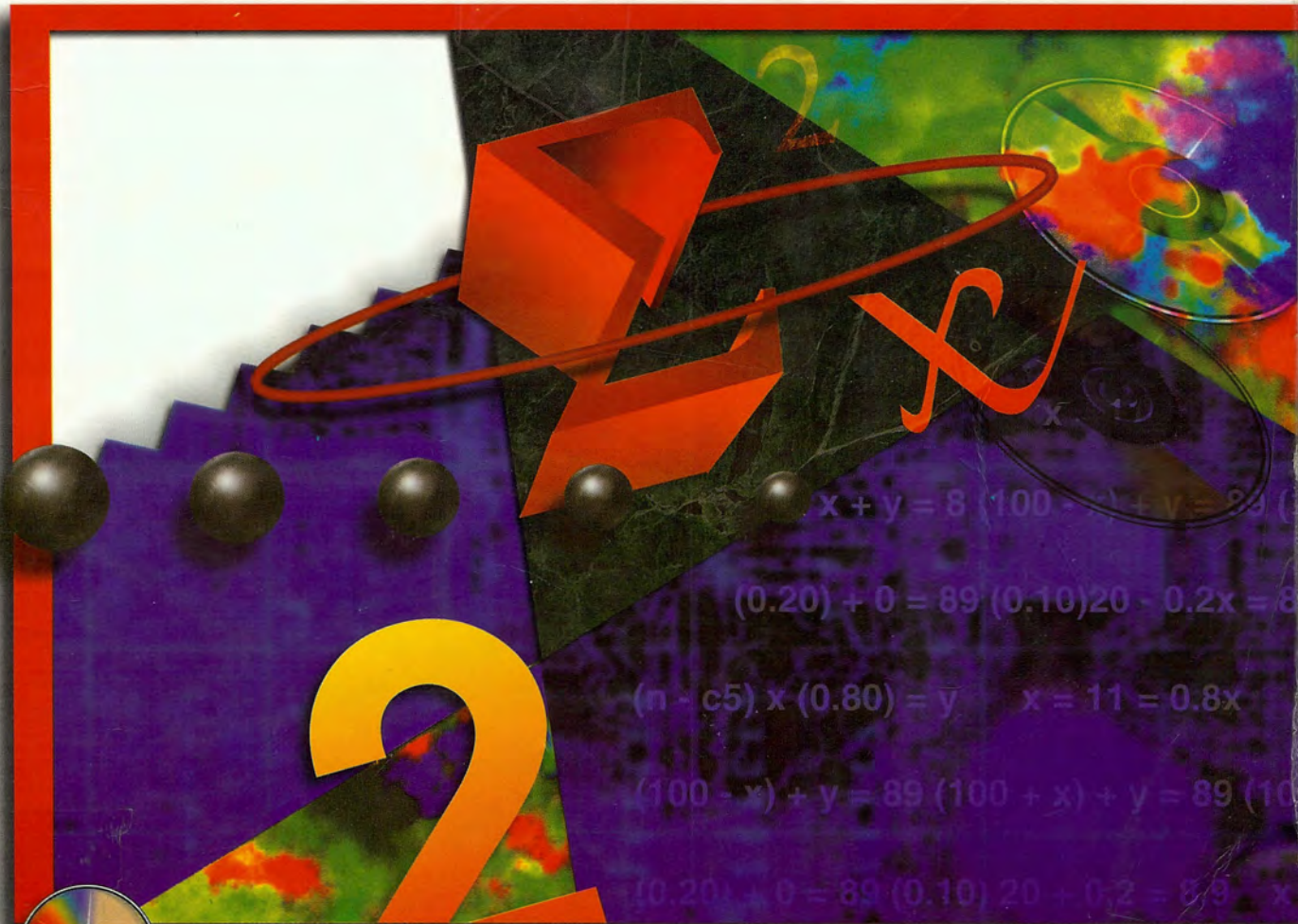


# MÉTODOS NUMÉRICOS

## APLICADOS A LA INGENIERÍA

Segunda edición



INCLUYE CD CON  
EJEMPLOS Y EJERCICIOS

**Antonio Nieves**  
**Federico C. Domínguez**

**CECSA**

# **Métodos Numéricos**

## **Aplicados a la Ingeniería**

**Antonio Nieves Hurtado**

**Federico C. Domínguez Sánchez**

*Profesores de la Academia de Matemáticas Aplicadas  
ESIQIE-IPN*

QUINTA REIMPRESIÓN  
MÉXICO, 2006

**COMPAÑÍA EDITORIAL CONTINENTAL**



Para establecer comunicación  
con nosotros puede hacerlo por:



correo:  
Renacimiento 180, Col. San Juan  
Tlihuaca, Azcapotzalco,  
02400, México, D.F.



fax pedidos:  
(01 55) 5561 4063 • 5561 5231



e-mail:  
info@patriacultural.com.mx



home page:  
www.patriacultural.com.mx

Dirección editorial: Javier Enrique Callejas  
Coordinación editorial: Elisa Pecina Rosas  
Diseño de interiores: Guillermo Rodríguez Luna  
Diseño de portada: Perla Alejandra López Romo

Colaboración especial:  
Dr. Guillermo Marroquín Suárez  
Profesor de la Academia de Matemáticas Aplicadas  
ESIQIE - Instituto Politécnico Nacional

Revisión técnica:  
M.C. José Luis Turriza  
Profesor de Matemáticas  
ESIME-IPN

*Métodos Numéricos, aplicados a la ingeniería*  
Derechos reservados respecto a la segunda edición:  
© 1995, 2002, Antonio Nieves Hurtado / Federico C. Domínguez Sánchez  
© 1995, COMPAÑÍA EDITORIAL CONTINENTAL, S.A. DE C.V.  
© 2002, GRUPO PATRIA CULTURAL, S.A. DE C.V.  
bajo el sello de Compañía Editorial Continental  
Renacimiento 180, Colonia San Juan Tlihuaca,  
Delegación Azcapotzalco, Código Postal 02400, México, D.F.

Miembro de la Cámara Nacional de la Industria Editorial  
Registro núm. 43

ISBN 970-24-0258-1 (segunda edición)  
(ISBN 968-26-1260-8 primera edición)

Queda prohibida la reproducción o transmisión total o parcial del contenido de la presente obra en cualesquiera formas, sean electrónicas o mecánicas, sin el consentimiento previo y por escrito del editor.

Impreso en México  
Printed in Mexico

**Primera edición: 1995**  
**Segunda edición: 2002**  
Cuarta reimpresión: 2005  
Quinta reimpresión: 2006

---

*A los Eggli:  
Violet (Mom), Fred,  
Josephine, Richard y David.*

*Gracias*

*Antonio*

*A mis hijos Alura,  
Alejandra y Federico,  
a mis hermanos, y a  
la memoria de mis padres.*

*Federico*



# CONTENIDO

|                                                                    |           |
|--------------------------------------------------------------------|-----------|
| <b>PREFACIO</b>                                                    | <b>xi</b> |
| <b>1 ERRORES</b>                                                   | <b>1</b>  |
| 1.1 Sistema numérico                                               | 2         |
| 1.2 Manejo de números en la computadora                            | 8         |
| 1.3 Errores                                                        | 11        |
| 1.4 Algoritmos y estabilidad                                       | 19        |
| Ejercicios                                                         | 20        |
| Problemas                                                          | 24        |
| <b>2 SOLUCIÓN DE ECUACIONES NO LINEALES</b>                        | <b>29</b> |
| 2.1 Método de punto fijo                                           | 30        |
| ALGORITMO 2.1 Método de punto fijo                                 | 35        |
| 2.2 Método de Newton-Raphson                                       | 44        |
| ALGORITMO 2.2 Método de Newton-Raphson                             | 47        |
| 2.3 Método de la secante                                           | 47        |
| ALGORITMO 2.3 Método de la secante                                 | 50        |
| 2.4 Método de posición falsa                                       | 51        |
| ALGORITMO 2.4 Método de posición falsa                             | 54        |
| 2.5 Método de la bisección                                         | 54        |
| 2.6 Problemas de los métodos de dos puntos y orden de convergencia | 56        |
| 2.7 Aceleración de convergencia                                    | 59        |
| ALGORITMO 2.5 Método de Steffensen                                 | 62        |
| 2.8 Búsqueda de valores iniciales                                  | 63        |
| 2.9 Raíces complejas                                               | 69        |
| ALGORITMO 2.6 Método de Müller                                     | 76        |
| 2.10 Polinomios y sus ecuaciones                                   | 77        |
| ALGORITMO 2.7 Método de Horner                                     | 79        |
| ALGORITMO 2.8 Método de Horner iterado                             | 82        |
| Ejercicios                                                         | 90        |
| Problemas                                                          | 117       |

|          |                                                                          |            |
|----------|--------------------------------------------------------------------------|------------|
| <b>3</b> | <b>MATRICES Y SISTEMAS DE ECUACIONES LINEALES</b>                        | <b>129</b> |
| 3.1      | <b>Matrices</b>                                                          | 130        |
|          | ALGORITMO 3.1 Multiplicación de matrices                                 | 136        |
| 3.2      | <b>Vectores</b>                                                          | 141        |
| 3.3      | <b>Independencia y ortogonalización de vectores</b>                      | 149        |
|          | ALGORITMO 3.2 Ortogonalización de Gram Schmidt                           | 159        |
| 3.4      | <b>Solución de sistemas de ecuaciones lineales</b>                       | 162        |
|          | ALGORITMO 3.3 Eliminación de Gauss                                       | 168        |
|          | ALGORITMO 3.4 Eliminación de Gauss con pivoteo                           | 172        |
|          | ALGORITMO 3.5 Método de Thomas                                           | 181        |
|          | ALGORITMO 3.6 Factorización directa                                      | 187        |
|          | ALGORITMO 3.7 Factorización con pivoteo                                  | 188        |
|          | ALGORITMO 3.8 Método de Doolittle                                        | 191        |
|          | ALGORITMO 3.9 Factorización de matrices simétricas                       | 193        |
|          | ALGORITMO 3.10 Método de Cholesky                                        | 196        |
| 3.5      | <b>Métodos iterativos</b>                                                | 206        |
|          | ALGORITMO 3.11 Métodos de Jacobi y Gauss-Seidel                          | 216        |
| 3.6      | <b>Valores y vectores propios</b>                                        | 222        |
|          | <b>Ejercicios</b>                                                        | 228        |
|          | <b>Problemas</b>                                                         | 238        |
| <b>4</b> | <b>SISTEMAS DE ECUACIONES NO LINEALES</b>                                | <b>255</b> |
| 4.1      | <b>Dificultades en la solución de sistemas de ecuaciones no lineales</b> | 256        |
| 4.2      | <b>Método de punto fijo multivariable</b>                                | 259        |
|          | ALGORITMO 4.1 Método de punto fijo multivariable                         | 265        |
| 4.3      | <b>Método de Newton-Raphson</b>                                          | 266        |
|          | ALGORITMO 4.2 Método de Newton-Raphson multivariable                     | 273        |
| 4.4      | <b>Método de Newton-Raphson modificado</b>                               | 275        |
|          | ALGORITMO 4.3 Método de Newton-Raphson modificado                        | 278        |
| 4.5      | <b>Método de Broyden</b>                                                 | 279        |
|          | ALGORITMO 4.4 Método de Broyden                                          | 283        |
| 4.6      | <b>Aceleración de convergencia</b>                                       | 283        |
|          | ALGORITMO 4.5 Método del descenso de máxima pendiente                    | 297        |
| 4.7      | <b>Método de Bairstow</b>                                                | 299        |
|          | <b>Ejercicios</b>                                                        | 304        |
|          | <b>Problemas</b>                                                         | 316        |

|          |                                                       |            |
|----------|-------------------------------------------------------|------------|
| <b>5</b> | <b>APROXIMACIÓN FUNCIONAL E INTERPOLACIÓN</b>         | <b>332</b> |
| 5.1      | <b>Aproximación polinomial simple e interpolación</b> | 325        |
|          | ALGORITMO 5.1 Aproximación polinomial simple          | 328        |
| 5.2      | <b>Polinomios de Lagrange</b>                         | 328        |
|          | ALGORITMO 5.2 Interpolación de polinomios de Lagrange | 333        |
| 5.3      | <b>Diferencias divididas</b>                          | 334        |
|          | ALGORITMO 5.3 Tabla de diferencias divididas          | 338        |
| 5.4      | <b>Aproximación polinomial de Newton</b>              | 338        |
|          | ALGORITMO 5.4 Interpolación polinomial de Newton      | 342        |
| 5.5      | <b>Polinomio de Newton en diferencias finitas</b>     | 343        |
| 5.6      | <b>Estimación de errores en la aproximación</b>       | 352        |
| 5.7      | <b>Aproximación polinomial segmentaria</b>            | 356        |
| 5.8      | <b>Aproximación polinomial con mínimos cuadrados</b>  | 362        |
|          | ALGORITMO 5.5 Aproximación con mínimos cuadrados      | 369        |
| 5.9      | <b>Aproximación multilínea con mínimos cuadrados</b>  | 370        |
|          | Ejercicios                                            | 373        |
|          | Problemas                                             | 382        |
| <b>6</b> | <b>INTEGRACIÓN Y DIFERENCIACIÓN NUMÉRICA</b>          | <b>393</b> |
| 6.1      | <b>Métodos de Newton Cotes</b>                        | 395        |
|          | ALGORITMO 6.1 Método trapezoidal compuesto            | 403        |
|          | ALGORITMO 6.2 Método de Simpson compuesto             | 406        |
| 6.2      | <b>Cuadratura de Gauss</b>                            | 415        |
|          | ALGORITMO 6.3 Cuadratura de Gauss-Lagrange            | 421        |
| 6.3      | <b>Integrales múltiples</b>                           | 422        |
|          | ALGORITMO 6.4 Integración doble por Simpson 1/3       | 428        |
| 6.4      | <b>Diferenciación numérica</b>                        | 429        |
|          | ALGORITMO 6.5 Derivación de polinomios de Lagrange    | 437        |
|          | Ejercicios                                            | 437        |
|          | Problemas                                             | 448        |
| <b>7</b> | <b>ECUACIONES DIFERENCIALES ORDINARIAS</b>            | <b>457</b> |
| 7.1      | <b>Formulación del problema de valor inicial</b>      | 459        |
| 7.2      | <b>Método de Euler</b>                                | 460        |
|          | ALGORITMO 7.1 Método de Euler                         | 463        |
| 7.3      | <b>Método de Taylor</b>                               | 463        |



|            |                                                                                                                                              |     |
|------------|----------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <b>7.4</b> | <b>Métodos de Euler modificado</b>                                                                                                           | 466 |
|            | ALGORITMO 7.2 Método de Euler modificado                                                                                                     | 468 |
| <b>7.5</b> | <b>Métodos de Runge-Kutta</b>                                                                                                                | 469 |
|            | ALGORITMO 7.3 Método de Runge-Kutta de cuarto orden                                                                                          | 473 |
| <b>7.6</b> | <b>Métodos de predicción-corrección</b>                                                                                                      | 474 |
|            | ALGORITMO 7.4 Método predictor-corrector                                                                                                     | 484 |
| <b>7.7</b> | <b>Ecuaciones diferenciales ordinarias de orden superior y sistemas de ecuaciones diferenciales ordinarias</b>                               | 485 |
|            | ALGORITMO 7.5 Método de Runge-Kutta de cuarto orden para un sistema de dos ecuaciones diferenciales ordinarias                               | 491 |
| <b>7.8</b> | <b>Formulación del problema de valores en la frontera</b>                                                                                    | 492 |
|            | Ejercicios                                                                                                                                   | 496 |
|            | Problemas                                                                                                                                    | 518 |
| <b>8</b>   | <b>ECUACIONES DIFERENCIALES PARCIALES</b>                                                                                                    | 527 |
| <b>8.1</b> | <b>Obtención de ecuaciones diferenciales parciales a partir de la modelación de fenómenos físicos (ecuación de calor y ecuación de onda)</b> | 528 |
| <b>8.2</b> | <b>Aproximación de derivadas por diferencias finitas</b>                                                                                     | 532 |
| <b>8.3</b> | <b>Solución de problemas de calor unidimensional</b>                                                                                         | 536 |
|            | ALGORITMO 8.1 Método explícito                                                                                                               | 541 |
|            | ALGORITMO 8.2 Método implícito                                                                                                               | 551 |
| <b>8.4</b> | <b>Convergencia (método explícito), estabilidad y consistencia</b>                                                                           | 553 |
| <b>8.5</b> | <b>Método de Crank-Nicholson</b>                                                                                                             | 556 |
|            | ALGORITMO 8.3 Método de Crank-Nicholson                                                                                                      | 560 |
| <b>8.6</b> | <b>Otros métodos para resolver el problema de conducción de calor en unidimensional</b>                                                      | 561 |
| <b>8.7</b> | <b>Solución de la ecuación de onda unidimensional</b>                                                                                        | 563 |
| <b>8.8</b> | <b>Tipos de condiciones frontera en procesos físicos y tratamientos de condiciones frontera irregulares</b>                                  | 569 |
|            | Ejercicios                                                                                                                                   | 573 |
|            | Problemas                                                                                                                                    | 579 |
|            | <b>RESPUESTAS A PROBLEMAS SELECCIONADOS</b>                                                                                                  | 585 |
|            | <b>ÍNDICE ANALÍTICO</b>                                                                                                                      | 597 |



# PREFACIO

## Objetivo del libro

El análisis numérico y sus métodos son una dialéctica entre el análisis matemático cualitativo y el análisis matemático cuantitativo. El primero nos dice, por ejemplo, que bajo ciertas condiciones algo existe, que es o no único, etcétera, mientras que el segundo complementa al primero, permitiendo calcular *aproximadamente* el valor de aquello que existe.

El análisis numérico es pues: una reflexión sobre los cursos tradicionales de cálculo, álgebra lineal, ecuaciones diferenciales entre otros, concretando en una serie de *métodos* o *algoritmos*, cuya característica principal es la posibilidad de obtener resultados *numéricos* de problemas matemáticos de cualquier tipo a partir de números y un número finito de operaciones aritméticas. La *finalidad* de este libro es el estudio y uso racional de dichos algoritmos en diferentes áreas de ingeniería y ciencias.

## Enfoque del libro

La noción de algoritmo es un concepto clásico en las matemáticas. Es un concepto muy anterior a la aparición de las computadoras y de las calculadoras. Por ejemplo, en el Papiro de Ahmes o de Rhind (de hacia el año 1650 a. C.) se encuentra la técnica de posición falsa aplicada a la solución de ecuaciones lineales y, en el *Jiuzhang suanshu* (el libro más famoso de la matemática china del año 200 a. C.) se resolvían sistemas de ecuaciones lineales con el método conocido hoy en día como eliminación de Gauss.

En realidad, en la enseñanza básica tradicional todos aprendimos algoritmos como el de la división, la multiplicación y la extracción de raíces cuadradas. Con el transcurso del tiempo, los dos primeros se convierten generalmente en las operaciones más conocidas y practicadas (aunque quizás también, en las más incomprensibles) y, el tercero, en la operación más fácilmente olvidada.

A fin de no caer en un curso más de recetas matemáticas desvinculadas y sin sentido, hemos desarrollado el material de este libro en torno a tres ideas fundamentales: el punto fijo, la eliminación de Gauss y la aproximación de funciones. Para instrumentarlas empleamos como recursos didácticos, en cada método o situación, diferentes sistemas de representación: el gráfico, el tabular y el algebraico, y promovemos el paso entre ellos. Con el fin de que el lector vea claramente la relación entre los métodos que estudia en el libro y su aplicación en el contexto real, se resuelven al final de cada capítulo alrededor de diez o más problemas de diferentes áreas de aplicación. De igual manera, hacemos énfasis en el uso de herramientas como la calculadora y la computadora así como la importancia de la visualización en los problemas. Dada la importancia de cada uno de estos aspectos los trataremos con cierto detalle a continuación.

# Los métodos numéricos y las herramientas computacionales

## COMPUTADORA

Dado que cada algoritmo implica numerosas operaciones lógicas, aritméticas y en múltiples casos graficaciones, la computadora es fundamental para el estudio de éstos. El bino-mio computadora-lenguaje de alto nivel (Fortran, Basic, C y otros) ha sido utilizado durante muchos años para la enseñanza y el aprendizaje de los métodos numéricos. Y si bien esta fórmula ha sido exitosa y sigue aún vigente, también es cierto que la aparición de paquetes comerciales como Mathcad, Maple, Matlab (por citar algunos de los más conocidos) permiten nuevos acercamientos al estudio de los métodos numéricos. Por ejemplo, han permitido que la programación sea más sencilla y rápida y han facilitado además la construcción directa de gráficas en dos y tres dimensiones, así como la exploración de conjeturas y la solución numérica directa de problemas matemáticos.

En respuesta a estas dos vertientes, se acompaña el libro con un CD donde se han mantenido los programas fuente de la primera edición (Fortran, Pascal y C) y se han incorporado programas en Visual Basic. En numerosos ejemplos, ejercicios y problemas utilizamos o sugerimos además el empleo de los paquetes mencionados arriba.

## CALCULADORAS GRAFICADORAS

Las calculadoras graficadoras (como la TI-89, TI-92, HP-48 o HP-49) disponen hoy en día de poderosos elementos como:

- a) Un sistema algebraico computarizado (CAS por sus siglas en inglés) que permite manipulaciones simbólicas y soluciones analíticas de problemas matemáticos.
- b) La graficación en dos y tres dimensiones con facilidades como el *zoom* y el *trace*.
- c) La posibilidad de resolver numéricamente problemas matemáticos.
- d) La posibilidad de programar y utilizar a través de dicha programación los recursos mencionados en los incisos anteriores, convirtiéndose así el conjunto lenguaje-recursos en una herramienta aún más poderosa que un lenguaje procedural como Basic o C.

Finalmente, su bajo costo, portabilidad y posibilidades de comunicación con sitios Web donde es posible actualizar, intercambiar y comprar programas e información, permiten plantear un curso de métodos numéricos sustentado en la calculadora o una combinación de calculadora y computadora. A fin de apoyar esta acción hemos incorporado en muchos de los ejemplos y ejercicios programas en la TI-92.

## Visualización

A raíz de las posibilidades gráficas que ofrecen las computadoras y las calculadoras, la visualización (un recurso natural del ser humano) ha tomado mayor importancia y se ha podido utilizar en las matemáticas de diferentes maneras como: en la aprehensión de los conceptos, en la solución de problemas, en la ilustración de los métodos y en general en darle un aspecto dinámico a diversas situaciones físicas. Así, hemos intentado aprovechar cada uno de estos aspectos y aplicarlos a lo largo del libro siempre que fue posible. Por ejemplo, en el capítulo 4, se presentan ilustraciones novedosas de los métodos para resolver sistemas de ecuaciones no lineales (inclusive se han puesto en color varias de esas gráficas).



ficas a fin de tener una mejor apreciación de las intersecciones de superficies y de las raíces), ilustraciones de conceptos abstractos como el criterio de convergencia del método de punto fijo univariable y la ponderación de pendientes en los métodos de Runge-Kutta. Además se incluyen varios ejercicios en Visual Basic donde se simula algún fenómeno como el de crecimiento de poblaciones (ejercicio 7.13), amortiguación en choques (ejercicio 7.11) y el desplazamiento de una cuerda vibrante (ejemplo 8.5). En estos últimos se pueden observar los resultados numéricos en tiempo real y la gráfica que van generando e incluso modificar los parámetros para hacer exploraciones propias. Todos ellos aparecen en el CD y se identifican con el icono correspondiente.

## **Prerrequisitos**

Generalmente los cursos de métodos numéricos siguen a los de cálculo en una variable, el de ecuaciones diferenciales ordinarias y el de programación. No obstante, consideramos sólo los cursos de cálculo y de programación como prerrequisitos. Los conocimientos de álgebra lineal requeridos así como los elementos básicos para estudiar las técnicas de la ecuaciones diferenciales parciales, se exponen en los capítulos correspondientes. Si bien los conceptos y técnicas analíticas de las ecuaciones diferenciales ordinarias serían benéficos y complementarios con los métodos de este curso, no son, sin embargo, material indispensable.

## **Secuencias sugeridas**

Como se dijo antes, el libro se desarrolla alrededor de tres ideas matemáticas fundamentales: punto fijo, eliminación de Gauss y aproximación de funciones. Las dos primeras se estudian en los capítulos 2 y 3, respectivamente, y junto con el capítulo 4 constituyen la parte algebraica del libro. Un primer curso (semestral) de métodos numéricos podría organizarse con los primeros cuatro capítulos del libro, seleccionando las secciones que correspondan a su programa de estudios o a las necesidades específicas del curso.

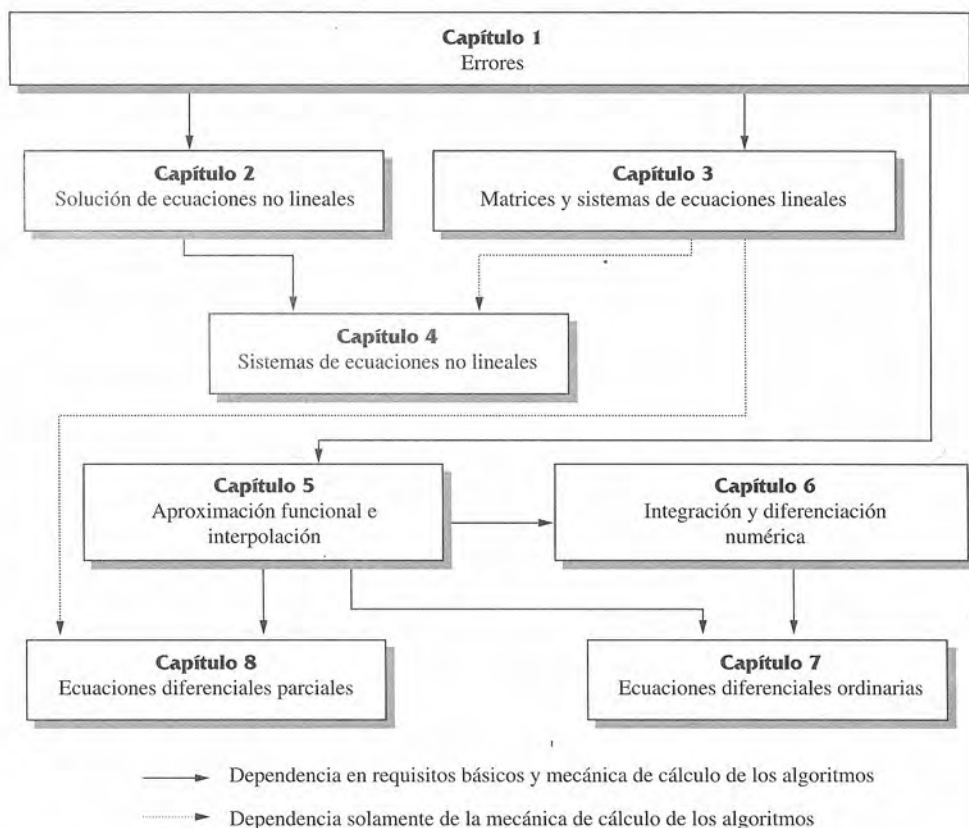
La tercera idea matemática clave en el libro es la de aproximación de funciones, la cual se presenta en el capítulo 5 y sustenta el material de análisis: integración y derivación numérica (capítulo 6), y más adelante será la base de la parte de dinámica: ecuaciones diferenciales ordinarias y parciales (capítulos 7 y 8, respectivamente). De este modo un curso semestral podría configurarse con los capítulos 1, 5, 6, 7 o bien 1, 5, 6, 7 y 8 (ver red de temas e interrelación).

Debido a esto y al hecho de que algunos tecnológicos y universidades sólo tiene un curso de un semestre de métodos numéricos, podría elaborarse éste con una secuencia como capítulos 1, 2, 3, 5 o bien 1, 2, 5, 6, por ejemplo.

Finalmente, recomendamos al maestro que cualquiera que sea la secuencia y las secciones elegidas en cada una de ellas, se discuta y trabaje con los ejemplos resueltos de final de cada capítulo.



## Red de temas e interrelación



## Cambios esenciales en esta edición

La mayoría de estos cambios responden a sugerencias de profesores que imparten métodos numéricos en diferentes instituciones del país.

- *A dónde nos dirigimos.* Se inicia cada capítulo con una introducción donde se describe brevemente qué estudiaremos, cómo lo vamos a hacer, qué relación guarda el material con el de los demás capítulos y, algunas veces, el tipo de problemas que pueden resolverse.
- *Guiones de Matlab.* Se incluyen en el libro y en la carpeta *Software* de cada capítulo del CD guiones de Matlab para distintos ejercicios, ejemplos y problemas.
- *Programas en la TI-92.* A lo largo del libro se dan programas para la TI-92.
- *Programas en Visual Basic.* A fin de aprovechar los aspectos visuales de los lenguajes actuales, en el CD se proporcionan programas en Visual Basic que pueden ser modificados para adaptarlos a otras situaciones.
- *Nuevos ejemplos, ejercicios y problemas.* Se han adicionado ejercicios y problemas de aplicación para darle mayor versatilidad al material.
- *Soluciones a ejemplos y ejercicios* en Matlab, Mathcad y Mathematica en el CD.
- *Sección de valores y vectores propios.* El material de valores y vectores propios que aparecía originalmente disperso se ha organizado como la sección 3.6.

- *Sección de problemas de valores a la frontera.* Se ha incorporado la sección 7.7 donde se estudian ecuaciones diferenciales ordinarias con valores a la frontera.
- *Método de Bairstow.* Se ha incorporado el método de Bairstow para encontrar raíces de ecuaciones polinomiales en la sección 4.7 como una aplicación de las técnicas de solución de ecuaciones no lineales.
- *Ecuación de onda unidimensional.* En la nueva sección 8.7 se resuelve la ecuación de onda de modo de trabajar con ecuaciones hiperbólicas y experimentar con el fenómeno de vibración de una cuerda.
- *Iconos utilizados en la segunda edición.* El libro se rediseñó íntegramente para facilitar su lectura. En particular, se incluyeron los iconos que aparecen a continuación para permitir al lector identificar con rapidez los nuevos apoyos con los que cuenta el libro.



Guiones de Matlab.



Programas para la calculadora TI-92.



Indica un programa en Visual Basic que se ha incluido en el CD y que le ayudan en la solución de ese ejercicio o ejemplo.



La solución se incluye en el CD (en Mathcad, Matlab y Mathematica).

## Materiales adicionales



**CD diseñado especialmente para la segunda edición con:**

- Programas fuente en Visual Basic y sus respectivos ejecutables que corren en Windows 95 o posterior para la solución de ejemplos y ejercicios.
- Documentos de Mathcad y guiones de Matlab. Los documentos en Mathcad permiten darle un sentido exploratorio a los métodos numéricos y los guiones segundos acceso a uno de los paquetes más poderosos para resolver problemas matemáticos.
- Algoritmos, descripción de los programas de cómputo y explicaciones detalladas de su uso.
- Ligas a sitios donde el lector encontrará tutoriales de Mathcad, Matlab y Mathematica, en los que podrá aprender a usar estos paquetes.
- Sugerencias de empleo de software comercial (Mathcad y Matlab, Mathematica) para resolver un gran número de ejemplos y ejercicios.

**www. Sitio Web con:**

- Actualizaciones del material del libro (nuevas soluciones, problemas, programas, etcétera).
- Registro del usuario a fin de recibir las actualizaciones del material del libro y los programas que se desarrollen más adelante (se están escribiendo los programas en Fortran 90 y en Visual C++), así como poder enviar sus propias sugerencias o establecer comunicación directa con los autores.
- Para tener acceso a este material, el lector sólo necesita entrar a la página de Grupo Patria Cultural (<http://www.patriacultural.com.mx/nieves/index.html>) e ingresar los datos de la obra.

## Agradecimientos

---

Esta obra tiene su origen en apuntes para los cursos de métodos numéricos en la carrera de Ingeniero Químico Industrial del Instituto Politécnico Nacional, desarrollados durante una estancia de año sabático en el Instituto Tecnológico de Celaya, y posteriormente, a raíz de un certamen organizado por el propio IPN, se convirtieron en una propuesta de libro que ganó el primer lugar en el Primer Certamen Editorial Politécnico en 1984. Desde entonces, con actualizaciones continuas, ha sido utilizado como texto para estos cursos en diferentes instituciones del país. Los autores agradecen al Instituto Politécnico Nacional la facilidad que otorgó para que la Editorial CECSA lo publicara.

Agradecemos también a las muchas personas que en distintas formas colaboraron para la realización de este libro. En especial al Dr. Guillermo Marroquín Suárez por el material, ideas y colaboración intensa que durante un año aportó a la elaboración de los capítulos 5 a 8; su valiosa ayuda permitió darle un enfoque interesante de aplicación al material matemático; sus observaciones a las soluciones enriquecieron el análisis de los ejercicios. Al Ing. Arturo Javier López García por el programa del método de Müller y al M. en C. Ramón Duarte Ramos por su asesoría en los problemas de vigas y columnas.

Nuestro agradecimiento especial al personal de la Editorial CECSA, y en particular a Elisa Pecina Rosas, nuestra editora, por su interés constante en lograr una óptima presentación técnica, de estilo y de diseño gráfico en el libro.



# CAPÍTULO 1

## ERRORES

### A dónde nos dirigimos

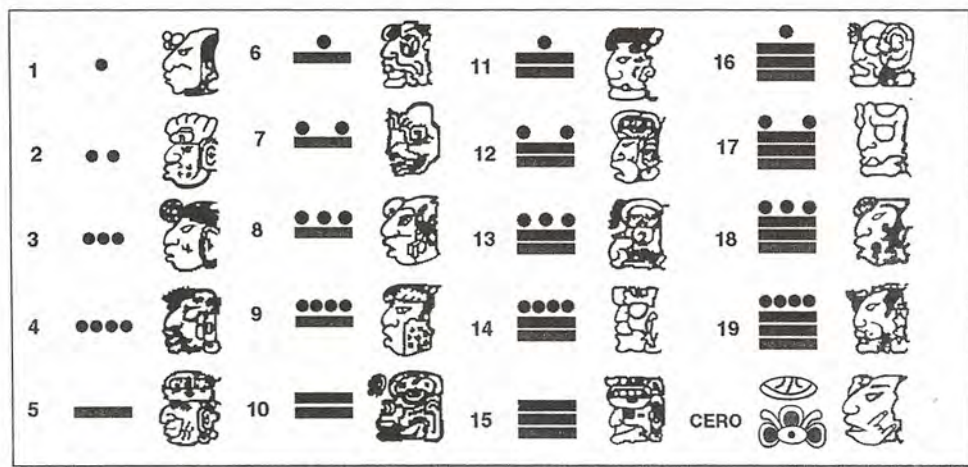
En este capítulo revisaremos tres de los sistemas numéricos posicionales más relevantes en el estudio de los métodos numéricos: binario, octal y decimal. Para esto analizaremos las conversiones entre ellos, la representación y manejo del sistema binario en la computadora, así como los diversos errores que ello puede ocasionar y algunas formas de evitarlos. Dada la naturaleza electrónica de las calculadoras y computadoras, los sistemas binario y octal resultan los más indicados a usarse en estos dispositivos; por lo que, a fin de tener una idea de los procesos numéricos internos en ellas, conviene hacer un estudio de tales sistemas y su conversión al decimal, ya que éste es finalmente nuestro medio de enlace con las máquinas.

Por un lado, dada la finitud de la palabra de memoria de las máquinas, es imposible representar a todos los números reales en ella. Así, números como  $\pi$ ,  $\sqrt{2}$ ,  $\frac{1}{3} = 0.333\dots$ , números muy pequeños\* (o muy grandes) se manejan usando números que son aproximaciones de ellos o simplemente no se manejan. Por otro lado, una de las características más sobresalientes de los métodos numéricos es el uso de los números reales en cálculos extensos. Cabe entonces preguntarse qué efecto tienen tales aproximaciones en los cálculos que hacemos con dichos números, en los resultados que obtenemos e incluso qué números reales pueden representarse exactamente en la computadora.

El conocimiento de todo esto nos ayudará a evitar cierto tipo de errores, analizar su propagación e incluso interpretar mejor los resultados dados por una máquina.

### Introducción

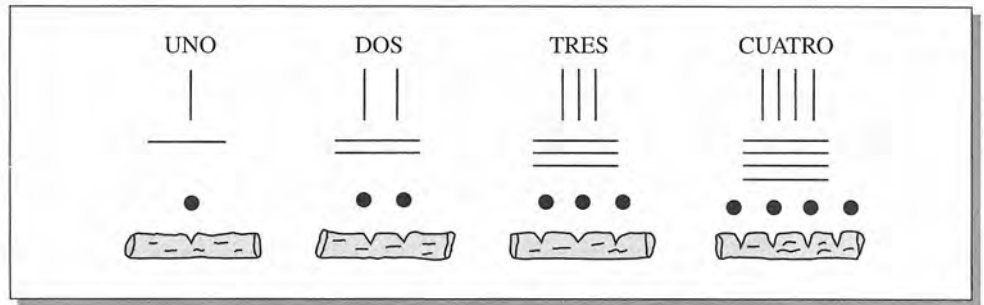
En la antigüedad, los números naturales se representaban con distintos tipos de símbolos o **numerales**. A continuación se presentan algunas muestras de numerales primitivos. (Fig. 1.1 y 1.2).



**Figura 1.1**  
Numerales  
usados por los  
mayas.

\* En estos casos la computadora envía un mensaje indicando que el número es muy pequeño (*underflow*) o muy grande (*overflow*) para su capacidad.

**Figura 1.2**  
Numerales  
primitivos.



En la figura 1.2 se puede observar que cada numeral es un conjunto de marcas **sencillas e iguales**. ¡Imagínese si así se escribiera el número de páginas del directorio telefónico de la Ciudad de México! No sería práctico por la enorme cantidad de tiempo y de espacio que requeriría tal sucesión de marcas iguales. Más aún, nadie podría reconocer, a primera vista, el número representado. Por ejemplo, ¿podría identificar rápidamente el siguiente numeral?

| | | | | | | | | | | | | | | | | = ?

Los antiguos egipcios evitaron algunos de los inconvenientes de los numerales representados por medio de marcas iguales, usando un solo jeroglífico o figura. Por ejemplo, en lugar de | | | | | | | | | |, usaron el símbolo  $\cap$ . Este jeroglífico representaba el hueso del talón. En la Fig. 1.3 se muestran otros numerales egipcios básicos relacionados con los del sistema decimal que les corresponden.

**Figura 1.3**  
Números  
egipcios  
antiguos.

| 1    | 10              | 100              | 1,000        | 10,000        | 100,000 | 1'000,000          |
|------|-----------------|------------------|--------------|---------------|---------|--------------------|
|      |                 |                  |              |               |         |                    |
| Raya | Hueso del talón | Cuerda enrollada | Flor de loto | Dedo señalado | Pez     | Hombre sorprendido |

## 1.1 Sistemas numéricos

### NUMERACIÓN CON BASE DOS (SISTEMA BINARIO)

Dado el siguiente conjunto de marcas simples e iguales

| | | | | | | | | | | ,

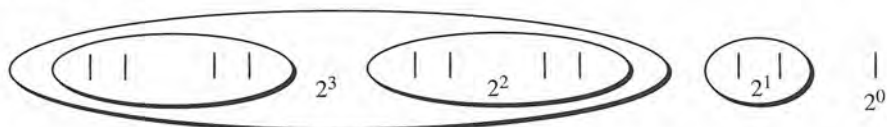
si se encierran en óvalos por parejas, a partir de la izquierda, se tiene.



A continuación, también empezando por la izquierda, se encierra cada par de óvalos en otro mayor.



Finalmente, se encierra cada par de óvalos en uno mayor todavía, comenzando también por la izquierda.



Nótese que el número de marcas dentro de cualquier óvalo es una potencia de 2.

El número representado por el numeral | | | | | | | | se

obtiene así  $2^3 + 2^1 + 2^0$ ,

o también  $(1 \times 2^3) + (1 \times 2^1) + (1 \times 2^0)$

Hay que observar que en esta suma no aparece  $2^2$ . Como  $0 \times 2^2 = 0$ , entonces la suma puede escribirse así

$$(1 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (1 \times 2^0)$$

Ahora puede formarse un nuevo símbolo para representar esta suma omitiendo los paréntesis, los signos de operación + y  $\times$ , y las potencias de 2, de la siguiente manera:

$$\begin{array}{cccc} (1 \times 2^3) & + & (0 \times 2^2) & + & (1 \times 2^1) & + & (1 \times 2^0) \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \text{Nuevo símbolo:} & 1 & 0 & 1 & 1 \end{array}$$

Ahora bien, ¿cómo interpretaremos este nuevo símbolo?

El significado de los números 1 en este nuevo símbolo depende del lugar que ocupan en el numeral. Así pues, el primero de derecha a izquierda representa una unidad; el segundo, un grupo de dos (o bien  $2^1$ ), el cuarto cuatro grupos de dos (8, o bien  $2^3$ ). El cero es el medio de asignarle a cada "1" su posición correcta. A los números o potencias de 2 que representan el "1" según su posición en el numeral, se les llama **valores de posición**; se dice que un sistema de numeración que emplea valores de posición es un **sistema posicional**.

El sistema de este ejemplo es un **sistema de base dos**, o sistema binario, porque emplea un grupo básico de dos símbolos: 0 y 1. Los símbolos "1" y "0" utilizados para escribir los numerales se denominan **dígitos binarios** o **bits**.

¿Qué número representa el numeral 101010<sub>dos</sub>?

(Se lee: "uno, cero, uno, cero, uno, cero, base dos").

Escribanse los valores de posición debajo de los dígitos:

|                     |       |       |       |       |       |                  |
|---------------------|-------|-------|-------|-------|-------|------------------|
| Dígitos binarios    | 1     | 0     | 1     | 0     | 1     | 0 <sub>dos</sub> |
| Valores de posición | $2^5$ | $2^4$ | $2^3$ | $2^2$ | $2^1$ | $2^0$            |

Multiplicando los valores de posición por los dígitos binarios correspondientes y sumándolos todos, se obtiene el equivalente en decimal.



$$101010_{\text{dos}} = (1 \times 2^5) + (0 \times 2^4) + (1 \times 2^3) + (0 \times 2^2) + (1 \times 2^1) + (0 \times 2^0) \\ = 42_{\text{diez}} \text{ (se lee: "cuatro, dos, base diez")}$$

El sistema de numeración más difundido en la actualidad es el **sistema decimal**. Es un sistema posicional que usa un grupo básico de diez (base diez).

Considérese por ejemplo el numeral  $582_{\text{diez}}$

|                     |                                                       |        |        |
|---------------------|-------------------------------------------------------|--------|--------|
| Dígitos decimales   | 5                                                     | 8      | 2      |
| Valores de posición | $10^2$                                                | $10^1$ | $10^0$ |
| Forma desarrollada  | $(5 \times 10^2) + (8 \times 10^1) + (2 \times 10^0)$ |        |        |

Al escribir números decimales se omite la palabra "diez" y se establece la convención de que un numeral con valor de posición, es un número decimal, sin necesidad de indicar la base. De ahí que siempre se anote 582 en lugar de  $582_{\text{diez}}$ .

El desarrollo y arraigo del sistema decimal, quizá se deba al hecho de tener siempre a la vista, los diez dedos de las manos. El sistema binario se emplea en las computadoras digitales, debido a que los alambres que forman los circuitos electrónicos presentan sólo dos estados: magnetizados o no magnetizados, ya sea que pase o no corriente por ellos.

### CONVERSIÓN DE NÚMEROS ENTEROS DEL SISTEMA DECIMAL A UN SISTEMA DE BASE $b$ Y VICEVERSA

Para convertir un número  $n$  del sistema decimal a un sistema con base  $b$ , se divide el número  $n$  entre la base  $b$  y se registra el cociente  $c_1$  y el residuo  $r_1$  resultantes; se divide  $c_1$  entre la base  $b$ , 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 residuo  $r_i$ . El número equivalente a  $n$  en el sistema con base  $b$  queda formado así:  $r_i \ r_{i-1} \ r_{i-2} \ \dots \ r_1$ .

#### Ejemplo 1.1

Convierta  $358_{10}$  al sistema octal.

#### Solución



La base del sistema octal\* es 8, por tanto

$$\begin{array}{rclcl} 358 & = & 8 & \times & 44 & + & 6 \\ & & & & c_1 & & r_1 \\ 44 & = & 8 & \times & 5 & + & 4 \\ & & & & c_2 & & r_2 \\ 5 & = & 8 & \times & 0 & + & 5 \\ & & & & c_3 & & r_3 \end{array}$$

Así que el número equivalente en octal es 546

\* El sistema octal usa un grupo básico de ocho símbolos: 0, 1, 2, 3, 4, 5, 6, 7.

**Ejemplo 1.2**Convierta  $358_{10}$  a binario (base 2).**Solución**

$$\begin{array}{rclcl}
 358 & = & 2 \times 179 & + & 0 \\
 179 & = & 2 \times 89 & + & 1 \\
 89 & = & 2 \times 44 & + & 1 \\
 44 & = & 2 \times 22 & + & 0 \\
 22 & = & 2 \times 11 & + & 0 \\
 11 & = & 2 \times 5 & + & 1 \\
 5 & = & 2 \times 2 & + & 1 \\
 2 & = & 2 \times 1 & + & 0 \\
 1 & = & 2 \times 0 & + & 1
 \end{array}$$

Por tanto  $358_{10} = 101100110_2$ 

Para convertir un entero  $m$  de un sistema con base  $b$  al sistema decimal, se multiplica cada dígito de  $m$  por la base  $b$  elevada a una potencia igual a la posición del dígito, tomando como posición cero la del dígito situado más a la derecha. De la suma resulta el equivalente decimal. Así

$$276_8 = 2 \times 8^2 + 7 \times 8^1 + 6 \times 8^0 = 190_{10}$$

$$\begin{aligned}
 1010001_2 &= 1 \times 2^6 + 0 \times 2^5 + 1 \times 2^4 + 0 \times 2^3 + 0 \times 2^2 \\
 &\quad + 0 \times 2^1 + 1 \times 2^0 = 81_{10}
 \end{aligned}$$

### CONVERSIÓN DE NÚMEROS ENTEROS DEL SISTEMA OCTAL AL BINARIO Y VICEVERSA

Dado un número del sistema octal, su equivalente en binario se obtiene sustituyendo cada dígito del número octal con los tres dígitos equivalentes del sistema binario.

| Base octal | Equivalente binario en tres dígitos |
|------------|-------------------------------------|
| 0          | 000                                 |
| 1          | 001                                 |
| 2          | 010                                 |
| 3          | 011                                 |
| 4          | 100                                 |
| 5          | 101                                 |
| 6          | 110                                 |
| 7          | 111                                 |

**Ejemplo 1.3**

Convierta  $546_8$  a binario.

**Solución**

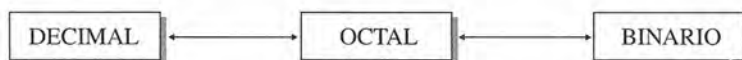
|     |     |     |
|-----|-----|-----|
| 5   | 4   | 6   |
| 101 | 100 | 110 |

Así que  $546_8 = 101100110_2$

Dado un número en binario, su equivalente en octal se obtiene formando ternas de dígitos, contando de derecha a izquierda y sustituyendo cada terna por su equivalente en octal. Así  
Convertir  $10011001_2$  a octal

|     |     |                  |                                |
|-----|-----|------------------|--------------------------------|
| 010 | 011 | 001 <sub>2</sub> | Por tanto $10011001_2 = 231_8$ |
| 2   | 3   | 1                |                                |

Dado que la conversión de octal a binario es simple y la de decimal a binario resulta muy tediosa, se recomienda usar la conversión a octal como paso intermedio al convertir un número decimal a binario.



Las flechas tienen dos sentidos porque es válido en ambas direcciones lo dicho.

**Ejemplo 1.4**

Convierta  $101100110_2$  a decimal.

**Solución**

a) Conversión directa

$$101100110_2 = 1 \times 2^8 + 0 \times 2^7 + 1 \times 2^6 + 1 \times 2^5 + 0 \times 2^4 + 0 \times 2^3 + 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0 = 358_{10}$$

b) Usando la conversión a octal como paso intermedio:

|                       |     |     |     |
|-----------------------|-----|-----|-----|
| 1) Conversión a octal | 101 | 100 | 110 |
|                       | 5   | 4   | 6   |

Por tanto  $101100110_2 = 546_8$

2) Conversión de octal a decimal

$$546_8 = 5 \times 8^2 + 4 \times 8^1 + 6 \times 8^0 = 358_{10}$$

### CONVERSIÓN DE NÚMEROS FRACCIONARIOS DEL SISTEMA DECIMAL A UN SISTEMA CON BASE $b$

Para convertir un número  $x_{10}$  fraccionario a un número con base  $b$ , se multiplica dicho número por la base  $b$ ; el resultado tiene una parte entera  $e_1$  y una parte fraccionaria  $f_1$ . Se multiplica ahora  $f_1$  por  $b$  y se obtiene un nuevo producto con parte entera  $e_2$  y fraccionaria  $f_2$ . Este procedimiento se repite un número suficiente de veces o hasta que se presenta  $f_i = 0$ . El equivalente de  $x_{10}$  con base  $b$  queda así  $0.e_1 e_2 e_3 e_4 \dots$



**Ejemplo 1.5**Convierta  $0.2_{10}$  a octal y binario.**Solución**

a) Conversión a octal

|            |            |            |            |            |
|------------|------------|------------|------------|------------|
| 0.2        | 0.6        | 0.8        | 0.4        | 0.2        |
| $\times 8$ | $\times 8$ | $\times 8$ | $\times 8$ | $\times 8$ |
| 1.6        | 4.8        | 6.4        | 3.2        | 1.6        |
| $e_1 f_1$  | $e_2 f_2$  | $e_3 f_3$  | $e_4 f_4$  | $e_5 f_5$  |

Después de  $e_4$  se van a repetir el  $e_1 e_2 e_3 e_4$  indefinidamente, por lo que  $0.2_{10} = 0.14631463\dots_8$

b) Conversión a binario

|            |            |            |            |            |
|------------|------------|------------|------------|------------|
| 0.2        | 0.4        | 0.8        | 0.6        | 0.2        |
| $\times 2$ | $\times 2$ | $\times 2$ | $\times 2$ | $\times 2$ |
| 0.4        | 0.8        | 1.6        | 1.2        | 0.4        |
| $e_1 f_1$  | $e_2 f_2$  | $e_3 f_3$  | $e_4 f_4$  | $e_5 f_5$  |

Igual que en el inciso a), después de  $e_4$  se repite  $e_1 e_2 e_3 e_4$  indefinidamente, por lo que  $0.2_{10} = 0.001100110011\dots_2$

Obsérvese que  $0.2_{10}$  pudo convertirse en binario simplemente tomando su equivalente en octal, y sustituyendo cada número por su terna equivalente en binario. Así

$$0.2_{10} = 0.1 \quad 4 \quad 6 \quad 3 \quad 1 \quad 4 \quad 6 \quad 3 \\ 0.001 \quad 100 \quad 110 \quad 011 \quad 001 \quad 100 \quad 110 \quad 011$$

y

$$0.2_{10} = 0.001100110011001100110011\dots_2$$

De lo anterior se puede observar que

$$358.2_{10} = 101100110.001100110011001100110011\dots_2$$

y cualquier número con parte entera y fraccionaria puede pasarse a otro sistema, cambiando su parte entera y fraccionaria de manera independiente, y al final integrándolos.



Para convertir números decimales enteros y fraccionarios a base 2, 3, ..., 9 puede usar el **PROGRAMA 1.1** del CD.

### CONVERSIÓN DE UN NÚMERO FRACCIONARIO EN SISTEMA BINARIO A SISTEMA DECIMAL

El procedimiento es similar al caso de números enteros, sólo hay que tomar en cuenta que la posición inicia con  $-1$ , a partir del punto.

**Ejemplo 1.6**Convierta  $0.010101110_2$  a decimal**Solución**

$$\begin{aligned} 0.010101110 &= 0 \times 2^{-1} + 1 \times 2^{-2} + 0 \times 2^{-3} + 1 \times 2^{-4} \\ &\quad + 0 \times 2^{-5} + 1 \times 2^{-6} + 1 \times 2^{-7} + 1 \times 2^{-8} + 0 \times 2^{-9} \\ &= 0.33984375_{10} \end{aligned}$$

**Ejemplo 1.7**Convierta  $0.010101110_2$  a decimal.**Solución** a) Conversión a octal

$$\begin{array}{ccc} 0.010 & 101 & 110 \\ 2 & 5 & 6 \end{array}$$

y

$$0.010101110_2 = 0.256_8$$

b) Conversión a decimal

$$0.256_8 = 2 \times 8^{-1} + 5 \times 8^{-2} + 6 \times 8^{-3} = 0.33984375_{10}$$

## 1.2 Manejo de números en la computadora

Por razones prácticas, sólo puede manejarse una cantidad finita de bits para cada número en una computadora, y esta cantidad o longitud varía de una máquina a otra. Por ejemplo, cuando se realizan cálculos de ingeniería y ciencias, es mejor trabajar con una longitud grande; por otro lado, una longitud pequeña es más económica y útil para cálculos y procesamiento administrativos.

Para una computadora dada, el número de bits generalmente se llama palabra. Las palabras van desde ocho bits hasta 64 bits. Para facilitar su manejo, la palabra se divide en partes más cortas denominadas bytes; por ejemplo, una palabra de 32 bits puede dividirse en cuatro bytes (ocho bits cada uno).

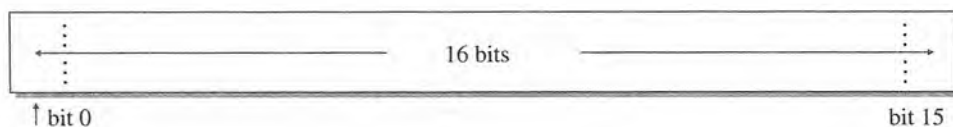
### NÚMEROS ENTEROS

Cada palabra, cualquiera que sea su longitud, almacena un número, aunque en ciertas circunstancias se usan varias palabras para contener un número. Por ejemplo, considérese una palabra de 16 bits para almacenar números enteros. De los 16 bits, el primero representa el signo del número; un cero es un signo más y un uno un signo menos. Los 15 bits restantes pueden usarse para guardar números binarios desde 000000000000000 hasta 111111111111111 (véase figura 1.1). Al convertir este número en decimal se obtiene

$$(1 \times 2^{14}) + (1 \times 2^{13}) + (1 \times 2^{12}) + \dots + (1 \times 2^1) + (1 \times 2^0)$$

que es igual a  $32767 (2^{15} - 1)$ . Por tanto, cada palabra de 16 bits puede contener un número cualquiera del intervalo  $-32768$  a  $+32767$  (véase Prob. 1.10).

**Figura 1.4**  
Esquema de una palabra de 16 bits para un número entero.



**Ejemplo 1.8**

Represente el número  $-26$  en una palabra de 16 bits.

**Solución**  $-26_{10} = -11010_2$ , y su almacenamiento en una palabra de 16 bits quedaría así

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**Ejemplo 1.9**

Represente el número  $525_{10}$ , en una palabra de 16 bits.

**Solución**  $525_{10} = 1015_8 = 1000001101_2$ , y su almacenamiento quedaría así

|   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

**NÚMEROS REALES (PUNTO FLOTANTE)**

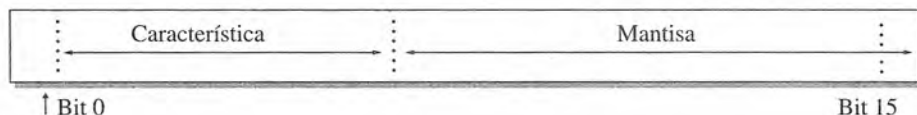
Cuando se desea almacenar un número real, se emplea en su representación binaria, llamada de punto flotante, la notación

$$0.d_1d_2d_3d_4d_5d_6d_7d_8 \times 2^{d'_1d'_2d'_3d'_4d'_5d'_6d'_7}$$

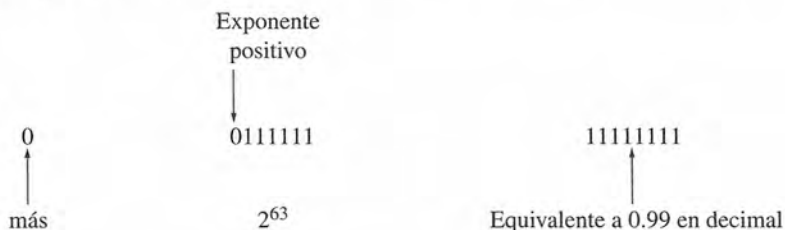
donde  $d_1 = 1$  y  $d_i$  y  $d'_j$  con  $i = 2, \dots, 8$  y  $j = 1, 2, \dots, 7$  pueden ser ceros o unos, y se guarda en una palabra como se muestra en la figura 1.2

**Figura 1.5**

Esquema de una palabra de 16 bits para un número de punto flotante.



Igual que antes, el bit cero se usa para guardar el signo del número. En los bits del uno al siete se almacena el exponente de la base 2 y en los ocho bits restantes la fracción.\* Según el lenguaje de los logaritmos, la fracción es llamada **mantisa** y el exponente **característica**. El número mayor que puede guardarse en una palabra de 16 bits usando la notación de punto flotante es



\* El exponente es un número binario de seis dígitos, ya que el bit uno se emplea para su signo. En algunas computadoras el exponente se almacena en base ocho (octal) o base 16 (hexadecimal) en lugar de base 2.



y los números que se pueden guardar en punto flotante binario van de alrededor de  $2^{-64}$  (si la característica es negativa) a cerca de  $2^{63}$ ; en decimal, de  $10^{-19}$  a cerca de  $10^{18}$  en magnitud (incluyendo números positivos, negativos y cero).

### Ejemplo 1.10

El número decimal  $-125.32$  que en binario es

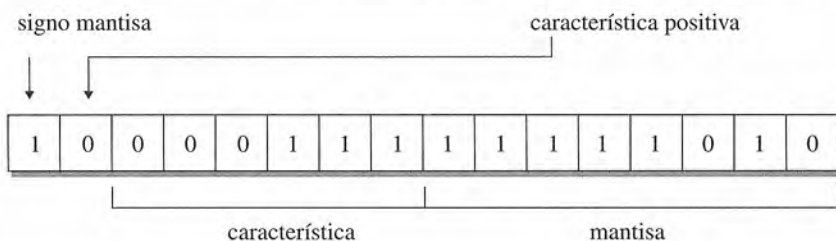
$$-1111101.010100011110101,$$

normalizado queda así

$$-.1111101010100011110101 \times 2^{+111}$$

bits truncados en el almacenamiento

y la palabra de memoria de 16 bits donde se almacena este valor quedaría como



Nótese que primero se normaliza el número, después se almacenan los primeros ocho bits y se truncan los restantes.

El número decimal  $+0.2$ , que en binario es

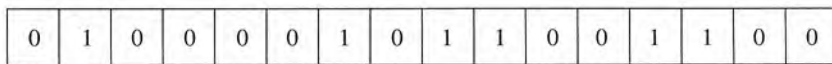
$$0.0011001100110011\dots$$

y que normalizado queda

$$.1100110011001100\dots \times 2^{-10}$$

bits truncados

se almacena así



### DOBLE PRECISIÓN

La doble precisión es un esfuerzo para aumentar la exactitud de los cálculos adicionando más bits a la mantisa. Esto se hace al utilizar dos palabras, la primera en la forma expuesta anteriormente, y los bits de la segunda para aumentar la mantisa de la primera. Entonces, con una palabra de 16 bits puede usarse en doble precisión una mantisa de  $8 + 16 = 24$  bits. Los 24 bits de la mantisa permiten expresar alrededor de 7 dígitos de exactitud en un número decimal, en lugar de 3 de la precisión sencilla.

La desventaja del uso de la doble precisión es que se emplean más palabras, con lo cual se acrecenta el uso de memoria por un programa.

## ERROR DE REDONDEO

Para finalizar esta sección, se analizarán brevemente algunas consecuencias de utilizar el sistema binario y una longitud de palabra finita.

Como no es posible guardar un número binario de longitud infinita o un número de más dígitos de los que posee la mantisa de la computadora que se está empleando, se almacena sólo un número finito de estos dígitos; como consecuencia, se comete automáticamente un pequeño error, conocido como error de redondeo, que al repetirse muchas veces puede llegar a ser considerable. Por ejemplo, si se desea guardar la fracción decimal 0.0001 que en binario es la fracción infinita

$$0.000000000000011010001101101110001011101011000...$$

quedaría, después de normalizarse, almacenado en una palabra de 16 bits como

$$.11010001 \times 2^{-1101}$$

Si se desea sumar el número 0.0001 con él mismo diez mil veces, usando una computadora, naturalmente que no se esperará obtener 1 como resultado, ya que los números que se adicionen no serían realmente 0.0001 sino valores aproximados a él (véase Prob. 1.16).

## 1.3 Errores

### ERROR ABSOLUTO, ERROR RELATIVO Y ERROR EN POR CIENTO

Si  $p^*$  es una aproximación a  $p$ , el error se define como

$$E = p^* - p$$

Sin embargo, para facilitar el manejo y el análisis se emplea el error absoluto definido como

$$EA = |p^* - p|$$

y el error relativo como

$$ER = \frac{|p^* - p|}{p}, \text{ si } p \neq 0$$

y como por ciento de error a

$$ERP = ER \times 100$$

En otros libros las definiciones pueden ser diferentes; por ejemplo, algunos autores definen el error  $E$  como  $p - p^*$ ; por tanto, sugerimos que al consultar las distintas bibliografías se busquen las definiciones de error dadas.

#### Ejemplo 1.11

Suponga que el valor para un cálculo debería ser

$$p = 0.10 \times 10^2 \text{ pero se obtuvo el resultado } p^* = 0.08 \times 10^2, \text{ entonces}$$

$$EA = |0.08 \times 10^2 - 0.10 \times 10^2| = 0.2 \times 10^1$$

$$ER = \frac{|0.08 \times 10^2 - 0.10 \times 10^2|}{0.10 \times 10^2} = 0.2 \times 10^0$$

$$ERP = ER \times 100 = 20\%$$

Por lo general, interesa el error absoluto y no el error relativo; pero cuando el valor exacto de una cantidad es “muy pequeño” o “muy grande”, los errores relativos son más significativos.

Por ejemplo si

$$p = 0.24 \times 10^{-4} \text{ y } p^* = 0.12 \times 10^{-4},$$

entonces:

$$EA = |0.12 \times 10^{-4} - 0.24 \times 10^{-4}| = 0.12 \times 10^{-4}$$

Sin reparar en las cantidades que se comparan, puede pensarse que el error absoluto es muy pequeño y, lo más grave, aceptar  $p^*$  como una buena aproximación a  $p$ .

Si, por otro lado, se calcula el error relativo

$$ER = \frac{|0.12 \times 10^{-4} - 0.24 \times 10^{-4}|}{0.24 \times 10^{-4}} = 0.5 \times 10^0$$

se observa que la “aproximación” es tan sólo la mitad del valor verdadero y, por tanto, está muy lejos de ser aceptable como aproximación a  $p$ . Finalmente

$$ERP = 50\%$$

De igual manera puede observarse que si

$$p = 0.46826564 \times 10^6 \text{ y } p^* = 0.46830000 \times 10^6,$$

entonces:

$$EA = 0.3436 \times 10^2,$$

y si de nueva cuenta no se toman en consideración las cantidades en cuestión, puede creerse que el  $EA$  es muy grande y que se tiene una mala aproximación a  $p$ . Sin embargo, al calcular el error relativo

$$ER = 0.7337715404 \times 10^{-4},$$

se advierte que el error es muy pequeño, como en realidad ocurre.

**ADVERTENCIA:** Cuando se manejan cantidades “muy grandes” o “muy pequeñas”, el error absoluto puede ser engañoso, mientras que el error relativo es más significativo en esos casos.

### DEFINICIÓN

Se dice que el número  $p^*$  aproxima a  $p$  con  $t$  dígitos significativos si  $t$  es el entero más grande no negativo para el cual se cumple

$$\frac{|p^* - p|}{p} < 5 \times 10^{-t}$$



Supóngase, por ejemplo, el número 10. Para que  $p^*$  aproxime a 10 con dos cifras significativas, usando la definición,  $p^*$  debe cumplir con

$$\begin{aligned}\frac{|p^* - 10|}{10} &< 5 \times 10^{-2} \\ -5 \times 10^{-2} &< \frac{p^* - 10}{10} < 5 \times 10^{-2} \\ 10 - 5 \times 10^{-1} &< p^* < 5 \times 10^{-1} + 10 \\ 9.5 &< p^* < 10.5\end{aligned}$$

esto es, cualquier valor de  $p^*$  en el intervalo (9.5, 10.5) cumple la condición.

En general para  $t$  dígitos significativos

$$\begin{aligned}\frac{|p^* - p|}{p} &< 5 \times 10^{-t} \quad \text{si } p > 0 \\ |p^* - p| &< 5 p \times 10^{-t} \\ p - 5 p \times 10^{-t} &< p^* < p + 5 p \times 10^{-t}\end{aligned}$$

Si, por ejemplo,  $p = 1000$  y  $t = 4$

$$\begin{aligned}1000 - 5 \times 1000 \times 10^{-4} &< p^* < 1000 + 5 \times 1000 \times 10^{-4} \\ 999.5 &< p^* < 1000.5\end{aligned}$$

## CAUSAS DE ERRORES GRAVES EN COMPUTACIÓN

Existen muchas causas de errores en la ejecución de un programa de cómputo, de las cuales se discutirán ahora algunas de las más serias. Para esto, vamos a pensar en una computadora imaginaria que trabaja con números en el sistema decimal, en forma tal que tiene una mantisa de cuatro dígitos decimales, y una característica de dos dígitos decimales, el primero de los cuales es usado para el signo. Sumados estos seis al bit empleado para el signo del número, se tendrá una palabra de siete bits. Los números que se van a guardar deben normalizarse primero en la siguiente forma

$$\begin{aligned}3.0 &= .3000 \times 10^1 \\ 7956000 &= .7956 \times 10^7 \\ -0.0000025211 &= -.2521 \times 10^{-5}\end{aligned}$$

Empleando esta computadora imaginaria, podemos estudiar algunos de los errores más serios que se cometen en su empleo.

### a) Suma de números muy distintos en magnitud

Vamos a suponer que se trata de sumar 0.002 a 600 en la computadora decimal imaginaria.

$$\begin{aligned}0.002 &= .2000 \times 10^{-2} \\ 600 &= .6000 \times 10^3\end{aligned}$$

Estos números *normalizados* no pueden sumarse directamente y, por tanto, la computadora debe desnormalizarlos antes de efectuar la suma.

$$\begin{array}{r} .000002 \times 10^3 \\ + .600000 \times 10^3 \\ \hline .600002 \times 10^3 \end{array}$$

Como sólo puede manejar cuatro dígitos, los últimos dos son eliminados y la respuesta es  $.6000 \times 10^3$  o 600. Por el resultado, la suma nunca se realizó.

Este tipo de errores, cuyo origen es el redondeo, es muy común y se recomienda, de ser posible, no sumar o restar dos números muy diferentes (véase ejercicio 1.2).

### b) Resta de números casi iguales

Supóngase que la computadora decimal va a restar 0.2144 de 0.2145.

$$\begin{array}{r} .2145 \times 10^0 \\ - .2144 \times 10^0 \\ \hline .0001 \times 10^0 \end{array}$$

Como la mantisa de la respuesta está desnormalizada, la computadora automáticamente la normaliza y el resultado se almacena como  $.1000 \times 10^{-3}$ .

Hasta aquí no hay error, pero en la respuesta sólo hay un dígito significativo; por tanto, se sugiere no confiar en su exactitud, ya que un pequeño error en alguno de los números originales produciría un error relativo muy grande en la respuesta de un problema que involucrara este error, como se ve a continuación.

Supóngase que la siguiente expresión aritmética es parte de un programa

$$X = (A - B) * C$$

Considérese ahora que los valores de  $A$ ,  $B$  y  $C$  son

$$A = 0.2145 \times 10^0, \quad B = 0.2144 \times 10^0, \quad C = 0.1000 \times 10^5$$

Al efectuarse la operación se obtiene el valor de  $X = 1$ , que es correcto. Sin embargo, supóngase que  $A$  fue calculada en el programa con un valor de  $0.2146 \times 10^0$  (error absoluto 0.0001, error relativo 0.00046 y  $ERP = 0.046\%$ ). Usando este valor de  $A$  en el cálculo de  $X$ , se obtiene como respuesta  $X = 2$ . Un error de 0.046% de pronto provoca un error de 100%. Aun más, este error puede pasar desapercibido.

### c) Overflow y Underflow

Con frecuencia una operación aritmética con dos números válidos da como resultado un número tan grande o tan pequeño que la computadora no puede manejarlo; como consecuencia se produce un *overflow* o un *underflow*, respectivamente.

Por ejemplo, al multiplicar  $0.5000 \times 10^8$  por  $0.2000 \times 10^9$  se tiene

$$\begin{array}{r} 0.5000 \times 10^8 \\ \times 0.2000 \times 10^9 \\ \hline 0.1000 \times 10^{17} \end{array}$$

Cada uno de los números que se multiplican puede guardarse en la palabra de la computadora imaginaria; sin embargo, su producto es muy grande y no puede almacenarse porque la característica requiere tres dígitos. Entonces se dice que ha llevado a cabo un *overflow*.

Otro caso de *overflow* puede ocurrir en la división; por ejemplo

$$\frac{2000000}{0.000005} = \frac{0.2000 \times 10^7}{0.5000 \times 10^{-5}} = 0.4000 \times 10^{12}$$

Las computadoras por lo general reportan esta circunstancia con un mensaje que varía dependiendo de cada máquina.

El *underflow* puede aparecer en la multiplicación o división, y por lo general no es tan serio como el *overflow*; las computadoras casi nunca envían mensaje de *underflow*. Por ejemplo:

$$(0.3000 \times 10^{-5}) \times (0.02000 \times 10^{-3}) = 0.006 \times 10^{-8} = 0.6000 \times 10^{-10}$$

Como el exponente  $-10$  está excedido en un dígito, no puede guardarse en la computadora y este resultado se expresa como valor cero. Este error expresado como error relativo es muy pequeño y a menudo no es serio. No obstante, puede ocurrir, por ejemplo:

$$A = 0.3000 \times 10^{-5}, \quad B = 0.0200 \times 10^{-3}, \quad C = 0.4000 \times 10^7,$$

y que se desee en algún punto del programa calcular el producto de  $A$ ,  $B$  y  $C$

$$X = A * B * C$$

Se multiplican primero  $A$  y  $B$ . El resultado parcial es cero. La multiplicación de este resultado por  $C$  da también cero. Si, en cambio, se arregla la expresión como

$$X = A * C * B$$

se multiplica  $A$  por  $C$  y se obtiene  $0.1200 \times 10^2$ . La multiplicación siguiente da la respuesta correcta:  $0.2400 \times 10^{-3}$ . De igual manera, un arreglo en una división puede evitar *underflow*.

#### d) División entre un número muy pequeño

Como se dijo, la división entre un número muy pequeño puede causar *overflow*.

Supóngase que se realiza en la computadora una división válida y que no se comete error alguno en la operación; pero considérese que ocurrió un pequeño error de redondeo previamente en el programa, cuando se calculó el denominador. Si el numerador es grande y el denominador pequeño, puede presentarse un error absoluto considerable en el cociente. Si éste se resta después, de otro número del mismo tamaño relativo, puede presentarse un error mayor en la respuesta final.

Como ejemplo considérese la siguiente instrucción en un programa

$$X = A - B/C$$

donde:

$$A = 0.1120 \times 10^9 = 112000000$$

$$B = 0.1000 \times 10^6 = 100000$$

$$C = 0.900 \times 10^{-3} = 0.0009$$

Si el cálculo se realiza en la computadora decimal de cuatro dígitos, el cociente  $B/C$  es  $0.1111 \times 10^9$ , y  $X$  es  $0.0009 \times 10^9$  o, después de ser normalizado,  $X = 0.9000 \times 10^6$ . Nótese que sólo hay un dígito significativo.

Vamos a imaginar ahora que se cometió un pequeño error de redondeo al calcular  $C$  en algún paso previo y resultó un valor  $C^* = 0.9001 \times 10^{-3}$  ( $EA = 0.0001 \times 10^{-3}$ ;  $ER = 10^{-4}$  y  $ERP = 0.01\%$ ).

Si se calcula  $B/C^*$  se obtiene como cociente  $0.1110 \times 10^9$  y  $X^* = 0.1000 \times 10^7$ . El valor correcto de  $X$  es  $0.9000 \times 10^6$ .



Entonces:

$$EA = |1000000 - 900000| = 100000$$

$$ER = \frac{|1000000 - 900000|}{900000} = 0.11$$

$$ERP = 0.11 \times 100 = 11\%$$

El error relativo se ha multiplicado cerca de 1100 veces. Como ya se dijo, estos cálculos pueden conducir a un resultado final carente de significado o sin relación con la respuesta verdadera.

### e) Error de discretización

Dado que un número específico no se puede almacenar exactamente como número binario de punto flotante, el error generado se conoce como error de discretización (error de cuantificación), ya que los números expresados exactamente por la máquina (números máquina) no forman un conjunto continuo sino discreto.

### Ejemplo 1.12

Cuando se suma 10000 veces 0.0001 con él mismo, debe resultar 1; sin embargo, el número 0.0001 en binario resulta en una sucesión infinita de ceros y unos, que se trunca al ser almacenada en una palabra de memoria, con lo que se perderá información y el resultado de la suma ya no será 1. Se obtuvieron los siguientes resultados que corroboran lo anterior, utilizando una PC, precisión sencilla y Visual Basic.

### Solución



$$a) \sum_{i=1}^{10000} 0.0001 = 1.000054$$

$$b) 1 + \sum_{i=1}^{10000} 0.0001 = 2.000166$$

$$c) 1000 + \sum_{i=1}^{10000} 0.0001 = 1001.221$$

$$d) 10000 + \sum_{i=1}^{10000} 0.0001 = 10000$$

Nótese que en los tres últimos incisos, además del error de discretización, se generó el error de sumar un número muy grande con un número muy pequeño (véase Prob. 1.16 y 1.17). El programa se ejecutó iniciando primero a una variable con el valor entero 0, 1, 1000 y 10000; después se fue acumulando a esa variable 0.0001 diez mil veces.

### f) Errores de salida

Aun cuando no se haya cometido error alguno durante la fase de cálculos de un programa, puede presentarse un error al imprimir resultados.

Por ejemplo, supóngase que la respuesta de un cálculo particular es exactamente 0.015625. Cuando este número se imprime con un formato tal como F10.6 o E14.6 (de FORTRAN), se obtiene la respuesta correcta. Si, por el contrario, se decide usar F8.3, se impi-

mirá el número 0.016 (si la computadora redondea), o bien 0.015 (si la computadora trunca), con lo cual se presenta un error.

## PROPAGACIÓN DE ERRORES

Una vez que se sabe cómo se producen los errores en un programa de cómputo, podría pensarse en tratar de determinar el error cometido en cada paso, y conocer de esa manera el error total en la respuesta final. Sin embargo, esto no es práctico. Resulta más adecuado analizar las operaciones individuales realizadas por la computadora para ver cómo se propagan los errores de dichas operaciones.

### a) Suma

Se espera que al sumar  $a$  y  $b$ , se obtenga el valor correcto de  $c = a + b$ ; no obstante, se tiene en general un valor de  $c$  incorrecto debido a la longitud finita de palabra que se emplea. Puede considerarse que este error fue causado por una operación incorrecta de la computadora  $+$  (el punto indica que es suma con error). Entonces el error es:

$$\text{Error} = (a \dot{+} b) - (a + b)$$

La magnitud de este error depende de las magnitudes relativas, de los signos de  $a$  y  $b$ , y de la forma binaria en que  $a$  y  $b$  son almacenados en la computadora. Esto último varía dependiendo de la computadora, y por tanto es un error muy difícil de analizar y no lo estudiaremos aquí.

Si por otro lado  $a$  y  $b$  de entrada son inexactos, hay un segundo error posible. Por ejemplo, considérese que en lugar del valor verdadero de  $a$ , la computadora tiene el valor  $a^*$ , el cual presenta un error  $\epsilon_a$

$$a^* = a + \epsilon_a$$

y de igual forma para  $b$

$$b^* = b + \epsilon_b$$

Como consecuencia se tendría, aun si no se cometiera error en la adición, un error en el resultado:

$$\begin{aligned} \text{Error} &= (a^* + b^*) - (a + b) \\ &= (a + \epsilon_a + b + \epsilon_b) - (a + b) = \epsilon_a + \epsilon_b = \epsilon_c \end{aligned}$$

o sea  $c^* = c + \epsilon_c$

El error absoluto es:

$$|(a^* + b^*) - (a + b)| = |\epsilon_a + \epsilon_b| \leq |\epsilon_a| + |\epsilon_b|$$

o bien

$$|\epsilon_c| \leq |\epsilon_a| + |\epsilon_b|$$

Se dice que los errores  $\epsilon_a$  y  $\epsilon_b$  se han extendido a  $c$ , y  $\epsilon_c$  se conoce como el error de propagación.

Dicho error es causado por valores inexactos de los valores iniciales y se propaga en los cálculos siguientes, con lo cual causa un error en el resultado final.

**b) Resta**

El error de propagación ocasionado por valores inexactos iniciales  $a^*$  y  $b^*$ , puede darse de manera similar que en la adición, con un simple cambio de signo (véase Prob. 1.24).

**c) Multiplicación**

Si se multiplican los números  $a^*$  y  $b^*$ , se obtiene (ignorando el error causado por la operación misma):

$$\begin{aligned}(a^* \times b^*) &= (a + \epsilon_a) \times (b + \epsilon_b) \\ &= (a \times b) + (a \times \epsilon_b) + (b \times \epsilon_a) + (\epsilon_a \times \epsilon_b)\end{aligned}$$

Si  $\epsilon_a$  y  $\epsilon_b$  son suficientemente pequeños, puede considerarse que su producto es muy pequeño en comparación con los otros términos y, por tanto, eliminar el último término. Se obtiene entonces el error del resultado final

$$(a^* \times b^*) - (a \times b) \approx (a \times \epsilon_b) + (b \times \epsilon_a)$$

Esto hace posible encontrar el valor absoluto del error relativo del resultado dividiendo ambos lados entre  $a \times b$ .

$$\left| \frac{(a^* \times b^*) - (a \times b)}{(a \times b)} \right| \approx \left| \frac{\epsilon_b}{b} + \frac{\epsilon_a}{a} \right| \leq \left| \frac{\epsilon_b}{b} \right| + \left| \frac{\epsilon_a}{a} \right|$$

El error de propagación relativo en valor absoluto en la multiplicación es aproximadamente igual o menor a la suma de los errores relativos de  $a$  y  $b$  en valor absoluto.

**d) División**

Puede considerarse la división de  $a^*$  y  $b^*$  como sigue:

$$\begin{aligned}a^* / b^* &= (a + \epsilon_a) / (b + \epsilon_b) \\ &= (a + \epsilon_a) \frac{1}{(b + \epsilon_b)}\end{aligned}$$

Multiplicando numerador y denominador por  $b - \epsilon_b$

$$\begin{aligned}a^* / b^* &= \frac{(a + \epsilon_a)(b - \epsilon_b)}{(b + \epsilon_b)(b - \epsilon_b)} \\ &= \frac{ab - a\epsilon_b + \epsilon_a b - \epsilon_a \epsilon_b}{b^2 - \epsilon_b^2}\end{aligned}$$

Si, como en la multiplicación, se considera el producto  $\epsilon_a \epsilon_b$  muy pequeño y, por las mismas razones, a  $\epsilon_b^2$  y se desprecian, se tiene.

$$\begin{aligned}a^* / b^* &\approx \frac{ab}{b^2} + \frac{\epsilon_a b}{b^2} - \frac{a\epsilon_b}{b^2} \\ &\approx \frac{a}{b} + \frac{\epsilon_a}{b} - \frac{a\epsilon_b}{b^2}\end{aligned}$$

El error es entonces:

$$a^* / b^* - \frac{a}{b} \approx \frac{\epsilon_a}{b} - \frac{a\epsilon_b}{b^2}$$



Dividiendo entre  $a/b$  se obtiene el error relativo. Al tomar el valor absoluto del error relativo, se tiene

$$\left| \frac{a^*/b^* - \frac{a}{b}}{a/b} \right| \approx \left| \frac{\frac{\epsilon_a}{b} - \frac{a\epsilon_b}{b^2}}{a/b} \right| \approx \left| \frac{\epsilon_a}{a} - \frac{\epsilon_b}{b} \right| \leq \left| \frac{\epsilon_a}{a} \right| + \left| \frac{\epsilon_b}{b} \right|$$

Se concluye que el error de propagación relativo del cociente en valor absoluto es aproximadamente igual o menor a la suma de los errores relativos en valor absoluto de  $a$  y  $b$ .

### e) Evaluación de funciones

Por último, se estudiará la propagación del error (asumiendo operaciones básicas  $+$ ,  $-$ ,  $\times$  y  $/$  ideales o sin errores), cuando se evalúa una función  $f(x)$  en un punto  $x = a$ . En general, se dispone de un valor de  $a$  aproximado:  $a^*$ ; la intención es determinar el error resultante

$$\epsilon_f = f(a^*) - f(a)$$

La figura 1.3 muestra la gráfica de la función  $f(x)$  en las cercanías de  $x = a$ . A continuación se determina la relación entre  $\epsilon_a$  y  $\epsilon_f$ .

Si  $\epsilon_a$  es pequeño, puede aproximarse la curva  $f(x)$  por su tangente en un entorno de  $x = a$ . Se sabe que la pendiente de esta tangente es  $f'(a)$  o aproximadamente  $\epsilon_f / \epsilon_a$ ; esto es:

$$\epsilon_f / \epsilon_a \approx f'(a)$$

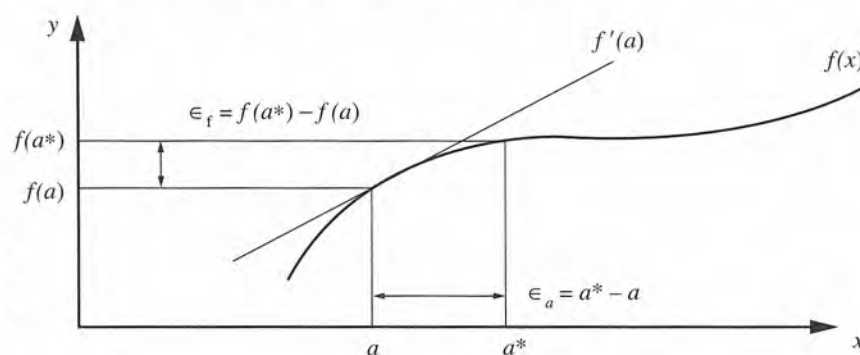
y

$$\epsilon_f \approx \epsilon_a f'(a) \approx \epsilon_a f'(a^*)$$

En valor absoluto

$$|\epsilon_f| \approx |\epsilon_a f'(a^*)| \approx |\epsilon_a| |f'(a^*)|$$

El error al evaluar una función en un argumento inexacto es proporcional a la primera derivada de la función en el punto donde se ha evaluado.



**Figura 1.6**  
Gráfica de una  
función y su  
primera  
derivada en  $a$ .

## 1.4 Algoritmos y estabilidad

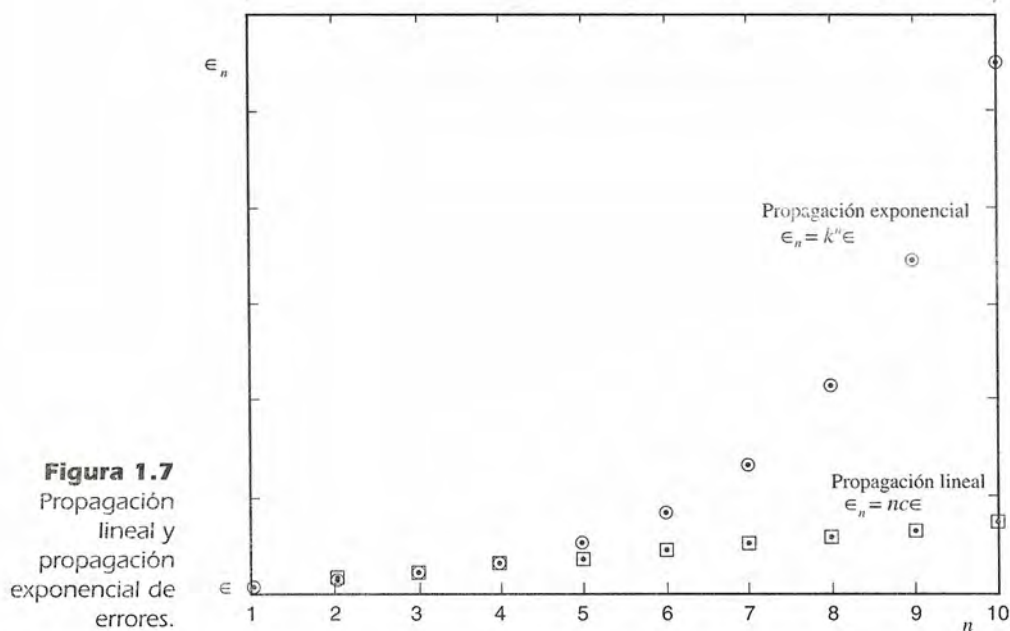
El tema fundamental de este libro es el estudio, selección y aplicación de algoritmos, que se definen como secuencias de operaciones algebraicas y lógicas para obtener la solución de un problema. Por lo general, se dispone de varios algoritmos para resolver un problema par-

titular; 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.

Supóngase que un error  $\epsilon$  se introduce en algún paso en los cálculos, y que el error de propagación de  $n$  operaciones subsiguientes se denota por  $\epsilon_n$ . En la práctica, por lo general se presentan dos casos.

- $|\epsilon_n| \approx n c \epsilon$ , donde  $c$  es una constante independiente de  $n$ ; se dice entonces que la propagación del error es lineal.
- $|\epsilon_n| \approx k^n \epsilon$ , 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  $\epsilon$  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  $\epsilon$ . Como consecuencia, se dice que un algoritmo con crecimiento lineal del error es estable, mientras que un algoritmo con una propagación exponencial es inestable (véase Fig. 1.4).



## Ejercicios



### 1.1 Error de redondeo al restar dos números casi iguales.

Vamos a considerar las ecuaciones

$$31.69x + 14.31y = 45.00 \quad (1)$$

$$13.05x + 5.89y = 18.53 \quad (2)$$

La única solución de este sistema de ecuaciones es (redondeando a cinco cifras decimales)  $x = 1.25055$ ,  $y = 0.37527$ . Un método para resolver este tipo de problemas es multiplicar la ecuación (1) por el coeficiente de  $x$  de la ecuación (2), multiplicar la ecuación

(2) por el coeficiente de  $x$  de la ecuación (1) y después restar las ecuaciones resultantes. Para este sistema se obtendría (como los coeficientes tienen dos cifras decimales, todas las operaciones intermedias se efectúan redondeando a dos cifras decimales):

$$\begin{aligned} [13.05 (14.31) - 31.69 (5.89)] y &= 13.05 (45.00) - 31.69 (18.53) \\ (186.75 - 186.65) y &= 587.25 - 587.22 \\ 0.10 y &= 0.03 \end{aligned}$$

de donde  $y = 0.3$ , luego

$$x = \frac{(18.53) - 5.89 (0.3)}{13.05} = \frac{18.53 - 1.77}{13.05} = \frac{16.76}{13.05} = 1.28$$

Para la variable  $x$

$$EA = |1.28 - 1.25| = 0.03; \quad ER = 0.03/1.25 = 0.024; \quad ERP = 2.4\%$$

Para la variable  $y$

$$EA = |0.3 - 0.38| = 0.08; \quad ER = 0.08/0.38 = 0.21; \quad ERP = 21\%$$

## 1.2 Error de redondeo al sumar un número grande y uno pequeño.

Considere la sumatoria infinita

$$s = \sum_{n=1}^{\infty} \frac{1}{n^2} = \frac{1}{1} + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \frac{1}{25} + \dots + \frac{1}{100} + \dots$$

resulta (usando precisión sencilla y 5000 como valor final de  $n$ ) 1.644725 si se suma de izquierda a derecha, pero resulta 1.644834 si se suma de derecha a izquierda, a partir de  $n = 5000$ .

Debe notarse que el resultado de sumar de derecha a izquierda es más exacto, ya que en todos los términos se suman valores de igual magnitud.

Por el contrario, al sumar de izquierda a derecha, una vez que se avanza en la sumatoria, se sumarán números cada vez más grandes con números más pequeños.

Lo anterior se corrobora si se realiza la suma en ambos sentidos, pero ahora con doble precisión. El resultado obtenido es 1.64473408684689 (estos resultados pueden variar de máquina a máquina).

## 1.3 Reducción de errores.

Para resolver la ecuación cuadrática

$$100 x^2 - 10011 x + 10.011 = 0,$$

el método común sería usar la fórmula

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a},$$

después de dividir la ecuación entre 100.

$$\begin{aligned} x^2 - 100.11 x + 0.10011 &= 0 \\ x &= \frac{100.11 \pm \sqrt{(-100.11)^2 - 4(0.10011)}}{2} \end{aligned}$$



Trabajando con aritmética de cinco dígitos

$$\begin{aligned} x &= \frac{100.11 \pm \sqrt{10022 - 0.40044}}{2} = \frac{100.11 \pm \sqrt{10022}}{2} \\ &= \frac{100.11 \pm 100.11}{2} = \begin{cases} \frac{200.22}{2} = 100.11 \\ 0 \end{cases} \end{aligned}$$

Las soluciones verdaderas, redondeadas a cinco dígitos decimales son 100.11 y 0.00100.

El método empleado fue adecuado para la solución mayor, pero no del todo para la solución menor. Si las soluciones fueran divisores de otras expresiones, la solución  $x = 0$  hubiese causado problemas serios.

Se restaron dos números “casi iguales” (números iguales en aritmética de cinco dígitos) y sufrieron pérdida de exactitud.

¿Cómo evitar esto? Una forma sería reescribir la expresión para la solución de una ecuación cuadrática a fin de evitar la resta de números “casi iguales”.

El problema, en este caso, se da en el signo negativo asignado a la raíz cuadrada; esto es

$$\frac{-b - \sqrt{b^2 - 4ac}}{2a}$$

Multiplicando numerador y denominador por  $-b + \sqrt{b^2 - 4ac}$ , queda

$$\begin{aligned} \frac{(-b - \sqrt{b^2 - 4ac})(-b + \sqrt{b^2 - 4ac})}{2a(-b + \sqrt{b^2 - 4ac})} &= \frac{(-b)^2 - (b^2 - 4ac)}{2a(-b + \sqrt{b^2 - 4ac})} \\ &= \frac{4ac}{2a(-b + \sqrt{b^2 - 4ac})} = \frac{2c}{(-b + \sqrt{b^2 - 4ac})} \end{aligned}$$

Usando esta expresión con  $a = 1$ ,  $b = -100.11$ , y  $c = 0.10011$ , se obtiene

$$\frac{2(0.10011)}{100.11 + \sqrt{10022}} = \frac{0.20022}{200.22} = 0.001 \text{ (en aritmética de cinco dígitos)}$$

que es el valor verdadero, redondeado a cinco dígitos decimales.

Esta forma alternativa para calcular una raíz pequeña de una ecuación cuadrática, casi siempre produce una respuesta más exacta que la de la fórmula usual (véase Prob. 2.31).

#### 1.4 Más sobre reducción de errores.

Se desea evaluar la expresión  $A / (1 - \sin x)$ , en  $x = 89^\circ 41'$ . En tablas con cinco cifras decimales,  $\sin 89^\circ 41' = 0.99998$ . Con aritmética de cinco dígitos y redondeando se tiene

$$\sin x = 0.99998 \text{ y } 1 - \sin x = 0.00002$$

El valor de  $\sin x$  sólo tiene cuatro dígitos exactos (confiables). Por otro lado, el único dígito que no es cero en  $1 - \sin x$  se ha calculado con el dígito no confiable de  $\sin x$ , por lo que se pudo perder la exactitud en la resta.

Esta situación de arriba puede mejorarse observando que

$$1 - \sin x = \frac{(1 - \sin x)(1 + \sin x)}{1 + \sin x} = \frac{1 - \sin^2 x}{1 + \sin x} = \frac{\cos^2 x}{1 + \sin x}$$

Por esto, es posible escribir  $1 - \sin x$  de una forma que no incluye la resta de dos números casi iguales.

### 1.5 Comparaciones seguras.

En los métodos numéricos, a menudo la comparación de igualdad de dos números en notación de punto flotante permitirá terminar la repetición de un conjunto de cálculos (proceso cíclico o iterativo). En vista de los errores observados, es recomendable comparar la diferencia de los dos números en valor absoluto contra una tolerancia  $\epsilon$  apropiada, usando por ejemplo el operador de relación menor o igual ( $\leq$ ). Esto se ilustra enseguida.

En lugar de

SI  $X = Y$  ALTO; En caso contrario REPETIR las instrucciones 5 a 9

Deberá usarse:

SI  $\text{ABS}(X - Y) \leq \epsilon$  ALTO; en caso contrario REPETIR las instrucciones 5 a 9.

En lugar de

REPETIR

{ Pasos de un ciclo }

HASTA QUE  $X = Y$

Deberá usarse

REPETIR

{ Pasos de un ciclo }

HASTA QUE  $\text{ABS}(X - Y) \leq \epsilon$

donde  $\epsilon$  es un número pequeño (generalmente menor que uno, pero puede ser mayor, dependiendo el contexto en que se trabaje) e indicará la cercanía de  $X$  con  $Y$  que se aceptará como “igualdad” de  $X$  y  $Y$ .

### 1.6 Análisis de resultados.

Al ejecutar las siguientes instrucciones en Visual Basic con doble precisión y en Matlab, se tiene, respectivamente:



```
Dim Y As Double, A as Double
Y=1000.2
A=Y-1000
Print A
```

Se obtiene:  
0.2000000000000045



```
format long
Y=1000.2;
A=Y-1000
```

Se obtiene:  
0.200000000000005

Ejecute las mismas instrucciones pero usando  $Y = 1000.25$ . Los resultados ahora son correctos. Explíquelo.

En doble precisión pueden manejarse alrededor de quince dígitos decimales de exactitud, de modo que la resta de arriba se representa

$$1000.200 - 1000.000$$

La computadora convierte Y a binario dando un número infinito de ceros y unos, y almacena un número distinto a 1000.2 (véase Prob. 1.6 b).

Por otro lado, 1000 sí se puede almacenar o representar exactamente en la computadora en binario en punto flotante (los números con esta característica se llaman números máquina). Al efectuarse la resta se obtiene un número diferente de 0.2. Esto muestra por qué deberá analizarse siempre un resultado de un dispositivo digital antes de aceptarlo.

### 1.7 Más sobre análisis de resultados.

El método de posición falsa (véase sección 2.4) obtiene su algoritmo al encontrar el punto de corte de la línea recta que pasa por los puntos  $(x_D, y_D)$ ,  $(x_I, y_I)$  y el eje  $x$ . Pueden obtenerse dos expresiones para encontrar el punto de corte  $x_M$

$$\text{i) } x_M = \frac{x_I y_D - x_D y_I}{y_D - y_I} \quad \text{ii) } x_M = x_D - \frac{(x_D - x_I) y_D}{y_D - y_I}$$

Si  $(x_D, y_D) = (2.13, 4.19)$  y  $(x_I, y_I) = (1.96, 6.87)$  y usando aritmética de tres dígitos y redondeando, ¿cuál es la mejor expresión y por qué?

