mathbf{t}}\mathbf{b} = \begin{pmatrix} 1 & 1 & 1 \\ 5 & 2 & 2 \\ 5 & 1 & 3 \end{pmatrix}\begin{pmatrix} 3 \\ 2 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 21 \\ 20 \end{pmatrix}$$

$$\Rightarrow \begin{pmatrix} 3 & 9 & 9 \\ 9 & 33 & 33 \\ 9 & 33 & 35 \end{pmatrix} \longrightarrow \begin{pmatrix} 1.732051 & 5.196152 & 5.196152 \\ 5.196152 & 2.449490 & 2.449490 \\ 5.196152 & 2.449490 & 1.414214 \end{pmatrix}$$

donde:

$$L = \begin{pmatrix} 1.732051 & 0 & 0 \\ 5.196152 & 2.449490 & 0 \\ 5.196152 & 2.449490 & 1.414214 \end{pmatrix} \quad \text{y} \quad L^t = \begin{pmatrix} 1.732051 & 5.196152 & 5.196152 \\ 0 & 2.449490 & 2.449490 \\ 0 & 0 & 1.414214 \end{pmatrix}$$

$$\begin{array}{rcr} x_1 & \approx & 0.500000 \\ x_2 & \approx & 1.000000 \\ x_3 & \approx & -0.500000 \end{array}$$

Ejercicio 4.5 Una florista ofrece tres tamaños de arreglos florales que contienen rosas, margaritas y crisantemos. Cada arreglo pequeño contiene una rosa, tres margaritas y tres crisantemos. Cada arreglo mediano contiene dos rosas, cuatro margaritas y seis crisantemos. Cada arreglo grande contiene cuatro rosas, ocho margaritas y seis crisantemos. Un día la florista advierte que ha utilizado un total de 24 rosas, 50 margaritas y 48 crisantemos para preparar ordenes de estos tres tipos de arreglos. ¿Cuántos arreglos de cada tipo habrá hecho?. Resolver por el método de Crout.

Solución:
Sean x_1, x_2 y x_3, los tipos de arreglos pequeño, mediano y grande, respectivamente. Puesto que para un arreglo pequeño se ocupa una rosa, se necesita de un total de x_1 rosas por día. De manera similar necesitan un total de dos rosas para el arreglo mediano y 4 para el grande: $2x_2$ y $4x_3$. Puesto que se quiere acabar con las 24 rosas, se obtiene la siguiente ecuación:

$$x_1 + 2x_2 + 4x_3 = 24$$

Del mismo modo de obtienen las ecuaciones correspondientes para las margaritas y los crisantemos.

$$\begin{array}{ccccccc} 3x_1 & + & 4x_2 & + & 8x_3 & = & 50 \\ 3x_1 & + & 6x_2 & + & 6x_3 & = & 48 \end{array}$$

Así se tiene un sistema de tres ecuaciones lineales con tres variables. (det $A = 12$). Resolviendo por el método de Crout, se obtiene:

$$\Rightarrow \begin{pmatrix} 1 & 2 & 4 \\ 3 & 4 & 8 \\ 3 & 6 & 6 \end{pmatrix} \longrightarrow \begin{pmatrix} 1 & 2 & 4 \\ 3 & -2 & 2 \\ 3 & 0 & -6 \end{pmatrix}$$

donde:

$$L = \begin{pmatrix} 1 & 0 & 0 \\ 3 & -2 & 0 \\ 3 & 0 & -6 \end{pmatrix} \quad \text{y} \quad U = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\begin{array}{ccc} x_1 & = & 2 \\ x_2 & = & 3 \\ x_3 & = & 4 \end{array}$$

Es decir la florista realizó dos arreglos pequeños, tres medianos y cuatro grandes.

Conclusiones

Con la elaboración de éste material de apoyo se logra poner al alcance de los estudiantes de la licenciatura en Matemáticas Aplicadas y Computación, una guía para el estudio independiente a la bibliografía básica y complementaria que aparece en los programas de la asignatura, sin pretender en ningún momento que éste la sustituya; es simplemente un complemento en donde se explican de forma clara y sencilla los temas de la asignatura de Métodos Numéricos I del plan 2006.

Cabe mencionar que pensando en que se seguirán actualizando los planes de estudio; este material más adelante puede ser corregido, perfeccionado y actualizado sin mayores complicaciones, ajustándose a las nuevas modificaciones que vayan solicitando la implementación de nuevos planes de estudio.

Bibliografía

[1] *Akai, T.*, Métodos numéricos aplicados a la ingeniería, Limusa Wiley, México, 2000.
[2] *Burden R. L. y Faires J. D.*, Analisis numérico, International Thomson, México, 2002.
[3] *Chapra S. y Canale R.*, Métodos numéricos para ingenieros, McGraw Hill, México, 2003.
[4] *Grossman S.*, Álgebra lineal, McGraw Hill, México, 1996.
[5] *Kincaid D. y Cheney W.*, Analisis numérico. Las matemáticas del cálculo científico, Addison-Wesley Iberoamericana, Estados Unidos, 1994.
[6] *Leithold Louis*, El cálculo con geometría analítica, Harla, México, 1992.
[7] *Luthe R., Olivera A. y Schutz F.*, Métodos numéricos, Limusa, México,1986.
[8] *Maron M. y López R.*, Análisis Numérico. Un enfoque práctico, Cecsa, México,1995.
[9] *Mathews J. y Fink Kurtis*, Métodos numéricos con Matlab, Prentice Hall, España, 2000.
[10] *Nakamura S.*, Métodos numéricos aplicados con software, Pearson Education, México, 1992.
[11] *Nieves A. y Domínguez F.*, Métodos numéricos aplicadas a la Ingeniería, Cecsa, México, 1995.
[12] *Poole D.*, Álgebra lineal. Una introducción moderna, Thomson, México, 2004.
[13] *Wheatley y Gerald*, Métodos numéricos con aplicaciones, Prentice Hall, México, 2000.
[14] http://www.unalmed.edu.co/%7Eifasmar/ , Fecha de consulta: Septiembre 2006
[15] http://www.uv.es/ diaz/mn/fmn.html, Fecha de consulta: Septiembre 2006
[16] http://proton.ucting.udg.mx/posgrado/cursos/metodos/temario.html, Fecha de consulta: Septiembre 2006
[17] http://www.chillan.udec.cl/ jsandov/cnsel2.htm, Fecha de consulta: Septiembre 2006
[18] http://docentes.uacj.mx/gtapia/AN/default.htm, Fecha de consulta: Septiembre 2006
[19]http://www.monografias.com/trabajos18/sistemas-ecuaciones/sistemas-ecuaciones.shtml#sistemas, Fecha de consulta: Septiembre 2006
[20] http://ma1.eii.us.es/Material/Alg_Num_ii_Bol.pdf , Fecha de consulta: Septiembre 2006

[21]
http://www.dm.uba.ar/materias/elementos_calculo_numerico_M/2004/1/apunte.pdf, Fecha de consulta: Septiembre 2006
[22] http://www.escet.urjc.es/ matemati/mm_iq/tema4.pdf , Fecha de consulta: Septiembre 2006
[23] http://www.etsimo.uniovi.es/ antonio/uned/ieee754/formato.html, Fecha de consulta: Septiembre 2006