**Solución** Sustituyendo en i) y en ii)

$$\text{i) } x_M = \frac{1.96 (4.19) - 2.13 (6.87)}{4.19 - 6.87} = 2.38$$

$$\text{ii) } x_M = 2.13 - \frac{(2.13 - 1.96) 4.19}{4.19 - 6.87} = 2.40$$

Al calcular los errores absoluto y relativo, y tomando como valor verdadero a 2.395783582, el cual se calculó con aritmética de 13 dígitos, se tiene:

$$\text{i) } EA = 2.395783582 - 2.38 = 0.015783582$$

$$ER = \frac{0.015783582}{2.395783582} = 0.006588066$$

$$\text{ii) } EA = 2.395783582 - 2.40 = 0.004216418$$

$$ER = \frac{0.004216418}{2.395783582} = 0.001759932$$

de donde es evidente que la forma ii) es mejor. Se sugiere al lector reflexionar sobre el por qué.

## Problemas

- 1.1** Proporcione los símbolos o numerales romanos correspondientes a los siguientes símbolos arábigos

10, 100, 1000, 10000, 100000, 1000000



- 1.2 Convierta\* los siguientes números decimales a los sistemas con base 2 y base 8, y viceversa.  
 a) 536      b) 923      c) 1536      d) 8      e) 2      f) 10      g) 0
- 1.3 Convierta los siguientes números enteros del sistema octal a binario y viceversa.  
 a) 777      b) 573      c) 7      d) 2      e) 10      f) 0
- 1.4 Resuelva las siguientes preguntas.  
 a) ¿El número 101121 pertenece al sistema binario?  
 b) ¿El número 3852 pertenece al sistema octal?
- Si su respuesta es NO en alguno de los incisos, explique por qué; si es SÍ, conviértalo(s) a decimal.
- 1.5 Convierta los siguientes números dados en binario a decimal y viceversa, usando la conversión a octal como paso intermedio  
 a) 1000      b) 10101      c) 111111
- 1.6 Convierta los siguientes números fraccionarios dados en decimal, a binario y octal  
 a) 0.8      b) 0.2      c) 0.973      d) 0.356      e) 0.713      f) 0.10
- 1.7 Convierta los siguientes números fraccionarios, dados en binario, a decimal  
 a) 0.1      b) 0.010101      c) 0.0001      d) 0.11111      e) 0.00110011  
 f) 0.0110111
- 1.8 Repita los incisos a) a f) del problema 1.7, pero pasando a octal como paso intermedio.
- 1.9 Convierta los siguientes números, dados en decimal, a octal y binario.  
 a) 985.34      b) 10.1      c) 888.222      d) 3.57      e) 977.93      f) 0.357  
 g) 0.9389      h) -0.9389
- 1.10 En la sección 1.2 se dijo que cada palabra de 16 bits puede contener un número entero cualquiera del intervalo  $-32768$  a  $+32767$ . Investigue por qué se incluye al  $-32768$ , o bien por qué el intervalo no inicia en  $-32767$ .
- 1.11 Considere una computadora con una palabra de 8 bits. ¿Qué rango de números enteros puede contener dicha palabra?
- 1.12 Represente el número  $-26$  en una palabra de 8 bits.
- 1.13 Dados los siguientes números máquina en una palabra de 16 bits

|    |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| a) | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 0 | 0 |
| b) | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| c) | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |

¿Qué decimales representan?

- 1.14 Normalice los siguientes números

a) 723.5578      b)  $-15.324$       c) 0.003485      d)  $8 \times 10^3$

\* Puede usar el Programa 1.1 del disco para comprobar sus resultados.

**SUGERENCIA:** Pasar los números a binario y después normalizarlos.

- 1.15** Represente en doble precisión el número decimal del ejemplo 1.10
- 1.16** Elabore un programa para la calculadora o el dispositivo de cálculo con el que se cuente, de modo que el número 0.0001 se sume diez mil veces consigo mismo

$$\begin{array}{ccccccc} 0.0001 & + & 0.0001 & + & \dots & + & 0.0001 \\ 1 & & 2 & & & & 10000 \end{array}$$

El resultado deberá imprimirse. Interprete este resultado de acuerdo con los siguientes lineamientos

- a) Si es 1, ¿cómo es posible si se sumaron diez mil valores que no son realmente 0.0001?
  - b) En caso de obtener 1, explore con el valor 0.00001, 0.000001, etc., hasta obtener un resultado diferente de 1.
  - c) ¿Es posible obtener un resultado menor de 1? ¿Por qué?
- 1.17** Con el programa del problema 1.16 efectúe los cálculos de los incisos a) a d) del ejemplo 1.12 y obtenga los resultados de la siguiente manera
- a) Inicialice la variable SUMA con 0, 1, 1000 y 10000 en los incisos a), b), c) y d), respectivamente, y luego en un ciclo súpese a ese valor diez mil veces el 0.0001. Anote sus resultados.
  - b) Inicialice la variable SUMA con 0 para los cuatro incisos y al final del ciclo donde se habrá sumado 0.0001 consigo mismo 10000 veces, sume a ese resultado los números 0, 1, 1000 y 10000 e imprima los resultados.

Interprete las diferencias de los resultados.

- 1.18** La mayoría de las calculadoras científicas almacenan dos o tres dígitos de seguridad más de los que despliegan. Por ejemplo, una calculadora que despliega ocho dígitos puede almacenar realmente diez (dos dígitos de seguridad); por tanto, será un dispositivo de diez dígitos. Para encontrar la exactitud real de su calculadora, realice las siguientes operaciones.

Divida 10 entre 3, al resultado réstele 3.

Divida 100 entre 3, al resultado réstele 33.

Divida 1000 entre 3, al resultado réstele 333.

Divida 10000 entre 3, al resultado réstele 3333.

Notará que la cantidad de los números 3 desplegados se va reduciendo.

La cantidad de 3 desplegada en cualquiera de las operaciones anteriores, sumada al número de ceros utilizados con el 1, indica el número de cifras significativas que maneja su calculadora. Por ejemplo, si con la segunda operación despliega 0.3333333 la calculadora maneja nueve cifras significativas de exactitud (7 + 2 ceros que tiene 100).

**NOTA:** Si su calculadora es del tipo intérprete BASIC, no realice las operaciones como 1000/3-333 porque obtendrá otros resultados.

- 1.19** Evalúe la expresión  $A / (1 - \cos x)$ , en un valor de  $x$  cercano a  $0^\circ$ . ¿Cómo podría evitar la resta de dos números casi iguales en el denominador?
- 1.20** Determine en su calculadora o microcomputadora si muestra un mensaje de *overflow* o no.
- 1.21** Deduzca las expresiones para  $x_M$  dadas en el ejercicio 1.7.
- 1.22** Un número máquina para una calculadora o computadora es un número real que se almacena exactamente (en forma binaria de punto flotante). El número -125.32 del ejemplo 1.10,

evidentemente no es un número de máquina (si el dispositivo de cálculo tiene una palabra de 16 bits). Por otro lado, el número  $-26$  del ejemplo 1.8 sí lo es, empleando una palabra de 16 bits. Determine 10 números de máquina en el intervalo  $[10^{-19}, 10^{18}]$  cuando se emplea una palabra de 16 bits.

- 1.23 Investigue cuántos números máquina positivos es posible representar en una palabra de 16 bits.
- 1.24 Haga el análisis de la propagación de errores para la resta (véase análisis de la suma, en la sección 1.3).
- 1.25 Se desea evaluar la función  $e^{5x}$  en el punto  $x = 1.0$ ; sin embargo, si el valor de  $x$  se calculó en un paso previo con un pequeño error y se tiene  $x^* = 1.01$ ; determine  $\epsilon_f$  con las expresiones dadas en la evaluación de funciones de la sección 1.3. Luego determine  $\epsilon_f$  como  $f(1) - f(1.01)$  y compare los resultados.
- 1.26 Resuelva el siguiente sistema de ecuaciones, usando dos cifras decimales para guardar los resultados intermedios y finales.

$$21.76x + 24.34y = 1.24 \qquad 14.16x + 15.84y = 1.15$$

y determine el error cometido. La solución exacta (redondeada a 5 cifras decimales es)  $x = -347.89167$ ,  $y = 311.06667$ .

- 1.27 Codifique el siguiente algoritmo en su microcomputadora (use precisión sencilla)

PASO 1. Leer A.

PASO 2. Mientras  $A > 0$ , repetir los pasos 3 y 4.

PASO 3. IMPRIMIR  $\text{Ln}(\text{Exp}(A)) - A$ ,  $\text{Exp}(\text{Ln}(A)) - A$

PASO 4. Leer A.

PASO 5. TERMINAR.

Ejecútelo con diferentes valores de A, por ejemplo 0.2, 0.25, 1, 1.5, 1.8, 2.5, 3.14159, 0.008205, etc., y observe los resultados.

- 1.28 Modifique el programa del problema del ejemplo 1.27 usando doble precisión para A y compare los resultados.
- 1.29 Modifique el paso 3 del programa del problema 1.27 para que quede así

$$\text{IMPRIMIR } \text{SQR}(A^2) - A, \text{SQR}(A)^2 - A$$

y vuelva a ejecutarlo con los mismos valores.

- 1.30 Realice la modificación indicada en el problema 1.29 al programa del problema 1.28. Compare los resultados.
- 1.31 Repita los problemas 1.27 a 1.30 con lenguaje Pascal (puede usar Delphi, por ejemplo), con lenguaje Visual C++ y compare los resultados con los obtenidos en Basic.



# CAPÍTULO 2

## SOLUCIÓN DE ECUACIONES NO LINEALES

### A dónde nos dirigimos

En este capítulo estudiaremos diversos métodos para resolver ecuaciones no lineales en una incógnita,  $f(x) = 0$ , aprovechando los conceptos básicos del cálculo y las posibilidades gráficas y de cómputo de la tecnología moderna. A lo largo del texto, recurriremos sistemáticamente a la interpretación gráfica de los métodos, a fin de mostrar visualmente su funcionamiento y de enriquecer las imágenes asociadas con ellos; de igual manera, se generan tablas en la aplicación de cada técnica para analizar el comportamiento numérico y eventualmente detener el proceso.

Se ha organizado el material como métodos de uno y de dos puntos, usando como prototipo de los primeros el de punto fijo, y de los segundos el de posición falsa. Esto, junto con el concepto de orden de convergencia, nos permitirá tener los elementos suficientes para seleccionar la técnica más adecuada para una situación dada. Finalizamos el capítulo con las técnicas para resolver ecuaciones polinomiales. Algunas de ellas son adaptaciones de las que estudiamos anteriormente y otras particulares para esta familia.

El propósito de este capítulo es que el lector cuente con los elementos básicos, computacionales y de criterio, apropiados para resolver el problema algebraico clásico de encontrar las raíces reales y complejas de la ecuación  $f(x)=0$ , en donde las técnicas algebraicas de “despejar” la incógnita no sean aplicables, como es el caso de  $\cos x - 3x = 0$  o  $e^x - 3x = 0$ , o bien resulten imprácticas. Por último, es importante señalar lo difícil que resulta pensar en un tópico de matemáticas o ingeniería que no involucre ecuaciones de esta naturaleza.

### Introducción

Uno de los problemas más frecuentes en ingeniería es encontrar las raíces de ecuaciones de la forma  $f(x) = 0$ , donde  $f(x)$  es una función real de una variable  $x$ , como un polinomio en  $x$

$$f(x) = 4x^5 + x^3 - 8x + 2$$

o una función trascendente\*

$$f(x) = e^x \sin x + \ln 3x + x^3$$

Existen distintos algoritmos para encontrar las raíces o ceros de  $f(x) = 0$ , pero ninguno es general. es decir, no hay un algoritmo que funcione con todas las ecuaciones; por ejemplo, se puede pensar en un algoritmo que funcione perfectamente para encontrar las raíces de  $f_1(x) = 0$ , pero al aplicarlo no se pueden encontrar los ceros de una ecuación distinta  $f_2(x) = 0$ .

\* Las funciones trascendentes contienen términos trigonométricos, exponenciales o logarítmicos de la variable independiente.

Sólo en muy pocos casos será posible obtener las raíces exactas de  $f(x) = 0$ , como cuando  $f(x)$  es un polinomio factorizable, tal como

$$f(x) = (x - \bar{x}_1)(x - \bar{x}_2) \dots (x - \bar{x}_n),$$

donde  $\bar{x}_i$ ,  $1 \leq i \leq n$  denota la  $i$ -ésima raíz de  $f(x) = 0$ . Sin embargo, se pueden obtener soluciones aproximadas al utilizar algunos de los métodos numéricos de este capítulo. Se empezará con el método de punto fijo (también conocido como de aproximaciones sucesivas, de iteración funcional, etc.), por ser el prototipo de todos ellos.

## 2.1 Método de punto fijo

Sea la ecuación general

$$f(x) = 0, \quad (2.1)$$

de la cual se desea encontrar una raíz real\*  $\bar{x}$ .

El primer paso consiste en transformar algebraicamente la ecuación 2.1 a la forma equivalente

$$x = g(x) \quad (2.2)$$

Por ejemplo para la ecuación

$$f(x) = 2x^2 - x - 5 = 0 \quad (2.3)$$

cuyas raíces son 1.850781059 y -1.350781059, algunas posibilidades de  $x = g(x)$  son

- |                                          |                                      |       |
|------------------------------------------|--------------------------------------|-------|
| a) $x = 2x^2 - 5$                        | “despejando” el segundo término.     |       |
| b) $x = \sqrt{\frac{x+5}{2}}$            | “Despejando” $x$ del primer término. | (2.4) |
| c) $x = \frac{5}{2x-1}$                  | Factorizando $x$ y “despejándola”.   |       |
| d) $x = 2x^2 - 5$                        | Sumando $x$ a cada lado.             |       |
| e) $x = x - \frac{2x^2 - x - 5}{4x - 1}$ | Véase sección 2.2                    |       |

Una vez que se ha determinado una forma equivalente (Ec. 2.2), el siguiente paso es **tantear** una raíz; esto puede hacerse por observación directa de la ecuación (por ejemplo en la Ec. 2.3 se ve directamente que  $x = 2$  es un valor cercano a una raíz).\*\* Se denota el valor de tanteo o valor de inicio como  $x_0$ . Otros métodos de tanteo se estudiarán en la sección 2.8.

Una vez que se tiene  $x_0$ , se evalúa  $g(x)$  en  $x_0$ , denotándose el resultado de esta evaluación como  $x_1$ ; esto es,

$$g(x_0) = x_1$$

El valor de  $x_1$  comparado con  $x_0$  presenta los dos siguientes casos:

\* En las secciones 2.9 y 2.10 se analizará el caso de raíces complejas.

\*\* Puede graficar usando un paquete comercial.

**CASO 1.** QUE  $x_1 = x_0$ 

Esto indica que se ha elegido como valor inicial una raíz y el problema queda concluido. Para aclararlo, recuérdese que si  $\bar{x}$  es raíz de la ecuación 2.1, se cumple que

$$f(\bar{x}) = 0,$$

y como la ecuación 2.2 es sólo un rearrreglo de la ecuación 2.1, también es cierto que

$$g(\bar{x}) = \bar{x}.$$

Si se hubiese elegido como  $x_0 = 1.850781059$  para la ecuación 2.3, el lector puede verificar que cualquiera que sea la  $g(x)$  seleccionada,  $g(1.850781059) = 1.850781059$ ; esto se debe a que 1.850781059 es una raíz de la ecuación 2.3. Esta característica de  $g(x)$  de fijar su valor en una raíz  $\bar{x}$  ha dado a este método el nombre que lleva.

**CASO 2.** QUE  $x_1 \neq x_0$ 

Es el caso más frecuente e indica que  $x_1$  y  $x_0$  son distintos de  $\bar{x}$ . Esto es fácil de explicar, ya que si  $\dot{x}$  no es una raíz de 2.1, se tiene que

$$f(\dot{x}) \neq 0,$$

y por otro lado, evaluando  $g(x)$  en  $\dot{x}$ , se tiene

$$g(\dot{x}) \neq \dot{x}.$$

En estas circunstancias se procede a una segunda evaluación de  $g(x)$ , ahora en  $x_1$ , denotándose el resultado como  $x_2$

$$g(x_1) = x_2$$

Este proceso se repite y se obtiene el siguiente esquema iterativo

|                           |                    |              |       |
|---------------------------|--------------------|--------------|-------|
| Valor inicial:            | $x_0$              | $f(x_0)$     |       |
| Primera iteración:        | $x_1 = g(x_0)$     | $f(x_1)$     |       |
| Segunda iteración:        | $x_2 = g(x_1)$     | $f(x_2)$     |       |
| Tercera iteración:        | $x_3 = g(x_2)$     | $f(x_3)$     |       |
| .                         | .                  | .            |       |
| .                         | .                  | .            |       |
| .                         | .                  | .            | (2.5) |
| $i$ -ésima iteración:     | $x_i = g(x_{i-1})$ | $f(x_i)$     |       |
| $i + 1$ -ésima iteración: | $x_{i+1} = g(x_i)$ | $f(x_{i+1})$ |       |
| .                         | .                  | .            |       |
| .                         | .                  | .            |       |
| .                         | .                  | .            |       |

Aunque hay excepciones, generalmente se encuentra que los valores  $x_0, x_1, x_2, \dots$  se van acercando a  $\bar{x}$  de manera que  $x_i$  está más cerca de  $\bar{x}$  que  $x_{i-1}$ , o bien se van alejando de  $\bar{x}$  de modo que cualquiera está más lejos que el valor anterior.

Si para la ecuación 2.3 se emplea  $x_0 = 2.0$  como valor inicial y las  $g(x)$  de los incisos a) y b) de la ecuación 2.4 se obtiene, respectivamente:



$$x_0 = 2; g(x) = 2x^2 - 5$$

| $i$ | $x_i$ | $g(x_i)$ |
|-----|-------|----------|
| 0   | 2     | 3        |
| 1   | 3     | 13       |
| 2   | 13    | 333      |
| 3   | 333   | 221773   |

$$x_0 = 2; g(x) = \sqrt{\frac{x+5}{2}}$$

| $i$ | $x_i$   | $g(x_i)$ |
|-----|---------|----------|
| 0   | 2.00000 | 1.87083  |
| 1   | 1.87083 | 1.85349  |
| 2   | 1.85349 | 1.85115  |
| 3   | 1.85115 | 1.85083  |

Puede apreciarse que la sucesión diverge con la  $g(x)$  del inciso a), y converge a la raíz 1.850781059 con la  $g(x)$  del inciso b).

Finalmente, para determinar si la sucesión  $x_0, x_1, x_2, \dots$  está convergiendo o divergiendo de una raíz  $\bar{x}$ , cuyo valor se desconoce, puede calcularse en el proceso 2.5 la sucesión  $f(x_0), f(x_1), f(x_2), \dots$ . Si dicha sucesión tiende a cero, el proceso 2.5 converge a  $\bar{x}$  y dicho proceso se continuará hasta que  $|f(x_i)| < \varepsilon_1$ , donde  $\varepsilon_1$  es un valor pequeño e indicativo de la exactitud o cercanía de  $x_i$  con  $\bar{x}$ . Se toma a  $x_i$  como la raíz y el problema de encontrar una raíz real queda concluido. Si por el contrario  $f(x_0), f(x_1), f(x_2), \dots$  no tiende a cero, la sucesión  $x_0, x_1, x_2, \dots$  diverge de  $\bar{x}$ , y el proceso deberá detenerse y ensayarse uno nuevo con una  $g(x)$  diferente.

### Ejemplo 2.1

Encuentre una aproximación a una raíz real de la ecuación

$$\cos x - 3x = 0$$

#### Solución

Dos posibilidades de  $g(x) = x$  son

$$a) x = \cos x - 2x \quad b) x = \cos x / 3$$

Graficando por separado las funciones  $\cos x$  y  $3x$ , se obtiene la figura 2.1

(Para graficar puede usar: el guión (script) de Matlab, las indicaciones para la TI-92 Plus o algún otro software comercial).

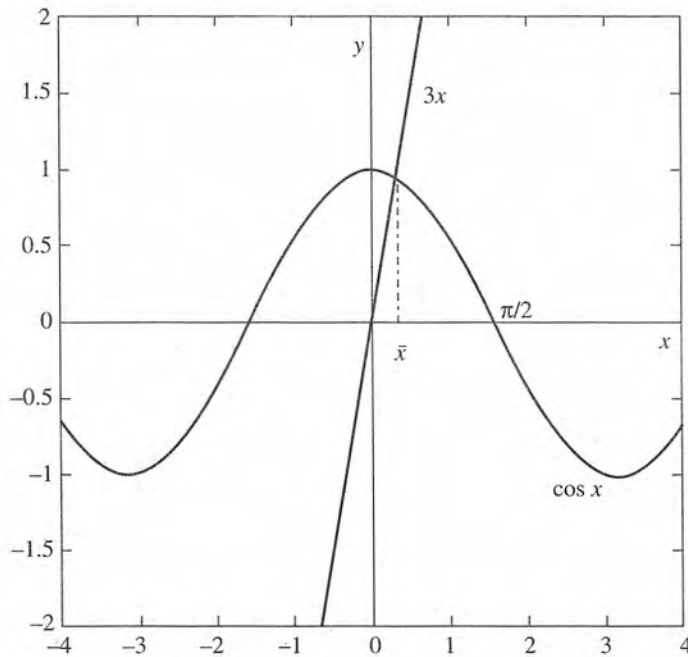


```
x= -4: 0.1:4;
y=cos (x);
z= 3*x;
t=zeros (size(x));
plot (x,y)
axis([-4 4 -2 2])
hold on
plot(x,z)
plot(x,t)
```

Invoque el editor Y=►W  
Escriba en y1 = la primera función a graficar: cos(x)  
Escriba en y2 = la segunda función a graficar: 3\*x  
Grafique con zoom estándar (F2 6)  
Lleve el cursor gráfico al punto donde se cruzan las dos funciones.  
Haga un acercamiento (F2 2)↓  
Use el trazador (F3) para ubicar la raíz.

De donde un valor cercano a  $\bar{x}$  es  $x_0 = (\pi/2) / 4^*$ . Iterando se obtiene para la forma del inciso a).

\* En el caso de funciones trigonométricas  $x$  debe estar en radianes.



**Figura 2.1**  
Gráfica de  $\cos x$   
y de  $3x$ .

| $i$ | $x_i$    | $g(x_i)$ | $ f(x_i) $ |
|-----|----------|----------|------------|
| 0   | $\pi/8$  | 0.13848  | 0.25422    |
| 1   | 0.13848  | 0.71346  | 0.57498    |
| 2   | 0.71346  | -0.67083 | 1.38429    |
| 3   | -0.67083 | 2.12496  | 2.79579    |
| 4   | 2.12496  | -4.77616 | 6.90113    |

Se detiene el proceso en la cuarta iteración, porque  $f(x_0), f(x_1), f(x_2), \dots$  no tiende a cero. Se emplea el valor absoluto de  $f(x)$  para manejar la idea de distancia.

Se inicia un nuevo proceso con  $x_0 = (\pi/2)/4$  y la forma equivalente del inciso b).

| $i$ | $x_i$   | $g(x_i)$ | $ f(x_i) $ |
|-----|---------|----------|------------|
| 0   | $\pi/8$ | 0.30796  | 0.25422    |
| 1   | 0.30796 | 0.31765  | 0.02907    |
| 2   | 0.31765 | 0.31666  | 0.00298    |
| 3   | 0.31666 | 0.31676  | 0.00031    |
| 4   | 0.31676 | 0.31675  | 0.00003    |

y la aproximación de la raíz es:

$$\bar{x} \approx x_4 = 0.31675$$

Para llevar a cabo los cálculos que se muestran en la tabla anterior, puede usar Matlab o la TI-92 Plus:



```
format long
x0=pi / 8 ;
for i = 1 : 5
    x=cos (x0) / 3;
    f=abs (cos (x0) - 3*x0);
    disp ( [x0, x, f] )
    x0=x;
end
```

```
E2_1()
Prgm
ClrIO : 3.1416/8→x0
For i, 1, 5
cos (x0) /3→x
abs (cos (x0) -3*x0)→f
string (x0) &" "&string(x)→a
a&" "&string(f)→a
Disp a: Pause : x→x0
EndFor
EndPrgm
```

Matlab posee una función que resuelve ecuaciones no lineales, suministrando la función y un valor inicial. Para este caso la instrucción quedaría:

`fzero('cos(x) -3*x', pi/8)`

con lo que se obtiene:

`ans = 0.3168`

y en formato largo (format long)

`ans = 0.31675082877122`

La calculadora TI-92 Plus también tiene una función que resuelve ecuaciones no lineales. La instrucción es:

`nSolve(cos(x) = 3*x, x)`

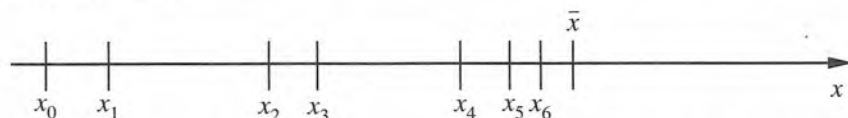
y el resultado es 0.316751

## CRITERIO DE CONVERGENCIA

Se estudiará un criterio más de convergencia del proceso iterativo 2.5, basado en que

$$g(\bar{x}) = \bar{x},$$

por lo cual puede suponerse que si la sucesión  $x_0, x_1, x_2, \dots$  converge a  $\bar{x}$ , los valores consecutivos  $x_i$  y  $x_{i+1}$  irán acercándose entre sí conforme el proceso iterativo avanza, como puede verse enseguida



Un modo práctico de saber si los valores consecutivos se acercan es ir calculando la distancia entre ellos

$$d_i = |x_{i+1} - x_i|$$

Si la sucesión  $d_1, d_2, d_3, \dots$  tiende a cero, puede pensarse que el proceso 2.5 está convergiendo a una raíz  $\bar{x}$  y debe continuarse hasta que  $d_i < \epsilon$ , y tomar a  $x_{i+1}$  como la raíz buscada. Si  $d_1, d_2, d_3, \dots$  no converge para un número "grande" de iteraciones (llámense



MAXIT), entonces  $x_0, x_1, x_2, \dots$  diverge de  $\bar{x}$ , y se detiene el proceso para iniciar uno nuevo, modificando la función  $g(x)$ , el valor inicial o ambos.

Este criterio de convergencia se utiliza ampliamente en el análisis numérico y resulta más sencillo de calcular que el que emplea la sucesión  $f(x_0), f(x_1), f(x_2), \dots$  pero también es menos seguro, como se verá más adelante.

Para finalizar esta sección se da un algoritmo del método de punto fijo en forma propia para lenguajes de programación.

### ALGORITMO 2.1 Método de punto fijo

Para encontrar una raíz real de la ecuación  $g(x) = x$  proporcionar la función  $G(X)$  y los

DATOS: Valor inicial  $X_0$ , criterio de convergencia EPS y número máximo de iteraciones MAXIT.

RESULTADOS: La raíz aproximada  $X$  o un mensaje de falla.

PASO 1. Hacer  $I = 1$

PASO 2. Mientras  $I < \text{MAXIT}$ , realizar los pasos 3 a 6.

PASO 3. Hacer  $X = G(X_0)$  (calcular  $(x_i)$ ).

PASO 4. Si  $|\text{ABS}(X - X_0)| \leq \text{EPS}$  entonces IMPRIMIR  $X$  y TERMINAR. De otro modo CONTINUAR.

PASO 5. Hacer  $I = I + 1$ .

PASO 6. Hacer  $X_0 = X$  (actualiza  $X_0$ ).

PASO 7. IMPRIMIR mensaje de falla: "EL MÉTODO NO CONVERGE A UNA RAÍZ" y TERMINAR.

### EL CRITERIO $|g'(x)| < 1$

Es importante analizar por qué algunas formas equivalentes  $x = g(x)$  de  $f(x) = 0$  conducen a una raíz en el método de punto fijo y otras no, aun empleando el mismo valor inicial en ambos casos.

Se inicia el análisis aplicando el teorema del punto medio\* a la función  $g(x)$  en el intervalo comprendido entre  $x_{i-1}$  y  $x_i$ .

$$g(x_i) - g(x_{i-1}) = g'(\xi_i)(x_i - x_{i-1}) \quad (2.6)$$

donde:

$$\xi_i \in (x_i, x_{i-1})$$

Como

$$g(x_i) = x_{i+1} \text{ y } g(x_{i-1}) = x_i$$

sustituyendo se obtiene:

$$x_{i+1} - x_i = g'(\xi_i)(x_i - x_{i-1})$$

Tomando valor absoluto en ambos miembros

$$|x_{i+1} - x_i| = |g'(\xi_i)| |x_i - x_{i-1}| \quad (2.7)$$

Para  $i = 1, 2, 3, \dots$  la ecuación 2.7 queda así:

$$|x_2 - x_1| = |g'(\xi_1)| |x_1 - x_0| \quad \xi_1 \in (x_1, x_0)$$

\* Se supone que  $g(x)$  satisface las condiciones de aplicabilidad de este teorema.

$$\begin{aligned}
 |x_3 - x_2| &= |g'(\xi_2)| |x_2 - x_1| & \xi_2 \in (x_2, x_1) \\
 |x_4 - x_3| &= |g'(\xi_3)| |x_3 - x_2| & \xi_3 \in (x_3, x_2) \\
 &\vdots \\
 &\vdots
 \end{aligned} \tag{2.8}$$

Supóngase ahora que en la región que comprende a  $x_0, x_1, \dots$  y en  $\bar{x}$  misma, la función  $g'(x)$  está acotada; esto es

$$|g'(x)| \leq M,$$

para algún número  $M$ . Entonces:

$$\begin{aligned}
 |x_2 - x_1| &\leq M |x_1 - x_0| \\
 |x_3 - x_2| &\leq M |x_2 - x_1| \\
 |x_4 - x_3| &\leq M |x_3 - x_2| \\
 &\vdots \\
 &\vdots
 \end{aligned} \tag{2.9}$$

Si se sustituye la primera desigualdad en la segunda, se tiene

$$|x_3 - x_2| \leq M |x_2 - x_1| \leq MM |x_1 - x_0|$$

o bien

$$|x_3 - x_2| \leq M^2 |x_1 - x_0|$$

Si se sustituye este resultado en la tercera desigualdad de la ecuación 2.9 se tiene

$$|x_4 - x_3| \leq M |x_3 - x_2| \leq MM^2 |x_1 - x_0|$$

o

$$|x_4 - x_3| \leq M^3 |x_1 - x_0|$$

Procediendo de igual manera se llega a

$$|x_{i+1} - x_i| \leq M^i |x_1 - x_0| \tag{2.10}$$

El proceso 2.5 puede converger por razones muy diversas, pero es evidente que si  $M < 1$ , dicho proceso convergirá, ya que  $M^i$  tenderá a cero al tender  $i$  a un número grande.

En conclusión, el proceso 2.5 puede converger si  $M$  es grande y convergirá si  $M < 1$  en un entorno de  $x$  que incluya  $x_0, x_1, x_2, \dots$ . Entonces  $M < 1$  es una condición suficiente, pero no necesaria para la convergencia.

Un método práctico de emplear este resultado es obtener distintas formas  $x = g(x)$  de  $f(x) = 0$ , y calcular  $|g'(x)|$ ; las que satisfagan el criterio  $|g'(x_0)| < 1$  prometerán convergencia al aplicar el proceso 2.5.

### Ejemplo 2.2

Calcule una raíz real de la ecuación\*

$$f(x) = x^3 + 2x^2 + 10x - 20 = 0,$$

empleando como valor inicial  $x_0 = 1$ .

\*Resuelta por Leonardo de Pisa en 1225.

**Solución** Dos formas  $x = g(x)$  de esta ecuación son



$$a) \ x = \frac{20}{x^2 + 2x + 10}$$

y

$$b) \ x = x^3 + 2x^2 + 11x - 20$$

de donde

$$g'(x) = \frac{-20(2x+2)}{(x^2+2x+10)^2}$$

y

$$g'(x) = 3x^2 + 4x + 11$$

Sustituyendo  $x_0 = 1$ .

$$|g'(1)| = \left| \frac{-80}{169} \right| = 0.47$$

y

$$|g'(1)| = 8$$

De donde la forma  $a)$  promete convergencia y la forma  $b)$  no.

Aplicando el proceso 2.5 y el criterio  $\varepsilon = 10^{-3}$  a  $|x_{i+1} - x_i|$  en caso de convergencia, se tiene:

| $i$ | $x_i$   | $ x_{i+1} - x_i $ | $ g'(x_i) $ |
|-----|---------|-------------------|-------------|
| 0   | 1.00000 |                   | 0.47337     |
| 1   | 1.53846 | 0.53846           | 0.42572     |
| 2   | 1.29502 | 0.24344           | 0.45100     |
| 3   | 1.40183 | 0.10681           | 0.44047     |
| 4   | 1.35421 | 0.04762           | 0.44529     |
| 5   | 1.37530 | 0.02109           | 0.44317     |
| 6   | 1.36593 | 0.00937           | 0.44412     |
| 7   | 1.37009 | 0.00416           | 0.44370     |
| 8   | 1.36824 | 0.00184           | 0.44389     |
| 9   | 1.36906 | 0.00082           | 0.44380     |

Para llevar a cabo los cálculos que se muestran en la tabla anterior, puede usarse Matlab o la TI-92 Plus.



```
format long
x0=1;
for i=1 : 9
    x=20/(x0^2+2*x0+10);
    dist=abs(x - x0);
    dg=abs(-20*(2*x+2)/(x^2+2*x+10)^2);
    disp([x, dist, dg])
    x0=x;
end
```

```
e2_2( )
Prgm
Define g(x)=20/(x^2+2*x+10)
ClrIO: 1→x0
For i, 1, 9
    g(x0)→x: abs(x - x0)→d
    Disp string(x)&" "&string(d)
    Pause: x→x0
EndFor
EndPrgm
```



Obsérvese que  $|g'(x_i)|$  se mantiene menor de uno. Una vez que  $|x_{i+1} - x_i| < 10^{-3}$ , se detiene el proceso y se toma como raíz a  $x_9$

$$\bar{x} \approx 1.36906$$

Si se hubiese tomado la forma equivalente

$$x = \frac{-x^3 - 2x^2 + 20}{10}$$

para la cual, se tiene

$$g'(x) = \frac{-3x^2 - 4x}{10}$$

y con  $x_0 = 1$

$$|g'(1)| = \left| \frac{-7}{10} \right| = 0.7,$$

lo cual indica posibilidad de convergencia, pero al aplicar el proceso 2.5 se tiene

| $i$ | $x_i$   | $ x_{i+1} - x_i $ | $ g'(x_i) $ |
|-----|---------|-------------------|-------------|
| 0   | 1.00000 |                   | 0.70000     |
| 1   | 1.70000 | 0.70000           | 1.54700     |
| 2   | 0.93070 | 0.76930           | 0.63214     |
| 3   | 1.74614 | 0.81544           | 1.61316     |
| 4   | 0.85780 | 0.88835           | 0.56386     |
| 5   | 1.78972 | 0.93192           | 1.67682     |

Una divergencia lenta, ya que  $|g'(x_i)|$  toma valores mayores de 1 en algunos puntos.

La condición de que el valor absoluto de  $g'(x)$  sea menor que 1 en la región que comprende la raíz buscada  $\bar{x}$  y los valores  $x_i$ , se interpreta geoméricamente a continuación.

En caso de contar con software comercial pueden graficarse las funciones  $g'(x)$  correspondientes a los incisos a) y b) y la recta  $y = x$ , y observar los valores de  $g'(x)$  en las  $x_i$  del proceso iterativo.

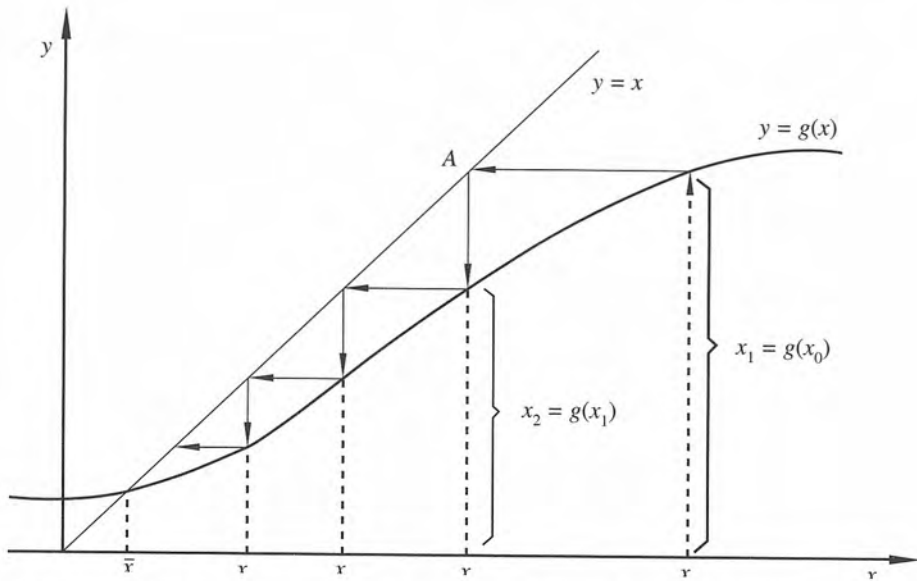
### INTERPRETACIÓN GEOMÉTRICA DE $|g'(x)| < 1$

Al graficar los dos miembros de la ecuación 2.2 como las funciones  $y = x$  y  $y = g(x)$ , la raíz buscada  $\bar{x}$  es la abscisa del punto de cruce de dichas funciones (véase Fig. 2.2).

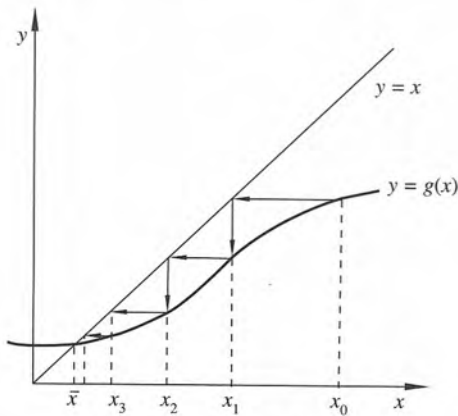
El proceso 2.5 queda geoméricamente representado en la figura 2.2, la cual muestra un caso de convergencia, ya que  $|g'(x)|$  es menor que 1 en  $x_0, x_1, \dots, \bar{x}$ .

Para ver esto se trazan las tangentes a  $g(x)$  en  $(x_0, x_1), (x_1, x_2), \dots$  y se observa que todas tienen un ángulo de inclinación menor que la función  $y = x$  cuya pendiente es 1.

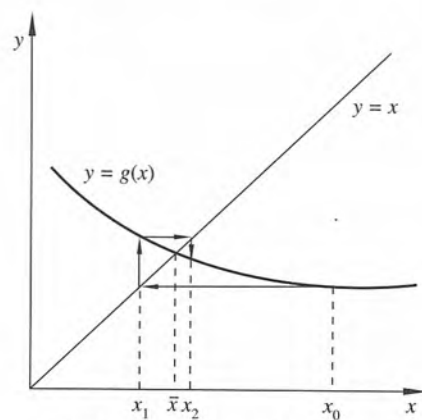
**Figura 2.2**  
Interpretación  
geométrica de  
 $|g'(x)| > 1$ .



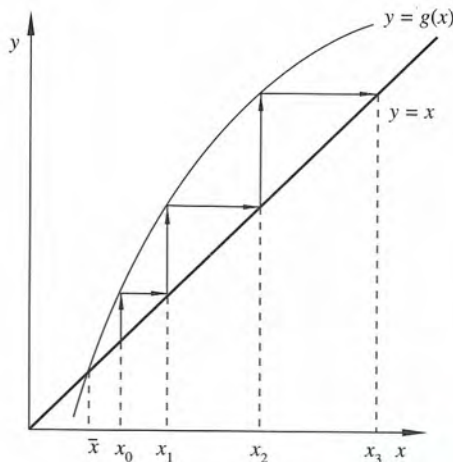
A continuación se presentan geoméricamente los casos posibles de convergencia y divergencia.



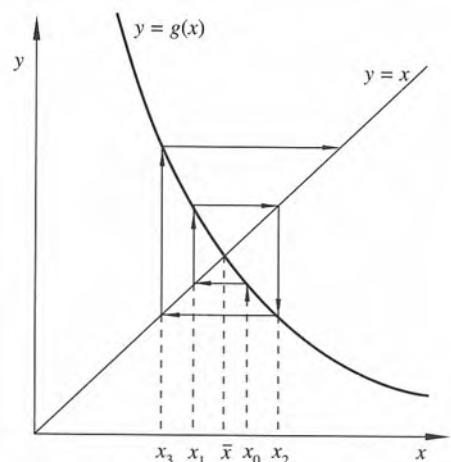
a) Convergencia monotónica



b) Convergencia oscilatoria



c) Divergencia monotónica



d) Divergencia oscilatoria

**Figura 2.3**  
Cuatro casos  
posibles de  
convergencia y  
divergencia en  
la iteración  
 $x = g(x)$ .

- Caso 1.** La figura 2.3a ilustra qué ocurre si  $g'(x)$  se encuentra entre 0 y 1. Incluso si  $x_0$  está lejos de la raíz  $\bar{x}$  —que se encuentra en el cruce de las curvas  $y = x$ ,  $y = g(x)$ — los valores sucesivos  $x_i$  se acercan a la raíz por un solo lado. Esto se conoce como convergencia monotónica.
- Caso 2.** La figura 2.3b muestra la situación en que  $g'(x)$  está entre  $-1$  y  $0$ . Aun si  $x_0$  está alejada de la raíz  $\bar{x}$ , los valores sucesivos  $x_i$  se aproximan por el lado derecho e izquierdo de la raíz. Esto se conoce como convergencia oscilatoria.
- Caso 3.** En la figura 2.3c se ve la divergencia cuando  $g'(x)$  es mayor que 1. Los valores sucesivos  $x_i$  se alejan de la raíz por un solo lado. Esto se conoce como divergencia monotónica.
- Caso 4.** La figura 2.3d presenta la divergencia cuando  $g'(x)$  es menor que  $-1$ . Los valores sucesivos  $x_i$  se alejan de la raíz oscilando alrededor de ella. Esto se conoce como divergencia oscilatoria.

Un excelente ejercicio es **crear** ecuaciones  $f(x) = 0$ , obtener para cada una de ellas varias alternativas de  $x = g(x)$  y graficarlas para obtener el punto de intersección (aproximación de la raíz); obtener las correspondientes  $g'(x)$  y graficarlas alrededor del punto de intersección. Una vez hecho esto se puede ver si alrededor de la raíz la gráfica de  $g'(x)$  queda dentro de la banda  $y = -1$  y  $y = 1$ ; de ser así, un valor inicial cercano a la raíz prometería convergencia. Si la gráfica de  $g'(x)$  queda fuera de la banda  $y = -1$ ,  $y = 1$  no sería recomendable iniciar el proceso iterativo.

### Ejemplo 2.3

Utilizando la ecuación del ejemplo 2.2 elabore las gráficas de las  $g'(x)$  de los incisos a y b. Agregue a las gráficas la banda constituida por  $y = -1$  y  $y = 1$ .

### Solución

$$a) \quad g(x) = \frac{20}{x^2 + 2x + 10}$$

$$g'(x) = \frac{-20(2x + 2)}{(x^2 + 2x + 10)^2}$$

Las gráficas de  $y = g(x)$  y  $y = x$  se intersectan alrededor de  $x = 1$ .

Utilizando Matlab o la TI-92 Plus se obtiene la gráfica de  $g'(x)$  y la banda  $y = -1$ ,  $y = 1$ .



```
x=0 : 0 . 05 : 2;
dg=-20*(2*x+2)/(x.^2+2*x+10).^2;
y=ones (size (x) ) ;
z=-ones (size (x) ) ;
ymin=nanmin (dg) ;
ymax=nanmax (dg) ;

if ymin > -1
```

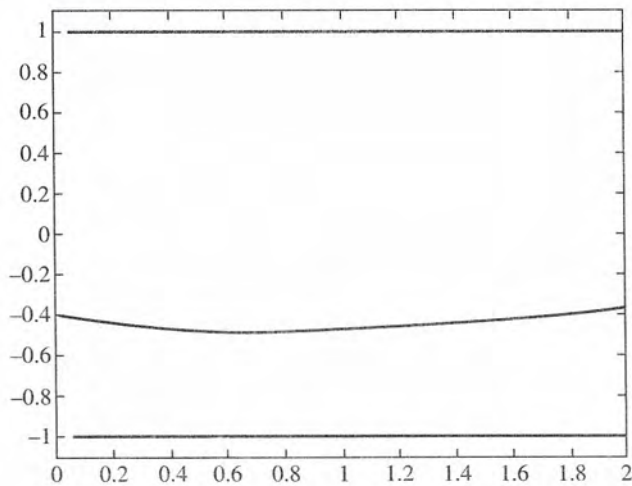
Invoke el editor Y= ♦W  
 Escriba en y1= la expresión de  $g'(x)$ :  
 $Y1=-20*(2*x+2)/(x^2+2*x+10)^2$   
 Escriba en y2= la cota inferior:  $y2=-1$   
 Escriba en y3= la cota superior:  $y3=1$   
 Muestre la gráfica con acercamiento  
 normal (F2 6).  
 Haga un acercamiento (F2 2.)



```

    ymin= -1.1;
end
if ymax < 1
    ymax=1.1 ;
end
plot (x, dg, 'k')
hold on
plot (x, y, 'k')
plot (x, z, 'k')
axis ([0 2 ymin ymax])

```



Como puede verse, la gráfica de  $g'(x)$  alrededor de  $x_0 = 1$  queda dentro de la banda  $y = 1$ ,  $y = -1$ , por lo que un valor inicial dentro del intervalo  $(0, 2)$  prometería convergencia, la cual se daría en caso de que la sucesión  $x_0, x_1, x_2, \dots, x_p, \dots$  sea tal que  $g'(x_i)$  se mantenga en dicha banda.

$$b) \quad g(x) = x^3 + 2x^2 + 11x - 20$$

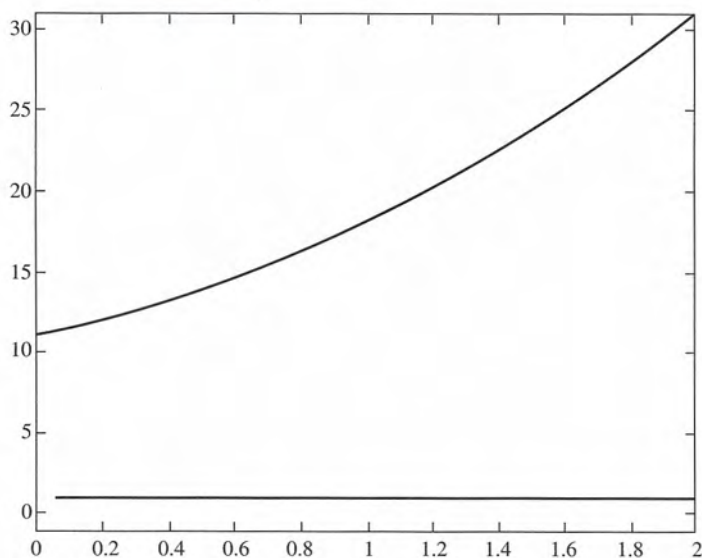
$$g'(x) = 3x^2 + 4x + 11$$

$$x_0 = 1$$

Utilizando el guión de Matlab anterior con el cambio (o el correspondiente para la TI-92 Plus)

$$dg = 3*x.^2 + 4*x + 11$$

se obtiene:



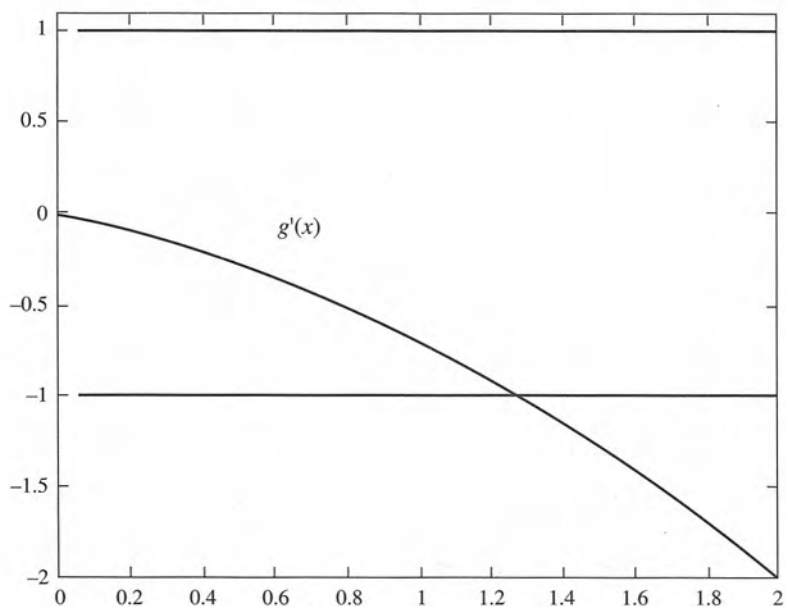
Como se puede observar, la gráfica de  $g'(x)$  queda totalmente fuera de la banda  $y = -1$ ,  $y = 1$ , por lo que no es recomendable utilizar esta  $g(x)$ .

Si por otro lado ensayamos la forma equivalente

$$g(x) = \frac{-x^3 - 2x^2 + 20}{10} \quad \text{con}$$

$$g'(x) = \frac{-3x^2 - 4x}{10}$$

La gráfica queda ahora:



Podría pensarse que al tomar un valor inicial dentro de la banda, por ejemplo  $x_0 = 1$  tendríamos posibilidades de convergencia. No obstante, en el proceso iterativo (ver ejemplo 2.2), se observa una divergencia oscilatoria. Explíquela utilizando la segunda tabla del ejemplo 2.2 y la gráfica anterior.

### ORDEN DE CONVERGENCIA

Ahora se verá que la magnitud de  $g'(x)$  no sólo indica si el proceso converge o no, sino que además puede usarse como indicador de cuán rápida es la convergencia.

Sea  $\epsilon_i$  el error en la  $i$ -ésima iteración; esto es

$$\epsilon_i = x_i - \bar{x}$$

Si se conoce el valor de la función  $g(x)$  y sus derivadas en  $\bar{x}$ , puede expandirse  $g(x)$  alrededor de  $\bar{x}$  en serie de Taylor y encontrar así el valor de  $g(x)$  en  $x_i$

$$g(x_i) = g(\bar{x}) + g'(\bar{x})(x_i - \bar{x}) + g''(\bar{x}) \frac{(x_i - \bar{x})^2}{2!} + g'''(\bar{x}) \frac{(x_i - \bar{x})^3}{3!} + \dots$$

o bien

$$g(x_i) - g(\bar{x}) = g'(\bar{x})(x_i - \bar{x}) + g''(\bar{x}) \frac{(x_i - \bar{x})^2}{2!} + g'''(\bar{x}) \frac{(x_i - \bar{x})^3}{3!} + \dots$$

Como

$$x_{i+1} = g(x_i)$$

y

$$\bar{x} = g(\bar{x}),$$

también puede escribirse la última ecuación como

$$x_{i+1} - \bar{x} = g'(\bar{x})\epsilon_i + g''(\bar{x}) \frac{\epsilon_i^2}{2!} + g'''(\bar{x}) \frac{\epsilon_i^3}{3!} + \dots$$

El miembro de la izquierda es el error en la  $(i+1)$ -ésima iteración y, por tanto, se expresa como  $\epsilon_{i+1}$  de modo que

$$\epsilon_{i+1} = g'(\bar{x})\epsilon_i + g''(\bar{x}) \frac{\epsilon_i^2}{2!} + g'''(\bar{x}) \frac{\epsilon_i^3}{3!} + \dots \quad (2.11)$$

donde puede observarse que si después de las primeras iteraciones  $\epsilon_i$  tiene un valor pequeño ( $|\epsilon_i| < 1$ ), entonces  $\epsilon_i^2, \epsilon_i^3, \epsilon_i^4, \dots$  serán valores más pequeños que  $|\epsilon_i|$ , de modo que si  $g'(\bar{x}) \neq 0$ , la magnitud del primer término de la ecuación 2.11 generalmente domina las de los demás términos y  $\epsilon_{i+1}$  es proporcional a  $\epsilon_i$ ; en cambio si  $g'(\bar{x}) = 0$  y  $g''(\bar{x}) \neq 0$ , la magnitud del segundo término de la ecuación 2.11 predomina sobre la de los términos restantes y  $\epsilon_{i+1}$  es proporcional a  $\epsilon_i^2$ . Si  $g'(\bar{x}) = g''(\bar{x}) = 0$  y  $g'''(\bar{x}) \neq 0$ ,  $\epsilon_{i+1}$  es proporcional a  $\epsilon_i^3$ , etcétera.

Se dice entonces que en caso de convergencia, el proceso 2.5 tiene orden uno si  $g'(\bar{x}) \neq 0$ , orden dos si  $g'(\bar{x}) = 0$  y  $g''(\bar{x}) \neq 0$ , orden tres si  $g'(\bar{x}) = g''(\bar{x}) = 0$  y  $g'''(\bar{x}) \neq 0$ , etc. Una vez determinado el orden  $n$  se tiene que  $\epsilon_{i+1} \propto \epsilon_i^n$  y el error  $\epsilon_{i+1}$  será más pequeño que  $\epsilon_i$  entre más grande sea  $n$  y la convergencia por tanto más rápida.



Obsérvese que en los ejemplos resueltos  $g'(\bar{x}) \neq 0$ , y el orden ha sido uno. Como al iniciar el proceso sólo se cuenta con  $x_0$  y algunas formas  $g(x)$ , puede obtenerse  $g'(x)$  para cada forma y las que satisfagan la condición  $|g'(x_0)| < 1$  prometerán convergencia. Dicha convergencia será más rápida para aquellas donde  $|g'(x_0)|$  sea más cercano a cero y más lenta entre más próximo esté dicho valor a 1. Así pues, para la ecuación 2.3, las formas 2.4 y el valor inicial  $x_0 = 2$  se obtiene, respectivamente:

- |                                                              |                       |
|--------------------------------------------------------------|-----------------------|
| a) $g'(x) = 4x$                                              | y $ g'(2)  = 8$       |
| b) $g'(x) = \frac{1}{4 \left( \frac{x+5}{2} \right)^{1/2}}$  | y $ g'(2)  = 0.1336$  |
| c) $g'(x) = \frac{-10}{(2x-1)^2}$                            | y $ g'(2)  = 1.111$   |
| d) $g'(x) = 4x$                                              | y $ g'(2)  = 8$       |
| e) $g'(x) = 1 - \frac{(4x-1)(4x-1) - (2x^2-x-5)4}{(4x-1)^2}$ | y $ g'(2)  = 0.08163$ |

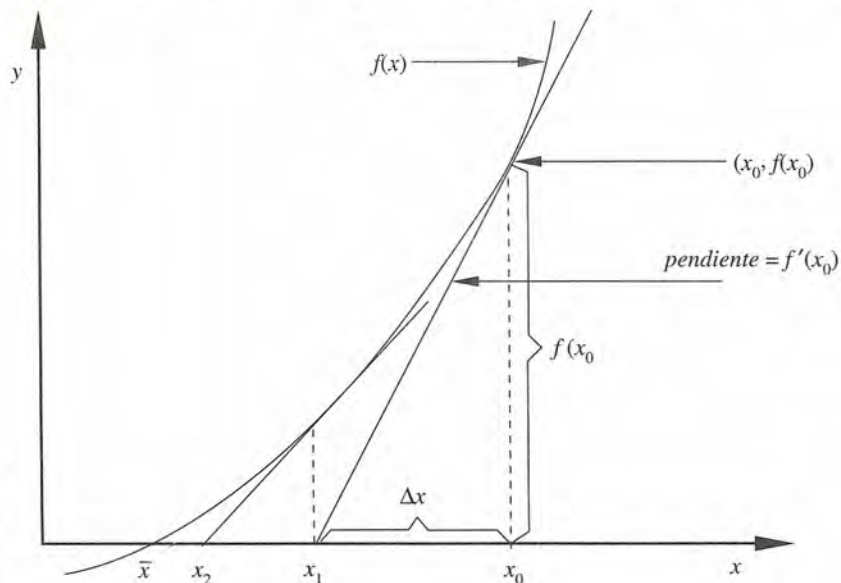
Las formas de los incisos b) y e) quedan con posibilidades de convergencia, y la e) como la mejor opción porque su valor está más cercano a cero.

Se deja al lector encontrar una raíz real de la ecuación 2.3 con el método de punto fijo, con la forma e) y detener la iteración una vez que  $|f(x_i)| \leq 10^{-4}$ , en caso de convergencia, o desde un principio si observa divergencia en las primeras iteraciones.

## 2.2 Método de Newton-Raphson

Ahora se estudiará un método de segundo orden de convergencia cuando se trata de raíces reales no repetidas. Consiste en un procedimiento que lleva la ecuación  $f(x) = 0$  a la forma  $x = g(x)$ , de modo que  $g'(\bar{x}) = 0$ . Su deducción se presenta enseguida.

En la figura 2.4 se tiene la gráfica de  $f(x)$  cuyo cruce con el eje  $x$  es una raíz real  $\bar{x}$ .



**Figura 2.4**  
Derivación del  
método de  
Newton-  
Raphson.

Vamos a suponer un valor inicial  $x_0$  que se sitúa en el eje horizontal. Trácese una tangente a la curva en el punto  $(x_0, f(x_0))$  y a partir de ese punto sígase por la tangente hasta su intersección con el eje  $x$ ; el punto de corte  $x_1$  es una nueva aproximación a  $\bar{x}$  (hay que observar que se ha reemplazado la curva  $f(x)$  por su tangente en  $(x_0, f(x_0))$ ). El proceso se repite comenzando con  $x_1$ , se obtiene una nueva aproximación  $x_2$  y así sucesivamente, hasta que un valor  $x_i$  satisfaga  $|f(x_i)| \leq \varepsilon_i$ ,  $|x_{i+1} - x_i| < \varepsilon$  o ambos. Si lo anterior no se cumpliera en un máximo de iteraciones (MAXIT), debe reiniciarse con un nuevo valor  $x_0$ .

La ecuación central del algoritmo se obtiene así

$$x_1 = x_0 - \Delta x$$

La pendiente de la tangente a la curva en el punto  $(x_0, f(x_0))$  es:

$$f'(x_0) = \frac{f(x_0)}{\Delta x}$$

así que

$$\Delta x = \frac{f(x_0)}{f'(x_0)}$$

y sustituyendo

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

en general

$$x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = g(x_i) \quad (2.12)$$

Este método es de orden 2, porque  $g'(\bar{x}) = 0$  y  $g''(\bar{x}) \neq 0$  (véase Prob. 2.11).

### Ejemplo 2.4

Encuentre una raíz real de la ecuación

$$f(x) = x^3 + 2x^2 + 10x - 20$$

mediante el método de Newton-Raphson,  $x_0 = 1$ , con  $\varepsilon = 10^{-3}$  aplicado a  $|x_{i+1} - x_i|$

**Solución** Se sustituyen  $f(x)$  y  $f'(x)$  en (2.12)



$$x_{i+1} = x_i - \frac{x_i^3 + 2x_i^2 + 10x_i - 20}{3x_i^2 + 4x_i + 10}$$

**Primera iteración**

$$x_1 = 1 - \frac{(1)^3 + 2(1)^2 + 10(1) - 20}{3(1)^2 + 4(1) + 10} = 1.41176$$

Como  $x_1 \neq x_0$ , se calcula  $x_2$

**Segunda iteración**

$$x_2 = 1.41176 - \frac{(1.41176)^3 + 2(1.41176)^2 + 10(1.41176) - 20}{3(1.41176)^2 + 4(1.41176) + 10} = 1.36934$$

Con este proceso se obtiene la tabla 2.1

**Tabla 2.1** Resultados del ejemplo 2.3.

| $i$ | $x_i$          | $ x_{i+1} - x_i $ | $ g'(x_i) $             |
|-----|----------------|-------------------|-------------------------|
| 0   | 1.00000        |                   | 0.24221                 |
| 1   | 1.41176        | 0.41176           | 0.02446                 |
| 2   | <u>1.36934</u> | 0.04243           | 0.00031                 |
| 3   | 1.36881        | 0.00053           | $4.6774 \times 10^{-8}$ |
| 4   | 1.36881        | 0.00000           | $9.992 \times 10^{-16}$ |

Para llevar a cabo los cálculos que se muestran en la tabla anterior, se puede emplear Matlab o la TI-92 Plus.



```
Format long
x0=1 ;
for I=1 : 4
    f=x0^3+2*x0^2+10*x0 - 20 ;
    df=3*x0^2+4*x0+10;
    x=x0 - f/df;
    dist = abs (x - x0) ;
    dg=abs (1 - ((3*x^2+4*x+10)^2 - ...
        (x^3+2*x^2+10*x-20)*(6*x+4))/ ...
        (3*x^2+4*x+10)^2);
    disp ([x, dist, dg])
    x0=x;
end
```

```
e2_4()
Prgm
Define f (x)=x^3+2*x^2+10*x-20
Define df (x) = 3*x^2+4*x+10
Define dg (x) = 1 - (df(x)^2-f
(x) * (6*x+4) ) / df (x)^2
ClrIO: 1.→x0
For i, 1, 4
    x0-f (x0) / df (x0)→x
    abs (x-x0)→dist
    Disp string(x)&" "&string(dist)&"
    "&string (abs (dg (x)))
    x→x0
EndFor
EndPrgm
```

Se requirieron sólo tres iteraciones para satisfacer el criterio de convergencia; además, se obtuvo una mejor aproximación a  $\bar{x}$  que en el ejemplo 2.2, ya que  $f(1.36881)$  se encuentra más cercana a cero que  $f(1.36906)$ , como se ve a continuación

$$f(1.36881) = (1.36881)^3 + 2(1.36881)^2 + 10(1.36881) - 20 = -0.00004$$

$$|f(1.36881)| = 0.00004 \quad \text{y} \quad |f(1.36906)| = 0.00531$$

Hay que observar que  $x_4$  ya no cambia con respecto a  $x_3$  en cinco cifras decimales y que  $g'(x_4)$  es prácticamente cero.



En el CD encontrará el **PROGRAMA 2.7** (Raíces de Ecuaciones), escrito para la versión 6 de Visual Basic. Con este programa se pueden resolver diferentes ecuaciones y obtener una visualización gráfica de los métodos de Newton-Raphson, de Bisección y de Posición Falsa; los últimos dos se verán más adelante.



**ALGORITMO 2.2** Método de Newton-Raphson

Para encontrar una raíz real de la ecuación  $f(x) = 0$ , proporcionar la función  $F(X)$  y su derivada  $DF(X)$  y los

DATOS: Valor inicial  $X_0$ , criterio de convergencia EPS, criterio de exactitud EPS1 y número máximo de iteraciones MAXIT.

RESULTADOS: La raíz aproximada  $X$  o un mensaje de falla.

PASO 1. Hacer  $I = 1$

PASO 2. Mientras  $I < \text{MAXIT}$ , repetir los pasos 3 a 7.

PASO 3. Hacer  $X = X_0 - F(X_0) / DF(X_0)$  (calcula  $x_i$ ).

PASO 4. Si  $\text{ABS}(X - X_0) < \text{EPS}$ , entonces IMPRIMIR  $X$  y TERMINAR. De otro modo CONTINUAR.

PASO 5. Si  $\text{ABS}(F(X)) < \text{EPS1}$ , entonces IMPRIMIR  $X$  y TERMINAR. De otro modo CONTINUAR.

PASO 6. Hacer  $I = I + 1$ .

PASO 7. Hacer  $X_0 = X$ .

PASO 8. IMPRIMIR mensaje de falla "EL MÉTODO NO CONVERGE A UNA RAÍZ" y TERMINAR.

**FALLAS DEL MÉTODO DE NEWTON-RAPHSON**

Cuando el método de Newton-Raphson converge se obtienen los resultados en relativamente pocas iteraciones, ya que para raíces no repetidas este método converge con orden 2 y el error  $\in_{i+1}$  es proporcional al cuadrado del error anterior\*  $\in_i$ . Para precisar más, supóngase que el error en una iteración es  $10^{-n}$ , el error siguiente —que es proporcional al cuadrado del error anterior— es entonces aproximadamente  $10^{-2n}$ , el que sigue será aproximadamente  $10^{-4n}$ , etc. De esto puede afirmarse que cada iteración duplica aproximadamente el número de dígitos correctos.

Sin embargo, algunas veces el método de Newton-Raphson no converge sino que oscila. Esto puede ocurrir si no hay raíz real como se ve en la figura 2.5a; si la raíz es un punto de inflexión como en la figura 2.5b, o si el valor inicial está muy alejado de la raíz buscada y alguna otra parte de la función "atrapa" la iteración, como en la figura 2.5c.

El método de Newton-Raphson requiere la evaluación de la primera derivada de  $f(x)$ . En la mayoría de los problemas de los textos este requisito es trivial, pero éste no es el caso en problemas reales donde, por ejemplo, la función  $f(x)$  está dada en forma tabular.

Es importante discutir algunos métodos para resolver  $f(x) = 0$  que no requieran el cálculo de  $f'(x)$ , pero que retengan algunas de las propiedades favorables de convergencia del método de Newton-Raphson. A continuación se estudian algunos métodos que tienen estas características y que se conocen como métodos de dos puntos.

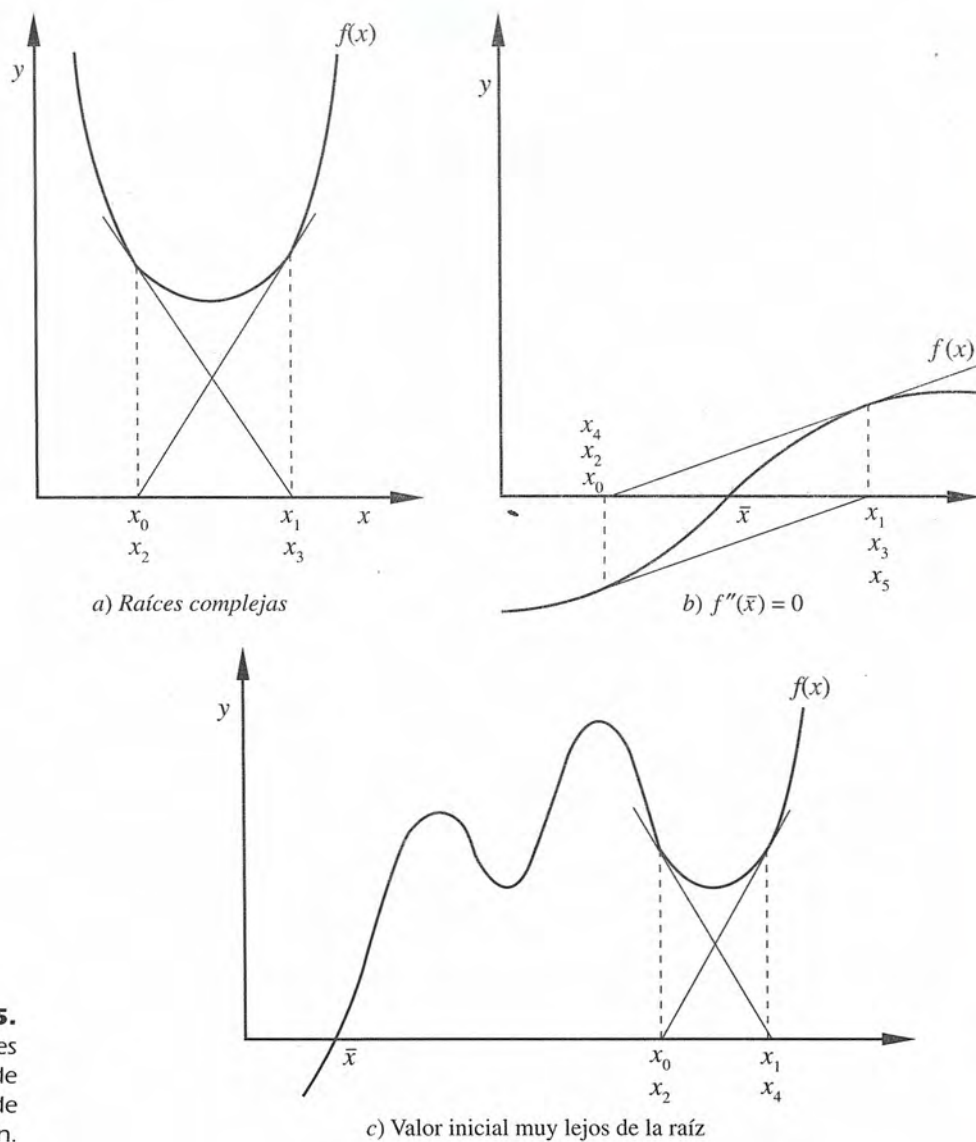
**2.3 Método de la secante**

El método de la secante consiste en aproximar la derivada  $f'(x_i)$  de la ecuación 2.12 por el cociente\*\*

$$\frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}},$$

\* Véase Prob. 2.13.

\*\* Nótese que este cociente es la derivada numérica de  $f(x)$ .



**Figura 2.5.**  
Funciones  
donde puede  
fallar el método de  
Newton-Raphson.

formado con los resultados de las dos iteraciones anteriores  $x_{i-1}$  y  $x_i$ . De esto resulta la fórmula

$$x_{i+1} = x_i - \frac{(x_i - x_{i-1}) f(x_i)}{f(x_i) - f(x_{i-1})} = g(x_i) \quad (2.13)$$

Para la primera aplicación de la ecuación 2.13 e iniciar el proceso iterativo, se requerirán dos valores iniciales:  $x_0$  y  $x_1$ .<sup>\*</sup> La siguiente aproximación,  $x_2$ , está dada por:

$$x_2 = x_1 - \frac{(x_1 - x_0) f(x_1)}{f(x_1) - f(x_0)},$$

<sup>\*</sup> Que pueden obtenerse por el método de punto fijo.

$x_3$  por:

$$x_3 = x_2 - \frac{(x_2 - x_1)f(x_2)}{f(x_2) - f(x_1)},$$

y así sucesivamente hasta que  $g(x_i) \approx x_{i+1}$  o una vez que

$$|x_{i+1} - x_i| < \varepsilon$$

o

$$|f(x_{i+1})| < \varepsilon_1$$

### Ejemplo 2.5

Use el método de la secante para encontrar una raíz real de la ecuación polinomial

$$f(x) = x^3 + 2x^2 + 10x - 20 = 0$$

**Solución** Con la ecuación 2.13 se obtiene



$$x_{i+1} = x_i - \frac{(x_i - x_{i-1})(x_i^3 + 2x_i^2 + 10x_i - 20)}{(x_i^3 + 2x_i^2 + 10x_i - 20) - (x_{i-1}^3 + 2x_{i-1}^2 + 10x_{i-1} - 20)}$$

Mediante  $x_0 = 0$  y  $x_1 = 1$  se calcula  $x_2$

$$x_2 = 1 - \frac{(1 - 0)(1^3 + 2(1)^2 + 10(1) - 20)}{(1^3 + 2(1)^2 + 10(1) - 20) - (0^3 + 2(0)^2 + 10(0) - 20)} = 1.53846$$

Los valores de las iteraciones subsecuentes se encuentran en la tabla 2.2. Si bien no se convergió a la raíz tan rápido como en el caso del método de Newton-Raphson, la velocidad de convergencia no es tan lenta como en el método de punto fijo (véase ejemplo 2.2); entonces se tiene para este ejemplo una velocidad de convergencia intermedia.

**Tabla 2.2** Resultados del ejemplo 2.4.

| $i$ | $x_i$   | $ x_{i+1} - x_i $ |
|-----|---------|-------------------|
| 0   | 0.00000 |                   |
| 1   | 1.00000 | 1.00000           |
| 2   | 1.53846 | 0.53846           |
| 3   | 1.35031 | 0.18815           |
| 4   | 1.36792 | 0.01761           |
| 5   | 1.36881 | 0.00090           |

$$|x_{i+1} - x_i| \leq \varepsilon = 10^{-3}$$

Para llevar a cabo los cálculos que se muestran en la tabla anterior, puede emplearse el siguiente guión de Matlab.





```
format long
x0=0 ; x1=1;
for i=1 : 4
    f0 = x0^3+2*x0^2+10*x0-20;
    f1 = x1^3+2*x1^2+10*x1 - 20;
    x2 = x1 - (x1 - x0) *f1 / (f1 - f0) ;
    dist=abs (x2-x1);
    disp ([x2, dist])
    x0=x1 ; x1=x2 ;
end
```

### ALGORITMO 2.3 Método de la secante

Para encontrar una raíz real de la ecuación  $f(x) = 0$ , dada  $f(x)$  analíticamente, proporcionar la función  $F(X)$  y los

**DATOS:** Valores iniciales  $X_0, X_1$ ; criterio de convergencia EPS, criterio de exactitud EPS1 y número máximo de iteraciones MAXIT.

**RESULTADOS:** La raíz aproximada  $X$  o un mensaje de falla.

PASO 1. Hacer  $I = 1$ .

PASO 2. Mientras  $I < \text{MAXIT}$ , repetir los pasos 3 a 8.

PASO 3. Hacer  $X = X_0 - (X_1 - X_0) * F(X_0) / (F(X_1) - F(X_0))$ .

PASO 4. Si  $\text{ABS}(X - X_1) < \text{EPS}$  entonces IMPRIMIR  $X$  y TERMINAR.

PASO 5. Si  $\text{ABS}(F(X)) < \text{EPS1}$  entonces IMPRIMIR  $X$  y TERMINAR.

PASO 6. Hacer  $X_0 = X_1$ .

PASO 7. Hacer  $X_1 = X$ .

PASO 8. Hacer  $I = I + 1$ .

PASO 9. IMPRIMIR mensaje de falla “EL MÉTODO NO CONVERGE A UNA RAÍZ” y TERMINAR.

### INTERPRETACIÓN GEOMÉTRICA DEL MÉTODO DE LA SECANTE

Los dos miembros de la ecuación  $x = g(x)$  se grafican por separado, como se ve en la figura 2.6.

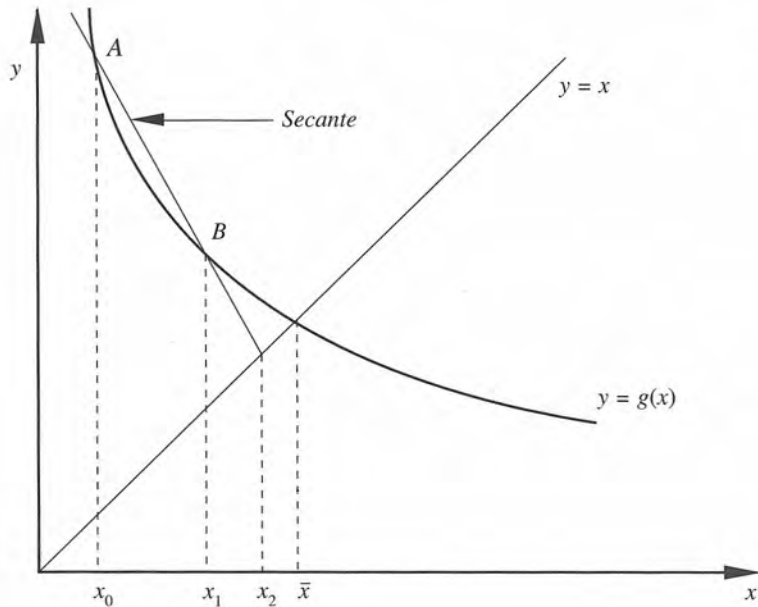
Se eligen dos puntos del eje  $x$ :  $x_0$  y  $x_1$  como primeras aproximaciones a  $\bar{x}$ .

Se evalúa  $g$  en  $x_0$  y en  $x_1$ , y se obtienen los puntos  $A$  y  $B$  de coordenadas  $(x_0, g(x_0))$  y  $(x_1, g(x_1))$ , respectivamente.

Los puntos  $A$  y  $B$  se unen con una línea recta [secante a la curva  $y = g(x)$ ] y se sigue por la secante hasta su intersección con la recta  $y = x$ . La abscisa correspondiente al punto de intersección es  $x_2$ , la nueva aproximación a  $\bar{x}$ .

Para obtener  $x_3$  se repite el proceso comenzando con  $x_1$  y  $x_2$  en lugar de  $x_0$  y  $x_1$ .

Este método **no garantiza** la convergencia a una raíz, lo cual puede lograrse con ciertas modificaciones que dan lugar a los métodos de posición falsa y de bisección.



**Figura 2.6**  
Interpretación  
geométrica del  
método de la  
secante.

## 2.4 Método de posición falsa

El método de posición falsa, también llamado de Regula-Falsi, al igual que el algoritmo de la secante, aproxima la derivada  $f'(x_i)$  de la ecuación 2.12 por el cociente

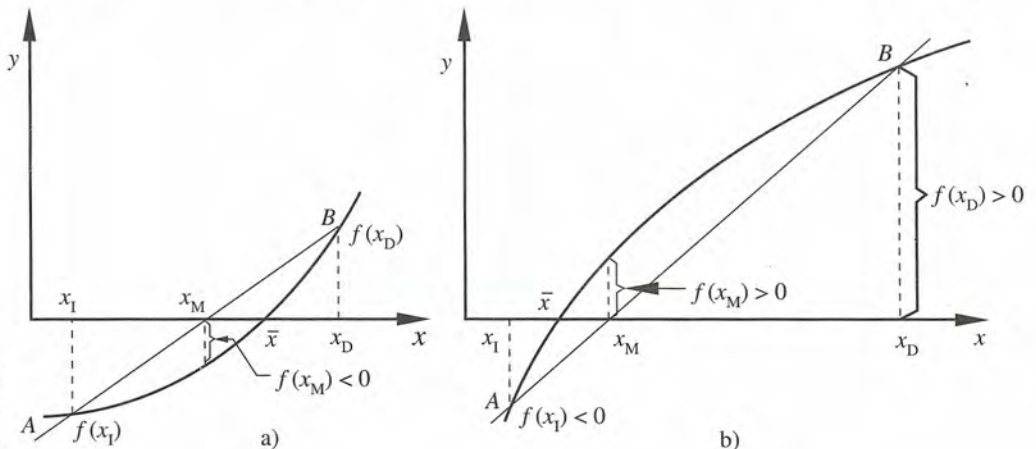
$$\frac{f(x_i) - f(x_{i-1})}{x_i - x_{i-1}}$$

pero en este caso los valores de  $x_i$  y  $x_{i-1}$  se encuentran en lados opuestos de la raíz buscada, de modo tal que sus valores funcionales correspondientes tienen signos opuestos, esto es:

$$f(x_i) \times f(x_{i-1}) < 0$$

Se denotan  $x_i$  y  $x_{i+1}$  como  $x_D$  y  $x_P$ , respectivamente.

Para ilustrar el método se utilizará la figura 2.7 y se partirá del hecho que se tienen dos valores iniciales  $x_D$  y  $x_I$  definidos arriba, y de que la función es continua en  $(x_P, x_D)$ .



**Figura 2.7**  
Método de  
posición falsa.

Se traza una línea recta que une los puntos  $A$  y  $B$  de coordenadas  $(x_I, f(x_I))$  y  $(x_D, f(x_D))$ , respectivamente. Se reemplaza  $f(x)$  en el intervalo  $(x_I, x_D)$  con el segmento de recta  $AB$  y el punto de intersección de este segmento con el eje  $x$ ,  $x_M$ , será la siguiente aproximación a  $\bar{x}$ .

Se evalúa  $f(x_M)$  y se compara su signo con el de  $f(x_D)$ . Si son iguales, se actualiza  $x_D$  sustituyendo su valor con el de  $x_M$ ; si los signos son diferentes, se actualiza  $x_I$  sustituyendo su valor con el de  $x_M$ . Nótese que el objetivo es mantener los valores descritos  $(x_D$  y  $x_I)$  cada vez más cercanos entre sí y la raíz entre ellos.

Se traza una nueva línea secante entre los puntos actuales  $A$  y  $B$ , y se repite el proceso hasta que se satisfaga el criterio de exactitud  $|f(x_M)| < \varepsilon_1$  tomándose como aproximación a  $x$  el valor último de  $x_M$ . Para terminar el proceso también puede usarse el criterio  $|x_D - x_I| < \varepsilon$ . En este caso se toma como aproximación a  $\bar{x}$  la media entre  $x_D$  y  $x_I$ .

Para calcular el valor de  $x_M$  se sustituye  $x_D$  por  $x_I$  y  $x_I$  por  $x_{I-1}$  en la ecuación 2.13, con lo que se llega a

$$x_M = x_D - \frac{(x_D - x_I) f(x_D)}{f(x_D) - f(x_I)} = \frac{x_I f(x_D) - x_D f(x_I)}{f(x_D) - f(x_I)} \quad (2.14)$$

el algoritmo de posición falsa.

### Ejemplo 2.6

Utilice el método de posición falsa para obtener una raíz real del polinomio

$$f(x) = x^3 + 2x^2 + 10x - 20$$

### Solución



Para obtener  $x_I$  y  $x_D$  se puede, por ejemplo, evaluar la función en algunos puntos donde este cálculo sea fácil o bien se grafica. Así:

$$\begin{aligned} f(0) &= -20 \\ f(1) &= -7 \\ f(-1) &= -29 \\ f(2) &= 16 \end{aligned}$$

De acuerdo con el teorema de Bolzano hay una raíz real, por lo menos, en el intervalo  $(1, 2)$ ; por tanto,

$$\begin{aligned} x_I &= 1 ; f(x_I) = -7 \\ x_D &= 2 ; f(x_D) = 16 \end{aligned}$$

Al aplicar la ecuación 2.14 se obtiene  $x_M$

$$x_M = 2 - \frac{(2 - 1)(16)}{16 - (-7)} = 1.30435$$

y

$$f(x_M) = (1.30435)^3 + 2(1.30435)^2 + 10(1.30435) - 20 = -1.33476$$

Como  $f(x_M) < 0$  (igual signo que  $f(x_I)$ ), se reemplaza el valor de  $x_I$  con el de  $x_M$ , con lo cual queda el nuevo intervalo como  $(1.30435, 2)$ . Por tanto:

$$x_I = 1.30435 ; f(x_I) = -1.33476$$

$$x_D = 2 ; f(x_D) = 16$$



Se calcula una nueva  $x_M$

$$x_M = 2 - \frac{(2 - 1.30435) 16}{16 - (-1.33476)} = 1.35791,$$

$$f(x_M) = (1.35791)^3 + 2(1.35791)^2 + 10(1.35791) - 20 = -0.22914$$

Como  $f(x_M) < 0$ , el valor actual de  $x_I$  se reemplaza con el último valor de  $x_M$ ; así el intervalo queda reducido a  $(1.35791, 2)$ . La tabla 2.3 muestra los cálculos llevados a cabo hasta satisfacer el criterio de exactitud

$$|f(x_M)| < 10^{-3}$$

**Tabla 2.3** Resultados del ejemplo 2.5.

| $i$ | $x_I$   | $x_D$   | $x_M$   | $ f(x_M) $ |
|-----|---------|---------|---------|------------|
| 0   | 1.00000 | 2.00000 |         |            |
| 1   | 1.00000 | 2.00000 | 1.30435 | 1.33476    |
| 2   | 1.30435 | 2.00000 | 1.35791 | 0.22914    |
| 3   | 1.35791 | 2.00000 | 1.36698 | 0.03859    |
| 4   | 1.36698 | 2.00000 | 1.36850 | 0.00648    |
| 5   | 1.36850 | 2.00000 | 1.36876 | 0.00109    |
| 6   | 1.36876 | 2.00000 | 1.36880 | 0.00018    |

Para llevar a cabo los cálculos que se muestran en la tabla anterior puede emplearse Matlab o la TI-92 Plus.



```
format long
xi=1; xd= 2; Eps= 0.001 ;
fi=xi^3+2*xi^2+10*xi-20;
fd=xd^3+2*xd^2+10*xd-20;
fm=1;
while abs (fm) > Eps
    xm=xd-fd*(xd-xi)/(fd-fi);
    fm=xm^3+2*xm^2+10*xm-20;
    disp ([xi, xd, xm, abs (fm) ])
    if fd*fm > 0 xd=xm; fd=fm;
    else xi=xm; fi=fm;
end
```

```
e2_6 ( )
Prgm
Define f (x)= x^3+2*x^2+10*x-20
ClrIO: 1.→xi: 2.→xd: 0.001→Eps
Disp " xi xd xm |f(xm)|"
Loop
    xd-f (xd)*(xd-xi)/(f(xd)-f(xi))→xm
    format (xi,"f5")&" "&format (xd,"f5")→a
    a&" "&format (xm, "f5") &" "→a
    a&format (abs (f (xm), "f6")→a
    disp a
    If abs (f(xm)) < Eps
Exit
If F(xd)*F(xm) > 0 Then
    xm→xi : Else: xm→xd
EndIf
EndLoop
EndPrgm
```

**NOTA** El GC proporciona los métodos de punto fijo, Newton-Raphson, posición falsa y bisección, de modo tal que pueden verse las iteraciones gráfica y numéricamente al resolver una ecuación dada. También hay calculadoras que disponen de algunos de estos métodos con los cuales auxiliarse.

### ALGORITMO 2.4 Método de posición falsa

Para encontrar una raíz real de la ecuación  $f(x) = 0$ , dada  $f(x)$  analíticamente, proporcionar la función  $F(X)$  y los

**DATOS:** Valores iniciales  $XI$  y  $XD$  que forman un intervalo, en donde se halla una raíz  $\bar{x}$  ( $F(XI) * F(XD) < 0$ ), criterio de convergencia EPS, criterio de exactitud EPS1 y número máximo de iteraciones MAXIT.

**RESULTADOS:** La raíz aproximada  $X$  o un mensaje de falla.

PASO 1. Hacer  $I = 1$ ;  $FI = F(XI)$ ;  $FD = F(XD)$ .

PASO 2. Mientras  $I < \text{MAXIT}$ , repetir los pasos 3 a 8.

PASO 3. Hacer  $XM = (XI * FD - XD * FI) / (FD - FI)$ ;  $FM = F(XM)$ .

PASO 4. Si  $\text{ABS}(FM) < \text{EPS1}$ , entonces IMPRIMIR  $XM$  y TERMINAR.

PASO 5. Si  $\text{ABS}(XD - XI) < \text{EPS}$ , entonces hacer  $XM = (XD + XI) / 2$ ; IMPRIMIR "LA RAÍZ BUSCADA ES", IMPRIMIR  $XM$  y TERMINAR.

PASO 6. Si  $FD * FM > 0$ , hacer  $XD = XM$  (actualiza  $XD$ ) y  $FD = FM$  (actualiza  $FD$ ).

PASO 7. Si  $FD * FM < 0$ , hacer  $XI = XM$  (actualiza  $XI$ ) y  $FI = FM$  (actualiza  $FI$ ).

PASO 8. Hacer  $I = I + 1$ .

PASO 9. IMPRIMIR mensaje de falla "EL MÉTODO NO CONVERGE A UNA RAÍZ" y TERMINAR.

## 2.5 Método de la bisección

El método de la bisección es muy similar al de posición falsa, aunque algo más simple. Como en el método de posición falsa también se requieren dos valores iniciales para ambos lados de la raíz, y que sus valores funcionales correspondientes sean de signos opuestos.

En este caso, el valor de  $x_M$  se obtiene como el punto medio entre  $x_I$  y  $x_D$ .

$$x_M = (x_I + x_D) / 2$$

Dependiendo de la función que se tenga en particular, el método de la bisección puede converger ligeramente más rápido o más lentamente que el método de posición falsa. Su gran ventaja sobre el método de posición falsa es que proporciona el tamaño exacto del intervalo en cada iteración (en ausencia de errores de redondeo). Para aclarar esto, nótese que en este método después de cada iteración el tamaño del intervalo se reduce a la mitad; después de  $n$  iteraciones, el intervalo original se habrá reducido  $2^n$  veces. Por lo anterior, si el intervalo original es de tamaño  $a$  y el criterio de convergencia aplicado al valor absoluto de la diferencia de dos  $x_M$  consecutivas es  $\varepsilon$ , entonces se requerirán  $n$  iteraciones, donde  $n$  se calcula con la igualdad de la expresión

$$\frac{a}{2^n} \leq \varepsilon,$$

de donde:

$$n = \frac{\ln a - \ln \varepsilon}{\ln 2} \quad (2.15)$$

Por esto se dice que se puede saber de antemano cuántas iteraciones se requieren.

### Ejemplo 2.7

Utilice el método de la bisección para obtener una raíz real del polinomio

$$f(x) = x^3 + 2x^2 + 10x - 20$$

**Solución** Con los valores iniciales obtenidos en el ejemplo 2.6

$$x_I = 1 ; f(x_I) = -7,$$

$$x_D = 2 ; f(x_D) = 16,$$

Si  $\varepsilon = 10^{-3}$ , el número de iteraciones  $n$  será

$$n = \frac{\ln a - \ln \varepsilon}{\ln 2} = \frac{\ln (2 - 1) - \ln 10^{-3}}{\ln 2} = 9.96$$

o bien

$$n \approx 10$$

#### Primera iteración

$$x_M = \frac{1 + 2}{2} = 1.5$$

$$f(1.5) = 2.88$$

Como  $f(x_M) > 0$  (distinto signo de  $f(x_I)$ ), se reemplaza el valor de  $x_I$  con el de  $x_M$ , con lo cual queda un nuevo intervalo (1, 1.5). Entonces:

$$x_D = 1 ; f(x_D) = -7$$

$$x_D = 1.5 ; f(x_D) = 2.88$$

#### Segunda iteración

$$x_M = \frac{1 + 1.5}{2} = 1.25$$

y

$$f(x_M) = -2.42$$

Como ahora  $f(x_M) < 0$  (igual signo que  $f(x_I)$ ), se reemplaza el valor de  $x_D$  con el valor de la nueva  $x_M$ ; de esta manera queda como intervalo (1.25, 1.5).

La tabla 2.4 muestra los cálculos, llevados a cabo trece veces, a fin de hacer ciertas observaciones.

El criterio  $|x_{i+1} - x_i| \leq 10^{-3}$  se satisface en diez iteraciones.

Nótese que si  $\varepsilon$  se hubiese aplicado sobre  $|f(x_M)|$ , se habrían requerido 13 iteraciones en lugar de 10. En general, se necesitarán más iteraciones para satisfacer un valor de  $\varepsilon$  sobre  $|f(x_M)|$  que cuando se aplica a  $|x_{i+1} - x_i|$ .



**Tabla 2.4** Resultados del ejemplo 2.7

| $i$ | $x_I$   | $x_D$   | $x_M$   | $ x_{Mi} - x_{Mi+1} $ | $ f(x_M) $ |
|-----|---------|---------|---------|-----------------------|------------|
| 0   | 1.00000 | 2.00000 |         |                       |            |
| 1   | 1.00000 | 2.00000 | 1.50000 |                       | 2.87500    |
| 2   | 1.00000 | 1.50000 | 1.25000 | 0.25000               | 2.42188    |
| 3   | 1.25000 | 1.50000 | 1.37500 | 0.12500               | 0.13086    |
| 4   | 1.25000 | 1.37500 | 1.31250 | 0.06250               | 1.16870    |
| 5   | 1.31250 | 1.37500 | 1.34375 | 0.03125               | 0.52481    |
| 6   | 1.34375 | 1.37500 | 1.35938 | 0.01563               | 0.19846    |
| 7   | 1.35938 | 1.37500 | 1.36719 | 0.00781               | 0.03417    |
| 8   | 1.36719 | 1.37500 | 1.37109 | 0.00391               | 0.04825    |
| 9   | 1.36719 | 1.37109 | 1.36914 | 0.00195               | 0.00702    |
| 10  | 1.36719 | 1.36914 | 1.36816 | 0.00098               | 0.01358    |
| 11  | 1.36816 | 1.36914 | 1.36865 | 0.00049               | 0.00329    |
| 12  | 1.36865 | 1.36914 | 1.36890 | 0.00025               | 0.00186    |
| 13  | 1.36865 | 1.36890 | 1.36877 | 0.00013               | 0.00071    |

Utilizando el guión de Matlab del ejemplo 2.6, con la modificación apropiada, puede obtenerse la tabla anterior.

## 2.6 Problemas de los métodos de dos puntos y orden de convergencia

A continuación se mencionan algunos problemas que se presentan en la aplicación de los métodos de dos puntos.

1. El hecho de requerir dos valores iniciales. Esto resulta imposible de satisfacer (en bisección y posición falsa) si se tienen raíces repetidas por parejas ( $\bar{x}_1$  y  $\bar{x}_2$ ), o muy difícil si la raíz buscada se encuentra muy cerca de otra ( $\bar{x}_3$  y  $\bar{x}_4$ ) (véase Fig. 2.8). En el último caso, uno de los valores iniciales debe estar entre las dos raíces o de otra manera no se detectará ninguna de ellas.\*
2. Debido a los errores de redondeo  $f(x_M)$  se calcula con un ligero error. Esto no es un problema sino hasta que  $x_M$  está muy cerca de la raíz  $\bar{x}$ , y  $f(x_M)$  resulta ser positiva cuando debería ser negativa o viceversa, o bien resulta ser cero.
3. En el método de la secante no hay necesidad de tener valores iniciales para ambos lados de la raíz que se busca. Esto constituye una ventaja, pero puede ser peligroso, ya que en la ecuación 2.13

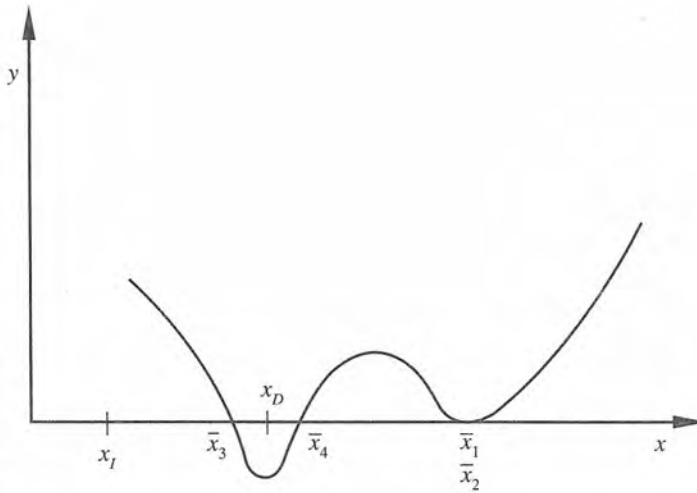
\* Para estos casos un graficador con capacidad de acercamiento (zoom) y rastreo (trace) puede ser de ayuda.

$$x_{i+1} = x_i - \frac{(x_i - x_{i-1})f(x_i)}{f(x_i) - f(x_{i-1})} = \frac{f(x_i)x_{i-1} - f(x_{i-1})x_i}{f(x_i) - f(x_{i-1})}$$

la diferencia

$$f(x_i) - f(x_{i-1})$$

**Figura 2.8**  
Raíces repetidas  
por parejas y  
muy cercanas  
entre sí.



puede causar problemas al evaluar  $x_{i-1}$ , pues  $f(x_i)$  y  $f(x_{i-1})$  no tienen necesariamente signos opuestos y su diferencia al ser muy cercana a cero producir *overflow*. Por último, debe decirse que en el método de la secante no hay certeza de convergencia.

### ORDEN DE CONVERGENCIA

Se determinará el orden de convergencia del método de la secante solamente, ya que para los demás métodos de dos puntos vistos, se siguen las mismas ideas.

Si, como antes,  $\epsilon_i$  representa el error en la  $i$ -ésima iteración

$$\epsilon_{i-1} = x_{i-1} - \bar{x}$$

$$\epsilon_i = x_i - \bar{x}$$

$$\epsilon_{i+1} = x_{i+1} - \bar{x}$$

Al sustituir en la ecuación 2.13  $x_{i+1}$ ,  $x_i$ ,  $x_{i-1}$  despejadas de las ecuaciones de arriba, se tiene:

$$\bar{x} + \epsilon_{i+1} = \bar{x} + \epsilon_i - \frac{(\bar{x} + \epsilon_i - \bar{x} - \epsilon_{i-1})f(\bar{x} + \epsilon_i)}{f(\bar{x} + \epsilon_i) - f(\bar{x} + \epsilon_{i-1})}$$

o bien

$$\epsilon_{i+1} = \epsilon_i - \frac{(\epsilon_i - \epsilon_{i-1})f(\bar{x} + \epsilon_i)}{f(\bar{x} + \epsilon_i) - f(\bar{x} + \epsilon_{i-1})} \quad (2.17)$$

Si se expande en serie de Taylor a  $f(\bar{x} + \epsilon_i)$  y  $f(\bar{x} + \epsilon_{i-1})$  alrededor de  $\bar{x}$  se tiene:

$$f(\bar{x} + \epsilon_i) = f(\bar{x}) + \epsilon_i f'(\bar{x}) + \frac{\epsilon_i^2}{2!} f''(\bar{x}) + \dots$$

$$f(\epsilon_{i-1} + \bar{x}) = f(\bar{x}) + \epsilon_{i-1} f'(\bar{x}) + \frac{\epsilon_{i-1}^2}{2!} f''(\bar{x}) + \dots$$

Sustituyendo estas expansiones en la ecuación 2.17 y como  $f(\bar{x}) = 0$ , queda:

$$\epsilon_{i+1} = \epsilon_i - \frac{(\epsilon_i - \epsilon_{i-1})(\epsilon_i f'(\bar{x}) + \epsilon_i^2 f''(\bar{x})/2! + \dots)}{(\epsilon_i - \epsilon_{i-1})f'(\bar{x}) + \frac{1}{2!}(\epsilon_i^2 - \epsilon_{i-1}^2)f''(\bar{x}) + \dots}$$

Factorizando a  $(\epsilon_i - \epsilon_{i-1})$  en el denominador y cancelándolo con el mismo factor del numerador queda:

$$\begin{aligned} \epsilon_{i+1} &= \epsilon_i - \frac{(\epsilon_i f'(\bar{x}) + \epsilon_i^2 f''(\bar{x})/2! + \dots)}{f'(\bar{x}) + \frac{1}{2!}(\epsilon_i + \epsilon_{i-1})f''(\bar{x}) + \dots} \\ &= \epsilon_i - \frac{(\epsilon_i f'(\bar{x}) + \epsilon_i^2 f''(\bar{x})/2! + \dots)}{f'(\bar{x})} \left(1 + \frac{1}{2!}(\epsilon_i + \epsilon_{i-1}) \frac{f''(\bar{x})}{f'(\bar{x})} + \dots\right)^{-1} \end{aligned}$$

Por el teorema binomial:

$$\begin{aligned} \epsilon_{i+1} &= \epsilon_i - \frac{1}{f'(\bar{x})} (\epsilon_i f'(\bar{x}) + \frac{\epsilon_i^2}{2!} f''(\bar{x}) + \dots) \left(1 - \frac{1}{2!}(\epsilon_i + \epsilon_{i-1}) \frac{f''(\bar{x})}{f'(\bar{x})} + \dots\right) \\ &= \epsilon_i - \frac{1}{f'(\bar{x})} (\epsilon_i f'(\bar{x}) + \frac{1}{2!} \epsilon_i^2 f''(\bar{x}) + \dots - \frac{\epsilon_i}{2!} (\epsilon_i + \epsilon_{i-1}) f''(\bar{x}) + \dots) \\ &= \epsilon_i - \frac{1}{f'(\bar{x})} (\epsilon_i f'(\bar{x}) - \frac{1}{2!} \epsilon_i \epsilon_{i-1} f''(\bar{x}) + \dots) \\ &= \frac{1}{2!} \epsilon_i \epsilon_{i-1} \frac{f''(\bar{x})}{f'(\bar{x})} + \dots \end{aligned}$$

o bien:

$$\epsilon_{i+1} \approx \frac{1}{2!} \frac{f''(\bar{x})}{f'(\bar{x})} \epsilon_i \epsilon_{i-1},$$

donde se aprecia que el error en la  $(i + 1)$ -ésima iteración es proporcional al producto de los errores de las dos iteraciones previas.

El error en el método de Newton-Raphson está dado así (véase Prob. 2.13)

$$\epsilon_{i+1} \approx \frac{f''(\bar{x})}{2! f'(\bar{x})} \epsilon_i^2,$$

donde por comparación puede observarse que el error en el método de la secante es ligeramente mayor que en el de Newton-Raphson; por tanto, su orden de convergencia será ligeramente menor, pero con la ventaja de que no hay que derivar la función  $f(x)$ .

Por otro lado, en los métodos de primer orden el error en la iteración  $(i + 1)$ -ésima es proporcional al error de la iteración previa solamente, por lo que puede decirse que los métodos de dos puntos son **superlineales** (orden de convergencia mayor de uno pero menor de dos).



## 2.7 Aceleración de convergencia

Se han visto métodos cuyo orden de convergencia es uno y dos, o bien un valor intermedio (superlineales). Existen métodos de orden 3 (véase Prob. 2.14) y de orden superior; sin embargo, es importante dar otro giro a la búsqueda de raíces reales y averiguar si la convergencia de los métodos vistos se puede acelerar.

### MÉTODOS DE UN PUNTO

Si en alguno de los métodos vistos se tiene que la sucesión  $x_0, x_1, x_2, \dots$ , converge muy lentamente a la raíz buscada, pueden tomarse, entre otras, las siguientes decisiones:

- Continuar el proceso hasta satisfacer alguno de los criterios de convergencia preestablecidos.
- Ensayar con una  $g(x)$  distinta; es decir, buscar una nueva  $g(x)$  en punto fijo o cambiar de método.
- Utilizar la sucesión de valores  $x_0, x_1, x_2, \dots$  para generar otra sucesión:  $x_0', x_1', x_2', \dots$  que converja más rápidamente a la raíz  $\bar{x}$  que se busca.

Los incisos *a)* y *b)* son suficientemente claros, mientras que la sucesión  $x_0', x_1', x_2', \dots$  de la parte *c)* se basa en que en ciertas condiciones de  $g'(x)$ ,\* se tiene que

$$\lim_{i \rightarrow \infty} \frac{\epsilon_{i+1}}{\epsilon_i} = g'(\bar{x}) \quad (2.18)$$

donde:  $\epsilon_i = x_i - \bar{x}$  es el error en la  $i$ -ésima iteración.

Para valores finitos de  $i$ , la ecuación 2.18 puede escribirse como:

$$\frac{\epsilon_{i+1}}{\epsilon_i} \approx g'(\bar{x})$$

o

$$x_{i+1} - \bar{x} \approx g'(\bar{x}) (x_i - \bar{x}) \quad (2.19)$$

o también:

$$x_{i+2} - \bar{x} \approx g'(\bar{x}) (x_{i+1} - \bar{x}) \quad (2.20)$$

Restando la ecuación 2.19 de la 2.20 se tiene:

$$x_{i+2} - x_{i+1} \approx g'(\bar{x}) (x_{i+1} - x_i),$$

de donde:

$$g'(\bar{x}) \approx \frac{x_{i+2} - x_{i+1}}{x_{i+1} - x_i} \quad (2.21)$$

Despejando  $\bar{x}$  de la ecuación 2.19

$$\bar{x} \approx \frac{x_{i+1} - g'(\bar{x}) x_i}{1 - g'(\bar{x})}$$

\* Véase Problema 2.22.

sustituyendo la ecuación 2.21 en la última ecuación, se llega a:

$$\bar{x} \approx x_i - \frac{(x_{i+1} - x_i)^2}{x_{i+2} - 2x_{i+1} + x_i}$$

que da aproximaciones a  $\bar{x}$  a partir de los valores ya obtenidos en alguna sucesión. Llámese a esta nueva sucesión  $x'_0, x'_1, x'_2, \dots$

$$x'_i = x_i - \frac{(x_{i+1} - x_i)^2}{x_{i+2} - 2x_{i+1} + x_i} \quad i \geq 0 \quad (2.22)$$

Por ejemplo,  $x'_0$  requiere de  $x_0, x_1, x_2$ , ya que

$$x'_0 = x_0 - \frac{(x_1 - x_0)^2}{x_2 - 2x_1 + x_0}$$

y  $x'_1$  de  $x_1, x_2, x_3$ , pues

$$x'_1 = x_1 - \frac{(x_2 - x_1)^2}{x_3 - 2x_2 + x_1}$$

y así sucesivamente.

Este proceso conducirá, en la mayoría de los casos, a la solución buscada  $\bar{x}$  más rápido que si se siguiera el inciso a); asimismo, evita la búsqueda de una nueva  $g(x)$  y el riesgo de no obtener convergencia con esa nueva  $g(x)$ . A este proceso se le conoce como aceleración de convergencia y se presenta como algoritmo de Aitken.

### ALGORITMO DE AITKEN

Dada una sucesión de número  $x_0, x_1, x_2, \dots$  a partir de ella se genera una nueva sucesión  $x'_0, x'_1, x'_2, \dots$  con la ecuación 2.22.

Si se emplea la notación

$$\Delta x_i = x_{i+1} - x_i, \quad i = 0, 1, 2, \dots$$

donde  $\Delta$  es un operador\* de diferencias cuyas potencias (o más propiamente su orden) se pueden obtener así

$$\Delta(\Delta x_i) = \Delta^2 x_i = \Delta(x_{i+1} - x_i) = \Delta x_{i+1} - \Delta x_i$$

o

$$\Delta^2 x_i = x_{i+2} - 2x_{i+1} + x_i$$

la ecuación 2.22 adquiere la forma simplificada

$$x'_i = x_i - \frac{(\Delta x_i)^2}{\Delta^2 x_i} \quad (2.23)$$

\* Véase capítulo 5.

**Ejemplo 2.8**

Acelerar la convergencia de la sucesión del ejemplo 2.2, mediante el algoritmo de Aitken.

**Solución** Con la ecuación 2.22 o 2.23 con  $x_0 = 1$ ,  $x_1 = 1.53846$ ,  $x_2 = 1.29502$  se tiene:

$$x'_0 = 1 - \frac{(1.53846 - 1)^2}{1.29502 - 2(1.53846) + 1} = 1.37081$$

Ahora, con la ecuación 2.22 y con  $x_1 = 1.53846$ ,  $x_2 = 1.29502$  y  $x_3 = 1.40183$ , resulta:

$$x'_1 = 1.53846 - \frac{(1.29502 - 1.53846)^2}{1.40183 - 2(1.29502) + 1.53846} = 1.36566$$

En una tercera iteración se obtiene:

$$x'_2 = 1.36889$$

Obsérvese que  $x'_1$  está prácticamente tan cerca de la raíz real de la ecuación como el valor de  $x_6$  del ejemplo 2.2, y  $x'_2$  mejora tanto la aproximación que es preciso comparar este valor con el de  $x_3$  del ejemplo 2.4. La comparación puede establecerse mediante  $|f(x'_i)|$  y  $|f(x_i)|$ .

Se ha encontrado que el método de Aitken es de segundo orden\* y se emplea normalmente para acelerar la convergencia de cualquier sucesión de valores que converge linealmente, cualquiera que sea su origen. La aplicación del método de Aitken a la iteración de punto fijo da el procedimiento conocido como método de Steffensen, que se ilustra a continuación.

**Ejemplo 2.9**

Encuentre una raíz real de la ecuación

$$f(x) = x^3 + 2x^2 + 10x - 20 = 0$$

con el método de Steffensen, usando  $\varepsilon = 10^{-3}$  aplicado a  $|f(x'_i)|$ .

**Solución** Se pasa primero la ecuación  $f(x) = 0$  a la forma  $g(x) = x$ . Al igual que en el ejemplo 2.2, se factoriza  $x$  en la ecuación y luego se “despeja”.

$$x = \frac{20}{x^2 + 2x + 10}$$

**Primera iteración**

Se elige un valor inicial  $x_0 = 1$  y se calcula  $x_1$  y  $x_2$

$$x_1 = 1.53846$$

$$x_2 = 1.29502$$

Se aplica ahora la ecuación 2.22 para acelerar la convergencia:

$$x'_0 = 1 - \frac{(1.53846 - 1)^2}{1.29502 - 2(1.53846) + 1} = 1.37081$$

Como  $|f(x'_0)| = (1.37081)^3 + 2(1.37081)^2 + 10(1.37081) - 20 = 0.04234 > 10^{-3}$ , se pasa a la

\* Henrici, P., *Elements of Numerical Analysis*. John Wiley & Sons, Inc. (1964). pp. 91-92.



### Segunda iteración

Con el valor de  $x'_0$  que ahora se denota como  $x_3$  y con la  $g(x)$  que se tiene, resulta:

$$x_4 = 1.36792$$

$$x_5 = 1.36920$$

Aplicando nuevamente la ecuación 2.22 a  $x_3$ ,  $x_4$  y  $x_5$  se llega a:

$$\begin{aligned} x'_1 = x_6 &= 1.37081 - \frac{(1.36792 - 1.37081)^2}{1.36920 - 2(1.36792) + 1.37081} \\ &= 1.36881 \end{aligned}$$

Luego, con el criterio de exactitud se tiene:

$$|f(x_6)| = 0.0000399 < 10^{-3}$$

y el problema queda resuelto.

Para llevar a cabo los cálculos, puede usarse Matlab o la TI-92 Plus, con los siguientes programas basados en el algoritmo 2.5



```
format long
x0=1;eps=0.001;
for i=1:10
    x1=20/(x0^2+2*x0+10);
    x2=20/(x1^2+2*x1+10);
    x=x0-(x1-x0)^2/(x2-2*x1+x0);
    dist=abs(x-x0);
    disp([x1, x2, x])
    if dist < eps
        break
    end
    x0=x;
end
```

```
e2_9( )
Prgm
Define g(x)=20/(x^2+2*x+10)
ClrIO : 1.→x0 : 1.e-4→eps
Loop
    g(x0) →x1: g(x1) →x2
    x0-(x1-x0)^2/(x2-2*x1+x0) →x
    Disp format (x, "f5")
    If abs (x-x0)<eps
        Exit
    x→x0
EndLoop
EndPrgm
```

A continuación se da el algoritmo de Steffensen.

#### ALGORITMO 2.5 Método de Steffensen

Para encontrar una raíz real de la ecuación  $g(x) = x$ , proporcionar la función  $G(X)$  y los

DATOS: Valor inicial  $X_0$ , criterio de convergencia EPS y número máximo de iteraciones MAXIT.  
 RESULTADOS: La raíz aproximada  $X$  o un mensaje de falla.

PASO 1. Hacer  $I = 1$ .

PASO 2. Mientras  $I < \text{MAXIT}$ , repetir los pasos 3 a 6.

PASO 3. Hacer:

$$X_1 = G(X_0)$$

$$X_2 = G(X_1).$$

$$X = X_0 - (X_1 - X_0)^2 / (X_2 - 2X_1 + X_0).$$

PASO 4. SI  $ABS(X - X_0) < EPS$ , IMPRIMIR  $X$  y TERMINAR. De otro modo CONTINUAR.

PASO 5. Hacer  $I = I + 1$

PASO 6. Hacer  $X_0 = X$  (actualiza  $X_0$ )

PASO 7. IMPRIMIR mensaje de falla: "EL MÉTODO NO CONVERGE A UNA RAÍZ" y TERMINAR.

## MÉTODOS DE DOS PUNTOS

Los métodos de dos puntos bisección y posición falsa garantizan convergencias, pero ya que puede ser muy lenta en algunos casos, conviene acelerarla. Enseguida se estudia una modificación de posición falsa que cumple con este cometido.

### MÉTODO ILLINOIS\*

Esta técnica difiere del método de posición falsa (véase algoritmo 2.4) en que los valores  $(X_p, F_p)$ ,  $(X_D, F_D)$  de las sucesivas iteraciones se determinan de acuerdo con las siguientes reglas:

a) Si  $F_D * F_M > 0$ , hacer  $X_D = X_p$ ,  $F_D = F_p$

b) Si  $F_D * F_M < 0$ , hacer  $F_D = F_D/2$

y en ambos casos se sustituye a  $X_I$  con  $X_M$  y  $F_I$  con  $F_M$ .

El empleo de  $F_D/2$  en lugar de  $F_D$  evita que uno de los extremos  $X_I$  o  $X_D$  se mantenga fijo (caso frecuente en posición falsa). Esta modificación acelera considerablemente la convergencia del método. Los valores funcionales  $F_p$ ,  $F_D$  empleados conservan sus signos opuestos. El algoritmo correspondiente puede obtenerse sustituyendo los pasos 6 y 7 en el algoritmo 2.4 con los incisos (a) y (b), respectivamente, y además un paso donde se sustituye a  $X_I$  con  $X_M$  y  $F_I$  con  $F_M$ .

## 2.8 Búsqueda de valores iniciales

El uso de cualquier algoritmo numérico para encontrar las raíces de  $f(x) = 0$ , requiere uno o más valores iniciales; además, en métodos como el de la bisección y el de posición falsa, los dos valores iniciales requeridos deben estar a los lados de la raíz buscada y sus valores funcionales correspondientes tienen que ser de signos opuestos.

A continuación se dan algunos lineamientos generales para obtener valores aproximados a las raíces de  $f(x) = 0$ .

1. Por lo general, la ecuación cuyas raíces se buscan tiene algún significado físico; entonces, a partir de consideraciones físicas pueden estimarse valores aproximados a las raíces. Este razonamiento es particular para cada ecuación. A continuación se presenta un ejemplo para ilustrar esta idea.

\* Dowel M. and Jarrat P., *A Modified Regula Falsi Method for Computing the Root of an Equation*. BIT. Vol. 11 p. 168 (1971).

**Ejemplo 2.10**

Determine el valor inicial en la solución de una ecuación de estado.

**Solución**

El cálculo del volumen molar de un gas dado, a cierta presión y temperatura también dadas, es un problema común en termodinámica. Para realizar dicho cálculo se emplea alguna de las ecuaciones de estado conocidas. Una de ellas es la ecuación de Beattie-Bridgeman:

$$P = \frac{RT}{V} + \frac{\beta}{V^2} + \frac{\gamma}{V^3} + \frac{\delta}{V^4} \quad (2.24)$$

donde los parámetros  $\beta$ ,  $\gamma$  y  $\delta$  quedan determinados al fijar el gas de que se trata, su temperatura  $T$  y su presión  $P$ .

En las condiciones expuestas, el problema se reduce a encontrar el o los valores de  $V$  que satisfagan la ecuación 2.24, o en otros términos, a determinar las raíces del polinomio en  $V$ .

$$f(V) = P V^4 - R T V^3 - \beta V^2 - \gamma V - \delta = 0, \quad (2.25)$$

que resulta de multiplicar por  $V_4$  la ecuación 2.24 y pasar todos sus términos a un solo miembro.

La solución de la ecuación 2.25 tiene como primer problema encontrar cuando menos un valor inicial  $V_0$  cercano al volumen buscado  $V$ . Este valor  $V_0$ , se obtiene a partir de la ley de los gases ideales; así:

$$V_0 = \frac{RT}{P},$$

que generalmente es una primera aproximación razonable.

Como puede verse, el razonamiento es sencillo y se basa en el sentido común y las leyes básicas del fenómeno involucrado.

2. Otra manera de conseguir información sobre la función, que permita determinar valores iniciales “adecuados”, consiste en obtener su gráfica aproximada mediante un análisis de  $f(x)$ , a la manera clásica del cálculo diferencial e integral, o bien como se ha venido sugiriendo, con algún software comercial y, en el mejor de los casos, empleando ambos. A continuación se presentan los pasos del análisis de la función  $f(x)$  y de la construcción de su gráfica en la forma clásica.
  - a) Determinar el dominio de definición de la función.
  - b) Determinar un subintervalo de a), que puede ser a) mismo. Es un intervalo donde se presupone que es de interés analizar la función. Evalúese la función en los siguientes puntos de ese subintervalo: puntos extremos y aquéllos donde sea fácil el cálculo de  $f(x)$ . En los siguientes pasos todo estará referido a este subintervalo.
  - c) Encontrar los puntos singulares de la función (puntos en los cuales es infinita o no está definida).
  - d) La primera y la segunda derivadas dan información muy útil sobre la forma de la función, aun más útil que información de valores computados; por ejemplo, dan los intervalos de crecimiento y decrecimiento de la función. Por esto, obténgase la primera derivada y evalúese en puntos apropiados, en particular en puntos cercanos a aquéllos donde la función ya está evaluada y en los que es fácil esta evaluación.
  - e) Encontrar los puntos máximo y mínimo, así como los valores de la función en esos puntos.



- f) Los dominios de concavidad y convexidad de la curva, y los puntos de inflexión es información cualitativa y cuantitativa, que se obtienen a partir de la segunda derivada y son imprescindibles para este análisis.
- g) Obtener las asíntotas de la función. Éstas, en caso de existir, indican cierta regularidad en los comportamientos de la gráfica de  $y = f(x)$  al tender  $x$  o  $y$  hacia infinito.
- h) Descomponer la función en sus partes más sencillas que se sumen o se multipliquen. Graficar cada parte y construir la gráfica de la función original, combinando las gráficas de las partes y la información conseguida en los pasos anteriores.

**Ejemplo 2.11**

**Análisis de una función.** A continuación se presenta el análisis clásico de la función

$$f(x) = x - e^{1-x} (1 + \ln x)$$

hecho por Pizer.\*

Nótese que  $\ln x$  está definida sólo para  $x > 0$ , así que  $f(x)$  está definida sólo en  $(0, \infty)$ .

En este ejemplo ilustrativo, se analiza la función en todo el dominio de definición; es decir, el intervalo de interés será  $(0, \infty)$ .

Un punto donde es fácil evaluar la función es en  $x = 1$ , ya que la parte exponencial y la parte logarítmica se determinan fácilmente en ese punto.

$$f(1) = 1 - e^{1-1} (1 + \ln 1) = 0$$

De esta forma, se ha encontrado una raíz de la ecuación  $\bar{x}_1 = 1$ .

En  $x = 10$

$$f(10) = 10 - e^{-9} (1 + \ln 10) \approx 10$$

En  $x = 100$

$$f(100) = 100 - e^{-99} (1 + \ln 100) \approx 100$$

Con esta información puede adelantarse que la función tiene la asíntota  $y = x$ , la función identidad.

Un punto donde la función no está definida es en el extremo  $x = 0$ . Al analizarlo se advierte que cuando  $x \rightarrow 0$ , el  $\ln x \rightarrow -\infty$  y  $f(x) \rightarrow \infty$  y se encuentra una asíntota más de la función, que es la parte positiva del eje  $y$ . Por un lado,  $x \rightarrow \infty$ ,  $\ln x \rightarrow \infty$ , pero  $e^{1-x}$  se acerca más rápidamente a cero y, por tanto, el producto  $e^{1-x} (1 + \ln x)$  tiende a cero, dejando como resultado global que  $f(x) \rightarrow \infty$ . Se concluye que  $f(x) \rightarrow \infty$  cuando  $x \rightarrow 0$ , o cuando  $x \rightarrow \infty$ . Como  $f(x)$  no tiene otros puntos singulares, se da por terminado el inciso c).

Al calcular la primera y segunda derivadas de  $f(x)$ , se tiene que

$$f'(x) = 1 - e^{1-x} (1/x - 1 - \ln x)$$

y

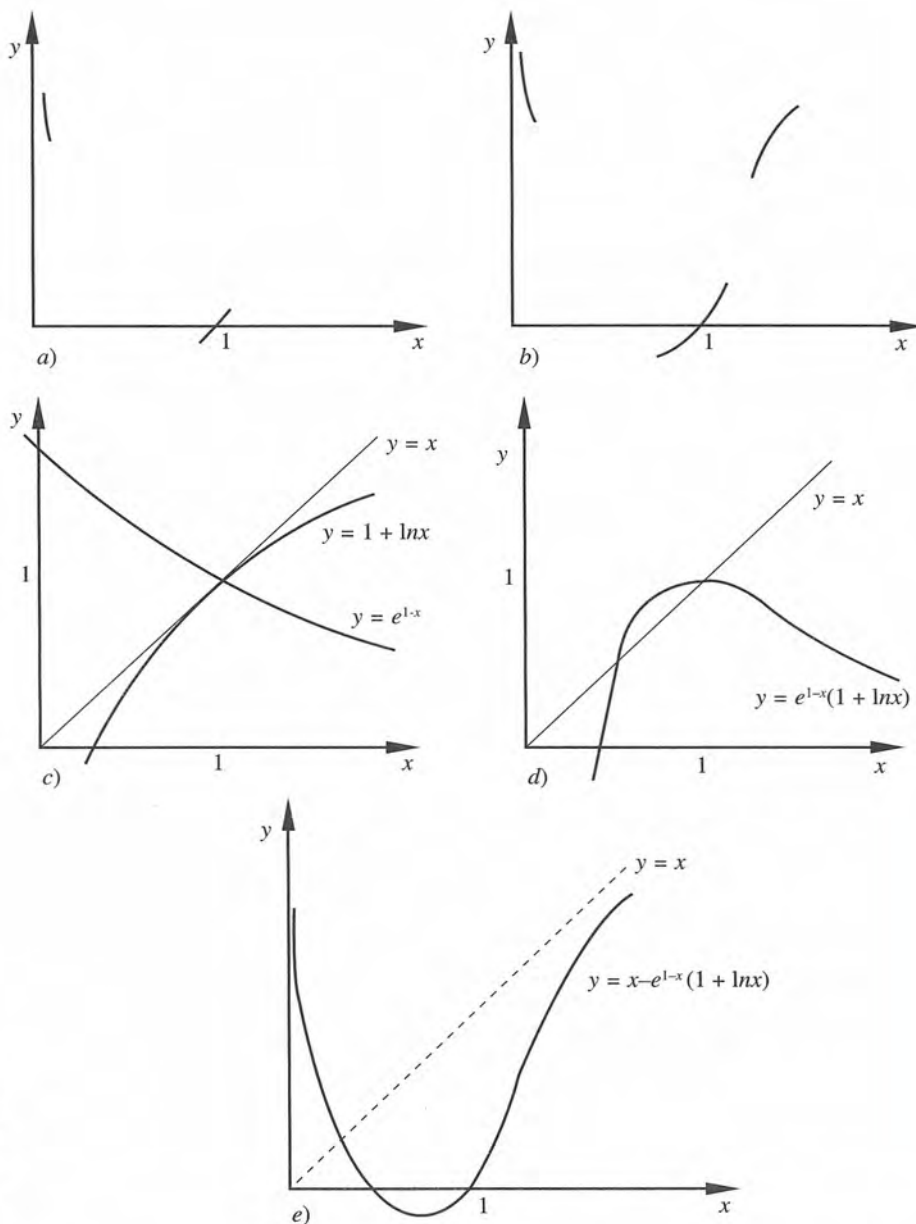
$$f''(x) = e^{1-x} (2/x + 1/x^2 - 1 - \ln x)$$

Al evaluar  $f'(x)$  en  $x = 1$ , se obtiene  $f'(1) = 1$ .

Cuando  $x \rightarrow \infty$ ,  $f'(x) \rightarrow 1$ .

Lo que se sabe hasta aquí de la función, se muestra en la figura 2.9 a. Como  $f(x)$  es continua (todas las funciones sencillas que la forman lo son) en  $(0, \infty)$ , deberá haber por lo menos otra raíz de  $f(x)$  en  $(0, 1)$ .

\* Stephen, M. Pizer. *Numerical Computing and Mathematical Analysis*. S.R.A., (1975) pp. 176-179.



**Figura 2.9**  
Construcción  
de la gráfica de  
 $f(x) = x - e^{1-x}(1 + \ln x)$ .

El inciso e) del análisis de la función no procede en este caso, ya que sería tan complejo como encontrar las raíces de  $f(x)$ . En su lugar se analiza la forma de la curva con la segunda derivada. Evaluando  $f''(x)$  en valores muy grandes de  $x$ , se tiene que  $f''(x) < 0$ , o sea que la función es convexa para valores muy grandes de  $x$  (también se dice que la curva gira su convexidad hacia la parte positiva del eje  $y$ ). Además, se tiene  $f''(1) = 2$ , lo que indica que la función es cóncava en  $x = 1$  (o en otras palabras gira su convexidad hacia la parte negativa del eje  $y$ ). La información se muestra en la figura 2.9 b).

Se puede obtener aún más información de  $f(x)$ , analizando las funciones elementales que la componen, como  $x$ ,  $e^{1-x}$ , y  $1 + \ln x$ . La familiaridad con las gráficas de las funciones elementales es útil cuando se consideran funciones más complejas. Las partes en que se puede descomponer  $f(x)$  se muestran en la figura 2.9 c). Primero nótese que la gráfica de  $1 + \ln x$  es la de  $\ln x$  aumentada en una unidad, y que la gráfica de  $e^{1-x}$  es la de  $e^{-x}$  llevada

una unidad a la derecha. Multiplicando  $e^{1-x}$  y  $1 + \ln x$  entre sí (Fig. 2.9 d), se ve que este producto es negativo entre cero y algún valor menor que 1, tiende a cero cuando  $x$  aumenta y permanece debajo de  $y = x$  para  $x > 1$ .

Como la derivada del producto es cero en  $x = 1$ , la curva del producto tiene ahí un máximo y el resto de la gráfica puede obtenerse como se ilustra en la figura 2.9 e).

Nótese que los ceros de  $f(x)$  son los puntos donde el producto  $e^{1-x}(1 + \ln x)$  y la función identidad  $y = x$  se intersectan. Esto significa que sólo hay dos raíces de la función. También puede concluirse que hay una raíz en  $x = 1$  y otra cerca de  $x = 0.5$ , por lo que 0.5 sería un valor inicial adecuado para calcular esta segunda raíz.

Actualmente se puede recurrir a programas comerciales con facilidades de graficación para visualizar funciones matemáticas; no obstante, es necesario verlos como auxiliares en esta tarea y no como algo que permita sustituir el análisis tradicional y mucho menos los conceptos. Por ejemplo, si graficamos con Matlab la función en el intervalo  $[-3, 3]$  obtendríamos la figura 2.10.

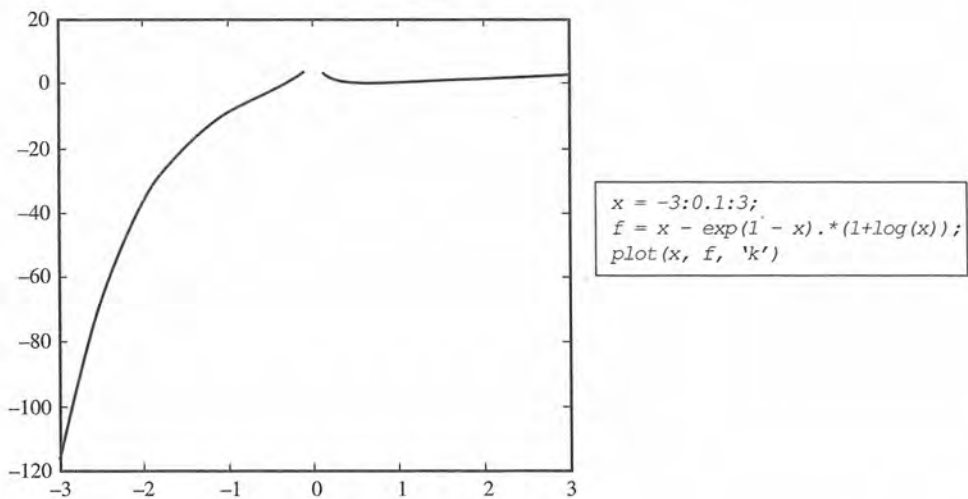


Figura 2.10

Lo cual ciertamente es algo que requiere una lectura en términos de lo que se “ve”, de lo que el programa hace, de los mensajes de error que pudieran darse y de los conceptos involucrados. En este caso, Matlab muestra las siguientes advertencias:

*Warning: Log of zero.*

> In D: \Archivos de programa\Matlab\bin\e2\_10.m at line 2

*Warning: Imaginary parts of complex X and/or Y arguments ignored.*

> In D: \Archivos de programa\Matlab\bin\e2\_10.m at line 3

La primera significa que se está evaluando la función en  $x = 0$ , y como ya es sabido,  $\ln(0)$  no está definido y de ahí la interrupción de la gráfica alrededor de ese punto; de igual manera, el logaritmo de números negativos genera valores complejos, sin embargo, Matlab ignora la parte imaginaria de los valores de la función y con la parte real continúa la graficación. De no haber hecho estas consideraciones pensaríamos que existe gráfica a la izquierda y a la derecha de cero, y además que hay una raíz negativa.

Si se elige el intervalo  $(0, 5)$  se obtiene la gráfica de la figura 2.11.



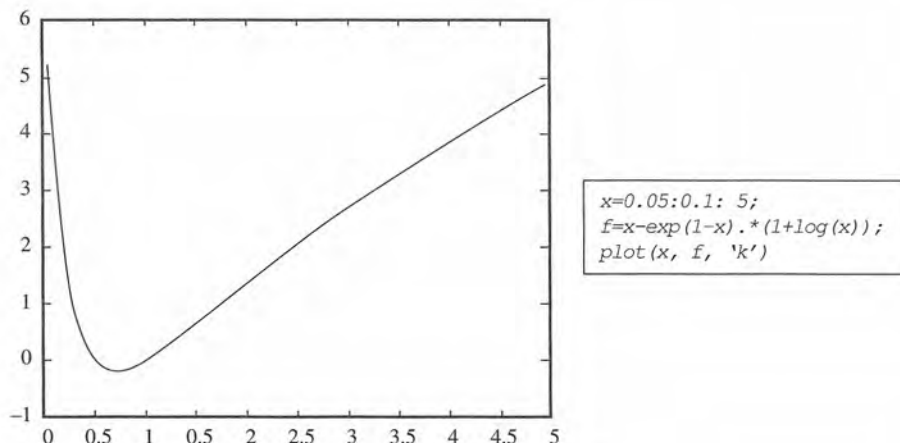


Figura 2.11

La cual, aunque ya es muy parecida a la mostrada en la figura 2.9 e, no revela por sí misma, por ejemplo, que es cóncava hacia arriba en el intervalo comprendido entre  $x = 0$  y  $x \approx 1.6$ , y que es cóncava hacia abajo después de  $x = 1.6$ . Lo anterior se puede comprobar obteniendo la segunda derivada y observando el signo de dicha derivada. En este caso

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

donde puede observarse que en  $0 < x \leq 1$  la segunda derivada es positiva; y para algún valor de  $x$  alrededor de 1.6, la segunda derivada es negativa y se mantiene con ese signo al aumentar  $x$ . El valor de  $x$ , donde la segunda derivada es cero, es el punto donde la curva cambia de concavidad, y encontrar este valor implica resolver una ecuación no lineal en una incógnita:  $f''(x) = 0$ . Resolviendo esta ecuación utilizando como valor inicial 1.6 observado en la gráfica y con la instrucción

```
fzero('exp(1-x)*(2/x+1/x^2-1-log(x))',1.6)
```

Matlab reporta  $\text{ans} = 1.6952$

Podemos apreciar también en la gráfica de la figura 2.11 dos raíces; sin embargo, ¿cómo podríamos saber que son las únicas? Una forma sería extender el intervalo de graficación en el sentido positivo del eje  $x$  y/o ver si la función es creciente o si tiene asíntotas. Una forma de auxiliarnos con un graficador sería graficar para valores de  $x$  muy grandes, por ejemplo  $[0, 200]$ , con lo que se obtiene la gráfica de la figura 2.12.

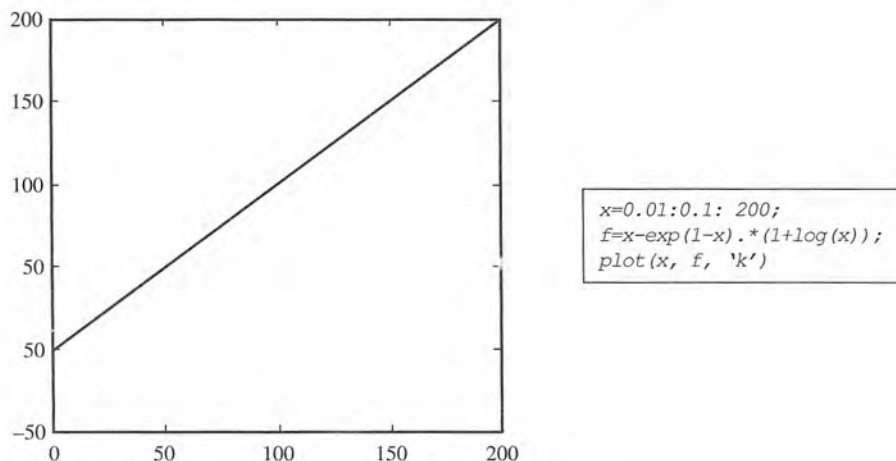


Figura 2.12

Nuevamente una lectura de esta gráfica revela que la función es creciente y que se acerca a la función  $y = x$ ; no obstante, es necesario precisar esto como lo hicimos en el análisis clásico.

Una vez que hayamos determinado las dos asíntotas de las funciones, podemos asegurar que sólo hay dos raíces reales en el intervalo  $(0, 1.5)$ , cuya obtención puede hacerse con alguno de los métodos vistos. Con la instrucción Matlab.

```
fzero('x - exp(1-x).*(1+log(x));0.3)
```

Se obtiene

```
ans = 0.4967
```

## 2.9 Raíces complejas

Hasta ahora se han discutido sólo técnicas para encontrar raíces reales de ecuaciones de la forma  $f(x) = 0$ . Sin embargo, a menudo se presentan ecuaciones polinomiales con coeficientes reales, cuyas raíces son complejas, o bien polinomios complejos, y ecuaciones trascendentes con raíces reales y complejas.

Generalmente, dichas ecuaciones pueden resolverse por el método de Newton-Raphson (Sec. 2.2), pero proponiendo un valor inicial  $x_0$  complejo o bien por algún otro método.

### MÉTODO DE NEWTON-RAPHSON

Supóngase que se tiene

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \quad (2.26)$$

con todos los coeficiente  $a_i$  reales.  $f'(x)$  es un polinomio de grado  $(n-1)$  y de coeficientes también reales

$$f'(x) = n a_n x^{n-1} + (n-1) a_{n-1} x^{n-2} + \dots + 2a_2 x + a_1 \quad (2.27)$$

Si el valor inicial  $x_0$  es real, entonces

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

también será real y todos los valores  $x_i$  siguientes. Consecuentemente no se puede encontrar una raíz compleja de la ecuación 2.26 si se inicia con un valor  $x_0$  real.

Si por el contrario, el valor inicial  $x_0$  es complejo,  $x_1$  entonces será complejo,  $x_2$  también, y así sucesivamente. De esta manera, si el proceso converge, puede encontrarse una raíz  $\bar{x}$  compleja.

#### Ejemplo 2.12

Encuentre las raíces complejas de la ecuación

$$f(x) = x^2 + 4 = 0,$$

con el método de Newton-Raphson.

**Solución** Al derivar  $f(x)$  se tiene

$$f'(x) = 2x$$

Sea  $x_0 = j$  el valor inicial propuesto. Aplicando la ecuación 2.12 con este valor inicial, se tiene:

$$x_1 = j - \frac{(j^2 + 4)}{2(j)}$$

pero  $(j)^2 = -1$ , entonces:  $x_1 = j - \frac{-1 + 4}{2j} = j - \frac{3}{2j}$

Multiplicando y dividiendo por  $j$  el término  $3/(2j)$ , se obtiene:

$$x_1 = j - (-1.5j) = 2.5j$$

$$x_2 = 2.5j - \frac{(2.5j)^2 + 4}{2(2.5j)} = 2.05j$$

$$x_3 = 2.5j - \frac{(2.5j)^2 + 4}{2(2.05j)} = 2.001j$$

La sucesión de valores complejos  $x_0, x_1, \dots$ , va acercándose rápidamente a la raíz  $\bar{x}_1 = 2j$

$$f(\bar{x}_1) = f(2j) = (2j)^2 + 4 = -4 + 4 = 0$$

Para evaluar la distancia entre dos valores complejos consecutivos, se utiliza

$$|x_{i+1} - x_i|,$$

donde las barras representan el módulo del número complejo  $x_{i+1} - x_i$ . Esto es, si

$$x_{i+1} - x_i = a + bj$$

Entonces

$$|x_{i+1} - x_i| = \sqrt{a^2 + b^2}$$

Por lo que se tiene para la sucesión previa

$$|x_1 - x_0| = |2.5j - j| = \sqrt{0^2 + (1.5)^2} = 1.5$$

$$|x_2 - x_1| = |2.05j - 2.5j| = \sqrt{0^2 + (-0.45)^2} = 0.45$$

$$|x_3 - x_2| = |2.001j - 2.05j| = \sqrt{0^2 + (-0.049)^2} = 0.049$$

y la convergencia es notoria.

En caso de tener raíces complejas una ecuación polinomial con coeficientes reales, éstas aparecen en parejas; es decir, si  $x = a + bj$  es raíz, también lo será  $x = a - bj$  (toda vez que al multiplicarlos deben producir los coeficientes reales).

Por esto:

$$\bar{x}_2 = -2j$$

es la segunda raíz que se busca.

$$f(\bar{x}_2) = f(-2j) = (-2j)^2 + 4 = -4 + 4 = 0$$

El problema queda terminado.

Para realizar los cálculos de este ejemplo, puede usar el guión de Matlab dado en el Ejemplo 2.4, con el valor inicial  $x_0=1i$ , con lo que Matlab realiza los cálculos con aritmética compleja. El lector puede apreciar aún más la utilidad de Matlab con este ejemplo.



Si bien se resolvió una ecuación cuadrática que no representa dificultad, el método también puede emplearse para un polinomio de mayor grado, siguiendo los mismos pasos. El lector puede crear un programa para el algoritmo en algún lenguaje de alto nivel o en un pizarrón electrónico como Mathcad.

## MÉTODO DE MÜLLER

Un método deducido por Müller,\* se ha puesto en práctica en las computadoras con éxito sorprendente. Se puede usar para encontrar cualquier tipo de raíz, real o compleja, de una función arbitraria. Converge casi cuadráticamente en un intervalo cercano a la raíz y, a diferencia del método de Newton-Raphson, no requiere la evaluación de la primera derivada de la función, y obtiene raíces reales y complejas aun cuando estas raíces sean repetidas.

La aplicación del método requiere valores iniciales y es una extensión del método de la secante, el cual aproxima la gráfica de la función  $f(x)$  por una línea recta que pasa por los puntos  $(x_{i-1}, f(x_{i-1}))$  y  $(x_i, f(x_i))$ . El punto de intersección de esta línea con el eje  $x$  da la nueva aproximación  $x_{i+1}$ .

En lugar de aproximar  $f(x)$  por una función lineal (línea recta o polinomio de grado 1), resulta natural tratar de obtener una convergencia más rápida aproximando  $f(x)$  por un polinomio  $p(x)$  de grado  $n > 1$  que coincida con  $f(x)$  en los puntos de abscisas  $x_i, x_{i-1}, \dots, x_{i-n}$ , y determinar  $x_{i+1}$  como una de las raíces de  $p(x)$ .

A continuación se describe el caso  $n = 2$ , donde el estudio detallado de Müller encontró que la elección de  $n$  da resultados satisfactorios.

Se toman tres valores iniciales  $x_0, x_1, x_2$  y se halla el polinomio  $p(x)$  de segundo grado que pasa por los puntos  $(x_0, f(x_0))$ ,  $(x_1, f(x_1))$  y  $(x_2, f(x_2))$ , y se toma una de las raíces de  $p(x)$ , la más cercana a  $x_2$  como la siguiente aproximación  $x_3$ . Se repite la operación con los nuevos valores iniciales  $x_1, x_2, x_3$ , y se termina el proceso tan pronto como se satisfaga algún criterio de convergencia. La figura 2.13 ilustra este método.

Sean  $x_i, x_{i-1}, x_{i-2}$ , tres aproximaciones distintas a una raíz de  $f(x) = 0$ . Usando la siguiente notación:

$$f_i = f(x_i)$$

$$f_{i-1} = f(x_{i-1})$$

$$f_{i-2} = f(x_{i-2})$$

en el capítulo 5 se demostrará que con

$$f[x_i, x_{i-1}] = \frac{f_i - f_{i-1}}{x_i - x_{i-1}} \quad (2.28)$$

$$f[x_{i-1}, x_{i-2}] = \frac{f_{i-1} - f_{i-2}}{x_{i-1} - x_{i-2}}$$

$$f[x_i, x_{i-1}, x_{i-2}] = \frac{f[x_i, x_{i-1}] - f[x_{i-1}, x_{i-2}]}{x_i - x_{i-2}} \quad (2.29)$$

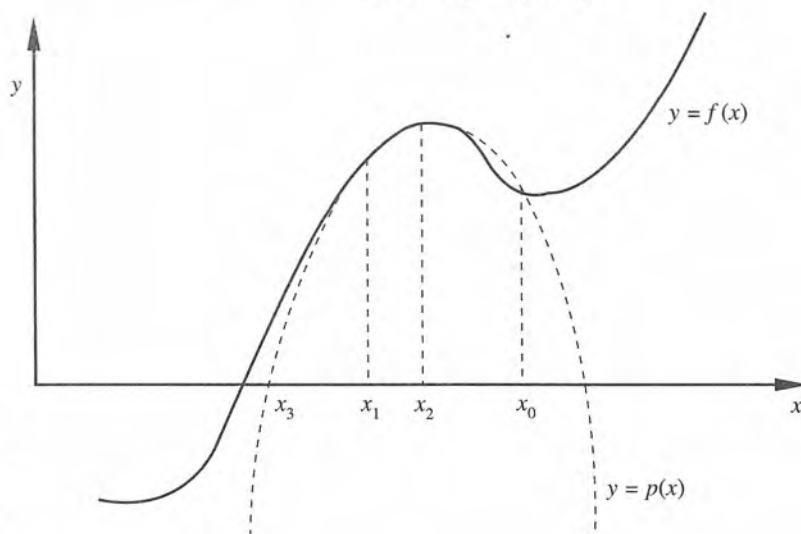
la función

$$p(x) = f_i + f[x_i, x_{i-1}](x - x_i) + f[x_i, x_{i-1}, x_{i-2}](x - x_i)(x - x_{i-1}) \quad (2.30)$$

\* Müller, D.E. "A Method of Solving Algebraic Equations Using an Automatic Computer". *Mathematical Tables and Other Aids to Computation* (MTAC), 10. pp. 208-215 (1956).

es la parábola única que pasa por los puntos  $(x_i, f_i)$ ,  $(x_{i-1}, f_{i-1})$  y  $(x_{i-2}, f_{i-2})$ . El lector recordará que la manera usual de escribir un polinomio de segundo grado o parábola es:

$$p(x) = a_0 + a_1x + a_2x^2$$



**Figura 2.13**  
Interpretación  
gráfica del  
método de  
Müller.

Al comparar esta última expresión con la ecuación 2.30 se establece la siguiente identificación:

$$a_2 = f[x_i, x_{i-1}, x_{i-2}]$$

$$a_1 = f[x_i, x_{i-1}] - (x_i + x_{i-1})a_2$$

$$a_0 = f_i - x_i(f[x_i, x_{i-1}] - x_{i-1}a_2)$$

Una vez calculados los valores de  $a_0$ ,  $a_1$  y  $a_2$ , las raíces de  $p(x)$  se determinan a partir de la fórmula cuadrática

$$x_{i+1} = \frac{2a_0}{-a_1 \pm (a_1^2 - 4a_0a_2)^{1/2}} \quad (2.31)$$

cuya explicación se encuentra en el problema 2.31, y en el ejercicio 1.3 del capítulo 1.

Se selecciona el signo que precede al radical de manera que el denominador sea máximo en magnitud,\* y la raíz correspondiente es la siguiente aproximación  $x_{i+1}$ . La razón para escribir la fórmula cuadrática de esta manera es obtener mayor exactitud (véase Prob. 2.31), ya disminuida por las diferencias de las ecuaciones 2.28 y 2.29, que se utilizan en el cálculo de  $a_0$ ,  $a_1$  y  $a_2$ , y que son aproximaciones a las derivadas de la función  $f(x)$ .

Puede ocurrir que la raíz cuadrada en la ecuación 2.31 sea compleja. Si  $f(x)$  no está definida para valores complejos, el algoritmo deberá reiniciarse con nuevos valores iniciales. Si  $f(x)$  es un polinomio, la posibilidad de raíces complejas es latente y el valor de  $x$  puede considerarse como aproximación a alguna de estas raíces y, por tanto, deberá emplearse en la siguiente iteración.

\* Con esto se encuentra el valor más cercano a  $x_i$ .

**Ejemplo 2.13**

Encuentre una raíz real de la ecuación polinomial

$$f(x) = x^3 + 2x^2 + 10x - 20 = 0,$$

con el método de Müller.

**Solución Primera iteración**

Al seleccionar como valores iniciales a

$$x_0 = 0; \quad x_1 = 1; \quad x_2 = 2$$

y evaluar la función  $f(x)$  en estos puntos, se tiene

$$f_0 = -20; \quad f_1 = -7; \quad f_2 = 16$$

Se calculan ahora los coeficientes del polinomio de segundo grado

$$f[x_1, x_0] = \frac{f_1 - f_0}{x_1 - x_0} = \frac{-7 + 20}{1 - 0} = 13$$

$$f[x_2, x_1] = \frac{f_2 - f_1}{x_2 - x_1} = \frac{16 + 7}{2 - 1} = 23$$

$$f[x_2, x_1, x_0] = \frac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0} = \frac{23 - 13}{2 - 0} = 5$$

Por tanto

$$a_2 = f[x_2, x_1, x_0] = 5$$

$$a_1 = f[x_2, x_1] - (x_2 + x_1)a_2 = 23 - (2 + 1)5 = 8$$

$$a_0 = f_2 - x_2(f[x_2, x_1] - x_1 a_2) = 16 - 2(23 - 1(5)) = -20$$

Se calculan los denominadores de la ecuación 2.31

$$-a_1 + (a_1^2 - 4a_0a_2)^{1/2} = -8 + (64 + 400)^{1/2} = 13.54066$$

$$-a_1 - (a_1^2 - 4a_0a_2)^{1/2} = -8 - (64 + 400)^{1/2} = -29.54066$$

Como el segundo es mayor en valor absoluto, se usa en la ecuación 2.31, de donde

$$x_3 = \frac{2a_0}{-a_1 - (a_1^2 - 4a_0a_2)^{1/2}} = \frac{2(-20)}{-29.54066} = 1.35407$$

**Segunda iteración**Recorriendo ahora los subíndices de  $x$ , se tiene:

$$x_0 = 1; \quad x_1 = 2; \quad x_2 = 1.35407$$

$$f_0 = -7; \quad f_1 = 16; \quad f_2 = -0.30959$$

En consecuencia:

$$f[x_1, x_0] = \frac{16 + 7}{2 - 1} = 23$$

$$f[x_2, x_1] = \frac{-0.30959 - 16}{1.35407 - 2} = 25.24978$$



$$f[x_2, x_1, x_0] = \frac{25.24978 - 23}{1.35407 - 1} = 6.35405$$

De donde:

$$a_2 = f[x_2, x_1, x_0] = 6.35405$$

$$a_1 = f[x_2, x_1] - (x_2 + x_1)a_2 =$$

$$25.24978 - (1.35407 + 2)6.35405 = 3.87077$$

$$a_0 = f_2 - x_2 (f[x_2, x_1] - x_1 a_2) =$$

$$-0.30959 - 1.35407 (25.24978 - 2 (6.35405)) = -17.29190$$

Calculando los denominadores de la ecuación 2.31

$$-a_1 + (a_1^2 - 4a_0a_2)^{1/2} = 17.39295$$

$$-a_1 - (a_1^2 - 4a_0a_2)^{1/2} = -25.26855$$

Como el segundo es mayor en valor absoluto, se usa en la ecuación 2.31, de donde

$$x_3 = \frac{2a_0}{-a_1 - (a_1^2 - 4a_0a_2)^{1/2}} = 1.36865$$

La tabla 2.5 se obtiene repitiendo el procedimiento.

**Tabla 2.5**

| $i$ | $x_i$   | $ x_{i+1} - x_i $ |
|-----|---------|-------------------|
| 0   | 0       |                   |
| 1   | 1       | 1.00000           |
| 2   | 2       | 1.00000           |
| 3   | 1.35407 | 0.64593           |
| 4   | 1.36865 | 0.01458           |
| 5   | 1.36881 | 0.00016           |

Para llevar a cabo los cálculos que se muestran en la tabla anterior, puede emplearse Matlab o la TI-92 Plus.



```
eps=0.001;eps1=0.0001;
x0=0; x1=1; x2=2;
for i=1 : 5
    f0=x0^3+2*x0^2+10*x0-20;
    f1=x1^3+2*x1^2+10*x1-20;
    f2=x2^3+2*x2^2+10*x2-20;
    f10=(f1-f0)/(x1-x0);
    f21=(f2-f1)/(x2-x1);
    f210=(f21-f10)/(x2-x0);
    a2=f210;
    a1=f21-(x2+x1)*a2;
    a0=f2-x2*(f21-x1*a2);
```

```
e2_13( )
Prgm
Define F(x)=x^3+2*x^2+10*x-20
.001→eps : .001→eps1: 0.→x0
1.→x1 : 2.→x2
For i, 1, 5
    f (x0)→f0 : f (x1)→f1
    f (x2)→f2 : (f1-f0)/(x1-x0)→f10
    (f2-f1)/(x2-x1)→f21
    (f21-f10)/(x2-x0)→f210 : f210→a2
    f21-(x2+x1)*a2→a1
    f2-x2*(f21-x1*a2) →a0
```

|                                                                                                                                                                                                                                                                                                |                                                                                                                                                                                                                                                                                                            |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <pre> d1=-a1+(a1^2-4*a0*a2)^0.5; d2=-a1-(a1^2-4*a0*a2)^0.5; if abs (d1) &gt;abs (d2)     x3=2*a0/d1; else     x3=2*a0/d2; end f3=x3^3+2*x3^2+10*x3-20; dist=abs (x3-x2) ; disp ([x3, dist]) if or ((dist&lt;eps), (abs (f3)&lt;eps1))     break else     x0=x1;x1=x2;x2=x3; end end end </pre> | <pre> -a1+√(a1^2-4*a0*a2)→d1 -a1-√(a1^2-4*a0*a2)→d2 If abs (d1) &gt;abs (d2) then     2*a0/d1→x3 Else     2*a0/d2→x3 EndIf abs (x3-x2)→dist format (x3, "f5") &amp; " " →d d&amp;format (dist, "f5")→d Disp d If dist&lt;eps or abs (f (x3) )&lt;eps1     Exit x1→x0 : x2→x1 : x3→x2 EndFor EndPrgm </pre> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

**Ejemplo 2.14**

Encuentre las raíces complejas de la ecuación polinomial del ejemplo 2.12

$$f(x) = x^2 + 4 = 0,$$

con el método de Müller.

**Solución Primera iteración**

A1 elegir como valores iniciales:

$$x_0 = 0; \quad x_1 = 1; \quad x_2 = -1$$

y evaluar la función en estos puntos, se tiene

$$f_0 = 4; \quad f_1 = 5; \quad f_2 = 5.$$

Se calculan ahora los coeficientes del polinomio de segundo grado

$$f[x_1, x_0] = \frac{f_1 - f_0}{x_1 - x_0} = \frac{5 - 4}{1 - 0} = 1$$

$$f[x_2, x_1] = \frac{f_2 - f_1}{x_2 - x_1} = \frac{5 - 5}{-1 - 1} = 0$$

$$f[x_2, x_1, x_0] = \frac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0} = \frac{0 - 1}{-1 - 0} = 1$$

Por tanto

$$a_2 = f[x_2, x_1, x_0] = 1$$

$$a_1 = f[x_2, x_1] - (x_2 + x_1)a_2 = 0 - (-1 + 1)(1) = 0$$

$$a_0 = f_2 - x_2(f[x_2, x_1] - x_1a_2) = 5 - (-1)(0 - 1(1)) = 4$$

Calculando los denominadores de la ecuación 2.31

$$-a_1 + (a_1^2 - 4a_0a_2)^{1/2} = 0 + (0 - 4(4)(1))^{1/2} = (-16)^{1/2} = 4j$$

$$-a_1 - (a_1^2 - 4a_0a_2)^{1/2} = 0 - (0 - 4(4)(1))^{1/2} = -(-16)^{1/2} = -4j$$

Como son de igual magnitud se usa cualquiera, por ejemplo  $4j$ . Entonces:

$$x_3 = \frac{2a_0}{-a_1 + (a_1^2 - 4a_0a_2)^{1/2}} = \frac{2(4)}{4j} = \frac{2}{j}$$

al multiplicar numerador y denominador por  $j$ , queda

$$x_3 = \frac{2}{j} \cdot \frac{j}{j} = \frac{2j}{-1} = -2j$$

Hay que observar que aun cuando  $x_0, x_1$  y  $x_2$  son números reales,  $x_3$  ha resultado un número complejo y además es la raíz buscada, lo cual resulta lógico, ya que la ecuación polinomial

$$f(x) = x^2 + 4 = 0$$

es una parábola y el método de Müller consiste, en el caso  $n = 2$ , en usar una parábola para sustituir la función.

La otra raíz es el complejo conjugado de  $x_3$ , o sea  $2j$ .

Para realizar los cálculos de este ejemplo, puede usar el guión de Matlab dado en el ejemplo 2.13, con los valores iniciales  $x_0 = 0$ ;  $x_1 = 1$ ;  $x_2 = -1$  y los cambios correspondientes de la función.

A continuación se proporciona el algoritmo del método de Müller para el caso  $n = 2$ .

### ALGORITMO 2.6 Método de Müller

Para encontrar una raíz real o compleja de la ecuación  $f(x) = 0$ , incluir la función  $f(x)$  y los

DATOS: Valores iniciales  $X_0, X_1, X_2$ ; criterio de convergencia EPS, criterio de exactitud EPS1 y número máximo de iteraciones MAXIT.  
RESULTADOS: La raíz aproximada  $X$  o un mensaje de falla.

PASO 1. Hacer  $I = 1$

PASO 2. Mientras  $I < \text{MAXIT}$ , repetir los pasos 3 a 7.

PASO 3. Hacer  $F10 = (F(X1) - F(X0)) / (X1 - X0)$ .

$F21 = (F(X2) - F(X1)) / (X2 - X1)$ .

$F210 = (F21 - F10) / (X2 - X0)$ .

$A2 = F210$ .

$A1 = F21 - (X2 + X1) * A2$ .

$A0 = F(X2) - X2 * (F21 - X1 * A2)$ .

$D1 = -A1 + (A1^2 - 4 * A0 * A2)^{0.5}$ .

$D2 = -A1 - (A1^2 - 4 * A0 * A2)^{0.5}$ .

PASO 4. Si  $\text{ABS}(D1) > \text{ABS}(D2)$  hacer  $X3 = 2 * A0 / D1$  En caso contrario hacer  $X3 = 2 * A0 / D2$ .

PASO 5. Si  $\text{ABS}(X3 - X0) < \text{EPS}$  o  $\text{ABS}(F(X3)) < \text{EPS1}$ .

IMPRIMIR  $X3$  y TERMINAR.

De otro modo, continuar.

PASO 6. Hacer  $X0 = X1$ .

$X1 = X2$  (actualización de valores iniciales).

$X2 = X3$ .

PASO 7. Hacer  $I = I + 1$ .

PASO 8. IMPRIMIR mensaje de falla: "EL MÉTODO NO CONVERGE A UNA RAÍZ" y TERMINAR.



La siguiente sección puede omitirse sin pérdida de continuidad en el resto del material.

## 2.10 Polinomios y sus ecuaciones

### EVALUACIÓN DE POLINOMIOS

#### Método de Horner

Se desea evaluar un polinomio  $p(x)$  en un valor particular de  $x$ . Por ejemplo, sea el polinomio

$$p(x) = 4x^4 + 3x^3 - 2x^2 + 4x - 8, \quad (2.32)$$

que se desea evaluar en  $x = 2$ .

Factorícese  $x$  en los primeros cuatro términos

$$p(x) = (4x^3 + 3x^2 - 2x + 4)x - 8$$

Dentro de los paréntesis, factorizar  $x$  en los primeros tres términos

$$p(x) = ((4x^2 + 3x - 2)x + 4)x - 8$$

Dentro de los paréntesis interiores, factorícese  $x$  en los primeros dos términos

$$p(x) = ((4x + 3)x - 2)x + 4)x - 8$$

El método de Horner consiste en evaluar, secuencialmente, los paréntesis en esta expresión:

$$\text{Paso 1. Evaluar } (4x + 3) \quad \text{en } x = 2: \quad 4(2) + 3 = 11$$

$$\text{Paso 2. Evaluar } ((11)x - 2) \quad \text{en } x = 2: \quad (11)2 - 2 = 20$$

$$\text{Paso 3. Evaluar } ((20)x + 4) \quad \text{en } x = 2: \quad (20)2 + 4 = 44$$

$$\text{Paso 4. Evaluar } ((44)x - 8) \quad \text{en } x = 2: \quad (44)2 - 8 = 80$$

$$\text{Así, } p(2) = 80.$$

Este proceso puede llevarse a cabo sin las factorizaciones.

$$\text{Escribase } p_4(x) = a_4x^4 + a_3x^3 + a_2x^2 + a_1x + a_0$$

$$\text{Para la ecuación 2.32, } a_4 = 4, a_3 = 3, a_2 = -2, a_1 = 4 \text{ y } a_0 = -8$$

Conviene almacenar los valores intermedios de la evaluación de esta ecuación: 11, 20, 44 y 80, como  $b_3, b_2, b_1$  y  $b_0$ , respectivamente. Sea además, por conveniencia,  $b_4 = a_4 (= 4)$ .

Ahora dispónganse los coeficientes, el valor de  $x$  donde se desea evaluar el polinomio y  $b_4$  en la siguiente forma:

|         |       |       |       |       |
|---------|-------|-------|-------|-------|
| $a_4$   | $a_3$ | $a_2$ | $a_1$ | $a_0$ |
| $x = 2$ | 4     | 3     | -2    | 4     |
|         |       |       |       | -8    |

$$b_4 = 4$$

En la columna de  $a_3$ , se desarrolla el paso 1:  $4(2) + 3 = 11$ . Esto puede verse como multiplicar  $b_4$  por el valor de  $x (= 2)$  y sumar el producto a  $a_3$ . Llámese este resultado  $b_3$ . Esto es:

|         |           |            |            |       |       |
|---------|-----------|------------|------------|-------|-------|
|         | $a_4$     | $a_3$      | $a_2$      | $a_1$ | $a_0$ |
| $x = 2$ | 4         | 3          | -2         | 4     | -8    |
|         |           | +          |            |       |       |
|         |           |            | $4(2) = 8$ |       |       |
|         | $b_4 = 4$ | $b_3 = 11$ |            |       |       |

En la columna de  $a_2$ , se desarrolla el Paso 2:  $(11)2 - 2 = 20$ . Esto es, multiplíquese  $b_3$  por el valor de  $x (= 2)$  y súmese el producto a  $a_2$ . Llámese este resultado  $b_2$ . Lo anterior se ilustra así

|         |           |            |              |       |       |
|---------|-----------|------------|--------------|-------|-------|
|         | $a_4$     | $a_3$      | $a_2$        | $a_1$ | $a_0$ |
| $x = 2$ | 4         | 3          | -2           | 4     | -8    |
|         |           |            | +            |       |       |
|         |           |            | $11(2) = 22$ |       |       |
|         | $b_4 = 4$ | $b_3 = 11$ | $b_2 = 20$   |       |       |

Repetiendo este proceso hasta calcular  $b_0$  se tiene

|         |           |            |            |              |              |
|---------|-----------|------------|------------|--------------|--------------|
|         | $a_4$     | $a_3$      | $a_2$      | $a_1$        | $a_0$        |
| $x = 2$ | 4         | 3          | -2         | 4            | -8           |
|         |           |            |            | +            | +            |
|         |           |            |            | $20(2) = 40$ | $44(2) = 88$ |
|         | $b_4 = 4$ | $b_3 = 11$ | $b_2 = 20$ | $b_1 = 44$   | $b_0 = 80$   |

El valor  $p(2)$  resulta en  $b_0$ .

### Ejemplo 2.15

Evalúe el polinomio

$$x^5 - 4x^3 + 2x + 3 \quad \text{en } x = 3,$$

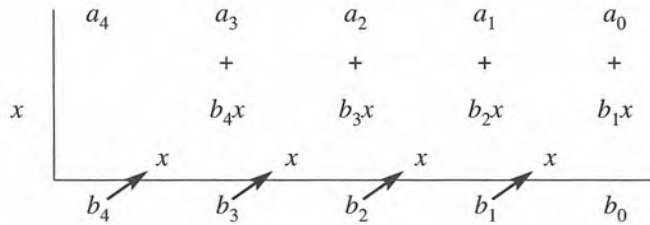
mediante el método de Horner.

**Solución** La no aparición de los términos en  $x^4$  y en  $x^2$  del polinomio significa que sus coeficientes son cero; para fines del método en estudio, dichos ceros deben aparecer en el arreglo  
Coeficientes de:

|         |           |            |            |             |              |                       |
|---------|-----------|------------|------------|-------------|--------------|-----------------------|
|         | $x^5$     | $x^4$      | $x^3$      | $x^2$       | $x$          | Término independiente |
|         | $a_5 = 1$ | $a_4 = 0$  | $a_3 = -4$ | $a_2 = 0$   | $a_1 = 2$    | $a_0 = 3$             |
|         |           | +          | +          | +           | +            |                       |
| $x = 3$ |           | $1(3) = 3$ | $3(3) = 9$ | $5(3) = 15$ | $15(3) = 45$ | $47(3) = 141$         |
|         | $b_5 = 1$ | $b_4 = 3$  | $b_3 = 5$  | $b_2 = 15$  | $b_1 = 47$   | $b_0 = 144$           |

De aquí  $p(3) = 144$ .

Se generaliza este método con polinomios de cuarto grado; sin embargo la extensión a cualquier grado, es inmediata. Así:



donde puede verse que:

$$b_4 = a_4,$$

$$b_3 = a_3 + b_4x, \quad b_2 = a_2 + b_3x, \quad b_1 = a_1 + b_2x, \quad b_0 = a_0 + b_1x;$$

esto es

$$b_4 = a_4 \text{ y } b_k = a_k + b_{k+1}x, \text{ para } k = 3, 2, 1, 0 \quad (2.33)$$

Mediante una sustitución regresiva puede verse con claridad por qué  $p(x) = b_0$ .

Sustituyendo en  $b_0 = a_0 + b_1x$  a  $b_1$  por  $a_1 + b_2x$ , se tiene:

$$b_0 = a_0 + (a_1 + b_2x)x$$

y ahora se reemplaza en la última expresión  $b_2$  con  $a_2 + b_3x$  y así sucesivamente, con lo cual se obtiene:

$$b_0 = ((a_4x + a_3)x + a_2)x + a_1)x + a_0 = p(x)$$

Las ecuaciones 2.33 representan un algoritmo programable y, como se verá más adelante, de elevada eficiencia para evaluar un polinomio  $p(x)$  en algún valor particular de  $x$ .

Se describe enseguida el algoritmo de Horner.

### ALGORITMO 2.7 Método de Horner

Para evaluar el polinomio

$$p(x) = a_nx^n + a_{n-1}x^{n-1} + \dots + a_1x + a_0 \quad \text{proporcionar los}$$

**DATOS:**  $n$ : Grado del polinomio.  
 $a_n, a_{n-1}, \dots, a_0$ : Coeficientes del polinomio.  
 $t$ : Valor de  $x$  en donde se desee evaluar  $p(x)$

**RESULTADOS:**  $p(t)$  en  $b_0$ .

PASO 1. Hacer  $b_n = a_n$

PASO 2. Para  $k = n-1, n-2, \dots, 0$  realizar el paso 3.

PASO 3. Hacer  $b_k = b_{k+1}t + a_k$ .

PASO 4. IMPRIMIR  $b_0$ .



### Método de Horner iterado

El método de Horner tiene otras características, que se analizarán enseguida.

Tómese de nuevo el polinomio general de cuarto grado  $p_4(x)$  y divídase entre  $(x - t)$ , donde  $t$  es un valor particular de  $x$ , lo que se expresa como:

$$p_4(x) = (x - t) q(x) + R, \quad (2.34)$$

donde  $q(x)$  es el polinomio cociente (en este caso de tercer grado) y  $R$  una constante llamada residuo.

Sustituyendo  $x$  con  $t$  se obtiene  $p_4(t) = R$ , de modo que el polinomio evaluado en un valor particular de  $x$  es igual al residuo  $R$  de la división,  $R = b_0$ .

Al derivar la ecuación 2.34 con respecto a  $x$  (recuérdese que  $t$  y  $R$  son constantes), se tiene:

$$p_4'(x) = (x - t)q'(x) + q(x)$$

Haciendo  $x = t$  resulta

$$p_4'(t) = q(t), \quad (2.35)$$

esto es, la derivada del polinomio  $p_4(x)$  evaluada en  $x = t$  es el cociente  $q(x)$  evaluado en  $t$ , toda vez que

$$p_4'(t) = q(t) = b_4t^3 + b_3t^2 + b_2t + b_1$$

y en general

$$q(x) = b_4x^3 + b_3x^2 + b_2x + b_1 \quad (2.36)$$

donde  $b_4, b_3, b_2$  y  $b_1$  son los valores intermedios que resultan en la evaluación de  $p_4(x)$  en  $t$  por el método de Horner (véase Ej. 2.15). Así pues, si después de evaluar  $p_4(x)$  en  $t$  se desea evaluar también  $p_4'(x)$  en  $t$ , puede aplicarse una vez más el método de Horner a los valores intermedios  $b_4, b_3, b_2$  y  $b_1$ , como se ilustra enseguida.

#### Ejemplo 2.16

Sea  $p(x) = 3x^3 - 4x - 1$ . Evalúe

a)  $p(2)$

b)  $p'(2)$

**Solución**

a) Para evaluar  $p(2)$ , se tiene

|         |           |            |             |             |
|---------|-----------|------------|-------------|-------------|
|         | $a_3$     | $a_2$      | $a_1$       | $a_0$       |
| $x = 2$ | 3         | 0          | -4          | -1          |
|         |           | +          | +           | +           |
|         |           | $3(2) = 6$ | $6(2) = 12$ | $8(2) = 16$ |
|         | $b_3 = 3$ | $b_2 = 6$  | $b_1 = 8$   | $b_0 = 15$  |

y  $p(2) = 15$ .

b) Como se dijo

$$p'(t) = b_3t^2 + b_2t + b_1$$

Para evaluar  $p'(2)$  se emplea de nuevo el método de Horner. Esto se logra eficientemente, repitiendo los pasos de los cálculos descritos; esto es bajo  $b_3$ ,  $b_2$  y  $b_1$  del arreglo anterior. Para almacenar los nuevos valores intermedios de esta evaluación se emplean  $c_3$ ,  $c_2$  y  $c_1$ . Nótese que como  $b_1$  es el término independiente de  $p'(x)$ , el proceso de evaluación termina una vez que se obtuvo  $c_1$ , y éste es el valor buscado de  $p'(2)$ .

|         |           |            |              |             |
|---------|-----------|------------|--------------|-------------|
| $x = 2$ | $a_3$     | $a_2$      | $a_1$        | $a_0$       |
|         | 3         | 0          | -4           | -1          |
|         |           | +          | +            | +           |
|         |           | $3(2) = 6$ | $6(2) = 12$  | $8(2) = 16$ |
| $x = 2$ | $b_3 = 3$ | $b_2 = 6$  | $b_1 = 8$    | $b_0 = 15$  |
|         |           | +          | +            |             |
|         |           | $3(2) = 6$ | $12(2) = 24$ |             |
|         | $c_3 = 3$ | $c_2 = 12$ | $c_1 = 32$   |             |

De esto,  $p'(2) = 32$ . El lector puede verificar el resultado derivando  $p(x)$  y evaluando la derivada en  $x = 2$ .

En la práctica, los cálculos suelen disponerse sin tantos comentarios.

### Ejemplo 2.17

Evalúe  $5x^3 - 2x^2 + 10$  y su primera derivada en  $x = 0.5$

#### Solución

|     |       |       |       |        |
|-----|-------|-------|-------|--------|
| 0.5 | $a_3$ | $a_2$ | $a_1$ | $a_0$  |
|     | 5     | -2    | 0     | 10     |
|     |       | +     | +     | +      |
|     |       | 2.5   | 0.25  | 0.125  |
| 0.5 | $b_3$ | $b_2$ | $b_1$ | $b_0$  |
|     | 5     | 0.5   | 0.25  | 10.125 |
|     |       | +     | +     |        |
|     |       | 2.5   | 1.50  |        |
|     | $c_3$ | $c_2$ | $c_1$ |        |
|     | 5     | 3     | 1.75  |        |

De esto  $p(0.5) = 10.125$  y  $p'(0.5) = 1.75$ .

En este punto conviene presentar el algoritmo de Horner iterado para evaluar un polinomio y su primera derivada en un valor  $t$ .

**ALGORITMO 2.8** Método de Horner iterado

Para evaluar el polinomio

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

y su primera derivada  $p'(x)$  en  $x = t$ , proporcionar los

**DATOS:**  $n$ : Grado del polinomio,  
 $a_n, a_{n-1}, \dots, a_0$ : Coeficientes del polinomio.  
 $t$ : Valor de  $x$  en donde se desea evaluar  $p(x)$  y  $p'(x)$ .

**RESULTADOS:**  $p(t)$  en  $b_0$  y  $p'(t)$  en  $c_1$ .

PASO 1. Hacer  $b_n = a_n$  y  $c_n = b_n$ .

PASO 2. Para  $k = n-1, n-2, \dots, 1$  realizar los pasos 3 y 4.

PASO 3. Hacer  $b_k = b_{k+1} t + a_k$ .

PASO 4. Hacer  $c_k = c_{k+1} t + b_k$ .

PASO 5. Hacer  $b_0 = b_1 t + a_0$ .

PASO 6. IMPRIMIR  $b_0$  y  $c_1$ .

**CUENTA DE OPERACIONES**

Si bien la implementación del método de Horner en una computadora es una de sus ventajas, no lo es menos su eficiencia, que se verá a continuación, contando las operaciones en el método de evaluación usual y comparando su número con el del método de Horner. Tomando de nuevo el polinomio general de cuarto grado

$$p_4(x) = a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0$$

**a) Método usual**

$a_4 x^4$  requiere cuatro multiplicaciones

$a_3 x^3$  requiere tres multiplicaciones

$a_2 x^2$  requiere dos multiplicaciones

$a_1 x$  requiere una multiplicación

$a_4 x^4 + a_3 x^3 + a_2 x^2 + a_1 x + a_0$  necesita cuatro sumas/restas.

En total se realizan 10 multiplicaciones y cuatro sumas/restas.

**b) Método de Horner**

$$\left. \begin{array}{l} b_4 = a_4 \\ b_3 = b_4 x + a_3 \\ b_2 = b_3 x + a_2 \\ b_1 = b_2 x + a_1 \\ b_0 = b_1 x + a_0 \end{array} \right\} \text{ Se requiere una multiplicación y una suma para cada } b.$$

En total cuatro multiplicaciones y cuatro sumas/restas.

Hay una reducción de 60% en el número de multiplicaciones requeridas y, consecuentemente, un error de redondeo menor.



A continuación se verá una aplicación del método de Horner en la búsqueda de raíces reales de ecuaciones de la forma  $f(x) = 0$ , donde  $f(x)$  es un polinomio de grado  $n$ .

Combinando las ecuaciones 2.34 y 2.36 y el resultado  $R = b_0$ ,

$$f(x) = (x - t) (b_4x^3 + b_3x^2 + b_2x + b_1) + b_0$$

y como  $f(t) = b_0$ ,

$$f(x) = (x - t) (b_4x^3 + b_3x^2 + b_2x + b_1) + f(t)$$

Si  $t$  es una raíz de  $f(x) = 0$ , se tiene  $f(t) = 0$  y la expresión resultante

$$f(x) = (x - t) (b_4x^3 + b_3x^2 + b_2x + b_1)$$

indica que  $x = t$  es una raíz (lo cual ya se sabía), pero lo más importante es que las raíces restantes de  $f(x) = 0$  son las raíces de

$$b_4x^3 + b_3x^2 + b_2x + b_1 = 0, \quad (2.37)$$

una ecuación polinomial de tercer grado y, por tanto, más fácil de manejar que la ecuación original; además, sus coeficientes son los valores ya citados  $b_4$ ,  $b_3$ ,  $b_2$  y  $b_1$ .

Si se sospecha que la raíz  $t$  se repite (es decir  $t$  es raíz de la ecuación 2.37), véase el valor de  $c_1$  del método de Horner iterado, ya que éste será muy cercano a cero si así fuera; esto es  $p'(t) = 0$  en ese caso.

Ahora, desarróllese el método de Newton-Raphson con el método de Horner iterado, llamado método de Birge-Vieta.

## RAÍCES DE UNA ECUACIÓN POLINOMIAL $P_n(x) = 0$

### Método de Birge-Vieta\*

De los métodos vistos para encontrar raíces, el de Newton-Raphson resulta el más adecuado para ser empleado en conjunción con el método de Horner iterado. Se resuelve a continuación un ejemplo con esta combinación.

#### Ejemplo 2.18

Aproxime las raíces reales del polinomio

$$p(x) = 4x^4 + 3x^3 - 2x^2 + 4x - 8$$

#### Solución



PASO 1. Al analizar gráficamente la función se advierte que tiene dos raíces reales, una alrededor de 1, y la otra alrededor de -2.

PASO 2. Se elige 0 como valor inicial para encontrar la primera raíz.

PASO 3. Con el método de Horner  $b_0 = -8$  y  $c_1 = 4$

PASO 4. Con el método de Newton-Raphson  $t_1 = t_0 - b_0/c_1 = 0 - (-8)/4 = 2$

PASO 5. Al repetir los pasos 3 y 4 se tiene

$$t_2 = t_1 - b_0/c_1 = 2 - 8/16 = 1.5$$

$$t_3 = t_2 - b_0/c_1 = 1.5 - 23.875/72.25 = 1.1696$$

$$t_4 = t_3 - b_0/c_1 = 1.1696 - 6.2258/37.2287 = 1.0023$$

\* Este método también se conoce como Newton-Raphson-Horner.

Este proceso converge al valor 0.9579

PASO 6. Se toma 0.9579 como primera raíz del polinomio.

PASO 7. El polinomio de menor grado que se obtiene con esta raíz conduce a  $p(x) = 4x^3 + 6.831315x^2 + 4.5435x + 8.3518$

PASO 2. Se elige nuevamente 0 como valor inicial para encontrar la segunda raíz.

PASO 3. Con el método de Horner  $b_0 = 8.3518$  y  $c_1 = 4.5435$

PASO 4. Con el método de Newton-Raphson

$$t_1 = t_0 - b_0/c_1 = 0 - 8.3518/4.5435 = -1.8382$$

PASO 5. Al repetir los pasos 3 y 4 se tiene

$$t_2 = t_1 - b_0/c_1 = -1.8382 - (-1.7623) / 19.9772 = -1.7500$$

$$t_3 = t_2 - b_0/c_1 = -1.7500 - (-1.1575) / 17.3841 = -1.7434$$

Este proceso converge al valor -1.7433

PASO 6. Se toma -1.7433 como segunda raíz del polinomio.

PASO 7. El polinomio disminuido con esta raíz conduce a  $p(x) = 4x^2 - 0.141885x + 4.79085$

Que tiene las raíces:  $0.01774 \pm 1.09426j$



Matlab posee una función que calcula todas las raíces de ecuaciones polinomiales, suministrando los coeficientes del polinomio.

Para este caso la instrucción quedaría:

```
Roots([4 3 -2 4 -8])
```

Y se obtiene como respuesta:

```
ans =
-1.74332500029465
0.01773561271143 + 1.09425660030108i
0.01773561271143 - 1.09425660030108i
0.95785377487178
```

La calculadora TI-92 Plus también obtiene estas raíces; las reales las obtiene con la instrucción

```
Solve (4x^4+3x^3-2x^2+4x=8, x)
```

y las complejas con

```
cSolve (4x^4+3x^3-2x^2+4x=8, x)
```

En el CD se encuentra el **PROGRAMA 2.1** para este algoritmo.

En cada etapa se ha calculado una aproximación a cada una de las raíces reales de  $p(x) = 0$ ; conforme se avanza en las etapas, los coeficientes  $b_1, b_2, \dots, b_n$  de cada etapa se alejan de los valores verdaderos, debido a la propagación de errores, y las aproximaciones a las raíces correspondientes también son más inexactas. Para disminuir la pérdida de

exactitud se ha sugerido trabajar primero con la raíz más pequeña en valor absoluto, luego con la raíz real restante más pequeña en magnitud, y así sucesivamente.

### Método de Lin\*

En 1941, S.N. Lin publicó un procedimiento que se fundamenta en el resultado

$$R = f(t) = b_0 = b_1 t + a_0$$

y en que si  $t$  es una raíz de  $p_n(x) = 0$ , entonces:

$$R = 0 = b_1 t + a_0$$

o

$$t = -a_0/b_1(t) \quad (2.38)$$

Se ha escrito  $b_1(t)$  en lugar de  $b_1$  para hacer énfasis en que el valor de  $b_1$  (y de las demás  $b$ ) depende del valor  $t$  donde se evalúa  $f(x)$  y así ver el lado derecho de la ecuación 2.38 como una función de  $t$ . Lo que puede escribirse como:

$$t = -a_0/b_1(t) = g(t) \quad (2.39)$$

y se le puede aplicar el método de punto fijo, empezando con un valor inicial  $t_0$  cercano a la raíz  $t$ , de modo que:

$$t_1 = -a_0/b_1(t_0) = g(t_0)$$

Restando en ambos lados  $t_0$

$$t_1 - t_0 = -\frac{a_0 + t_0 b_1(t_0)}{b_1(t_0)}$$

o

$$t_1 = t_0 - \frac{R(t_0)}{b_1(t_0)} \quad (2.40)$$

y se obtiene el algoritmo de Lin. Este método no requiere el cálculo de las  $c$  como el de Birge-Vieta, por lo que el trabajo por iteración se reduce a la mitad. Esta reducción contrasta con un orden bajo de convergencia y la inestabilidad propia del método de punto fijo.

### Ejemplo 2.19

Encuentre una raíz real de la ecuación

$$x^4 - 3x^3 + 2x - 1 = 0,$$

con el método de Lin y un valor inicial  $t_0 = 2.8$

### Solución Primera iteración

$$R(2.8) = 0.2096; \quad b_1(2.8) = 0.432;$$

$$t_1 = t_0 - R(t_0) / b_1(t_0) = 2.8 - 0.2096 / 0.432 = 2.3148$$

\* En el capítulo 4 se desarrolla el método de Bairstow.



**Segunda iteración**

$$R(2.3148) = -4.8692; b_1(2.3148) = -1.6715;$$

$$t_2 = t_1 - R(t_1) / b_1(t_1) = 2.3148 - (-4.8692) / (-1.6715) = -0.5983$$

Al continuar las iteraciones se advierte que el método es inestable y no llega a la raíz 2.78897.

La estabilidad\* del método puede mejorarse en una raíz  $\bar{x}_k$ , si se conoce una buena aproximación a  $\bar{x}_k$ . Para esto se incorpora el parámetro  $\lambda$  a la ecuación 2.40 de Lin y queda

$$t_1 = t_0 - \lambda \frac{R}{b_1}$$

donde:

$$\lambda = - \frac{f(0)}{t_0 f'(t_0)}$$

Con  $t_0 = 2.8$ ,  $\lambda = 0.018555$  y la fórmula modificada de Lin, en general es

$$t = t - \lambda \frac{R}{b_1} \quad (2.41)$$

**Ejemplo 2.20**

Con la fórmula modificada de Lin, aproxime una raíz real de la ecuación

$$f(x) = x^4 - 3x^3 + 2x - 1 = 0,$$

use como valor inicial  $t_0 = 2.8$

**Solución**

$$f(0) = -1; \quad f'(2.8) = 19.248$$

$$\lambda = -(-1) / 2.8 / 19.248 = 0.018555$$

**Primera iteración**

$$R(2.8) = 0.2096; \quad b_1(2.8) = 0.432;$$

$$t_1 = t_0 - \lambda R(t_0) / b_1(t_0) = 2.8 - 0.018555 (0.2096) / 0.432 \\ = 2.791$$

**Segunda iteración**

$$R(2.791) = 0.03808; \quad b_1(2.791) = 0.37194;$$

$$t_2 = t_1 - \lambda R(t_1) / b_1(t_1) \\ = 2.791 - 0.018555 (0.03808) / (0.37194) = 2.7891$$

Al continuar las iteraciones se encuentra la raíz 2.78897

\* Hildebrand. *Introduction to Numerical Analysis*. McGraw Hill, 2a. Ed., pp. 591-595.

Los métodos anteriores son válidos para raíces reales y complejas. Sin embargo, para las segundas deberá inicializarse con un número complejo y llevar a cabo las operaciones complejas correspondientes. Cuando los coeficientes de  $p_n(x) = 0$  son reales, las raíces complejas aparecen en pares conjugados

$$\bar{x}_k = a + bj, \bar{x}_{k+1} = a - bj,$$

lo que se puede aprovechar buscando en  $p_n(x) = 0$  el factor cuadrático

$$(x - \bar{x}_k)(x - \bar{x}_{k+1}) = x^2 - 2ax + (a^2 + b^2)$$

de coeficientes reales que genera  $\bar{x}_k$  y  $\bar{x}_{k+1}$

### Factores cuadráticos. Método de Lin

Sea el polinomio

$$f(x) = x^n + a_{n-1}x^{n-1} + \dots + a_2x^2 + a_1x + a_0 \quad (2.42)$$

Si  $a_n$  no es uno,  $f(x)$  puede dividirse entre  $a_n$  para obtener la ecuación 2.42.

Al dividir la ecuación 2.42 entre la expresión cuadrática

$$x^2 + px + q \quad (2.43)$$

$$\begin{aligned} f(x) &= x^n + a_{n-1}x^{n-1} + \dots + a_2x^2 + a_1x + a_0 \\ &= (x^2 + px + q)(x^{n-2} + b_{n-3}x^{n-3} + \dots + b_1x + b_0) + Rx + S, \end{aligned} \quad (2.44)$$

donde  $Rx + S$  es el residuo lineal de la división, y  $R$  y  $S$  dependen de  $p$  y  $q$ .

Para que la ecuación 2.43 sea un factor cuadrático de la 2.42 (es decir, que la divida exactamente) es necesario que el residuo lineal sea cero o simbólicamente que

$$R(p, q) = 0 \quad \text{y} \quad S(p, q) = 0 \quad (2.45)$$

De donde nuestro objetivo será encontrar  $p$  y  $q$ , tales que se cumplan las ecuaciones 2.45.

Conviene tener un método que permita calcular  $R$  y  $S$  sin verificar la división de la 2.42 por la 2.43. Para obtenerlo, se igualan los coeficientes de las mismas potencias de  $x$  en los dos miembros de la ecuación 2.44

$$\left. \begin{aligned} a_{n-1} &= b_{n-3} + p \\ a_{n-2} &= b_{n-4} + p b_{n-3} + q \\ a_{n-3} &= b_{n-5} + p b_{n-4} + q b_{n-3} \\ &\vdots \\ a_k &= b_{k-2} + p b_{k-1} + q b_k \\ &\vdots \\ a_1 &= p b_0 + q b_1 + R \\ a_0 &= q b_0 + S \end{aligned} \right\} \quad (2.46)$$

Despejando  $b_k$  de la expresión general (usando para ello  $a_{k+2} = b_k + p b_{k+1} + q b_{k+2}$ ) se obtiene

$$b_k = a_{k+2} - p b_{k+1} - q b_{k+2} \quad \text{para } k = n-3, n-4, \dots, 0 \quad (2.47)$$

con

$$b_{n-1} = 0 \quad \text{y} \quad b_{n-2} = 1 \quad (2.48)$$

el algoritmo buscado para obtener los coeficientes del polinomio cociente de la 2.44 y además

$$R = a_1 - p b_0 - q b_1 \quad (2.49)$$

$$S = a_0 - q b_0 \quad (2.50)$$

Al emplear las condiciones de la ecuación 2.45

$$\begin{aligned} a_1 - p b_0 - q b_1 &= 0 \\ a_0 - q b_0 &= 0 \end{aligned} \quad (2.51)$$

se pueden obtener, despejando, valores de  $p$  y  $q$  para formar una nueva expresión 2.43, quizás más cercana al factor cuadrático que andamos buscando.

El método de Lin consiste en:

- PASO 1. Proponer aproximaciones iniciales de los valores desconocidos  $p$  y  $q$  (pueden llamarse  $p_0$  y  $q_0$ ).
- PASO 2. Emplear las ecuaciones 2.47 para obtener aproximaciones de  $b_{n-3}, b_{n-4}, \dots, b_1, b_0$ .
- PASO 3. Calcular  $R$  y  $S$ . Si son cero o suficientemente cercanas a éste, el problema está terminado. En caso contrario, se estiman nuevos valores de  $p$  y  $q$  (pueden llamarse  $p_1$  y  $q_1$ )

$$p_1 = \frac{a_1 - q_0 b_1}{b_0} \quad \text{y} \quad q_1 = \frac{a_0}{b_0}$$

para volver al paso 2.

### Ejemplo 2.21

Encuentre los factores cuadráticos de la ecuación polinomial de grado cuatro

$$f(x) = x^4 - 8x^3 + 39x^2 - 62x + 50 = 0$$

### Solución

PASO 1. Se propone  $p = 0$  y  $q = 0$ .

PASO 2.  $b_3 = 0; b_2 = 1;$

$$b_1 = a_3 - p b_2 - q b_3 = -8; \quad b_0 = a_2 - p b_1 - q b_2 = 39.$$

PASO 3.  $R = a_1 - p b_0 - q b_1 = -62 \quad S = a_0 - q b_0 = 50$

$$p_1 = \frac{a_1 - q b_1}{b_0} = \frac{-62}{39} = -1.5897;$$

$$q_1 = a_0 / b_0 = 50 / 39 = 1.2821$$

Al repetir los pasos 2 y 3 se encuentra la siguiente sucesión de valores:

| $p$     | $q$    | $R$      | $S$     |
|---------|--------|----------|---------|
| -1.9358 | 1.8164 | -10.0204 | 14.7086 |
| -2.0109 | 1.9708 | -1.4494  | 3.9171  |
| -2.0090 | 2.0011 | 0.0469   | 0.7586  |
| -2.0034 | 2.0030 | 0.1396   | 0.0458  |
| -2.0009 | 2.0013 | 0.0632   | -0.0410 |
| -2.0001 | 2.0004 | 0.0187   | -0.0235 |



Por lo que el factor cuadrático es

$$x^2 - 2x + 2$$

Para llevar a cabo los cálculos que se muestran en la tabla anterior, puede usar Matlab o la TI-92 Plus.



```
% Método de Lin
format short
% Datos
n=5; a=[50 -62 39 -8 1];
p=0; q=0;
b(n-1)=0; b(n-2)=1; i=1; R=1; S=1;
while or (abs(R)>=0.01,abs(S)>=0.01),i>10)
    for L=1: n-3
        k=n-L-2;
        b(k)=a(k+2)-p*b(k+1)-q*b(k+2);
    end
    R=a(2)-p*b(1)-q*b(2);
    S=a(1)-q*b(1);
    p=(a(2)-q*b(2))/b(1); q=a(1)/b(1);
    disp([p,q,R,S])
    i=i+1;
end
```



```
e2_21( )
Prgm
5→n : 50→a[1] : -62→a[2] : 39→a[3] : -8→a[4]
1→a[5] : 0→b[1] : 0→b[2] : 1→b[n-2]
0→b[n-1]
0→p : 0→q : 1→i : 1→r : 1→s : ClrIO
Disp "   p   q   R   S"
Loop
    For L, 1, n-3
        n-L-2→k
        a[k+2]-p*b[k+1]-q*b[k+2]→b[k]
    EndFor
    a[2]-p*b[1]-q*b[2]→r
    a[1]-q*b[1]→s
    (a[2]-q*b[2])/b[1]→p
    a[1]/b[1]→q
    format(p,"f4")&" "&format(q,"f4")&" "&format(r,"f4")&" "&format(s,"f4")→d
    Disp d
    If abs(r)<.01 or abs(s)<.01
        Exit
    EndLoop
EndPrgm
```

## Ejercicios



**2.1** La ecuación de estado de Van der Waals para un gas real es:

$$\left(P + \frac{a}{V^2}\right)(V - b) = RT \quad (1)$$

donde:

$P$  = presión en *atm*

$T$  = temperatura en *K*

$R$  = constante universal de los gases en *atm-L / (gmol K)* = 0.08205

$V$  = volumen molar del gas en *L/gmol*

$a, b$  = constantes particulares para cada gas

Para los siguientes gases, calcule  $V$  a 80 °C para presiones de 10, 20, 30 y 100 *atm*.

| Gas             | $a$     | $b$     |
|-----------------|---------|---------|
| CO <sub>2</sub> | 3.599   | 0.04267 |
| Dimetilamina    | 37.49   | 0.19700 |
| He              | 0.03412 | 0.02370 |
| Óxido nítrico   | 1.34    | 0.02789 |

**Solución** La ecuación 1 también puede escribirse como



$$PV^3 - bPV^2 - RTV^2 + aV - ab = 0 \quad (2)$$

que es un polinomio cúbico en el volumen molar  $V$ ; entonces, para una  $P$  y una  $T$  dadas, puede escribirse como una función de la variable  $V$

$$f(V) = pV^3 - (Pb + RT)V^2 + aV - ab = 0 \quad (3)$$

Esta ecuación se resuelve con el método de posición falsa para encontrar el volumen molar.

### VALORES INICIALES

El **PROGRAMA 2.2** del CD realiza los cálculos necesarios para resolver esta ecuación, usando como intervalo inicial:  $V_1 = 0.8 \nu$  y  $V_D = 1.2 \nu$ , donde  $\nu = RT / P$ , el volumen molar ideal. (Se resuelve sólo el caso del CO<sub>2</sub> a 10 atm y 80 °C, dejando como ejercicio para el lector los demás casos.)

Los valores obtenidos para las diferentes iteraciones son los siguientes:

| iteración | $V_M$ (L / gmol) | $ f(V_M) $              |
|-----------|------------------|-------------------------|
| 1         | 2.603856         | $0.1362 \times 10^2$    |
| 2         | 2.734767         | $0.5711 \times 10^1$    |
| 3         | 2.785884         | $0.2141 \times 10^1$    |
| 4         | 2.804528         | 0.7685                  |
| 5         | 2.811156         | 0.2716                  |
| 6         | 2.813489         | $0.9546 \times 10^{-1}$ |
| 7         | 2.814309         | $0.3348 \times 10^{-1}$ |
| 8         | 2.814596         | $0.1173 \times 10^{-1}$ |
| 9         | 2.814697         | $0.4113 \times 10^{-2}$ |
| 10        | 2.814732         | $0.1441 \times 10^{-2}$ |
| 11        | 2.814744         | $0.5050 \times 10^{-3}$ |
| 12        | 2.814749         | $0.1769 \times 10^{-3}$ |
| 13        | 2.814750         | $0.6200 \times 10^{-4}$ |

Se utilizó el criterio de exactitud

$$|f(V)| < 10^{-4}$$

aunque puede verse que desde la iteración 7, el cambio en los valores de  $V_M$  son solamente en la cuarta cifra decimal, que en este caso representan décimas de mililitro.

Resultado: el volumen molar del  $\text{CO}_2$  a una presión de 10 atm y una temperatura de 80 °C (= 353.2 K) es 2.81475 L/gmol.

Los cálculos pueden realizarse con el siguiente guión de Matlab:



```
P=10; R=0.08205; T=80+273.2;
a=3.599; b=0.04267;
v=R*T/P;
vi=0.8*v; vd=1.2*v; Eps=0.0001;
fi=P*vi^3-(P*b+R*T)*vi^2+a*vi-a*b;
fd=P*vd^3-(P*b+R*T)*vd^2+a*vd-a*b;
fm=1; k=0;
while abs(fm) > Eps
    k=k+1;
    vm=(vi*fd-vd*fi)/(fd-fi);
    fm=P*vm^3-(P*b+R*T)*vm^2+a*vm-a*b;
    fprintf('%3d %8.6f %8.4e\n',k,vm,abs(fm))
    if fd*fm > 0
        vd=vm; fd=fm;
    else
        vi=vm; fi=fm;
    end
end
```



También puede resolverse este problema con la función `fzero` de Matlab, para lo cual es necesario escribir la siguiente función y grabarla en el área de trabajo de Matlab con el nombre `Vander.m`:

```
function f=Vander(V)
P=10; R=0.08205; T=80+273.2;
a=3.599; b=0.04267;
f=P*V^3-(P*b+R*T)*V^2+a*V-a*b;
```

Ahora use:

```
P=10; R=0.08205; T=80+273.2;
fzero('Vander', R*T/P)
```

En este caso el resultado que proporciona Matlab es:

```
ans = 2.8148
```

Dado que la función de este problema es un polinomio de tercer grado en el volumen, también puede usarse la función `roots` que calcula **todas** las raíces de un polinomio. Para ello use el guión que se proporciona enseguida:

```
P=10; R=0.08205; T=80+273.2;
a=3.599; b=0.04267;
roots ( [P -(P*b+R*T) a -a*b] )
```

En este caso el resultado que proporciona Matlab es:

```
ans =
    2.8148
    0.0630 + 0.0386i
    0.0630 - 0.0386i
```

Dado que el polinomio es de tercer grado, siempre tendremos tres raíces. Las posibilidades matemáticas son: tres raíces reales distintas, tres raíces reales iguales, dos raíces reales iguales y una distinta, una raíz real y dos complejas conjugadas. En cada uno de estos casos, ¿cuál de las tres correspondería al volumen buscado y qué significado tendrían las dos restantes?

En el caso de la TI-92 Plus puede usar:



```
Al usar
Solve(10v^3-(.4267+.08205*353.2)v^2+3.599v-3.599*.04267=0,v)
se obtiene:
v=2.81475
y al usar
cSolve(10v^3-(.4267+.08205*353.2)v^2+3.599v-3.599*.04267=0,v)
se obtiene:
v=.062962+.038622i or v=.062962-.038622i or v=2.81475
```

2.2 La fórmula de Bazin para la velocidad de un fluido en canales abiertos está dada por:

$$v = c (re)^{1/2}$$

con

$$c = \frac{87}{0.552 + \frac{m}{(r)^{1/2}}}$$

donde:

$m$  = coeficiente de rugosidad

$r$  = radio hidráulico en pies (área dividida entre el perímetro mojado)

$e$  = pendiente de la superficie del fluido

$v$  = velocidad del fluido en pies/segundos

Calcule el radio hidráulico correspondiente a los siguientes datos (dados en unidades consistentes) por el método de Steffensen.

$$m = 1.1; \quad e = 0.001; \quad v = 5$$

### Solución



$$\text{Sustituyendo } c \text{ en } v: v = \frac{87 (re)^{1/2}}{0.552 + \frac{m}{(r)^{1/2}}}$$

o bien

$$\left(0.552 + \frac{m}{(r)^{1/2}}\right) v = 87 (r)^{1/2} (e)^{1/2}$$

multiplicando ambos lados por  $(r)^{1/2}$

$$[0.552(r)^{1/2} + m] v = 87 (e)^{1/2} r$$

y “despejando”  $r$  se llega a:

$$r = \frac{[0.552(r)^{1/2} + m] v}{87 (e)^{1/2}}$$

una de las formas de  $g(r) = r$ , necesaria para el método de Steffensen. Sin embargo, antes de usar el método, conviene averiguar el comportamiento de  $g'(r)$

$$g'(r) = \frac{0.552 v}{174 (r)^{1/2} (e)^{1/2}}$$

sustituyendo valores:

$$g'(r) = \frac{0.5}{(r)^{1/2}}$$

Como el radio hidráulico debe ser mayor de cero, ya que un valor negativo o cero no tendría significado físico y como  $|g'(r)| < 1$  para  $(r)^{1/2} > 0.5$ , o  $r > 0.25$ , se selecciona como valor inicial de  $r$  a 1.0. Con esto:

$$g'(1) = 0.5$$

y el método puede aplicarse con cierta garantía de convergencia:

### Primera iteración

$$\begin{aligned}
 r_0 &= 1 \\
 r_1 &= g(r_0) = \frac{[0.552 (1)^{1/2} + 1.1] (5)}{87 (0.001)^{1/2}} = 3.00235 \\
 r_2 &= g(r_1) = \frac{[0.552 (3.00235)^{1/2} + 1.1] (5)}{87 (0.001)^{1/2}} = 3.73742 \\
 r_3 &= r_0 - \frac{(r_1 - r_0)^2}{r_2 - 2r_1 + r_0} = 1 - \frac{(3.00235 - 1)^2}{3.73742 - 2(3.00235) + 1} = 4.16380
 \end{aligned}$$

### Segunda iteración

Tomando ahora como nuevo valor inicial  $r_3 = 4.16380$ , se tiene:

$$\begin{aligned}
 r_0 &= 4.16380 \\
 r_1 &= g(r_0) = \frac{[0.552 (4.16380)^{1/2} + 1.1] (5)}{87 (0.001)^{1/2}} = 4.04622 \\
 r_2 &= g(r_1) = \frac{[0.552 (4.04622)^{1/2} + 1.1] (5)}{87 (0.001)^{1/2}} = 4.01711 \\
 r_3 &= 4.16380 - \frac{(4.04622 - 4.16380)^2}{4.01711 - 2(4.04622) + 4.16380} \\
 &= 4.00753
 \end{aligned}$$

Dado que la sucesión es convergente y que se trata del radio de un canal abierto, donde la exactitud después del primer decimal no es necesaria, se toma como valor a  $r = 4$  pies.

Los cálculos pueden realizarse con Matlab o con la TI-92 Plus.



```

m=1.1; e=0.001; v=5;
r0=1;
for i=1:4
    r1=(0.552*r0^0.5+m)*v/(87*e^0.5);
    r2=(0.552*r1^0.5+m)*v/(87*e^0.5);
    r=r0-(r1-r0)^2/(r2-2*r1+r0);
    fprintf('%2d %8.5f %8.5f %8.5f\n', i, r1, r2, r);
    r0=r;
end

```





```

P2_2( )
Prgm
  1.1 → m : .001 → e : 5. → v : 1. → r0 : ClrIO
  For i, 1, 4
    (.552*√(r0) + m)*v / (87*√(e)) → x
    (.552*√(x) + m)*v / (87*√(e)) → y
    r0 ← (x-r0)^2 / (y-2*x+r0) → r
    format(i, "f0") & " " & format(x, "f5") → d
    d & " " & format(y, "f5") & " " & format(r, "f5") → d
    Disp d
    r → r0
  EndFor
EndPrgm

```

**2.3** La siguiente fórmula es atribuida a Francis\* y se aplica a un vertedor con contracciones

$$Q = 3.33(B - 0.2 H)(H^3)^{1/2},$$

donde:

$Q$  = cantidad de agua que pasa por el vertedor en *pies*<sup>3</sup>/s

$B$  = ancho del vertedor en *pies*

$H$  = carga sobre la cresta del vertedor en *pies*

Si se sabe que  $B$  varía de 0 a 5 y  $Q$  de 0 a 33, calcule los valores de  $H$  correspondientes a las siguientes parejas de valores de  $B$  y  $Q$  (las unidades son consistentes), con el método de Newton-Raphson.

|     |    |    |    |     |
|-----|----|----|----|-----|
| $B$ | 3  | 2  | 4  | 3.6 |
| $Q$ | 12 | 20 | 13 | 30  |

**Solución** Se escribe la ecuación en la forma



$$f(H) = 3.33 (B - 0.2 H) (H^3)^{1/2} - Q = 0$$

Se deriva

$$f'(H) = 3.33 (B - 0.2 H) (1.5) H^{1/2} + (H^3)^{1/2} (3.33) (-0.2)$$

y sustituyendo en la fórmula de Newton-Raphson a  $f(H)$  y  $f'(H)$ , se tiene:

$$H_{i+1} = H_i - \frac{3.33(B - 0.2H_i) (H_i^3)^{1/2} - Q}{4.995(B - 0.2H_i) H_i^{1/2} - 0.666 (H_i^3)^{1/2}}$$

Para elegir un valor inicial de  $H$  en cada caso, se considera que por cuestión de diseño  $H$  debe ser menor que  $B$ . Por lo anterior, se sugiere utilizar como valor inicial  $H_0 = B/2$ .

Para la pareja  $B = 3$ ,  $Q = 12$

**Primera iteración**

$$H_0 = B/2 = 3/2 = 1.5$$

\* J. Lipka. Computaciones gráficas y mecánicas. CECSA (1972) pp. 139-141.

$$H_1 = 1.5 - \frac{3.33 [3 - 0.2 (1.5)] (1.5^3)^{1/2} - 12}{4.995[3 - 0.2 (1.5)] (1.5)^{1/2} - 0.666 (1.5^3)^{1/2}} = 1.2046$$

**Segunda iteración**

$$H_2 = 1.2046 - \frac{3.33 [3 - 0.2 (1.2046)] (1.2046^3)^{1/2} - 12}{4.995[3 - 0.02 (1.2046)] (1.2046)^{1/2} - 0.666 (1.2046^3)^{1/2}} = 1.1942$$

**Tercera iteración**

$$H_3 = 1.1942 - \frac{3.33 [3 - 0.2 (1.1942)] (1.1942^3)^{1/2} - 12}{4.995[3 - 0.2 (1.1942)] (1.1942)^{1/2} - 0.666 (1.1942^3)^{1/2}} = 1.1942$$

El método ha convergido al valor 1.1942 y se toma como carga sobre la cresta del vertedor 1.2 pies; las demás cifras significativas no interesan, por el sentido físico que tiene  $H$ .

Los resultados para las siguientes parejas de  $B$  y  $Q$  se dan a continuación:

| $B$ | 2   | 4   | 3.6 |
|-----|-----|-----|-----|
| $Q$ | 20  | 13  | 30  |
| $H$ | 2.5 | 1.0 | 2.0 |

Los cálculos pueden realizarse con Matlab o con la TI92-Plus.



```

B=3; Q=12;
H=B/2; Eps=0.01; dist=1;
While dist > Eps
    F=3.33*(B-0.2*H)*H^1.5-Q;
    DF=4.995*(B-0.2*H)*H^0.5-0.666*H^1.5;
    H1=H-F/DF;
    dist=abs (H1-H) ;
    disp([H1,dist] ) ;
    H=H1;
end

```



```

P2_3( )
Prgm
Define f(h) =3.33 (b-.2h) h^1.5-q
Define df(h) =4.995 (b-.2h) √(h) -.66h^1.5
3.→b : 12.→q : b/2→h : .01→eps : ClrIO
Loop
    h-f(h)/df(h)→h1
    abs(h1-h)→dist
    format(h, "f3") &" "&format(dist, "f5")→d
    Disp d
    h1→h
    If Dist<eps
        Exit
    EndLoop
EndPrgm

```

- 2.4 En la solución de problemas de valor inicial en ecuaciones diferenciales por transformadas de Laplace\* se presentan funciones racionales del tipo

$$F(s) = \frac{p_1(s)}{p_2(s)}$$

donde  $p_1$  y  $p_2$  son polinomios con: grado  $p_1 \leq$  grado  $p_2$ .

La expresión de  $F(s)$  en fracciones parciales es parte importante del proceso de solución y se realiza descomponiendo primero  $p_2(s)$  en sus factores más sencillos posibles.

En la solución de un problema de valor inicial\*\* (PVI) que modela un sistema de control lineal, la función de transferencia es (obtenida al aplicar la transformada de Laplace al PVI)

$$F(s) = \frac{C(s)}{R(s)} = \frac{24040(s+25)}{s^4 + 125s^3 + 5100s^2 + 65000s + 598800}$$

Para expresar en términos más sencillos a  $F(s)$  se resuelve primero la ecuación polinomial

$$s^4 + 125s^3 + 5100s^2 + 65000s + 598800 = 0$$

Con el método de Müller (programa 2.3 del disco), se tiene

$$s_1 = -6.6 + 11.4i$$

$$s_2 = -6.6 - 11.4i$$

$$s_3 = -55.9 + 18i$$

$$s_4 = -55.9 - 18i$$

Para obtener estos resultados se puede usar la siguiente instrucción de Matlab

```
roots ( [1 125 5100 65000 598800] )
```

Con lo que se obtiene:

```
ans = -55.8899 + 18.0260i
      -55.8899 - 18.0260i
      -6.6101 + 11.3992i
      -6.6101 - 11.3992i
```

y los factores buscados son:

$$(s + 6.6 - 11.4i)(s + 6.6 + 11.4i)(s + 55.9 - 18i)(s + 55.9 + 18i)$$

con lo que  $F(s)$  queda

$$F(s) = \frac{24040(s+25)}{F_1 F_2 F_3 F_4}$$

donde:

$$F_1 = s - s_1; \quad F_2 = s - s_2; \quad F_3 = s - s_3; \quad F_4 = s - s_4$$

\* Spiegel, Murray R., Applied Differential Equations. 2nd Ed Prentice Hall, Inc (1967), pp. 263-270.

\*\* Véase capítulo 7.



El segundo paso que completa la descomposición pedida es encontrar los valores de  $A_1, A_2, A_3, A_4$  que satisfagan la ecuación

$$F(s) = \frac{A_1}{F_1} + \frac{A_2}{F_2} + \frac{A_3}{F_3} + \frac{A_4}{F_4}$$

Esto se logra pasando el denominador de  $F(s)$  al lado derecho

$$24040(s + 25) = A_1 F_2 F_3 F_4 + A_2 F_1 F_3 F_4 + A_3 F_1 F_2 F_4 + A_4 F_1 F_2 F_3$$

y dando valores a  $s$ , por ejemplo  $s = s_1$ . Así

$$24040(-6.6 + 11.4i + 25) =$$

$$A_1(-6.6 + 11.4i + 6.6 + 11.4i)(-6.6 + 11.4i + 55.9 - 18i)(-6.6 + 11.4i + 55.9 + 18i),$$

ya que:  $A_2 F_1 F_3 F_4 = A_3 F_1 F_2 F_4 = A_4 F_1 F_2 F_3 = 0$ .

Al despejar  $A_1$  y realizar operaciones, se encuentra su valor

$$A_1 = 1.195 - 7.904i$$

Procediendo de igual manera, se calcula  $A_2, A_3$ , y  $A_4$  con  $s = s_2, s = s_3$  y  $s = s_4$ , respectivamente.

Esto se deja como ejercicio para el lector.

- 2.5** Una vez descompuesto  $F(s) = C(s) / R(s)$  en fracciones parciales (véase ejercicio anterior), se les aplica el proceso de “transformación” inversa de Laplace, que da como resultado la solución del problema de valor inicial.

Sea esta solución

$$F(t) = 1.21e^{-6.6t} \text{ sen}(11.4t - 111.7^\circ) + 0.28e^{-55.9t} \text{ sen}(18t + 26.1^\circ)$$

La solución obtenida debe analizarse matemáticamente e interpretarse físicamente si procede.

## BREVE ANÁLISIS CLÁSICO

Si  $t$  es el tiempo, el intervalo de interés es  $t > 0$ .

En los términos primero y segundo de  $F(t)$  aparece la función seno, que es oscilatoria, afectada de la función exponencial. Ésta tiende a cero cuando  $t$  tiene valores superiores a 1; se lleva tanto sus factores como la función  $F(t)$  a dicho valor, con lo cual la gráfica  $F(t)$  se confunde con el eje  $t$  para  $t \geq 1$ .

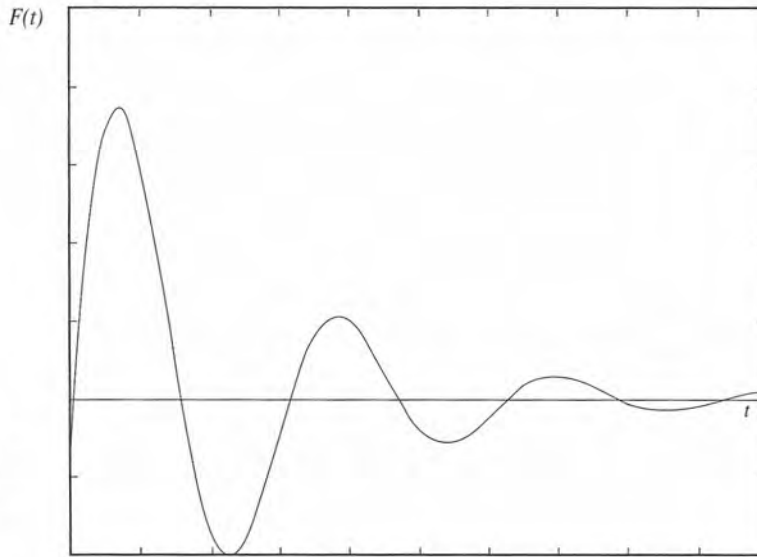
Estas funciones son conocidas como oscilatorias amortiguadas y sus gráficas son del tipo mostrado en la figura 2.14.

Si, por el contrario, el exponente de  $e$  es positivo, al tender  $t$  a infinito, la función es creciente y tiende rápidamente a infinito; lo cual se conoce como función oscilatoria no amortiguada.

Por otro lado, obsérvese que la contribución numérica del segundo término de  $F(t)$  es despreciable, y que el análisis y la gráfica de  $F(t)$  pueden obtenerse sin menoscabo de exactitud con el primer término.

Si se dan algunos valores particulares a  $t$  se obtiene:

| $t$    | 0.0    | 0.2   | 0.4   | 0.6    | 0.8   | 1.0                    |
|--------|--------|-------|-------|--------|-------|------------------------|
| $F(t)$ | -1.124 | 0.105 | 0.044 | -0.023 | 0.005 | $-4.22 \times 10^{-5}$ |



**Figura 2.14**  
Comportamiento  
de una función  
oscilatoria  
amortiguada.

Estos valores señalan claramente la presencia de raíces reales en los intervalos (0,0.2), (0.4,0.6), (0.6,0.8), (0.8,1.0): (véase Fig. 2.15). Utilizando como valores iniciales 0.1, 0.5, 0.7, y 0.9 y el método de Newton-Raphson se obtiene, respectivamente:

$$\bar{t}_1 = 0.171013; \quad \bar{t}_2 = 0.44659; \quad \bar{t}_3 = 0.72217; \quad \bar{t}_4 = 0.99775$$

Los posibles máximos y mínimos de esta función se consiguen resolviendo la ecuación que resulta de igualar con cero la primera derivada de  $F(t)$

$$F'(t) = 13.794e^{-6.6t} \cos(11.4t - 111.7^\circ) - 7.986e^{-6.6t} \operatorname{sen}(11.4t - 111.7^\circ) \\ + 5.04e^{-55.9t} \cos(18t + 26.1^\circ) - 15.652e^{-55.9t} \operatorname{sen}(18t + 26.1^\circ) = 0$$

Aprovechando las evaluaciones que se hicieron de  $F'(t)$  en el método de Newton-Raphson, se tiene:

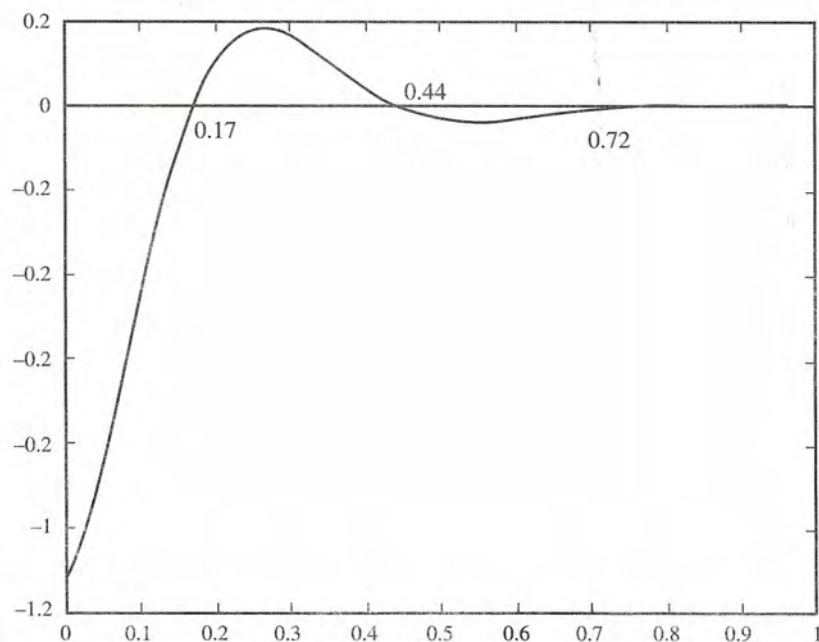
| $t$     | 0.0    | 0.1   | 0.3    | 0.6   | 0.9    | 1.0      |
|---------|--------|-------|--------|-------|--------|----------|
| $F'(t)$ | -0.040 | 7.849 | -0.900 | 0.196 | -0.035 | -0.00184 |

Con los valores iniciales dados a la izquierda, se obtuvieron las raíces anotadas a la derecha

$$\begin{array}{ll} t_0 = 0 & \bar{t}_1 = 0.00175 \\ t_0 = 0.2 & \bar{t}_2 = 0.26277 \\ t_0 = 0.45 & \bar{t}_3 = 0.53834 \\ t_0 = 0.75 & \bar{t}_4 = 0.81399 \end{array}$$

Con los valores de la función en diferentes puntos, sus raíces y puntos máximos y mínimos, la gráfica aproximada de  $F(t)$  se muestra en la figura 2.12

Este análisis se puede comprobar con el **PROGRAMA 2.2** del CD o el Matlab, por ejemplo.



**Figura 2.15**  
Gráfica de la  
función  $F(t)$ .



- 2.6** Determine la cantidad de vapor  $V$  (moles/hr) y la cantidad de líquido  $L$  (moles/hr) que se generan en una vaporización instantánea continua a una presión de 1600 psia y una temperatura de 120 °F de la siguiente mezcla.

| Componente                               | Composición $z_i$ | $K_i = y_i/x_i$ |
|------------------------------------------|-------------------|-----------------|
| CO <sub>2</sub>                          | 0.0046            | 1.65            |
| CH <sub>4</sub>                          | 0.8345            | 1.80            |
| C <sub>2</sub> H <sub>6</sub>            | 0.0381            | 0.94            |
| C <sub>3</sub> H <sub>8</sub>            | 0.0163            | 0.55            |
| <i>i</i> -C <sub>4</sub> H <sub>10</sub> | 0.0050            | 0.40            |
| <i>n</i> -C <sub>4</sub> H <sub>10</sub> | 0.0074            | 0.38            |
| C <sub>5</sub> H <sub>12</sub>           | 0.0287            | 0.22            |
| C <sub>6</sub> H <sub>14</sub>           | 0.0220            | 0.14            |
| C <sub>7</sub> H <sub>16</sub>           | 0.0434            | 0.09            |

**Solución** Con base en la figura 2.16



Un balance total de materia da  $F = L + V$

(1)

Un balance de materia para cada componente da:

$$Fz_i = Lx_i + Vy_i \quad i = 1, 2, \dots, n$$

(2)

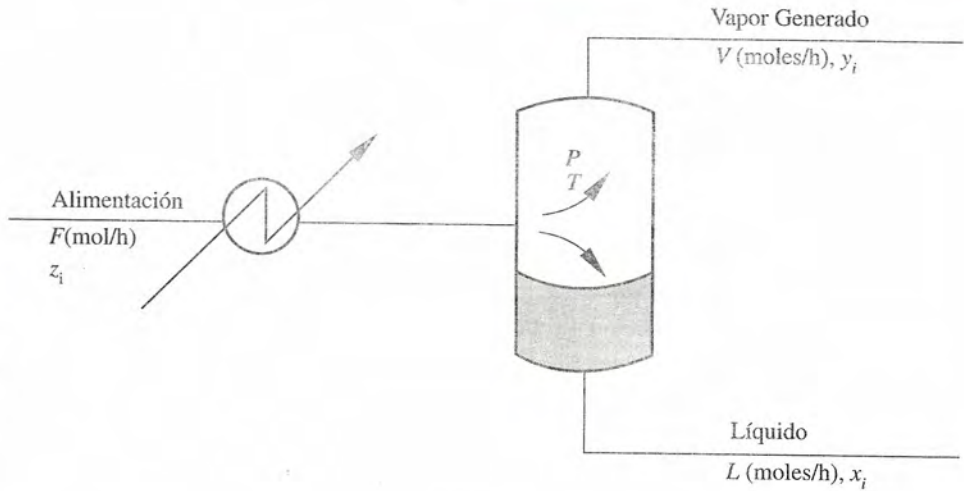
Las relaciones de equilibrio líquido-vapor establecen

$$K_i = \frac{y_i}{x_i} \quad i = 1, 2, \dots, n$$

(3)



**Figura 2.16**  
Esquema de una vaporización instantánea (*flash*) de una mezcla multicomponente.



Sustituyendo la ecuación 3 en la 2 se obtiene

$$Fz_i = Lx_i + VK_i x_i \quad i = 1, 2, \dots, n \quad (4)$$

o bien

$$Fz_i = x_i (L + VK_i) \quad i = 1, 2, \dots, n$$

de donde

$$x_i = \frac{Fz_i}{L + VK_i} \quad i = 1, 2, \dots, n$$

Sustituyendo la ecuación 1 en esta última se obtiene:

$$x_i = \frac{Fz_i}{F + V(K_i - 1)} \quad i = 1, 2, \dots, n \quad (5)$$

Las restricciones de composición establecen

$$\sum_{i=1}^n x_i = 1 \quad ; \quad \sum_{i=1}^n y_i = 1$$

Por lo que puede escribirse:

$$\sum_{i=1}^n y_i - \sum_{i=1}^n x_i = 0$$

o bien:

$$\sum_{i=1}^n K_i x_i - \sum_{i=1}^n x_i = 0$$

o simplemente:

$$\sum_{i=1}^n x_i (K_i - 1) = 0 \quad (6)$$

sustituyendo la ecuación 5 en la 6 se obtiene:

$$\sum_{i=1}^n \frac{Fz_i (K_i - 1)}{F + V(K_i - 1)} = 0 \quad (7)$$

**VALORES INICIALES**

El valor de  $V$  que satisface la ecuación 7 está comprendido en el intervalo  $0 \leq V \leq F$ , por lo que la estimación de un valor inicial de  $V$  es difícil, ya que el valor de  $F$  puede ser muy grande. Esta dificultad se reduce normalizando el valor de  $V$ ; esto es, dividiendo numerador y denominador de la ecuación 7 entre  $F$ , para obtener

$$\sum_{i=1}^n \frac{z_i (K_i - 1)}{1 + \psi (K_i - 1)} \quad (8)$$

donde  $\psi = V/F$

La ecuación 8 equivale a la 7, pero expresada en la nueva variable  $\psi$  cuyos límites son:

$$0 \leq \psi \leq 1$$

La ecuación 8 es no lineal en una sola variable ( $\psi$ ), que se resolverá con el método de Newton-Raphson. Hay que observar que esta ecuación es monotónica decreciente, por lo que el valor inicial puede ser cualquier número dentro del intervalo  $[0, 1]$ , por ejemplo  $\psi_0 = 0$ .

El PROGRAMA 2.4 del CD emplea  $\psi_0 = 0$  como estimado inicial y

$$f'(\psi) = \sum_{i=1}^n \frac{-z_i (K_i - 1)^2}{[1 + \psi (K_i - 1)]^2}$$

A continuación se muestran los valores que adquiere  $\psi$  y  $f(\psi)$  a lo largo de las iteraciones realizadas.

| Iteración | $\psi$    | $f(\psi)$              |
|-----------|-----------|------------------------|
| 1         | 0.9328799 | $-8.79 \times 10^{-2}$ |
| 2         | 0.8968149 | $-1.29 \times 10^{-2}$ |
| 3         | 0.8895657 | $-3.5 \times 10^{-4}$  |
| 4         | 0.8893582 | $-2.68 \times 10^{-7}$ |
| 5         | 0.8893580 | $-1.5 \times 10^{-13}$ |

**RESULTADOS**

Para  $F = 1$  moles/h

Vapor generado:  $V = 0.889358$  moles/h

Líquido generado:  $L = 0.110642$  moles/h

Composiciones del líquido y del vapor generados:

| Componente                               | Líquido ( $x_i$ ) | Vapor ( $y_i$ ) |
|------------------------------------------|-------------------|-----------------|
| CO <sub>2</sub>                          | 0.00291           | 0.00481         |
| CH <sub>4</sub>                          | 0.48759           | 0.87766         |
| C <sub>2</sub> H <sub>6</sub>            | 0.04025           | 0.03783         |
| C <sub>3</sub> H <sub>8</sub>            | 0.02718           | 0.01495         |
| <i>i</i> -C <sub>4</sub> H <sub>10</sub> | 0.01072           | 0.00429         |
| <i>n</i> -C <sub>4</sub> H <sub>10</sub> | 0.01650           | 0.00627         |
| C <sub>5</sub> H <sub>12</sub>           | 0.09370           | 0.02061         |
| C <sub>6</sub> H <sub>14</sub>           | 0.09356           | 0.01310         |
| C <sub>7</sub> H <sub>16</sub>           | 0.22760           | 0.02048         |

Los cálculos de este problema pueden realizarse con Matlab o con la TI-92 Plus.



```
Z=[0.0046 0.8345 0.0381 0.0163 0.0050...
    0.0074 0.0287 0.0220 0.0434];
K=[1.65 1.80 0.94 0.55 0.40...
    0.38 0.22 0.14 0.09];
Eps=1e-3; f=1; Fi=0; i =0;
fprintf(' Fi f(Fi)\n')
while and(abs(f)>Eps,i<10)
    f=sum((Z.*(K-1))./(1+Fi*(K-1)));
    df=sum((-Z.*(K-1).^2)./(1+Fi*(K-1)).^2);
    Fi=Fi-f/df;
    fprintf('%10.6f %8.2e\n',Fi,f)
    Fi=Fi; i=i+1;
end
fprintf('Fi= %10.6f\n',Fi)
fprintf('x(i) y(i)\n')
for i=1:9
    X(i)=Z(i)/(1+Fi*(K(i)-1));
    Y(i)=K(i)*X(i);
    fprintf('%10.5f %10.5f\n',X(i),Y(i))
end
```



```
p2_6( )
Prgm
.0046→z[1] : .8345→z[2] : .0381→z[3]
.0163→z[4] : .0050→z[5] : .0074→z[6]
.0287→z[7] : .0220→z[8] : .0434→z[9]
1.65→k[1] : 1.8→k[2] : .94→k[3]
.55→k[4] : .40→k[5] : .38→k[6]
.22→k[7] : .14→k[8] : .09→k[9]
.001→eps : 1→f : 0→fi : 0→i: ClrIO
While abs(f)>eps or i<10
    0→f : 0→d
    For i,1,9
        f+z[i]*(k[i]-1)/(1+fi*(k[i]-1))→f
        d-z[i]*(k[i]-1)^2/(1+fi*(k[i]-1))^2→d
    EndFor
    fi=f/d→fi
    Disp fi
EndWhile
EndPrgm
```

También puede utilizarse la función *fzero* de Matlab:

Con su editor de texto escriba el siguiente guión y grábalo con el nombre *Flash.m* en el directorio de trabajo de Matlab:



```
function f=Flash(Fi)
Z=[0.0046 0.8345 0.0381 0.0163 0.0050...
    0.0074 0.0287 0.0220 0.0434];
K=[1.65 1.80 0.94 0.55 0.40...
    0.38 0.22 0.14 0.09];
f=sum((Z.*(K-1))./(1+Fi*(K-1)));
```

Ahora use el gui3n dado enseguida para resolver el problema:

```
Z=[0.0046 0.8345 0.0381 0.0163 0.0050...
    0.0074 0.0287 0.0220 0.0434];
K=[1.65 1.80 0.94 0.55 0.40...
    0.38 0.22 0.14 0.09];
Fi=fzero('Flash',0)
for i=1:9
X(i)=Z(i)/(1+Fi*(K(i)-1));
Y(i)=K(i)*X(i);
fprintf('%10.5f %10.5f\n',X(i),Y(i))
end
```



- 2.7 Considere un líquido en equilibrio con su vapor. Si el líquido está formado por los componentes 1, 2, 3 y 4, con los datos dados a continuación calcule la temperatura y la composición del vapor en el equilibrio a la presión total de 75 psia.

| Componente | Composición del líquido % mol | Presión de vapor de componente puro (psia) |          |
|------------|-------------------------------|--------------------------------------------|----------|
|            |                               | a 150 °F                                   | a 200 °F |
| 1          | 10.0                          | 25.0                                       | 200.0    |
| 2          | 54.0                          | 14.7                                       | 60.0     |
| 3          | 30.0                          | 4.0                                        | 14.7     |
| 4          | 6.0                           | 0.5                                        | 5.0      |

Utilice la siguiente ecuación para la presión de vapor:

$$\ln(p_i^0) = A_i + B_i/T; \quad i = 1, 2, 3, 4; \quad T \text{ en } ^\circ R$$

### Solución

La presión total del sistema será:  $P_T = \sum_{i=1}^n P_i$  (1)



Si se considera que la mezcla de estos cuatro componentes, a las condiciones de presión y temperatura de este sistema, obedece las leyes de Raoult y de Dalton

$$P_T = \sum_{i=1}^4 p_i^0 x_i \quad (2)$$

donde:

$p_i^0$  = presión de vapor de cada componente.

$P_T$  = presión total del sistema.

$P_i$  = presión parcial de cada componente

$x_i$  = fracción mol de cada componente en el líquido

De la ecuación de presión de vapor se tiene que

$$p_i^0 = \exp (A_i + B_i / T) \quad i = 1, 2, 3, 4 \quad (3)$$

de las ecuaciones 1 y 2 resulta

$$P_T = \sum_{i=1}^4 x_i \exp (A_i + B_i / T) \quad (4)$$

de donde puede establecerse

$$f(T) = P_T - \sum_{i=1}^4 x_i \exp (A_i + B_i / T) = 0 \quad (5)$$

$A_i$  y  $B_i$  pueden obtenerse como sigue

Si se hace  $p_{1,i}^0$  = presión de vapor del componente  $i$  a  $T_1 = 150^\circ \text{F} = 609.56^\circ \text{R}$

$p_{2,i}^0$  = presión de vapor del componente  $i$  a  $T_2 = 200^\circ \text{F} = 659.56^\circ \text{R}$

entonces

$$\ln (p_{1,i}^0) = A_i + B_i / T_1 \quad i = 1, 2, 3, 4 \quad (6)$$

y

$$\ln (p_{2,i}^0) = A_i + B_i / T_2 \quad i = 1, 2, 3, 4 \quad (7)$$

restando la ecuación 7 de la 6 se tiene

$$\ln \left( \frac{p_{1,i}^0}{p_{2,i}^0} \right) = B_i (1 / T_1 - 1 / T_2)$$

de donde

$$B_i = \frac{\ln \left( \frac{p_{1,i}^0}{p_{2,i}^0} \right)}{\frac{1}{T_1} - \frac{1}{T_2}} \quad (8)$$

Conociendo  $B_i$  se puede obtener  $A_i$  de la ecuación 6

$$A_i = \ln (p_{1,i}^0) - B_i / T_1 \quad i = 1, 2, 3, 4 \quad (9)$$

### VALORES INICIALES

Para estimar un valor inicial de  $T$  para resolver la ecuación 5, se considera el componente dominante de la mezcla, en este caso el componente 2, y se usa  $P_T$  en lugar de  $p_2^0$  en la ecuación de presión de vapor

$$\ln (P_T) = A_2 + B_2 / T$$

de donde

$$T = \frac{B_2}{\ln (P_T) - A_2} \quad (10)$$

Con este resultado inicial y las consideraciones ya anotadas, el **PROGRAMA 2.5** del CD, utiliza el método de Newton-Raphson con

$$f'(T) = - \sum_{i=1}^4 x_i \exp(A_i + B_i/T) (-B_i/T^2) \quad (11)$$

y reporta los siguientes resultados después de cuatro iteraciones

Temperatura del sistema = 209.07 °F = 668.63 °R

(temperatura de burbuja)

Composición del vapor en equilibrio

| Componente (i) | $y_i$  |
|----------------|--------|
| 1              | 0.3761 |
| 2              | 0.5451 |
| 3              | 0.0729 |
| 4              | 0.0059 |

Los cálculos de este problema pueden realizarse con el siguiente guión de Matlab:

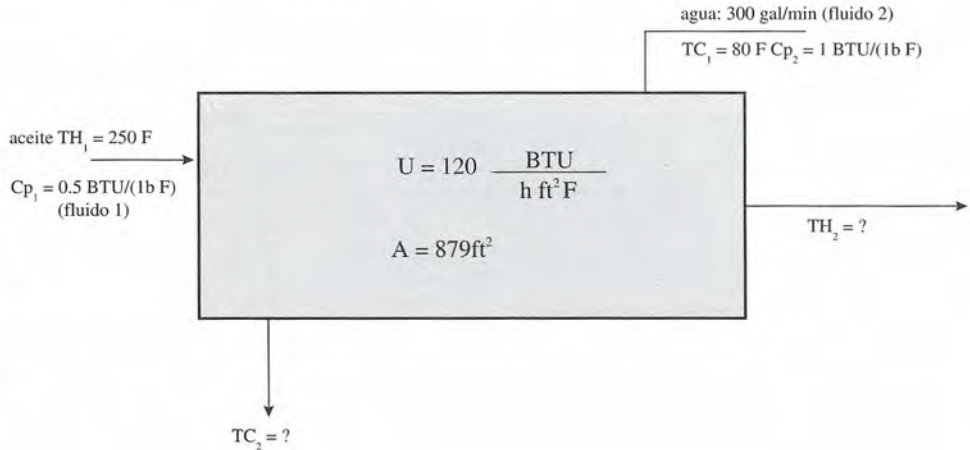


```
P1=[25 14.7 4 0.5];
P2=[200 60 14.7 5];
T1=150+459.56;
T2=200+459.56;
B=log(P1./P2)/(1/T1-1/T2);
A=log(P1)-B/T1;
X=[0.10 0.54 0.30 0.06];
PT=75; i=0; f=1; Eps=0.000001;
T=B(2)/(log(PT)-A(2));
fprintf(' T f(T)\n',T,f)
while and(abs(f)>Eps,i<10)
    f=PT-sum(X.*exp(A+B/T));
    df=sum(X.*exp(A+B/T).*(B/T^2));
    T1=T-f/df;
    fprintf('%10.2f %8.2e\n',T,f)
    T=T1;
    i=i+1;
end
fprintf(' y(i)\n')
for i=1:4
    Y(i)=(X(i)*exp(A(i)+B(i)/T))/PT;
    fprintf('%10.4f \n',Y(i))
end
```



- 2.8** Se emplea un intercambiador de calor (Fig. 2.17) para enfriar aceite. Encuentre la temperatura de salida del aceite y del agua enfriadora ( $TH_2$  y  $TC_2$ , respectivamente), para gastos de aceite de 105,000; 80,000; 50,000; 30,000 y 14,000 lbm/h.





**Figura 2.17**  
Esquema de un  
intercambiador  
de calor con  
flujo a  
contracorriente.

**Solución** Un balance de calor para el aceite da:

$$Q_1 = w_1 C_{p1} (TH_1 - TH_2) \quad (1)$$

Un balance de calor para el agua da:

$$Q_2 = w_2 C_{p2} (TC_2 - TC_1) \quad (2)$$

La ecuación que rige la transferencia de calor a través de este equipo es:

$$Q = U A \Delta Tm \quad (3)$$

donde:

$U$  = coeficiente global de transferencia de calor

$A$  = área total de transferencia de calor

$$\Delta Tm = \frac{(TH_1 - TC_2) - (TH_2 - TC_1)}{\ln \left( \frac{TH_1 - TC_2}{TH_2 - TC_1} \right)} \quad (4)$$

Para encontrar  $TH_2$  y  $TC_2$  debe cumplirse que  $Q_1 = Q_2 = Q$ , o bien:

$$\frac{Q}{Q_1} - 1 = 0 \quad (5)$$

Pero  $Q$  sólo podrá calcularse cuando se conozcan todas las temperaturas. Para resolver este problema se propone el siguiente procedimiento.

Establecer que  $TH_2$  sea la única variable; entonces,  $Q_2$  puede escribirse en función de  $TH_2$  como sigue

$$Q_2 = Q_1 = w_1 C_{p1} (TH_1 - TH_2) = w_2 C_{p2} (TC_2 - TC_1) \quad (6)$$

de donde puede despejarse  $TC_2$

$$TC_2 = \frac{w_1 C_{p1}}{w_2 C_{p2}} (TH_1 - TH_2) + TC_1 \quad (7)$$

Con todo esto ya puede establecerse  $Q$  en función de  $TH_2$ , y así escribir la ecuación 5 también en función de dicha variable única:

$$f(TH_2) = \frac{UA \left( TH_1 - \frac{w_1 C_{p1}}{w_2 C_{p2}} (TH_1 - TH_2) - TC_1 \right) - (TH_2 - TC_1)}{\ln \left( \frac{(TH_1 - \frac{w_1 C_{p1}}{w_2 C_{p2}} (TH_1 - TH_2) - TC_1)}{TH_2 - TC_1} \right)} - 1 = 0 \quad (8)$$

### VALORES INICIALES

Para estimar un valor inicial de  $TH_2$  cabe apoyarse en la figura 2.18, la cual muestra una gráfica de temperaturas en este tipo de intercambiadores de calor.

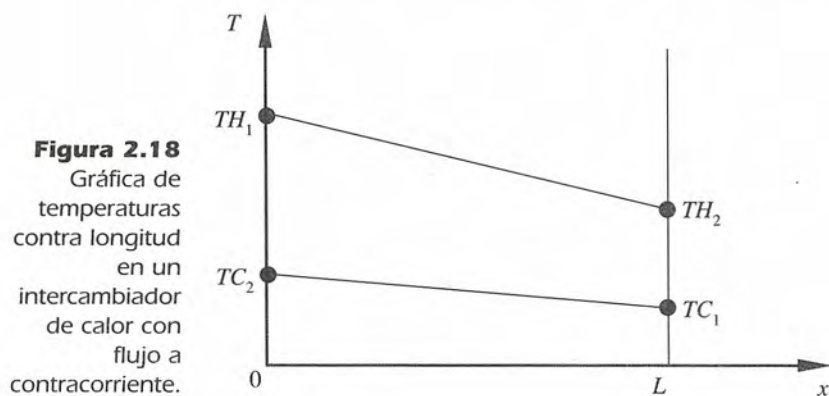
De acuerdo con esta gráfica, se tienen las siguientes restricciones

$$TC_1 < TH_2 < TH_1$$

y

$$TC_1 < TC_2 < TH_1$$

Como en este caso no se dispone de mayor información, el **PROGRAMA 2.6** del CD emplea el método de la bisección con  $TH_{21} = TC_1 + 0.5$  y  $TH_{2D} = TH_1 - 0.5$  para resolver la ecuación 8.



Para un gasto de aceite de 105,000 lbm/h

### RESULTADOS

$$TH_2 = 113$$

$$TC_2 = 128$$

Los cálculos de este problema pueden realizarse con el siguiente guión de Matlab:



```

w1=105000; w2=300*60*2.2*3.785;
Cp1=0.5; Cp2=1; U=120; A=879;
TH1=250; TC1=80; Eps=0.0001;
TH2i=TC1+0.5; TH2d=TH1-0.5; fm=1;
Q1=w1*Cp1*(TH1-TH2i); TC2=Q1/(w2*Cp2)+TC1;
DTm=((TH1-TC2)-(TH2i-TC1))/log((TH1-TC2)/(TH2i-TC1));
fi=U*A*DTm/Q1-1;
Q1=w1*Cp1*(TH1-TH2d); TC2=Q1/(w2*Cp2)+TC1;
DTm=((TH1-TC2)-(TH2d-TC1))/log((TH1-TC2)/(TH2d-TC1));
fd=U*A*DTm/Q1-1;
if fi*fd<0
    while abs(fm) > Eps
        TH2m=(TH2i+TH2d)/2;
        Q1=w1*Cp1*(TH1-TH2m); TC2=Q1/(w2*Cp2)+TC1;
        DTm=((TH1-TC2)-(TH2m-TC1))/log((TH1-TC2)/(TH2m-TC1));
        fm=U*A*DTm/Q1-1;
        if fi*fm < 0
            TH2d=TH2m; fd=fm;
        else
            TH2i=TH2m; fi=fm;
        end
    end
else
    disp('TH2i y TH2d no encierran una raíz')
break
end
TH2=(TH2i+TH2d)/2;
Q1=w1*Cp1*(TH1-TH2); TC2=Q1/(w2*Cp2)+TC1;
fprintf('TH2= %8.2f TC2= %8.2f\n',TH2,TC2)

```

**2.9** El siguiente circuito representa en forma muy simplificada un generador de impulsos para probar el aislamiento de un transformador en circuito abierto.

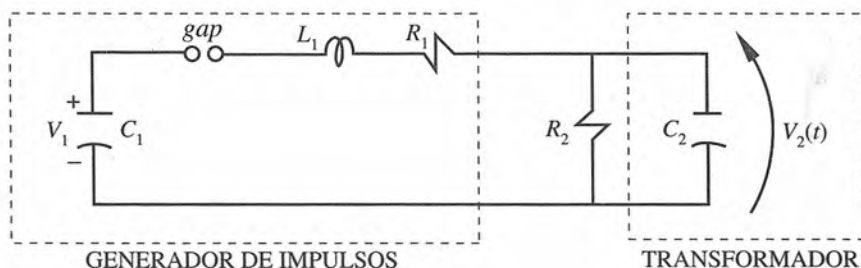
Considérese el *gap* como un interruptor.

Las condiciones iniciales en el transformador y la inductancia son cero. Use los siguientes datos para encontrar  $v_2(t)$ :

$$C_1 = 12.5 \times 10^{-9} \text{ fd}, \quad C_2 = 0.3 \times 10^{-9} \text{ fd}, \quad L_1 = 0.25 \times 10^{-3} \text{ Hy}$$

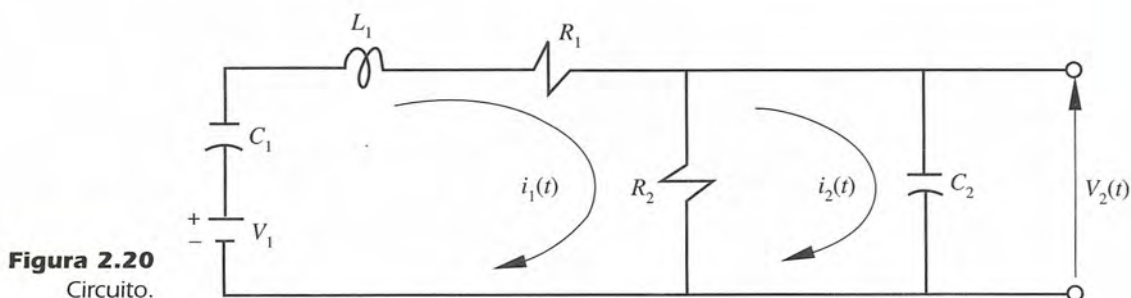
$$R_1 = 2 \text{ Kohms}, \quad R_2 = 3 \text{ Kohms}, \quad V_1 = 300 \text{ Kv}$$

**Figura 2.19**  
Circuito  
representativo  
de un  
generador de  
impulsos.





**Solución** Estableciendo las ecuaciones para el circuito.



$$v_1(t) = (R_1 + R_2)i_1(t) + L_1 \frac{di_1(t)}{dt} + \frac{1}{C_1} \int i_1(t) dt - R_2 i_2(t) \quad (1)$$

$$0 = -R_2 i_1(t) + R_2 i_2(t) + \frac{1}{C_2} \int i_2(t) dt \quad (2)$$

$$v_2(t) = \frac{1}{C_2} \int i_2(t) dt \quad (3)$$

Aplicando la transformada de Laplace y considerando que las condiciones iniciales son cero, se tiene:

$$\frac{V_1}{s} = (R_1 + R_2)I_1(s) + L_1 s I_1(s) + \frac{1}{C_1} \frac{I_1(s)}{s} - R_2 I_2(s) \quad (4)$$

$$0 = -R_2 I_1(s) + R_2 I_2(s) + \frac{1}{C_2} \frac{I_2(s)}{s} \quad (5)$$

Despejando  $I_1(s)$  de la ecuación 5 y sustituyendo en la ecuación cuatro se tiene:

$$I_2(s) = \frac{V_1}{(R_1 + L_1 s + 1/C_1 s)(s + 1/R_2 C_2) + 1/C_2} \quad (6)$$

al aplicar la transformada inversa de Laplace a la ecuación 3 y recordando que las condiciones iniciales son cero

$$V_2(s) = \frac{1}{C_2} \frac{I_2(s)}{s} \quad (7)$$

Se sustituye la ecuación 6 en la 7

$$V_2(s) = \frac{\frac{V_1}{C_2}}{s \left( (R_1 + L_1 s + \frac{1}{C_1 s}) (s + \frac{1}{R_2 C_2}) + \frac{1}{C_2} \right)}$$

y simplificando se llega a:

$$V_2(s) = \frac{V}{s^3 + P_1 s^2 + P_2 s + P_3} \quad (8)$$

con

$$P_1 = \frac{R_1 R_2 C_2 + L_1}{R_2 C_2 L_1} = 9.1111 \times 10^6$$

$$P_2 = \frac{R_1 C_1 + R_2 C_2 + R_2 C_1}{R_2 C_1 C_2 L_1} = 22.5422 \times 10^{12}$$

$$P_3 = \frac{1}{R_2 C_1 C_2 L_1} = 355.556 \times 10^{15}$$

$$V = \frac{V_1}{C_2 L_1} = \frac{V_1}{75 \times 10^{-15}}$$

La ecuación 8 puede escribirse:

$$V_2(s) = \frac{V}{(s+a)(s+b)(s+c)}$$

cuya transformada inversa de Laplace es

$$v_2(t) = V \left( \frac{e^{-at}}{(b-a)(c-a)} + \frac{e^{-bt}}{(c-b)(a-b)} + \frac{e^{-ct}}{(a-c)(b-c)} \right) \quad (9)$$

donde  $a$ ,  $b$  y  $c$  son las raíces de la ecuación

$$s^3 + P_1 s^2 + P_2 s + P_3 = 0$$

La primera raíz, obtenida con el programa 2.3 del apéndice, es

$$a = -1.5874547 \times 10^4$$

Se reduce el grado del polinomio y aplicando la fórmula cuadrática, se tiene

$$b = -4.547618 \times 10^6 + 1.310346 \times 10^6 i$$

$$c = -4.547618 \times 10^6 - 1.310346 \times 10^6 i$$

Recuerde que puede utilizar la función *roots* de Matlab.

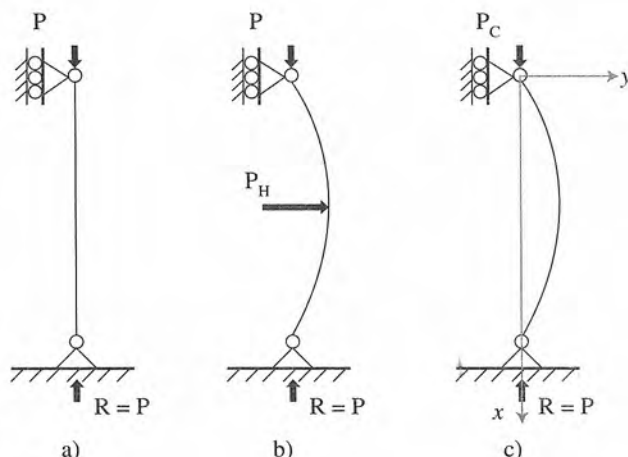
Estos valores se sustituyen en la ecuación 9 y se tiene:

$$v_2(t) = 300 \left( 0.6e^{-1.5874547 \times 10^4 t} - e^{-4.547618 \times 10^6 t} \left[ 0.6 \cos(1.310346 \times 10^6 t) + 2.072102 \sin(1.310346 \times 10^6 t) \right] \right)$$

donde  $t$  está en segundos y  $v_2(t)$  en *Kvolts*.

- 2.10** Se tiene una columna articulada en ambos extremos (véase Fig. 2.21a). Aplicando una carga vertical pequeña  $P$  (de modo que no se pandee), se obtiene una reacción  $R$  de igual magnitud y de sentido contrario en la base. Si ahora se aplica una carga horizontal  $P_H$  se obtiene un pandeo infinitesimal (imperceptible a la vista), que se ha magnificado en la figura 2.21b, con fines de ilustración. Si se empieza a “jugar” aumentando  $P$  y disminuyendo  $P_H$ , de modo que se mantenga el mismo pandeo en la columna, va a llegar un momento

**Figura 2.21**  
Columna  
articulada.  $R$  es  
la reacción de  
la carga.



en que  $P_H$  valga cero y el valor de  $P$  correspondiente será llamado la carga crítica de pandeo  $P_C$ . Para obtener esta carga crítica se hace un análisis de la estabilidad de la columna, usando la proposición de Jacobo Bernoulli: la curvatura producida en una viga debida a la flexión es directamente proporcional al momento flexionante e inversamente proporcional a la rigidez, es decir:

$$\kappa = \frac{M}{EI},$$

donde  $\kappa$  es la curvatura,  $M$  el momento flexionante,  $E$  es el módulo de elasticidad del material e  $I$  el momento de inercia que depende de la forma de la sección transversal de la columna. El producto  $EI$  se conoce como la rigidez a la flexión de la columna. Por otro lado

$$\kappa = \frac{\frac{d^2y}{dx^2}}{\sqrt{1 + \left(\frac{dy}{dx}\right)^2}}$$

y ya que el pandeo es infinitesimal  $\frac{dy}{dx}$  (las pendientes de las tangentes a la curva elástica) son muy pequeñas, despreciándose y quedando entonces la curvatura aproximada por

$$\kappa \approx \frac{d^2y}{dx^2}$$

De igual modo se tiene que  $M = -Py$ , donde el signo es convencional.

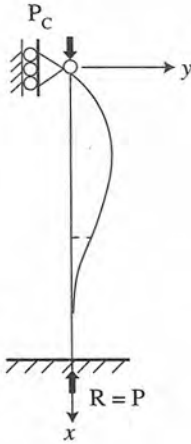
Sustituyendo se tiene

$$\frac{d^2y}{dx^2} = -\frac{Py}{EI}, \text{ o } \frac{d^2y}{dx^2} + \frac{Py}{EI} = 0$$

Haciendo  $\lambda = \frac{P}{EI}$  y observando que el desplazamiento  $y$  de la columna en ambos extremos es nulo se tiene el siguiente problema de valores en la frontera:

$$\begin{aligned} \frac{d^2y}{dx^2} + \lambda^2 y &= 0 \\ y(0) &= 0 \\ y(L) &= 0 \end{aligned}$$



**Figura 2.22**

Columna articulada y empotrada en la base.

cuya solución analítica da lugar a

$$P = \frac{\pi^2 EI}{L^2}$$

Si ahora se tiene una columna articulada por arriba y empotrada en el piso (véase Fig. 2.22) y se quiere conocer la carga crítica de pandeo correspondiente, el análisis de la estabilidad de la columna conduce al problema de valores en la frontera siguiente:

$$\frac{d^2 y}{dx^2} + \lambda^2 y = \frac{H_A}{EI} x$$

$$y(0) = 0$$

$$y(L) = 0$$

$$y'(L) = 0$$

La solución analítica de este tipo de problemas produce, en pasos intermedios, ecuaciones no lineales en una incógnita. Así, para nuestro problema, se tiene:

$$\lambda L = \tan \lambda L$$

que habrá que resolver para encontrar  $y$  en función de  $x$ .

**Solución** Con el objeto de simplificar, haremos,  $x = \lambda L$  con lo que la ecuación anterior queda:

$$\tan x = x$$

Resolviendo con el método de Newton-Raphson se obtiene  $x = 4.493409$ .



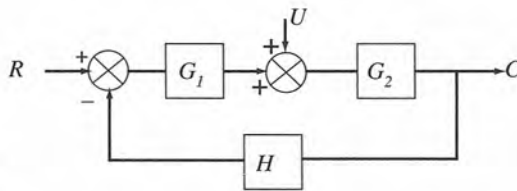
**2.11** La respuesta de un sistema de control de retroalimentación simple, mostrado en la figura 2.23, está dada por la expresión\*

$$C = \frac{G_1 G_2}{1 + G} R + \frac{G_2}{1 + G} U,$$

donde  $G = G_1 G_2 H$ . Cuando el factor del denominador,  $1 + G$ , se iguala a cero, se obtiene la ecuación característica del sistema de lazo cerrado. Las raíces de la ecuación característica determinan la forma o tipo de la respuesta  $C(t)$  a cualquier función forzante particular  $R(t)$  o  $U(t)$ .

\* Cughanowr, *Process Systems Analysis and Control*, Second Edition, McGraw Hill International Editions.

**Figura 2.23**  
Sistema  
de control de  
retroalimentación  
simple.



El método de lugar geométrico de las raíces es un procedimiento gráfico para encontrar las raíces de  $1 + G = 0$ , cuando uno de los parámetros de  $G$  varía continuamente. En este caso el parámetro que variará es la ganancia (o sensibilidad)  $K_C$  del controlador. En el diagrama de bloques de la figura 2.23

$$G_1 = K_C$$

$$G_2 = \frac{1}{(\tau_1 s + 1)(\tau_2 s + 1)}$$

$$H = \frac{1}{\tau_3 s + 1}$$

Para este caso, la función de transferencia de lazo abierto es:

$$G = \frac{K_c}{(\tau_1 s + 1)(\tau_2 s + 1)(\tau_3 s + 1)}$$

que puede escribirse en la forma

$$G(s) = \frac{K}{(s - p_1)(s - p_2)(s - p_3)}$$

$$\text{donde } K = \frac{K_c}{\tau_1 \tau_2 \tau_3}, p_1 = -\frac{1}{\tau_1}, p_2 = -\frac{1}{\tau_2}, p_3 = -\frac{1}{\tau_3}$$

Se llama a los términos  $p_1, p_2$  y  $p_3$  los polos de la función de transferencia de lazo abierto. Un polo de  $G(s)$  es cualquier valor de  $s$  para el cual  $G(s)$  es infinito. Por tanto  $p_1 = -1/\tau_1$  es un polo de  $G(s)$ .

La ecuación característica del sistema de lazo cerrado es:

$$1 + \frac{K}{(s - p_1)(s - p_2)(s - p_3)} = 0$$

Esta expresión puede escribirse:

$$(s - p_1)(s - p_2)(s - p_3) + K = 0$$

Si por ejemplo, los polos fueran  $-1, -2$  y  $-3$ , respectivamente, tendríamos:

$$(s + 1)(s + 2)(s + 3) + K = 0$$

donde  $K = 6K_c$ .

Expandiendo el producto de esta ecuación resulta:

$$s^3 + 6s^2 + 11s + (K + 6) = 0,$$

| K     | Raíz real | a        | b        |
|-------|-----------|----------|----------|
| 0.000 | -1.00000  | -2.00000 | -3.00000 |
| 0.100 | -1.05435  | -1.89897 | -3.04668 |
| 0.200 | -1.12111  | -1.79085 | -3.08803 |
| 0.300 | -1.21352  | -1.66106 | -3.12542 |
| 0.400 | -3.15970  | -1.42015 | -0.09320 |
| 0.500 | -3.40140  | -1.40426 | -0.05440 |

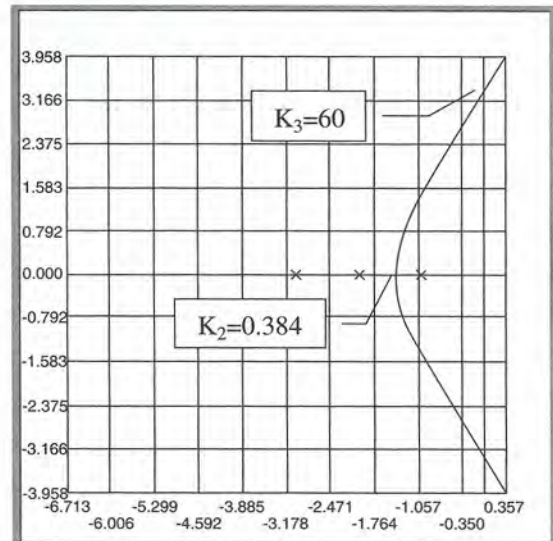
  

| K     | Raíz real | a        | b        |
|-------|-----------|----------|----------|
| 0.381 | -1.37583  | -1.47078 | -3.15340 |
| 0.382 | -1.38220  | -1.46407 | -3.15373 |
| 0.383 | -1.38984  | -1.45610 | -3.15407 |
| 0.384 | -1.40000  | -1.44560 | -3.15440 |
| 0.386 | -3.15507  | -1.42247 | -0.02520 |
| 0.387 | -3.15540  | -1.42230 | -0.03481 |

| K      | Raíz real | a        | b       |
|--------|-----------|----------|---------|
| 59.996 | -5.99991  | -0.00004 | 3.31655 |
| 59.997 | -5.99994  | -0.00003 | 3.31657 |
| 59.998 | -5.99996  | -0.00002 | 3.31659 |
| 59.999 | -5.99998  | -0.00001 | 3.31661 |
| 60.000 | -6.00000  | -0.00000 | 3.31662 |
| 60.001 | -6.00002  | -0.00001 | 3.31664 |
| 60.002 | -6.00004  | -0.00002 | 3.31666 |
| 60.003 | -6.00006  | -0.00003 | 3.31668 |
| 60.004 | -6.00008  | -0.00004 | 3.31670 |

a)



b)

**Figura 2.24**

a) Tabla de valores de las raíces; arriba con incrementos de 0.1 para K; abajo con incrementos de 0.001.

b) Diagrama del lugar geométrico de las raíces.

una ecuación polinomial de tercer grado en  $s$ . Para cualquier valor particular de la ganancia del controlador  $K_c$ , podemos obtener las raíces de la ecuación característica. Por ejemplo, si  $K_c = 4.41$  ( $K = 26.5$ ), tenemos:

$$s^3 + 6s^2 + 11s + 32.5 = 0$$

Resolviendo por el método de Newton Raphson se encuentra una raíz real; posteriormente se degrada el polinomio con división sintética y se resuelve la ecuación cuadrática resultante, dando en este caso:

$$r_1 = -5.10, r_2 = -0.45 - 2.5j, r_3 = -0.45 + 2.5j.$$

Seleccionando otros valores de  $K$ , se obtienen otros conjuntos de raíces. Para facilitar los cálculos se elaboró el **PROGRAMA 2.8** (lugar geométrico de las raíces) que permite obtener estos conjuntos de raíces para diferentes valores de  $K$ , desde un valor inicial  $K = 0$ , hasta algún valor seleccionado y con incrementos también seleccionados. El programa también grafica estos conjuntos de raíces, con lo que puede verse el lugar geométrico de las raíces. A continuación mostramos un segmento de la tabla generada por el programa para un valor máximo de  $K = 100$  con incrementos de 0.1 y la gráfica respectiva. Las celdas con fondo blanco representan raíces reales; las celdas con fondo amarillo representan la parte real y las azules la parte imaginaria de las raíces complejas, que aparecen siempre conjugadas  $a \pm bj$ . Para simplificar la presentación, el programa escribe en la tabla sólo los valores de  $a$  y  $b$ .

Nótese que hay tres ramas correspondientes a las tres raíces y que dichas ramas “emergen” o empiezan (para  $K = 0$ ) en los polos de la función de transferencia de lazo abierto  $(-1, -2, -3)$ . El diagrama del lugar geométrico de las raíces es simétrico con respecto al eje real para cualquier sistema. Esto se debe al hecho de que la ecuación característica para un sistema físico tiene coeficientes reales y, por tanto, las raíces complejas de dicha ecuación aparecen en pares conjugados.

El diagrama del lugar geométrico de las raíces tiene la ventaja de dar una idea a primera vista del tipo de respuesta cuando se cambia continuamente la ganancia del contro-



lador. Por ejemplo, el diagrama de la figura 2.24b revela dos valores críticos de  $K$ ; uno es donde se hacen iguales dos de las raíces, y el otro es donde dos de las raíces son imaginarios puros. Por tanto, si las raíces son todas reales, lo cual ocurre para  $K < K_2 = 0.384$  (Fig. 2.24b), la respuesta será no oscilatoria. Si dos de las raíces son complejas y tienen partes reales negativas ( $K_2 < K < K_3$ ), la respuesta comprende términos senoidales amortiguados, que produce una respuesta oscilatoria. Si  $K > K_3$  dos de las raíces son complejas y tienen partes reales positivas, y la respuesta es senoidal creciente.

- 2.12** Es difícil situar el origen de los métodos numéricos; sin embargo, se conocía ya en Babilonia el método para calcular aproximaciones de raíces cuadradas, que contiene todos los elementos que caracterizan los métodos numéricos de hoy en día, excepto quizás por el uso de la computadora. Veamos por ejemplo cómo aproximaban  $\sqrt{2}$ .

Tomaban un primer valor inicial:  $\frac{3}{2} = 1;30^* [ \frac{3}{2} = 1;5 ]$

Como resultaba una aproximación mayor que el valor correcto ya que

$$\left(\frac{3}{2}\right)^2 = \frac{9}{4} = 2;15 > 2^{**}, [ \left(\frac{3}{2}\right)^2 = \frac{9}{4} = 2.25 ].$$

Obtenían un segundo valor inicial que quedaba por abajo del valor correcto, dividiendo 2 por  $3/2$ . El resultado es:

$$\frac{2}{\frac{3}{2}} = \frac{4}{3} = 1;20 [ \frac{2}{\frac{3}{2}} = \frac{4}{3} = 1.333333... ] \text{ (el resultado no es un decimal exacto)}$$

Ahora se tienen dos valores, uno mayor y uno menor que el valor correcto; se obtiene una mejor aproximación de  $\sqrt{2}$  sacando la media aritmética de ellos: media de  $1;30$  y

$1;20$  es  $1;25 [ \frac{1.5 + 1.333333...}{2} = 1.416666... ]$ , que resultaba mayor al valor correcto:

$$(1;25)^2 = 2;0,25 [ (1.416666...)^2 = 2.006943... ]$$

Por tanto, 2 dividido por  $1;25$  da  $1;24,42,21 [ \frac{2}{1.416666...} = 1.411765... ]$ , que es más peque-

ño que el valor correcto:  $[(1.411765...)^2 = 1.993080...]$ . El valor medio de estas dos últimas aproximaciones que encierran el valor correcto es:

$$1;25 \text{ y } 1;24,42,21 \text{ es } 1;24,51,10 \left[ \frac{1.416666... + 1.411765...}{2} \right] = 1.414215...,$$

$$(1.414215...)^2 = 2.000004066225$$

este valor resulta ser la aproximación encontrada en nuestros textos.  $\sqrt{2} = 1.414213562373$  en calculadoras modernas.

**Comentarios.** Conviene destacar varios aspectos, como por ejemplo:

- a) Se emplea un sistema numérico posicional (base 60) desarrollado por los babilonios para sus trabajos astronómicos y matemáticos.

\* Nótese el uso del sistema sexagesimal (base 60): horas; minutos, segundos o grados; minutos, segundos y que la fracción 30 viene de multiplicar la fracción decimal 0.5 por 60. Entre [ ] se escriben los valores en el sistema decimal.

\*\* La fracción sexagesimal 15 viene de multiplicar 0.25 por 60.

- b) Es un método de dos puntos que encierran el valor buscado para luego tomar el punto medio (bisección).
- c) El orden de convergencia parece ser cuadrático porque el número de cifras significativas correctas se duplica en cada iteración.
- d) Hay un criterio de terminación sustentado en la exactitud requerida por ellos para sus cálculos.
- e) Este método puede extenderse para resolver cierto tipo de ecuaciones polinómicas (véase Prob. 2.19).
- f) Es un algoritmo que puede programarse fácilmente en una computadora.

## Problemas

- 2.1 Dadas las siguientes expresiones para  $x = g(x)$ , obtenga  $g'(x)$  y dos valores iniciales que satisfagan la condición  $|g'(x)| < 1$

$$\begin{array}{lll} a) x = \frac{1}{(x+1)^2} & b) x = 4 + \left(\frac{x-1}{x+1}\right) & c) x = \sin x \\ d) \tan x = \ln x & e) x = \left(\frac{6-x-x^3}{4}\right)^{1/2} & f) x = \frac{\sec x}{2} \end{array}$$

- 2.2 Determine una  $g(x)$  y un valor inicial  $x_0$  tales que  $|g'(x)| < 1$  en las siguientes ecuaciones:

$$\begin{array}{lll} a) 2x = 4x^2 - 1 & b) x^3 - 10x - 5 = 0 & c) \sin x + \ln x = 0 \\ d) e^x - \tan x = 0 & e) \sqrt{x^3} \sin x - \ln(\cos x) = 3 \end{array}$$

- 2.3 Resuelva por el método de punto fijo las ecuaciones de los problemas anteriores.

- 2.4 Generalmente hay muchas maneras de pasar de  $f(x) = 0$  a  $x = g(x)$  e incluso se pueden obtener distintas formas de  $g(x)$  al “despejar”  $x$  de un mismo término de  $f(x)$ . Por ejemplo, en la ecuación polinomial

$$x^3 - 2x - 2 = 0$$

al “despejar”  $x$  del primer término se puede llegar a:

$$a) x = \sqrt[3]{2x+2} \quad b) x = \sqrt{2+2/x} \quad c) x = \frac{2}{x} + \frac{2}{x^2}$$

¿Cuál  $g(x)$  sería más ventajosa para encontrar la raíz que está en el intervalo  $(1, 2)$ ?

Calcule con un mismo valor inicial dicha raíz empleando las tres  $g(x)$  y compare resultados.

- 2.5 Utilice la fórmula de Francis (véase ejercicio 2.3)

- a) Encuentre una expresión  $H = g(H)$  tal que usando como valor inicial  $H = B/2$ , el método de punto fijo prometa convergencia (quizá sea necesaria una  $g(H)$  distinta para cada pareja  $B, Q$  dada).
- b) Con los valores de  $B$  y  $Q$  dados en el ejercicio 2.3 y los resultados del inciso a), calcule los respectivos valores de  $H$  que satisfacen la ecuación de Francis.

- 2.6** Sea el polinomio de grado  $n$  en su forma más general

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + a_{n-2} x^{n-2} + \dots + a_1 x + a_0$$

- a) Calcule el número de multiplicaciones y sumas algebraicas necesarias para evaluar  $f(x)$  en un punto dado mediante el método de Horner.
- b) Calcule el número de multiplicaciones y sumas algebraicas requeridas para evaluar  $f(x)$  en un punto dado usando la forma tradicional. Al comparar las cantidades de los incisos (a) y (b), encontrará que el número de multiplicaciones y sumas algebraicas en el arreglo de Horner se reduce prácticamente a la mitad. Como cada multiplicación involucra errores de redondeo, este método de evaluación es más exacto y rápido.
- 2.7** Elabore un programa que evalúe polinomios según la regla de Horner.
- 2.8** Resuelva las siguientes ecuaciones por medio del método de Newton-Raphson.

$$\begin{array}{ll} a) \ln x - x + 2 = 0 & b) x e^x - 2 = 0 \\ c) x - 2 \cos x = 0 & d) x^3 - 5x = -1 \end{array}$$

- 2.9** Resuelva los siguientes sistemas de ecuaciones por medio del método de Newton-Raphson

$$\begin{array}{ll} a) 2x^3 - y = 0 & b) 2x^2 - y = 0 \\ x^3 - 2 - y^3 = 0 & x = 2 - y^2 \\ c) x^2 + 5xy^2 - 3z + 1 = 0 & d) (x-1)^{1/2} + yx - 5 = 0 \\ x - \sin y = 1 & y - \sin x^2 = 0 \\ y - e^{-z} = 0 & \end{array}$$

- 2.10** La manera más simple de evitar el cálculo de  $f'(x)$  en el método de Newton-Raphson es reemplazar  $f'(x)$  en la ecuación 2.12 con un valor constante  $m$ . La fórmula resultante

$$x_{i+1} = x_i - \frac{f(x_i)}{m}$$

define un método de convergencia lineal para  $m$  en cierto intervalo de valores.

- a) Utilice este algoritmo, conocido como el método de Wittaker, para encontrar una raíz real de la ecuación

$$f(x) = x^3 + 2x^2 + 10x - 20 = 0$$

- b) Con este algoritmo encuentre una raíz en el intervalo (1.5, 2.5) de la ecuación

$$f(x) = x^3 - 12x^2 + 36x - 32 = 0$$

- 2.11** Demuestre que en el método de Newton-Raphson  $g'(\bar{x}) = 0$  y  $g''(\bar{x}) \neq 0$  para raíces reales no repetidas.
- 2.12** Dado un polinomio de grado  $n$

$$p_n(x) = a_n x^n + a_{n-1} x^{n-1} + a_{n-2} x^{n-2} + \dots + a_0 \quad (1)$$

elabore un programa para encontrar todas las raíces reales y complejas de  $p_n(x)$ , mediante el método de Newton-Raphson.

El programa deberá tener incorporada la división sintética para:

- a) Evaluar polinomios.
- b) Degradar polinomios cada vez que se encuentre una raíz (véase Sec. 2.10).



- 2.13** Demuestre que en el método de Newton-Raphson

$$\epsilon_{i+1} \approx \frac{f''(\bar{x})}{2! f'(\bar{x})} \epsilon_i^2$$

**SUGERENCIA:** Utilice la ecuación

$$\epsilon_{i+1} = g'(\bar{x}) \epsilon_i + g''(\bar{x}) \frac{\epsilon_i^2}{2!} + g'''(\bar{x}) \frac{\epsilon_i^3}{3!} + \dots$$

y los resultados del problema 2.11

- 2.14** El siguiente algoritmo se conoce como método de Richmond y es de tercer orden

$$x_{i+1} = x_i - \frac{2f(x_i)f'(x_i)}{2[f'(x_i)]^2 - f(x_i)f''(x_i)}$$

Resuelva las ecuaciones de los problemas 2.8 y 2.9 con este algoritmo y compare los resultados con los obtenidos con el método de Newton Raphson; por ejemplo, la velocidad de convergencia y el número de cálculos por iteración.

- 2.15** Obtenga la expresión 2.14 del algoritmo de posición falsa, utilizando la semejanza de los triángulos rectángulos cuyos vértices son:  $A x_i x_M$  y  $B x_D x_M$ , en la figura 2.7
- 2.16** La expresión 2.13, puede escribirse también

$$x_{i+1} = \frac{x_{i-1}f(x_i) - x_i f(x_{i-1})}{f(x_i) - f(x_{i-1})}$$

Explique por qué, en general, es más eficiente la ecuación 2.13 que la ecuación anterior en la aplicación del método de la secante.

- 2.17** Resuelva por el método de la secante, posición falsa o bisección las siguientes ecuaciones

- a)  $x \log x - 10 = 0$
- b)  $\sin x - \csc x + 1 = 0$
- c)  $e^x + 2^{-x} + 2 \cos x - 6 = 0$
- d)  $e^x + x^3 + 2x^2 + 10x - 20 = 0$

**SUGERENCIA:** Utilice un análisis preliminar de estas funciones para obtener valores iniciales apropiados.

- 2.18** Elabore un programa para encontrar una raíz de  $f(x) = 0$ , por el método de posición falsa, dada  $f(x)$  como una tabla de valores.
- 2.19** Encuentre una aproximación a  $\sqrt[3]{2}$  y a  $\sqrt{3}$  mediante el método de la bisección. El cálculo deberá ser correcto en cuatro dígitos significativos.  
Sugerencia: considere  $f(x) = x^3 - 2 = 0$  y  $f(x) = x^2 - 3 = 0$ , respectivamente.
- 2.20** Utilice la expresión 2.15 para hallar el número aproximado de iteraciones  $n$  a fin de encontrar una raíz de

$$x^2 + 10 \cos x = 0$$

con una aproximación de  $10^{-3}$ . Encuentre además dicha raíz.

- 2.21** Aplique el método de bisección y el de posición falsa a la ecuación

$$\frac{7x - 3}{(x - 0.45)^2} = 0$$

Use los intervalos (0.4, 0.5) y (0.39, 0.53). Explique gráficamente los resultados.

- 2.22** Demuestre que en el caso de convergencia de una sucesión de valores  $x_0, x_1, x_2, \dots$  a una raíz  $x$  en el método de punto fijo se cumple que

$$\lim_{i \rightarrow \infty} \frac{\epsilon_{i+1}}{\epsilon_i} = g'(\bar{x})$$

- 2.23** Las siguientes sucesiones convergen y los límites de convergencia de cada una se dan al lado derecho

$$a) \ x_k = \frac{(-1)^k}{k} \qquad \lim_{k \rightarrow \infty} \{x_k\} = 0$$

$$b) \ x_n = n \ln(1 + 1/n) \qquad \lim_{n \rightarrow \infty} \{x_n\} = 1$$

$$c) \ x_k = \frac{2^{k+1} + (-1)^k}{2^k} \qquad \lim_{k \rightarrow \infty} \{x_k\} = 2$$

$$d) \ x_k = 1 + e^{-k} \qquad \lim_{k \rightarrow \infty} \{x_k\} = 1$$

Genere en cada inciso la sucesión finita:  $x_1, x_2, x_3, \dots, x_{10}$

Aplique después el algoritmo de Aitken a estas sucesiones para generar las nuevas sucesiones  $x'_1, x'_2, x'_3, \dots$  observe qué ocurre y dé sus conclusiones.

- 2.24** Modifique el algoritmo 2.5 de Steffensen, incorporando una prevención para el caso en que el denominador de la ecuación 2.22 sea muy cercano a cero.

- 2.25** Encuentre una aproximación  $\sqrt[3]{2}$  y a  $\sqrt{3}$  con el método de Steffensen. El cálculo deberá ser correcto en cuatro dígitos significativos. Compare los resultados con los obtenidos en el problema 2.19.

- 2.26** Aproxime una solución para cada una de las siguientes ecuaciones con una aproximación de  $10^{-5}$ , usando el método de Steffensen con  $x_0 = 0$ .

$$a) \ 3x - x^2 + e^x - 2 = 0 \qquad b) \ 4.1 x^2 - 1.3 e^x = 0 \qquad c) \ x^2 + 2xe^x - e^{2x} = 0$$

- 2.27** Encuentre la gráfica aproximada de las siguientes funciones en los intervalos indicados

$$a) \ f(x) = e^{x^2} + x - 1000; (1, 10)$$

$$b) \ f(x) = x^4 - 2x + 10; (-\infty, \infty)$$

$$c) \ f(x) = 4(x-2)^{1/3} + \sin(3x); [0, \infty)$$

$$d) \ f(x) = \frac{1}{\sqrt{2}} e^{-x^2/2}; -\infty < x < \infty$$

$$e) \ f(x) = x^2 - 4 + \ln 3x + 5 \sin x$$

- 2.28** Utilizando el método del Newton-Raphson con valores iniciales complejos ( $a + bi$ ), encuentre las raíces complejas del polinomio

$$f(x) = x^3 + 4x + 3x^2 + 12$$

- 2.29** Utilizando el método de Müller con valores iniciales reales, encuentre las raíces complejas del polinomio del problema 2.28.

- 2.30** Encuentre las raíces faltantes de la ecuación polinomial usada a lo largo del capítulo para ilustrar los distintos métodos

$$f(x) = x^3 + 2x^2 + 10x - 20 = 0,$$

pero usando ahora el método de Newton-Raphson con valores iniciales complejos.

- 2.31** La solución general de la ecuación polinomial

$$p(x) = a_0 + a_1 x + a_2 x^2$$

es 
$$x_1 = \frac{-a_1 + \sqrt{a_1^2 - 4a_0a_2}}{2a_2}; \quad x_2 = \frac{-a_1 - \sqrt{a_1^2 - 4a_0a_2}}{2a_2}$$

a) Demuestre que  $x_1 x_2 = a_0/a_2$

b) Utilizando a), demuestre que una forma alterna para encontrar las raíces de

$$p(x) = a_0 + a_1 x + a_2 x^2 = 0$$

es 
$$x_1 = \frac{2a_0}{-a_1 + \sqrt{a_1^2 - 4a_0a_2}}; \quad x_2 = \frac{2a_0}{-a_1 - \sqrt{a_1^2 - 4a_0a_2}}$$

c) Calcule la raíz  $x_2$  de

$$p(x) = x^2 + 81x - 0.5 = 0$$

usando aritmética de cuatro dígitos con las dos formas presentadas y sustituya ambos resultados en  $p(x)$ . Compare la exactitud de los resultados y explique la diferencia. Puede usar Mathematica o Fortan.

d) Calcule la raíz  $x_1$  de

$$p(x) = x^2 + 81x - 0.5 = 0$$

- 2.32** Elabore un programa de propósito general para encontrar todas las raíces reales y complejas de una ecuación polinomial de la forma

$$p_n(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

con el método de Müller.

- 2.33** El siguiente algoritmo, de orden tres, es conocido como método de Laguerre

$$x_{i+1} = x_i - \frac{np(x_i)}{p'(x_i) \pm \sqrt{H(x_i)}}, \quad i = 0, 1, 2, \dots$$

donde  $n$  es el grado de la ecuación polinomial  $p(x) = 0$ , cuyas raíces se desea encontrar

$$H(x_i) = (n-1) [(n-1) (p'(x_i))^2 - np(x_i) p''(x_i)]$$

y el signo del radical queda determinado por el signo de  $p'(x_i)$ .

Este método, que funciona con orden 3 para polinomios cuyas raíces son todas reales y distintas, converge sólo linealmente para raíces múltiples. En el caso de raíces complejas poco se sabe del orden de convergencia; no obstante, ésta es alta para raíces complejas simples. Finalmente, se hace la observación de que un valor de  $x_i$  real puede producir una  $H(x_i)$  negativa y, por tanto, generar un valor de  $x_{i+1}$  complejo y eventualmente llevar a una raíz compleja de la ecuación  $p(x) = 0$ .

Resuelva las siguientes ecuaciones con el método de Laguerre

a)  $x^4 - 8.2x^3 + 39.41x^2 - 62.26x + 30.25 = 0$

b)  $x^4 - 15.2x^3 + 59.7x^2 - 81.6x + 36 = 0$

c)  $x^5 - 10x^4 + 40x^3 - 80x^2 + 79x - 30 = 0$

d)  $x^5 - 3.7x^4 + 7.4x^3 - 10.8x^2 + 10.8x - 6.8 = 0$

- 2.34** Se ha encontrado una simplificación\* al algoritmo de Müller (véase algoritmo 2.6), y es

\* Hildebrand, B. *Introduction to Numerical Analysis*. 2a. ed. McGraw-Hill (1974) pp. 580-581.



$$x_{i+1} = x_i - \frac{2\lambda_i}{1 + \sqrt{1 - 4\lambda_i\mu_i}} \quad i = 2, 3, 4 \dots \quad (1)$$

donde:

$$\lambda_i = \frac{f_i}{w_i}, \mu_i = \frac{f[x_i, x_{i-1}, x_{i-2}]}{w_i}$$

y

$$\begin{aligned} w_i &= f[x_i, x_{i-1}] + (x_i - x_{i-1}) f[x_i, x_{i-1}, x_{i-2}] \\ &= f[x_i, x_{i-1}] + (f_i - f_{i-1}) \frac{f[x_i, x_{i-1}, x_{i-2}]}{f[x_i, x_{i-1}]} \end{aligned}$$

Para esta modificación el orden de convergencia está dado por:

$$\epsilon_{i+1} \approx - \frac{f'''(\bar{x})}{6f'(\bar{x})} \epsilon_i \epsilon_{i-1} \epsilon_{i-2}$$

Resuelva las ecuaciones dadas en los problemas 2.17, 2.26 y 2.33 con este algoritmo.

- 2.35** Con la fórmula 1 del problema 2.34 y algunas consideraciones teóricas que se omiten por ser más bien tema del análisis numérico, se llega a modificaciones del método de la secante, con lo cual se consigue en éstas un orden de convergencia mayor de 2.

a) La primera modificación está dada por la ecuación

$$x_{i+1} = x_i - \frac{2\lambda_i}{1 + \sqrt{1 - 4\mu_i\lambda_i}}, \quad i = 1, 2, \dots \quad (1)$$

pero ahora

$$\lambda_i = \frac{f_i}{f'_i}, \quad \mu_i = \frac{f'_i - f[x_i, x_{i-1}]}{(x_i - x_{i-1}) f'_i}$$

y

$$w_i = f'_i$$

La interpretación geométrica de este método consiste en reemplazar la función  $f(x)$  en cierto intervalo con una parábola que pasa por el punto  $(x_{i-1}, f_{i-1})$  y es tangente a la curva de  $f(x)$  en  $(x_i, f'_i)$ . Para la ecuación 1 se tiene que

$$\epsilon_{i+1} \approx - \frac{f'''(\bar{x})}{6f'(\bar{x})} \epsilon_i^2 \epsilon_{i-1}$$

y se ha encontrado que es aproximadamente de orden 2.41

b) La segunda modificación está dada por la expresión

$$x_{i+1} = x_{i-1} + \frac{2(f_i / f'_i)}{1 + \sqrt{1 - 2[f_i f''_i / (f'_i)^2]}} \quad (2)$$

y en ésta el orden de convergencia es 3, y se sabe que:

$$\epsilon_{i+1} \approx - \frac{f'''(\bar{x})}{6f'(\bar{x})} \epsilon_i^3$$

Aquí, la curva que reemplaza a  $f(x)$  en cierto intervalo es una parábola que coincide con la curva de  $f(x)$  en  $x_i$  y tiene la misma pendiente y curvatura que  $f(x)$  en  $x_i$ . Resuelva las ecuaciones dadas en los problemas 2.17, 2.26 y 2.33, usando las modificaciones de los incisos (a) y (b).

- c) De estas fórmulas pueden obtenerse otras más simples mediante aproximaciones. Por ejemplo, si  $f_i$  es pequeña, puede hacerse

$$\left(1 - 2 \frac{f_i f_i''}{(f_i')^2}\right)^{1/2} \approx 1 - \frac{f_i f_i''}{(f_i')^2}$$

en la ecuación 2 y obtener la fórmula simplificada

$$x_{i+1} = x_i - \frac{f_i / f_i'}{1 - f_i f_i'' / 2 (f_i')^2}$$

para la cual

$$\epsilon_{i+1} \approx \left[ \left( \frac{f''(\bar{x})}{2f'(\bar{x})} \right)^2 - \frac{f'''(\bar{x})}{6f'(\bar{x})} \right] \epsilon_i^3$$

obsérvese que también es de tercer orden, pero sin raíz cuadrada. Esta fórmula se atribuye a Halley. Los métodos iterativos basados en esta expresión algunas veces se denominan métodos de Bailey o métodos de Lambert.

- d) Si se aproxima

$$\left[1 - \frac{f_i f_i''}{2 (f_i')^2}\right]^{-1} \approx 1 + \frac{f_i f_i''}{2 (f_i')^2}$$

en la fórmula de Halley, se obtiene la iteración

$$x_{i+1} = x_i - \frac{f_i}{f_i'} \left[ 1 + \frac{f_i f_i''}{2 (f_i')^2} \right] \quad (4)$$

con

$$\epsilon_{i+1} \approx \left[ 2 \left( \frac{f''(\bar{x})}{2f'(\bar{x})} \right)^2 - \frac{f'''(\bar{x})}{6f'(\bar{x})} \right] \epsilon_i^3$$

cuyo orden es tres también y se llama fórmula de Chebyshev.

Resuelva las ecuaciones dadas en los problemas 2.17, 2.26 y 2.33, empleando los algoritmos de Halley y Chebyshev cuando sean aplicables y compare los resultados obtenidos con los algoritmos de los incisos (a) y (b).

**2.36** La ecuación de estado de Beattie-Bridgeman en su forma virial es:

$$PV = RT + \frac{\beta}{V} + \frac{\gamma}{V^2} + \frac{\delta}{V^3}$$

donde:

P = presión de atm

T = temperatura en K

V = volumen molar en L/gmol

R = Constante universal de los gases en atm L/(gmol K)

$\beta = RTB_0 - A_0 - Rc/T^2$

$\gamma = -RTB_0b + A_0a - RB_0c/T^2$

$$\delta = R B_0 b c / T^2, \text{ y}$$

$A_0, B_0, a, b, c =$  constantes particulares para cada gas.

Calcule el volumen molar  $V$  a 50 atm y 100°C para los siguientes gases

| Gas            | $A_0$  | $a$      | $B_0$   | $b$      | $c \times 10^{-4}$ |
|----------------|--------|----------|---------|----------|--------------------|
| He             | 0.0216 | 0.05984  | 0.01400 | 0.000000 | 0.0040             |
| H <sub>2</sub> | 0.1975 | -0.00506 | 0.02096 | -0.43590 | 0.0504             |
| O <sub>2</sub> | 1.4911 | 0.02562  | 0.04624 | 0.004208 | 4.8000             |



2.37 La ecuación de estado de Redlich-Kwong es:

$$\left[ P + \frac{a}{T^{1/2} V (V + b)} \right] (V - b) = RT$$

donde:

$P =$  presión en atm

$T =$  temperatura en K

$V =$  volumen molar en L/gmol

$R =$  constante universal de los gases en atm-L/(gmol K)

$$a = 0.4278 \frac{R^2 T_c^{2.5}}{P_c}, \quad b = 0.0867 \frac{R T_c}{P_c}$$

Calcule el volumen molar  $V$  a 50 atm y 100 °C para los siguientes gases

| Gas            | $P_c$ (atm) | $T_c$ ( K ) |
|----------------|-------------|-------------|
| He             | 2.26        | 5.26        |
| H <sub>2</sub> | 12.80       | 33.30       |
| O <sub>2</sub> | 49.70       | 154.40      |

Compare los resultados obtenidos con los del problema 2.36.

2.38 Mediante la ecuación de estado de Van der Waals (véase ejercicio 2.1), encuentre el volumen molar  $V$  del CO<sub>2</sub> a 80 °C y 10 atm, utilizando los métodos de Newton-Raphson y de Richmond (véase Probl. 2.14).

2.39 Descomponga en fracciones parciales las siguientes funciones racionales

$$a) F(s) = \frac{52.5 s (s + 1) (s + 1.5) (s + 5)}{s^4 + 20.75 s^3 + 92.6 s^2 + 73.69 s}$$

$$b) F(s) = \frac{10 A}{s^3 + 101.4 s^2 + 142.7 s + 100}$$

$$c) F(s) = \frac{0.47 K_G (s^3 + 4.149 s^2 + 6.362 s + 4.255)}{s^4 + 7 s^3 + 11 s^2 + 5 s}$$

$$d) F(s) = \frac{100 (s^2 + 3.4 s + 2.8)}{s^5 + 10 s^4 + 32 s^3 + 38 s^2 + 15 s}$$



-  **2.40** Una forma alterna para resolver el problema de vaporización instantánea (véase ejercicio 2.6) es:

Tomando en cuenta que  $\sum x_i = 1$  y que  $\sum y_i = 1$ , o bien  $\sum K_i x_i = 1$ , puede escribirse

$$\frac{\sum K_i x_i}{\sum x_i} = 1 \quad (\text{todas las sumatorias sobre } i \text{ son de } 1 \text{ a } n)$$

o también

$$\ln \frac{\sum K_i x_i}{\sum x_i} = 0$$

Siguiendo la secuencia mostrada en el ejercicio 2.6, se llega a la expresión:

$$\ln \frac{\sum \frac{K_i z_i}{1 + \psi(K_i - 1)}}{\sum \frac{z_i}{1 + \psi(K_i - 1)}} = 0$$

Utilice el método de posición falsa y los datos del ejercicio 2.6 para resolver esta última ecuación.

-  **2.41** Para el cálculo de la temperatura de burbuja de una mezcla multicomponente a la presión total  $P$  se utiliza la ecuación

$$f(T) = \sum_{i=1}^n K_i x_i - 1 = 0 \quad (1)$$

donde  $x_i$  y  $K_i$ ,  $i = 1, 2, \dots, n$  son la fracción mol en la fase líquida y la relación de equilibrio del componente  $i$ , respectivamente, y  $T$  (la raíz de la ecuación 1) es la temperatura de burbuja.

Determine la temperatura de burbuja a 10 atm de presión total de una mezcla cuya composición en la fase líquida es 45% mol de n-butano, 30% mol de n-pentano y 25% mol de n-hexano. Los valores de  $K_i$  a 10 atm son

| Componente | $K(T)$ con $T$ en °C para $35 \leq T \leq 205$ °C                |
|------------|------------------------------------------------------------------|
| n-butano   | $-0.17809 + 1.2479 \times 10^{-2} T + 3.7159 \times 10^{-5} T^2$ |
| n-pentano  | $0.13162 - 1.9367 \times 10^{-3} T + 7.1373 \times 10^{-5} T^2$  |
| n-hexano   | $0.13985 - 3.8690 \times 10^{-3} T + 5.5604 \times 10^{-5} T^2$  |

-  **2.42** Para el cálculo de la temperatura de rocío de una mezcla multicomponente a la presión total  $P$  se utiliza la ecuación.

$$f(T) = \sum_{i=1}^n \frac{y_i}{K_i} - 1 = 0 \quad (1)$$

donde  $y_i$  y  $K_i$ ,  $i = 1, 2, \dots, n$  son la fracción mol en la fase vapor y la relación de equilibrio del componente  $i$ , respectivamente, y  $T$  (la raíz de la ecuación) es la temperatura de rocío.

Determine la temperatura de rocío a 10 atm de presión total de una mezcla cuya composición en la fase líquida es 45% mol de n-butano, 30% mol de n-pentano y 25% mol, de n-hexano. Los valores de  $K_i$  a 10 atm se proporcionan en el problema 2.41.

-  **2.43** Para obtener la temperatura de burbuja de una solución líquida de  $\text{CCl}_4$  y  $\text{CF}_4$  en equilibrio con su vapor, se llegó a la ecuación:

$$760 = 0.75 [ 10^{6.898-1221.8 / (T+227.4)} ] + 0.25 [ 10^{6.195-376.71 / (T+241.2)} ]$$

Aplicando un método iterativo de dos puntos, encuentre la temperatura de burbuja  $T$  con una aproximación de  $10^{-2}$  aplicado a  $f(T)$ .

- 2.44** En la solución de ecuaciones diferenciales ordinarias con coeficientes constantes, es necesario resolver la “ecuación auxiliar asociada”, que resulta ser un polinomio cuyo grado es igual al orden de la ecuación diferencial. Así, si la ecuación diferencial está dada por

$$y^{IV} + 2y'' - 8y = 0 \quad (1)$$

la ecuación auxiliar asociada es

$$m^4 + 2m^2 - 8 = 0$$

cuyas cuatro raíces:  $m_1$ ,  $m_2$ ,  $m_3$  y  $m_4$  se emplean de la siguiente manera

$$y = c_1 e^{m_1 x} + c_2 e^{m_2 x} + c_3 e^{m_3 x} + c_4 e^{m_4 x}$$

para dar la solución general de la ecuación 1.

Encuentre la solución general de la ecuación 1 y de las siguientes ecuaciones diferenciales

$$y^{VI} + 2y^{IV} + y'' = 0$$

$$y''' - 4y'' + 4y' = 0$$

- 2.45** La ecuación 4 del ejercicio 2.8 se aplica para calcular la  $\Delta T_m$ , cuando

$$TC_1 - TC_2 \neq TH_2 - TC_1$$

Cuando el gradiente  $TH_1 - TC_2$  es muy cercano al gradiente  $TH_2 - TC_1$  se deberá utilizar la siguiente expresión para el cálculo de  $\Delta T_m$

$$\Delta T_m = \frac{(TH_1 - TC_2) + (TH_2 - TC_1)}{2}$$

Modifique el programa 2.6 del ejercicio 2.8 de modo que se utilice la  $\Delta T_m$  dada arriba cuando

$$| (TH_1 - TC_2) - (TH_2 - TC_1) | < 10^{-2}$$

y la ecuación (4) del ejercicio 2.8 en caso contrario.

- 2.46** Si el cambiador de calor del ejercicio 2.8, se opera en paralelo, esto es:



Encuentre  $TH_2$  y  $TC_2$  en estas nuevas condiciones de operación.

- 2.47** Suponga que el fenómeno de la transmisión de calor en un cierto material obedece en forma aproximada al modelo

$$T = T_0 + \frac{q}{k} \left( \beta \left( \frac{\infty T}{\pi} \right)^{1/2} e^{-x^2 / (4\alpha t)} \right)$$

Calcule el tiempo requerido para que la temperatura a la distancia  $x$  alcance un valor dado. Use la siguiente información

$$T_0 = 25\text{ }^{\circ}\text{C}; q = 300\text{ BTU/h ft}^2;$$

$$\alpha = 0.04\text{ ft}^2/\text{h}; x = 1\text{ ft};$$

$$k = 1\text{ BTU/h ft}^2\text{ }^{\circ}\text{F}$$

$$T = 120\text{ }^{\circ}\text{F}$$

$$\beta = 2 \frac{^{\circ}\text{F ft}}{\text{h}^{1/2}}\text{ }^{\circ}\text{C}^{1/2}$$

- 2.48 El factor de fricción  $f$  para fluidos pseudoplásticos que siguen el modelo de Ostwald-De-Waele se calcula mediante la siguiente ecuación

$$\frac{1}{f} = \frac{4}{n^{0.75}} \log(\text{Re } f^{1-0.5n}) - \frac{0.4}{n^{1.2}}$$

Encuentre el factor de fricción  $f$ , si se tiene un número de Reynolds  $\text{Re}$  de 6000 y un valor de  $n = 0.4$ .

- 2.49 La siguiente relación entre el factor de fricción  $f$  y el número de Reynolds  $\text{Re}$  se cumple cuando hay flujo turbulento de un fluido en un tubo liso

$$\frac{1}{f} = -0.4 + 1.74 \ln(\text{Re} \sqrt{f})$$

Construya una tabla de valores de  $f$  correspondientes a números de Reynolds de  $10^4$  hasta  $10^6$  con intervalos de  $10^4$ .

- 2.50 Para determinar la constante de nacimientos de una población se necesita calcular  $\lambda$  en la siguiente ecuación

$$1.546 \times 10^6 = 10^6 e^{\lambda} + \frac{0.435 \times 10^6}{\lambda} (e^{\lambda} - 1)$$

con una aproximación de  $10^{-3}$ .

- 2.51 Graficar por separado las funciones  $y = x$  y  $y = \tan x$  (véase ejercicio 2.10). Encuentre las raíces en el intervalo  $(0, 35)$  ¿nota usted alguna relación entre ellas? ¿Podría explicar esta relación?



# CAPÍTULO 3

## MATRICES Y SISTEMAS DE ECUACIONES LINEALES

### A dónde nos dirigimos

En este capítulo estudiaremos las técnicas de solución de sistemas de ecuaciones lineales cuadrados  $Ax = b$ . Para ello, primero realizaremos un repaso de álgebra de matrices y, para sustentar teóricamente los métodos, revisaremos las ideas de ortogonalización de vectores.

Posteriormente se exponen las dos ideas sobre las que se desarrollan, en los métodos numéricos, las soluciones de los sistemas: la eliminación de Gauss para los métodos directos y la iteración de Jacobi para los iterativos.

Para establecer la velocidad de cálculo y el “trabajo computacional” en los métodos directos, se analiza el número de operaciones de éstos y con base en ello se determinan sus necesidades de memoria. Como consecuencia de lo anterior, se da particular atención a los sistemas especiales: simétricos, bandedos y dispersos, entre otros. Así, estudiaremos los métodos que aprovechan estas características para lograr reducir con esto el número de operaciones y los requerimientos de máquina.

Los métodos iterativos se vinculan con el método de punto fijo del capítulo 2, aprovechando las ideas ahí desarrolladas como la de aceleración de la convergencia.

Al final del capítulo, se presenta una comparación entre ambas familias para brindar al lector los elementos necesarios para seleccionar la más adecuada a su problema en particular.

Dado que el mundo real puede verse como grupos de objetos o partes trabajando en conjunto o bien conectadas de alguna manera que forman un todo, creemos que con estos conocimientos lograremos brindar al lector una mejor comprensión de la extraordinaria cantidad de situaciones que pueden representarse con los sistemas o grupos de ecuaciones donde cada una de ellas corresponde a alguna de sus partes, por ejemplo: en circuitos, estructuras, columnas de destilación a régimen permanente.

### Introducción

La solución de sistemas de ecuaciones lineales es un tema clásico de las matemáticas, rico en ideas y conceptos, y de gran utilidad en ramas del conocimiento tan diversas como economía, biología, física, psicología, etc. La resolución de sistemas casi de cualquier número de ecuaciones (10, 100, 1000, etc.) es una realidad hoy día gracias a las computadoras, lo cual proporciona un atractivo especial a las técnicas de solución directas e iterativas: su programación, la cuenta de los cálculos necesarios, la propagación de errores, etcétera.

Sin embargo, todo lo anterior requiere una revisión de los conceptos básicos sobre matrices, ortogonalización de vectores, y la existencia y unicidad de las soluciones; por tanto, estos conceptos dan inicio al capítulo.

### 3.1 Matrices

Una matriz es un conjunto de elementos ordenados en filas y columnas como:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ a_{3,1} & a_{3,2} & a_{3,3} & \cdots & a_{3,n} \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & a_{m,3} & \cdots & a_{m,n} \end{bmatrix}$$

Los elementos  $a_{i,j}$  son números reales o complejos, o funciones de una o varias variables. En este libro sólo se tratarán matrices cuyos elementos son números reales.

Para denotar matrices se utilizarán las primeras letras mayúsculas del alfabeto en *cur-sivas*  $A$ ,  $B$ ,  $C$ , etc. Cuando se hace referencia a una matriz es conveniente especificar su número de filas y columnas. Así, la expresión  $A$  de  $m \times n$ , indica que se trata de una matriz de  $m$  filas y  $n$  columnas o de  $m \times n$  elementos. A " $m \times n$ " se le conoce como las dimensiones de  $A$ . Si el número de filas y de columnas es el mismo; esto es  $m = n$ , se tiene una matriz cuadrada de orden  $n$  o simplemente una matriz de orden  $n$ .

Para ciertas demostraciones es más conveniente la notación  $[a_{i,j}]$ ,  $[b_{i,j}]$ , etc., en lugar de  $A$ ,  $B$ , etcétera.

Dos matrices son iguales cuando tienen el mismo número de filas y columnas (las mismas dimensiones) y, además, los elementos correspondientes son iguales.

Por ejemplo, las matrices

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix}$$

son de orden tres y tienen los mismos elementos. Aun así son distintas, ya que los elementos correspondientes no son todos iguales. El elemento de la segunda fila y la segunda columna de  $A$ ,  $a_{2,2}$  es 5 y el correspondiente de  $B$ ,  $b_{2,2}$  es 5; pero el elemento de la segunda fila y la primera columna de  $A$ ,  $a_{2,1}$  es 4 y el correspondiente a  $B$ ,  $b_{2,1}$ , es 2.

#### OPERACIONES ELEMENTALES CON MATRICES Y SUS PROPIEDADES

Se definirán dos operaciones en el conjunto establecido de las matrices.

#### SUMA DE MATRICES

Para sumar dos matrices  $A$  y  $B$  han de ser de las mismas dimensiones; si esto es cierto, la suma es una matriz  $C$  de iguales dimensiones que  $A$  y que  $B$ , y sus elementos se obtienen sumando los elementos correspondientes de  $A$  y  $B$ . Para mayor claridad

$$\begin{array}{c}
 A \quad + \quad B \quad = \quad C \\
 \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} + \begin{bmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,n} \\ b_{2,1} & b_{2,2} & \dots & b_{2,n} \\ \vdots & \vdots & & \vdots \\ b_{m,1} & b_{m,2} & \dots & b_{m,n} \end{bmatrix} = \begin{bmatrix} a_{1,1} + b_{1,1} & a_{1,2} + b_{1,2} & \dots & a_{1,n} + b_{1,n} \\ a_{2,1} + b_{2,1} & a_{2,2} + b_{2,2} & \dots & a_{2,n} + b_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} + b_{m,1} & a_{m,2} + b_{m,2} & \dots & a_{m,n} + b_{m,n} \end{bmatrix} \\
 = \begin{bmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,n} \\ c_{2,1} & c_{1,2} & \dots & c_{2,n} \\ \vdots & \vdots & & \vdots \\ c_{m,1} & c_{m,2} & \dots & c_{m,n} \end{bmatrix}
 \end{array} \quad (3.1)$$

o también

$$[a_{i,j}] + [b_{i,j}] = [a_{i,j} + b_{i,j}] = [c_{i,j}] \quad (3.2)$$

### Ejemplo 3.1

Sumar las matrices

$$\begin{bmatrix} 4 & 8.5 & -3 \\ 2 & -1.3 & 7 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} -1 & 2 & -4 \\ 5 & 8 & 3 \end{bmatrix}$$

**Solución**



$$\begin{array}{ccc}
 \begin{bmatrix} 4 & 8.5 & -3 \\ 2 & -1.3 & 7 \end{bmatrix} & + & \begin{bmatrix} -1 & 2 & -4 \\ 5 & 8 & 3 \end{bmatrix} = \begin{bmatrix} 4-1 & 8.5+2 & -3-4 \\ 2+5 & -1.3+8 & 7+3 \end{bmatrix} = \begin{bmatrix} 3 & 10.5 & -7 \\ 7 & 6.7 & 10 \end{bmatrix} \\
 2 \times 3 & & 2 \times 3 \qquad \qquad \qquad 2 \times 3 \qquad \qquad \qquad 2 \times 3
 \end{array}$$

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
A=[4 8.5 -3; 2 -1.3 7]
B=[-1 2 -4; 5 8 3]
C=A+B
```

```
[4, 8.5, -3; 2, -1.3, 7]→a
[-1, 2, -4; 5, 8, 3]→b
a+b→c
```

La conmutatividad y asociatividad de la suma de matrices son propiedades heredadas de las propiedades de la suma de los números reales. Así, la conmutatividad puede verse claramente en la ecuación 3.1, ya que:

$$a_{i,j} + b_{j,j} = b_{i,j} + a_{i,j} = c_{i,j}$$

donde  $a_{i,j}$  representa un elemento cualquiera de  $A$  y  $b_{i,j}$  su correspondiente en  $B$ . Por tanto, es cierto que

$$A + B = B + A = C$$

De igual manera puede verse la asociatividad

$$(a_{i,j} + b_{i,j}) + d_{i,j} = a_{i,j} + (b_{i,j} + d_{i,j})$$



o bien

$$(A + B) + D = A + (B + D)$$

donde  $D$  es una matriz de las mismas dimensiones que  $A$  y que  $B$ .

Además, si se denota con  $O$  a la matriz cuyos elementos son todos cero (matriz cero); es decir,

$$O = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ \cdot & \cdot & \cdot & & \cdot \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}$$

y por  $-A$  la matriz cuyos elementos son los mismos que  $A$ , pero de signo contrario

$$-A = \begin{bmatrix} -a_{1,1} & -a_{1,2} & \dots & -a_{1,n} \\ -a_{2,1} & -a_{2,2} & \dots & -a_{2,n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ -a_{m,1} & -a_{m,2} & \dots & -a_{m,n} \end{bmatrix}$$

se tiene:

$$A + O = A, \quad (3.3)$$

$$A + (-A) = O \quad (3.4)$$

A partir de la ecuación 3.4, puede definirse la resta entre  $A$  y  $B$  como

$$A + (-B)$$

o más simple

$$A - B$$

### PRODUCTO DE MATRICES POR UN ESCALAR

Así como se ha definido la suma de matrices, también se puede formar el producto de un número real  $\alpha$  y una matriz  $A$ . El resultado, denotado por  $\alpha A$ , es la matriz cuyos elementos son los componentes de  $A$  multiplicados por  $\alpha$ . Así, se tiene

$$\alpha A = \alpha \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} = \begin{bmatrix} \alpha a_{1,1} & \alpha a_{1,2} & \dots & \alpha a_{1,n} \\ \alpha a_{2,1} & \alpha a_{2,2} & \dots & \alpha a_{2,n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \alpha a_{m,1} & \alpha a_{m,2} & \dots & \alpha a_{m,n} \end{bmatrix} \quad (3.5)$$

o bien

$$\alpha [a_{i,j}] = [\alpha a_{i,j}] \quad (3.6)$$

**Ejemplo 3.2**

Multiplique la matriz  $\begin{bmatrix} 5.8 & -2.3 & 2 \\ 4 & 7.2 & 10 \\ 43 & -13 & 5 \end{bmatrix}$  por 2.

**Solución**

$$2 \begin{bmatrix} 5.8 & -2.3 & 2 \\ 4 & 7.2 & 10 \\ 43 & -13 & 5 \end{bmatrix} = \begin{bmatrix} 2(5.8) & 2(-2.3) & 2(2) \\ 2(4) & 2(7.2) & 2(10) \\ 2(43) & 2(-13) & 2(5) \end{bmatrix} = \begin{bmatrix} 11.6 & -4.6 & 4 \\ 8 & 14.4 & 20 \\ 86 & -26 & 10 \end{bmatrix}$$

Las principales propiedades algebraicas de esta multiplicación son:

$$\alpha(A + B) = \alpha A + \alpha B, \text{ distributividad respecto a la suma de matrices.} \quad (3.7)$$

$$(\alpha + \beta)A = \alpha A + \beta A, \text{ distributividad respecto a la suma de escalares.} \quad (3.8)$$

$$(\alpha\beta)A = \alpha(\beta A), \text{ asociatividad.} \quad (3.9)$$

$$1A = A, \quad (3.10)$$

donde  $\alpha$  y  $\beta$  son dos escalares cualesquiera, y  $A$  y  $B$  dos matrices sumables (con igual número de filas e igual número de columnas).

Las ecuaciones 3.7 a 3.10, se comprueban con facilidad a partir de las definiciones de suma de matrices y multiplicación por un escalar. Sólo se demostrará la 3.9; las otras quedan como ejercicio para el lector.

De la definición (Ec. 3.5), aplicada al lado izquierdo de la ecuación 3.9.

$$(\alpha\beta)A = \begin{bmatrix} (\alpha\beta)a_{1,1} & (\alpha\beta)a_{1,2} & \dots & (\alpha\beta)a_{1,n} \\ (\alpha\beta)a_{2,1} & (\alpha\beta)a_{2,2} & \dots & (\alpha\beta)a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ (\alpha\beta)a_{m,1} & (\alpha\beta)a_{m,2} & \dots & (\alpha\beta)a_{m,n} \end{bmatrix}$$

De la asociatividad de la multiplicación de los números reales se tiene:

$$(\alpha\beta)A = \begin{bmatrix} \alpha(\beta a_{1,1}) & \alpha(\beta a_{1,2}) & \dots & \alpha(\beta a_{1,n}) \\ \alpha(\beta a_{2,1}) & \alpha(\beta a_{2,2}) & \dots & \alpha(\beta a_{2,n}) \\ \vdots & \vdots & \ddots & \vdots \\ \alpha(\beta a_{m,1}) & \alpha(\beta a_{m,2}) & \dots & \alpha(\beta a_{m,n}) \end{bmatrix}$$

Al aplicar la ecuación 3.5 en sentido inverso dos veces:

$$(\alpha\beta)A = \alpha \begin{bmatrix} \beta a_{1,1} & \beta a_{1,2} & \dots & \beta a_{1,n} \\ \beta a_{2,1} & \beta a_{2,2} & \dots & \beta a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \beta a_{m,1} & \beta a_{m,2} & \dots & \beta a_{m,n} \end{bmatrix}$$

$$(\alpha\beta)A = \alpha \left[ \beta \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} \right]$$

se llega al lado derecho de la ecuación 3.9, con lo cual concluye la demostración.

## MULTIPLICACIÓN DE MATRICES

Dos matrices  $A$  y  $B$  son conformes en ese orden (primero  $A$  y después  $B$ ), si  $A$  tiene el mismo número de columnas que  $B$  tiene de filas.

Se definirá la multiplicación sólo para matrices conformes. Dada una matriz  $A$  de  $m \times n$  y una matriz  $B$  de  $n \times p$ , el producto es una matriz  $C$  de  $m \times p$  cuyo elemento general  $c_{ij}$  se obtiene por la suma de los productos de los elementos de la  $i$ -ésima fila de  $A$  y la  $j$ -ésima columna de  $B$ . Si:

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ a_{i,1} & a_{i,2} & \dots & a_{i,n} \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} b_{1,1} & b_{1,2} & \dots & b_{1,j} & \dots & b_{1,p} \\ b_{2,1} & b_{2,2} & \dots & b_{2,j} & \dots & b_{2,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{n,1} & b_{n,2} & \dots & b_{n,j} & \dots & b_{n,p} \end{bmatrix}$$

$$AB = C = \begin{bmatrix} c_{1,1} & c_{1,2} & \dots & c_{1,j} & \dots & c_{1,p} \\ c_{2,1} & c_{2,2} & \dots & c_{2,j} & \dots & c_{2,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{i,1} & c_{i,2} & \dots & c_{i,j} & \dots & c_{i,p} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{m,1} & c_{m,2} & \dots & c_{m,j} & \dots & c_{m,p} \end{bmatrix}$$

donde:

$$c_{i,j} = a_{i,1}b_{1,j} + a_{i,2}b_{2,j} + \dots + a_{i,n}b_{n,j}$$

o bien:

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



**Ejemplo 3.3**

Multiplicar las matrices  $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & -4 & -5 \end{bmatrix}$  y  $B = \begin{bmatrix} 0 & 1 & -2 \\ -1 & 2 & 3 \\ 4 & 2 & 1 \end{bmatrix}$

**Solución**

$$\begin{matrix} A & B & C \\ \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & -4 & -5 \end{bmatrix} & \begin{bmatrix} 0 & 1 & -2 \\ -1 & 2 & 3 \\ 4 & 2 & 1 \end{bmatrix} & = \begin{bmatrix} 0-2+12 & 1+4+6 & -2+6+3 \\ 0-3+16 & 2+6+8 & -4+9+4 \\ 0+4-20 & 3-8-10 & -6-12-5 \end{bmatrix} = \begin{bmatrix} 10 & 11 & 7 \\ 13 & 16 & 9 \\ -16 & -15 & -23 \end{bmatrix} \end{matrix}$$

En orden inverso

$$\begin{matrix} B & A & C \\ \begin{bmatrix} 0 & 1 & -2 \\ -1 & 2 & 3 \\ 4 & 2 & 1 \end{bmatrix} & \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 4 \\ 3 & -4 & -5 \end{bmatrix} & = \begin{bmatrix} 0+2-6 & 0+3+8 & 0+4+10 \\ -1+4+9 & -2+6-12 & -3+8-15 \\ 4+4+3 & 8+6-4 & 12+8-5 \end{bmatrix} = \begin{bmatrix} -4 & 11 & 14 \\ 12 & -8 & -10 \\ 11 & 10 & 15 \end{bmatrix} \end{matrix}$$

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus



```
A=[1 2 3; 2 3 4; 3 -4 -5]
B=[0 1 -2; -1 2 3; 4 2 1]
disp('C = A * B')
C=A*B
disp('C = B * A')
```

```
[1,2,3;2,3,4;3,-4,-5]→a
[0,1,-2;-1,2,3;4,2,1]→b
a*b→c
b*a→c
```

Obsérvese que  $AB \neq BA$ ; es decir, la multiplicación de matrices no es conmutativa. Este hecho deberá tenerse siempre en cuenta al multiplicar matrices.

A continuación se verán las propiedades de distributividad y asociatividad del producto de matrices.

$$A(B + C) = AB + AC \quad (3.11)$$

$$(AB)C = A(BC) \quad (3.12)$$

Con la notación de sumatoria se comprobará la ecuación 3.11; la 3.12 queda como ejercicio para el lector.

**Demostración de la ecuación 3.11.** Sea  $e_{ij}$  un elemento cualquiera de la matriz producto  $A$   $B$ , esto es:

$$e_{ij} = \sum_{k=1}^n a_{i,k} b_{k,j}$$

y  $d_{i,j}$  el elemento correspondiente del producto  $A C$

$$d_{i,j} = \sum_{k=1}^n a_{i,k} c_{k,j}$$

Al sumarlos se obtiene el elemento correspondiente del lado derecho de la ecuación 3.11

$$e_{i,j} + d_{i,j} = \sum_{k=1}^n a_{i,k} b_{k,j} + \sum_{k=1}^n a_{i,k} c_{k,j} = \sum_{k=1}^n a_{i,k} (b_{k,j} + c_{k,j}),$$

el cual es igual al elemento de la  $i$ -ésima fila y la  $j$ -ésima columna del lado izquierdo de la ecuación 3.11, con lo que finaliza la demostración.

A continuación se da el algoritmo para multiplicar matrices.

### ALGORITMO 3.1 Multiplicación de matrices

Para multiplicar las matrices  $A$  y  $B$ , proporcionar los

DATOS: Número de filas y columnas de  $A$  y  $B$ ;  $N$ ,  $M$ ,  $N1$ ,  $M1$ , respectivamente, y sus elementos.

RESULTADOS: La matriz producto  $C$  de dimensiones  $N \times M1$  o el mensaje "LAS MATRICES  $A$  Y  $B$  NO PUEDEN MULTIPLICARSE".

- PASO 1. Si  $M = N1$  continuar, de otro modo IMPRIMIR "LAS MATRICES  $A$  Y  $B$  NO SE PUEDEN MULTIPLICAR" y TERMINAR.
- PASO 2. Hacer  $I = 1$ .
- PASO 3. Mientras  $I \leq N$ , repetir los pasos 4 a 12.
- PASO 4. Hacer  $J = 1$ .
- PASO 5. Mientras  $J \leq M1$ , repetir los pasos 6 a 11.
- PASO 6. Hacer  $C(I, J) = 0$ .
- PASO 7. Hacer  $K = 1$ .
- PASO 8. Mientras  $K \leq M$ , repetir los pasos 9 y 10.
- PASO 9. Hacer
- $$C(I, J) = C(I, J) + A(I, K) * B(K, J).$$
- PASO 10. Hacer  $K = K + 1$ .
- PASO 11. Hacer  $J = J + 1$ .
- PASO 12. Hacer  $I = I + 1$ .
- PASO 13. IMPRIMIR las matrices  $A$ ,  $B$  y  $C$  y TERMINAR.

### Ejemplo 3.4

Elaborar un programa para multiplicar matrices, utilizando el algoritmo 3.1

### Solución

Ver el PROGRAMA 3.1 del CD.



**SUGERENCIA:** Este material puede complementarse e incluso enriquecerse si se cuenta con un pizarrón electrónico, por ejemplo el Mathcad, ya que permite, una vez entendida la mecánica de las operaciones matriciales, averiguar sus propiedades e incluso motivar algunas demostraciones. En adelante se hará referencia al Mathcad y a Matlab, pero puede usarse un software equivalente.

## MATRICES ESPECIALES

En una matriz cuadrada  $A$ , el conjunto de elementos en donde el primero y el segundo subíndices son iguales —es decir,  $i = j$ — forman la diagonal principal. Por ejemplo, en la matriz de  $4 \times 4$  que se da a continuación, los elementos dentro de la banda constituyen la **diagonal principal**.

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \\ a_{4,1} & a_{4,2} & a_{4,3} & a_{4,4} \end{bmatrix}$$

Una matriz de orden  $n$  con todos sus elementos debajo de la diagonal principal iguales a cero se llama **matriz triangular superior**. Si todos los elementos por encima de la diagonal principal son cero en una matriz, entonces será una **matriz triangular inferior**; en caso de que una matriz tenga únicamente ceros arriba y abajo de la diagonal principal, se tiene una **matriz diagonal** y, si en particular, todos los elementos de la diagonal son 1, entonces se obtiene la **matriz unitaria** o **matriz identidad**.

Matriz triangular superior

$$\begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} & \cdots & a_{1,n} \\ 0 & a_{2,2} & a_{2,3} & \cdots & a_{2,n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \cdots & a_{n-1,n} \\ 0 & 0 & \cdots & a_{n,n} \end{bmatrix}$$

Matriz triangular inferior

$$\begin{bmatrix} a_{1,1} & 0 & 0 & \cdots & 0 \\ a_{2,1} & a_{2,2} & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & 0 \\ a_{n,1} & a_{n,2} & a_{n,3} & \cdots & a_{n,n} \end{bmatrix}$$

Matriz diagonal

$$\begin{bmatrix} a_{1,1} & 0 & 0 & \cdots & 0 \\ 0 & a_{2,2} & 0 & \cdots & 0 \\ 0 & 0 & a_{3,3} & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & a_{n,n} \end{bmatrix}$$

Matriz unitaria o identidad

$$\begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

A continuación se dan algunos casos particulares de matrices cuadradas especiales

Triangular superior

$$\begin{bmatrix} 1 & 3 & -4 \\ 0 & 6 & 2 \\ 0 & 0 & -5 \end{bmatrix}$$

Triangular inferior

$$\begin{bmatrix} 4 & 0 & 0 \\ -2 & -1 & 0 \\ 7 & 5 & 3 \end{bmatrix}$$

Diagonal

$$\begin{bmatrix} 2 & 0 & 0 \\ 0 & -6 & 0 \\ 0 & 0 & 8 \end{bmatrix}$$

Unitaria

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



La matriz unitaria se denota, independientemente de su orden, como  $I$ .

Dada una matriz  $A$  de  $m \times n$ , la matriz de  $n \times m$  que se obtiene de  $A$  intercambiando sus filas por sus columnas se denomina **matriz transpuesta** de  $A$  y se denota por  $A^T$ . Esto es

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix}, \quad A^T = \begin{bmatrix} a_{1,1} & a_{2,1} & \cdots & a_{m,1} \\ a_{1,2} & a_{2,2} & \cdots & a_{m,2} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{1,n} & a_{2,n} & \cdots & a_{m,n} \end{bmatrix}$$

### Ejemplo 3.5

Dada la matriz  $A$ , encuentre su transpuesta.

$$A = \begin{bmatrix} 1 & 0 & 3 & 4 & 1 \\ 2 & 3 & 5 & 7 & 9 \\ 8 & 6 & 2 & 5 & 0 \end{bmatrix}$$

**Solución**

$3 \times 5$

$$A^T = \begin{bmatrix} 1 & 2 & 8 \\ 0 & 3 & 6 \\ 3 & 5 & 2 \\ 4 & 7 & 5 \\ 1 & 9 & 0 \end{bmatrix}$$

$5 \times 3$

Una matriz cuadrada para la que  $A^T = A$ , recibe el nombre de **matriz simétrica**. Por ejemplo

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

son iguales, y por tanto  $A$  es simétrica.

Si  $A$  y  $B$  son dos matrices cuadradas, tales que

$$AB = I = BA,$$

se dice que  $B$  es la **inversa** de  $A$  y se representa generalmente como  $A^{-1}$ .

### Ejemplo 3.6

Demuestre que  $B$  es la inversa de  $A$ , si:

$$A = \begin{bmatrix} 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{bmatrix} \quad \text{y} \quad B = \begin{bmatrix} 7 & -3 & -3 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}$$

## Solución

$$A B = \begin{bmatrix} 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{bmatrix} \begin{bmatrix} 7 & -3 & -3 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

Por tanto,

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

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus



```
A=[1 3 3; 1 4 3; 1 3 4]
B=[7 -3 -3; -1 1 0; -1 0 1]
disp('I = A * B')
I=A*B
```

```
[1,3,3;1,4,3;1,3,4]→a
[7,-3,-3;-1,1,0;-1,0,1]→b
a*b→I
```

En particular si  $A$  es diagonal; es decir,

$$A = \begin{bmatrix} a_{1,1} & 0 & 0 & \dots & 0 \\ 0 & a_{2,2} & 0 & \dots & 0 \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ 0 & 0 & \dots & a_{n,n} \end{bmatrix}, \text{ entonces } A^{-1} = \begin{bmatrix} 1/a_{1,1} & 0 & 0 & \dots & 0 \\ 0 & 1/a_{2,2} & 0 & \dots & 0 \\ \cdot & \cdot & & & \cdot \\ \cdot & \cdot & & & \cdot \\ 0 & 0 & \dots & 1/a_{n,n} \end{bmatrix}$$

La demostración se deja como ejercicio para el lector.

Es importante señalar que no todas las matrices tienen inversa. Si una matriz la tiene, se dice también que es **no singular**, y **singular** en caso contrario.

Más adelante se ven métodos para encontrar la inversa de una matriz.

### MATRIZ PERMUTADORA

Una matriz, cuyos elementos son ceros y unos, y donde sólo hay un uno por cada fila o columna, se conoce como **matriz permutadora** o **intercambiadora**; por ejemplo, las matrices

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

son casos particulares de matrices intercambiadoras.

El efecto de multiplicar una matriz permutadora  $P$  por una matriz  $A$  en ese orden es intercambiar las filas de  $A$ ; al multiplicar en orden inverso, se intercambian las columnas de  $A$ .

**Ejemplo 3.7**

Multiplique la matriz  $A$  del ejemplo 3.6 por la matriz permutadora  $P$  de  $3 \times 3$  dada arriba.

**Solución** a) Cálculo de  $PA$ :

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

$P \qquad \qquad \qquad A \qquad \qquad \qquad C$

Obsérvese que la matriz producto  $C$  es la matriz  $A$  con la primera y segunda filas intercambiadas

b) Cálculo de  $AP$

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

$A \qquad \qquad \qquad P \qquad \qquad \qquad D$

Obsérvese que la matriz producto  $D$  es la matriz  $A$  con la primera y segunda columnas intercambiadas.

Para realizar los cálculos puede usarse el siguiente guión de Matlab o la TI-92 Plus:



```
A=[1 3 3; 1 4 3; 1 3 4]
P=[0 1 0; 1 0 0; 0 0 1]
disp('C=P * A')
C=P*A
disp('D=A * P')
D=A*P
```

```
[1,3,3;1,4,3;1,3,4]→a
[0,1,0;1,0,0;0,0,1]→p
p*a→c
a*p→d
```

La matriz identidad es un caso particular de matriz permutadora y su efecto es dejar igual la matriz por la que se multiplica (ya sea por la derecha o por la izquierda). Este hecho, junto con el ejemplo 3.7, manifiesta que cuando aparece un 1 en la diagonal principal de una matriz permutadora, la fila o columna correspondiente de la matriz por la que se multiplica no sufre cambio alguno.

Véase que hay un 1 en la posición (3, 3) de la matriz  $P$  y que la fila 3 y la columna 3 de  $A$  no sufrieron intercambio en los incisos a) y b), respectivamente, en el ejemplo 3.7.

**Ejemplo 3.8**

Sin multiplicar diga qué efecto tendrá sobre una matriz cualquiera  $A$  de  $4 \times 4$  la siguiente matriz:

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



**Solución** **Análisis de la multiplicación  $PA$ .** Los unos en las posiciones (1, 1) y (3, 3) indican que las filas 1 y 3 de  $A$  no sufrirán efecto alguno. Por otro lado, los unos de la segunda y cuarta filas cuyas posiciones son (2, 4) y (4, 2) indican que las filas 2 y 4 de  $A$  se intercambiarán (nótese que en el ejemplo 3.7, los unos fuera de la diagonal ocupan las posiciones (1, 2) y (2, 1) y las filas 1 y 2 se intercambian).

El lector puede generalizar estos resultados de manera muy sencilla.

## 3.2 Vectores

Las matrices donde  $m > 1$  y  $n = 1$  (es decir, están formadas por una sola columna) son llamadas matrices columna o vectores. De igual manera, si  $m = 1$  y  $n > 1$ , se tiene una matriz fila o vector. Los vectores se denotarán con las letras minúsculas en negritas: **a**, **b**, **x**, etc. En estos casos no será necesaria la utilización de doble subíndice para la identificación de sus elementos y un vector **x** de  $m$  elementos (en columna) queda simplemente como:

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}$$

Un vector **y** de  $n$  elementos (en fila) queda como:

$$\mathbf{y} = [y_1 \ y_2 \ \dots \ y_n]$$

Por ejemplo, los siguientes vectores están en columna:

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 3 \\ 1 \\ 0 \\ 5 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

y estos en fila

$$[0 \ 1 \ 0], [3 \ 5 \ 7 \ 2], [0 \ 0 \ 0 \ 0 \ 0].$$

Obsérvese que si se tiene un vector columna, la transpuesta será un vector fila y vice-versa.

$$\text{Dado } \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \quad \mathbf{x}^T = [x_1 \ x_2 \ \dots \ x_m]$$

### Ejemplo 3.9

Obtener la transpuesta de los vectores columna y fila dados arriba.

## Solución

$$\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}^T = [1 \ 0 \ 0], \quad \begin{bmatrix} 3 \\ 1 \\ 0 \\ 5 \end{bmatrix}^T = [3 \ 1 \ 0 \ 5]$$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}^T = [00000]$$

$$[0 \ 1 \ 0]^T = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \quad [3 \ 5 \ 7 \ 2]^T = \begin{bmatrix} 3 \\ 5 \\ 7 \\ 2 \end{bmatrix}$$

$$[0 \ 0 \ 0 \ 0 \ 0]^T = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Como en el texto resulta generalmente difícil expresar un vector en columna, se usará algunas veces su transpuesta.

## MULTIPLICACIÓN DE VECTORES

Dado que los vectores son sólo casos particulares de las matrices, siguen las mismas reglas de multiplicación que éstas. Sea por ejemplo  $\mathbf{a} = [a_1 \ a_2 \ \dots \ a_n]$  y  $\mathbf{b}^T = [b_1 \ b_2 \ \dots \ b_n]$ , el producto  $\mathbf{a} \mathbf{b}$  es:

$$\mathbf{a} \mathbf{b} = \begin{matrix} [a_1 \ a_2 \ \dots \ a_n] \\ 1 \times n \end{matrix} \begin{matrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\ n \times 1 \end{matrix} = \begin{matrix} a_1 b_1 + a_2 b_2 + \dots + a_n b_n \\ 1 \times 1 \end{matrix}$$

El producto de  $\mathbf{a}$  por  $\mathbf{b}$  es el número real  $a_1 b_1 + a_2 b_2 + \dots + a_n b_n$ , que también puede verse como una matriz de  $1 \times 1$ .

Multiplicando en orden inverso

$$\mathbf{b} \mathbf{a} = \begin{matrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\ n \times 1 \end{matrix} \begin{matrix} [a_1 \ a_2 \ \dots \ a_n] \\ 1 \times n \end{matrix} = \begin{matrix} \begin{bmatrix} b_1 a_1 & b_1 a_2 & \dots & b_1 a_n \\ b_2 a_1 & b_2 a_2 & \dots & b_2 a_n \\ \vdots & \vdots & & \vdots \\ b_n a_1 & b_n a_2 & \dots & b_n a_n \end{bmatrix} \\ n \times n \end{matrix}$$

se obtiene una matriz de  $n \times n$ .

**Ejemplo 3.10**

Dados  $\mathbf{a} = [1 \ 5 \ 7]$  y  $\mathbf{b}^T = [0 \ -2 \ 3]$ , obtener  $\mathbf{ab}$  y  $\mathbf{ba}$

**Solución**

$$\mathbf{a} \mathbf{b} = \begin{matrix} [1 \ 5 \ 7] \\ 1 \times 3 \end{matrix} \begin{matrix} \begin{bmatrix} 0 \\ -2 \\ 3 \end{bmatrix} \\ 3 \times 1 \end{matrix} = 1(0) + 5(-2) + 7(3) = 11$$

y

$$\mathbf{b} \mathbf{a} = \begin{bmatrix} 0 \\ -2 \\ 3 \end{bmatrix} [1 \ 5 \ 7] = \begin{bmatrix} 0(1) & 0(5) & 0(7) \\ -2(1) & -2(5) & -2(7) \\ 3(1) & 3(5) & 3(7) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ -2 & -10 & -14 \\ 3 & 15 & 21 \end{bmatrix}$$

Puede multiplicarse también un vector por una matriz y viceversa si las dimensiones son adecuadas.

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
a=[1 5 7]
b=[0; -2; 3]
ab=a*b
ba=b*a
```

```
[1,5,7]→a
[0;-2,3]→b
a*b→ab
b*a→ba
```

**Ejemplo 3.11**

Multiplique el vector  $\mathbf{a} = [1 \ -2 \ 3]$  por la matriz  $\mathbf{B} = \begin{bmatrix} 0 & 4 & 3 \\ -1 & 8 & 2 \\ 3 & 1 & 5 \end{bmatrix}$

**Solución**

$$\begin{matrix} \mathbf{a} \\ [1 \ -2 \ 3] \\ 1 \times 3 \end{matrix} \begin{matrix} \mathbf{B} \\ \begin{bmatrix} 0 & 4 & 3 \\ -1 & 8 & 2 \\ 3 & 1 & 5 \end{bmatrix} \\ 3 \times 3 \end{matrix} = \begin{matrix} \mathbf{c} \\ [11 \ -9 \ 14] \\ 1 \times 3 \end{matrix}$$

los elementos de  $\mathbf{c}$  se calculan como:

$$\begin{aligned} 1(0) &+ (-2)(-1) + 3(3) = 11 \\ 1(4) &+ (-2)(8) + 3(1) = -9 \\ 1(3) &+ (-2)(2) + 3(5) = 14 \end{aligned}$$

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.





```
a=[1 -2 3]
B=[0 4 3; -1 8 2; 3 1 5]
C=a*B
```

```
[1,-2,3]→a
[0,4,3;-1,8,2;3,1,5]→B
a*B→C
```

Efectuar la multiplicación en orden inverso ( $B \mathbf{a}$ ) no es posible, por no ser conformes en ese orden. En cambio, sí puede multiplicarse  $B$  por algún vector columna  $\mathbf{d}$  de tres elementos. Así:

$$\begin{array}{ccc} \begin{bmatrix} 0 & 4 & 3 \\ -1 & 8 & 2 \\ 3 & 1 & 5 \end{bmatrix} & \begin{bmatrix} 1 \\ 0 \\ 2 \end{bmatrix} & = \begin{bmatrix} 0(1) + 4(0) + 3(2) \\ -1(1) + 8(0) + 2(2) \\ 3(1) + 1(0) + 5(2) \end{bmatrix} = \begin{bmatrix} 6 \\ 3 \\ 13 \end{bmatrix} \\ 3 \times 3 & 3 \times 1 & 3 \times 1 \end{array}$$

### PRODUCTO PUNTO DE VECTORES

**Definición.** Dados dos vectores  $\mathbf{a}$  y  $\mathbf{b}$  con igual número de elementos, por ejemplo  $n$ , su producto punto (o escalar), denotado por  $\mathbf{a} \cdot \mathbf{b}$ , es un número real obtenido de la siguiente manera

$$\mathbf{a} \cdot \mathbf{b} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n \quad (3.13)$$

### Ejemplo 3.12

Si  $\mathbf{a} = \begin{bmatrix} 2 \\ 1 \\ 6 \end{bmatrix}$  y  $\mathbf{b} = \begin{bmatrix} -3 \\ 0 \\ 2.5 \end{bmatrix}$ , obtenga el producto punto.

### Solución

$$\mathbf{a} \cdot \mathbf{b} = 2(-3) + 1(0) + 6(2.5) = 9$$



Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
a=[2 1 6]
b=[-3; 0; 2.5]
ab=a*b
```

```
[2,1,6]→a
[-3;0;2.5]→b
a*b→ab
```

Este producto punto así definido tiene las siguientes propiedades

$$a) \mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a} \text{ conmutatividad.} \quad (3.14)$$

$$b) (\mathbf{a} + \mathbf{b}) \cdot \mathbf{c} = \mathbf{a} \cdot \mathbf{c} + \mathbf{b} \cdot \mathbf{c} \text{ distributividad.} \quad (3.15)$$

$$c) (\alpha \mathbf{a}) \cdot \mathbf{b} = \alpha (\mathbf{a} \cdot \mathbf{b}) \text{ para cualquier número real } \alpha. \text{ Asociatividad.} \quad (3.16)$$

$$d) \mathbf{a} \cdot \mathbf{a} \geq 0 \text{ y } \mathbf{a} \cdot \mathbf{a} = 0 \text{ si y sólo si } \mathbf{a} = \mathbf{0}. \text{ Positividad de la definición.} \quad (3.17)$$

Sólo se demostrará la propiedad (a) y se dejarán las restantes como ejercicio para el lector.

Demostración de (a)

$$\mathbf{a} \cdot \mathbf{b} = a_1 b_1 + a_2 b_2 + \dots + a_n b_n$$

y

$$\mathbf{b} \cdot \mathbf{a} = b_1 a_1 + b_2 a_2 + \dots + b_n a_n$$

Por la conmutatividad de la multiplicación de los números reales se tiene que

$$a_i b_i = b_i a_i, \quad 1 \leq i \leq n$$

y por tanto

$$\mathbf{a} \cdot \mathbf{b} = \mathbf{b} \cdot \mathbf{a}$$

Enseguida se definirán conceptos tan importantes como la longitud de un vector, el ángulo entre dos vectores cualesquiera y distancia entre vectores en función del producto punto.

Cada una de estas ideas tiene un significado bien definido en los vectores de dos elementos en la geometría analítica, y es razonable pedir que cualquier definición que se adopte se reduzca a la ya conocida. Con esto en mente, se pueden obtener definiciones aceptables extendiendo las fórmulas correspondientes de la geometría analítica a vectores de  $n$  elementos.

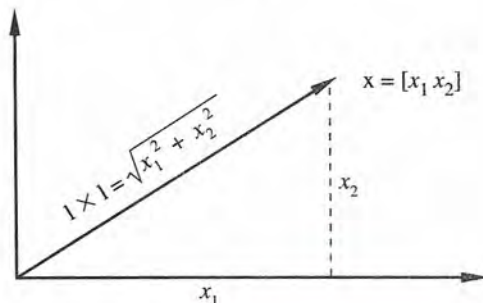
### LONGITUD DE UN VECTOR

La noción de longitud para vectores de dos elementos está dada por la siguiente definición:

Sea  $\mathbf{x}$  un vector cualquiera de dos elementos, su longitud denotada por  $|\mathbf{x}|$  es el número real no negativo\*

$$|\mathbf{x}| = \sqrt{x_1^2 + x_2^2} \quad (3.18)$$

Gráficamente se representa así



**Figura 3.1**  
Interpretación  
gráfica de la  
longitud de un  
vector.

\* Se dice que un número real es no negativo cuando sólo puede ser cero o positivo.

La ecuación 3.18 puede escribirse en términos del producto punto como:

$$| \mathbf{x} | = \sqrt{\mathbf{x} \cdot \mathbf{x}} \quad (3.19)$$

lo cual está bien definido para vectores de  $n$  elementos y puede, por tanto, tomarse como longitud de estos últimos.

**Definición.** La longitud (o norma) de un vector  $\mathbf{x}$  de  $n$  componentes, con  $n \geq 1$ , está dada por el número real no negativo.\*

$$\begin{aligned} | \mathbf{x} | &= \sqrt{\mathbf{x} \cdot \mathbf{x}} \\ | \mathbf{x} | &= \sqrt{x_1^2 + x_2^2 + \dots + x_n^2} \end{aligned} \quad (3.20)$$

### Ejemplo 3.13

Si  $\mathbf{a} = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix}$ , encuentre su norma.

### Solución

$$| \mathbf{a} | = \sqrt{25 + 9 + 16} = 7.0711$$

Para realizar los cálculos puede usarse el siguiente guión de Matlab o la TI-92 Plus.



```
a=[5 3 4]
Norma=norm(a)
```



```
[5,3,4]→a
norm(a)→norma
```

## ÁNGULO ENTRE VECTORES

Hay que recordar que si se tienen dos vectores de dos componentes, ambos distintos del vector cero, la fórmula

$$\cos \theta = \frac{\mathbf{x} \cdot \mathbf{y}}{| \mathbf{x} | | \mathbf{y} |} \quad 0 \leq \theta \leq \pi \quad (3.21)$$

es una consecuencia inmediata de la ley de los cosenos. Como la expresión

$$\frac{\mathbf{x} \cdot \mathbf{y}}{| \mathbf{x} | | \mathbf{y} |},$$

está bien definida para vectores distintos del vector cero, de  $n$  componentes, parece conveniente usarla como definición del ángulo entre vectores de más de dos componentes. Sin embargo, sería necesario probar primero que el rango o codominio de esta expresión —usando vectores  $\mathbf{x}$ ,  $\mathbf{y}$  de  $n$  componentes— es el intervalo cerrado  $[-1, 1]$ , para que así se guarde consistencia con el primer miembro de la ecuación 3.21.\*\*

\* Se conoce también como **norma euclídeana** y algunos autores la representan por  $L_2$ .

\*\* Recuerdese que la función  $\cos$  tiene como rango el intervalo  $[-1, 1]$ .



La demostración está fuera de los objetivos de este libro, pero el lector interesado puede encontrarla en Kreider *et. al.*\*

**Definición.** Si  $\mathbf{x}$  y  $\mathbf{y}$  son vectores distintos del vector 0, con  $n$  componentes, el coseno del ángulo entre ellos se define como:

$$\cos \theta = \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}| |\mathbf{y}|}$$

Si alguno de los vectores es el vector cero, se hace  $\cos \theta$  igual a cero.

### Ejemplo 3.14

Si  $\mathbf{x}^T = [2 \ -3 \ 4 \ 1]$  y  $\mathbf{y}^T = [-1 \ 2 \ 4 \ 2]$ , calcule el ángulo entre ellos.

### Solución

$$\cos \theta = \frac{2(-1) + (-3)(2) + 4(4) + 1(2)}{\sqrt{4 + 9 + 16 + 1} \sqrt{1 + 4 + 16 + 4}} = 0.3651$$

de donde  $\theta = 68.58$

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
x=[2 -3 4 1]
y=[-1 2 4 2]
ct=(x*y')/(norm(x)*norm(y))
teta=acos(ct)/pi*180
```

```
[2,-3,4,1]→x
[-1,2,4,2]→y
dotp(x,y)/(norm(x)*norm(y))→ct
cos-1(ct)/π*180→teta
```

### DISTANCIA ENTRE DOS VECTORES

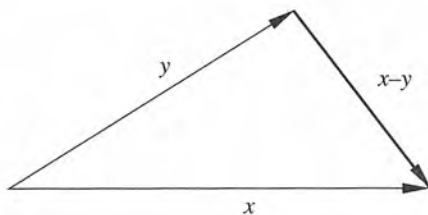
Uno de los tres conceptos que aún no se analiza es el de distancia entre dos vectores de  $n$  componentes. De nueva cuenta esto se hará “copiando” la definición dada en la geometría analítica, donde la distancia entre  $\mathbf{x}$  y  $\mathbf{y}$  es la longitud del vector  $(\mathbf{x} - \mathbf{y})$  (véase Fig. 3.2).

**Definición.** La distancia entre dos vectores  $\mathbf{x}$  y  $\mathbf{y}$  de  $n$  componentes es

$$d(\mathbf{x}, \mathbf{y}) = |\mathbf{x} - \mathbf{y}| \quad (3.22)$$

definición que satisface las siguientes propiedades

**Figura 3.2**  
Resta de  
vectores en el  
plano.



\* Kreider, Kuller, Ostberg, Perkins. *An Introduction to Linear Analysis*. Addison-Wesley (1966).

- a) La distancia entre dos vectores es un número real no negativo que es cero si y sólo si se trata del mismo vector; es decir,

$$d(\mathbf{x}, \mathbf{y}) \geq 0 \text{ y } d(\mathbf{x}, \mathbf{y}) = 0 \text{ si y sólo si } \mathbf{x} = \mathbf{y} \quad (3.23)$$

- b) Es independiente del orden en que se tomen los vectores; esto es

$$d(\mathbf{x}, \mathbf{y}) = d(\mathbf{y}, \mathbf{x})$$

- c) Finalmente, satisface la desigualdad del triángulo, conocida en la geometría en los términos: **la suma de las longitudes de los catetos de un triángulo es mayor o igual a la longitud de la hipotenusa**; esto es

$$d(\mathbf{x}, \mathbf{y}) + d(\mathbf{y}, \mathbf{z}) \geq d(\mathbf{x}, \mathbf{z})$$

para tres vectores cualesquiera  $\mathbf{x}$ ,  $\mathbf{y}$  y  $\mathbf{z}$ .

### Ejemplo 3.15

Calcule la distancia entre  $\mathbf{x}$  y  $\mathbf{y}$  dadas por:

$$\mathbf{x}^T = [0 \ 3 \ 5 \ 1], \quad \mathbf{y}^T = [-2 \ 1 \ -3 \ 1]$$

**Solución** Primero se obtiene  $\mathbf{x} - \mathbf{y}$

$$\mathbf{x} - \mathbf{y} = \begin{bmatrix} 0 \\ 3 \\ 5 \\ 1 \end{bmatrix} - \begin{bmatrix} -2 \\ 1 \\ -3 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 2 \\ 8 \\ 0 \end{bmatrix}$$

La norma de este vector es:

$$\|\mathbf{x} - \mathbf{y}\| = \sqrt{2^2 + 2^2 + 8^2 + 0^2} = \sqrt{72} = 8.4853,$$

y, por tanto, la distancia entre  $\mathbf{x}$  y  $\mathbf{y}$  es 8.4853 unidades de longitud.

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
x=[0 3 5 1]
y=[-2,1,-3,1]
dist=norm(x-y)
```

```
[0,3,5,1]→x
[-2,1,-3,1]→y
norm(x-y)→dist
```

Obsérvese que ninguno de estos tres conceptos tiene representación geométrica cuando el número de componentes de los vectores es mayor de tres.

**SUGERENCIA:** Explore con Mathcad, Matlab o algún software disponible, las operaciones vistas y sus propiedades.

### 3.3 Independencia y ortogonalización de vectores

Una expresión de la forma

$$\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \dots + \alpha_n \mathbf{x}_n, \quad (3.26)$$

donde  $\alpha_1, \alpha_2, \dots, \alpha_n$  son números reales y  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  son vectores de  $m$  elementos cada uno, se llama combinación lineal de los vectores  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ .

#### Ejemplo 3.16

¿La expresión

$$2.5 \begin{bmatrix} 1 \\ 0 \\ 4 \\ 3 \end{bmatrix} + 3 \begin{bmatrix} -4 \\ 2 \\ 1.6 \\ 5 \end{bmatrix} + (-7) \begin{bmatrix} 5 \\ -2 \\ 0 \\ 1 \end{bmatrix}$$

es una combinación lineal?

**Solución** Sí; es una combinación lineal de  $[1 \ 0 \ 4 \ 3]^T$ ,  $[-4 \ 2 \ 1.6 \ 5]^T$  y  $[5 \ -2 \ 0 \ 1]^T$ , con los escalares 2.5, 3 y -7, respectivamente.

A menudo los elementos de un vector  $\mathbf{x}_i$  de una combinación lineal, tendrán dos subíndices; el primero indica la fila a que pertenece y el segundo se refiere al vector a que corresponde, así

$$\mathbf{x}_i = \begin{bmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{mi} \end{bmatrix}$$

Se dice que un vector  $\mathbf{x} = [x_1, x_2, \dots, x_m]^T$ , depende linealmente de un conjunto de vectores de  $m$  elementos  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , si se pueden encontrar escalares  $\alpha_1, \alpha_2, \dots, \alpha_n$ , tales que se cumpla la siguiente ecuación vectorial

$$\mathbf{x} = \alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2 + \dots + \alpha_n \mathbf{x}_n \quad (3.27)$$

Si, por el contrario, no existen escalares que satisfagan tal ecuación,  $\mathbf{x}$  es un vector linealmente independiente de  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ . En otras palabras,  $\mathbf{x}$  es linealmente dependiente de  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  si y sólo si  $\mathbf{x}$  es una combinación lineal de  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ .

#### Ejemplo 3.17

Dado el conjunto de dos vectores de dos elementos:

$$\mathbf{x}_1 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \quad \text{y} \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 2 \end{bmatrix},$$

demuestre que el vector  $\mathbf{x}^T = [0 \ 8]^T$  es linealmente dependiente de dicho conjunto.

**Solución** Es suficiente encontrar dos escalares  $\alpha_1$  y  $\alpha_2$  tales que la combinación  $\alpha_1 \mathbf{x}_1 + \alpha_2 \mathbf{x}_2$  reproduzca a  $\mathbf{x}$ . Por observación se advierte que los números  $\alpha_1 = 1$  y  $\alpha_2 = 2$  cumplen este requisito.



$$\begin{bmatrix} 0 \\ 8 \end{bmatrix} = (1) \begin{bmatrix} 4 \\ 4 \end{bmatrix} + (2) \begin{bmatrix} -2 \\ 2 \end{bmatrix}$$

Generalmente, encontrar los escalares o la demostración de que no existen es un problema difícil que requiere una técnica específica, misma que se desarrolla más adelante.

### INDEPENDENCIA DE CONJUNTOS DE VECTORES

Un conjunto de vectores dado  $y_1, y_2, \dots, y_n$ , es linealmente dependiente si por lo menos uno de ellos es combinación lineal de alguno o todos los vectores restantes. Si ninguno lo es, se dice que es un conjunto linealmente independiente.

#### Ejemplo 3.18

Sea el siguiente conjunto de cuatro vectores de tres elementos cada uno.

$$y_1 = \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix}, \quad y_2 = \begin{bmatrix} 0.1 \\ -3 \\ 0 \end{bmatrix}, \quad y_3 = \begin{bmatrix} 0.5 \\ -15 \\ 0 \end{bmatrix}, \quad y_4 = \begin{bmatrix} 0.03 \\ -0.9 \\ 0 \end{bmatrix}$$

Determine si es linealmente dependiente o independiente.

**Solución** Este conjunto es linealmente dependiente, ya que  $y_3$  se obtiene de la combinación

$$y_3 = 5y_2 = 5 \begin{bmatrix} 0.1 \\ -3 \\ 0 \end{bmatrix},$$

y  $y_4$  se obtiene de combinar  $y_1$  y  $y_2$  en la siguiente forma

$$\begin{bmatrix} 0.03 \\ -0.9 \\ 0 \end{bmatrix} = 0 \begin{bmatrix} 1 \\ 4 \\ 2 \end{bmatrix} + 0.3 \begin{bmatrix} 0.1 \\ -3 \\ 0 \end{bmatrix}$$

Si se considera el conjunto formado sólo por  $y_1$  y  $y_2$ , se tiene que es linealmente independiente, ya que ninguno se obtiene multiplicando al otro por algún escalar.

Cualquier conjunto que tenga el vector cero (vector cuyos componentes son todos cero) como uno de sus elementos, es linealmente independiente, ya que dicho vector podrá obtenerse siempre de cualquier otro vector del conjunto por la combinación

$$0 = \begin{bmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix} = 0 \begin{bmatrix} x_{1,i} \\ x_{2,i} \\ \cdot \\ \cdot \\ \cdot \\ x_{n,i} \end{bmatrix}$$

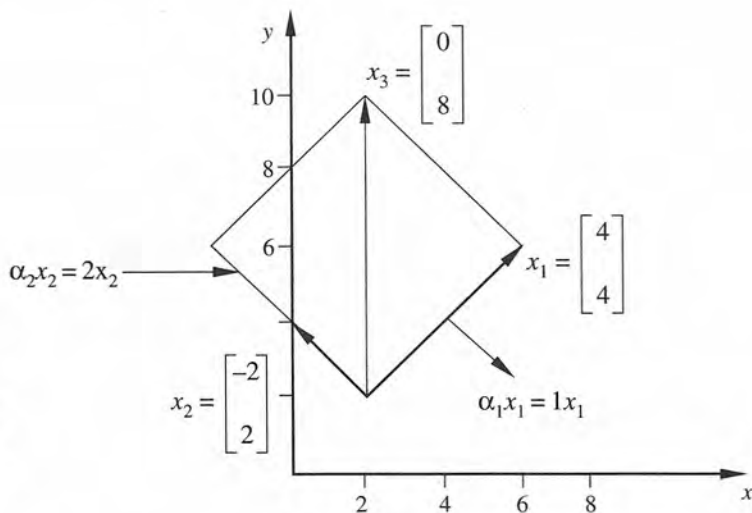
Un conjunto formado por un solo vector (distinto de  $0$ ) es linealmente independiente.

### INTERPRETACIÓN GEOMÉTRICA DE LA INDEPENDENCIA LINEAL

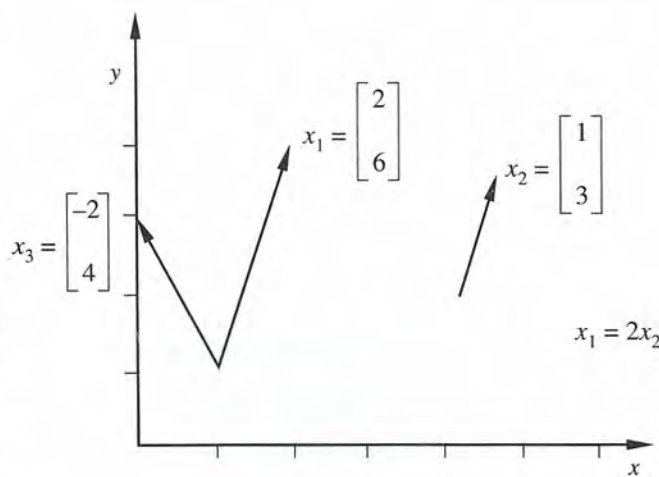
Es conveniente estudiar la independencia lineal desde el punto de vista geométrico, aunque esto sólo valga para vectores de dos y tres componentes. Considérense los tres vectores del ejemplo 3.17 en el plano  $x$ - $y$  (Fig. 3.3). Por la geometría se sabe que dos vectores que se cortan forman un plano (por ejemplo  $\mathbf{x}_1$  y  $\mathbf{x}_2$  forman el plano  $x$ - $y$ ). Por tanto, es natural pensar que si se tiene un tercer vector del plano  $x$ - $y$ , éste pueda obtenerse de alguna combinación de los que se cortan, por ejemplo  $\mathbf{x}_3$  de  $\mathbf{x}_1$  y  $\mathbf{x}_2$ , aplicando la ley del paralelogramo.

Si, por otro lado, se tienen dos vectores de dos componentes linealmente dependientes, esto se manifiesta geométricamente como paralelismo (véanse los vectores  $\mathbf{x}_1$  y  $\mathbf{x}_2$  de la Fig. 3.4). Es evidente que estos vectores paralelos no forman un plano y un tercer vector  $\mathbf{x}_3$  que no sea paralelo a ellos no podrá generarse con una combinación lineal de  $\mathbf{x}_1$  y  $\mathbf{x}_2$ .

En conclusión, la característica geométrica de dos vectores linealmente independientes es que se cortan en un punto. En cambio, dos vectores linealmente dependientes son paralelos.



**Figura 3.3**  
Interpretación  
geométrica de  
independencia  
lineal en el  
plano.



**Figura 3.4**  
Interpretación  
geométrica de  
dependencia  
lineal en el  
plano.

**CONJUNTOS ORTOGONALES DE VECTORES**

Dos vectores de igual número de componentes son ortogonales o perpendiculares si el coseno del ángulo entre ellos es cero. De acuerdo con esta definición, el vector cero es ortogonal con cualquier otro vector; en general,  $\mathbf{x}$  y  $\mathbf{y}$  son ortogonales si y sólo si

$$\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = 0,$$

derivada esta expresión del hecho que

$$\cos \theta = \frac{\mathbf{x} \cdot \mathbf{y}}{|\mathbf{x}| |\mathbf{y}|}$$

A continuación se generaliza la definición de ortogonalidad.

Un conjunto de vectores  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  forma un conjunto ortogonal si  $\mathbf{x}_i \neq \mathbf{0}$  para  $1 \leq i \leq n$ , y

$$\mathbf{x}_i \cdot \mathbf{y}_j = 0 \quad 1 \leq j \leq n \quad (3.28)$$

siempre que  $i \neq j$ .

**Ejemplo 3.19**

Determine si los vectores  $\mathbf{x}_1$  y  $\mathbf{x}_2$  del ejemplo 3.17 son ortogonales.

**Solución**

$$\mathbf{x}_1 \cdot \mathbf{x}_2 = \begin{bmatrix} 4 \\ 4 \end{bmatrix} \cdot \begin{bmatrix} -2 \\ 2 \end{bmatrix} = -8 + 8 = 0$$

Son perpendiculares en el sentido usual del término (véase Fig. 3.3) y esto es lo que significa la definición, dada para cualquier número de componentes.

**Ejemplo 3.20**

¿El conjunto siguiente es ortogonal?

$$\mathbf{x}_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

**Solución** Sí, ya que

$$\mathbf{x}_1 \cdot \mathbf{x}_2 = \mathbf{x}_1 \cdot \mathbf{x}_3 = \mathbf{x}_2 \cdot \mathbf{x}_3 = 0$$

En cambio, si se adiciona a este conjunto el vector

$$\mathbf{x}_4 = \begin{bmatrix} 2 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

el conjunto resultante  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$  no es ortogonal, pues

$$\mathbf{x}_4 \cdot \mathbf{x}_2 = 1 \neq 0$$



**Ejemplo 3.21**

Corrobore si el siguiente conjunto de vectores es ortogonal

$$\mathbf{x}_1 = \begin{bmatrix} -3 \\ 4 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 2 \\ 2 \\ -2.0003 \end{bmatrix}$$

**Solución**

$$\mathbf{x}_1 \cdot \mathbf{x}_2 = (-3)(2) + 4(2) + 1(-2.0003) = -0.0003$$

Obsérvese que los vectores son “casi” ortogonales. Esto ocurre con frecuencia y en los cálculos prácticos será preciso decidir con qué cercanía a cero se aceptará que un producto punto de dos vectores “es cero” y, por tanto, que los vectores son ortogonales. De nuevo  $\varepsilon$  denotará el límite de aceptación o de rechazo. El valor que tome  $\varepsilon$  estaría en función del instrumento con que se lleven a cabo los cálculos. Por ejemplo, para una calculadora de nueve dígitos de exactitud  $\varepsilon$  puede ser  $10^{-4}$ . Con  $\varepsilon = 10^{-4}$  los vectores de este ejemplo no son ortogonales. Así pues,  $\varepsilon$  usado de esta manera puede llamarse **criterio de ortogonalidad**.

**ORTOGONALIZACIÓN**

Se ha llegado al punto central de esta sección, donde es posible construir un conjunto de vectores ortogonales (ortogonalización) a partir de un conjunto de vectores linealmente independientes. Enseguida se considerará uno de los métodos más difundidos, la ortogonalización de Gram-Schmidt, aunque pueda representar ciertas dificultades computacionales.

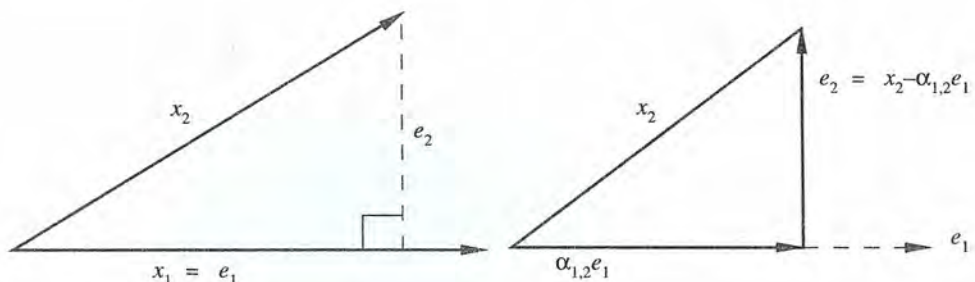
**MÉTODO DE GRAM-SCHMIDT**

En lugar de empezar con el caso más general, se introducirá el proceso de ortogonalización con dos ejemplos; el primero se tiene cuando se toman dos vectores  $\mathbf{x}_1$  y  $\mathbf{x}_2$  del plano  $x$ - $y$ , linealmente independientes y a partir de ellos se forma el conjunto ortogonal  $\mathbf{e}_1$  y  $\mathbf{e}_2$ . La figura 3.5 muestra la manera natural de resolver este caso; simplemente se toma  $\mathbf{e}_1 = \mathbf{x}_1$  y  $\mathbf{e}_2$  como la “componente” de  $\mathbf{x}_2$  perpendicular a  $\mathbf{x}_1$ . Así, se escribe  $\mathbf{e}_2$  en la forma

$$\mathbf{e}_2 = \mathbf{x}_2 - \alpha_{1,2}\mathbf{e}_1 \quad (3.29)$$

y sólo queda determinar  $\alpha_{1,2}$  de manera que la condición  $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$  se cumpla. Esto da la ecuación

$$\mathbf{e}_2 \cdot \mathbf{e}_1 = 0 = \mathbf{x}_2 \cdot \mathbf{e}_1 - \alpha_{1,2} \mathbf{e}_1 \cdot \mathbf{e}_1 \quad (3.30)$$



**Figura 3.5**  
Ortogonalización  
en el plano  $x$ - $y$ .

y finalmente:

$$\alpha_{1,2} = \frac{\mathbf{x}_2 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1} \quad (3.31)$$

De este modo  $\mathbf{e}_2$  queda determinado en función de  $\mathbf{x}_1$  y  $\mathbf{x}_2$ , y el conjunto  $\mathbf{x}_1, \mathbf{x}_2$  se ha ortogonalizado.

### Ejemplo 3.22

Ortogonalice  $\mathbf{x}_1 = [2 \ 2]^T$  y  $\mathbf{x}_2 = [3 \ 0]^T$

#### Solución

$$\mathbf{e}_1 = [2 \ 2]^T$$

y

$$\mathbf{e}_2 = \mathbf{x}_2 - \alpha_{1,2}\mathbf{e}_1$$

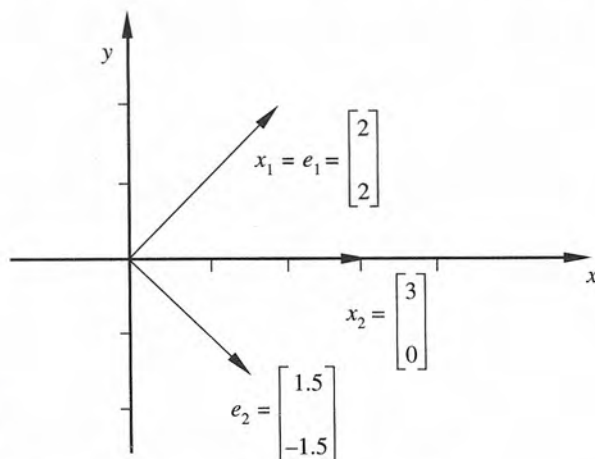
con

$$\alpha_{1,2} = \frac{[2 \ 2]^T \cdot [3 \ 0]^T}{[2 \ 2]^T \cdot [2 \ 2]^T} = \frac{6}{4+4} = \frac{3}{4}$$

Sustituyendo queda:

$$\mathbf{e}_2 = [3 \ 0]^T - \frac{3}{4} [2 \ 2]^T = [3 \ 0]^T - \left[\frac{3}{2} \ \frac{3}{2}\right]^T = [1.5 \ -1.5]^T$$

Al graficar estos vectores se obtiene la siguiente figura



**Figura 3.6**  
Ortogonalización  
de vectores.

Obsérvese la perpendicularidad de  $\mathbf{e}_1$  y  $\mathbf{e}_2$ .

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
x1=[2; 2]
x2=[3; 0]
e1=x1
alfa12=(e1'*x2)/(e1'*e1)
e2=x2-alfa12*e1
```

```
[2;2]→x1
[3;0]→x2
x1→e1
dotP(e1,x2)/dotP(e1,e1)→a12
x2-a12*e1→e2
```

Como segundo ejemplo se ortogonalizará el conjunto arbitrario  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$  de vectores lineales independientes de tres componentes. El procedimiento es esencialmente igual al que se usó antes, y se empieza escogiendo  $\mathbf{e}_1 = \mathbf{x}_1$ . El segundo paso es determinar  $\mathbf{e}_2$  de acuerdo con el par de ecuaciones

$$\mathbf{e}_2 \cdot \mathbf{e}_1 = 0, \mathbf{e}_2 = \mathbf{x}_2 - \alpha_{1,2} \mathbf{e}_1 \quad (3.32)$$

de las que se obtiene nuevamente que

$$\alpha_{1,2} = \frac{\mathbf{x}_2 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1} \quad (3.33)$$

Obsérvese que  $\mathbf{e}_2 \neq 0$ ; de lo contrario se cumpliría la primera de las ecuaciones 3.32 y en la segunda se tendría que  $\mathbf{x}_2 = \alpha_{1,2} \mathbf{e}_1 = \alpha_{1,2} \mathbf{x}_1$ . O sea que  $\mathbf{x}_2$  estaría en función de  $\mathbf{x}_1$ , lo cual es imposible por la independencia lineal de  $\mathbf{x}_1$  y  $\mathbf{x}_2$ .

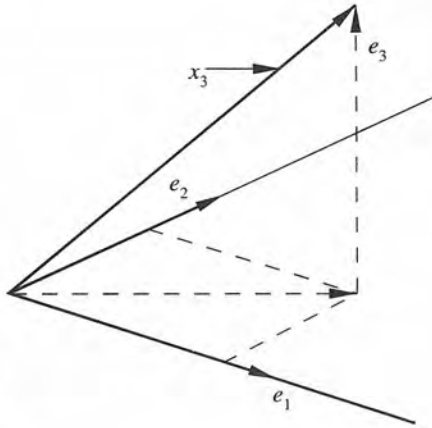
Para el tercer vector se recurre nuevamente a una representación geométrica, en donde se verá que el proceso de ortogonalización puede completarse tomando  $\mathbf{e}_3$  como la componente de  $\mathbf{x}_3$  perpendicular al plano formado por los vectores  $\mathbf{e}_1$  y  $\mathbf{e}_2$  (Fig.3.7).\*

De esto se tiene

$$\mathbf{e}_3 = \mathbf{x}_3 - \alpha_{1,3} \mathbf{e}_1 - \alpha_{2,3} \mathbf{e}_2 \quad (3.34)$$

y se puede encontrar  $\alpha_{1,3}$  y  $\alpha_{2,3}$  por medio de las condiciones de ortogonalidad

$$\mathbf{e}_1 \cdot \mathbf{e}_2 = \mathbf{e}_1 \cdot \mathbf{e}_3 = \mathbf{e}_2 \cdot \mathbf{e}_3 = 0$$



**Figura 3.7**  
Ortogonalización  
en el espacio  
x-y-z.

Multiplicando en forma punto los dos miembros de la ecuación 3.34 por  $\mathbf{e}_1$  y después por  $\mathbf{e}_2$ , se obtiene el par de ecuaciones

$$\begin{aligned} \mathbf{e}_3 \cdot \mathbf{e}_1 &= 0 = \mathbf{x}_3 \cdot \mathbf{e}_1 - \alpha_{1,3} \mathbf{e}_1 \cdot \mathbf{e}_1 - \alpha_{2,3} \mathbf{e}_2 \cdot \mathbf{e}_1 \\ \mathbf{e}_3 \cdot \mathbf{e}_2 &= 0 = \mathbf{x}_3 \cdot \mathbf{e}_2 - \alpha_{1,3} \mathbf{e}_1 \cdot \mathbf{e}_2 - \alpha_{2,3} \mathbf{e}_2 \cdot \mathbf{e}_2 \end{aligned} \quad (3.35)$$

o bien

$$\begin{aligned} \mathbf{x}_3 \cdot \mathbf{e}_1 &= \alpha_{1,3} \mathbf{e}_1 \cdot \mathbf{e}_1 \\ \mathbf{x}_3 \cdot \mathbf{e}_2 &= \alpha_{2,3} \mathbf{e}_2 \cdot \mathbf{e}_2 \end{aligned} \quad (3.36)$$

\* Recuérdese que dos líneas que se cortan solamente en un punto forman un plano.



resolviendo para  $\alpha_{1,3}$  y para  $\alpha_{2,3}$ , se tiene

$$\alpha_{1,3} = \frac{\mathbf{x}_3 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1}, \quad \alpha_{2,3} = \frac{\mathbf{x}_3 \cdot \mathbf{e}_2}{\mathbf{e}_2 \cdot \mathbf{e}_2}$$

y con esto termina la ortogonalización del conjunto  $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ .

### Ejemplo 3.23

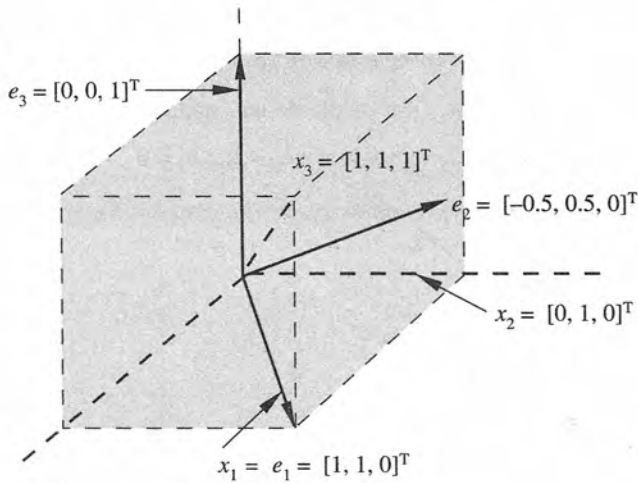
Ortogonalice los vectores

$$\mathbf{x}_1 = [1 \ 1 \ 0]^T, \quad \mathbf{x}_2 = [0 \ 1 \ 0]^T, \quad \text{y} \quad \mathbf{x}_3 = [1 \ 1 \ 1]^T$$

**Solución**

$$\mathbf{e}_1 = \mathbf{x}_1, \quad \mathbf{e}_2 = \mathbf{x}_2 - \alpha_{1,2} \mathbf{e}_1, \quad \text{y}$$

$$\mathbf{e}_3 = \mathbf{x}_3 - \alpha_{1,3} \mathbf{e}_1 - \alpha_{2,3} \mathbf{e}_2,$$



**Figura 3.8**  
Ortogonalización  
en el espacio.

donde  $\alpha_{1,2}$ ,  $\alpha_{1,3}$ , y  $\alpha_{2,3}$

se obtienen de las ecuaciones

$$\alpha_{1,2} = \frac{\mathbf{x}_2 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1}, \quad \alpha_{1,3} = \frac{\mathbf{x}_3 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1}, \quad \alpha_{2,3} = \frac{\mathbf{x}_3 \cdot \mathbf{e}_2}{\mathbf{e}_2 \cdot \mathbf{e}_2}$$

Al verificar los cálculos se llega a

$$\alpha_{1,2} = 1/2, \quad \alpha_{1,3} = 1, \quad \alpha_{2,3} = 0$$

y sustituyendo

$$\mathbf{e}_1 = [1 \ 1 \ 0]^T, \quad \mathbf{e}_2 = [-1/2 \ 1/2 \ 0]^T, \quad \mathbf{e}_3 = [0 \ 0 \ 1]^T$$

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```

x1=[1; 1; 0]
x2=[0; 1; 0]
x3=[1; 1; 1]
e1=x1
alfa12=(e1'*x2)/(e1'*e1)
e2=x2-alfa12*e1
alfa13=(e1'*x3)/(e1'*e1)
alfa23=(e2'*x3)/(e2'*e2)
e3=x3-alfa13*e1-alfa23*e2

```

```

[1;1;0]→x1
[0;1;0]→x2
[1;1;1]→x3

x1→e1
dotP(e1,x2)/dotP(e1,e1)→a12
x2-a12*e1→e2
dotP(e1,x3)/dotP(e1,e1)→a13
dotP(e2,x3)/dotP(e2,e2)→a23
x3-a13*e1-a23*e2→e3

```

Una vez realizado lo anterior, se puede pasar el caso general de ortogonalizar un conjunto de  $n$  vectores linealmente independientes  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  de  $n$  componentes cada uno. Primero se efectuará  $\mathbf{e}_1 = \mathbf{x}_1$ , después  $\mathbf{e}_2 = \mathbf{x}_2 - \alpha_{1,2} \mathbf{e}_1$ , donde  $\alpha_{1,2}$  se escoge de manera que  $\mathbf{e}_1 \cdot \mathbf{e}_2 = 0$ .

De aquí que

$$\alpha_{1,2} = \frac{\mathbf{x}_2 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1},$$

y la independencia lineal de  $\mathbf{x}_1$  y  $\mathbf{x}_2$  implica que  $\mathbf{e}_2 \neq \mathbf{0}$ .

Únicamente queda por demostrar que este proceso puede continuar hasta obtener un conjunto ortogonal  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$ . Para ello, supóngase que se llegó al conjunto ortogonal  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$  con  $m < n$ . Para continuar un paso más efectúese

$$\mathbf{e}_{m+1} = \mathbf{x}_{m+1} - \alpha_{1,m+1} \mathbf{e}_1 - \dots - \alpha_{m,m+1} \mathbf{e}_m$$

y determínese  $\alpha_{1,m+1}, \alpha_{2,m+1}, \dots, \alpha_{m,m+1}$ , de manera que  $\mathbf{e}_{m+1}$  sea ortogonal a cada elemento del conjunto  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_m$ . Consecuentemente el conjunto de ecuaciones es

$$\begin{aligned} \mathbf{x}_{m+1} \cdot \mathbf{e}_1 - \alpha_{1,m+1} (\mathbf{e}_1 \cdot \mathbf{e}_1) &= 0, \\ \mathbf{x}_{m+1} \cdot \mathbf{e}_2 - \alpha_{2,m+1} (\mathbf{e}_2 \cdot \mathbf{e}_2) &= 0, \\ &\vdots \\ \mathbf{x}_{m+1} \cdot \mathbf{e}_m - \alpha_{m,m+1} (\mathbf{e}_m \cdot \mathbf{e}_m) &= 0, \end{aligned}$$

y por tanto:

$$\alpha_{1,m+1} = \frac{\mathbf{x}_{m+1} \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1}, \quad \alpha_{2,m+1} = \frac{\mathbf{x}_{m+1} \cdot \mathbf{e}_2}{\mathbf{e}_2 \cdot \mathbf{e}_2}, \dots, \quad \alpha_{m,m+1} = \frac{\mathbf{x}_{m+1} \cdot \mathbf{e}_m}{\mathbf{e}_m \cdot \mathbf{e}_m}$$

que determinan  $\mathbf{e}_{m+1}$ . De nuevo, la independencia lineal de  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{m+1}$ , implica que  $\mathbf{e}_{m+1} \neq \mathbf{0}$ . Por tanto, el proceso de ortogonalización se ha aumentado en un paso y con el mismo argumento puede continuarse hasta tener  $m = n$ . Lo anterior queda condensado en el siguiente teorema.

**TEOREMA 3.1** Sean  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ , un conjunto de vectores linealmente independientes de  $n$  componentes cada uno. A partir de ellos se puede construir un conjunto ortogonal  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_n$  de la siguiente manera

$$\mathbf{e}_1 = \mathbf{x}_1 \quad (3.37)$$

y

$$\mathbf{e}_{i+1} = \mathbf{x}_{i+1} - \alpha_{1,i+1} \mathbf{e}_1 - \dots - \alpha_{i,i+1} \mathbf{e}_i \quad 1 \leq i \leq n-1$$

donde

$$\alpha_{1,i+1} = \frac{\mathbf{x}_{i+1} \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1}, \quad \alpha_{2,i+1} = \frac{\mathbf{x}_{i+1} \cdot \mathbf{e}_2}{\mathbf{e}_2 \cdot \mathbf{e}_2}, \quad \alpha_{i,i+1} = \frac{\mathbf{x}_{i+1} \cdot \mathbf{e}_i}{\mathbf{e}_i \cdot \mathbf{e}_i} \quad (3.38)$$

**Ejemplo 3.24**

Ortogonalice el siguiente conjunto de vectores linealmente independientes

$$\mathbf{x}_1 = \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

**Solución**

$$\mathbf{e}_1 = \mathbf{x}_1, \quad \mathbf{e}_2 = \mathbf{x}_2 - \alpha_{1,2} \mathbf{e}_1,$$

donde:

$$\alpha_{1,2} = \frac{\mathbf{x}_2 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1} = \frac{\begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}} = \frac{6}{5}$$

Sustituyendo

$$\mathbf{e}_2 = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} - \frac{6}{5} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 3/5 \\ 2 \\ -6/5 \end{bmatrix}$$

$$\mathbf{e}_3 = \mathbf{x}_3 - \alpha_{1,3} \mathbf{e}_1 - \alpha_{2,3} \mathbf{e}_2, \text{ donde}$$

$$\alpha_{1,3} = \frac{\mathbf{x}_3 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1} = \frac{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}} = \frac{3}{5}, \quad \alpha_{2,3} = \frac{\mathbf{x}_3 \cdot \mathbf{e}_2}{\mathbf{e}_2 \cdot \mathbf{e}_2} = \frac{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 3/5 \\ 2 \\ -6/5 \end{bmatrix}}{\begin{bmatrix} 3/5 \\ 2 \\ -6/5 \end{bmatrix} \cdot \begin{bmatrix} 3/5 \\ 2 \\ -6/5 \end{bmatrix}} = \frac{35}{145}$$

Sustituyendo

$$\mathbf{e}_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \frac{3}{5} \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix} - \frac{35}{145} \begin{bmatrix} 3/5 \\ 2 \\ -6/5 \end{bmatrix} = \begin{bmatrix} -10/29 \\ 15/29 \\ 20/29 \end{bmatrix}$$

Para los cálculos puede auxiliarse del guión del ejemplo 3.23, con los cambios pertinentes.



A continuación se presenta un algoritmo para ortogonalizar un conjunto de  $n$  vectores de  $n$  componentes cada uno por el método visto.

### ALGORITMO 3.2 Ortogonalización de Gram-Schmidt

Para ortogonalizar un conjunto de  $N$  vectores linealmente independientes de  $N$  componentes cada uno, proporcionar los

DATOS: El número  $N$  y los vectores  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_N$ .

RESULTADOS: El conjunto de vectores ortogonales  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N$ .

PASO 1. Hacer  $\mathbf{e}_1 = \mathbf{x}_1$ .

PASO 2. Hacer  $I = 1$ .

PASO 3. Mientras  $I \leq N - 1$ , repetir los pasos 4 a 10.

PASO 4. Hacer  $\mathbf{e}(I+1) = \mathbf{x}(I+1)$ .

PASO 5. Hacer  $J = 1$ .

PASO 6. Mientras  $J \leq I$ , repetir los pasos 7 a 9.

PASO 7. Hacer  $\alpha(J, I+1) = (\mathbf{x}(I+1) \cdot \mathbf{e}(J)) / (\mathbf{e}(J) \cdot \mathbf{e}(J))$ .

PASO 8. Hacer  $\mathbf{e}(I+1) = \mathbf{e}(I+1) - \alpha(J, I+1) * \mathbf{e}(J)$ .

PASO 9. Hacer  $J = J + 1$ .

PASO 10. Hacer  $I = I + 1$ .

PASO 11. IMPRIMIR los vectores  $\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N$  y TERMINAR.

NOTA: En el paso 7, el punto indica producto escalar de dos vectores.

En el 8,  $\alpha(J, I+1)$  es un escalar que multiplica al vector  $\mathbf{e}(J)$  y la resta es vectorial.

En los pasos 1, 4, 7 y 8 se trata de asignaciones de todos los componentes de un vector a otro.

**SUGERENCIA:** Es recomendable trabajar con un programa desarrollado en un lenguaje de alto nivel (véase Prob. 3.14) basado en el algoritmo 3.2 o en un pizarrón electrónico (Mathcad, por ejemplo) para evitar cálculos y analizar la ortogonalización más finamente.

Una aplicación importante de los resultados obtenidos es determinar la independencia o dependencia lineal de un conjunto dado de vectores. Para esto se partirá de un conjunto linealmente dependiente particular, obsérvese qué ocurre en el proceso de ortogonalización.

Sean  $\mathbf{x}_1 = [1 \ 2]^T$  y  $\mathbf{x}_2 = [-2 \ -4]^T$ . Obviamente  $\mathbf{x}_2 = -2 \mathbf{x}_1$

Efectuando  $\mathbf{e}_1 = \mathbf{x}_1 = [1 \ 2]^T$  y

$$\begin{aligned} \mathbf{e}_2 &= \mathbf{x}_2 - \frac{\mathbf{x}_2 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1} \mathbf{e}_1 = [-2 \ -4]^T - \frac{[-2 \ -4]^T \cdot [1 \ 2]^T}{[1 \ 2]^T \cdot [1 \ 2]^T} [1 \ 2]^T \\ &= [-2 \ -4]^T - (-2) [1 \ 2]^T = [0 \ 0]^T \end{aligned}$$

y, por tanto,  $\mathbf{e}_2 = \mathbf{0}$

Si  $\mathbf{x}_1$  y  $\mathbf{x}_2$  son vectores linealmente dependientes cualesquiera, al aplicar el proceso de ortogonalización se tiene:

$$\mathbf{e}_1 = \mathbf{x}_1,$$

$$e_2 = x_2 - \frac{x_2 \cdot e_1}{e_1 \cdot e_1} e_1$$

$$\text{como } x_2 = \beta x_1 = \beta e_1$$

$$e_2 = \beta e_1 - \frac{\beta e_1 \cdot e_1}{e_1 \cdot e_1} e_1 = \beta e_1 - \beta \frac{e_1 \cdot e_1}{e_1 \cdot e_1} e_1$$

$$\text{pero } \frac{e_1 \cdot e_1}{e_1 \cdot e_1} = 1, \text{ por tanto, } e_2 = 0 \text{ y } |e_2| = 0.$$

Generalmente, para determinar si un conjunto dado  $x_1, x_2, \dots, x_n$  es linealmente dependiente o independiente, se le aplica el proceso de ortogonalización de Gram-Schmidt. Supóngase que se han obtenido en dicho proceso  $e_1, e_2, \dots, e_i$ , a partir de  $x_1, x_2, \dots, x_i$ . Si al querer obtener  $e_{i+1}$  resulta que  $|e_{i+1}| = 0$ , o en términos prácticos su cercanía a cero satisface un criterio de ortogonalidad preestablecido  $|e_{i+1}| < \varepsilon$ , el vector  $x_{i+1}$  es linealmente dependiente de los vectores  $x_1, x_2, \dots, x_i$ ; como consecuencia, el conjunto dado es linealmente dependiente. Si, por el contrario, se obtienen  $e_1, e_2, \dots, e_n$  tales que  $|e_j| > \varepsilon$  para  $1 \leq j \leq n$ , el conjunto en cuestión es linealmente independiente.

### Ejemplo 3.25

Analice si los siguientes vectores son linealmente independientes.

$$x_1 = \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix}, \quad x_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \quad x_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

**Solución** Se aplica el proceso de Gram-Schmidt

$$e_1 = x_1 = \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix}$$

$$e_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} - \frac{\begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix}} \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 0 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix} = 0$$

lo cual implica que  $x_2$  es linealmente dependiente de  $x_1$ . El conjunto es linealmente dependiente. Sin embargo, el proceso de ortogonalización puede continuar para ver si  $x_3$  es linealmente dependiente de  $x_1$

$$e_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix}} \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} - \frac{1}{5} \begin{bmatrix} 0 \\ 5 \\ 5 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Obsérvese que en el cálculo de  $e_3$  se ignora a  $e_2$ . Como  $e_3 \neq 0$ ,  $x_1$  y  $x_3$  son linealmente independientes.

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
x1=[0; 5; 5; 0];
x2=[0; 1; 1; 0]
x3=[1; 1; 1; 1]
e1=x1
alfa12=(e1'*x2)/(e1'*e1)
e2=x2-alfa12*e1
alfa13=(e1'*x3)/(e1'*e1)
if norm(e2) >= 1e-5
    alfa23=(e2'*x3)/(e2'*e2)
    e3=x3-alfa13*e1-alfa23*e2
else
    e3=x3-alfa13*e1
end
```

```
e3_25()
Prøgm
[0;5;5;0]→x1 : [0;1;1;0]→x2
[1;1;1;1]→x3 : ClrIO
x1→e1 : Disp e1 : Pause
dotP(e1,x2)/(dotP(e1,e1)→a12
x2-a12*e1→e2 : Disp e2 : Pause
dotP(e1,x3)/(dotP(e1,e1)→a13
If norm(e2) >= 1E-5 Then
    dotP(e2,x3)/(dotP(e2,e2)→a23
    x3-a13*e1-a23*e2→e3
Else
    x3-a13*e1→e3
EndIf
Disp e3
EndPrøgm
```

## RANGO

El número de vectores linealmente independientes de un conjunto dado recibe el nombre de rango o característica del conjunto. Así, el conjunto del ejemplo 3.25 tiene un rango de 2.

Para un conjunto de  $m$  vectores, cada uno de  $n$  componentes, el rango puede ser como máximo igual al menor de  $m$  o  $n$ .

## RANGO DE UNA MATRIZ

Una matriz puede verse como un conjunto de vectores; más claramente, la matriz.

$$A = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{m,1} & a_{m,2} & \cdots & a_{m,n} \end{bmatrix},$$



se puede tratar como un conjunto de  $n$  vectores columna de  $m$  componentes cada uno (o bien  $m$  vectores fila de  $n$  componentes cada uno); es decir, como  $A = [x_1 \ x_2 \ \dots \ x_n]$  donde

$$x_1 = \begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix}, \quad x_2 = \begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{bmatrix}, \quad \dots, \quad x_n = \begin{bmatrix} a_{1,n} \\ a_{2,n} \\ \vdots \\ a_{m,n} \end{bmatrix}$$

o como:

$$A = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

donde

$$y_1 = [a_{1,1} \ a_{1,2} \ \dots \ a_{1,n}], \quad y_2 = [a_{2,1} \ a_{2,2} \ \dots \ a_{2,n}], \quad \dots, \quad y_m = [a_{m,1} \ a_{m,2} \ \dots \ a_{m,n}]$$

En estas condiciones puede hablarse del rango de una matriz, en donde el rango de una matriz  $A$  está dado por el número máximo de vectores columna o vectores fila, linealmente independientes.\*

Así la matriz

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

cuyas columnas son los elementos del conjunto dado en el ejemplo 3.25, tiene rango 2.

Cuando el rango de una matriz cuadrada de orden  $n$  es menor que  $n$ , se dice que la matriz es singular. Lo cual significa también que su determinante es cero (véase Prob. 3.18). Si las columnas de la matriz son “casi” linealmente dependientes, recibe el nombre de **casi singular** o **mal condicionada** (véase sistemas de ecuaciones mal condicionadas, Sec. 3.4).

En esta sección se ha considerado una serie de conceptos teóricos que, además de su interés por sí mismos, forman un marco que permitirá explicar de manera lógica ciertos algoritmos importantes de las matemáticas y también conceptos de existencia y unicidad de las soluciones de los problemas que resuelven dichos algoritmos.

### 3.4 Solución de sistemas de ecuaciones lineales

Gran número de problemas prácticos de ingeniería se reduce al problema de resolver un sistema de ecuaciones lineales. Por ejemplo, pueden citarse la solución de sistemas de ecuaciones no lineales, la aproximación polinomial, la solución de ecuaciones diferenciales parciales, entre otros.

\* Puede demostrarse que el número máximo de vectores columna linealmente independientes de una matriz  $A$ , es igual al número máximo de vectores fila linealmente independientes.

Un sistema de  $m$  ecuaciones lineales en  $n$  incógnitas tiene la forma general

$$\begin{array}{ccccccc} a_{1,1}x_1 & + & a_{1,2}x_2 & + & \dots & + & a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 & + & a_{2,2}x_2 & + & \dots & + & a_{2,n}x_n & = & b_2 \\ \cdot & & \cdot & & & & \cdot & & \cdot \\ \cdot & & \cdot & & & & \cdot & & \cdot \\ \cdot & & \cdot & & & & \cdot & & \cdot \\ a_{m,1}x_1 & + & a_{m,2}x_2 & + & \dots & + & a_{m,n}x_n & = & b_m \end{array} \quad (3.39)$$

Con la notación matricial se puede escribir la ecuación anterior como:

$$\begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_m \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{bmatrix}$$

y concretamente como  $A \mathbf{x} = \mathbf{b}$ .

Donde  $A$  es la **matriz coeficiente** del sistema,  $\mathbf{x}$  el **vector incógnita** y  $\mathbf{b}$  el **vector de términos independientes**.

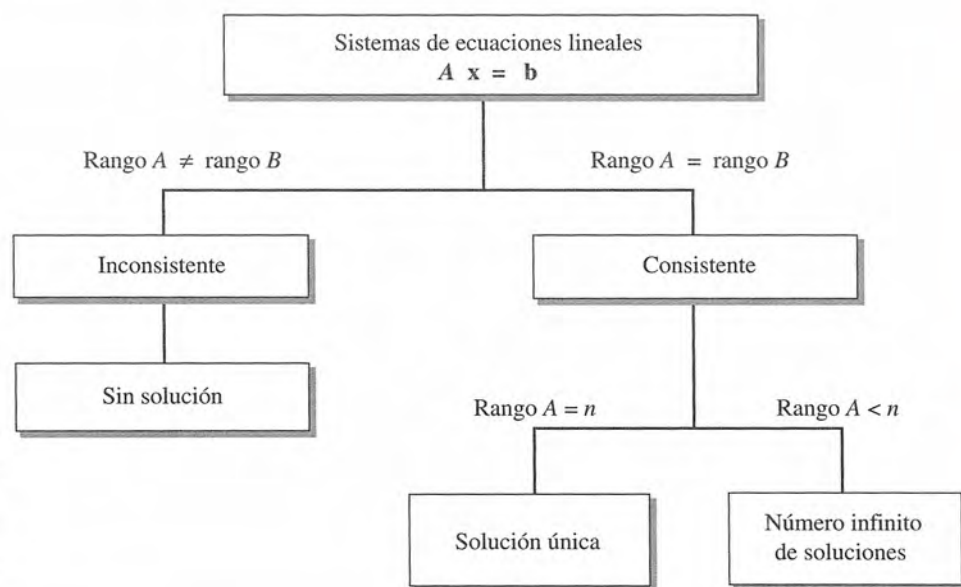
Dados  $A$  y  $\mathbf{b}$ , se entiende por resolver el sistema (Ec. 3.39) encontrar los vectores  $\mathbf{x}$  que lo satisfagan. Antes de estudiar las técnicas que permiten encontrar  $\mathbf{x}$  se expondrán algunas consideraciones teóricas.

## EXISTENCIA Y UNICIDAD DE SOLUCIONES

Si  $\mathbf{b}$  es el vector cero, la ecuación 3.39 es un **sistema homogéneo**. Si por el contrario,  $\mathbf{b} \neq \mathbf{0}$ , el sistema es **no homogéneo**. A continuación se define la **matriz aumentada**  $B$ , formada con los elementos de la matriz coeficiente  $A$  y los del vector  $\mathbf{b}$  de la siguiente manera:

$$B = \left[ \begin{array}{cccc|c} a_{1,1} & a_{1,2} & \dots & a_{1,n} & b_1 \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} & b_2 \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ \cdot & \cdot & & \cdot & \cdot \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} & b_m \end{array} \right] = [A \mid \mathbf{b}]$$

Si el rango de la matriz coeficiente  $A$  y de la matriz aumentada  $B$  son iguales, se dice que el sistema (Ec. 3.39) es **consistente**. Si no ocurre esto, el sistema es inconsistente (por tanto, un sistema homogéneo siempre es consistente). Un sistema **inconsistente** no tiene solución, mientras que uno consistente tiene una solución única o un número infinito de soluciones, según como sea el rango de  $A$  en comparación con el número de incógnitas  $n$ . Si el rango de  $A$  es igual al número de incógnitas, la solución es única; si el rango de  $A$  es menor que dicho número, hay un número infinito de soluciones (véase Fig. 3.9).



**Figura 3.9**  
Solución  
de sistemas de  
ecuaciones  
lineales.

### Ejemplo 3.26

Sea el sistema

$$\begin{aligned} 2x_1 + 4x_2 &= 6 \\ 3x_1 + 6x_2 &= 5 \end{aligned}$$

La matriz aumentada es:

$$\left[ \begin{array}{cc|c} 2 & 4 & 6 \\ 3 & 6 & 5 \end{array} \right]$$

Puede verse fácilmente que: rango de  $A = 1$ , rango de  $B = 2$ ; como  $\text{rango } A \neq \text{rango } B$ , el sistema no tiene solución.

Si el sistema es homogéneo

$$\begin{aligned} 2x_1 + 4x_2 &= 0 \\ 3x_1 + 6x_2 &= 0, \end{aligned}$$

la matriz aumentada es:

$$\left[ \begin{array}{cc|c} 2 & 4 & 0 \\ 3 & 6 & 0 \end{array} \right]$$

y  $\text{rango } A = 1$ ,  $\text{rango } B = 1$ ,  $\text{rango } A < 2 = n$ ; en este caso existe un número infinito de soluciones.

Para realizar los cálculos puede usarse el siguiente guión de Matlab:



```
A=[2 4; 3 6]
rangoA=rank(A)
B=[2 4 6; 3 6 5]
rangoB=rank(B)
```



**Ejemplo 3.27**

Sea el sistema

$$\begin{aligned} 2x_1 + 3x_2 + x_3 &= 0 \\ 0x_1 + 2x_2 + x_3 &= 1 \\ x_1 + 0x_2 + x_3 &= 0, \end{aligned}$$

donde la matriz aumentada es:

$$\left[ \begin{array}{ccc|c} 2 & 3 & 1 & 0 \\ 0 & 2 & 1 & 1 \\ 1 & 0 & 1 & 0 \end{array} \right]$$

Obsérvese que la matriz coeficiente son los vectores del ejemplo 3.24, que son linealmente independientes y, por tanto,  $\text{rango } A = 3$ .

Al aplicar el método de Gram-Schmidt para ortogonalizar el vector de términos independientes se observa que es linealmente dependiente y, por tanto,  $\text{rango } B = 3$ . El sistema es consistente y como  $\text{rango } A = \text{número de incógnitas} = 3$ , puede esperarse solución única del sistema.

Esta comprobación se deja como ejercicio para el lector.

**MÉTODOS DIRECTOS DE SOLUCIÓN**

El prototipo de todos estos métodos se conoce como la eliminación de Gauss y se presenta a continuación.

**ELIMINACIÓN DE GAUSS**

Considérese un sistema general de tres ecuaciones lineales con tres incógnitas.

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 &= b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + a_{2,3}x_3 &= b_2 \\ a_{3,1}x_1 + a_{3,2}x_2 + a_{3,3}x_3 &= b_3 \end{aligned} \quad (3.40)$$

Como primer paso, se reemplaza la segunda ecuación con lo que resulte de sumarle la primera ecuación multiplicada por  $(-a_{2,1}/a_{1,1})$ . De manera similar se sustituye la tercera ecuación con el resultado de sumarle la primera ecuación multiplicada por  $(-a_{3,1}/a_{1,1})$ .

Esto da lugar al nuevo sistema

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 &= b_1 \\ a'_{2,2}x_2 + a'_{2,3}x_3 &= b'_2 \\ a'_{3,2}x_2 + a'_{3,3}x_3 &= b'_3 \end{aligned} \quad (3.41)$$

en donde las  $a'$  y las  $b'$  son los nuevos elementos que se obtienen de las operaciones ya mencionadas, y en donde  $x_1$  se ha eliminado en la segunda y tercera ecuaciones. Ahora, multiplicando la segunda ecuación de 3.41 por  $(-a'_{3,2}/a'_{2,2})$  y sumando el resultado a la tercera ecuación de 3.41, se obtiene el sistema triangular

$$\begin{aligned} a_{1,1}x_1 + a_{1,2}x_2 + a_{1,3}x_3 &= b_1 \\ a'_{2,2}x_2 + a'_{2,3}x_3 &= b'_2 \\ a''_{3,3}x_3 &= b''_3 \end{aligned} \quad (3.42)$$

donde  $a''_{3,3}$  y  $b''_3$ , resultaron de las operaciones realizadas y  $x_2$  se ha eliminado de la tercera ecuación.

El proceso de llevar el sistema de ecuaciones 3.40 a la forma de la ecuación 3.42 se conoce como **triangularización**.

El sistema en la forma de la ecuación 3.42 se resuelve despejando de su última ecuación  $x_3$ , sustituyendo  $x_3$  en la segunda ecuación y despejando  $x_2$  de ella. Por último, con  $x_3$  y  $x_2$  sustituidas en la primera ecuación de 3.42 se obtiene  $x_1$ . Esta parte del proceso se llama **sustitución regresiva**.

Antes de ilustrar la eliminación de Gauss con un ejemplo particular, nótese que no es necesario conservar  $x_1$ ,  $x_2$  y  $x_3$  en la triangularización y que ésta puede llevarse a cabo usando solamente la matriz coeficiente  $A$  y el vector  $b$ . Para mayor simplicidad se empleará la matriz aumentada  $B$ .

$$B = \left[ \begin{array}{ccc|c} a_{1,1} & a_{1,2} & a_{1,3} & b_1 \\ a_{2,1} & a_{2,2} & a_{2,3} & b_2 \\ a_{3,1} & a_{3,2} & a_{3,3} & b_3 \end{array} \right] = [A \mid b]$$

Con esto se incorpora la notación matricial y todas sus ventajas a la solución de sistemas de ecuaciones lineales.

### Ejemplo 3.28

Resuelva por eliminación de Gauss el sistema

$$\begin{aligned} 4x_1 - 9x_2 + 2x_3 &= 5 \\ 2x_1 - 4x_2 + 6x_3 &= 3 \\ x_1 - x_2 + 3x_3 &= 4 \end{aligned} \quad (3.43)$$

**Solución** La matriz aumentada del sistema es:



$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 2 & -4 & 6 & 3 \\ 1 & -1 & 3 & 4 \end{array} \right] \quad (3.44)$$

### Triangularización

Al sumar la primera ecuación multiplicada por  $(-2/4)$  a la segunda, y la primera ecuación multiplicada por  $(-1/4)$  a la tercera, resulta

$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 0 & 0.5 & 5 & 0.5 \\ 0 & 1.25 & 2.5 & 2.75 \end{array} \right] \quad (3.45)$$

Obsérvese que en este paso la primera fila se conserva sin cambio.

Sumando la segunda fila multiplicada por  $(-1.25/0.5)$  a la tercera se obtiene la matriz\*

$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 0 & 0.5 & 5 & 0.5 \\ 0 & 0 & -10 & 1.5 \end{array} \right] \quad (3.46)$$

que en términos de sistemas de ecuaciones quedaría como

\* Nótese que los vectores columna de  $A$  se han ortogonalizado en la triangularización.

$$\begin{aligned} 4x_1 - 9x_2 + 2x_3 &= 5 \\ 0.5x_2 + 5x_3 &= 0.5 \\ -10x_3 &= 1.5 \end{aligned} \quad (3.47)$$

Un proceso de sustitución regresiva produce el resultado buscado. La tercera ecuación de 3.47 da el valor de  $x_3 = -0.15$ ; de la segunda ecuación se obtiene entonces:

$$0.5 x_2 = 0.5 - 5x_3 = 1.25$$

y por tanto  $x_2 = 2.5$

finalmente al sustituir  $x_2$  y  $x_3$  en la primera ecuación de la forma 3.47 resulta

$$4 x_1 = 5 + 9 x_2 - 2 x_3 = 27.8,$$

de modo que  $x_1 = 6.95$

Con la sustitución de estos valores en el sistema original se verifica la exactitud de los resultados.\*

Para realizar los cálculos puede usarse el siguiente gui3n de Matlab:



```
A=[4 -9 2 5; 2 -4 6 3; 1 -1 3 4]
A(2,:)=A(2,:)-A(1,:)*A(2,1)/A(1,1);
A(3,:)=A(3,:)-A(1,:)*A(3,1)/A(1,1);
A
A(3,:)=A(3,:)-A(2,:)*A(3,2)/A(2,2);
x(3)=A(3,4)/A(3,3);
x(2)=(A(2,4)-A(2,3)*x(3))/A(2,2);
x(1)=(A(1,4)-A(1,2:3)*x(2:3))/A(1,1)
```

Tambi3n puede obtenerse la soluci3n directamente con las instrucciones siguientes:



```
A= [4 -9 2; 2 -4 6; 1 -1 3]
b=[5; 3; 4]
x=A\b
```

```
simult([4,-9,2;2,-4,6;1,-1,3], [5; 3; 4])
```

Como producto secundario de este trabajo, se puede calcular f3cilmente el **determinante de la matriz A** del sistema original. La matriz coeficiente A pasa de la forma original a la matriz triangular superior:

$$\begin{bmatrix} 4 & -9 & 2 \\ 0 & 0.5 & 5 \\ 0 & 0 & -10 \end{bmatrix} \quad (3.48)$$

mediante operaciones que, de acuerdo con las reglas de los determinantes, no alteran el valor de  $|A|$ . El determinante de la ecuaci3n 3.48 es s3lo el producto de los elementos de la diagonal principal, de modo que el resultado es

$$|A| = 4(0.5)(-10) = -20$$

\* V3ase matrices mal condicionadas (Sec. 3.4.).



Las ecuaciones para la triangularización, sustitución regresiva y cálculo del determinante de un sistema de  $n$  ecuaciones en  $n$  incógnitas  $A \mathbf{x} = \mathbf{b}$  por el método de eliminación de Gauss son:

### TRIANGULARIZACIÓN

Para  $1 \leq i \leq n-1$

Para  $i+1 \leq k \leq n$

$$b_k = b_k - (a_{k,i}/a_{i,i})b_i \quad (3.49)$$

Para  $i+1 \leq j \leq n$

$$a_{k,j} = a_{k,j} - \frac{a_{k,i}}{a_{i,i}} a_{i,j}$$

### SUSTITUCIÓN REGRESIVA

$$x_n = b_n/a_{n,n}$$

Para  $i = n-1, n-2, \dots, 1$

$$x_i = \frac{1}{a_{i,i}} [b_i - \sum_{j=i+1}^n a_{i,j} x_j] \quad (3.50)$$

### CÁLCULO DEL DETERMINANTE

$$\det A = \prod_{i=1}^n a_{i,i} = a_{1,1} a_{2,2} \dots a_{n,n} \quad (3.51)$$

El algoritmo para resolver  $A \mathbf{x} = \mathbf{b}$  por eliminación de Gauss queda entonces

#### ALGORITMO 3.3 Eliminación de Gauss

Para obtener la solución de un sistema de ecuaciones lineales  $A \mathbf{x} = \mathbf{b}$  y el determinante de  $A$ , proporcionar los

DATOS:  $N$  número de ecuaciones,  $A$  matriz coeficiente y  $\mathbf{b}$  vector de términos independientes.

RESULTADOS: El vector solución  $\mathbf{x}$  y el determinante de  $A$  o mensaje de falla "HAY UN CERO EN LA DIAGONAL PRINCIPAL".

PASO 1. Hacer  $DET = 1$ .

PASO 2. Hacer  $I = 1$ .

PASO 3. Mientras  $1 \leq N-1$ , repetir los pasos 4 a 14.

PASO 4. Hacer  $DET = DET * A(I, I)$ .

PASO 5. Si  $DET = 0$  IMPRIMIR mensaje "HAY UN CERO EN LA DIAGONAL PRINCIPAL" y TERMINAR. De otro modo continuar.

PASO 6. Hacer  $K = I + 1$ .

PASO 7. Mientras  $K \leq N$ , repetir los pasos 8 a 13.

PASO 8. Hacer  $J = I + 1$ .

PASO 9. Mientras  $J \leq N$ , repetir los pasos 10 y 11.

PASO 10. Hacer  $A(K, J) = A(K, J) - A(K, I) * A(I, J) / A(I, I)$ .

PASO 11. Hacer  $J = J + 1$ .

PASO 12. Hacer  $b(K) = b(K) - A(K, I) * b(I) / A(I, I)$ .

- PASO 13. Hacer  $K = K + 1$ .  
 PASO 14. Hacer  $I = I + 1$ .  
 PASO 15. Hacer  $DET = DET * A(N, N)$ .  
 PASO 16. Si  $DET = 0$  IMPRIMIR mensaje "HAY UN CERO EN LA DIAGONAL PRINCIPAL" y TERMINAR.  
 De otro modo continuar.  
 PASO 17. Hacer  $x(N) = b(N) / A(N, N)$ .  
 PASO 18. Hacer  $I = N - 1$ .  
 PASO 19. Mientras  $I \geq 1$ , repetir los pasos 20 a 26.  
 PASO 20. Hacer  $x(I) = b(I)$ .  
 PASO 21. Hacer  $J = I + 1$ .  
 PASO 22. Mientras  $J \leq N$ , repetir los pasos 23 y 24.  
 PASO 23. Hacer  $x(I) = x(I) - A(I, J) * x(J)$ .  
 PASO 24. Hacer  $J = J + 1$ .  
 PASO 25. Hacer  $x(I) = x(I) / A(I, I)$ .  
 PASO 26. Hacer  $I = I - 1$ .  
 PASO 27. IMPRIMIR  $x$  y  $DET$  y TERMINAR.

### ELIMINACIÓN DE GAUSS CON PIVOTEO

En la eliminación de  $x_1$  de la segunda y tercera ecuaciones de la forma 3.40 se tomó como base la primera, por lo cual se denomina ecuación pivote o, en términos de la notación matricial, **fila pivote**. Para eliminar  $x_2$  de la tercera ecuación de la forma 3.41, la fila **pivote** utilizada fue la segunda. El coeficiente de la incógnita que se va a eliminar en la fila pivote se llama **pivote**. En la eliminación que dio como resultado el sistema de ecuaciones 3.42, los pivotes fueron  $a_{1,1}$  y  $a'_{2,2}$ . Esta elección natural de los pivotes  $a_{1,1}$ ,  $a'_{2,2}$ ,  $a'_{3,3}$ , etc., es muy conveniente tanto para trabajar con una calculadora como con una computadora; desafortunadamente falla cuando alguno de esos elementos es cero, puesto que los multiplicadores quedarían indeterminados [por ejemplo si  $a_{1,1}$  fuera cero, el multiplicador  $(-a_{2,1} / a_{1,1})$  no está definido]. Una manera de evitar esta posibilidad es seleccionar como pivote el coeficiente de máximo valor absoluto en la columna relevante de la matriz reducida. Como antes, se tomarán las columnas en orden natural de modo que se vayan eliminando las incógnitas también en orden natural  $x_1, x_2, x_3$ , etc. Esta técnica, llamada **pivoteo parcial**, se ilustra con la solución del siguiente sistema.

#### Ejemplo 3.29

Resuelva el sistema

$$\begin{aligned} 10x_1 + x_2 - 5x_3 &= 1 \\ -20x_1 + 3x_2 + 20x_3 &= 2 \\ 5x_1 + 3x_2 + 5x_3 &= 6. \end{aligned} \quad (3.52)$$

**Solución** La matriz aumentada es:

$$\left[ \begin{array}{ccc|c} 10 & 1 & -5 & 1 \\ -20 & 3 & 20 & 2 \\ 5 & 3 & 5 & 6 \end{array} \right] \quad (3.53)$$

El primer pivote debe ser  $(-20)$ , ya que es el elemento de máximo valor absoluto en la primera columna. Se elimina entonces  $x_1$  de la primera y tercera filas de la ecuación 3.52. Para ello, se suma a la primera fila la segunda multiplicada por  $(-10 / (-20))$ , y a la tercera fila la segunda multiplicada por  $(-5 / (-20))$ . Con esto se obtiene la matriz reducida:

$$\left[ \begin{array}{ccc|c} 0 & 2.5 & 5 & 2 \\ -20 & 3 & 20 & 2 \\ 0 & 3.75 & 10 & 6.5 \end{array} \right] \quad (3.54)$$

El siguiente pivote debe seleccionarse entre la primera y tercera filas (segunda columna) y en este caso es (3.75). Sumando a la primera fila la tercera multiplicada por  $(-2.5 / 3.75)$ , resulta

$$\left[ \begin{array}{ccc|c} 0 & 0 & -1.666 & -2.333 \\ -20 & 3 & 20 & 2 \\ 0 & 3.75 & 10 & 6.5 \end{array} \right] \quad (3.55)$$

que puesta en forma de sistema de ecuaciones queda:

$$\begin{aligned} -1.666x_3 &= -2.333 \\ -20x_1 + 3x_2 + 20x_3 &= 2 \\ 3.75x_2 + 10x_3 &= 6.5 \end{aligned} \quad (3.56)$$

De la primera ecuación de 3.56

$$x_3 = \frac{-2.333}{-1.666} = 1.4,$$

de la tercera ecuación

$$x_2 = \frac{6.5 - 10(1.4)}{3.75} = -2,$$

y finalmente de la segunda ecuación

$$x_1 = \frac{2 - 3(-2) - 20(1.4)}{-20} = 1.$$

Para realizar los cálculos puede usarse el siguiente guión de Matlab:



```
A=[10 1 -5 1; -20 3 20 2; 5 3 5 6]
copia=A(2,:); A(2,:)=A(1,:); A(1,:)=copia;
A(2,:)=A(2,:)-A(1,:)*A(2,1)/A(1,1);
A(3,:)=A(3,:)-A(1,:)*A(3,1)/A(1,1);
A
A(3,:)=A(3,:)-A(2,:)*A(3,2)/A(2,2);
A
x(3)=A(3,4)/A(3,3);
x(2)=(A(2,4)-A(2,3)*x(3))/A(2,2);
x(1)=(A(1,4)-A(1,2:3)*x(2:3))/A(1,1)
x
```

**Otra alternativa** para solucionar el sistema de ecuaciones 3.52 es utilizar el mismo criterio de selección de los pivotes, pero llevando las filas pivote a las posiciones de modo que se obtenga la forma triangular en la eliminación. Para esto es necesario, por ejemplo, en la



ecuación 3.52, intercambiar la segunda fila (donde se encuentra el elemento de máximo valor absoluto) con la primera, con lo que se obtiene:

$$\left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ 10 & 1 & -5 & 1 \\ 5 & 3 & 5 & 6 \end{array} \right] \quad (3.53')$$

que se reduce en la primera eliminación a

$$\left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ 0 & 2.5 & 5 & 2 \\ 0 & 3.75 & 10 & 6.5 \end{array} \right] \quad (3.54')$$

Como el siguiente pivote es (3.75), se intercambian la segunda y la tercera filas de la ecuación 3.54', para obtener:

$$\left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ 0 & 3.75 & 10 & 6.5 \\ 0 & 2.5 & 5 & 2 \end{array} \right] \quad (3.54'')$$

la cual se reduce al eliminar  $x_2$  a

$$\left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ 0 & 3.75 & 10 & 6.5 \\ 0 & 0 & -1.666 & -2.333 \end{array} \right] \quad (3.55')$$

que tiene ya la forma triangular y está lista para la sustitución regresiva. **En adelante, cualquier referencia a la eliminación con pivoteo que se haga, entraña la segunda alternativa.**

La sustitución regresiva proporciona los siguientes valores

$$x_3 = 1.4, \quad x_2 = -2, \quad x_1 = 1$$

El determinante de  $A$  se calcula de nuevo, multiplicando entre sí los elementos de la diagonal principal de la matriz triangularizada (Ec. 3.55'), pero dicho producto es afectado por un cambio de signo por cada intercambio de filas que se verifique en la triangularización. En el caso en estudio

$$\det A = (-1)^2 (-20) (3.75) (-1.666) = 125$$

ya que hubo dos intercambios de fila para llegar a la ecuación 3.55'.

A fin de elaborar el algoritmo de este método, se utilizarán las ecuaciones 3.49 para la triangularización después de cada búsqueda del elemento de máximo valor absoluto y del intercambio de filas correspondiente. Una vez realizada la triangularización, se hará la sustitución regresiva con las ecuaciones 3.50 y el cálculo del determinante de la siguiente forma

$$\det A = (-1)^r \prod_{i=1}^n a_{i,i} \quad (3.57)$$

donde  $r$  es el número de intercambios de filas que hubo en el proceso de triangularización.

**ALGORITMO 3.4** Eliminación de Gauss con pivoteo

Para obtener la solución de un sistema de ecuaciones lineales  $A \mathbf{x} = \mathbf{b}$  y el determinante de  $A$ , proporcionar los

DATOS:  $N$  número de ecuaciones,  $A$  matriz coeficiente y  $\mathbf{b}$  vector de términos independientes.

RESULTADOS: El vector solución  $\mathbf{x}$  y el determinante de  $A$  o mensaje "MATRIZ SINGULAR, SISTEMA SIN SOLUCION".

PASO 1. Hacer  $DET = 1$ .

PASO 2. Hacer  $R = 0$ .

PASO 3. Hacer  $I = 1$ .

PASO 4. Mientras  $I \leq N - 1$  repetir los pasos 5 a 12.

PASO 5. Encontrar PIVOTE (elemento de mayor valor absoluto en la parte relevante de la columna  $I$  de  $A$ ) y  $P$  la fila donde se encuentra PIVOTE.

PASO 6. Si  $PIVOTE = 0$  IMPRIMIR "MATRIZ SINGULAR SISTEMA SIN SOLUCION" y TERMINAR. En caso contrario continuar.

PASO 7. Si  $P = I$  ir al paso 10. De otro modo realizar los pasos 8 y 9.

PASO 8. Intercambiar la fila  $I$  con la fila  $P$ .

PASO 9. Hacer  $R = R + 1$ .

PASO 10. Hacer  $DET = DET * A(I, I)$ .

PASO 11. Realizar los pasos 6 a 13 del algoritmo 3.3.

PASO 12. Hacer  $I = I + 1$ .

PASO 13. Hacer  $DET = DET * A(N, N) * (-1)^{**R}$

PASO 14. Realizar los pasos 17 a 26 del algoritmo 3.3

PASO 15. IMPRIMIR  $\mathbf{x}$  y  $DET$  y TERMINAR.

Para terminar el tema, se compararán las técnicas de eliminación de Gauss con pivoteo y sin éste. Por brevedad, la primera se denominará GP y la segunda G.

1. La búsqueda del coeficiente de mayor valor absoluto que se usará como pivote y el intercambio de filas significa mayor programación en GP.
2. Los factores  $(a_{k,i} / a_{i,i})$  de las ecuaciones 3.49 siempre serán menores que la unidad en valor absoluto en GP, con esto los elementos de  $A | \mathbf{b}$  se conservan dentro de cierto intervalo, circunstancia valiosa en los cálculos computacionales.
3. Encontrar en GP un pivote igual a cero significaría que se trata de una matriz coeficiente  $A$  singular ( $\det A = 0$ ) y que el sistema  $A \mathbf{x} = \mathbf{b}$  no tiene solución única. Encontrar en G un pivote igual a cero, no proporciona información alguna acerca del determinante de  $A$  y si detendría el proceso de triangularización.

A pesar de la programación adicional y el mayor tiempo de máquina que se emplea en el método de Gauss con pivoteo, sus otras ventajas borran totalmente estas desventajas en la práctica; por tanto, el pivoteo natural se emplea sólo en circunstancias especiales, por ejemplo cuando se sabe por adelantado que no hay pivotes más grandes que los que van resultando en la diagonal principal.

**ELIMINACIÓN DE JORDAN**

Es posible extender los métodos vistos de modo que las ecuaciones se reduzcan a una forma en que la matriz coeficiente del sistema sea diagonal y ya no se requiera la sustitución regresiva. Los pivotes se eligen como en el método de Gauss con pivoteo, y una vez intercambiadas las filas se eliminan los elementos arriba y debajo del pivote. El sistema del ejemplo 3.28 ilustra este método.

**Ejemplo 3.30**

Por eliminación de Jordan, resuelva el sistema

$$\begin{aligned} 4x_1 - 9x_2 + 2x_3 &= 5 \\ 2x_1 - 4x_2 + 6x_3 &= 3 \\ x_1 - x_2 + 3x_3 &= 4 \end{aligned}$$

**Solución** La matriz aumentada del sistema es

$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 2 & -4 & 6 & 3 \\ 1 & -1 & 3 & 4 \end{array} \right]$$

Como en la primera columna el elemento de máximo valor absoluto se encuentra en la primera fila, ningún intercambio es necesario y el primer paso de eliminación produce

$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 0 & 0.5 & 5 & 0.5 \\ 0 & 1.25 & 2.5 & 2.75 \end{array} \right]$$

El elemento de máximo valor absoluto en la parte relevante de la segunda columna (filas 2 y 3) es 1.25; por tanto, la fila 3 debe intercambiarse con la 2.

$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 0 & 1.25 & 2.5 & 2.75 \\ 0 & 0.5 & 5 & 0.5 \end{array} \right]$$

Sumando la segunda fila multiplicada por  $(-(-9)/1.25)$ , a la primera fila y la segunda multiplicada por  $(-0.5/1.25)$  a la tercera se obtiene el nuevo arreglo

$$\left[ \begin{array}{ccc|c} 4 & 0 & 20 & 24.8 \\ 0 & 1.25 & 2.5 & 2.75 \\ 0 & 0 & 4 & -0.6 \end{array} \right]$$

donde se han eliminado los elementos de arriba y abajo del pivote (nótese que en este paso el primer pivote no se modifica porque sólo hay ceros debajo de él).

Por último, sumando la tercera multiplicada por  $(-20/4)$  a la primera fila y a la tercera multiplicada por  $(-2.5/4)$  a la segunda

$$\left[ \begin{array}{ccc|c} 4 & 0 & 0 & 27.8 \\ 0 & 1.25 & 0 & 3.125 \\ 0 & 0 & 4 & -0.6 \end{array} \right]$$

que escrita de nuevo como sistema de ecuaciones da

$$\begin{aligned} 4x_1 &= 27.8 \\ 1.25x_2 &= 3.125 \\ 4x_3 &= -0.6 \end{aligned}$$

de donde el resultado final se obtiene fácilmente

$$x_1 = \frac{27.8}{4} = 6.95, \quad x_2 = \frac{3.125}{1.25} = 2.5, \quad x_3 = \frac{-0.6}{4} = -0.15$$



El determinante también puede calcularse

$$|A| = (-1)^1 (4) (1.25) (4) = -20,$$

donde la potencia 1 indica que sólo hubo un intercambio de filas.

Para realizar los cálculos puede usarse el siguiente guión de Matlab:



```
A=[4 -9 2 5; 2 -4 6 3; 1 -1 3 4]
% No es necesario intercambiar filas
for i=2:3
    A(i,:)=A(i,:) - A(1,:)*A(i,1)/A(1,1);
end
A
% Se intercambia la fila 2 con la fila 3
copia=A(3,:);A(3,:)=A(2,:);A(2,:)=copia;
for i=1:3
    if i ~=2
        A(i,:)=A(i,:) - A(2,:)*A(i,2)/A(2,2);
    end
end
A
for i=1:2
    A(i,:)=A(i,:)-A(3,:)*A(i,3)/A(3,3);
end
A
for i=1:3
    x(i)=A(i,4)/A(i,i);
end
x
```

También puede obtenerse el determinante directamente con las instrucciones siguientes:



```
A=[4 -9 2; 2 -4 6; 1 -1 3]
det(A)
```

```
det([4,-9,2;2,-4,6;1,-1,3])
```

Si sólo se requiere calcular  $|A|$  y no la solución del sistema, el método de Jordan requiere mayor trabajo que el método de eliminación de Gauss con pivoteo.

### CÁLCULO DE INVERSAS

Si se tienen varios sistemas por resolver que comparten la misma matriz coeficiente; es decir,

$$A \mathbf{x}_1 = \mathbf{b}_1, \quad A \mathbf{x}_2 = \mathbf{b}_2, \text{ etc.}$$

pueden resolverse todos a un tiempo si se aplica al arreglo

$$[A \mid \mathbf{b}_1 \mid \mathbf{b}_2 \dots]$$

el proceso de eliminación, como antes y después, se realiza una sustitución regresiva particular para cada columna del lado derecho de  $A$ . Como caso particular es factible encon-

trar  $A^{-1}$  si  $\mathbf{b}_1 = \mathbf{e}_1, \mathbf{b}_2 = \mathbf{e}_2, \dots, \mathbf{b}_n = \mathbf{e}_n$ .<sup>\*</sup> Las  $n$  soluciones obtenidas forman las  $n$  columnas de la matriz inversa  $A^{-1}$ .

### CÁLCULO DE LA INVERSA CON EL MÉTODO DE GAUSS CON PIVOTEO

Como ejemplo se usará la matriz coeficiente del sistema (3.44) para obtener su inversa. Primero se forma el arreglo

$$\left[ \begin{array}{ccc|ccc} 4 & -9 & 2 & 1 & 0 & 0 \\ 2 & -4 & 6 & 0 & 1 & 0 \\ 1 & -1 & 3 & 0 & 0 & 1 \end{array} \right], \quad (3.58)$$

nótese que a la derecha de  $A$  se tiene la matriz identidad correspondiente. Eliminando los elementos debajo del primer pivote (4), se llega al sistema

$$\left[ \begin{array}{ccc|ccc} 4 & -9 & 2 & 1 & 0 & 0 \\ 0 & 0.5 & 5 & -0.5 & 1 & 0 \\ 0 & 1.25 & 2.5 & -0.25 & 0 & 1 \end{array} \right], \quad (3.59)$$

Se intercambian la segunda y tercera filas.

$$\left[ \begin{array}{ccc|ccc} 4 & -9 & 2 & 1 & 0 & 0 \\ 0 & 1.25 & 2.5 & -0.25 & 0 & 1 \\ 0 & 0.5 & 5 & -0.5 & 1 & 0 \end{array} \right] \quad (3.60)$$

Ahora se elimina el segundo elemento de la tercera fila y el arreglo cambia a

$$\left[ \begin{array}{ccc|ccc} 4 & -9 & 2 & 1 & 0 & 0 \\ 0 & 1.25 & 2.5 & -0.25 & 0 & 1 \\ 0 & 0 & 4 & -0.4 & 1 & -0.4 \end{array} \right]$$

Con la sustitución regresiva para el primer vector al lado derecho de la matriz triangular resulta:

$$4x_3 = -0.4, \text{ de donde } x_3 = -0.1;$$

al sustituir  $x_3$  en la fila 2 se tiene

$$1.25x_2 = -0.25 - 2.5(-0.1) \text{ y } x_2 = 0;$$

y reemplazando  $x_3$  y  $x_2$  en la fila 1, se obtiene

$$4x_1 = 1 + 9(0) - 2(-0.1) = 1.2 \text{ y } x_1 = 0.3$$

Este primer vector solución representa la primera columna de  $A^{-1}$ . Del mismo modo se calculan la segunda y la tercera columnas de  $A^{-1}$  con el segundo y tercer vectores del lado derecho de la matriz triangular

$$A^{-1} = \begin{bmatrix} 0.3 & -1.25 & 2.3 \\ 0 & -0.5 & 1.0 \\ -0.1 & 0.25 & -0.1 \end{bmatrix}$$

<sup>\*</sup> En este caso  $\mathbf{e}_1, \mathbf{e}_2$ , etc., son vectores de  $n$  elementos cuyo único elemento distinto de cero es el de la fila 1, 2, etc., y su valor es 1.

**CÁLCULO DE LA INVERSA CON EL MÉTODO DE JORDAN**

Se parte del mismo arreglo (Ec. 3.58) y también se eliminan los elementos debajo del primer pivote para llegar a la ecuación 3.59. Se intercambian la segunda y tercera filas y se llega al sistema de ecuaciones 3.60. En este último arreglo se eliminan los elementos arriba y debajo del pivote (1.25) para llegar a:

$$\left[ \begin{array}{ccc|ccc} 4 & 0 & 20 & -0.8 & 0 & 7.2 \\ 0 & 1.25 & 2.5 & -0.25 & 0 & 1 \\ 0 & 0 & 4 & -0.4 & 1 & -0.4 \end{array} \right]$$

arreglo que todavía se reduce a:

$$\left[ \begin{array}{ccc|ccc} 4 & 0 & 0 & 1.2 & -5 & 9.2 \\ 0 & 1.25 & 0 & 0 & -0.625 & 1.25 \\ 0 & 0 & 4 & -0.4 & 1 & -0.4 \end{array} \right],$$

y que con la primera columna a la derecha de la matriz diagonal produce

$$x_1 = \frac{1.2}{4} = 0.3, \quad x_2 = \frac{0}{1.25} = 0, \quad x_3 = \frac{-0.4}{4} = -0.1,$$

con la segunda columna

$$x_1 = \frac{-5}{4} = -1.25, \quad x_2 = \frac{-0.625}{1.25} = -0.5, \quad x_3 = 0.25$$

De igual manera con la tercera columna para llegar a

$$A^{-1} = \begin{bmatrix} 0.3 & -1.25 & 2.3 \\ 0 & -0.5 & 1.0 \\ -0.1 & 0.25 & -0.1 \end{bmatrix}$$

Los métodos de eliminación vistos proporcionan la solución del sistema  $A \mathbf{x} = \mathbf{b}$ , el  $\det A$  y  $A^{-1}$ , siempre que  $A$  sea no singular.

Obsérvese por otro lado que si se tiene un conjunto de vectores  $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$  de  $n$  componentes cada uno y se quieren ortogonalizar, se aplica alguna de las eliminaciones vistas al conjunto dado tomado como una matriz. La técnica de Gauss con pivoteo también puede aplicarse —por ejemplo— para determinar si dicho conjunto es linealmente independiente o no (cuando un elemento pivote  $a_{i,i}$  es igual a cero, la fila correspondiente es linealmente dependiente de las filas anteriores).

La sección que sigue puede omitirse sin pérdida de continuidad en los siguientes temas.

**CUENTA DE OPERACIONES**

Para establecer la velocidad de cálculo y el “trabajo computacional”, se requiere conocer cuántos cálculos de los diferentes tipos se realizan. Considérese para ello la reducción del sistema general

$$\begin{array}{rcl} a_{1,1}x_1 + a_{1,2}x_2 + \dots + a_{1,n}x_n & = & b_1 \\ a_{2,1}x_1 + a_{2,2}x_2 + \dots + a_{2,n}x_n & = & b_2 \\ \vdots & & \vdots \\ a_{n,1}x_1 + a_{n,2}x_2 + \dots + a_{n,n}x_n & = & b_n \end{array} \quad (3.61)$$



a la forma triangular

$$\begin{array}{ccccccc}
 t_{1,1}x_1 & + & t_{1,2}x_2 & + & \dots & + & t_{1,n}x_n & = & c_1 \\
 & & t_{2,2}x_2 & + & \dots & + & t_{2,n}x_n & = & c_2 \\
 & & & & & & \cdot & & \\
 & & & & & & \cdot & & \\
 & & & & & & \cdot & & \\
 & & & & & & t_{n,n}x_n & = & c_n
 \end{array} \quad (3.62)$$

o en notación matricial más compacta de  $[A | \mathbf{b}]$  a  $[T | \mathbf{c}]$ , matrices ambas de  $n \times (n + 1)$ . Sea

$M_n$  = número de multiplicaciones o divisiones

$S_n$  = número de sumas y restas

necesarias para ir del sistema 3.61 al 3.62.

Evidentemente  $M_1 = 0$  y  $S_1 = 0$ , ya que cualquier matriz  $A$  de  $1 \times 1$  es triangular. Si  $n > 1$ , se considera la eliminación en la primera columna. Si la primera columna de  $A$  es distinta del vector cero, generalmente se intercambian filas a fin de llevar el elemento de máximo valor absoluto de la primera columna a la posición  $(1, 1)$ . Denomínese de nuevo  $[A | \mathbf{b}]$  el sistema resultante de este intercambio. Ahora debe restarse un múltiplo de la nueva primera fila:

$$a_{1,1} \ a_{1,2} \ a_{1,3} \ \dots \ a_{1,n} \ b_1$$

de cada fila:

$$a_{i,1} \ a_{i,2} \ a_{i,3} \ \dots \ a_{i,n} \ b_i \quad 2 \leq i \leq n \quad (3.63)$$

para producir filas de la forma:

$$0 \ a'_{i,2} \ a'_{i,3} \ \dots \ a'_{i,n} \ b'_i \quad 2 \leq i \leq n \quad (3.64)$$

Explícitamente, si  $r_i = a_{i,1} / a_{1,1}$ ,

$$\begin{aligned}
 a'_{i,j} &= a_{i,j} - r_i a_{1,j} \\
 b'_i &= b_i - r_i b_1
 \end{aligned} \quad 2 \leq j \leq n \quad (3.65)$$

Se efectúa una división para producir  $r_i$ . La fórmula 3.65 requiere  $n$  multiplicaciones y un número igual de restas. Como se forman  $(n-1)$  filas, la eliminación en la primera columna se logra con:

$$\begin{aligned}
 (n+1)(n-1) &\text{ divisiones o multiplicaciones y} \\
 n(n-1) &\text{ restas.}
 \end{aligned} \quad (3.66)$$

La primera columna ya tiene ceros debajo de la posición  $(1, 1)$ . Queda por reducir la matriz de  $(n-1) \times n$ , matriz debajo de la primera fila y a la derecha de la primera columna. De la fórmula 3.66, se obtienen las fórmulas

$$\begin{aligned}
 M_n &= (n+1)(n-1) + M_{n-1} \\
 S_n &= n(n-1) + S_{n-1}
 \end{aligned} \quad n \geq 2 \quad (3.67)$$

Como  $M_1 = S_1 = 0$ , se tiene para  $n \geq 2$

$$\begin{aligned}
 M_n &= (2+1)1 + (3+1)2 + \dots + (n+1)(n-1) \\
 S_n &= 2(1) + 3(2) + \dots + n(n-1)
 \end{aligned} \quad (3.68)$$

Fácilmente se verifica por inducción que:

$$\sum_{t=1}^{n-1} t = \frac{1}{2} (n-1) n; \quad \sum_{t=1}^n t^2 = \frac{1}{6} (n-1) n (2n-1);$$

Por tanto, como

$$M_n = \sum_{t=1}^{n-1} (t+1+t)$$

y

$$S_n = \sum_{t=1}^{n-1} (t+1)t$$

Entonces:

$$M_n = \frac{1}{6} (n-1)n (2n-1) + (n-1)n \quad (3.69)$$

$$S_n = \frac{1}{6} (n-1)n (2n-1) + \frac{1}{2} (n-1)n$$

Se determinará el número  $m_n$  de multiplicaciones o divisiones y el número  $s_n$  de sumas o restas requeridas para resolver el sistema triangular  $[T|x] = c$ . Sean  $n \geq 2$  y todas las  $t_{i,i} \neq 0$ . Supóngase que se han calculado  $x_n, x_{n-1}, \dots, x_2$ ; llámense  $m_{n-1}$  y  $s_{n-1}$  las operaciones realizadas para ello.

Sea ahora

$$x_1 = \frac{c_1 - t_{1,2}x_2 - \dots - t_{1,n}x_n}{t_{1,1}} \quad (3.70)$$

El cálculo de  $x_1$  requiere  $(n-1)$  multiplicaciones, una división y  $(n-1)$  restas. Entonces, para  $n \geq 2$

$$m_n = (n-1+1) + m_{n-1} \quad (3.71)$$

$$s_n = (n-1) + s_{n-1}$$

Como  $m_1 = 1$  y  $s_1 = 0$ , se tiene

$$\begin{aligned} m_n &= 1 + 2 + 3 + \dots + n = \frac{1}{2} n (n+1) \\ s_n &= 1 + 2 + 3 + \dots + (n-1) = \frac{1}{2} (n-1) n \end{aligned} \quad (3.72)$$

El resultado final se resume a continuación.

El sistema 3.61 con matriz coeficiente  $A$  y determinante distinto de cero, puede resolverse por el método de eliminación con pivoteo con

$$\begin{aligned} u_n &= M_n + m_n = \frac{1}{6} (n-1)n (2n-1) + (n-1)n + \frac{1}{2} n (n+1) \\ &= \frac{1}{3} n^3 + n^2 - \frac{1}{3} n \text{ multiplicaciones o divisiones y} \\ v_n &= S_n + s_n = \frac{1}{6} (n-1)n (2n-1) + \frac{1}{2} (n-1)n + \frac{1}{2} (n-1)n \\ &= \frac{1}{3} n^3 + \frac{1}{2} n^2 - \frac{5}{6} n \text{ sumas o restas.} \end{aligned} \quad (3.73)$$

Obviamente, el "trabajo computacional" para resolver la ecuación 3.61 es función del número de operaciones necesarias (Ec. 3.73); por tanto, puede decirse que es proporcional a  $n^3$ . Por otro lado, las necesidades de memoria de máquina serán proporcionales a  $n^2$ .

## SISTEMAS ESPECIALES

Con frecuencia, la matriz coeficiente del sistema  $A \mathbf{x} = \mathbf{b}$  por resolver es simétrica, o bien gran número de sus componentes son cero (matrices dispersas). En estos casos algunos de los métodos conocidos pueden adaptarse, con lo cual se reduce el trabajo computacional y la memoria de máquina. Primero se tratará el caso de las matrices bandeadas (matrices dispersas particulares); las matrices simétricas serán abordadas como un caso particular de los métodos L-U.

Primero se darán algunos ejemplos particulares de matrices bandeadas

$$\begin{bmatrix} 2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 7 & 0 \\ 0 & 0 & 0 & 0 & 6 \end{bmatrix} \quad \begin{bmatrix} 4 & 0 & 0 & 0 & 0 \\ 7 & 8 & 1 & 0 & 0 \\ 0 & 0 & 5 & 2 & 0 \\ 0 & 0 & 1 & 3 & 5 \\ 0 & 0 & 0 & 3 & 4 \end{bmatrix} \quad \begin{bmatrix} 8 & 7 & 6 & 0 & 0 \\ 9 & 3 & 0 & -2 & 0 \\ 3 & -1 & 8 & 9 & 10 \\ 0 & 0 & 3 & 5 & 8 \\ 0 & 0 & 7 & 4 & 0 \end{bmatrix}$$

Diagonal

Tridiagonal

Pentadiagonal

Generalizando: Una matriz  $A$  de  $n \times n$  es tridiagonal si

$$a_{i,j} = 0 \text{ siempre que } |i - j| > 1,$$

pentadiagonal si

$$a_{i,j} = 0 \text{ siempre que } |i - j| > 2, \text{ etcétera.}$$

El ancho de banda es 1, 3, 5, etc., en las matrices diagonales, tridiagonales, pentadiagonales, etc., respectivamente.

Enseguida se adapta la eliminación de Gauss para la solución del sistema tridiagonal  $A \mathbf{x} = \mathbf{b}$ ; es decir,  $A$  es tridiagonal.

## MÉTODO DE THOMAS

Sea el sistema tridiagonal de tres ecuaciones en tres incógnitas

$$\begin{aligned} b_1 x_1 + c_1 x_2 &= d_1 \\ a_2 x_1 + b_2 x_2 + c_2 x_3 &= d_2 \\ a_3 x_2 + b_3 x_3 &= d_3 \end{aligned}$$

## TRIANGULARIZACIÓN

Si  $b_1 \neq 0$ , se elimina  $x_1$  **sólo** en la segunda ecuación, con lo que se obtiene como nueva segunda ecuación:

$$b'_2 x_2 + c'_2 x_3 = d'_2$$

con

$$b'_2 = b_2 - a_2 c_1 / b_1; \quad c'_2 = c_2; \quad d'_2 = d_2 - a_2 d_1 / b_1$$

Si  $b'_2 \neq 0$ ,  $x_2$  se elimina **sólo** en la tercera ecuación, y así se obtiene como nueva tercera ecuación

$$b'_3 x_3 = d'_3$$

con

$$b'_3 = b_3 - a_3 c'_2 / b'_2; \quad d'_3 = d_3 - a_3 d'_2 / b'_2$$

Generalizando: para un sistema tridiagonal de  $n$  ecuaciones en  $n$  incógnitas.



**TRIANGULARIZACIÓN**

Para  $i = 1, 2, \dots, n-1$

si  $b'_i \neq 0$  se elimina  $x_i$  sólo en la  $(i+1)$ -ésima ecuación, con lo que se obtiene como nueva  $(i+1)$ -ésima ecuación

$$b'_{i+1} x_{i+1} + c'_{i+1} x_{i+2} = d'_{i+1}$$

con

$$b'_{i+1} = b_{i+1} - a_{i+1} c'_i / b'_i; \quad c'_{i+1} = c_i; \quad d'_{i+1} = d_{i+1} - a_{i+1} d'_i / b'_i$$

**SUSTITUCIÓN REGRESIVA**

$$x_n = d'_n / b'_n$$

y para  $i = n-1, n-2, \dots, 1$

$$x_i = \frac{d'_i - c'_i x_{i+1}}{b'_i}$$

Esta simplificación del algoritmo de Gauss, válida para sistemas tridiagonales se conoce como método de Thomas. Con su aplicación se consiguen las siguientes ventajas:

- La memoria de máquina se reduce al no tener que almacenar los elementos de  $A$  que son cero. Obsérvese que en lugar de almacenar la matriz  $A$ , se guardan sólo los vectores  $\mathbf{a} = [a_1, a_2, \dots, a_n]$ ,  $\mathbf{b} = [b_1, b_2, \dots, b_n]$  y  $\mathbf{c} = [c_1, c_2, \dots, c_n]$  con  $a_i = c_n = 0$ , empleando  $3n$  localidades en lugar de  $n \times n$  localidades, ventaja muy importante cuando  $n$  es grande ( $n \geq 50$ ).
- No se requiere pivotear.
- Sólo se elimina durante el  $i$ -ésimo paso de la triangularización la variable  $x_i$  en la ecuación  $i+1$ , con lo que se reduce el número de operaciones.
- Por último, en la sustitución regresiva debe reemplazarse sólo  $x_{i+1}$  en la  $i$ -ésima ecuación para obtener  $x_i$ .

**Ejemplo 3.31**

Resuelva el sistema tridiagonal

$$\begin{aligned} 3x_1 - 2x_2 &= 1.0 \\ x_1 + 5x_2 - 0.2x_3 &= 5.8 \\ 4x_2 + 7x_3 &= 11.0, \end{aligned}$$

por el método de Thomas.

**Solución** En este sistema

$$\mathbf{b} = \begin{bmatrix} 3 \\ 5 \\ 7 \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} 0 \\ 1 \\ 4 \end{bmatrix}, \quad \mathbf{c} = \begin{bmatrix} -2 \\ -0.2 \\ 0 \end{bmatrix} \quad \text{y} \quad \mathbf{d} = \begin{bmatrix} 1.0 \\ 5.8 \\ 11.0 \end{bmatrix}$$

Como  $b_1 \neq 0$  se calculan las componentes de la nueva segunda fila

$$b'_2 = b_2 - a_2 c_1 / b_1 = 5 - 1(-2) / 3 = 5.6666$$

y

$$c'_2 = c_2 = -0.2$$

$$d'_2 = d_2 - a_2 d_1 / b_1 = 5.8 - 1(1/3) = 5.4666$$

Como  $b'_2 \neq 0$ , se forma la nueva tercera fila

$$b'_3 = b_3 - a_3 \cdot c'_2 / b'_2 = 7 - 4(-0.2) / 5.666 = 7.141176$$

$$d'_3 = d_3 - a_3 \cdot d'_2 / b'_2 = 11.0 - 4(5.4666) / 5.6666 = 7.1411760$$

El sistema equivalente resultante es:

$$3x_1 - 2x_2 = 1.0$$

$$5.6666x_2 - 0.2x_3 = 5.4666$$

$$7.141176x_3 = 7.141176$$

y por sustitución regresiva se llega a:

$$x_3 = d'_3 / b'_3 = 7.141176 / 7.141176 = 1$$

$$x_2 = (d'_2 - c'_2 x_3) / b'_2 = (5.4666 - 0.2(1)) / 5.6666 = 1$$

$$x_1 = (d'_1 - c'_1 x_2) / b'_1 = (1.0 - (-2)(1)) / 3 = 1$$

Nótese que  $d'_1 = d_1$  y  $b'_1 = b_1$

Para realizar los cálculos puede usarse el siguiente guión de Matlab:



```
b=[3 5 7]
a=[0 1 4]
c=[-2 -0.2 0]
d=[1 5.8 11]
for i=2:3
    b(i)=b(i)-a(i)*c(i-1)/b(i-1);
    d(i)=d(i)-a(i)*d(i-1)/b(i-1);
end
x(3)=d(3)/b(3);
for i=2:-1:1
    x(i)=(d(i)-c(i)*x(i+1))/b(i);
end
x
```

A continuación se da el algoritmo de Thomas.

### ALGORITMO 3.5 Método de Thomas

Para obtener la solución  $\mathbf{x}$  del sistema triadiagonal  $A \mathbf{x} = \mathbf{b}$  proporcionar los

DATOS: El número de ecuaciones  $N$ , los vectores  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ , y el vector de términos independientes  $\mathbf{d}$ .

RESULTADOS: El vector solución  $\mathbf{x}$  o mensaje de falla "EL SISTEMA NO TIENE SOLUCIÓN".

PASO 1. Hacer  $I=1$ .

PASO 2. Mientras  $I \leq N-1$ , repetir los pasos 3 a 6.

PASO 3. Si  $b(I) \neq 0$  continuar. De otro modo IMPRIMIR el mensaje "EL SISTEMA NO TIENE SOLUCIÓN" y TERMINAR.

PASO 4. Hacer  $b(I+1) = b(I+1) - a(I+1) * c(I) / b(I)$ .

- PASO 5. Hacer  $d(I+1) = d(I+1) - a(I+1) * d(I) / b(I)$   
 PASO 6. Hacer  $I = I + 1$   
 PASO 7. Si  $b(N) \neq 0$  continuar. De otro modo IMPRIMIR mensaje "EL SISTEMA NO TIENE SOLUCIÓN" y TERMINAR.  
 PASO 8. Hacer  $x(N) = d(N) / b(N)$   
 PASO 9. Hacer  $I = N - 1$   
 PASO 10. Mientras  $I \geq 1$ , repetir los pasos 11 y 12.  
 PASO 11. Hacer  $x(I) = (d(I) - c(I) * x(I+1) / b(I) * x(I+1)) / b(I)$   
 PASO 12. Hacer  $I = I - 1$   
 PASO 13. IMPRIMIR el vector solución  $x$  y TERMINAR.

## MÉTODOS DE FACTORIZACIÓN

### Factorización de matrices en matrices triangulares

La eliminación de Gauss aplicada al sistema (véase ejemplo 3.28)

$$\begin{aligned} 4x_1 - 9x_2 + 2x_3 &= 5 \\ 2x_1 - 4x_2 + 6x_3 &= 3 \\ x_1 - x_2 + 3x_3 &= 4 \end{aligned}$$

condujo en su fase de triangularización al sistema equivalente

$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 0 & 0.5 & 5 & 0.5 \\ 0 & 0 & -10 & 1.5 \end{array} \right]$$

donde se aprecia una matriz triangular superior de orden 3 que se denotará como  $U$

$$U = \begin{bmatrix} 4 & -9 & 2 \\ 0 & 0.5 & 5 \\ 0 & 0 & -10 \end{bmatrix}$$

Ahora se define una matriz triangular inferior  $L$  de orden 3, con números 1 a lo largo de la diagonal principal y con  $l_{ij}$  igual al factor que permitió eliminar el elemento  $a_{ij}$  del sistema 3.43 (por ejemplo, a fin de eliminar  $a_{2,1} = 2$  se utilizó el factor  $l_{2,1} = 2/4$ ; para eliminar  $a_{3,1} = 1$ , el factor  $l_{3,1} = 1/4$ , y para hacer cero a  $a_{3,2} = -1$  se empleó  $l_{3,2} = 1.25/0.5$ ). Así, la matriz  $L$  queda

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 2/4 & 1 & 0 \\ 1/4 & 1.25/0.5 & 1 \end{bmatrix},$$

cuyo producto con  $U$  resulta en

$$\begin{matrix} \begin{bmatrix} 1 & 0 & 0 \\ 2/4 & 1 & 0 \\ 1/4 & 1.25/0.5 & 1 \end{bmatrix} & \begin{bmatrix} 4 & -9 & 2 \\ 0 & 0.5 & 5 \\ 0 & 0 & -10 \end{bmatrix} & = A, \\ L & U & \end{matrix}$$

la matriz coeficiente del sistema original.



Esta descomposición de  $A$  en los factores  $L$  y  $U$  es cierta en general cuando la eliminación de Gauss puede aplicarse al sistema  $A \mathbf{x} = \mathbf{b}$  sin intercambio de filas, o equivalentemente si y sólo si los determinantes de las submatrices de  $A$  son todos distintos de cero

$$|a_{1,1}| \neq 0, \begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} \neq 0, \dots, \begin{vmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ a_{n,1} & \dots & a_{n,n} \end{vmatrix} \neq 0$$

El resultado anterior permite revolver el sistema  $A \mathbf{x} = \mathbf{b}$ , ya que sustituyendo  $A$  por  $LU$  se tiene

$$LU \mathbf{x} = \mathbf{b}$$

**Solución** Se hace  $U \mathbf{x} = \mathbf{c}$ , donde  $\mathbf{c}$  es un vector desconocido  $[c_1 \ c_2 \ c_3 \ \dots \ c_n]^T$ , que se puede obtener fácilmente resolviendo el sistema

$$L \mathbf{c} = \mathbf{b},$$

con sustitución progresiva o hacia adelante, ya que  $L$  es triangular inferior (en el sistema del Ejemplo 3.28,  $\mathbf{c}$  resulta  $[5 \ 0.5 \ 1.5]^T$ ).

Una vez calculado  $\mathbf{c}$ , se resuelve

$$U \mathbf{x} = \mathbf{c}$$

con sustitución regresiva, ya que  $U$  es triangular superior y de esa manera se obtiene el vector solución  $\mathbf{x}$  (el sistema particular que se ha trabajado da  $\mathbf{x} = [6.95 \ 2.5 \ -0.15]^T$ ).

### MÉTODOS DE DOOLITTLE Y CROUT

Aun cuando las matrices  $L$  y  $U$  pueden obtenerse en la triangularización de la matriz aumentada  $[A \mid \mathbf{b}]$ , es deseable encontrar un método más directo para su determinación. Esto es factible analizando la factorización de  $A$  en las matrices generales de orden tres  $L$  y  $U$ , dadas a continuación.

$$\begin{bmatrix} l_{1,1} & 0 & 0 \\ l_{2,1} & l_{2,2} & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} \end{bmatrix} \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} \\ 0 & u_{2,2} & u_{2,3} \\ 0 & 0 & u_{3,3} \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \\ a_{3,1} & a_{3,2} & a_{3,3} \end{bmatrix}$$

#### Análisis:

Se multiplican

a) Primera fila de  $L$  por las tres columnas de  $U$

$$l_{1,1}u_{1,1} = a_{1,1}$$

$$l_{1,1}u_{1,2} = a_{1,2}$$

$$l_{1,1}u_{1,3} = a_{1,3}$$

b) Segunda fila de  $L$  por las tres columnas de  $U$

$$l_{2,1}u_{1,1} = a_{2,1}$$

$$l_{2,1}u_{1,2} + l_{2,2}u_{2,2} = a_{2,2}$$

$$l_{2,1}u_{1,3} + l_{2,2}u_{2,3} = a_{2,3}$$

c) Tercera fila de  $L$  por las tres columnas de  $U$

$$\begin{aligned} l_{3,1} u_{1,1} &= a_{3,1} \\ l_{3,1} u_{1,2} + l_{3,2} u_{2,2} &= a_{3,2} \\ l_{3,1} u_{1,3} + l_{3,2} u_{2,3} + l_{3,3} u_{3,3} &= a_{3,3}, \end{aligned}$$

se llega a un sistema de nueve ecuaciones en 12 incógnitas  $l_{1,1}, l_{2,1}, l_{2,2}, l_{3,1}, l_{3,2}, l_{3,3}, u_{1,1}, u_{1,2}, u_{1,3}, u_{2,2}, u_{2,3}, u_{3,3}$ , por lo que será necesario establecer tres condiciones arbitrarias sobre las incógnitas para resolver dicho sistema. La forma de seleccionar las condiciones ha dado lugar a diferentes métodos; por ejemplo, si se toman de modo que  $l_{1,1} = l_{2,2} = l_{3,3} = 1$ , se obtiene el **método de Doolittle**; si en cambio se selecciona  $u_{1,1} = u_{2,2} = u_{3,3} = 1$ , el algoritmo resultante es llamado **método de Crout**.

Se continuará el desarrollo de la factorización. Tómese

$$l_{1,1} = l_{2,2} = l_{3,3} = 1$$

Con estos valores se resuelven las ecuaciones directamente en el orden en que están dadas

$$\text{De (a)} \quad u_{1,1} = a_{1,1}, \quad u_{1,2} = a_{1,2}, \quad u_{1,3} = a_{1,3}, \quad (3.74)$$

De (b) y sustituyendo los resultados (Ec. 3.74)

$$\begin{aligned} l_{2,1} &= a_{2,1} / u_{1,1} = a_{2,1} / a_{1,1}, \\ u_{2,2} &= a_{2,2} - l_{2,1} u_{1,2} = a_{2,2} - \frac{a_{2,1}}{a_{1,1}} a_{1,2} \\ u_{2,3} &= a_{2,3} - l_{2,1} u_{1,3} = a_{2,3} - \frac{a_{2,1}}{a_{1,1}} a_{1,3} \end{aligned} \quad (3.75)$$

De (c) y sustituyendo los resultados de las ecuaciones 3.74 y 3.75

$$\begin{aligned} l_{3,1} &= a_{3,1} / u_{1,1} = a_{3,1} / a_{1,1} \\ l_{3,2} &= \frac{a_{3,2} - l_{3,1} u_{1,2}}{u_{2,2}} = \frac{a_{3,2} - \frac{a_{3,1}}{a_{1,1}} a_{1,2}}{a_{2,2} - \frac{a_{2,1}}{a_{1,1}} a_{1,2}} \\ u_{3,3} &= a_{3,3} - l_{3,1} u_{1,3} - l_{3,2} u_{2,3} = \\ &= a_{3,3} - \frac{a_{3,1}}{a_{1,1}} a_{1,3} - \left[ \frac{a_{3,2} - \frac{a_{3,1}}{a_{1,1}} a_{1,2}}{a_{2,2} - \frac{a_{2,1}}{a_{1,1}} a_{1,2}} \right] \left[ a_{2,3} - \frac{a_{2,1}}{a_{1,1}} a_{1,3} \right] \end{aligned} \quad (3.76)$$

Las ecuaciones 3.74, 3.75 y 3.76, convenientemente generalizadas constituyen un método directo para la obtención de  $L$  y  $U$ , con la ventaja sobre la triangularización de que no se tiene que escribir repetidamente las ecuaciones o arreglos modificados de  $A \mathbf{x} = \mathbf{b}$ . A continuación se resuelve un ejemplo.

**Ejemplo 3.32**

Resuelva por el método de Doolittle el sistema

$$4x_1 - 9x_2 + 2x_3 = 5$$

$$2x_1 - 4x_2 + 6x_3 = 3$$

$$x_1 - x_2 + 3x_3 = 4$$

**Solución** Con  $l_{1,1} = l_{2,2} = l_{3,3} = 1$ , se procede alcálculo de la primera fila de  $U$ 

$$u_{1,1} = 4; u_{1,2} = -9; u_{1,3} = 2,$$

cálculo de la primera columna de  $L$ 

$$l_{1,1} = 1 \text{ (dato)}; l_{2,1} = 2/4 = 0.5; l_{3,1} = 1/4 = 0.25$$

cálculo de la segunda fila de  $U$ 

$$u_{2,1} = 0 \text{ (recuérdese que } U \text{ es triangular superior)}$$

$$u_{2,2} = -4 - (2/4)(-9) = 0.5; u_{2,3} = 6 - (2/4)(2) = 5$$

cálculo de la segunda columna de  $L$ 

$$l_{1,2} = 0 \text{ (ya que } L \text{ es triangular inferior)}$$

$$l_{2,2} = 1 \text{ (dato)}, l_{3,2} = (-1 - (1/4)(-9)) / (-4 - (2/4)(-9)) = 2.5$$

cálculo de la tercera fila de  $U$ , o más bien sus elementos faltantes, ya que por ser triangular superior

$$u_{3,1} = u_{3,2} = 0$$

$$u_{3,3} = 3 - (1/4)(2) - [(-1 - (1/4)(-9)) / (-4 - (2/4)(-9))](6 - (2/4)(2)) = -10$$

Con esto se finaliza la factorización\*.

Las matrices  $L$  y  $U$  quedan como sigue

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.25 & 2.5 & 1 \end{bmatrix}; \quad U = \begin{bmatrix} 4 & -9 & 2 \\ 0 & 0.5 & 5 \\ 0 & 0 & -10 \end{bmatrix}$$

cuyo producto, como ya se comprobó, da  $A$ .Se resuelve el sistema  $L\mathbf{c} = \mathbf{b}$ , donde  $\mathbf{b}$  es el vector de términos independientes del sistema original

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 0.25 & 2.5 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 3 \\ 4 \end{bmatrix}$$

$$c_1 = 5; c_2 = 3 - 0.5(5) = 0.5$$

$$c_3 = 4 - 0.25(5) - 2.5(0.5) = 1.5,$$

y, finalmente, al resolver el sistema  $U\mathbf{x} = \mathbf{c}$  se tiene la solución del sistema original:

\* Los cálculos se han llevado en el orden fila-columna, fila-columna, etc., por convenir a la elaboración de los algoritmos correspondientes.



$$\begin{bmatrix} 4 & -9 & 2 \\ 0 & 0.5 & 5 \\ 0 & 0 & -10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 5 \\ 0.5 \\ 1.5 \end{bmatrix}$$

$$x_3 = -0.15$$

$$x_2 = (0.5 - 5(-0.15)) / 0.5 = 2.5$$

$$x_1 = (5 + 9(2.5) - 2(-0.15)) / 4 = 6.95$$

$$\mathbf{x} = \begin{bmatrix} 6.95 \\ 2.5 \\ -0.15 \end{bmatrix}$$

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
clear
format short
A=[4 -9 2 ; 2 -4 6 ; 1 -1 3 ]
b=[5 3 4]
[L,U,P]=lu(A);
L
U
% Se multiplica b por P para que refleje los
% intercambios hechos en A para llegar a L y U
b=b*P
c(1)=b(1);
c(2)=b(2)-L(2,1)*c(1);
c(3)=b(3)-L(3,1)*c(1)-L(3,2)*c(2);
c
x(3)=c(3)/U(3,3);
x(2)=(c(2)-U(2,3)*x(3))/U(2,2);
x(1)=(c(1)-U(1,2)*x(2)-U(1,3)*x(3))/U(1,1);
x
```



```
e3_32()
Prgrm
[4,-9,2;2,-4,6;1,-1,3]→a:[5,3,4]→b:ClrIO
LU a,L,u,p:Disp L:Pause:Disp u:Pause
b*P→b:b[1,1]→c[1]:b[1,2]-L[2,1]*c[1]→c[2],
b[1,3]-L[3,1]*c[1]-L[3,2]*c[2]→c[3]
Disp c:Pause
0→x[1]:0→x[2]: c[3]/u[3,3]→x[3]
(c[2]-u[2,3]*x[3])/u[2,2]→x[2]
(c[1]-u[1,2]*x[2]-u[1,3]*x[3])/u[1,1]→x[1]
Disp x
```

Los resultados intermedios difieren de los anotados anteriormente, debido a que Matlab realiza intercambios de fila para llegar a las matrices  $L$  y  $U$ . Sin embargo, los resultados finales son los mismos.

Las ecuaciones 3.74, 3.75 y 3.76 se generalizan para factorizar la matriz coeficiente del sistema  $A \mathbf{x} = \mathbf{b}$ , que puede resolverse por eliminación de Gauss sin intercambio de filas; se tiene entonces

$$\begin{aligned} u_{i,j} &= a_{i,j} - \sum_{k=1}^{j-1} l_{i,k} u_{k,j}; j = i, i+1, \dots, n \\ l_{i,j} &= \frac{1}{u_{j,j}} (a_{i,j} - \sum_{k=1}^{j-1} u_{k,j} l_{i,k}); i = j+1, \dots, n \\ l_{i,i} &= 1; i = 1, 2, \dots, n \end{aligned} \quad (3.77)$$

con la convención en las sumatorias que  $\sum_{k=1}^0 = 0$ .

Puede observarse al seguir las ecuaciones 3.74, 3.75 y 3.76 o bien las ecuaciones 3.77, que una vez empleada  $a_{i,j}$  en el cálculo de  $u_{i,j}$  o  $l_{i,j}$  según sea el caso, esta componente de  $A$  no vuelve a emplearse como tal, por lo que las componentes de  $L$  y  $U$  generadas pueden guardarse en  $A$  y ahorrar memoria de esa manera. El siguiente algoritmo de factorización de  $A$  ilustra esto.

### ALGORITMO 3.6 Factorización directa

Para factorizar una matriz  $A$  de orden  $N$  en el producto de las matrices  $L$  y  $U$  triangulares inferior y superior, respectivamente, con  $l_{i,i} = 1; i = 1, 2, \dots, N$ , proporcionar los

DATOS: El orden  $N$  y las componentes de la matriz  $A$ .

RESULTADOS: Las matrices  $L$  y  $U$  en  $A$  o mensaje de falla "LA FACTORIZACIÓN NO ES POSIBLE".

- PASO 1. Si  $A(1,1) = 0$  IMPRIMIR "LA FACTORIZACIÓN NO ES POSIBLE" y TERMINAR. De otro modo continuar.
- PASO 2. Hacer  $J = 1$ .
- PASO 3. Mientras  $J \leq N$ , repetir los pasos 4 a 25.
- PASO 4. Hacer  $I = J$ .
- PASO 5. Mientras  $I \leq N$ , repetir los pasos 6 a 13.
- PASO 6. Hacer  $SUMAT = 0$ .
- PASO 7. Si  $J = 1$  ir al paso 12. De otro modo continuar.
- PASO 8. Hacer  $K = 1$ .
- PASO 9. Mientras  $K \leq J - 1$ , repetir los pasos 10 y 11.
- PASO 10. Hacer  $SUMAT = SUMAT + A(J,K) * A(K,I)$ .
- PASO 11. Hacer  $K = K + 1$ .
- PASO 12. Hacer  $A(J,I) = A(J,I) - SUMAT$ .
- PASO 13. Hacer  $I = I + 1$ .
- PASO 14. Si  $J = N$  ir al paso 25. De otro modo continuar.
- PASO 15. Hacer  $I = J + 1$ .
- PASO 16. Mientras  $I \leq N$ , repetir los pasos 17 a 24.
- PASO 17. Hacer  $SUMAT = 0$ .
- PASO 18. Si  $J = 1$  ir al paso 23. De otro modo continuar.
- PASO 19. Hacer  $K = 1$ .
- PASO 20. Mientras  $K \leq J - 1$ , repetir los pasos 21 y 22.
- PASO 21. Hacer  $SUMAT = SUMAT + A(K,J) * A(I,K)$ .
- PASO 22. Hacer  $K = K + 1$ .
- PASO 23. Hacer  $A(I,J) = (A(I,J) - SUMAT) / A(J,J)$ .
- PASO 24. Hacer  $I = I + 1$ .

PASO 25. Hacer  $J = J + 1$

PASO 26. Si  $A(N,N) = 0$  IMPRIMIR "LA FACTORIZACIÓN NO ES POSIBLE" y TERMINAR. De otro modo continuar.

PASO 27. IMPRIMIR  $A$  y TERMINAR.

Obsérvese que cualquier elemento  $a_{i,i} = 0$ , impediría emplear este algoritmo; por otro lado, al no pivotar no se reducen en lo posible los errores de redondeo. Para hacer eficiente este algoritmo, debe incluirse un intercambio de filas como en la eliminación de Gauss con pivoteo. A continuación se presenta el algoritmo anterior, pero ahora con estas modificaciones.

### ALGORITMO 3.7 Factorización con pivoteo

Para factorizar una matriz  $A$  de orden  $N$  en el producto de las matrices  $L$  y  $U$  triangulares inferior y superior respectivamente, con  $l_{i,i} = 1$ ;  $i = 1, 2, \dots, N$ , con pivoteo parcial, proporcionar los

DATOS: El orden  $N$  y las componentes de la matriz  $A$ .

RESULTADOS: Las matrices  $L$  y  $U$  en  $A$  o mensaje de falla "LA FACTORIZACIÓN NO ES POSIBLE".

PASO 1. Hacer  $R = 0$  ( $R$  registra el número de intercambios de fila que se llevan a cabo).

PASO 2. Hacer  $J = 1$ .

PASO 3. Mientras  $J \leq N$ , repetir los pasos 4 a 11.

PASO 4. Si  $J = N$  ir al paso 10.

PASO 5. Encontrar PIVOTE y  $P$  (ver paso 5 de algoritmo 3.4).

PASO 6. Si PIVOTE = 0 IMPRIMIR "LA FACTORIZACIÓN NO ES POSIBLE" y TERMINAR. De otro modo continuar.

PASO 7. Si  $P = J$  ir al paso 10. De otro modo continuar.

PASO 8. Intercambiar la fila  $J$  con la fila  $P$  de  $A$ .

PASO 9. Hacer  $R = R + 1$ .

PASO 10. Realizar los pasos 4 a 24 de algoritmo 3.6.

PASO 11. Hacer  $J = J + 1$ .

PASO 12. Si  $A(N,N) = 0$  IMPRIMIR "LA FACTORIZACIÓN NO ES POSIBLE" y TERMINAR. De otro modo continuar.

PASO 13. IMPRIMIR  $A$  y TERMINAR.

A continuación se resuelve un sistema por el método de Doolittle usando la factorización con pivoteo.

### Ejemplo 3.33

Resuelva el sistema del ejemplo 3.29

$$\begin{aligned} 10x_1 + x_2 - 5x_3 &= 1 \\ -20x_1 + 3x_2 + 20x_3 &= 2 \\ 5x_1 + 3x_2 + 5x_3 &= 6 \end{aligned}$$

por el método de Doolittle, con pivoteo parcial

**Solución** Al intercambiar la primera y segunda filas resulta la matriz aumentada siguiente:

$$A = \left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ 10 & 1 & -5 & 1 \\ 5 & 3 & 5 & 6 \end{array} \right]$$



Como la nueva  $a_{1,1} \neq 0$ , se forma la primera fila de  $U$  y se guarda como primera fila de  $A$ .

$$a_{1,1} = u_{1,1} = -20, a_{1,2} = u_{1,2} = 3, a_{1,3} = u_{1,3} = 20$$

Cálculo de la primera columna de  $L$  y su registro, excepto  $l_{1,1}$ , como primera columna de  $A$

$$l_{1,1} = 1 \text{ (dato),}$$

$$a_{2,1} = l_{2,1} = 10/(-20) = -0.5$$

$$a_{3,1} = l_{3,1} = 5/(-20) = -0.25$$

La matriz  $A$  resultante entonces es:

$$A = \left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ -0.5 & 1 & -5 & 1 \\ -0.25 & 3 & 5 & 6 \end{array} \right]$$

Se busca el nuevo pivote en la parte relevante de la segunda columna (segunda y tercera filas) y resulta ser el elemento  $a_{3,2}$ .

Se intercambia la segunda fila con la tercera y entonces queda:

$$A = \left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ -0.25 & 3 & 5 & 6 \\ -0.5 & 1 & -5 & 1 \end{array} \right]$$

Cálculo de la segunda fila de  $U$  (mejor dicho de los elementos distintos de cero de dicha fila y almacenamiento de éstos en las posiciones correspondientes de  $A$ ):

$$a_{2,2} = u_{2,2} = 3(-0.25)(3) = 3.75$$

$$a_{2,3} = u_{2,3} = 5 - (-0.25)(20) = 10.0$$

Cálculo de la segunda columna de  $L$ ; es decir, de los elementos debajo de  $l_{2,2}$  y almacenamiento de éstos en las posiciones correspondientes de  $A$ .

$$a_{3,2} = l_{3,2} = \frac{1 - (-0.5)(3)}{3.75} = 0.666666$$

Con estos valores la matriz  $A$  resultante es

$$A = \left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ -0.25 & 3.75 & 10 & 6 \\ -0.5 & 0.6666 & -5 & 1 \end{array} \right]$$

Como  $a_{3,3} \neq 0$ , se calcula  $u_{3,3}$ , que constituye la parte relevante de la tercera fila de  $U$ , y se almacena en  $a_{3,3}$ .

$$a_{3,3} = u_{3,3} = -5 - (-0.5)(20) - (0.66666)(10) = -1.6666$$

con lo cual la matriz aumentada queda como sigue:

$$A = \left[ \begin{array}{ccc|c} -20 & 3 & 20 & 2 \\ -0.25 & 3.75 & 10 & 6 \\ -0.5 & 0.6666 & -1.6666 & 1 \end{array} \right]$$

Al resolver los sistemas

$$L \mathbf{c} = \mathbf{b}' \text{ con } L = \begin{bmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ -0.5 & 0.6666 & 1 \end{bmatrix} \text{ y } \mathbf{b}' = \begin{bmatrix} 2 \\ 6 \\ 1 \end{bmatrix}$$

se tiene:

$$c_1 = 2$$

$$c_2 = 6 + 0.25(2) = 6.5$$

$$c_3 = 1 + 0.5(2) - 0.6666(6.5) = -2.33329$$

y

$$U \mathbf{x} = \mathbf{c} \text{ con } U = \begin{bmatrix} -20 & 3 & 20 \\ 0 & 3.75 & 10 \\ 0 & 0 & -1.6666 \end{bmatrix} \text{ y } \mathbf{c} \text{ como arriba.}$$

se tiene

$$x_3 = \frac{-2.33329}{-1.66666} = 1.3999796$$

$$x_2 = \frac{6.5 - 10(1.3999796)}{3.75} = -1.9999456$$

$$x_1 = \frac{2 - 3(-1.9999456) - 20(1.3999796)}{-20} = 0.99999$$

Para realizar los cálculos puede usarse el siguiente guión de Matlab:



```
clear
A=[10 1 -5 ; -20 3 20 ; 5 3 5]
b=[1 2 6]
[L,U,P]=lu(A);
L
U
%Se multiplica P por b transpuesto para
%reflejar los intercambios hechos en A
b=P*b'
c(1)=b(1);
c(2)=b(2)-L(2,1)*c(1);
c(3)=b(3)-L(3,1)*c(1)-L(3,2)*c(2);
c
x(3)=c(3)/U(3,3);
x(2)=(c(2)-U(2,3)*x(3))/U(2,2);
x(1)=(c(1)-U(1,2)*x(2)-U(1,3)*x(3))/U(1,1);
x
```

A continuación se da el algoritmo de Doolittle

### ALGORITMO 3.8 Método de Doolittle

Para obtener la solución del sistema  $A \mathbf{x} = \mathbf{b}$  y el determinante de  $A$ , proporcionar los

DATOS:  $N$  el número de ecuaciones,  $A$  la matriz aumentada del sistema.

RESULTADOS: El vector solución  $\mathbf{x}$  y el determinante de  $A$  o mensaje "LA FACTORIZACIÓN NO ES POSIBLE".

- PASO 1. Realizar los pasos 1 al 12 del algoritmo 3.7.  
 PASO 2. Hacer  $c(1) = A(1, N+1)$ .  
 PASO 3. Hacer  $DET = A(1, 1)$ .  
 PASO 4. Hacer  $I = 2$ .  
 PASO 5. Mientras  $I \leq N$ , repetir los pasos 6 a 12.  
     PASO 6. Hacer  $DET = DET * A(I, I)$ .  
     PASO 7. Hacer  $c(I) = A(I, N+1)$ .  
     PASO 8. Hacer  $J = 1$ .  
     PASO 9. Mientras  $J \geq I-1$ , repetir los pasos 10 y 11.  
         PASO 10. Hacer  $c(I) = c(I) - A(I, J) * c(J)$ .  
         PASO 11. Hacer  $J = J + 1$ .  
     PASO 12. Hacer  $I = I + 1$ .  
 PASO 13. Hacer  $\mathbf{x}(N) = c(N)/A(N, N)$ .  
 PASO 14. Hacer  $I = N - 1$ .  
 PASO 15. Mientras  $I \geq 1$ , repetir los pasos 16 a 22.  
     PASO 16. Hacer  $\mathbf{x}(I) = c(I)$ .  
     PASO 17. Hacer  $J = I + 1$ .  
     PASO 18. Mientras  $J \leq N$ , repetir los pasos 19 y 20.  
         PASO 19. Hacer  $\mathbf{x}(I) = \mathbf{x}(I) - A(I, J) * \mathbf{x}(J)$ .  
         PASO 20. Hacer  $J = J + 1$ .  
     PASO 21. Hacer  $\mathbf{x}(I) = \mathbf{x}(I)/A(I, I)$ .  
     PASO 22. Hacer  $I = I - 1$ .  
 PASO 23. Hacer  $DET = DET * (-1) ** R$ .  
 PASO 24. IMPRIMIR  $\mathbf{x}$  y  $DET$  y TERMINAR.

### SISTEMAS SIMÉTRICOS

En el caso de que la matriz coeficiente del sistema  $A \mathbf{x} = \mathbf{b}$  sea simétrica, los cálculos de la factorización (si es posible) se simplifican, ya que la segunda de las ecuaciones 3.77 se reduce a:

$$l_{i,j} = \frac{a_{j,i}}{a_{j,j}} \quad i = j + 1, \dots, n; j = 1, 2, \dots, n-1 \quad (3.78)$$

Esto disminuye considerablemente el trabajo, en particular cuando  $n$  es grande.

### Ejemplo 3.34

Resuelva el sistema simétrico siguiente

$$\begin{bmatrix} 2 & 1 & 3 \\ 1 & 0 & 4 \\ 3 & 4 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$

**Solución** Cálculo de la primera fila de  $U$  y su registro en  $A$ .



$$\begin{aligned}a_{1,1} &= u_{1,1} = 2, \\a_{1,2} &= u_{1,2} = 1, \\a_{1,3} &= u_{1,3} = 3.\end{aligned}$$

Cálculo de los elementos relevantes de la primera columna de  $L$ , usando la ecuación 3.78 y su registro en  $A$ .

$$a_{2,1} = l_{2,1} = \frac{a_{1,2}}{a_{1,1}} = 0.5$$

$$a_{3,1} = l_{3,1} = \frac{a_{1,3}}{a_{1,1}} = 1.5$$

Cálculo de los elementos relevantes de la segunda fila de  $U$  y su registro en las posiciones correspondientes de  $A$

$$a_{2,2} = u_{2,2} = a_{2,2} - l_{2,1} u_{1,2} = 0 - 0.5 (1) = -0.5$$

$$a_{2,3} = u_{2,3} = a_{2,3} - l_{2,1} u_{1,3} = 4 - 0.5 (3) = 2.5$$

Cálculo de los elementos relevantes de la segunda columna de  $L$  mediante la ecuación 3.78 y su registro en las posiciones correspondientes de  $A$ .

$$a_{3,2} = l_{3,2} = \frac{a_{2,3}}{a_{2,2}} = -5$$

Finalmente se calcula la componente  $u_{3,3}$  (único elemento relevante de la tercera fila de  $U$ ) y se verifica su registro en  $a_{3,3}$ .

$$\begin{aligned}a_{3,3} = u_{3,3} &= a_{3,3} - l_{3,1} u_{1,3} - l_{3,2} u_{2,3} \\&= 3 - 1.5(3) - (-5)(2.5) = 11\end{aligned}$$

La factorización da como resultado

$$\begin{bmatrix} 2 & 1 & 3 \\ 0.5 & -0.5 & 2.5 \\ 1.5 & -5 & 11 \end{bmatrix}$$

Con la resolución del sistema  $L \mathbf{c} = \mathbf{b}$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0.5 & 1 & 0 \\ 1.5 & -5 & 1 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$$

se obtiene:  $\mathbf{c} = [0 \ 1 \ 8]^T$

y al resolver el sistema  $U \mathbf{x} = \mathbf{c}$

$$\begin{bmatrix} 2 & 1 & 3 \\ 0 & -0.5 & 2.5 \\ 0 & 0 & 11 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 8 \end{bmatrix}$$

se obtiene:

$$\mathbf{x} = \begin{bmatrix} -1.9091 \\ 1.6364 \\ 0.7273 \end{bmatrix}$$

Para realizar los cálculos puede usarse el siguiente gui3n de Matlab:



```
clear
A=[2 1 3; 1 0 4; 3 4 3]
b=[0 1 3]
A(2,1)=A(1,2)/A(1,1);
A(3,1)=A(1,3)/A(1,1);
A(2,2:3)=A(2,2:3)-A(2,1)*A(1,2:3);
A(3,2)=A(2,3)/A(2,2);
A(3,3)=A(3,3)-A(3,1:2)*A(1:2,3);
A
c(1)=b(1);
c(2)=b(2)-A(2,1)*c(1);
c(3)=b(3)-A(3,1:2)*c(1:2)';
c
x=[0 0 0];
x(3)=c(3)/A(3,3);
x(2)=(c(2)-A(2,3)*x(3))/A(2,2);
x(1)=(c(1)-A(1,2:3)*x(2:3)')/A(1,1);
x
```

Es importante observar que no se emplea pivoteo parcial y que si alguno de los elementos  $u_{i,i}$  resulta ser cero, este m3todo no es aplicable; como consecuencia, habr3 que recurrir al m3todo de Doolittle con pivoteo, por ejemplo, con lo cual se pierde la ventaja de que  $A$  es sim3trica.

A continuaci3n se da el algoritmo correspondiente.

### ALGORITMO 3.9 Factorizaci3n de matrices sim3tricas

Para factorizar una matriz  $A$  de orden  $n$  en el producto de las matrices  $L$  y  $U$  triangulares inferior y superior, respectivamente, con  $l_{i,i} = 1$ ;  $i = 1, 2, \dots, n$ , proporcionar los

DATOS: El orden  $N$  y las componentes de la matriz sim3trica  $A$ .

RESULTADOS: Las matrices  $L$  y  $U$  en  $A$  o mensaje de falla "LA FACTORIZACI3N NO ES POSIBLE".

PASO 1. Hacer  $J = 1$ .

PASO 2. Mientras  $J \leq N$ , repetir los pasos 3 al 15.

PASO 3. Hacer  $I = J$ .

PASO 4. Mientras  $I \leq N$ , repetir los pasos 5 a 13.

PASO 5. Hacer  $SUMAT = 0$ .

PASO 6. Si  $J = 1$  ir al paso 11. De otro modo continuar.

PASO 7. Hacer  $K = 1$ .

PASO 8. Mientras  $K \leq J - 1$ , repetir los pasos 9 y 10.

PASO 9. Hacer  $SUMAT = SUMAT + A(J,K) * A(K,I)$ .

PASO 10. Hacer  $K = K + 1$ .

PASO 11. Hacer  $A(J,I) = A(J,I) - SUMAT$ .

PASO 12. Si  $I > J$  Hacer  $A(I,J) = A(J,I) / A(J,J)$ . De otro modo continuar.

PASO 13. Hacer  $I = I + 1$ .

PASO 14. Si  $A(J,J) = 0$  IMPRIMIR "LA FACTORIZACIÓN NO ES POSIBLE" y TERMINAR. De otro modo continuar.

PASO 15. Hacer  $J = J + 1$ .

PASO 16. IMPRIMIR A y TERMINAR.

### MÉTODO DE CHOLSKY

Una matriz simétrica  $A$  cuyas componentes son números reales, es positiva definida si y sólo si los determinantes de las submatrices de  $A$  son positivos.

$$|a_{1,1}| > 0, \begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} > 0, \dots, \begin{vmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,n} \\ a_{2,1} & a_{2,2} & \dots & a_{2,n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_{n,1} & a_{n,2} & \dots & a_{n,n} \end{vmatrix} > 0$$

En el caso de tener un sistema  $A \mathbf{x} = \mathbf{b}$ , con  $A$  positiva definida, la factorización de  $A$  en la forma  $L U$  es posible y muy sencilla ya que toma la forma  $L L^T$ , donde  $L$  es triangular inferior:

$$L = \begin{vmatrix} l_{1,1} & 0 & \dots & 0 \\ l_{2,1} & l_{2,2} & \dots & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & 0 \\ l_{n,1} & l_{n,2} & \dots & l_{n,n} \end{vmatrix}$$

Los cálculos se reducen, ya que ahora basta estimar  $n(n+1)/2$  elementos (los  $l_{i,j} \neq 0$ ), en lugar de los  $n^2$  elementos de una factorización nominal (los  $l_{i,j}$  tales que  $i < j$  y los  $u_{i,j}$  tales que  $i \geq j$ ). El número de cálculos es prácticamente la mitad.

#### Ejemplo 3.35

Resuelva el sistema de ecuaciones lineales

$$\begin{bmatrix} 4 & 1 & 2 \\ 1 & 2 & 0 \\ 2 & 0 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

cuya matriz coeficiente es simétrica y positivamente definida.

#### Solución

##### Factorización de $A$

$$\begin{bmatrix} 4 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} l_{1,1} & 0 & 0 \\ l_{2,1} & l_{2,2} & 0 \\ l_{3,1} & l_{3,2} & l_{3,3} \end{bmatrix} \begin{bmatrix} l_{1,1} & l_{2,1} & l_{3,1} \\ 0 & l_{2,2} & l_{3,2} \\ 0 & 0 & l_{3,3} \end{bmatrix}$$



De la multiplicación de matrices se tiene

$$l_{1,1}^2 = a_{1,1}; \quad l_{1,1} = \pm \sqrt{a_{1,1}} = \pm 2 \text{ se toma el valor positivo de todas las raíces}$$

$$l_{1,1} = 2$$

$$l_{1,1} l_{2,1} = a_{1,2}; \quad l_{2,1} = a_{1,2} / l_{1,1} = 1/2 = 0.5$$

$$l_{1,1} l_{3,1} = a_{1,3}; \quad l_{3,1} = a_{1,3} / l_{1,1} = 2/2 = 1$$

$$l_{2,1}^2 + l_{2,2}^2 = a_{2,2}; \quad l_{2,2} = \sqrt{a_{2,2} - l_{2,1}^2}$$

$$l_{2,2} = \sqrt{2 - 0.5^2} = 1.32287$$

$$l_{2,1} l_{3,1} + l_{2,2} l_{3,2} = a_{2,3}; \quad l_{3,2} = \frac{-l_{2,1} l_{3,1} + a_{2,3}}{l_{2,2}}$$

$$l_{3,2} = -\frac{0.5(1)}{1.32287} = -0.37796$$

$$l_{3,1}^2 + l_{3,2}^2 + l_{3,3}^2 = a_{3,3}; \quad l_{3,3} = \sqrt{a_{3,3} - l_{3,1}^2 - l_{3,2}^2}$$

$$l_{3,3} = \sqrt{5 - 1 - 0.14286} = 1.96396$$

Al resolver el sistema:

$$L \mathbf{c} = \mathbf{b}$$

$$\begin{bmatrix} 2 & 0 & 0 \\ 0.5 & 1.32287 & 0 \\ 1 & -0.37796 & 1.96396 \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$$

$$c_1 = 0.5$$

$$c_2 = (2 - 0.5(0.5))/1.32287 = 1.32287$$

$$c_3 = (4 - 0.5 + 0.37796(1.32287))/1.96396 = 2.0367$$

Al resolver el sistema:

$$L^T \mathbf{x} = \mathbf{c}$$

$$\begin{bmatrix} 2 & 0.5 & 1 \\ 0 & 1.32287 & -0.37796 \\ 0 & 0 & 1.96396 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0.5 \\ 1.32287 \\ 2.0367 \end{bmatrix}$$

$$x_3 = 2.0367/1.96396 = 1.037$$

$$x_2 = (1.32287 + 0.37796(1.037))/1.32287 = 1.29629$$

$$x_1 = (0.5 - 0.5(1.29629) - 1.037)/2 = -0.59259$$

El vector solución es:

$$\mathbf{x} = \begin{bmatrix} -0.59259 \\ 1.29629 \\ 1.037 \end{bmatrix}$$

Para realizar los cálculos puede usarse el siguiente guión de Matlab que utiliza la función `chol(A)`, que devuelve la matriz triangular superior  $U$  de  $A$ , la cual deberá ser simétrica y positiva definida:



```
A= [4 1 2 ; 1 2 0; 2 0 5]
U=chol(A)
L=U'
b=[1 2 4]
c=inv(L)*b'
x=inv(L')*c
```

Las fórmulas de este algoritmo para un sistema de  $n$  ecuaciones son

$$\begin{aligned}
 l_{1,1} &= \sqrt{a_{1,1}} \\
 l_{i,1} &= a_{1,i} / l_{1,1} & i = 2, 3, \dots, n \\
 l_{i,i} &= \left( a_{i,i} - \sum_{k=1}^{i-1} l_{i,k}^2 \right)^{1/2} & i = 2, 3, \dots, n \\
 l_{j,i} &= \frac{1}{l_{i,i}} \left( a_{i,j} - \sum_{k=1}^{i-1} l_{i,k} l_{j,k} \right) & i = 2, 3, \dots, n \\
 & & j = i+1, i+2, \dots, n-1 \\
 l_{i,j} &= 0 & i < j
 \end{aligned}$$

A continuación se da el algoritmo para este método.

### ALGORITMO 3.10 Método de Cholesky

Para factorizar una matriz positiva definida en la forma  $L L^T$ , proporcionar los

DATOS:  $N$ , el orden de la matriz y sus elementos.

RESULTADOS: La matriz  $L$ .

PASO 1. Hacer  $L(1,1) = A(1,1) ** 0.5$ .

PASO 2. Hacer  $I = 2$ .

PASO 3. Mientras  $I \leq N$ , repetir los pasos 4 y 5.

PASO 4. Hacer  $L(I,1) = A(I,1)/L(1,1)$  y  $L(I,I) = 0$ .

PASO 5. Hacer  $I = I + 1$ .

PASO 6. Hacer  $I = 2$ .

PASO 7. Mientras  $I \leq N$ , repetir los pasos 8 a 24.

PASO 8. Hacer  $S = 0$ .

PASO 9. Hacer  $K = 1$ .

PASO 10. Mientras  $K \leq I-1$ , repetir los pasos 11 y 12.

PASO 11. Hacer  $S = S + L(I,K) ** 2$ .

- PASO 12. Hacer  $K = K + 1$ .  
 PASO 13. Hacer  $L(I,I) = (A(I,I) - S) ** 0.5$ .  
 PASO 14. Si  $I = N$  ir al paso 25.  
 PASO 15. Hacer  $J = I + 1$ .  
 PASO 16. Mientras  $J \leq N$ , repetir los pasos 17 a 23.  
     PASO 17. Hacer  $S = 0$ .  
     PASO 18. Hacer  $K = 1$ .  
     PASO 19. Mientras  $K \leq I-1$ , repetir los pasos 20 y 21.  
         PASO 20. Hacer  $S = S + L(I,K)*L(J,K)$ .  
         PASO 21. Hacer  $K = K + 1$ .  
     PASO 22. Hacer  $L(J,I)=(A(I,J)-S)/L(I,I)$  y  $L(I,J) = 0$ .  
     PASO 23. Hacer  $J = J + 1$ .  
 PASO 24. Hacer  $I = I + 1$ .  
 PASO 25. IMPRIMIR  $L$  y TERMINAR.

### SISTEMAS DE ECUACIONES MAL CONDICIONADOS

Algunos autores caracterizan los métodos de solución directos como aquellos con los que se obtiene la solución exacta  $\mathbf{x}$  del sistema  $A \mathbf{x} = \mathbf{b}$  mediante un número finito de operaciones, siempre y cuando no existan errores de redondeo. Como estos errores son prácticamente inevitables, se obtendrán en general soluciones aproximadas  $\mathbf{y}$ , cuya sustitución en el sistema producirá una aproximación del vector  $\mathbf{b}$ :  $\mathbf{b}'$

$$A \mathbf{y} = \mathbf{b}' \approx \mathbf{b}$$

En general, pequeños errores de redondeo producen sólo pequeños cambios en el vector solución; en estos casos se dice que el sistema está **bien condicionado**. Sin embargo, en algunos otros los errores de redondeo de los primeros pasos causan errores más adelante (se propagan), de modo que la solución obtenida  $\mathbf{y}$  resulta ser un vector distinto del vector solución; peor aún, en estos sistemas la sustitución de  $\mathbf{y}$  satisface prácticamente dicho sistema. Este tipo de sistemas se conocen como **mal condicionados**. A continuación se presentan dos ejemplos

Sea el sistema mal condicionado\*

$$\begin{bmatrix} 1.00 & 0.99 \\ 0.99 & 0.98 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1.99 \\ 1.97 \end{bmatrix} \quad (3.79)$$

cuya solución es  $x_1 = x_2 = 1.00$ , y sea la matriz aumentada siguiente

$$\left[ \begin{array}{cc|c} 1.00 & 0.9900 & 1.9900 \\ 0.00 & 0.0001 & 0.0001 \end{array} \right],$$

el resultado de la triangularización. Si se redondea o corta a tres dígitos la última fila, quedaría como fila de ceros y el sistema original como un sistema sin solución única.

Si, por otro lado, por un pequeño error en los cálculos se obtiene como solución de la ecuación 3.79

$$y_1 = 0, y_2 = 2,$$

\* Forsythe, G.E. y Moler, C.B. *Computer Solution of Linear Algebraic Systems*. Englewood Cliffs, N.J. Prentice Hall (1967).



que aunque distinta del vector solución da en la situación

$$\begin{bmatrix} 1.00 & 0.99 \\ 0.99 & 0.98 \end{bmatrix} \begin{bmatrix} 0 \\ 2 \end{bmatrix} = \begin{bmatrix} 1.98 \\ 1.96 \end{bmatrix}$$

prácticamente el vector **b**.

Aun una solución tan absurda como

$$y_1 = 100, y_2 = -99,$$

da resultados sorprendentemente cercanos a **b**

$$\begin{bmatrix} 1.00 & 0.99 \\ 0.99 & 0.98 \end{bmatrix} \begin{bmatrix} 100 \\ -99 \end{bmatrix} = \begin{bmatrix} 1.99 \\ 1.98 \end{bmatrix}$$

Algunas veces los elementos de **A** y **b** son generados por cálculos (véase algoritmos 5.1 y 5.5) y los valores resultantes de ambos son ligeramente erróneos.

Sea el sistema mal condicionado

$$\begin{aligned} 1.001 x_1 - x_2 &= 1 \\ x_1 - x_2 &= 0 \end{aligned} \quad (3.80)$$

que se desea resolver, pero por errores de redondeo o de otro tipo, se obtiene en su lugar

$$\begin{aligned} y_1 - 0.9999 y_2 &= 1.001 \\ y_1 - 1.0001 y_2 &= 0, \end{aligned} \quad (3.80')$$

que difiere sólo "ligeramente" del sistema 3.80

Las soluciones exactas son, respectivamente,

$$\mathbf{x} = \begin{bmatrix} 1000 \\ 1000 \end{bmatrix}; \quad \mathbf{y} = \begin{bmatrix} 5005.5005 \\ 5005.0000 \end{bmatrix}$$

cuya diferencia es notable a pesar de que los sistemas son casi idénticos. Para entender esto se da a continuación una interpretación geométrica de los sistemas mal condicionados.

## INTERPRETACIÓN GEOMÉTRICA DE UN SISTEMA MAL CONDICIONADO DE ORDEN 2

La solución de un sistema de dos ecuaciones en dos incógnitas

$$\begin{aligned} a_{1,1} x_1 + a_{1,2} x_2 &= b_1 \\ a_{2,1} x_1 + a_{2,2} x_2 &= b_2 \end{aligned} \quad (3.81)$$

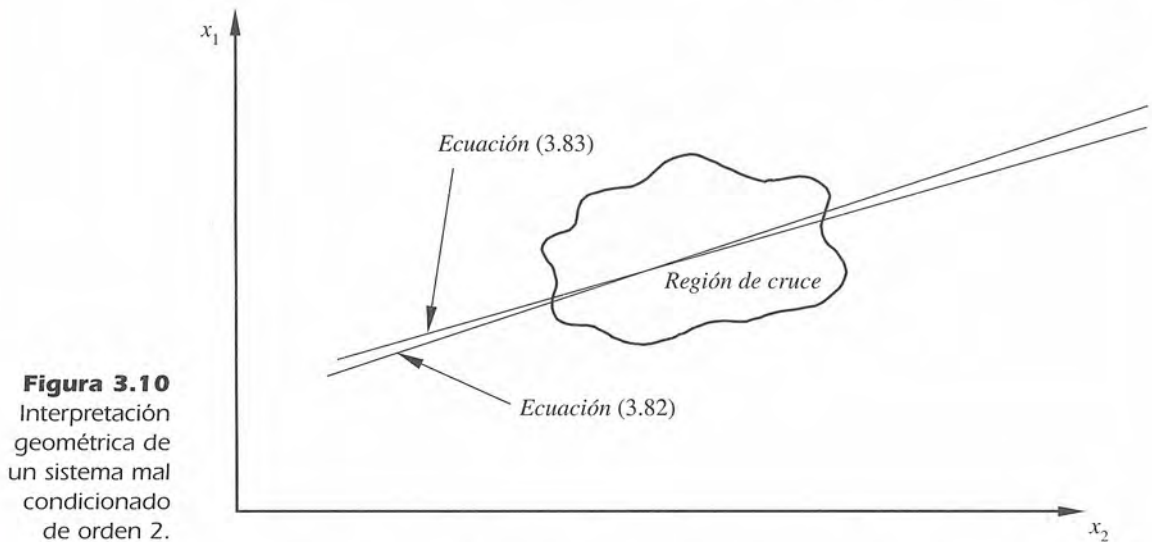
es el punto de intersección de las rectas

$$x_1 = \frac{b_1}{a_{1,1}} - \frac{a_{1,2}}{a_{1,1}} x_2 \quad (3.82)$$

$$x_1 = \frac{b_2}{a_{2,1}} - \frac{a_{2,2}}{a_{2,1}} x_2 \quad (3.83)$$

en el plano  $x_2 - x_1$ . Si el sistema 3.81 es mal condicionado, las rectas 3.82 y 3.83 son casi paralelas, pero resulta difícil decir dónde se cortan exactamente\* (véase Fig. 3.10).

\* Nótese que hay una solución única, pero resulta difícil decir dónde está.



**Figura 3.10**  
Interpretación  
geométrica de  
un sistema mal  
condicionado  
de orden 2.

Cualquier pequeño error de redondeo o de otro tipo puede alejar del vector solución, con lo que se produce una solución errónea  $\mathbf{y}$ . No obstante esto, si  $\mathbf{y}$  está en la región de cruce, el sistema 3.81 se satisface prácticamente con  $\mathbf{y}$ . Hay que observar que la región de cruce es muy amplia y que algunos de sus puntos pueden estar muy alejados del vector solución.

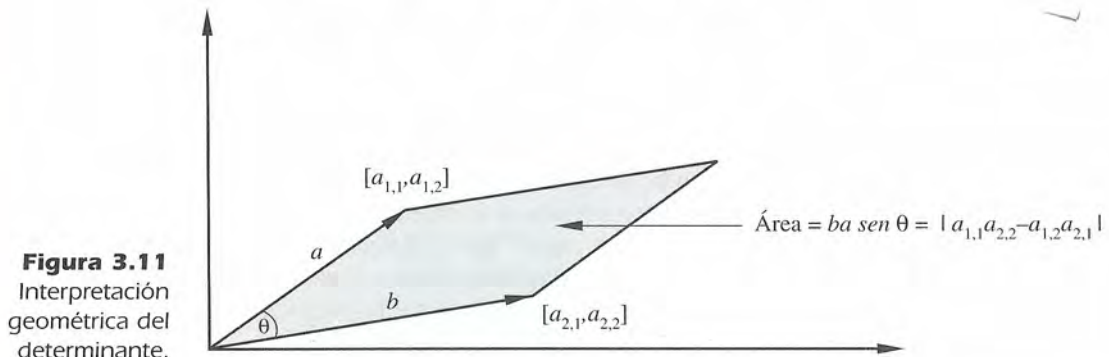
Una vez que se ha visto el comportamiento de los sistemas mal condicionados, resulta de interés determinar si un sistema dado está mal condicionado y qué hacer en tales casos para resolverlo. Hay varias formas de detectar si un sistema está mal o bien condicionado; pero quizá la más simple de ellas es la del determinante normalizado que se describe a continuación.

### MEDIDA DE CONDICIONAMIENTO USANDO EL DETERMINANTE NORMALIZADO

En el sistema 3.81 el determinante de la matriz coeficiente

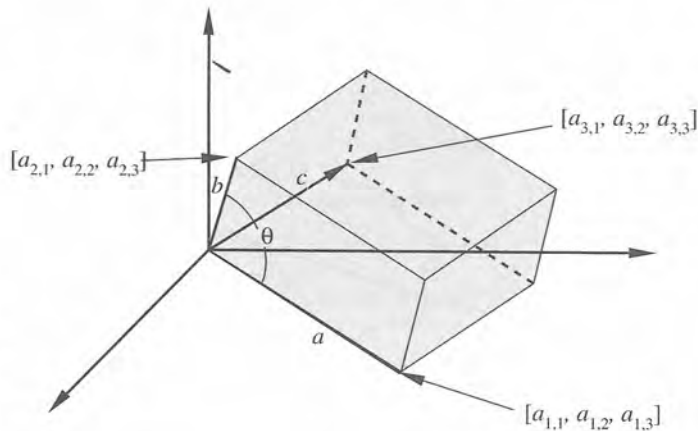
$$\begin{vmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{vmatrix} = a_{1,1}a_{2,2} - a_{1,2}a_{2,1},$$

puede interpretarse en valor absoluto como el área del paralelogramo cuyos lados son los vectores fila\*  $[a_{1,1} \ a_{1,2}]$  y  $[a_{2,1} \ a_{2,2}]$  (véase Fig. 3.11).



**Figura 3.11**  
Interpretación  
geométrica del  
determinante.

\* Puede decirse lo mismo para los vectores columna.



**Figura 3.12**  
Interpretación  
geométrica del  
determinante.

En el caso de un sistema general de orden 3, el determinante de la matriz coeficiente de dicho sistema es, en valor absoluto, el volumen del paralelepípedo cuyos lados son los vectores  $[a_{1,1} \ a_{1,2} \ a_{1,3}]$ ,  $[a_{2,1} \ a_{2,2} \ a_{2,3}]$  y  $[a_{3,1} \ a_{3,2} \ a_{3,3}]$  (véase Fig. 3.12).

Al multiplicar cada una de las filas del sistema 3.81 por un factor, el sistema resultante es equivalente, pero la matriz coeficiente se ha modificado y, por ende, su determinante. Si, por ejemplo, se divide la primera y segunda ecuaciones de 3.81, respectivamente, entre:

$$k_1 = \sqrt{a_{1,1}^2 + a_{1,2}^2} \quad k_2 = \sqrt{a_{2,1}^2 + a_{2,2}^2}$$

se obtiene como nueva matriz coeficiente

$$\begin{bmatrix} \frac{a_{1,1}}{k_1} & \frac{a_{1,2}}{k_1} \\ \frac{a_{2,1}}{k_2} & \frac{a_{2,2}}{k_2} \end{bmatrix},$$

cuyo determinante en valor absoluto es menor o igual a la unidad, ya que ahora  $|a| = 1$  y  $|b| = 1$  (véase Fig. 3.11). El determinante así obtenido se conoce como **determinante normalizado** y, en general, para sistemas de orden  $n$  la matriz coeficiente resultante de dividir la  $i$ -ésima fila por los factores\*

$$k_i = \sqrt{a_{i,1}^2 + a_{i,2}^2 + \dots + a_{i,n}^2} \quad i = 1, 2, \dots, n$$

tiene un determinante, en valor absoluto, menor o igual que la unidad.

Si el sistema 3.81 está mal condicionado, los vectores fila  $[a_{1,1} \ a_{1,2}]$  y  $[a_{2,1} \ a_{2,2}]$  son casi paralelos y el determinante normalizado estará muy cercano a cero (muy pequeño). Si, por otro lado, los vectores fila son casi ortogonales (perpendiculares), el determinante estará muy cercano a la unidad, en valor absoluto.

Resumiendo y precisando: para medir el condicionamiento de un sistema de orden  $n$ , se debe obtener el determinante normalizado de la matriz coeficiente de dicho sistema; y si su valor absoluto es “prominentemente menor” que 1, el sistema está mal condicionado en caso de tener un valor absoluto prominentemente cercano a 1, el sistema está bien condicionado. Esta lejanía o cercanía de 1 queda determinada por la precisión empleada.\*\*

\* Llamados factores de escalamiento.

\*\* Young, D. M. y Gregory, R. T. *A Survey Of Numerical Mathematics*, Vol. II Addison-Wesley (1973), pp. 812-820.



Si bien la técnica es útil, no resulta práctica en sistemas grandes, ya que el cálculo del determinante toma tiempo y es casi equivalente a resolver dichos sistemas. Entonces, si se sospecha que un sistema está mal condicionado, se analiza de la manera siguiente:

- a) Se resuelve el sistema original  $A \mathbf{x} = \mathbf{b}$ .
- b) Se modifican los componentes de  $A$  ligeramente y se resuelve el sistema resultante  $A' \mathbf{x} = \mathbf{b}$ .
- c) Si las dos soluciones son sustancialmente diferentes (estas diferencias se comparan con los cambios hechos en  $a_{i,j}$ ), el sistema está mal condicionado.

Una vez corroborado que un sistema grande está mal condicionado, deberán emplearse los métodos de solución vistos con ciertas recomendaciones.

- a) Aprovechar las características de la matriz coeficiente (matrices bandeadas, simétricas, diagonal dominantes, positivas definidas, etc.), para que el método seleccionado sea el más adecuado y se realicen, por ejemplo, menos cálculos.
- b) Emplear pivoteo parcial o total (véase Ej. 3.9).
- c) Emplear doble precisión en los cálculos.

Si aún después de seguir estas sugerencias persisten las dificultades, puede recurrirse a los métodos iterativos que se estudian más adelante y que son, en general, otra alternativa de solución de sistemas lineales mal y bien condicionados, con la ventaja de no ser tan sensibles a los errores de redondeo.

### MATRICES ELEMENTALES Y LOS MÉTODOS DE ELIMINACIÓN

Nótese que cualquiera de los métodos de eliminación vistos para resolver el sistema  $A \mathbf{x} = \mathbf{b}$  involucra las siguientes operaciones sobre una matriz:\*

- a) Intercambio de filas.
- b) Multiplicación de la fila por un escalar, y
- c) Sustitución de una fila por la suma de ésta y alguna otra fila de la matriz.

Estas operaciones pueden llevarse a cabo, mediante multiplicaciones de la matriz en cuestión, por ciertas matrices especiales; por ejemplo, la matriz permutadora permite intercambiar filas. Multiplicando en cambio por la izquierda una matriz  $B$  cualquiera por la matriz identidad correspondiente  $I$ , pero sustituido uno de sus elementos unitarios por  $m$  (la posición  $(i,i)$ , por ejemplo), se multiplica la  $i$ -ésima fila de  $B$  por  $m$ .

#### Ejemplo 3.36

Multipique la matriz general  $B$  de  $3 \times 4$  por la matriz identidad correspondiente  $I$ , donde se ha reemplazado el 1 de la posición  $(2, 2)$  con  $m$ .

#### Solución

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & m & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} \\ b_{2,1} & b_{2,2} & b_{2,3} & b_{2,4} \\ b_{3,1} & b_{3,2} & b_{3,3} & b_{3,4} \end{bmatrix} = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} & b_{1,4} \\ mb_{2,1} & mb_{2,2} & mb_{2,3} & mb_{2,4} \\ b_{3,1} & b_{3,2} & b_{3,3} & b_{3,4} \end{bmatrix}$$

Los resultados hablan por sí solos.

Finalmente, cuando se multiplica por la izquierda una matriz general  $B$  por la matriz identidad correspondiente  $I$ , en la que se ha sustituido uno de los ceros con  $m$  (el cero de

\* Generalmente se trata de la matriz aumentada  $[A \mid \mathbf{b}]$ .

la posición  $(i, j)$ , por ejemplo), se tiene el efecto de sustituir la fila  $i$ -ésima de  $B$  por la fila resultante de sumar ésta y la fila  $j$ -ésima de  $B$  multiplicada por  $m$ .

**Ejemplo 3.37**

Sustituya la segunda fila de la matriz general  $B$  de  $3 \times 3$  por el resultado de sumar dicha segunda fila con la primera fila de  $B$  multiplicada por  $m$ .

**Solución**

Se sustituye el cero de la posición  $(2, 1)$  de la matriz  $I$  de  $3 \times 3$  con  $m$  y se multiplica por la izquierda por  $B$ ; es decir:

$$\begin{bmatrix} 1 & 0 & 0 \\ m & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ b_{2,1} & b_{2,2} & b_{2,3} \\ b_{3,1} & b_{3,2} & b_{3,3} \end{bmatrix} = \begin{bmatrix} b_{1,1} & b_{1,2} & b_{1,3} \\ mb_{1,1} + b_{2,1} & mb_{1,2} + b_{2,2} & mb_{1,3} + b_{2,3} \\ b_{3,1} & b_{3,2} & b_{3,3} \end{bmatrix}$$

Si se desea intercambiar columnas, multiplicarlas por un escalar o sustituir una columna por la suma de ésta y alguna otra, se procede siguiendo las mismas ideas, pero con las multiplicaciones por la derecha sobre la matriz en cuestión.

Estas matrices se conocen como elementales y se denotan como:

Permutación:  $P$

Multiplicación por un escalar:  $M$

Sustitución:  $S$

Para aclarar la relación que existe entre estas matrices y los métodos de eliminación, se resuelve nuevamente el ejemplo 3.30, pero ahora con matrices elementales.

**Ejemplo 3.38**

Resuelva por eliminación de Jordan el sistema

$$\begin{aligned} 4x_1 - 9x_2 + 2x_3 &= 5 \\ 2x_1 - 4x_2 + 6x_3 &= 3 \\ x_1 - x_2 + 3x_3 &= 4 \end{aligned}$$

con matrices  $P$ ,  $M$  y  $S$ .

**Solución**

La matriz aumentada es



$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 2 & -4 & 6 & 3 \\ 1 & -1 & 3 & 4 \end{array} \right] = B$$

No se intercambian filas, ya que el elemento de máximo valor absoluto se encuentra en la primera. Para hacer cero el elemento  $(2, 1)$ , se suma la primera fila multiplicada por  $-1/2$  a la segunda; la siguiente matriz cumple con ese fin.

$$\left[ \begin{array}{ccc|c} 1 & 0 & 0 & 5 \\ -1/2 & 1 & 0 & 3 \\ 0 & 0 & 1 & 4 \end{array} \right] = S_1$$

Para hacer cero el elemento  $(3,1)$  se suma la primera multiplicada por  $-1/4$  a la tercera fila; esto es,

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -1/4 & 0 & 1 \end{bmatrix} = S_2$$

El efecto de  $S_1$  y  $S_2$  sobre  $B$  resulta en:

$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 0 & 0.5 & 5 & 0.5 \\ 0 & 1.25 & 2.5 & 2.75 \end{array} \right] = S_2 S_1 B$$

Como el elemento de máximo valor absoluto es 1.25, se intercambia la segunda y tercera filas, para lo cual se emplea la matriz

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} = P_1$$

y queda:

$$\left[ \begin{array}{ccc|c} 4 & -9 & 2 & 5 \\ 0 & 1.25 & 2.5 & 2.75 \\ 0 & 0.5 & 5 & 0.5 \end{array} \right] = P_1 S_2 S_1 B$$

Para hacer cero los elementos (1, 2) y (3, 2) se suma la segunda multiplicada por  $(-(-9)/1.25)$  a la primera fila, y la segunda multiplicada por  $(-0.5/1.25)$  a la tercera, proceso que se lleva a cabo con las matrices.

$$\begin{bmatrix} 1 & -(-9)/1.25 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = S_3 \quad \text{y} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -0.5/1.25 & 1 \end{bmatrix} = S_4$$

y queda como resultado:

$$\left[ \begin{array}{ccc|c} 4 & 0 & 20 & 24.8 \\ 0 & 1.25 & 2.5 & 2.75 \\ 0 & 0 & 4 & -0.6 \end{array} \right] = S_4 S_3 P_1 S_2 S_1 B$$

Para eliminar los elementos (1, 3) y (2, 3), se suma la tercera multiplicada por  $(-20/4)$  a la primera fila y la tercera multiplicada por  $(-2.5/4)$  a la segunda, lo cual se logra con  $S_5$  y  $S_6$ , respectivamente. (Se deja al lector determinar la forma que tienen  $S_5$  y  $S_6$ .) El resultado es:

$$\left[ \begin{array}{ccc|c} 4 & 0 & 0 & 27.8 \\ 0 & 1.25 & 0 & 3.125 \\ 0 & 0 & 4 & -0.6 \end{array} \right] = S_6 S_5 S_4 S_3 P_1 S_2 S_1 B$$

Todavía se puede multiplicar la primera fila por  $m_1 = 1/4$ , la segunda por  $m_2 = 1/1.25$  y la tercera por  $m_3 = 1/4$ , lo cual se consigue con

$$\begin{bmatrix} 1/4 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = M_1, \text{ etcétera}$$



finalmente queda:

$$\left[ \begin{array}{ccc|c} 1 & 0 & 0 & 6.95 \\ 0 & 1 & 0 & 2.5 \\ 0 & 0 & 1 & -0.15 \end{array} \right] = M_3 M_2 M_1 S_6 S_5 S_4 S_3 P_1 S_2 S_1 B$$

que puesta nuevamente como un sistema de ecuaciones da:

$$x_1 = 6.95$$

$$x_2 = 2.5$$

$$x_3 = -0.15,$$

directamente la solución del sistema original  $A \mathbf{x} = \mathbf{b}$ .

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
B=[4 -9 2 5; 2 -4 6 3; 1 -1 3 4]
S1=[1 0 0; -0.5 1 0; 0 0 1]
S2=[1 0 0; 0 1 0; -0.25 0 1]
B=S2*S1*B
P1=[1 0 0; 0 0 1; 0 1 0]
B=P1*B
S3=[1 9/1.25 0; 0 1 0; 0 0 1]
S4=[1 0 0; 0 1 0; 0 -0.5/1.25 1]
B=S4*S3*B
S5=[1 0 -5; 0 1 0; 0 0 1]
S6=[1 0 0; 0 1 -2.5/4; 0 0 1]
B=S6*S5*B
M1=[1/4 0 0; 0 1 0; 0 0 1]
M2=[1 0 0; 0 1/1.25 0; 0 0 1]
M3=[1 0 0; 0 1 0; 0 0 1/4]
B=M3*M2*M1*B
```



```
e3_38()
Prgm
[4,-9,2,5;2,-4,6,3;1,-1,3,4]→b
[1,0,0;-0.5,1,0;0,0,1]→s1
[1,0,0;0,1,0;-0.25,0,1]→s2
s2*s1*b→b : Disp b : Pause
[1,0,0;0,0,1;0,1,0]→p1
p1*b→b : Disp b : Pause
[1,9/1.25,0;0,1,0;0,0,1]→s3
[1,0,0;0,1,0;0,-0.5/1.25,1]→s4
s4*s3*b→b : Disp b : Pause
[1,0,-5;0,1,0;0,0,1]→s5
[1,0,0;0,1,-2.5/4;0,0,1]→s6
s6*s5*b→b : Disp b : Pause
[1/4,0,0;0,1,0;0,0,1]→m1
[1,0,0;0,1/1.25,0;0,0,1]→m2
[1,0,0;0,1,0;0,0,1/4]→m3
m3*m2*m1*b→b : Disp b
EndPrgm
```

Si el producto de las matrices elementales se denota por  $E$

$$E = M_3 M_2 M_1 S_6 S_5 S_4 S_3 P_1 S_2 S_1,$$

se tiene:

$$E B = E [A \mid \mathbf{b}] = [I \mid \mathbf{x}],$$

de donde:

$$E A = I \quad \text{y} \quad E B = \mathbf{x}$$

resulta que  $E$  es la inversa de  $A$

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

Por otro lado, se sabe que el determinante del producto de dos o más matrices es igual al producto de los determinantes de cada una de las matrices.

$$\det A B \dots = \det A \det B \dots$$

de donde:

$$\det E A = \det I$$

o bien:

$$\det E \det A = 1$$

y

$$\frac{1}{\det E} = \det A$$

de modo que el determinante de  $A$  está dado como la inversa del determinante de  $E$  y sólo queda obtener  $\det E$ . Esto parece complicado a simple vista; sin embargo, observando que en general (véase Prob. 3.52)

$\det P = -1$ , el determinante de una matriz permutadora es  $-1$ .

$\det M = m$ , el determinante de una matriz multiplicadora es el factor  $m$ , que deberá ser distinto de cero.

$\det S = 1$ , el determinante de una matriz del tipo  $S$  es 1.

Se tiene:

$$\det E = \det M_3 \det M_2 \det M_1 \det S_6 \det S_5 \det S_4 \det S_3 \det P_1 \det S_2 \det S_1$$

sustituyendo

$$\det E = m_3 m_2 m_1 (-1) = -\frac{1}{4} \left( \frac{1}{1.25} \right) \frac{1}{4} = -0.05$$

y

$$\det A = \frac{1}{-0.05} = -20$$

Finalmente, para obtener  $E$  y por tanto  $A^{-1}$  se toma  $S_1$  como matriz pivote y sobre ella se efectúan las operaciones de intercambio de filas, multiplicación por un escalar, etc., que vayan indicando las matrices a su izquierda. Así:

$$\begin{bmatrix} 1 & 0 & 0 \\ -1/2 & 1 & 0 \\ -1/4 & 0 & 1 \end{bmatrix} = S_2 S_1$$

ya que según se dijo,  $S_2$  tiene como efecto multiplicar la primera fila de  $S_1$  por  $-1/4$  y sumarla a la tercera fila de  $S_1$ .

Con  $P_1$  en cambio se tiene:

$$\begin{bmatrix} 1 & 0 & 0 \\ -1/4 & 0 & 1 \\ -1/2 & 1 & 0 \end{bmatrix} = P_1 (S_2 S_1)$$

ya que  $P_1$  intercambia las filas segunda y tercera de  $(S_2 S_1)$ .

Continuando este proceso se llega a:

$$\begin{bmatrix} 0.3 & -1.25 & 2.3 \\ 0 & -0.5 & 1.0 \\ -0.1 & 0.25 & -0.1 \end{bmatrix} = E = A^{-1}$$

### 3.5 Métodos iterativos

Al resolver un sistema de ecuaciones lineales por eliminación, la memoria de máquina requerida es proporcional al cuadrado del orden de  $A$ , y el trabajo computacional es proporcional al cubo del orden de la matriz coeficiente  $A$  (véase Secc. 3.4). Debido a esto, la solución de sistemas lineales grandes ( $n \geq 50$ ), con matrices coeficiente densas,\* se vuelve costoso y difícil en una computadora con los métodos de eliminación, ya que se requiere amplia memoria; además, como el número de operaciones que se debe ejecutar es muy grande, se pueden producir errores de redondeo también muy grandes. Sin embargo, se han resuelto sistemas de orden 1000, y aun mayor, con los métodos que se estudiarán en esta sección.

Estos sistemas de un número muy grande de ecuaciones se presentan en la solución numérica de ecuaciones diferenciales parciales, en la solución de los modelos resultantes en la simulación de columnas de destilación, etc. En favor de estos sistemas, puede decirse que tienen matrices con pocos elementos distintos de cero y que éstas poseen ciertas propiedades (simétricas, bandeadas, diagonal dominantes, entre otras), que permiten garantizar el éxito en la aplicación de los métodos de esta sección.

#### MÉTODOS DE JACOBI Y GAUSS-SEIDEL

Los métodos iterativos más sencillos y conocidos son una generalización del método de punto fijo, estudiado en el capítulo 2. Se puede aplicar la misma técnica a fin de elaborar métodos para la solución de  $A \mathbf{x} = \mathbf{b}$ , de la siguiente manera.

Se parte de  $A \mathbf{x} = \mathbf{b}$  para obtener la ecuación

$$A \mathbf{x} - \mathbf{b} = \mathbf{0}, \quad (3.84)$$

ecuación vectorial correspondiente a  $f(\mathbf{x}) = 0$ . Se busca ahora una matriz  $B$  y un vector  $\mathbf{c}$ , de manera que la ecuación vectorial

$$\mathbf{x} = B \mathbf{x} + \mathbf{c}, \quad (3.85)$$

sea sólo un arreglo de la ecuación 3.84; es decir, de manera que la solución de una sea también la solución de la otra. La ecuación 3.85 correspondería a  $\mathbf{x} = g(\mathbf{x})$ . A continuación se

\* Una matriz densa tiene pocos ceros como elementos.



propone un vector inicial  $\mathbf{x}^{(0)}$  como primera aproximación al vector solución  $\mathbf{x}$ . Luego, se calcula con la ecuación 3.85 la sucesión vectorial  $\mathbf{x}^{(1)} \mathbf{x}^{(2)}, \dots$ , de la siguiente manera

$$\mathbf{x}^{(k+1)} = \mathbf{B} \mathbf{x}^{(k)} + \mathbf{c}, \quad k = 0, 1, 2, \dots$$

donde:

$$\mathbf{x}^{(k)} = [x_1^k \ x_2^k \ \dots \ x_n^k]^T \quad (3.86)$$

Para que la sucesión  $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(n)}, \dots$ , converja al vector solución  $\mathbf{x}$  es necesario que eventualmente  $x_j^m$ ,  $1 \leq j \leq n$  (los componentes del vector  $\mathbf{x}^{(m)}$ ), se aproximen tanto a  $x_j$ ,  $1 \leq j \leq n$  (los componentes correspondientes a  $\mathbf{x}$ ), que todas las diferencias  $|x_j^m - x_j|$ ,  $1 \leq j \leq n$  sean menores que un valor pequeño previamente fijado, y que se conserven menores para todos los vectores siguientes de la iteración; es decir,

$$\lim_{m \rightarrow \infty} x_j^m = x_j \quad 1 \leq j \leq n \quad (3.87)$$

La forma como se llega a la ecuación 3.85 define el algoritmo y su convergencia. Dado el sistema  $\mathbf{A} \mathbf{x} = \mathbf{b}$ , la manera más sencilla es despejar  $x_1$  de la primera ecuación,  $x_2$  de la segunda, etc. Para ello, es necesario que todos los elementos de la diagonal principal de  $\mathbf{A}$ , por razones obvias, sean distintos de cero. Para ver esto en detalle considérese el sistema general de tres ecuaciones (naturalmente puede extenderse a cualquier número de ecuaciones).

Sea entonces:

$$\begin{aligned} a_{1,1} x_1 + a_{1,2} x_2 + a_{1,3} x_3 &= b_1 \\ a_{2,1} x_1 + a_{2,2} x_2 + a_{2,3} x_3 &= b_2 \\ a_{3,1} x_1 + a_{3,2} x_2 + a_{3,3} x_3 &= b_3 \end{aligned}$$

con  $a_{11}$ ,  $a_{22}$  y  $a_{33}$  distintos de cero.

Se despeja  $x_1$  de la primera ecuación,  $x_2$  de la segunda, y  $x_3$  de la tercera, con lo que se obtiene:

$$\begin{aligned} x_1 &= -\frac{a_{1,2}}{a_{1,1}} x_2 - \frac{a_{1,3}}{a_{1,1}} x_3 + \frac{b_1}{a_{1,1}} \\ x_2 &= -\frac{a_{2,1}}{a_{2,2}} x_1 - \frac{a_{2,3}}{a_{2,2}} x_3 + \frac{b_2}{a_{2,2}} \\ x_3 &= -\frac{a_{3,1}}{a_{3,3}} x_1 - \frac{a_{3,2}}{a_{3,3}} x_2 + \frac{b_3}{a_{3,3}} \end{aligned} \quad (3.88)$$

que en notación matricial queda:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 & -\frac{a_{1,2}}{a_{1,1}} & -\frac{a_{1,3}}{a_{1,1}} \\ -\frac{a_{2,1}}{a_{2,2}} & 0 & -\frac{a_{2,3}}{a_{2,2}} \\ -\frac{a_{3,1}}{a_{3,3}} & -\frac{a_{3,2}}{a_{3,3}} & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} + \begin{bmatrix} \frac{b_1}{a_{1,1}} \\ \frac{b_2}{a_{2,2}} \\ \frac{b_3}{a_{3,3}} \end{bmatrix} \quad (3.89)$$

y ésta es la ecuación 3.86 desarrollada, con

$$B = \begin{bmatrix} 0 & -\frac{a_{1,2}}{a_{1,1}} & -\frac{a_{1,3}}{a_{1,1}} \\ -\frac{a_{2,1}}{a_{2,2}} & 0 & -\frac{a_{2,3}}{a_{2,2}} \\ -\frac{a_{3,1}}{a_{3,3}} & -\frac{a_{3,2}}{a_{3,3}} & 0 \end{bmatrix} \quad \text{y } \mathbf{c} = \begin{bmatrix} \frac{b_1}{a_{1,1}} \\ \frac{b_2}{a_{2,2}} \\ \frac{b_3}{a_{3,3}} \end{bmatrix}$$

Una vez que se tiene la forma 3.89, se propone un vector inicial  $\mathbf{x}^{(0)}$  que puede ser  $\mathbf{x}^{(0)} = \mathbf{0}$ , o algún otro que sea aproximado al vector solución  $\mathbf{x}$ .

Para iterar existen dos variantes

### 1. Iteración de Jacobi (método de desplazamientos simultáneos)

Si

$$\mathbf{x}^{(k)} = \begin{bmatrix} x_1^k \\ x_2^k \\ x_3^k \end{bmatrix} \quad (3.90)$$

es el vector aproximación a la solución  $\mathbf{x}$  después de  $k$  iteraciones, entonces se tiene para la siguiente aproximación

$$\mathbf{x}^{(k+1)} = \begin{bmatrix} x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \end{bmatrix} = \begin{bmatrix} \frac{1}{a_{1,1}} (b_1 - a_{1,2} x_2^k - a_{1,3} x_3^k) \\ \frac{1}{a_{2,2}} (b_2 - a_{2,1} x_1^k - a_{2,3} x_3^k) \\ \frac{1}{a_{3,3}} (b_3 - a_{3,1} x_1^k - a_{3,2} x_2^k) \end{bmatrix} \quad (3.91)$$

O bien, para un sistema de  $n$  ecuaciones con  $n$  incógnitas y usando notación más compacta y de mayor utilidad en programación, se tiene:

$$x_i^{k+1} = -\frac{1}{a_{i,i}} \left[ -b_i + \sum_{\substack{j=1 \\ j \neq i}}^n a_{i,j} x_j^k \right], \text{ para } 1 \leq i \leq n \quad (3.92)$$

### 2. Iteración de Gauss-Seidel (método de desplazamientos sucesivos)

En este método los valores que se van calculando en la  $(k+1)$ -ésima iteración se emplean para calcular los valores faltantes de esa misma iteración; es decir, con  $\mathbf{x}^{(k)}$  se calcula  $\mathbf{x}^{(k+1)}$  de acuerdo con:

$$\mathbf{x}^{(k+1)} = \begin{bmatrix} x_1^{k+1} \\ x_2^{k+1} \\ x_3^{k+1} \end{bmatrix} = \begin{bmatrix} \frac{1}{a_{1,1}} (b_1 - a_{1,2} x_2^k - a_{1,3} x_3^k) \\ \frac{1}{a_{2,2}} (b_2 - a_{2,1} x_1^{k+1} - a_{2,3} x_3^k) \\ \frac{1}{a_{3,3}} (b_3 - a_{3,1} x_1^{k+1} - a_{3,2} x_2^{k+1}) \end{bmatrix} \quad (3.93)$$

O bien, para un sistema de  $n$  ecuaciones

$$x_i^{k+1} = -\frac{1}{a_{i,i}} \left[ -b_i + \sum_{j=1}^{i-1} a_{i,j} x_j^{k+1} + \sum_{j=i+1}^n a_{i,j} x_j^k \right], \text{ para } 1 \leq i \leq n \quad (3.94)$$

**SUGERENCIA:** El empleo de un pizarrón electrónico para los siguientes ejemplos o el de una calculadora programable, atenuaría considerablemente el trabajo de los cálculos.

### Ejemplo 3.39

Resuelva el siguiente sistema por los métodos de Jacobi y Gauss-Seidel

$$\begin{aligned} 4x_1 - x_2 &= 1 \\ -x_1 + 4x_2 - x_3 &= 1 \\ -x_2 + 4x_3 - x_4 &= 1 \\ -x_3 + 4x_4 &= 1 \end{aligned} \quad (3.95)$$

**Solución** Despejando  $x_1$  de la primera ecuación,  $x_2$  de la segunda, etc., se obtiene

$$\begin{aligned} x_1 &= x_2/4 + 1/4 \\ x_2 &= x_1/4 + x_3/4 + 1/4 \\ x_3 &= x_2/4 + x_4/4 + 1/4 \\ x_4 &= x_3/4 + 1/4 \end{aligned} \quad (3.96)$$

#### Vector inicial

Cuando no se tiene una aproximación al vector solución, se emplea generalmente como vector inicial el vector cero, esto es:

$$\mathbf{x}^{(0)} = [0 \ 0 \ 0 \ 0]^T$$

#### a) Método de Jacobi

El cálculo de  $\mathbf{x}^{(1)}$  en el método de Jacobi se obtiene reemplazando  $\mathbf{x}^{(0)}$  en cada una de las ecuaciones de 3.96

$$\begin{aligned} x_1 &= 0/4 + 1/4 = 1/4 \\ x_2 &= 0/4 + 0/4 + 1/4 = 1/4 \\ x_3 &= 0/4 + 0/4 + 1/4 = 1/4 \\ x_4 &= 0/4 + 1/4 = 1/4 \end{aligned}$$

y entonces  $\mathbf{x}^{(1)} = [1/4 \ 1/4 \ 1/4 \ 1/4]^T$

Para calcular  $\mathbf{x}^{(2)}$  se sustituye  $\mathbf{x}^{(1)}$  en cada una de las ecuaciones de 3.96. Para simplificar la notación se han omitido los superíndices.

$$\begin{aligned} x_1 &= 1/16 + 1/4 = 0.3125 \\ x_2 &= 1/16 + 1/16 + 1/4 = 0.3750 \\ x_3 &= 1/16 + 1/16 + 1/4 = 0.3750 \\ x_4 &= 1/16 + 1/4 = 0.3125 \end{aligned}$$



A continuación se presentan los resultados de subsecuentes iteraciones, en forma tabular

**Tabla 3.1** Solución del sistema 3.95 por el método de Jacobi.

| $k$ | $x_1^k$ | $x_2^k$ | $x_3^k$ | $x_4^k$ |
|-----|---------|---------|---------|---------|
| 0   | 0.0000  | 0.0000  | 0.0000  | 0.0000  |
| 1   | 0.2500  | 0.2500  | 0.2500  | 0.2500  |
| 2   | 0.3125  | 0.3750  | 0.3750  | 0.3125  |
| 3   | 0.3438  | 0.4219  | 0.4219  | 0.3438  |
| 4   | 0.3555  | 0.4414  | 0.4414  | 0.3555  |
| 5   | 0.3604  | 0.4492  | 0.4492  | 0.3604  |
| 6   | 0.3623  | 0.4524  | 0.4524  | 0.3623  |
| 7   | 0.3631  | 0.4537  | 0.4537  | 0.3631  |
| 8   | 0.3634  | 0.4542  | 0.4542  | 0.3634  |
| 9   | 0.3635  | 0.4544  | 0.4544  | 0.3635  |
| 10  | 0.3636  | 0.4545  | 0.4545  | 0.3636  |

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
% Método de Jacobi
clear
A=[4 -1 0 0; -1 4 -1 0; 0 -1 4 -1; 0 0 -1 4]
b=[1 1 1 1]
X0=zeros(1,4);
K=0;Norma=1;
fprintf('K X(1) X(2) X(3) X(4) Norma\n')
while Norma > 0.0001
    K=K+1;
    fprintf('%2d',K)
    for i=1:4
        suma=0;
        for j=1:4
            if i ~= j
                suma=suma+A(i,j)*X0(j);
            end
        end
        X(i)=(b(i)-suma)/A(i,i);
        fprintf('%10.4f',X(i))
    end
    Norma=norm(X0-X);
    fprintf('%10.4f\n',Norma)
    X0=X;
    if K > 25
        disp('No se alcanzó la convergencia')
        break
    end
end
```



```

e3_39a()
Prgrm
ClrIO : [4,-1,0,0;-1,4,-1,0;0,-1,4,-1;0,0,-1,4]→a
[1,1,1,1]→b : [0,0,0,0]→x0 : x0→x1 : 0→k : 1→norma
Disp"k x(1) x(2) x(3) x(4) norma"
While norma>1.E-4
  k+1→k : string(k)&"↵"→d
  For i,1,4
    0→suma
    For j,1,4
      If i≠j
        suma+a[i,j]*x0[l,j]→suma
      EndFor
      (b[l,i]-suma)/a[i,i]→x1[l,i]:d&format(x1[l,i],"f4")&"↵"→d
    EndFor
    norm(x1-x0)→norma : d&format(norma,"f5")→d : Disp d : x1→x0
  If k>25 Then
    Disp "No se alcanzó la convergencia"
    Exit
  EndIf
EndWhile
EndPrgrm

```

### b) Método de Gauss-Siedel

Para el cálculo del primer elemento del vector  $\mathbf{x}^{(1)}$ , se sustituye  $\mathbf{x}^{(0)}$  en la primera ecuación de 3.96, para simplificar la notación se han omitido los superíndices.

$$x_1 = 0/4 + 1/4 = 1/4$$

Para el cálculo de  $x_2$  de  $\mathbf{x}^{(1)}$ , se emplea el valor de  $x_1$  ya obtenido (1/4) y los valores  $x_2$ ,  $x_3$  y  $x_4$  de  $\mathbf{x}^{(0)}$ . Así,

$$x_2 = \frac{1}{4(4)} + 0/4 + 1/4 = 0.3125$$

Con los valores de  $x_1$  y  $x_2$  ya obtenidos, y con  $x_3$  y  $x_4$  de  $\mathbf{x}^{(0)}$  se evalúa  $x_3$  de  $\mathbf{x}^{(1)}$ .

$$x_3 = 0.3125/4 + 0/4 + 1/4 = 0.3281$$

Finalmente, con los valores de  $x_1$ ,  $x_2$  y  $x_3$  calculados previamente, y con  $x_4$  de  $\mathbf{x}^{(0)}$ , se obtiene la última componente de  $\mathbf{x}^{(1)}$

$$x_4 = 0.3281/4 + 1/4 = 0.3320$$

Entonces  $\mathbf{x}^{(1)} = [0.25 \ 0.3125 \ 0.3281 \ 0.3320]^T$

Para la segunda iteración (cálculo de  $\mathbf{x}^{(2)}$ ) se procede de igual manera.

$$x_1 = 0.3125/4 + 1/4 = 0.3281$$

$$x_2 = 0.3281/4 + 0.3281/4 + 1/4 = 0.4141$$

$$x_3 = 0.4141/4 + 0.3320/4 + 1/4 = 0.4365$$

$$x_4 = 0.4365/4 + 1/4 = 0.3591$$

Con lo que  $\mathbf{x}^{(2)} = [0.3281 \ 0.4141 \ 0.4365 \ 0.3591]^T$ .

En la tabla 3.2 se presentan los resultados de las iteraciones subsecuentes.

**Tabla 3.2** Solución del sistema 3.95 por el método de Gauss-Seidel.

| $k$ | $x_1^k$ | $x_2^k$ | $x_3^k$ | $x_4^k$ |
|-----|---------|---------|---------|---------|
| 0   | 0.0000  | 0.0000  | 0.0000  | 0.0000  |
| 1   | 0.2500  | 0.3125  | 0.3281  | 0.3320  |
| 2   | 0.3281  | 0.4141  | 0.4365  | 0.3591  |
| 3   | 0.3535  | 0.4475  | 0.4517  | 0.3629  |
| 4   | 0.3619  | 0.4534  | 0.4541  | 0.3635  |
| 5   | 0.3633  | 0.4544  | 0.4545  | 0.3636  |
| 6   | 0.3636  | 0.4545  | 0.4545  | 0.3636  |

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus:



```
% Método de Gauss-Seidel
clear; A=[4 -1 0 0; -1 4 -1 0; 0 -1 4 -1; 0 0 -1 4]; b=[1 1 1 1]
X0=zeros(1,4); X=X0; K=0; Norma=1;
fprintf(' K X(1) X(2) X(3) X(4) Norma\n')
while Norma > 0.0001
    K=K+1; fprintf(' %2d ',K)
    for i=1:4
        suma=0;
        for j=1:4
            if i ~= j
                suma=suma+A(i,j)*X(j);
            end
        end
        X(i)=(b(i)-suma)/A(i,i); fprintf('%10.4f',X(i))
    end
    Norma=norm(X0-X); fprintf('%10.4f\n',Norma)
    X0=X;
    if K > 17
        disp('No se alcanzó la convergencia')
        break
    end
end
```





```

e3_39b()
Prgm
ClrIO : [4,-1,0,0;-1,4,-1,0;0,-1,4,-1;0,0,-1,4]→a
[1,1,1,1]→b : [0,0,0,0]→x0 : x0→x1 : 0→k : 1→norma
Disp "k x(1) x(2) x(3) x(4) norma"
While norma>1.E-4
  k+1→k : string(k)&" "→d
  For i,1,4
    0→suma
    For j,1,4
      If i≠j
        suma+a[i,j]*x1[l,j]→suma
      EndFor
      (b[l,i]-suma)/a[i,i]→x1[l,i]:d&format(x1[l,i],"f4")&" "→d
    EndFor
    norm(x1-x0)→norma : d&format(norma,"f5")→d : Disp d : x1→x0
  If k>25 Then
    Disp "No se alcanzó la convergencia"
    Exit
  EndIf
EndWhile
EndPrgm

```

En la aplicación de estas dos variantes son válidas las preguntas siguientes:

1. ¿La sucesión de vectores  $\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \mathbf{x}^{(3)}, \dots$ , converge o se aleja del vector solución  $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]^T$ ?
2. ¿Cuándo detener el proceso iterativo?

Las respuestas correspondientes, conocidas como criterio de convergencia, se dan a continuación

1. Si la sucesión converge a  $\mathbf{x}$ , cabe esperar que los elementos de  $\mathbf{x}^{(k)}$  se vayan acercando a los elementos correspondientes de  $\mathbf{x}$ ; es decir,  $x_1^k$  a  $x_1$ ;  $x_2^k$  a  $x_2$ , etc., o que se alejen en caso contrario.
2. Cuando
  - a) Los valores absolutos  $|x_1^{k+1} - x_1^k|$ ,  $|x_2^{k+1} - x_2^k|$ , etc., sean todos menores de un número pequeño  $\varepsilon$  en cuyo valor será dado por el programador.
  - O bien
  - b) Si el número de iteraciones ha excedido un máximo predeterminado MAXIT.

Por otro lado, es natural pensar que si la sucesión  $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots$ , converge a  $\mathbf{x}$ , la distancia (véase Sec. 3.2) de  $\mathbf{x}^{(0)}$  a  $\mathbf{x}$ , de  $\mathbf{x}^{(1)}$  a  $\mathbf{x}$ , etc., se va reduciendo; también es cierto que la distancia entre cada dos vectores consecutivos  $\mathbf{x}^{(0)}$  y  $\mathbf{x}^{(1)}$ ,  $\mathbf{x}^{(1)}$  y  $\mathbf{x}^{(2)}$ , etc., se decrementa conforme el proceso iterativo avanza; esto es, la sucesión de números reales

$$\begin{aligned}
 &|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}| \\
 &|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}| \\
 &\vdots \\
 &|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}|
 \end{aligned} \tag{3.97}$$

convergirá a cero.

Si, por el contrario, esta sucesión de números diverge, entonces puede pensarse que el proceso diverge. Con esto, un criterio más es

$$c) \text{ Detener el proceso una vez que } |\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}| < \varepsilon$$

Al elaborar un programa de cómputo para resolver sistemas de ecuaciones lineales, generalmente se utilizan los criterios a), b) y c) o la combinación de a) y b), o la de b) y c).

Si se observan las columnas de las tablas 3.1 y 3.2, se advertirá que todas son sucesiones de números convergentes, por lo que ambos métodos convergen a un vector, presumiblemente la solución del sistema 3.95.

Si se tomara el criterio (a) con  $\varepsilon = 10^{-2}$  y el método de Jacobi,  $\varepsilon$  se satisface en la sexta iteración de la tabla 3.1; en cambio si  $\varepsilon = 10^{-3}$ , se necesitan 10 iteraciones.

Si se toma  $\varepsilon = 10^{-3}$ , el método de Gauss-Seidel y el criterio (a), se requerirían sólo seis iteraciones, como puede verse en la tabla 3.2.

Aunque hay ejemplos en los que Jacobi converge y Gauss-Seidel diverge, y viceversa, en general puede esperarse convergencia más rápida por Gauss-Seidel, o una manifestación más rápida de divergencia. Esto se debe al hecho de ir usando los valores más recientes de  $\mathbf{x}^{(k+1)}$  que permitirán acercarse o alejarse más rápidamente de la solución.

### REARREGLO DE ECUACIONES

Para motivar el rearreglo de ecuaciones, se propone resolver el siguiente sistema con el método de Gauss Seidel y con  $\varepsilon = 10^{-2}$  aplicado a  $|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}|$ .

$$\begin{aligned} -x_1 + 3x_2 + 5x_3 + 2x_4 &= 10 \\ x_1 + 9x_2 + 8x_3 + 4x_4 &= 15 \\ & x_2 + x_4 = 2 \\ 2x_1 + x_2 + x_3 - x_4 &= -3 \end{aligned} \quad (3.98)$$

Al resolver para  $x_1$  de la primera ecuación, para  $x_2$  de la segunda,  $x_3$  de la cuarta, y  $x_4$  de la tercera se obtiene:

$$\begin{aligned} x_1 &= 3x_2 + 5x_3 + 2x_4 - 10 \\ x_2 &= -x_1/9 - (8/9)x_3 - (4/9)x_4 + 15/9 \\ x_3 &= -2x_1 - x_2 + x_4 - 3 \\ x_4 &= -x_2 + 2 \end{aligned}$$

Con el vector cero como vector inicial, se tiene la siguiente sucesión de vectores. Nótese que el proceso diverge.

**Tabla 3.3** Aplicación del método de Gauss-Seidel al sistema 3.98.

| $k$ | $x_1^k$ | $x_2^k$ | $x_3^k$ | $x_4^k$ | $ \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} $ |
|-----|---------|---------|---------|---------|-------------------------------------------|
| 0   | 0.000   | 0.000   | 0.000   | 0.000   |                                           |
| 1   | -10.000 | 2.7778  | 14.222  | -0.7778 | 17.62                                     |
| 2   | 67.8889 | -18.172 | -121.2  | 20.17   | 159.0                                     |
| 3   | -631.1  | 170.7   | 1108.0  | -168.71 | 1439.05                                   |

Si el proceso iterativo diverge, como es el caso, un rearrreglo de las ecuaciones puede originar convergencia; por ejemplo, en lugar de despejar  $x_1$  de la primera ecuación,  $x_2$  de la segunda, etc., cabe despejar las diferentes  $x_i$  de diferentes ecuaciones, teniendo cuidado de que los coeficientes de las  $x_i$  despejadas sean distintos de cero.

Esta sugerencia presenta, para un sistema de  $n$  ecuaciones,  $n!$  distintas formas de rearrreglar dicho sistema. A fin de simplificar este procedimiento, se utilizará el siguiente teorema.

**TEOREMA 3.2** Los procesos de Jacobi y Gauss-Seidel convergirán si en la matriz coeficiente 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 (matriz diagonal dominante). Es decir, se asegura la convergencia si:

$$|a_{i,i}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{i,j}| \quad 1 \leq i \leq n$$

y

$$|a_{i,i}| > \sum_{\substack{j=1 \\ j \neq i}}^n |a_{j,i}| \quad 1 \leq i \leq n \quad (3.99)$$

Este teorema no será de mucha utilidad si se toma al pie de la letra, ya que contados sistemas de ecuaciones lineales poseen matrices coeficiente diagonalmente dominantes; sin embargo, si se arreglan las ecuaciones para tener el sistema lo más cercano posible a las condiciones del teorema, algún beneficio se puede obtener. Ésta es la pauta para reordenar las ecuaciones y obtener o mejorar la convergencia, en el mejor de los casos. A continuación se ilustra esto, rearrreglando el sistema 3.98, despejando  $x_1$  de la ecuación 4,  $x_2$  de la ecuación 2,  $x_3$  de la ecuación 1, y  $x_4$  de la ecuación 3, para llegar a:

$$\begin{aligned} x_1 &= -x_2/2 - x_3/2 + x_4/2 - 3/2 \\ x_2 &= -x_1/9 - 8x_3/9 - 4x_4/9 + 15/9 \\ x_3 &= x_1/5 - 3x_2/5 - 2x_4/5 + 10/5 \\ x_4 &= -x_2 + 2 \end{aligned}$$

Los resultados para las primeras 18 iteraciones con el vector cero como vector inicial se muestran en la tabla 3.4.

Antes de continuar las iteraciones, puede observarse en la tabla 3.4 que los valores de  $\mathbf{x}^{(18)}$  parecen converger al vector

$$\mathbf{x} = [-1 \ 0 \ 1 \ 2]^T$$

Con la sustitución de estos valores en el sistema 3.98, se comprueba que  $x_1 = 1$ ,  $x_2 = 0.0$ ,  $x_3 = 1$  y  $x_4 = 2$  es el vector solución, y por razones obvias se detiene el proceso.

Finalmente, las ecuaciones 3.99 son equivalentes (en sistemas de ecuaciones) a la expresión 2.10 del capítulo 2 que establece el criterio de convergencia del método iterativo para resolver  $f(x) = 0$ .



**Tabla 3.4** Aplicación del método de Gauss-Seidel al sistema 3.98, rearreglando las ecuaciones para obtener una aproximación a un sistema diagonal dominante.

| $k$ | $x_1^k$ | $x_2^k$ | $x_3^k$ | $x_4^k$ | $ x^{(k+1)} - x^{(k)} $ |
|-----|---------|---------|---------|---------|-------------------------|
| 0   | 0.0000  | 0.0000  | 0.0000  | 0.0000  |                         |
| 1   | -1.5000 | 1.8333  | 0.6000  | 0.1667  | 2.44                    |
| 2   | -2.6333 | 1.3519  | 0.5956  | 0.6481  | 1.32                    |
| 3   | -2.1496 | 1.0881  | 0.6580  | 0.9119  | 0.6140                  |
| 4   | -1.9171 | 0.8895  | 0.7181  | 1.1105  | 0.3695                  |
| 5   | -1.7486 | 0.7291  | 0.7686  | 1.2704  | 0.2867                  |
| 6   | -1.6134 | 0.5978  | 0.8102  | 1.4022  | 0.2337                  |
| 7   | -1.5030 | 0.4903  | 0.8444  | 1.5097  | 0.1907                  |
| 8   | -1.4125 | 0.4020  | 0.8724  | 1.5980  | 0.1567                  |
| 9   | -1.3382 | 0.3297  | 0.8953  | 1.6703  | 0.1285                  |
| 10  | -1.2774 | 0.2704  | 0.9142  | 1.7296  | 0.10529                 |
| 11  | -1.2275 | 0.2217  | 0.9296  | 1.7783  | 0.08643                 |
| 12  | -1.1865 | 0.1818  | 0.9423  | 1.8182  | 0.07089                 |
| 13  | -1.1530 | 0.1491  | 0.9527  | 1.8509  | 0.06162                 |
| 14  | -1.1254 | 0.1223  | 0.9612  | 1.8777  | 0.04764                 |
| 15  | -1.1029 | 0.1003  | 0.9682  | 1.8997  | 0.03903                 |
| 16  | -1.0844 | 0.0822  | 0.9739  | 1.9178  | 0.03209                 |
| 17  | -1.0692 | 0.0674  | 0.9786  | 1.9326  | 0.02629                 |
| 18  | -1.0567 | 0.0553  | 0.9824  | 1.9447  | 0.02152                 |

Se presenta a continuación un algoritmo para resolver sistemas de ecuaciones lineales con el método iterativo, en sus dos versiones: desplazamientos simultáneos y desplazamientos sucesivos.

### ALGORITMO 3.11 Métodos de Jacobi y Gauss-Seidel

Para encontrar la solución aproximada del sistema de ecuaciones  $A \mathbf{x} = \mathbf{b}$  proporcionar los

**DATOS:** El número de ecuaciones  $N$ , la matriz coeficiente  $A$ , el vector de términos independientes  $\mathbf{b}$ , el vector inicial  $\mathbf{x}_0$ , el número máximo de iteraciones  $\text{MAXIT}$ , el valor de  $\text{EPS}$  y  $M = 0$  para usar JACOBI o  $M \neq 0$  para usar GAUSS-SEIDEL.

**RESULTADOS:** La solución aproximada  $\mathbf{x}$  y el número de iteraciones  $K$  en que se alcanzó la convergencia o mensaje "NO SE ALCANZÓ LA CONVERGENCIA", la última aproximación a  $\mathbf{x}$  y  $\text{MAXIT}$ .

**PASO 1.** Arreglar la matriz aumentada de modo que la matriz coeficiente quede lo más cercana posible a la diagonal dominante (véase Prob. 3.55).

**PASO 2.** Hacer  $K = 1$

**PASO 3.** Mientras  $K \leq \text{MAXIT}$ , repetir los pasos 4 a 18.

- PASO 4. Si  $M = 0$  ir al paso 5. De otro modo Hacer  $*x = x_0$ .
- PASO 5. Hacer  $I = 1$ .
- PASO 6. Mientras  $I \leq N$ , repetir los pasos 7 a 14.
- PASO 7. Hacer  $SUMA = 0$ .
- PASO 8. Hacer  $J = 1$ .
- PASO 9. Mientras  $J \leq N$ , repetir los pasos 10 a 12.
- PASO 10. Si  $J = I$  ir al paso 12.
- PASO 11. Hacer  $SUMA = SUMA + A(I, J) * x_0(J)$
- PASO 12. Hacer  $J = J + 1$ .
- PASO 13. Si  $M = 0$ , hacer  $x(I) = (b(I) - SUMA) / A(I, I)$ .  
De otro modo hacer  $x_0(I) = (b(I) - SUMA) / A(I, I)$ .
- PASO 14. Hacer  $I = I + 1$ .
- PASO 15. Si  $|x - x_0| \leq EPS$  ir al paso 19. De otro modo continuar.
- PASO 16. Si  $M = 0$ , hacer  $x_0 = x$ .
- PASO 17. Hacer  $K = K + 1$ .
- PASO 18. IMPRIMIR mensaje "NO SE ALCANZÓ LA CONVERGENCIA", el vector  $x$ , MAXIT y el mensaje "ITERACIONES" y TERMINAR.
- PASO 19. IMPRIMIR el mensaje "VECTOR SOLUCIÓN",  $x$ ,  $K$  y el mensaje "ITERACIONES" y TERMINAR.

\* Operaciones vectoriales.

**SUGERENCIA:** Una vez más se recomienda programar el algoritmo 3.11 en un lenguaje de alto nivel (véase PROGRAMA 3.3 del CD), o bien en una calculadora o en un pizarrón electrónico o en Matlab, donde las operaciones vectoriales se ejecutan con sólo indicárlas.

### ACELERACIÓN DE CONVERGENCIA

Si aún después de arreglado el sistema por resolver  $Ax = b$ , conforme la pauta del teorema 3.2, no se obtiene convergencia por los métodos de Jacobi y Gauss-Seidel o es muy lenta (como sucedió con el sistema 3.98 de la sección anterior), puede recurrirse a los métodos de **relajación** que, como se hará notar posteriormente, son los métodos de Jacobi y Gauss-Seidel afectados por un factor de peso  $w$  que, elegido adecuadamente, puede producir convergencia o acelerarla si ya existe. Se describen a continuación estos métodos para un sistema de  $n$  ecuaciones en  $n$  incógnitas.

Llámesse  $N$  la matriz coeficiente del sistema por resolver, una vez que haya sido llevada a la forma más cercana posible a diagonal dominante, y después de dividir la primera fila entre  $a_{1,1}$ , la segunda entre  $a_{2,2}$ , ..., y la  $n$ -ésima entre  $a_{n,n}$ .  $N$  es una matriz con unos en la diagonal principal. A continuación descompóngase  $N$  en la siguiente forma:

$$N = L + I + U,$$

donde  $L$  es una matriz cuyos elementos por debajo de su diagonal principal son idénticos a los correspondientes de  $N$  y ceros en cualquier otro sitio,  $I$  es la matriz identidad y  $U$  una matriz cuyos elementos arriba de la diagonal principal son idénticos a los correspondientes de  $N$  y cero en cualquier otro sitio. Sustituyendo esta descomposición de  $N$ , el sistema que se quiere resolver quedaría:

$$(L + I + U)x = b \quad (3.100)$$

Si ahora se suma  $x$  a cada miembro de la ecuación 3.100 se obtiene

$$(L + I + U)x + x = b + x$$

“Despejando”  $\mathbf{x}$  del lado izquierdo, se llega al esquema siguiente

$$\mathbf{x} = \mathbf{x} + [\mathbf{b} - L\mathbf{x} - \mathbf{x} - U\mathbf{x}], \quad (3.101)$$

que puede utilizarse para iterar a partir de un vector inicial  $\mathbf{x}^{(0)}$ . Nótese que la ecuación 3.101, puede reducirse a la ecuación 3.89, ya que sólo es un rearrreglo de ésta.

Al aplicar la ecuación 3.101, pueden presentarse de nuevo las dos variantes que dieron lugar a los métodos de Jacobi y Gauss-Seidel, con lo que el esquema de desplazamientos simultáneos quedaría:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + [\mathbf{b} - L\mathbf{x}^{(k)} - \mathbf{x}^{(k)} - U\mathbf{x}^{(k)}] \quad (3.102)$$

y el de desplazamientos sucesivos así:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + [\mathbf{b} - L\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} - U\mathbf{x}^{(k)}] \quad (3.103)$$

Llegar al esquema 3.102 y 3.103 no es simplemente para tener una versión distinta de las ecuaciones 3.89, sino para someterlo a un análisis que permita proponer “nuevos métodos” o mejoras en los que ya se tienen. Por ejemplo, factorizando  $\mathbf{x}^{(k)}$  dentro del paréntesis rectangular de la ecuación 3.102, se tiene:

$$\mathbf{b} - (L + I + U)\mathbf{x}^{(k)} = \mathbf{b} - N\mathbf{x}^{(k)} = \mathbf{r}^{(k)} \quad (3.104)$$

vector que se denota como  $\mathbf{r}^{(k)}$  y se llama **vector residuo** de la  $k$ -ésima iteración y puede tomarse como una medida de la cercanía de  $\mathbf{x}^{(k)}$  al vector solución  $\mathbf{x}$ ; si las componentes de  $\mathbf{r}^{(k)}$  o  $|\mathbf{r}^{(k)}|$  son pequeñas,  $\mathbf{x}^{(k)}$  suele ser una buena aproximación a  $\mathbf{x}$ ; pero si los elementos de  $\mathbf{r}^{(k)}$  o  $|\mathbf{r}^{(k)}|$  son grandes, puede pensarse que  $\mathbf{x}^{(k)}$  no es muy cercana a  $\mathbf{x}$ . Aunque hay circunstancias donde esto no se cumple, por ejemplo, cuando el sistema por resolver está **mal condicionado** (véase Sec. 3.4), *es práctico tomar estos criterios como válidos*.

Al sustituir la ecuación 3.104, en la 3.102 queda

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{r}^{(k)} \quad (3.105)$$

que puede verse como un esquema iterativo donde el vector de la  $(k+1)$ -ésima iteración se obtiene a partir del vector de la  $k$ -ésima iteración y el residuo correspondiente.

Si la aplicación de la ecuación 3.105 a un sistema particular da convergencia lenta, entonces  $\mathbf{x}^{(k+1)}$  y  $\mathbf{x}^{(k)}$  están muy cercanas entre sí, y para que la convergencia se acelere puede intentarse afectar  $\mathbf{r}^{(k)}$  con un peso  $w > 1$  (**sobrerrelajar** el proceso); si, en cambio, el proceso diverge  $|\mathbf{r}^{(k)}|$  es grande y convendría afectar  $\mathbf{r}^{(k)}$  con un factor  $w < 1$  (**subrelajar** el proceso), para provocar la convergencia. El esquema 3.105 quedaría en general así:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + w\mathbf{r}^{(k)} \quad (3.106)$$

o

$$x_i^{k+1} = x_i^{(k)} + w \left[ b_i - \sum_{\substack{j=1 \\ j \neq i}}^n a_{ij} x_j^{(k)} \right] \quad 1 \leq i \leq n, \quad (3.107)$$

para desplazamientos simultáneos.

Para desplazamientos sucesivos, en cambio, quedaría

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + w [\mathbf{b} - L\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} - U\mathbf{x}^{(k)}] \quad (3.108)$$

o

$$x_i^{k+1} = x_i^{(k)} + w \left[ b_i - \sum_{j=1}^{i-1} l_{ij} x_j^{k+1} - x_i^{(k)} \sum_{j=i+1}^n u_{ij} x_j^{(k)} \right] \quad 1 \leq i \leq n, \quad (3.109)$$



Estos métodos se abrevian frecuentemente como SOR (del inglés *Successive Over-Relaxation*).

En general, el cálculo de  $w$  es complicado y sólo para sistemas especiales (matriz coeficiente positivamente definida y tridiagonal) se tiene una fórmula.\*

### Ejemplo 3.40

Resuelva el sistema 3.98

$$\begin{aligned} -x_1 + 3x_2 + 5x_3 + 2x_4 &= 10 \\ x_1 + 9x_2 + 8x_3 + 4x_4 &= 15 \\ x_2 &= 2 \\ 2x_1 + x_2 + x_3 - x_4 &= -3 \end{aligned}$$

con desplazamientos sucesivos,  $w = 1.3$  y con  $\varepsilon = 10^{-2}$  aplicado a  $\|x^{(k+1)} - x^{(k)}\|$ . (Puede seguir los cálculos con un pizarrón electrónico o con Matlab.)

**Solución** La matriz  $N$  y el vector de términos independientes correspondiente son

$$N = \begin{bmatrix} 1 & 1/2 & 1/2 & -1/2 \\ 1/9 & 1 & 8/9 & 4/9 \\ -1/5 & 3/5 & 1 & 2/5 \\ 0 & 1 & 0 & 1 \end{bmatrix}, \quad \mathbf{b} = [-3/2 \quad 15/9 \quad 10/5 \quad 2]^T$$

Descomposición de  $N$

$$L = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1/9 & 0 & 0 & 0 \\ -1/5 & 3/5 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad U = \begin{bmatrix} 0 & 1/2 & 1/2 & -1/2 \\ 0 & 0 & 8/9 & 4/9 \\ 0 & 0 & 0 & 2/5 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

### Primera iteración

Obtención de  $x^{(1)}$  a partir del vector inicial  $x^{(0)} = [0 \ 0 \ 0 \ 0]^T$  y empleando la ecuación 3.108.

Cálculo de  $x_1^1$ , esto es,  $i = 1$  y  $k + 1 = 1$

$$x_1^1 = x_1^0 + 1.3 \left[ b_1 - \sum_{j=1}^0 l_{1j} x_j^1 - x_1^0 - \sum_{j=2}^4 u_{1j} x_j^0 \right]$$

Obsérvese que en la primera sumatoria el valor inicial ( $j=1$ ) es mayor que el valor final (0); la convención en estos casos es que tal sumatoria no se realiza. Por tanto,

$$x_1^1 = 0 + 1.3[-3/2 - 0 - 1/2(0) - 1/2(0) + 1/2(0)] = -1.95$$

Cálculo de  $x_2^1$ , esto es,  $i = 2$  y  $k + 1 = 1$

$$x_2^1 = x_2^0 + 1.3 \left[ b_2 - \sum_{j=1}^1 l_{2j} x_j^1 - x_2^0 - \sum_{j=3}^4 u_{2j} x_j^0 \right]$$

$$= 0 + 1.3 [15/9 - 1/9(-1.95) - 0 - 8/9(0) - 4/9(0)] = 2.4483$$

Cálculo de  $x_3^1$ , esto es,  $i = 3$  y  $k + 1 = 1$

$$x_3^1 = x_3^0 + 1.3 \left[ b_3 - \sum_{j=1}^2 l_{3j} x_j^1 - x_3^0 - \sum_{j=4}^4 u_{3j} x_j^0 \right]$$

$$= 0 + 1.3 [10/5 - (-1/5)(-1.95) - (3/5)(2.4483) - 0 - 2/5(0)] = 0.1833$$

\* Burden, R.L. y Faires, J.D. *Análisis numérico*. Grupo Editorial Iberoamérica (1985), pp. 475.

Cálculo de  $x_4^1$ , esto es,  $i = 4$  y  $k+1 = 1$

$$x_4^1 = x_4^0 + 1.3 \left[ b_4 - \sum_{j=1}^3 l_{4,j} x_j^1 - x_4^0 - \sum_{j=5}^4 u_{4,j} x_j^0 \right]$$

$$= 0 + 1.3 [2 - 0(-1.95) - 1(2.4483) - 0(0.1833) - 0] = -0.5828$$

Cálculo de  $\|x^{(1)} - x^{(0)}\| = d_1$

$$d_1 = \sqrt{(x_1^1 - x_1^0)^2 + (x_2^1 - x_2^0)^2 + (x_3^1 - x_3^0)^2 + (x_4^1 - x_4^0)^2}$$

$$= \sqrt{(-1.95)^2 + (2.4483)^2 + (0.1833)^2 + (-0.5828)^2} = 3.1891$$

Los valores mostrados en la tabla 3.5 se encuentran continuando las iteraciones.

**Tabla 3.5** Resultados obtenidos con  $w = 1.3$ .

| $k$ | $x_1^k$ | $x_2^k$ | $x_3^k$ | $x_4^k$ | $\ x^{(k+1)} - x^{(k)}\ $ |
|-----|---------|---------|---------|---------|---------------------------|
| 0   | 0.0000  | 0.0000  | 0.0000  | 0.0000  | —                         |
| 1   | -1.9500 | 2.4483  | 0.1833  | -0.5828 | 3.1891                    |
| 2   | -3.4544 | 2.0561  | 0.3462  | 0.1020  | 1.7066                    |
| 3   | -2.4089 | 1.4388  | 0.6945  | 0.6989  | 1.3971                    |
| 4   | -2.1597 | 0.8406  | 0.8110  | 1.2976  | 0.8898                    |
| 5   | -1.5322 | 0.4489  | 0.9334  | 1.6271  | 0.8190                    |
| 6   | -1.3312 | 0.2055  | 0.9674  | 1.8447  | 0.3848                    |
| 7   | -1.1140 | 0.0822  | 0.9968  | 1.9397  | 0.2689                    |
| 8   | -1.0563 | 0.0220  | 1.0005  | 1.9895  | 0.0972                    |
| 9   | -1.0046 | -0.0004 | 1.0045  | 2.0037  | 0.0583                    |
| 10  | -0.9988 | -0.0074 | 1.0028  | 2.0084  | 0.0103                    |
| 11  | -0.9919 | -0.0070 | 1.0024  | 2.0066  | 0.0072                    |

Para realizar los cálculos puede usarse el siguiente guión de Matlab:



```
%Método SOR
clear
A=[2 1 1 -1; 1 9 8 4; -1 3 5 2; 0 1 0 1]
b=[-3 15 10 2]
for i=1:4
    N(i,:)=A(i,:)/A(i,i);
    b(i)=b(i)/A(i,i);
end
N
U=triu(N)
L=tril(N)
X0=zeros(1,4);
X=X0; w=1.3;
K=0; Norma=1;
fprintf(' K X(1) X(2) X(3) X(4) Norma\n')
while Norma>0.01
```

```

k=k+1;
fprintf('%2s',k)
for i=1:4
    sumaL=0; sumaU=0;
    for j=1:4
        if i~=j
            sumaL=sumaL+L(i,j)*X(j);
            sumaU=sumaU+U(i,j)*X0(j);
        end
    end
    X(i)=X(i)+w*(b(i)-sumaL-X(i)-sumaU);
    fprintf('%10.4f',X(i))
end
Norma=norm(X0-X);
fprintf('%10.4f\n',Norma)
X0=X;
if K > 17
    disp('No se alcanzó la convergencia')
    break
end
end
end

```

Al comparar estos resultados con los obtenidos en la tabla 3.4 (método de Gauss-Seidel aplicado al sistema que aquí se resuelve), se observa que la convergencia es acelerada y los cálculos se reducen a la mitad.

### COMPARACIÓN DE LOS MÉTODOS DIRECTOS E ITERATIVOS

Una parte importante el análisis numérico es conocer las características (ventajas y desventajas) de los métodos numéricos básicos que resuelven una familia de problemas (en este caso  $A \mathbf{x} = \mathbf{b}$ ), para así elegir el algoritmo más adecuado para cada problema.

A continuación se presentan las circunstancias donde pudiera verse como ventajosa la elección de un método iterativo y también a qué se renuncia con esta decisión.

**Tabla 3.6** Ventajas y desventajas de los métodos iterativos comparados con los métodos directos.

| Ventajas                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Desventajas                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                  |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ol style="list-style-type: none"> <li>1. Probablemente más eficientes que los directos para sistemas de orden muy alto.</li> <li>2. Más simples de programar.</li> <li>3. Puede aprovecharse una aproximación a la solución, si tal aproximación existe.</li> <li>4. Se obtienen fácilmente aproximaciones burdas de la solución.</li> <li>5. Son menos sensibles a los errores de redondeo (valioso en sistemas mal condicionados).</li> <li>6. Se requiere menos memoria de máquina. Generalmente las necesidades de memoria son proporcionales al orden de la matriz.</li> </ol> | <ol style="list-style-type: none"> <li>1. Si se tienen varios sistemas que comparten la matriz coeficiente, esto no representará ahorro de cálculos ni tiempo de máquina, ya que por cada vector a la derecha de <math>A</math> tendrá que aplicarse el método seleccionado.</li> <li>2. Aun cuando la convergencia esté asegurada, puede ser lenta y, por tanto, los cálculos requeridos para obtener una solución particular no son predecibles.</li> <li>3. El tiempo de máquina y la exactitud del resultado dependen del criterio de convergencia.</li> <li>4. Si la convergencia es lenta, los resultados deben interpretarse con cautela.</li> <li>5. No se tiene ventaja particular alguna (tiempo de máquina por iteración) si la matriz coeficiente es simétrica.</li> <li>6. No se obtiene <math>A^{-1}</math> ni <math>\det A</math>.</li> </ol> |



## 3.6 Valores y vectores propios

Si  $A$  es una matriz de números reales de orden  $n$  e  $I$  la matriz identidad de orden  $n$ , el polinomio definido por

$$p(\lambda) = \det(A - \lambda I) \quad (3.110)$$

se llama el **polinomio característico** de  $A$ .

Es fácil ver que  $p$  es un polinomio de  $n$ -ésimo grado en  $\lambda$  con coeficientes reales,\* y que, por tanto, la ecuación

$$p(\lambda) = 0 \quad (3.111)$$

tiene  $n$  raíces, de las cuales algunas suelen ser complejas. Los ceros de esta ecuación, conocidos como **valores característicos** o **propios** de  $A$ , están ligados con la solución del sistema  $A\mathbf{x} = \mathbf{b}$ . Por ejemplo, el método de Gauss-Seidel, independientemente del vector inicial que se emplee, converge a la solución de  $A\mathbf{x} = \mathbf{b}$  si y sólo si los valores propios de  $B$  son todos menores de uno en valor absoluto.\*\*

### Ejemplo 3.41

Dada la siguiente matriz, encuentre sus valores propios

$$A = \begin{bmatrix} 4 & -9 & 2 \\ 2 & -4 & 6 \\ 1 & -1 & 3 \end{bmatrix}$$

**Solución** Se forma  $A - \lambda I$



$$A - \lambda I = \begin{bmatrix} 4 & -9 & 2 \\ 2 & -4 & 6 \\ 1 & -1 & 3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 4 - \lambda & -9 & 2 \\ 2 & -4 - \lambda & 6 \\ 1 & -1 & 3 - \lambda \end{bmatrix}$$

Se obtiene el determinante de este último arreglo

$$\begin{aligned} \det(A - \lambda I) &= (4 - \lambda)(-4 - \lambda)(3 - \lambda) - 4 - 54 - \\ &= (2)(-4 - \lambda)(1) - (-9)(2)(3 - \lambda) - (6)(-1)(4 - \lambda) \end{aligned}$$

Al desarrollar e igualar con cero se obtiene

$$-\lambda^3 + 3\lambda^2 - 6\lambda - 20 = 0,$$

el polinomio característico de  $A$ , cuyos ceros  $\lambda_1, \lambda_2, \lambda_3$  son los valores buscados.

El hecho de ser un polinomio cúbico con coeficientes reales garantiza una raíz real por lo menos. Con el método de Newton-Raphson y un valor inicial de  $-2$  se llega a

$$\lambda_1 = -1.53968$$

El polinomio se degrada por división sintética

$$\begin{array}{r|rrrr} -1.53968 & -1 & 3 & -6 & -20 \\ & & 1.53968 & -6.98965 & 20 \\ \hline & -1 & 4.53968 & -12.98965 & 0 \end{array}$$

\* Véase problema 3.59.

\*\* J. N. Franklin, *Matrix Theory*. Prentice Hall, 1968.

El polinomio degradado es:

$$-\lambda^2 + 4.53968\lambda - 12.98965 = 0,$$

de donde, por aplicación de la fórmula cuadrática se tiene

$$\lambda_2 = 2.26984 + 2.799553 i$$

$$\lambda_3 = 2.26984 - 2.799553 i$$

Una vez obtenidos los valores propios de una matriz  $A$  de orden  $n$ , los vectores  $\mathbf{x} \neq \mathbf{0}$  que resuelven el sistema

$$\begin{aligned} A\mathbf{x} &= \lambda_i \mathbf{x}, & i &= 1, 2, \dots, n \\ (A - \lambda_i I)\mathbf{x} &= \mathbf{0} \end{aligned} \quad (3.112)$$

se denominan **vectores propios** de  $A$  correspondientes a  $\lambda_i$ . Como  $\det(A - \lambda_i I) = 0$  y el sistema es homogéneo, se tiene un número infinito de soluciones para cada  $\lambda_i$ .

### Ejemplo 3.42

Encuentre los vectores propios de la matriz del ejemplo 3.41, correspondientes al valor propio  $\lambda_1 = -1.53968$ .

### Solución

Al resolver el sistema por alguno de los métodos de eliminación



$$(A - \lambda_1 I)\mathbf{x} = \begin{bmatrix} 4 - (-1.53968) & -9 & 2 \\ 2 & -4 - (-1.53968) & 6 \\ 1 & -1 & 3 - (-1.53968) \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

resulta una matriz triangular superior, por lo menos con una fila de ceros.\* Para asegurar que esa(s) fila(s) de ceros sea(n) la(s) última(s) y que la submatriz no singular resultante esté lo mejor condicionada posible, se usa **pivoteo total** (intercambio de filas y columnas) y escalamiento.

Sea entonces la matriz por triangularizar

$$\begin{bmatrix} 5.53968 & -9 & 2 \\ 2 & -2.46032 & 6 \\ 1 & -1 & 4.53968 \end{bmatrix} = B$$

Nótese que el vector de términos independientes no se emplea porque todos sus componentes son cero.

En lugar de emplear la **norma euclídeana** para el escalamiento, se usará ahora la siguiente norma, definida para un vector cualquiera  $\mathbf{y} = [y_1, y_2, \dots, y_n]^T$ , como

$$\mathbf{y} = |y_1| + |y_2| + \dots + |y_n|$$

ya que es más sencilla de calcular que la euclídeana y que para la primera, segunda y tercera filas de  $A$  es, respectivamente,

$$\begin{bmatrix} 16.53968 \\ 10.46032 \\ 6.53968 \end{bmatrix}$$

\* Pizar, M. S. *Numerical Computing and Mathematical Analysis*. S.R.A. (1975)

Cada fila de la matriz  $B$  se divide entre su factor de escalamiento y se obtiene:

$$B' = \begin{bmatrix} 0.33493 & -0.54415 & 0.12092 \\ 0.19120 & -0.23520 & 0.57360 \\ 0.15291 & -0.15291 & 0.69417 \end{bmatrix}$$

En el pivoteo total es necesario registrar los cambios de columnas que se verifican, ya que éstos afectan el orden de las incógnitas. Para ello se utilizará un vector  $\mathbf{q}$ , en donde aparecen como elementos las columnas. Al principio están en orden natural y se tiene:

$$\mathbf{q} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$$

Se busca el elemento de máximo valor absoluto de  $B'$ . En este caso es  $b'_{3,3} = 0.69417$ . Se intercambian las filas 1 y 3, y las columnas 1 y 3 para llevar este elemento a la posición pivote (1, 1), teniendo cuidado de registrar los intercambios de columnas en  $\mathbf{q}$ . Los resultados son:

$$B'' = \begin{bmatrix} 0.69417 & -0.15291 & 0.15291 \\ 0.57360 & -0.23520 & 0.19120 \\ 0.12092 & -0.54415 & 0.33493 \end{bmatrix}, \quad \mathbf{q} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$$

Se eliminan los elementos de la primera columna que están debajo del elemento pivote, con lo cual se produce:

$$B''' = \begin{bmatrix} 0.69417 & -0.15291 & 0.15291 \\ 0.0 & -0.10885 & 0.06485 \\ 0.0 & -0.51751 & 0.30830 \end{bmatrix}$$

Se busca el elemento de máximo valor absoluto en las dos últimas filas; resulta ser  $b'''_{3,2} = -0.51751$ . Se intercambian las filas 2 y 3, y con esto se lleva a este elemento a la posición pivote (2, 2). Los resultados son:

$$B^{IV} = \begin{bmatrix} 0.69417 & -0.15291 & 0.15291 \\ 0.0 & -0.51751 & 0.30830 \\ 0.0 & -0.10885 & 0.06485 \end{bmatrix}$$

y  $\mathbf{q} = \begin{bmatrix} 3 \\ 2 \\ 1 \end{bmatrix}$ , ya que no hubo intercambio de columnas.

Se eliminan los elementos de la segunda columna que están debajo del elemento pivote y se produce

$$B^V = \begin{bmatrix} 0.69417 & -0.15291 & 0.15291 \\ 0.00000 & -0.51751 & 0.30830 \\ 0.00000 & 0.00000 & -0.00000 \end{bmatrix}$$

una matriz triangularizada con una fila de ceros, la última como se planeó. La submatriz no singular de la que se habló al principio está formada por los elementos (1, 1), (1, 2), (2, 1) y



(2, 2). Al escribir el sistema en términos de  $x_1, x_2$  y  $x_3$ , y considerar los cambios de columnas que hubo, se tiene

$$0.69417 x_3 - 0.15291 x_2 + 0.15291 x_1 = 0$$

$$0.00000 x_3 - 0.51751 x_2 + 0.30830 x_1 = 0$$

Un sistema homogéneo de dos ecuaciones en tres incógnitas, cuyas infinitas soluciones pueden obtenerse en términos de alguna de las incógnitas. El sistema se resuelve en términos de  $x_1$

$$0.69417 x_3 - 0.15291 x_2 = -0.15291 x_1$$

$$0.00000 x_3 - 0.51751 x_2 = -0.30830 x_1$$

de donde

$$x_2 = 0.59573 x_1$$

$$x_3 = -0.08905 x_1$$

Se da un valor particular a  $x_1$ , por ejemplo  $x_1 = 1$ , y resulta

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.59573 \\ -0.08905 \end{bmatrix}$$

uno de los infinitos vectores propios de  $A$  correspondientes a  $\lambda_1$ .

### Comprobación

Ya que por definición  $A \mathbf{x} = \lambda_1 \mathbf{x}$

$$\begin{bmatrix} 4 & -9 & 2 \\ 2 & -4 & 6 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 1 \\ 0.59573 \\ -0.08905 \end{bmatrix} = -1.53968 \begin{bmatrix} 1 \\ 0.59573 \\ -0.08905 \end{bmatrix}$$

### MÉTODO DE LAS POTENCIAS

El método de las potencias permite calcular el valor y el vector característicos dominantes de una matriz  $A$  de orden  $n$ , cuando dicha matriz tiene  $n$  vectores característicos linealmente independientes:  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$  y un valor característico  $\lambda_1$  estrictamente dominante en magnitud

$$|\lambda_1| > |\lambda_2| \geq |\lambda_3| \geq \dots \geq |\lambda_n|$$

Se muestra a continuación dicho método.

Dada la independencia lineal de los vectores característicos, cualquier vector  $\mathbf{v}$  de  $n$  componentes puede expresarse como una combinación lineal de ellos:

$$\mathbf{v} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_n \mathbf{v}_n$$

Multiplicando la ecuación anterior por la izquierda por  $A$  se tiene:

$$\begin{aligned} A\mathbf{v} &= a_1 A\mathbf{v}_1 + a_2 A\mathbf{v}_2 + \dots + a_n A\mathbf{v}_n \\ &= a_1 \lambda_1 \mathbf{v}_1 + a_2 \lambda_2 \mathbf{v}_2 + \dots + a_n \lambda_n \mathbf{v}_n \end{aligned}$$

Multiplicando repetidamente por  $A$  se llega a

$$A^k \mathbf{v} = a_1 \lambda_1^k \mathbf{v}_1 + a_2 \lambda_2^k \mathbf{v}_2 + \dots + a_n \lambda_n^k \mathbf{v}_n$$

y factorizando

$$A^k \mathbf{v} = \lambda_1^k \left[ a_1 \mathbf{v}_1 + a_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k \mathbf{v}_2 + \dots + a_n \left( \frac{\lambda_n}{\lambda_1} \right)^k \mathbf{v}_n \right]$$

y como  $\lambda_1$  es el mayor, todos los términos dentro del paréntesis rectangular tienden a cero cuando  $k$  tiende a  $\infty$ , excepto el primer término (si  $a_1 \neq 0$ ). Para  $k$  grande  $A^k \mathbf{v} \approx \lambda_1^k a_1 \mathbf{v}_1$ .

Al tomar la relación de cualesquiera componentes correspondientes a  $A^k \mathbf{v}$  y  $A^{k+1} \mathbf{v}$ , se obtiene una sucesión de valores convergentes a  $\lambda_1$ , ya que

$$\frac{\lambda_1^{k+1} a_1 \mathbf{v}_{1,j}}{\lambda_1^k a_1 \mathbf{v}_{1,j}} \approx \lambda_1 \quad (3.113)$$

Además, la sucesión  $\lambda_1^{-k} A^k \mathbf{v}$  convergirá al vector característico  $\mathbf{v}_1$  multiplicado por  $a_1$ .

### Ejemplo 3.43

Encuentre el valor característico y el vector característico dominantes de la matriz coeficiente del siguiente sistema, usando el método de las potencias

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

**Solución** Como generalmente no se conocen los vectores característicos, sino que ése es el propósito, se empieza a iterar con  $\mathbf{v} = \mathbf{e}_1 = [1 \ 0 \ 0]^T$ .

#### Primera iteración

Primero se calcula el producto  $A \mathbf{v}$

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

Ahora se calcula el producto  $A^2 \mathbf{v}$

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

Se calcula el primero de los valores de la ecuación 3.113, utilizando el primer componente de ambos productos

$$\lambda_{1,1} \approx 5/1 = 5$$

#### Segunda iteración

Se calcula el producto  $A^3 \mathbf{v}$

$$\begin{bmatrix} 1 & 2 & 0 \\ 2 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}^3 \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 13 \\ 14 \\ 0 \end{bmatrix}$$

El nuevo valor de la ecuación 3.113 es:

$$\lambda_{1,2} \approx 13/5 = 2.6$$

Al continuar las iteraciones se obtiene:

| $k$ | $\lambda_{1,k}$ |
|-----|-----------------|
| 1   | 5.00000         |
| 2   | 2.60000         |
| 3   | 3.15385         |
| 4   | 2.95122         |
| 5   | 3.01653         |
| 6   | 2.99452         |
| 7   | 3.00183         |
| 8   | 2.99939         |
| 9   | 3.00020         |
| 10  | 2.99993         |

El proceso converge al valor propio dominante  $\lambda_1 = 3$ . El lector puede repetir el proceso, usando la segunda componente de cada producto  $A^k \mathbf{v}$ .

Para encontrar uno de los vectores propios correspondientes a  $\lambda_1 = 3$ , se usa la fórmula  $\lambda_1^{-k} A^k \mathbf{v}$ , resultando

$$\mathbf{v}_1 = [9841.7 \quad 9841.3 \quad 0]^T, \text{ que normalizado da } \mathbf{v}_1 = [1 \quad 1 \quad 0]^T.$$

Debido a que  $A^k$  produce, por lo general, valores muy grandes o muy pequeños, conviene normalizar los productos  $A^k \mathbf{v}$  en cada iteración, dividiendo cada elemento del vector entre el elemento de máximo valor absoluto de dicho vector.

Los cálculos pueden realizarse con el siguiente guión de Matlab, en el que se obtiene la ecuación 3.113 con el elemento de máximo valor del segundo producto  $A^k \mathbf{v}$  obtenido en cada iteración.



```
A=[1 2 0;2 1 0;0 0 -1];
v= [1; 0; 0];
Dist=1;R=0
Eps=1e-5;K=0;
while Dist>Eps
    K=K+1;
    X=A^K*v;
    Y=A^(K+1)*v;
    [Z,I]=max(Y);
    R1=Y(I)/X(I);
    fprintf(' %2d %10.5f\n',K,R)
    Dist=abs(R1-R);
    R=R1;
end
v1=A^K*v/R
```



## Ejercicios



- 3.1 En una columna de cinco platos, se requiere absorber benceno contenido en una corriente de gas V, con un aceite L que circula a contracorriente del gas. Considérese que el benceno transferido no altera sustancialmente el número de moles de V y L fluyendo a contracorriente, que la relación de equilibrio está dada por la ley de Henry ( $y = mx$ ) y que la columna opera a régimen permanente. Calcule la composición del benceno en cada plato.

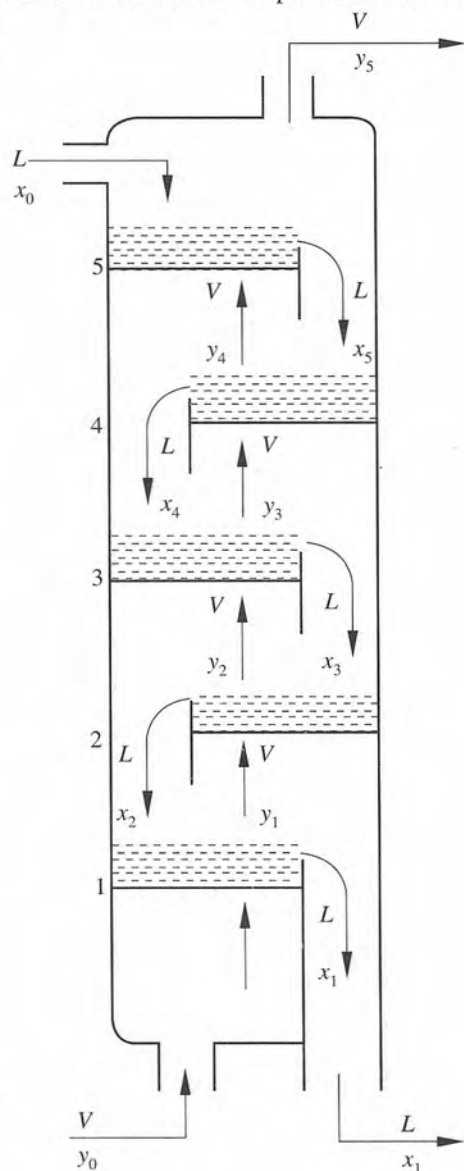
Datos:  $V = 100$  moles/min;  $L = 500$  moles/min.

$y_0 = 0.09$  fracción molar de benceno en V.

$x_0 = 0.0$  fracción molar de benceno en L (el aceite entra por el domo sin benceno).

$m = 0.12$

**Solución** Los balances de materia para el benceno en cada plato son (véase Fig. 3.13).



**Fig. 3.13**  
Columna de absorción de cinco platos.

| Plato | Balance de benceno                |
|-------|-----------------------------------|
| 5     | $L(x_0 - x_5) + V(y_4 - y_5) = 0$ |
| 4     | $L(x_5 - x_4) + V(y_3 - y_4) = 0$ |
| 3     | $L(x_4 - x_3) + V(y_2 - y_3) = 0$ |
| 2     | $L(x_3 - x_2) + V(y_1 - y_2) = 0$ |
| 1     | $L(x_2 - x_1) + V(y_0 - y_1) = 0$ |

Al sustituir la información que se tiene, las consideraciones hechas y reorganizando las ecuaciones, se llega a:

$$\begin{array}{rclclcl}
 512 x_5 & + & 12 x_4 & & & = & 0 \\
 500 x_5 & - & 512 x_4 & + & 12 x_3 & = & 0 \\
 & & 500 x_4 & - & 512 x_3 & + & 12 x_2 & = & 0 \\
 & & & & 500 x_3 & - & 512 x_2 & + & 12 x_1 & = & 0 \\
 & & & & & & - & 500 x_2 & + & 512 x_1 & = & 9
 \end{array}$$

Con el **PROGRAMA 3.2** del CD, se obtienen los siguientes resultados

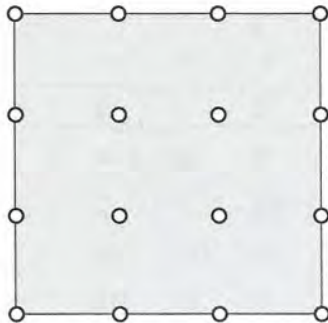
$$\begin{array}{lll}
 x_1 = 0.018, & x_2 = 4.32 \times 10^{-4}, & x_3 = 1.037 \times 10^{-5}, \\
 x_4 = 2.4869 \times 10^{-7}, & x_5 = 5.8286 \times 10^{-9}, &
 \end{array}$$

También pueden usarse las instrucciones en Matlab dadas en el ejemplo 3.28, con los cambios apropiados en los datos.



3.2

Supóngase que se tiene una estructura cuadrada. A fin de analizarla se forma una malla imaginaria sobre dicha estructura, como se muestra en la figura siguiente.



**Figura 3.14**  
Estructura cuadrada.

Y se numeran los nodos, por ejemplo, como se muestra a continuación.

$$\begin{array}{cccc}
 4 & 8 & 12 & 16 \\
 3 & 7 & 11 & 15 \\
 2 & 6 & 10 & 14 \\
 1 & 5 & 9 & 13
 \end{array}$$

Cada nodo se identifica con  $a_{ij}$ ; por ejemplo, el 4 con  $a_{1,1}$ , el 6 con  $a_{3,2}$ , etc., y así queda formada una matriz  $A$  representativa de la estructura.

Ciertas consideraciones de ingeniería determinan que  $a_{ij} \neq 0$  siempre que los nodos  $i$  y  $j$  sean vecinos o adyacentes.\* Para aclararlo, hay que observar que al nodo 5 le corresponde  $a_{4,2}$ , y como los nodos 4 y 2 no son vecinos  $a_{4,2} = 0$ ; al 11 en cambio le corresponde  $a_{2,3}$  y como 2 y 3 son vecinos,  $a_{2,3} \neq 0$ . Por último  $a_{3,3} \neq 0$ , ya que el nodo 3 puede considerarse vecino consigo mismo.

Estas consideraciones generan matrices o sistemas dispersos o frecuentemente bandeados. Estos sistemas suelen ser muy grandes, ya que las mallas se construyen con un gran número de nodos.

En la aplicación del método de las rigideces,\*\* para calcular los desplazamientos en los nodos de una estructurada dada al aplicarse una carga en uno de los nodos, se obtiene el siguiente sistema de ecuaciones:

$$10^5 \begin{bmatrix} 3.01687 & 0.00000 & 3.3750 & -3.0000 & 0.00000 & 0.0000 \\ 0.00000 & 3.01687 & 3.3750 & 0.0000 & -0.01687 & 3.3750 \\ 3.37500 & 3.37500 & 900.00 & 0.0000 & -3.37500 & 450.00 \\ -3.00000 & 0.00000 & 0.0000 & 3.0400 & 0.00000 & 6.0000 \\ 0.00000 & -0.01687 & -3.3750 & 0.0000 & 4.01687 & -3.3750 \\ 0.00000 & 3.37500 & 450.00 & 6.0000 & -3.37500 & 2100.0 \end{bmatrix} \begin{bmatrix} d_{xB} \\ d_{yB} \\ \theta_B \\ d_{xC} \\ d_{yC} \\ \theta_C \end{bmatrix} = \begin{bmatrix} 1600 \\ 0.00 \\ 0.00 \\ 0.00 \\ 0.00 \\ 0.00 \end{bmatrix}$$

donde la matriz coeficiente es conocida como la matriz de nodos,  $d_B = [d_{xB} \ d_{yB} \ \theta_B]^T$  y  $d_C = [d_{xC} \ d_{yC} \ \theta_C]^T$  son los vectores de desplazamiento de los nodos B y C, respectivamente. Resuelva dicho sistema.

**Solución** La solución obtenida con el PROGRAMA 3.2 del CD o con Matlab, se da a continuación

$$\begin{aligned} d_{xB} &= 0.47185; & d_{yB} &= 0.00259; & \theta_B &= -0.00125 \\ d_{xC} &= 0.46776; & d_{yC} &= -0.00194; & \theta_C &= -0.00108 \end{aligned}$$



**3.3** Determine las concentraciones molares de una mezcla de cinco componentes en solución a partir de los siguientes datos espectrofotométricos.

| Longitud de onda<br><i>i</i> | Absorbancia molar del componente<br><i>j</i> |     |    |     |      | Absorbancia total observada |
|------------------------------|----------------------------------------------|-----|----|-----|------|-----------------------------|
|                              | 1                                            | 2   | 3  | 4   | 5    |                             |
| 1                            | 98                                           | 9   | 2  | 1   | 0.5  | 0.1100                      |
| 2                            | 11                                           | 118 | 9  | 4   | 0.88 | 0.2235                      |
| 3                            | 27                                           | 27  | 85 | 8   | 2    | 0.2800                      |
| 4                            | 1                                            | 3   | 17 | 142 | 25   | 0.3000                      |
| 5                            | 2                                            | 4   | 7  | 17  | 118  | 0.1400                      |

Asúmase que la longitud de la trayectoria óptica es unitaria y que el solvente no absorbe a estas longitudes de onda.

**Solución** Si se considera que se cumple la ley de Beer, entonces a una longitud de onda dada,  $i$

$$A_{TOT \ i} = \sum_{j=1}^5 \epsilon_{ij} C_j$$

\* El nodo 7, por ejemplo, tiene como vecinos a los nodos 3, 6, 8 y 11.

\*\* Carlos Magdaleno. *Análisis matricial de estructuras reticulares*. Edición mimeográfica. ESIA, IPN.



donde:

$A_{\text{TOT}i}$  es la absorbancia total observada a la longitud de onda  $i$ .

$\epsilon_{ij}$  es la absorbancia molar del componente  $j$  a la longitud de onda  $i$ .

$C_j$  es la concentración molar del componente  $j$  en la mezcla.

Al sustituir los valores de la tabla se obtiene

$$\begin{aligned} 98 C_1 + 9 C_2 + 2 C_3 + C_4 + 0.5 C_5 &= 0.1100 \\ 11 C_1 + 118 C_2 + 9 C_3 + 4 C_4 + 0.88 C_5 &= 0.2235 \\ 27 C_1 + 27 C_2 + 85 C_3 + 8 C_4 + 2 C_5 &= 0.2800 \\ C_1 + 3 C_2 + 17 C_3 + 142 C_4 + 25 C_5 &= 0.3000 \\ 2 C_1 + 4 C_2 + 7 C_3 + 17 C_4 + 118 C_5 &= 0.1400 \end{aligned}$$

Un sistema de ecuaciones lineales con matriz coeficiente dominante. Esto sugiere resolver el sistema con el método de Gauss-Seidel.

El PROGRAMA 3.3 del CD utiliza el método de Gauss-Seidel para resolver un sistema de ecuaciones lineales. Este programa se utilizó con el vector cero como vector inicial, por la relativa cercanía de cero con cada uno de los valores del lado derecho del sistema. Los resultados obtenidos son

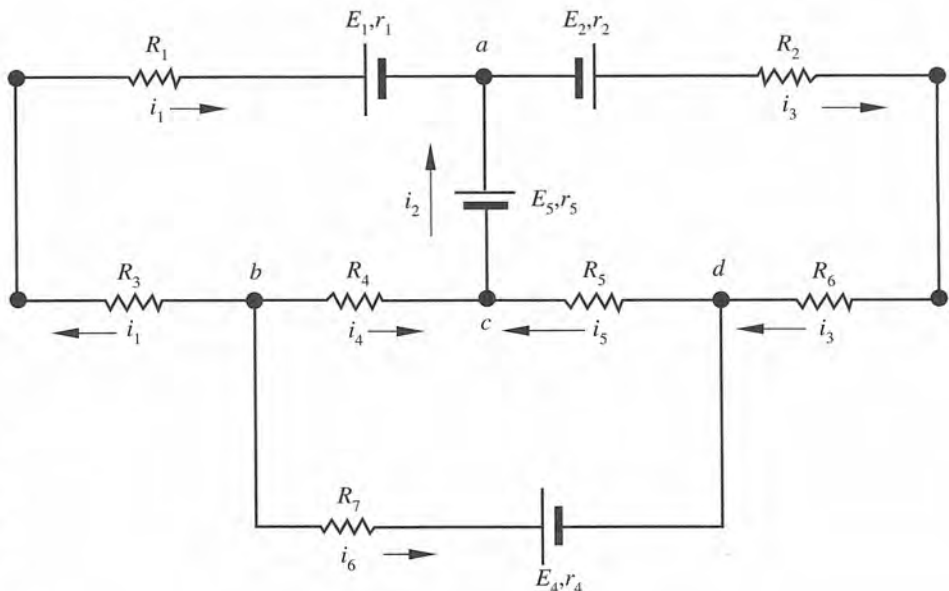
$$\begin{aligned} C_1 &= 0.000910 & C_2 &= 0.001569 & C_3 &= 0.002333 \\ C_4 &= 0.001664 & C_5 &= 0.000740 \end{aligned}$$



**3.4** Determine la intensidad de corriente en cada rama del circuito que se muestra en la figura 3.15.

### Solución

Se asigna un sentido y una letra a cada magnitud desconocida; los sentidos supuestos son enteramente arbitrarios. Hay que observar que la intensidad de corriente en  $R_3$ ,  $R_1$  y  $E_1$  es la misma y, por consiguiente, sólo se requiere una letra. Lo mismo ocurre para la intensidad de corriente en  $R_2$ ,  $E_2$  y  $R_6$ . Los nodos (puntos de la red en los cuales se unen tres o más conductores) se designan con las letras a, b, c, d.



**Figura 3.15**

Circuito eléctrico con resistencias y fuentes de poder.

Aplicación de la **regla de los nodos de Kirchhoff** a tres nodos cualesquiera

| Nodo | $\Sigma i = 0$         |
|------|------------------------|
| a    | $i_1 + i_2 - i_3 = 0$  |
| b    | $-i_1 - i_4 - i_6 = 0$ |
| c    | $i_4 + i_5 - i_2 = 0$  |

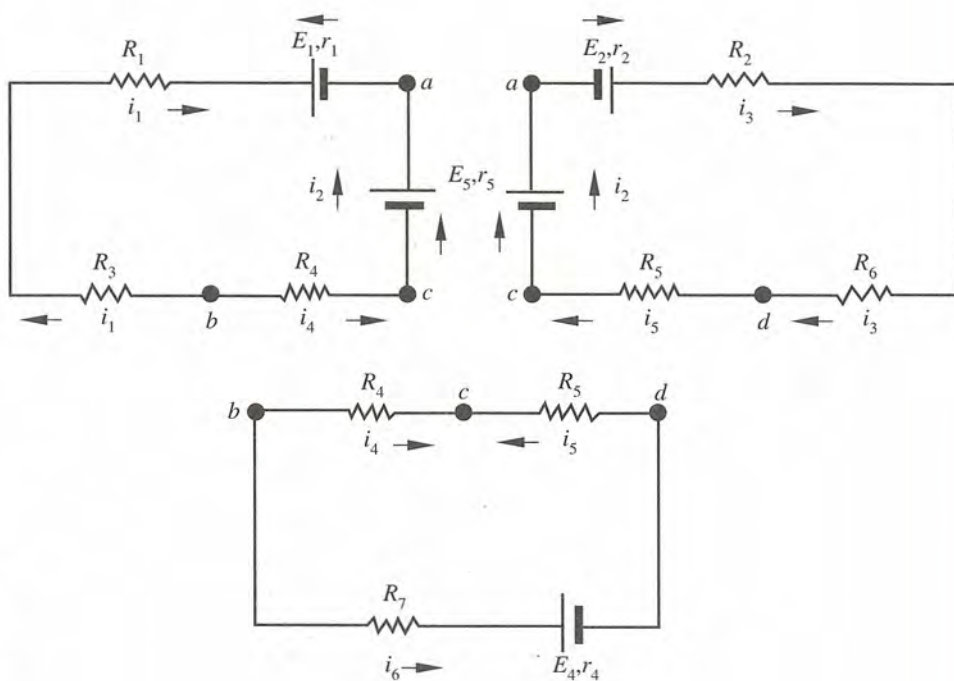
Si bien es cierto que hay un nodo más, el *d*, la aplicación de la regla daría una ecuación linealmente dependiente de las otras tres, esto es:

$$\text{Nodo d} \quad i_6 + i_3 - i_5 = 0,$$

ecuación que se obtiene sumando las tres primeras; por ello resulta redundante y en general se aplica dicha regla a  $n-1$  nodos solamente.

En la figura 3.16 se representa el circuito cortado en mallas. Considérese en cada malla como positivo el sentido de las agujas del reloj. La regla de las mallas de Kirchhoff ( $\Sigma E_k = \Sigma i_k R_k$ ) proporciona las siguientes ecuaciones:

| Malla | $\Sigma E_k = \Sigma i_k R_k$                                  |
|-------|----------------------------------------------------------------|
| I     | $-E_1 - E_5 = i_1 R_1 + i_1 r_1 - i_2 r_5 - i_4 R_4 + i_1 R_3$ |
| II    | $E_2 + E_5 = i_3 r_2 + i_3 R_2 + i_3 R_6 + i_5 R_5 + i_2 r_5$  |
| III   | $E_4 = i_4 R_4 - i_5 R_5 - i_6 r_4 - i_6 R_7$                  |



**Figura 3.16**  
Circuito de la  
figura 3.15  
cortado en  
mallas.

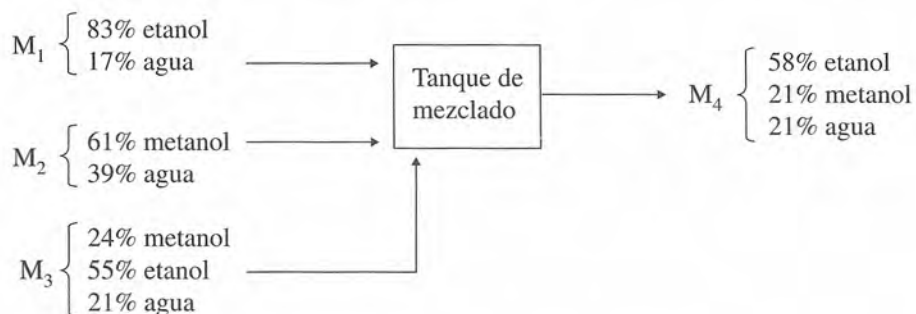
Se tienen ecuaciones independientes, donde conocidas las  $R_k$ , las  $E_k$  y las  $r_k$ , se pueden calcular las seis intensidades de corriente resolviendo el sistema. Para los siguientes datos, calcule las intensidades de corriente.

| $k$ | $E_k$ (volts) | $r_k$ ( $\Omega$ ) | $R_k$ ( $\Omega$ ) |
|-----|---------------|--------------------|--------------------|
| 1   | 12            | 0.1                | 25                 |
| 2   | 10            | 0.5                | 40                 |
| 3   | —             | —                  | 16                 |
| 4   | 12            | 0.5                | 20                 |
| 5   | 24            | 0.2                | 9                  |
| 6   | —             | —                  | 4                  |
| 7   | —             | —                  | 20                 |

Con el **PROGRAMA 3.2** del CD se obtienen los siguientes valores para las intensidades de corriente

| $k$   | 1        | 2      | 3      | 4       | 5       | 6        |
|-------|----------|--------|--------|---------|---------|----------|
| $i_k$ | -0.53811 | 1.1934 | 0.6553 | 0.68226 | 0.51115 | -0.14415 |

- 3.5 Con los datos del diagrama siguiente (donde los porcentajes están dados en peso), encuentre posibles valores de las corrientes  $M_1$ ,  $M_2$ ,  $M_3$  y  $M_4$ .



**Solución** Mediante balances de materia por componente y global, se tiene

| Componente | Balance de materia |   |            |                             |
|------------|--------------------|---|------------|-----------------------------|
| Etanol     | $0.83 M_1$         | + | $0.55 M_3$ | $- 0.58 M_4 = 0$            |
| Metanol    |                    |   | $0.61 M_2$ | $+ 0.24 M_3 - 0.21 M_4 = 0$ |
| Agua       | $0.17 M_1$         | + | $0.39 M_2$ | $+ 0.21 M_3 - 0.21 M_4 = 0$ |
| Global     | $M_1$              | + | $M_2$      | $+ M_3 - M_4 = 0$           |

Hay que observar que sólo se tienen tres ecuaciones linealmente independientes, pues la ecuación del balance global de materia es la suma de las otras tres. Por ser el sistema homogéneo es consistente, y como el rango de la matriz coeficiente es menor que el número de incógnitas, el sistema tiene un número infinito de soluciones. Fijando una base de cálculo, por ejemplo  $M_4 = 100\text{Kg}$ , se obtiene el sistema:



$$\begin{aligned}
 0.83 M_1 & \quad \quad \quad + 0.55 M_3 = 58 \\
 & \quad \quad \quad 0.61 M_2 + 0.24 M_3 = 21 \\
 0.17 M_1 + 0.39 M_2 & + 0.21 M_3 = 21
 \end{aligned}$$

cuya solución se deja al lector, utilizando alguno de los programas vistos.



- 3.6** Un granjero desea preparar una fórmula alimenticia para engordar ganado. Dispone de maíz, desperdicios, alfalfa y cebada, cada uno con ciertas unidades de ingredientes nutritivos, de acuerdo con la tabla siguiente

UNIDADES DE INGREDIENTES NUTRITIVOS POR kg DE CADA ALIMENTO DISPONIBLE

| Ingrediente nutritivo | ALIMENTO |             |         |        | Requerimiento diario<br>Unidades / kg |
|-----------------------|----------|-------------|---------|--------|---------------------------------------|
|                       | Maíz     | Desperdicio | Alfalfa | Cebada |                                       |
| Carbohidrato          | 80       | 15          | 35      | 60     | 230                                   |
| Proteína              | 28       | 72          | 57      | 25     | 180                                   |
| Vitaminas             | 20       | 20          | 12      | 20     | 80                                    |
| Celulosa              | 50       | 10          | 20      | 60     | 160                                   |
| Costo \$              | 18       | 5           | 7       | 20     | —                                     |

- a) Determine los kilogramos necesarios de cada material para satisfacer el requerimiento diario (presentado en la última columna).  
b) Determine el costo de la mezcla

**Nota:** La fórmula alimenticia debe contener los cuatro alimentos.

**Solución** Si se llama  $x_1$  a los kg de maíz necesarios,  $x_2$  los de desperdicio, ..., se tiene

$$\begin{aligned}
 80 x_1 + 15 x_2 + 35 x_3 + 60 x_4 &= 230 \\
 28 x_1 + 72 x_2 + 57 x_3 + 25 x_4 &= 180 \\
 20 x_1 + 20 x_2 + 12 x_3 + 20 x_4 &= 80 \\
 50 x_1 + 10 x_2 + 20 x_3 + 60 x_4 &= 160
 \end{aligned}$$

Con el **PROGRAMA 3.2** del CD o con Matlab se obtiene:

$$x_1 = 1.8524, x_2 = 1.0318, x_3 = 0.6178, x_4 = 0.745$$

De donde el costo de la mezcla es:

$$\text{Costo} = 18 \cdot 1.8524 + 5 \cdot 1.03 + 7 \cdot 0.61 + 20 \cdot 0.745 = \$ 57.66$$

- 3.7** En un sistema monofásico en equilibrio químico existen los siguientes compuestos: CO, H<sub>2</sub>, CH<sub>3</sub>OH, H<sub>2</sub>O y C<sub>2</sub>H<sub>6</sub>. Calcule el número de reacciones químicas independientes.

**Solución**



Se establece la matriz atómica listando los compuestos como cabezas de columna y los átomos como inicio de filas, de tal modo que la intersección muestre el número de átomos del compuesto correspondiente.

| Compuesto<br>Átomo | CO | H <sub>2</sub> | CH <sub>3</sub> OH | H <sub>2</sub> O | C <sub>2</sub> H <sub>6</sub> |
|--------------------|----|----------------|--------------------|------------------|-------------------------------|
| C                  | 1  | 0              | 1                  | 0                | 2                             |
| H                  | 0  | 2              | 4                  | 2                | 6                             |
| O                  | 1  | 0              | 1                  | 1                | 0                             |

Si N es el número de compuestos en equilibrio químico, R el número de reacciones independientes, se tiene la siguiente relación discutida por Jouguet, Brinkey y otros\*

$$R = N - C$$

donde C es el rango de la matriz atómica.

Para encontrar el rango se utilizará el método de ortogonalización de Gram-Schmidt, aplicado a las columnas de la matriz atómica. Para esto, llámense  $x_1, x_2, \dots, x_5$  las columnas CO, H<sub>2</sub>, ..., C<sub>2</sub>H<sub>6</sub>.

$$e_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix},$$

$$e_2 = x_2 - \alpha_{1,2} e_1, \text{ donde } \alpha_{1,2} = \frac{x_2 \cdot e_1}{e_1 \cdot e_1} = \frac{\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}} = 0$$

Por tanto:

$$e_2 = x_2 = \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}$$

Nótese que como  $x_2$  es ortogonal a  $x_1$ , el proceso da  $e_2 = x_2$ .

$$e_3 = x_3 - \alpha_{1,3} e_1 - \alpha_{2,3} e_2, \text{ donde } \alpha_{1,3} = \frac{x_3 \cdot e_1}{e_1 \cdot e_1} = \frac{\begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}} = 1$$

y

$$\alpha_{2,3} = \frac{x_3 \cdot e_2}{e_2 \cdot e_2} = \frac{\begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}} = 2$$

\* Jouguet, J. Ec. Polyt. París, 2, 62 (1921); Prigogine and Defay. J. Chem. Phys 15, 614 (1947).

Por tanto

$$\mathbf{e}_3 = \begin{bmatrix} 1 \\ 4 \\ 1 \end{bmatrix} - (1) \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - (2) \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Esto indica que  $\mathbf{x}_3$  es linealmente dependiente de  $\mathbf{x}_1$  y  $\mathbf{x}_2$ . Continuando el proceso de ortogonalización, pero sin tomar en cuenta a  $\mathbf{e}_3$ , se tiene:

$$\mathbf{e}_4 = \mathbf{x}_4 - \alpha_{1,4} \mathbf{e}_1 - \alpha_{2,4} \mathbf{e}_2, \text{ donde } \alpha_{1,4} = \frac{\mathbf{x}_4 \cdot \mathbf{e}_1}{\mathbf{e}_1 \cdot \mathbf{e}_1} = \frac{\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}} = \frac{1}{2}$$

$$\text{y } \alpha_{2,4} = \frac{\mathbf{x}_4 \cdot \mathbf{e}_2}{\mathbf{e}_2 \cdot \mathbf{e}_2} = \frac{\begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}}{\begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix}} = 1$$

Por tanto

$$\mathbf{e}_4 = \begin{bmatrix} 0 \\ 2 \\ 1 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} - (1) \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/2 \\ 0 \\ 1/2 \end{bmatrix}$$

Como el número de filas de la matriz atómica es 3, el máximo número de vectores linealmente independientes es 3 y como ya se ha encontrado que  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  y  $\mathbf{x}_4$  son linealmente independientes,  $\mathbf{x}_5$  es necesariamente dependiente de  $\mathbf{x}_1$ ,  $\mathbf{x}_2$  y  $\mathbf{x}_4$  y  $\mathbf{e}_5$  debe ser el vector cero (demostración que se deja al lector como ejercicio); entonces, el rango de la matriz atómica es 3.

Al aplicar la fórmula

$$R = N - C = 5 - 3 = 2$$

se tiene que el número de reacciones independientes para llegar al sistema en equilibrio químico mencionado es 2.

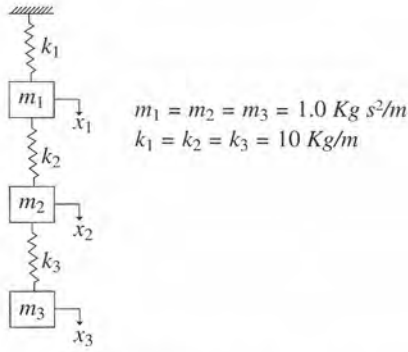
Los cálculos pueden hacerse con Matlab usando el guión del ejemplo 3.23.

- 3.8** Analicemos las características de la vibración libre no amortiguada del sistema de tres grados de libertad mostrado en la figura 3.17. El sistema consta de tres masas  $m_1$ ,  $m_2$  y  $m_3$ , conectadas mediante los tres resortes mostrados, siendo sus constantes elásticas  $k_1$ ,  $k_2$  y  $k_3$ . Los desplazamientos de las masas se definen mediante las coordenadas generalizadas  $x_1$ ,  $x_2$  y  $x_3$ , respectivamente, estando medido cada desplazamiento a partir de la posición de equilibrio estático de la masa respectiva.

Utilizando ya sea las ecuaciones de Lagrange o bien la segunda ley de Newton, se encuentra que las ecuaciones diferenciales de movimiento del sistema son:

$$\begin{aligned} m_1 x_1'' + (k_1 + k_2)x_1 - k_2 x_2 &= 0 \\ m_2 x_2'' - k_2 x_1 + (k_2 + k_3)x_2 - k_3 x_3 &= 0 \\ m_3 x_3'' - k_3 x_2 + k_3 x_3 &= 0 \end{aligned} \quad (1)$$





**Figura 3.17**  
Sistema de tres  
grados de  
libertad.

Sabemos, de la teoría de las vibraciones, que la solución del sistema de ecuaciones (1) se puede escribir en la forma:

$$\begin{aligned} x_1 &= X_1 \text{ sen } pt \\ x_2 &= X_2 \text{ sen } pt \\ x_3 &= X_3 \text{ sen } pt \end{aligned} \quad (2)$$

En donde  $X_1$ ,  $X_2$  y  $X_3$  son las *amplitudes* del movimiento de las masas respectivas, y  $p$  denota las frecuencias circulares naturales que corresponden a los modos principales de vibración del sistema.

Sustituyendo la ecuación 2 y las derivadas correspondientes a esas expresiones en la ecuación 1, y utilizando los valores de masas y de constantes elásticas mostrados en la figura 3.17, obtenemos el siguiente conjunto de ecuaciones algebraicas homogéneas:

$$\begin{aligned} (20 - p^2)X_1 - 10X_2 &= 0 \\ -10X_1 + (20 - p^2)X_2 - 10X_3 &= 0 \\ -10X_2 + (10 - p^2)X_3 &= 0 \end{aligned} \quad (3)$$

Para obtener una solución distinta de la trivial de la ecuación 3, el determinante de la matriz coeficiente del sistema, debe ser igual a cero, de manera que

$$\begin{vmatrix} (20 - p^2) & -10 & 0 \\ -10 & (20 - p^2) & -10 \\ 0 & -10 & (10 - p^2) \end{vmatrix} = 0 \quad (4)$$

El desarrollo de este determinante resulta en el polinomio característico

$$p^6 - 50p^4 + 600p^2 - 1000 = 0 \quad (5)$$

que se puede escribir como ecuación cúbica en  $p^2$ , de la forma

$$(p^2)^3 - 50(p^2)^2 + 600p^2 - 1000 = 0 \quad (6)$$

Se encuentra que las raíces de la ecuación 6 son:

$$\begin{aligned} p_1^2 &= 1.98 \text{ seg}^{-2} \\ p_2^2 &= 1.98 \text{ seg}^{-1} \\ p_3^2 &= 32.5 \text{ seg}^{-2} \end{aligned}$$

Estos valores característicos son los cuadrados de las frecuencias circulares del primero, segundo y tercer modos de vibración del sistema, respectivamente.

Como la ecuación 3 constituye un conjunto homogéneo de ecuaciones simultáneas, no se puede obtener un conjunto único de valores para  $X_1$ ,  $X_2$  y  $X_3$ . Sin embargo, se pueden

determinar varias relaciones para las amplitudes, que proporcionarán la configuración del sistema para los diferentes modos de vibración cuando se define una amplitud unitaria para cualquiera de las masas. Por ejemplo, sustituyendo  $p_1^2 = 1.98$  en la ecuación 3 se obtiene la siguiente configuración para el *primer* modo

$$\left. \begin{array}{l} X_2 = 1.80 X_1 \\ X_3 = 2.25 X_1 \end{array} \right\} \text{ primer modo} \quad (7)$$

En forma similar, las configuraciones del segundo y tercer modos, utilizando  $p_2^2 = 15.5$  y  $p_3^2 = 32.5$ , respectivamente, son

$$\left. \begin{array}{l} X_2 = 0.45 X_1 \\ X_3 = -0.80 X_1 \end{array} \right\} \text{ segundo modo} \quad (8)$$

$$\left. \begin{array}{l} X_2 = -1.25 X_1 \\ X_3 = 0.555 X_1 \end{array} \right\} \text{ tercer modo} \quad (9)$$

Se puede ver en las tres últimas ecuaciones que si la amplitud de cualquiera de las masas se conoce o se supone para un modo particular de vibración, se puede determinar la configuración del sistema para ese modo. Como las ecuaciones 7 a 9 consisten en relaciones de amplitudes  $X_i$ , la substitución de la ecuación 2 en estas ecuaciones indica que las relaciones mostradas son también las relaciones de los *desplazamientos*. Por ejemplo, cuando  $m_1$  tiene un desplazamiento de 1 cm y el sistema está vibrando en el segundo modo, los desplazamientos correspondientes de  $m_2$  y  $m_3$  serán 0.45 cm y 0.80 cm, respectivamente, y el movimiento de  $m_3$  estará  $180^\circ$  fuera de fase con el de  $m_1$ . Se puede agregar aquí que la configuración de un sistema, dada por las relaciones mostradas arriba, define también los desplazamientos iniciales que se tendrían que dar a las masas para que el sistema vibrara en el modo asociado con esa configuración, sin que estuvieran presentes otros armónicos como cuando el sistema se suelta a partir del reposo.

## Problemas

- 3.1 Elabore un algoritmo general para sumar y restar matrices.
- 3.2 Con el algoritmo del problema anterior, elabore uno de propósito general para sumar y restar matrices.
- 3.3 Demuestre, partiendo de la definición del producto de una matriz por un escalar, las ecuaciones 3.7, 3.8 y 3.10.
- 3.4 Demuestre la ecuación 3.12, utilizando la definición de multiplicación de matrices.
- 3.5 Con el programa 3.1 del disco multiplique las siguientes matrices

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 3 & 4 & 5 \\ 7 & 8 & 9 \\ 11 & 12 & 13 \\ 15 & 16 & 17 \end{bmatrix},$$

$$\begin{bmatrix} 2 & 3 & 4 & 5 \\ 0 & 6 & 7 & 8 \\ 0 & 0 & 5 & 3 \\ 0 & 0 & 0 & 4 \end{bmatrix} \quad \begin{bmatrix} 4 & 3 & 0 & 1 \\ 5 & 0.2 & -1 & 8 \\ 3 & 4 & 5 & 7 \\ 10 & 9 & 3 & -2 \end{bmatrix},$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix},$$

$$[0 \ 1 \ 2 \ 3 \ 4] \quad \begin{bmatrix} 3 \\ 8 \\ -2 \\ 5 \\ 1 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{bmatrix} \quad [3 \ 8 \ -2 \ 5 \ 1]$$

$$\begin{bmatrix} 3 & 4 & 5 \\ 7 & 8 & 9 \\ 11 & 12 & 13 \\ 15 & 16 & 17 \end{bmatrix} \quad \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- 3.6 La siguiente tabla representa las existencias en bodega de una agencia de refacciones para automóviles.

| Refacción | MARCA |    |    |    |    |    |
|-----------|-------|----|----|----|----|----|
|           | M1    | M2 | M3 | M4 | M5 | M6 |
| R1        | 5     | 13 | 23 | 8  | 15 | 98 |
| R2        | 16    | 45 | 11 | 54 | 10 | 86 |
| R3        | 34    | 22 | 77 | 21 | 65 | 2  |
| R4        | 21    | 19 | 83 | 2  | 16 | 37 |
| R5        | 8     | 97 | 69 | 27 | 14 | 3  |

En la siguiente tabla se dan los precios unitarios correspondientes a las refacciones de arriba.

| Refacción | MARCA |       |       |       |       |       |
|-----------|-------|-------|-------|-------|-------|-------|
|           | M1    | M2    | M3    | M4    | M5    | M6    |
| R1        | 65000 | 73450 | 82500 | 71245 | 62350 | 76450 |
| R2        | 3400  | 3560  | 2560  | 5790  | 4700  | 5000  |
| R3        | 12500 | 13450 | 16400 | 15600 | 11650 | 9500  |
| R4        | 895   | 940   | 780   | 950   | 645   | 1000  |
| R5        | 5350  | 7620  | 6700  | 3250  | 5890  | 7000  |

Determine la inversión en bodega de la agencia.

- 3.7 Responda las siguientes preguntas.

- ¿Una matriz no cuadrada puede ser simétrica?
- ¿Una matriz diagonal es triangular superior, triangular inferior o ambas?
- ¿Una matriz diagonal tiene inversa con uno de sus elementos de la diagonal principal igual a cero?



- 3.8** Multiplique una matriz permutadora (seleccione una cualquiera) por sí misma y observe el resultado. Generalice dicho resultado.
- 3.9** Demuestre las ecuaciones 3.15, 3.16 y 3.17.
- 3.10** Obtenga la ecuación 3.21 a partir de la ley de los cosenos.
- 3.11** Elabore un algoritmo tal que, dados dos vectores de igual número de componentes, se determine e imprima la norma euclídeana de estos vectores, su producto punto, el ángulo que guardan entre ellos y la distancia que hay entre ambos.
- 3.12** Codifique el algoritmo del problema 3.11 y verifique este programa con las siguientes parejas de vectores

$$\begin{aligned}
 a) \quad & \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad b) \quad \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \\ 5 \\ -2 \end{bmatrix} \quad c) \quad \begin{bmatrix} 5 \\ -2 \\ 8 \\ 0.1 \end{bmatrix}, \begin{bmatrix} 1.5 \\ -0.6 \\ 2.4 \\ 0.03 \end{bmatrix} \quad d) \quad \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \end{bmatrix}
 \end{aligned}$$

- 3.13** El teorema 3.1 puede y debe emplearse también para ortogonalizar un conjunto de  $m$  vectores linealmente independientes de  $n$  componentes cada uno, con  $m < n$ . Por otro lado, demuestre con el teorema mencionado, que cualquier conjunto de  $n$  vectores linealmente independientes con  $n$  componentes cada uno da como resultado un conjunto linealmente dependiente al adicionársele un vector  $\mathbf{x}_{n+1}$  de  $n$  componentes.

**NOTA:** Use como motivación algunos casos particulares sencillos: por ejemplo, a un conjunto particular de dos vectores linealmente independientes con dos componentes cada uno, añada un tercer vector y aplique la ortogonalización al conjunto resultante.

- 3.14** Elabore una subrutina de propósito general para ortogonalizar un conjunto de  $m$  vectores linealmente independientes de  $n$  componentes cada uno ( $m < n$ ) con el método de Gram-Schmidt.

**NOTA:** Puede usar el algoritmo 3.2 como base.

- 3.15** Con la subrutina del problema 3.14 ortogonalice los siguientes conjuntos de vectores.

$$a) \quad \mathbf{x}_1 = \begin{bmatrix} 1 \\ -2 \\ 5 \\ 7 \\ 8 \\ 0.3 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} -2 \\ 1 \\ 7 \\ 3 \\ 12 \\ 0 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 3 \\ 0.8 \\ 4 \\ 15 \\ 3 \\ 2 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 7 \\ -3 \\ 5 \\ 3.2 \\ 9 \\ 40 \end{bmatrix}, \quad \mathbf{x}_5 = \begin{bmatrix} 5 \\ 4 \\ 3 \\ 1 \\ 7 \\ 8 \end{bmatrix},$$

$$b) \quad \mathbf{x}_1 = \begin{bmatrix} 4 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} -9 \\ -4 \\ -1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 2 \\ 6 \\ 3 \end{bmatrix}$$

$$c) \quad \mathbf{x}_1 = \begin{bmatrix} 10 \\ -20 \\ 5 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 1 \\ 3 \\ 3 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} -5 \\ 20 \\ 5 \end{bmatrix}$$

d)

$$\mathbf{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \\ 2 \end{bmatrix}, \quad \mathbf{x}_2 = \begin{bmatrix} 3 \\ 9 \\ 1 \\ 1 \end{bmatrix}, \quad \mathbf{x}_3 = \begin{bmatrix} 5 \\ 8 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{x}_4 = \begin{bmatrix} 2 \\ 4 \\ 1 \\ 1 \end{bmatrix}.$$

3.16 Modifique el programa del problema 3.14 de modo que:

- Dado un conjunto cualquiera de  $m$  vectores de  $n$  componentes cada uno ( $m < n$ ), se vayan ortogonalizando los linealmente independientes y se descarte los que resulten linealmente dependientes.
- Imprima el número de vectores linealmente independientes del conjunto denotando este número como **rango** del conjunto.

Corra el programa para determinar el rango de las siguientes matrices o conjuntos de vectores columna

$$\begin{bmatrix} 4 & -9 & 2 \\ 2 & -4 & 6 \\ 1 & -1 & 3 \end{bmatrix}, \quad \begin{bmatrix} 10 & 1 & -5 \\ -20 & 3 & 20 \\ 5 & 3 & 5 \end{bmatrix}, \quad \begin{bmatrix} 10 & 1 & -5 & 1 \\ -20 & 3 & 20 & 2 \\ 5 & 3 & 5 & 6 \end{bmatrix}, \quad \begin{bmatrix} 4 & -1 & 0 & 0 \\ -1 & 4 & -1 & 0 \\ 0 & -1 & 4 & -1 \\ 0 & 0 & -1 & 4 \end{bmatrix}$$

3.17 Calcule el número de reacciones independientes en una reacción de pirólisis, en la cual se encuentran en equilibrio los siguientes compuestos  $\text{O}_2$ ,  $\text{H}_2$ ,  $\text{CO}$ ,  $\text{CO}_2$ ,  $\text{H}_2\text{CO}_3$ ,  $\text{CH}_3\text{OH}$ ,  $\text{C}_2\text{H}_5\text{OH}$ ,  $(\text{CH}_3)_2\text{CO}$ ,  $\text{CH}_4$ ,  $\text{CH}_3\text{CHO}$  y  $\text{H}_2\text{O}$ .

3.18 Dada una matriz  $A$  de orden  $n$ , los términos

- Matriz singular ( $\det A = 0$ )
- Rango  $A < n$
- Los vectores columna o fila de  $A$  son linealmente dependientes

están estrechamente relacionados.

Demuestre que a) implica tanto b) como c).

3.19 ¿La coincidencia del número de incógnitas con el número de ecuaciones en un sistema de ecuaciones lineales implica que éste tiene solución única? Justifique su respuesta.

3.20 Dado el siguiente sistema de ecuaciones, encuentre dos valores de  $w$  que permitan tener solución única y diga qué valores de  $w$  permiten un número infinito de soluciones.

$$\begin{bmatrix} w & 1 & 5 \\ 4 & 2 & -w \\ 0 & 3 & -7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 5 \\ 3 \end{bmatrix}$$

3.21 Si la matriz coeficiente del sistema  $A\mathbf{x} = \mathbf{0}$  es tal que  $\det A = 0$ ; ¿dicho sistema tiene por ese hecho un número infinito de soluciones?

3.22 El método de eliminación de Gauss usualmente hace la transformación conocida como triangularización.

$$\left[ \begin{array}{ccc|c} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \end{array} \right] \rightarrow \left[ \begin{array}{ccc|c} a'_{1,1} & a'_{1,2} & a'_{1,3} & a'_{1,4} \\ 0 & a'_{2,2} & a'_{2,3} & a'_{2,4} \\ 0 & 0 & a'_{3,3} & a'_{3,4} \end{array} \right]$$

En estas condiciones, una sustitución hacia atrás permite obtener la solución. Las ecuaciones 3.49 y 3.50 constituyen el algoritmo para el caso general.

Encuentre las ecuaciones correspondientes para resolver el sistema  $A \mathbf{x} = \mathbf{b}$ , pero ahora llevando a cabo la transformación.

$$\left[ \begin{array}{ccc|c} a_{1,1} & a_{1,2} & a_{1,3} & a_{1,4} \\ a_{2,1} & a_{2,2} & a_{2,3} & a_{2,4} \\ a_{3,1} & a_{3,2} & a_{3,3} & a_{3,4} \end{array} \right] \rightarrow \left[ \begin{array}{ccc|c} a'_{1,1} & 0 & 0 & a'_{1,4} \\ a'_{2,1} & a'_{2,2} & 0 & a'_{2,4} \\ a'_{3,1} & a'_{3,2} & a_{3,3} & a'_{3,4} \end{array} \right]$$

y posteriormente una sustitución hacia delante.

- 3.23** Modifique el algoritmo 3.4, de modo que una vez encontrado el elemento pivote e intercambiadas las filas (si procede), se divida la fila pivote entre el elemento pivote. En el caso de un sistema de orden 3, el resultado en la triangularización sería

$$\left[ \begin{array}{ccc|c} 1 & a_{1,2} & a_{1,3} & b_1 \\ 0 & 1 & a_{2,3} & b_2 \\ 0 & 0 & 1 & b_3 \end{array} \right],$$

y, por tanto, en la sustitución regresiva no se tendría que dividir entre los coeficientes de las incógnitas.

Por otro lado, para el cálculo del determinante deben guardarse los pivotes para su empleo en la expresión

$$\det A = (-1)^r \prod_{i=1}^n a_{i,i}$$

**SUGERENCIA:** Vea el **PROGRAMA 3.1** del CD.

- 3.24** Elabore un algoritmo para resolver un sistema de ecuaciones  $A \mathbf{x} = \mathbf{b}$  usando la eliminación de Jordan.
- 3.25** Calcule el número de multiplicaciones, divisiones o ambas y la cantidad de sumas, restas o ambas que se requieren para resolver un sistema tridiagonal por el método de Thomas. Determine también las necesidades de memoria para este algoritmo.
-  **3.26** Utilice el subprograma que se da en el ejercicio 8.2 del capítulo 8 para resolver los siguientes sistemas

$$\begin{array}{rcll} a) & 0.5 x_1 + & 0.25 x_2 & = 0.32 \\ & 0.3 x_1 + & 0.8 x_2 + & 0.4 x_3 = 0.77 \\ & & 0.2 x_2 + & x_3 + 0.6 x_4 = -0.6 \\ & & & x_3 - 3 x_4 = -2 \end{array}$$

$$\begin{array}{rcll} b) & x_1 - & x_2 & = 1 \\ & 2 x_1 - & x_2 + 3 x_3 & = 8 \\ & & x_2 + & x_3 = 4 \end{array}$$

$$\begin{array}{rcll} c) & 4 x_1 + & x_2 & = -1 \\ & -8 x_1 - & x_2 + & x_3 = 13 \\ & & 3 x_2 - 2 x_3 + & 4 x_4 = -3 \\ & & & x_3 - x_4 + & x_5 = 2.1 \\ & & & 2 x_4 + & 6 x_5 = 3.4 \end{array}$$

- 3.27** Una matriz tridiagonal por bloques (o partida) es una matriz de la forma



$$A = \begin{bmatrix} B_1 & C_1 & O & & & O \\ A_2 & B_2 & C_2 & O & & \\ O & A_3 & B_3 & C_3 & O & \\ O & O & & & & O \\ O & & & & & C_{n-1} \\ O & & & & O & A_n & B_n \end{bmatrix}$$

donde  $B_1, B_2, \dots, B_n$  son matrices de orden  $n_1, n_2, \dots, n_n$ , respectivamente.  $A_2, A_3, \dots, A_n$  son matrices de orden  $(n_2 \times n_1), (n_3 \times n_2), \dots, (n_n \times n_{n-1})$ , respectivamente, y  $C_1, C_2, \dots, C_{n-1}$  son matrices de orden  $(n_1 \times n_2), (n_2 \times n_3), \dots, (n_{n-1} \times n_n)$ , respectivamente.

Por ejemplo, las matrices

a)

$$A = \begin{bmatrix} B_1 & C_1 & O \\ A_2 & B_2 & C_2 \\ O & A_3 & B_3 \end{bmatrix} \quad \text{donde } B_i = \begin{bmatrix} 6 & -1 & 0 \\ -1 & 6 & -1 \\ 0 & -1 & 6 \end{bmatrix} \quad i = 1, 2, 3$$

$$\text{y } A_{i+1} = C_i = \begin{bmatrix} -2 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -2 \end{bmatrix} \quad i = 1, 2$$

b)

$$\begin{bmatrix} 1 & 5 & 3 & 5 & 8 & 9 & -2 & 0 & 0 & 0 & 0 & 0 \\ 2 & -1 & 0 & 1 & 4 & 0 & 7 & 0 & 0 & 0 & 0 & 0 \\ 4 & 3 & 6 & 7 & 3 & 2 & 3 & 0 & 0 & 0 & 0 & 0 \\ 7 & 3 & 6 & 4 & 5 & 8 & 9 & 4 & 5 & 5 & 4 & 3 \\ 2 & 2 & 5 & 7 & 6 & 3 & 2 & 2 & 7 & 8 & 9 & 1 \\ 3 & 7 & 3 & 4 & 1 & 0 & 1 & 0 & -3 & 5 & 7 & 2 \\ 1 & 1 & 2 & 4 & 3 & 2 & 5 & 4 & 5 & 7 & 9 & 5 \\ 0 & 0 & 0 & 5 & 7 & 9 & 5 & 0 & 5 & 7 & 4 & 2 \\ 0 & 0 & 0 & 4 & 8 & 2 & 2 & -1 & 7 & 9 & 7 & 8 \\ 0 & 0 & 0 & 3 & 2 & 1 & 1 & 4 & 8 & 4 & 3 & 2 \\ 0 & 0 & 0 & 5 & 1 & 5 & 4 & 2 & 7 & 4 & 5 & -1 \\ 0 & 0 & 0 & 2 & 9 & 7 & 3 & 3 & 2 & 7 & 2 & 2 \end{bmatrix}$$

son tridiagonales por bloques.

Hay que observar que una matriz tridigonal por bloques no es tridiagonal en el sentido de la definición original.

Elabore un algoritmo similar al algoritmo 3.5 para resolver sistemas tridiagonales por bloques  $A \mathbf{x} = \mathbf{b}$ .

SUGERENCIA: Para el sistema

$$\begin{bmatrix} B_1 & C_1 & O \\ A_2 & B_2 & C_2 \\ O & A_3 & B_3 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix}$$

donde se ha segmentado a  $\mathbf{x}$  y  $\mathbf{b}$  de modo tal que

$\mathbf{x}_1$  y  $\mathbf{b}_1$  son vectores de  $n_1$  componentes (el orden de  $B_1$ ),  
 $\mathbf{x}_2$  y  $\mathbf{b}_2$  son vectores de  $n_2$  componentes (el orden de  $B_2$ ),  
 $\mathbf{x}_3$  y  $\mathbf{b}_3$  son vectores de  $n_3$  componentes (el orden de  $B_3$ ),

forme la matriz aumentada

$$\begin{bmatrix} B_1 & C_1 & O \\ A_2 & B_2 & C_2 \\ O & A_3 & B_3 \end{bmatrix} \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix},$$

y elimine la matriz  $A_2$  por medio de los elementos de la diagonal principal de  $B_1$ ; posteriormente elimine la matriz  $A_3$  con los elementos diagonales de  $B_2$ . Para iniciar la sustitución regresiva, resuelva el sistema

$$B_3 \mathbf{x}_3 = \mathbf{b}_3,$$

con el resultado resuelva el sistema

$$B_2 \mathbf{x}_2 = \mathbf{b}_2 - C_2 \mathbf{x}_3$$

Finalmente, sustituyendo  $\mathbf{x}_2$ , resuelva

$$B_1 \mathbf{x}_1 = \mathbf{b}_1 - C_1 \mathbf{x}_2$$

Los sistemas pueden resolverse con alguno de los métodos vistos.

### 3.28 Resuelva el sistema tridiagonal por bloques

$$\begin{bmatrix} 6 & -1 & 0 & -2 & 0 & 0 & 0 & 0 & 0 \\ -1 & 6 & -1 & 0 & -2 & 0 & 0 & 0 & 0 \\ 0 & 1 & 6 & 0 & 0 & -2 & 0 & 0 & 0 \\ -2 & 0 & 0 & 6 & -1 & 0 & -2 & 0 & 0 \\ 0 & -2 & 0 & -1 & 6 & -1 & 0 & -2 & 0 \\ 0 & 0 & -2 & 0 & -1 & 6 & 0 & 0 & -2 \\ 0 & 0 & 0 & -2 & 0 & 0 & 6 & -1 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 1 & 6 & -1 \\ 0 & 0 & 0 & 0 & 0 & -2 & 0 & -1 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 3 \\ 1 \\ 0 \\ 1 \\ 3 \\ 2 \\ 3 \end{bmatrix}$$

Utilice la sugerencia del problema 3.27, el algoritmo de ese ejercicio o ambos.

- 3.29** En la simulación de una columna de destilación de NP platos que separa una mezcla de NC componentes, el balance de materia por componente en cada plato, el balance de entalpía en cada plato y la relación de equilibrio líquido-vapor de cada componente en cada plato, resultan en un sistema de  $NP(2NC+1)$  (dos veces el número de componentes más uno multiplicado por el número de platos) ecuaciones algebraicas no lineales. En la aplicación del método de Newton-Raphson para un sistema no lineal es necesario resolver un sistema tridiagonal por bloques de orden  $NP(2NC+1)$  en cada iteración. Para una columna de cinco platos y tres componentes, el sistema tridiagonal por bloques por resolver en cada iteración es:

$$\begin{bmatrix} B_1 & C_1 & & & \\ A_2 & B_2 & C_2 & & \\ & A_3 & B_3 & C_3 & \\ & & A_4 & B_4 & C_4 \\ & & & A_5 & B_5 \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \mathbf{x}_3 \\ \mathbf{x}_4 \\ \mathbf{x}_5 \end{bmatrix} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \\ \mathbf{b}_4 \\ \mathbf{b}_5 \end{bmatrix}$$

donde

$$A_2 = \begin{bmatrix} 0 & 0 & 0 & 12204.1 & 9216.1 & 1262.6 & 1768.8 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 0 & 0 & 0 & 10550.0 & 1187.3 & 1636.7 & 2291.6 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_4 = \begin{bmatrix} 0 & 0 & 0 & 9863.2 & 1529.5 & 2109.1 & 2951.1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$A_5 = \begin{bmatrix} 0 & 0 & 0 & 8341.5 & 2227.6 & 3073.1 & 4294.5 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$B_1 = \begin{bmatrix} 5522.3 & 6518.9 & 7105.4 & 18015.3 & 916.1 & 1262.6 & 1768.8 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ -0.065 & 0.9346 & 0.9346 & 2.93858 & 0.119 & -0.4894 & -0.4894 \\ 0.0643 & -0.9357 & 0.0643 & 0.30762 & -0.033 & 0.1481 & -0.0337 \\ 0.0011 & 0.0011 & -0.9989 & 0.00643 & -0.006 & -0.0006 & 0.0524 \end{bmatrix}$$

$$B_2 = \begin{bmatrix} 5777.5 & 6941.9 & 7659.2 & 28231.1 & 1187.3 & 1636.7 & 2291.6 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ -0.137 & 0.8625 & 0.8625 & 7.2004 & 0.8898 & -1.6296 & -1.6296 \\ 0.1314 & -0.8686 & 0.1314 & 1.6619 & 0.5605 & 0.5605 & -0.2482 \\ 0.0061 & 0.0061 & -0.9989 & 0.0979 & -0.0115 & -0.0115 & 0.2374 \end{bmatrix}$$



$$B_3 = \begin{bmatrix} 6099.7 & 7471.9 & 8357.4 & 27837.5 & 1529.5 & 2109.1 & 2951.1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ -0.2217 & 0.7783 & 0.7783 & 5.6209 & 1.6619 & -1.6521 & -1.6296 \\ 0.1917 & -0.8029 & 0.1971 & 2.1270 & -0.4184 & 0.7336 & -0.2482 \\ 0.0246 & 0.0246 & -0.9754 & 0.3359 & -0.0522 & -0.0522 & 0.3355 \end{bmatrix}$$

$$B_4 = \begin{bmatrix} 6557.3 & 8540.2 & 9778.8 & 18947.5 & 2227.6 & 3073.1 & 4294.5 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ -0.4351 & 0.5649 & 0.5649 & 1.7373 & 2.2062 & -0.8346 & -0.8346 \\ 0.3456 & -0.6544 & 0.3456 & 1.5331 & -0.5106 & 0.6927 & -0.5106 \\ 0.8923 & 0.8923 & -0.9108 & 0.5003 & -0.1322 & -0.1322 & 0.3058 \end{bmatrix}$$

$$B_5 = \begin{bmatrix} 7547.5 & 9801.1 & 11480.6 & 12961.3 & 3065.4 & 4231.4 & 5904.4 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ -0.6827 & 0.3173 & 0.3173 & 0.7585 & 29.335 & -3.9259 & -3.9059 \\ 0.4311 & -0.5689 & 0.4311 & 1.4233 & -5.3074 & 9.3998 & -5.3074 \\ 0.2516 & 0.2516 & -0.7484 & 1.0573 & -3.0980 & -3.0980 & 2.8411 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} -5777.5 & -6941.9 & -7659.2 & -17681.1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

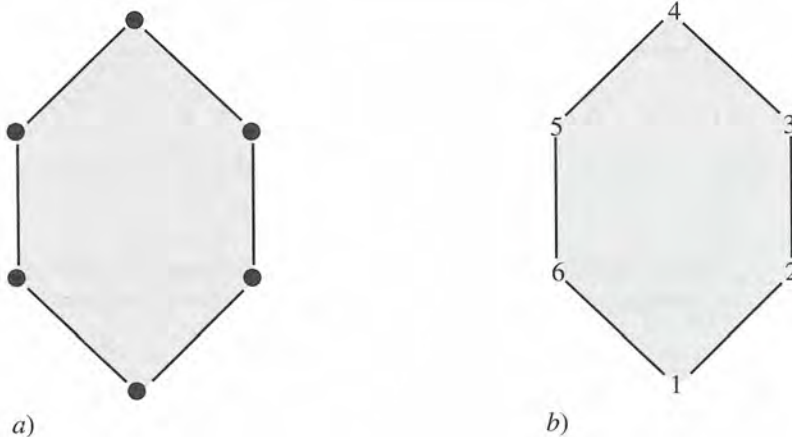
$$C_2 = \begin{bmatrix} -6099.6 & -7471.9 & -8357.4 & -17974.2 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$C_3 = \begin{bmatrix} -6757.4 & -8540.2 & -9778.7 & -10606.0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$C_4 = \begin{bmatrix} -7547.5 & -9801.1 & -11480.6 & -11886.0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$\mathbf{b}_1 = \begin{bmatrix} -3390307.1 \\ -6.7419 \\ 13.4936 \\ 1.2278 \\ -0.009835 \\ 0.000629 \\ 0.000566 \end{bmatrix}, \mathbf{b}_2 = \begin{bmatrix} -117198.1 \\ -70.6904 \\ -16.8926 \\ 0.1614 \\ 0.0142 \\ 0.00463 \\ -0.00256 \end{bmatrix}, \mathbf{b}_3 = \begin{bmatrix} -117288.9 \\ -16.5304 \\ -18.9351 \\ -2.1684 \\ 0.02723 \\ 0.00536 \\ -0.0020 \end{bmatrix}, \mathbf{b}_4 = \begin{bmatrix} -421289.9 \\ -3.9928 \\ -52.9235 \\ 2.2335 \\ -0.0021 \\ -0.00192 \\ 0.12736 \end{bmatrix}, \mathbf{b}_5 = \begin{bmatrix} 348305.0 \\ 59.4815 \\ 35.448 \\ 3.1614 \\ -0.01917 \\ -0.01459 \\ -0.00998 \end{bmatrix}$$

**3.30** Adapte la eliminación de Gauss a la solución del sistema pentadiagonal  $A \mathbf{x} = \mathbf{b}$  ( $A$  es una matriz pentadiagonal) y obtenga las ecuaciones correspondientes a esta adaptación.



**Figura 3.18** a)

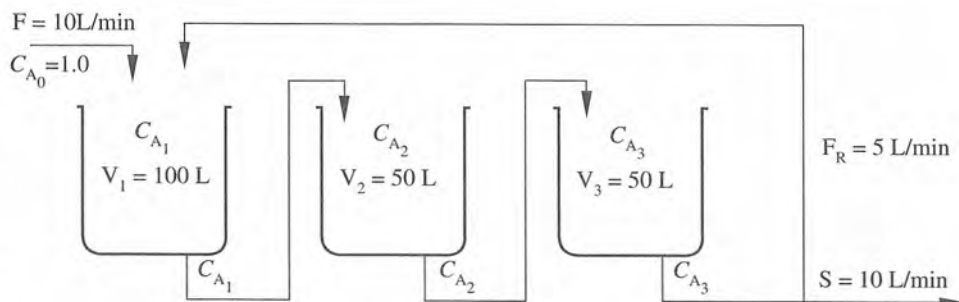
b)

- 3.31** Demuestre que la numeración de los nodos de la figura 3.14 con las consideraciones de que  $a_{ij} \neq 0$  siempre que los nodos sean vecinos, genera una matriz tridiagonal.
- 3.32** Considere la estructura hexagonal de la figura 3.18a (véase Ej. 3.2.). Numere los nodos en la forma mostrada en la figura 3.18b, por ejemplo, y con consideraciones físicas que determinan que  $a_{ij} \neq 0$ , cuando  $i$  y  $j$  son nodos vecinos, determine la matriz  $A$  representativa de dicha estructura.
- 3.33** Se tiene un sistema de tres rectores continuos tipo tanque perfectamente agitado trabajando en serie, en donde se lleva a cabo la reacción  $A \rightarrow \text{Productos}$  y se opera isotérmicamente (véase Fig. 3.19). Los volúmenes se mantienen constantes y son de 100, 50 y 50, litros respectivamente.
- Un balance de materia en cada reactor, de acuerdo con la ecuación de continuidad, conduce al siguiente sistema de ecuaciones:

| Entrada               | – Salida         | – lo que reacciona  | = Acumulación          |
|-----------------------|------------------|---------------------|------------------------|
| $FC_{A0} + FR C_{A3}$ | $-(F+F_R)C_{A1}$ | $-k_1 V_1 C_{A1}^n$ | $= \frac{dC_{A1}}{dt}$ |
| $(F+F_R)C_{A1}$       | $-(F+F_R)C_{A2}$ | $-k_1 V_2 C_{A2}^n$ | $= \frac{dC_{A2}}{dt}$ |
| $(F+F_R)C_{A2}$       | $-(F+F_R)C_{A3}$ | $-k_1 V_3 C_{A3}^n$ | $= \frac{dC_{A3}}{dt}$ |

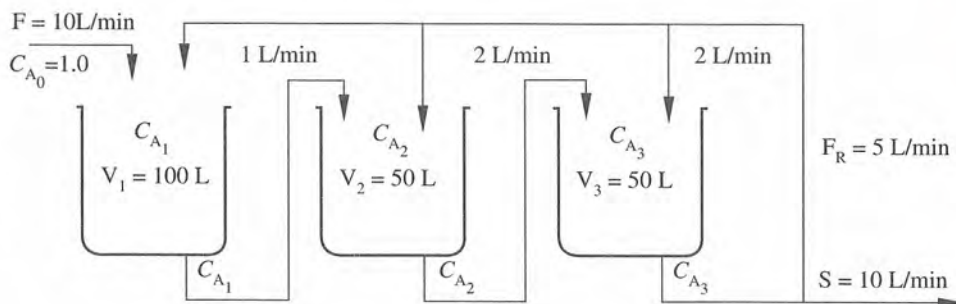
Calcule la concentración de A a régimen permanente en cada reactor si la reacción es de primer orden con respecto a A y la constante de velocidad de reacción  $k_1$  es  $0.1 \text{ min}^{-1}$ . Las composiciones están dadas en  $\text{mol/L}$ .

**Figura 3.19**  
Sistema de tres reactores continuos tipo tanque agitado en donde se lleva a cabo la reacción  $A \rightarrow \text{Productos}$ .



**3.34** Repita el problema 3.33, considerando que el reflujo es como se muestra en la figura 3.20

**Figura 3.20**  
Sistema de tres reactores continuos tipo tanque agitado, en donde se lleva a cabo la reacción  $A \rightarrow \text{Productos}$ .



**3.35** Calcule la composición del benceno en cada plato de la columna de absorción del ejercicio 3.1, si se modifica  $y_0$  a 0.2 de fracción molar. Use las consideraciones del mismo ejercicio.

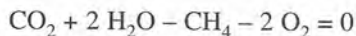
**3.36** Las reacciones químicas pueden escribirse como:

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

donde:  $x_i$  es el coeficiente estequiométrico del compuesto  $i$  y  $c_i$  el compuesto  $i$ .  
Por ejemplo,  $\text{CH}_4 + 2 \text{O}_2 \rightarrow \text{CO}_2 + 2 \text{H}_2\text{O}$



puede escribirse como:



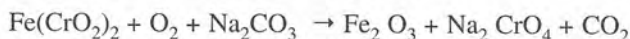
Dado que los átomos se conservan en una reacción química, la ecuación de conservación del elemento  $k$  es:

$$\sum_{i=1}^n x_i m_{i,k} = 0; \quad k = 1, 2, \dots, m$$

donde  $m_{i,k}$  es el número de átomos del elemento  $k$  en el compuesto  $i$ .

Esta última expresión representa un conjunto de ecuaciones lineales, donde  $x_i$  son las incógnitas. Lo anterior se conoce como el **método algebraico de balanceo de ecuaciones químicas**.

Utilice este método para balancear la ecuación química



**3.37** Factorice las siguientes en la forma  $LU$ , con el algoritmo 3.6

a) 
$$\begin{bmatrix} 4 & 1 & -1 \\ 2 & 5 & 0 \\ 3 & 8 & 9 \end{bmatrix}$$

b) 
$$\begin{bmatrix} 3.444 & 16100 & -9.1 \\ 1.9999 & 17.01 & 9.6 \\ 1.6 & 5.2 & 1.7 \end{bmatrix}$$

c) 
$$\begin{bmatrix} 5.8 & 3.2 & 11.25 \\ 4.3 & 3.4 & 9.625 \\ 2.5 & 5.2 & 9.625 \end{bmatrix}$$

d) 
$$\begin{bmatrix} 1.002 & 4 \times 10^{-4} & 5 \times 10^{-4} & 8 \times 10^{-6} \\ 0 & 2.3 & 3 \times 10^{-3} & 4 \times 10^{-5} \\ 0 & 0 & 5 & 0.01 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

e) 
$$\begin{bmatrix} 4 & 0 & 0 & 0 \\ 7 & 10 & 0 & 0 \\ 8 & 9 & 5 & 0 \\ 10 & 3 & 3 & 1 \end{bmatrix}$$

f) 
$$\begin{bmatrix} 0 & 6 & 1 & 12 \\ 7 & 5 & -2 & 5 \\ -4 & 5 & 7 & 8 \\ 3 & 9 & 12 & 24 \end{bmatrix}$$

**3.38** Factorice las matrices del problema 3.37, con el algoritmo 3.7.

**3.39** Resuelva los siguientes sistemas de ecuaciones con el algoritmo 3.8

a) 
$$\begin{aligned} 4x_1 + x_2 - x_3 &= 8 \\ 2x_1 + 5x_2 &= 5 \\ 3x_1 + 8x_2 + 9x_3 &= 0 \end{aligned}$$

b) 
$$\begin{aligned} 3.444x_1 + 16100x_2 - 9.1x_3 &= 0 \\ 1.9999x_1 + 17.01x_2 + 9.6x_3 &= 1 \\ 1.6x_1 + 5.2x_2 + 1.7x_3 &= 0 \end{aligned}$$

c) 
$$\begin{aligned} 5.8x_1 + 3.2x_2 + 11.24x_3 &= 20.24 \\ 4.3x_1 + 3.4x_2 + 9.625x_3 &= 17.325 \\ 2.5x_1 + 5.2x_2 + 9.625x_3 &= 17.325 \end{aligned}$$

d) 
$$\begin{aligned} 4x_1 + 5x_2 + 2x_3 - x_4 &= 3 \\ 5x_1 + 8x_2 + 7x_3 + 6x_4 &= 2 \\ 3x_1 + 7x_2 - 4x_3 - 2x_4 &= 0 \\ -x_1 + 6x_2 - 2x_3 + 5x_4 &= 1 \end{aligned}$$

e) 
$$\begin{aligned} 2.156x_1 + 4.102x_2 - 2.3217x_3 + 6x_4 &= 18 \\ -4.102x_1 + 6x_2 + &+ 1.2x_4 = 6.5931 \\ -x_1 - 5.7012x_2 + 1.2222x_3 &= 3.4 \\ 6.532x_1 + 7x_2 &- 4x_4 = 0 \end{aligned}$$

**3.40** Factorice las matrices simétricas siguientes, mediante el algoritmo 3.9.

$$a) \begin{bmatrix} -5 & 5 & 3 \\ 5 & 6 & 1 \\ 3 & 1 & 7 \end{bmatrix} \quad b) \begin{bmatrix} 3.33 & 4.81 & -2.22 \\ 4.81 & 10.0 & 7.45 \\ -2.22 & 7.45 & 15.0 \end{bmatrix}$$

$$c) \begin{bmatrix} 72.0 & 0.00 & 0.00 & 9.00 & 0.00 & 0.00 \\ 0.00 & 2.88 & 0.00 & 0.00 & 0.00 & -4.50 \\ 0.00 & 0.00 & 18.00 & 9.00 & 0.00 & 0.00 \\ 9.00 & 0.00 & 9.00 & 12.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 33.00 & 0.00 \\ 0.00 & -4.50 & 0.00 & 0.00 & 0.00 & 33.00 \end{bmatrix}$$

**3.41** Con el algoritmo 3.9, elabore un algoritmo para resolver sistemas lineales simétricos y resuelva con él los siguientes sistemas.

$$a) \begin{aligned} -5x_1 + 5x_2 + 3x_3 &= 1 \\ 5x_1 + 6x_2 + x_3 &= 2 \\ 3x_1 + x_2 + 7x_3 &= 3 \end{aligned}$$

$$b) \begin{aligned} 3.33x_1 + 4.81x_2 - 2.22x_3 &= 5 \\ 4.81x_1 + 10.00x_2 + 7.45x_3 &= 0 \\ -2.22x_1 + 7.45x_2 + 15.00x_3 &= 2 \end{aligned}$$

$$c) \begin{aligned} 72x_1 &+ 9x_4 &= 2 \\ &2.88x_2 &-4.5x_6 = 0.5 \\ &18x_3 + 9x_4 &= 1 \\ &9x_3 + 12x_4 &= 0 \\ &&33x_5 &= 1.2 \\ -4.5x_2 &&+ 33x_6 &= 5 \end{aligned}$$

**3.42** Use el algoritmo 3.10 para factorizar en la forma  $L L^T$  las siguientes matrices positivas definitivas.

$$a) \begin{bmatrix} 4 & -2 & 0 \\ -2 & 4 & -1 \\ 0 & -1 & 4 \end{bmatrix} \quad b) \begin{bmatrix} 5 & 1 & 2 & -1 \\ 1 & 7 & 0 & 3 \\ 2 & 0 & 5 & 1 \\ -1 & 3 & 1 & 8 \end{bmatrix} \quad c) \begin{bmatrix} 10 & 0 & 0 & -1 & 0 \\ 0 & 5 & 0 & 0 & 2 \\ 0 & 0 & 2 & 0 & 0 \\ -1 & 0 & 0 & 8 & 3 \\ 0 & 2 & 0 & 3 & 5 \end{bmatrix}$$

**3.43** Mediante el algoritmo 3.10 elabore un algoritmo para resolver sistemas lineales con matriz coeficiente positiva definida y resuelva con él los siguientes sistemas.

$$a) \begin{aligned} 4x_1 - 2x_2 &= 0 \\ -2x_1 + 4x_2 - x_3 &= 0.5 \\ -x_2 + 4x_3 &= 1 \end{aligned}$$

$$b) \begin{aligned} 5x_1 + x_2 + 2x_3 - x_4 &= 1 \\ x_1 + 7x_2 &+ 3x_4 = 2 \\ 2x_1 &+ 5x_3 + x_4 = 3 \\ -x_1 + 3x_2 + x_3 + 8x_4 &= 4 \end{aligned}$$

$$c) \begin{aligned} 10x_1 &- x_4 &= 0.2 \\ &5x_2 &+ 2x_4 &= 0.4 \\ &&2x_3 &= 1.0 \\ -x_1 &&+ 8x_4 + 3x_5 &= 0.6 \\ &2x_2 &+ 3x_4 + 5x_5 &= 0.8 \end{aligned}$$

- 3.44** Si la factorización de  $A$  en las matrices  $L$  y  $U$  es posible, puede imponerse que  $u_{i,i} = 1$  con  $i = 1, 2, \dots, n$ . Con estas condiciones obtenga las ecuaciones correspondientes a las ecuaciones 3.74, 3.75 y 3.76, para el caso del orden de  $A$  igual a 3. También obtenga las ecuaciones correspondientes a la ecuación 3.77 para el caso general, orden de  $A$  igual a  $n$ . Este método, como se recordará, es conocido como **algoritmo de Crout**.
- 3.45** Elabore un algoritmo para resolver un sistema de ecuaciones lineales con el método de Crout (véase algoritmo 3.8); resuelva los sistemas del problema 3.43 con el algoritmo encontrado.
- 3.46** Demuestre que en la solución del sistema lineal  $A \mathbf{x} = \mathbf{b}$ , donde  $A$  es positiva definida, con el método de Cholesky se requiere efectuar:

$n$  raíces cuadradas

$$\frac{n^3 + 9n^2 + 2n}{6} \text{ multiplicaciones o divisiones}$$

y

$$\frac{n^3 + 6n^2 - 7n}{6} \text{ sumas o restas}$$

cuando el orden de  $A$  es  $n$ .

- 3.47** Demuestre que si una matriz  $A$  es positiva definida, entonces  $a_{i,i} > 0$  para  $i = 1, 2, \dots, n$ .
- 3.48** Los algoritmos de factorización, cuando son aplicables, se pueden simplificar considerablemente en el caso de matrices bandedas, debido al gran número de ceros que aparecen en estas matrices. Adapte el método de Doolittle o el de Crout para sistemas tridiagonales y una vez obtenidas las ecuaciones correspondientes, elabore un algoritmo eficiente.
- 3.49** En la solución de una estructura doblemente empotrada se obtuvo el siguiente sistema:

$$\frac{1}{EI} \mathbf{c} + \frac{1}{EI} A \mathbf{p} = \mathbf{0}$$

donde  $EI$  es el módulo de elasticidad del elemento,

$$\mathbf{c} = \begin{bmatrix} -1.80 \\ 22.50 \\ -67.50 \\ 0.00 \\ 165.00 \\ 0.00 \end{bmatrix}; \quad \mathbf{p} = \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \\ p_5 \\ p_6 \end{bmatrix};$$

y

$$A = \begin{bmatrix} 72.00 & 0.00 & 0.00 & 9.00 & 0.00 & 0.00 \\ 0.00 & 2.88 & 0.00 & 0.00 & 0.00 & -4.50 \\ 0.00 & 0.00 & 18.00 & 9.00 & 0.00 & 0.00 \\ 9.00 & 0.00 & 9.00 & 12.00 & 0.00 & 0.00 \\ 0.00 & 0.00 & 0.00 & 0.00 & 33.00 & 0.00 \\ 0.00 & -4.50 & 0.00 & 0.00 & 0.00 & 33.00 \end{bmatrix}$$

Encuentre  $\mathbf{p}$ .

- 3.50** Determine si el sistema que sigue está mal condicionado.

$$\begin{bmatrix} 3.4440 & 16100 & -9.1 \\ 1.9999 & 17.01 & 9.6 \\ 1.6000 & 5.20 & 1.7 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 16000.00 \\ 29.00 \\ 8.42 \end{bmatrix}$$

Resuélvalo usando la eliminación de Gauss y aritmética de cinco dígitos.



3.51 La matriz  $H^{(n)}$  de orden  $n$  o matriz de Hilbert, definida por

$$h_{ij} = \frac{1}{i+j-1}; \quad 1 \leq i \leq n; \quad 1 \leq j \leq n,$$

es una matriz mal condicionada que surge, por ejemplo, al resolver las ecuaciones normales del método de aproximación por mínimos cuadrados (véase capítulo 5). Encuentre  $H^{(4)}$ ,  $H^{(5)}$  y sus inversas por alguno de los métodos vistos; además, resuelva el sistema.

$$H^{(4)} \mathbf{x} = [1 \ 0 \ 1 \ 0]^T$$

3.52 Demuestre que:

det  $P = -1$ , el determinante de una matriz permutadora es  $-1$ .

det  $M = m$ , el determinante de una matriz multiplicadora es el factor  $m$  ( $m \neq 0$ ).

det  $S = 1$ , el determinante de una matriz del tipo  $S$  es 1.

Sugerencia: Utilice la función determinante de una matriz de orden  $n$ .

3.53 Dados los siguientes sistemas de ecuaciones lineales:

$$a) \begin{bmatrix} 3 & 5 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 3 \end{bmatrix} \quad b) \begin{bmatrix} 3 & 1 & 4 \\ 2 & 0 & 1 \\ -1 & 5 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} \quad c) \begin{bmatrix} -2 & 4 & 5 \\ 4 & 8 & 3 \\ 5 & 3 & 6 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

calcule las inversas, determinantes y soluciones correspondientes, usando matrices elementales.

3.54 Resuelva los siguientes sistemas de ecuaciones lineales mediante los métodos de Gauss-Seidel y de Jacobi.

$$a) \begin{bmatrix} 5 & -1 & -1 \\ 1 & -1 & 2 \\ 3 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 3 \\ 0 \\ 4 \end{bmatrix} \quad b) \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \\ 4 \end{bmatrix}$$

c) inciso c) del problema 3.41.

$$d) \begin{bmatrix} 1 & 2 & 2^2 & 2^3 & 2^4 \\ 1 & 6 & 6^2 & 6^3 & 6^4 \\ 1 & 10 & 10^2 & 10^3 & 10^4 \\ 1 & 20 & 20^2 & 20^3 & 20^4 \\ 1 & 30 & 30^2 & 30^3 & 30^4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} 13.4 \\ 30.4 \\ 41.8 \\ 57.9 \\ 66.5 \end{bmatrix}$$

$$d) \begin{bmatrix} 8 & 0 & 6 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9 & 0 & 0 & 0 & 0 & 5 & 0 & 2 & 1 & 0 \\ 0 & 1 & 7 & 0 & 0 & 1 & 2 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 6 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 9 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 2 & 0 & 10 & 1 & 0 & 3 & 0 & 0 \\ 0 & 0 & 5 & 0 & 2 & 2 & 10 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 6 & 1 & 0 & 0 & 15 & 0 & 2 & 0 \\ 0 & 2 & 0 & 0 & 4 & 0 & 1 & 1 & 20 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 6 & 5 & 25 & 1 \\ 1 & 0 & 3 & 1 & 5 & 0 & 7 & 0 & 0 & 1 & 12 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \\ x_{11} \end{bmatrix} = \begin{bmatrix} 1 \\ 5 \\ 8 \\ 0 \\ 8 \\ 1 \\ 0 \\ 3 \\ 0 \\ 1 \\ 2 \end{bmatrix}$$

- 3.55 Elabore un algoritmo para arreglar la matriz aumentada de un sistema, de modo que la matriz coeficiente quede lo más cercana posible a diagonal dominante.
- 3.56 Elabore un algoritmo para resolver un sistema de ecuaciones lineales, usando los métodos SOR con  $w > 1$  y con  $w < 1$ .

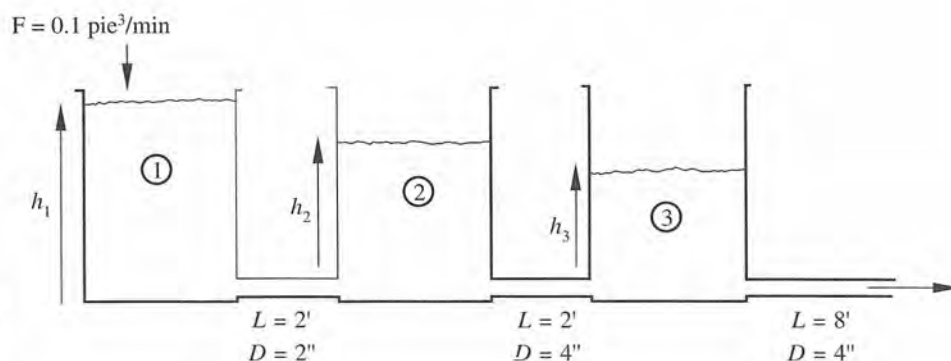
SUGERENCIA: Puede obtenerlo fácilmente modificando el algoritmo 3.11.

- 3.57 Aplique la segunda ley de Kirchhoff a la malla formada por el contorno del circuito de la figura 3.15; es decir, no considere  $E_5$ ,  $R_4$  y  $R_5$ . Demuestre que la ecuación resultante es linealmente dependiente de las tres obtenidas al seccionar en mallas dicho circuito.
- 3.58 Resuelva los sistemas de ecuaciones lineales del problema 3.54 con el algoritmo elaborado en el problema 3.56.
- 3.59 Demuestre que la ecuación 3.110 es un polinomio de grado  $n$  si  $A$  es una matriz de orden  $n$  e  $I$  es la matriz identidad correspondiente.
- 3.60 Encuentre los valores característicos (eigenvalores) de la matriz coeficiente del siguiente sistema.

$$\begin{bmatrix} 4 & 1 & 0.3 & -1 \\ 1 & 7 & 2 & 0 \\ -0.3 & 2 & 5 & 2 \\ -1 & 0 & -2 & 4 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 2 \\ 3 \end{bmatrix}$$

- 3.61 Encuentre los vectores característicos (eigenvectores) correspondientes al valor característico dominante (el de máximo valor absoluto) del ejemplo anterior.
- 3.62 Encuentre el valor característico dominante y los vectores característicos correspondientes del sistema de ecuaciones del problema 3.60.
- 3.63 Se tienen tres tanques cilíndricos iguales de 6 pies de diámetro, comunicados entre sí por medio de tubos de 4 pulgadas de diámetro y 2 pies de largo, como se muestra en la figura 3.21. El tercer tanque tiene una salida a través de un tubo de 4 pulgadas de diámetro y 8 pies de largo. Al primer tanque llega un fluido a razón de 0.1 pies cúbicos por minuto e inicialmente su nivel tiene una altura de 20 pies, mientras que el segundo y tercer tanques están vacíos. El fluido es un aceite viscoso cuya densidad es de  $51.45 \text{ lb}_m/\text{pie}^3$ , y viscosidad es 100 centipoises. Calcule la altura del fluido en cada tanque cuando se alcance el régimen permanente.

SUGERENCIA: Use la ecuación de Poiseuille para el cálculo de la velocidad media del fluido a través de los tubos.



**Figura 3.21**  
Tres tanques  
interconectados.

# CAPÍTULO 4

## SISTEMAS DE ECUACIONES

### NO LINEALES

#### A dónde nos dirigimos

En este capítulo estudiaremos las técnicas que nos permitirán resolver sistemas de ecuaciones no lineales,  $f(\mathbf{x}) = \mathbf{0}$ , vistas como la situación más general de los casos que analizamos en los capítulos 2 y 3. Para esto, utilizaremos sistemas de dos ecuaciones no lineales en dos incógnitas, lo cual no implica pérdida de generalidad y a cambio nos permitirá realizar los cálculos de manera más ágil y, sobre todo, presentar una interpretación geométrica del método. De esta manera, el lector tendrá frente a sí un reto de visualización y entenderá por qué estos métodos requieren de numerosos cálculos.

Además de las extensiones que se hacen del método de punto fijo, de Newton-Raphson y de bisección a sistemas no lineales, estudiaremos una fórmula que, mediante un planteamiento generalizado, nos permite proponer y explorar técnicas diferentes a las vistas anteriormente, como la del descenso de máxima pendiente y variantes a las ya conocidas como la de Newton-Raphson con optimización del tamaño de paso.

Con este capítulo terminamos la parte algebraica del libro en la que se basarán los siguientes capítulos correspondientes a la parte de análisis y dinámica, ya que muchos de los problemas que veremos más adelante se reducen a resolver problemas de tipo algebraico; por ejemplo: cuando se resuelve una ecuación diferencial parcial donde se termina resolviendo un sistema lineal o no lineal de ecuaciones, o en el caso en que es necesario resolver problemas de absorción, destilación, diseño de reactores, diseño de vigas u otros de gran interés para el ingeniero.

## Introducción

En el capítulo 2 se vio cómo encontrar las raíces de una ecuación de la forma:

$$f(x) = 0.$$

Por otro lado, en el capítulo 3 se estudiaron las técnicas iterativas de solución de un sistema de ecuaciones lineales  $Ax = b$ .

Estos dos son casos particulares de la situación más general, donde se tiene un sistema de varias ecuaciones con varias incógnitas, cuya representación es:

$$\begin{aligned} f_1(x_1, x_2, x_3, \dots, x_n) &= 0 \\ f_2(x_1, x_2, x_3, \dots, x_n) &= 0 \\ &\vdots \\ f_n(x_1, x_2, x_3, \dots, x_n) &= 0, \end{aligned} \tag{4.1}$$

donde  $f_i(x_1, x_2, x_3, \dots, x_n)$  para  $1 \leq i \leq n$  es una función (lineal o no) de las variables independientes  $x_1, x_2, x_3, \dots, x_n$ .



Si por ejemplo la ecuación 4.1 consiste sólo en una ecuación de una incógnita ( $n = 1$ ), se tiene la ecuación 2.1. En cambio la ecuación 4.1 se reducirá al caso (3.39) si  $n > 1$  y  $f_1, f_2, \dots, f_n$  son todas funciones lineales de  $x_1, x_2, x_3, \dots, x_n$ .

Por todo esto, es fácil entender que los métodos iterativos de solución de la ecuación 4.1 son extensiones de los métodos para ecuaciones no lineales con una incógnita y emplean las ideas que se aplicaron al desarrollar los algoritmos iterativos para resolver  $Ax = b$ .

A continuación se dan algunos ejemplos.

$$a) f_1(x_1, x_2) = x_1^2 + x_2^2 - 4 = 0$$

$$f_2(x_1, x_2) = x_2 - x_1^2 = 0$$

$$b) f_1(x_1, x_2) = 10(x_2 - x_1^2) = 0$$

$$f_2(x_1, x_2) = 1 - x_1 = 0$$

$$c) f_1(x_1, x_2, x_3) = x_1x_2x_3 - 10x_1^3 + x_2 = 0$$

$$f_2(x_1, x_2, x_3) = x_1 + 2x_2x_3 + \text{sen}x_2 - 15 = 0$$

$$f_3(x_1, x_2, x_3) = x_2^2 - 5x_1x_3 - 3x_3^3 + 3 = 0$$

## 4.1 Dificultades en la solución de sistemas de ecuaciones no lineales

Antes de desarrollar los métodos iterativos para resolver sistemas de ecuaciones no lineales con varias incógnitas, se destacarán algunas de las dificultades que se presentan al aplicar estos métodos.

- Es imposible graficar las superficies multidimensionales definidas por las ecuaciones de los sistemas para  $n > 2$ .
- No es fácil encontrar “buenos” valores iniciales.

Para atenuar estas dificultades proporcionamos algunas sugerencias antes de analizar un intento formal de solución de la ecuación 4.1.

### REDUCCIÓN DE ECUACIONES

Resulta muy útil tratar de reducir analíticamente el número de ecuaciones y de incógnitas antes de intentar una solución numérica. En particular, hay que intentar resolver alguna de las ecuaciones para alguna de las incógnitas. Después, sustitúyase la ecuación resultante para esa incógnita en todas las demás ecuaciones; con esto el sistema se reduce en una ecuación y una incógnita. Continúe de esta manera hasta donde sea posible.

Por ejemplo, en el sistema

$$f_1(x_1, x_2) = 10(x_2 - x_1^2) = 0$$

$$f_2(x_1, x_2) = 1 - x_1 = 0$$

se despeja  $x_1$  en la segunda ecuación

$$x_1 = 1$$

y se sustituye en la primera

$$10(x_2 - 1^2) = 0$$

cuya solución,  $x_2 = 1$ , conjuntamente con  $x_1 = 1$  proporciona una solución del sistema dado, sin necesidad de resolver dos ecuaciones con dos incógnitas.

### PARTICIÓN DE ECUACIONES

A veces resulta más sencillo dividir las ecuaciones en subsistemas menores y resolverlos por separado. Considérese por ejemplo el siguiente sistema de cinco ecuaciones con cinco incógnitas.

$$f_1(x_1, x_2, x_3, x_4, x_5) = 0$$

$$f_2(x_1, x_2, x_4) = 0$$

$$f_3(x_1, x_3, x_4, x_5) = 0$$

$$f_4(x_2, x_4) = 0$$

$$f_5(x_1, x_4) = 0$$

En vez de atacar las cinco ecuaciones al mismo tiempo, se resuelve el subsistema formado por  $f_2, f_4$  y  $f_5$ . Las soluciones de este subsistema se utilizan después para resolver el subsistema compuesto por las ecuaciones  $f_1$  y  $f_3$ .

En general, una partición de ecuaciones es la división de un sistema de ecuaciones en subsistemas llamados bloques. Cada bloque de la partición es el sistema de ecuaciones más pequeño que incluye todas las variables que es preciso resolver.

### TANTEO DE ECUACIONES

Supóngase que se quiere resolver el siguiente sistema de cuatro ecuaciones con cuatro incógnitas.

$$f_1(x_2, x_3) = 0$$

$$f_2(x_2, x_3, x_4) = 0$$

$$f_3(x_1, x_2, x_3, x_4) = 0$$

$$f_4(x_1, x_2, x_3) = 0$$

No se pueden dividir en subsistemas, sino que es preciso resolverlas simultáneamente; sin embargo, es posible abordar el problema por otro camino. Supóngase que se estima un valor de  $x_3$ . Se podría obtener así  $x_2$  a partir de  $f_1$ ,  $x_4$  de  $f_2$  y  $x_1$  de  $f_3$ . Finalmente, se comprobaría con  $f_4$  la estimación hecha de  $x_3$ . Si  $f_4$  fuese cero o menor en magnitud que un valor predeterminado o criterio de exactitud  $\varepsilon$ , la estimación  $x_3$  y los valores de  $x_2$ ,  $x_4$  y  $x_1$  obtenidos con ella, serían una aproximación a la solución del sistema dado. En caso contrario, habría que proponer un nuevo valor de  $x_3$  y repetir el proceso.

Nótese la íntima relación que guarda este método con el de punto fijo (Cap. 2), ya que un problema multidimensional se reduce a uno unidimensional en  $x_3$

$$h(x_3) = 0.$$

### VALORES INICIALES

#### a) De consideraciones físicas

Si el sistema de ecuaciones 4.1 tiene un significado físico, con frecuencia es posible acotar los valores de las incógnitas a partir de consideraciones físicas. Por ejemplo, si alguna de las variables  $x_i$  representa la velocidad de flujo de un fluido, ésta no podrá ser negativa. Por tanto,  $x_i \geq 0$ . En el caso de que  $x_i$  represente una concentración expresada como fracción peso o fracción molar de una corriente de alimentación, se tiene que  $0 \leq x_i \leq 1$ . (Para mayores detalles ver los ejercicios resueltos al final del capítulo).

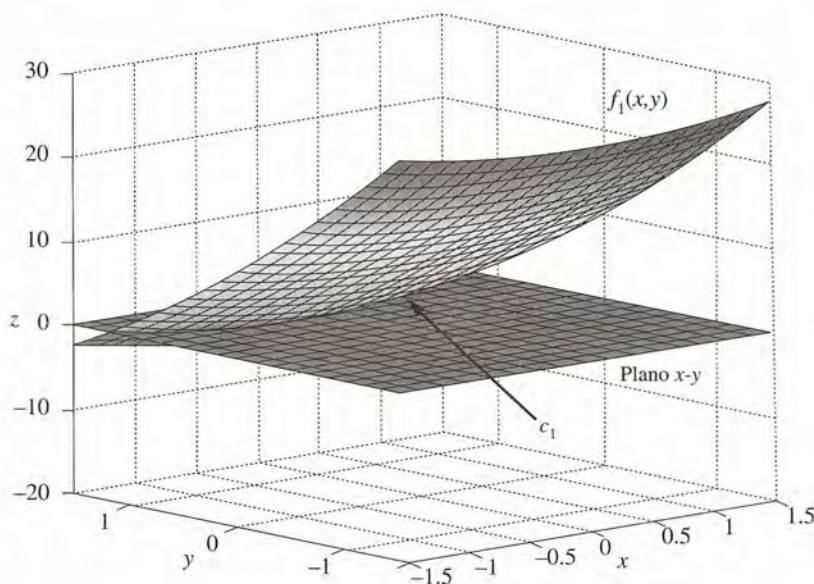
- b) Visualización de raíces en sistemas de dos ecuaciones no lineales con dos incógnitas.  
Sea el sistema

$$\begin{aligned} f_1(x, y) &= x^2 - 10x + y^2 + 8 = 0 \\ f_2(x, y) &= xy^2 + x - 10y + 8 = 0 \end{aligned} \quad (4.2)$$

Algebraicamente una solución o raíz del sistema 4.2 es una pareja  $\bar{x}, \bar{y}$ , tal que satisface cada una de las ecuaciones de dicho sistema. Nos permitiremos hacer la interpretación o visualización de una raíz a través de varias etapas:

Al graficar la función  $f_1(x, y)$ , se obtiene una superficie en el espacio, como se ve en la figura 4.1.

Para apreciar mejor ésta y otras gráficas, se ha incluido una inserción a color (con las Fig. 4.1 a la Fig. 4.8) en páginas más adelante.



**Figura 4.1**  
Intersección de  
la superficie  
 $f_1(x, y)$  con el  
plano  $x-y$ .

La intersección, si hay, de la superficie  $f_1(x, y)$  con el plano  $x-y$  puede resultar en una curva  $c_1$ , como se muestra en la figura 4.1. A lo largo de esta curva se da el hecho de que  $f_1(x, y) = 0$ , dicho de otra manera, que los puntos de esta curva son la solución de la ecuación  $f_1(x, y) = 0$ , no del sistema 4.2.

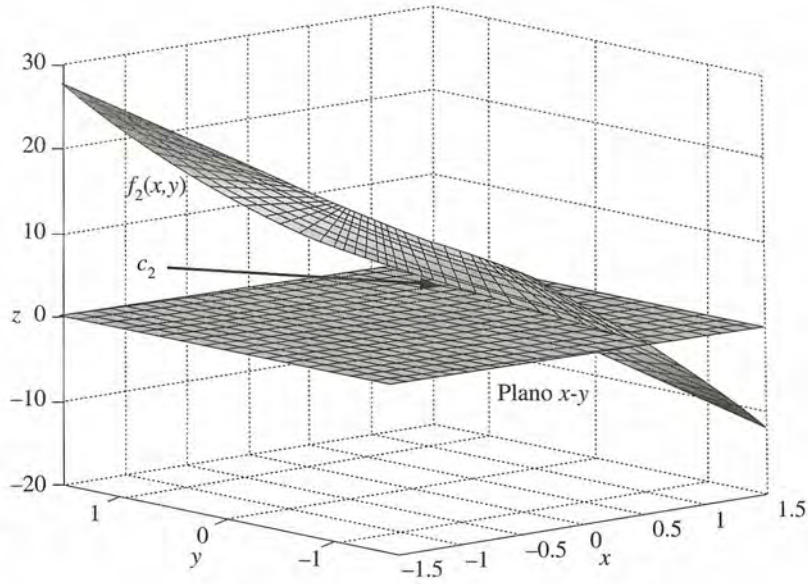
Repetiendo el mismo procedimiento con la superficie de la función  $f_2(x, y)$ , se obtiene otra curva  $c_2$  en el plano  $x-y$ , que ahora resulta ser la solución de la ecuación  $f_2(x, y) = 0$ , ver figura 4.2.

Finalmente, las intersecciones de las curvas  $c_1$  y  $c_2$  del plano  $x-y$ , resultan ser puntos comunes a las tres superficies:  $f_1(x, y)$ ,  $f_2(x, y)$  y el plano  $x-y$ ; dichos puntos satisfacen ambas ecuaciones del sistema 4.2 y son precisamente las raíces  $\bar{x}, \bar{y}$  que buscamos, ver figura 4.3. Partiendo de la raíz mostrada en la gráfica 4.3 se pueden proponer valores iniciales.

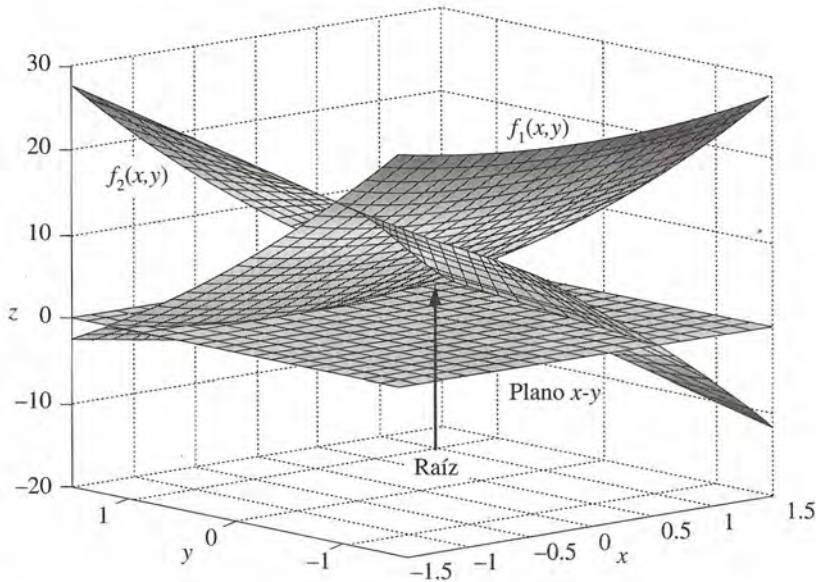
Por último, resulta muy conveniente conocer bien las características de cada método de solución del sistema 4.1 para efectuar la elección más adecuada para el mismo.

Se iniciará el estudio de dichos métodos con la extensión del método de punto fijo a sistemas de ecuaciones no lineales.





**Figura 4.2**  
Intersección de  
la superficie  
 $f_2(x, y)$  con el  
plano  $x-y$ .



**Figura 4.3**  
Intersección de  
las superficies  
 $f_1(x, y)$  y  $f_2(x, y)$   
con el plano  
 $x-y$ .

## 4.2 Método de punto fijo multivariable

Los algoritmos estudiados en este capítulo son, en principio, aplicables a sistemas de cualquier número de ecuaciones; sin embargo, para ser más concisos y evitar notación complicada, se considerará sólo el caso de dos ecuaciones con dos incógnitas. Éstas generalmente se escribirán como:

$$\begin{aligned} f_1(x, y) &= 0 \\ f_2(x, y) &= 0 \end{aligned} \tag{4.3}$$

y se tratará de encontrar pares de valores  $(x, y)$  que satisfagan ambas ecuaciones.

Como en el método de punto fijo (Sec. 2.1), y en los métodos de Jacobi y Gauss-Seidel (Sec. 3.5), se resolverá la primera ecuación para alguna de las variables,  $x$  por ejemplo, y la segunda para  $y$ .

$$\begin{aligned}x &= g_1(x, y) \\ y &= g_2(x, y)\end{aligned}\quad (4.4)$$

Al igual que en los métodos mencionados, se tratará de obtener la estimación  $(k+1)$ -ésima a partir de la estimación  $k$ -ésima con la expresión

$$\begin{aligned}x^{k+1} &= g_1(x^k, y^k) \\ y^{k+1} &= g_2(x^k, y^k)\end{aligned}\quad (4.5)$$

Se comienza con valores iniciales  $x^0, y^0$ , se calculan nuevos valores  $x^1, y^1$ , y se repite el proceso, esperando que después de cada iteración los valores de  $x^k, y^k$  se aproximen a la raíz buscada  $\bar{x}, \bar{y}$ , la cual cumple con:

$$\begin{aligned}\bar{x} &= g_1(\bar{x}, \bar{y}) \\ \bar{y} &= g_2(\bar{x}, \bar{y})\end{aligned}$$

Por analogía con los casos analizados, puede predecirse el comportamiento y las características de este método de punto fijo multivariable.

Como se sabe, en el caso de una variable la manera particular de pasar de  $f(x) = 0$  a  $x = g(x)$ , afecta la convergencia del proceso iterativo. Entonces debe esperarse que la forma en que se resuelve para  $x = g_1(x, y)$  y  $y = g_2(x, y)$  afecta la convergencia de las iteraciones (4.5).

Por otro lado, se sabe que el reordenamiento de las ecuaciones en el caso lineal afecta la convergencia, por lo que puede esperarse que la convergencia del método en estudio dependa de si se despeja  $x$  de  $f_2$  o de  $f_1$ .

Finalmente, como en el método iterativo univariable y en el de Jacobi y de Gauss-Seidel, la convergencia —en caso de existir— es de primer orden, cabe esperar que el método iterativo multivariable tenga esta propiedad.

### Ejemplo 4.1

Encuentre una solución del sistema de ecuaciones no lineales

$$\begin{aligned}f_1(x, y) &= x^2 - 10x + y^2 + 8 = 0 \\ f_2(x, y) &= xy^2 + x - 10y + 8 = 0.\end{aligned}$$

### Solución



Con el despeje de  $x$  del término  $(-10x)$  en la primera ecuación, y de  $y$  del término  $(-10y)$  en la segunda ecuación, resulta:

$$\begin{aligned}x &= \frac{x^2 + y^2 + 8}{10} \\ y &= \frac{xy^2 + x + 8}{10}\end{aligned}$$

o con la notación de la ecuación 4.5

$$\begin{aligned}x^{k+1} &= \frac{(x^k)^2 + (y^k)^2 + 8}{10} \\ y^{k+1} &= \frac{x^k (y^k)^2 + x^k + 8}{10}\end{aligned}$$

Con los valores iniciales  $x^0 = 0, y^0 = 0$ , se inicia el proceso iterativo.

**Primera iteración**

$$x^1 = \frac{0^2 + 0^2 + 8}{10} = 0.8$$

$$y^1 = \frac{0(0)^2 + 0 + 8}{10} = 0.8$$

**Segunda iteración**

$$x^2 = \frac{(0.8)^2 + (0.8)^2 + 8}{10} = 0.928$$

$$y^2 = \frac{0.8(0.8)^2 + 0.8 + 8}{10} = 0.9312$$

Al continuar el proceso iterativo, se encuentra la siguiente sucesión de vectores:

| $k$ | $x^k$   | $y^k$   |
|-----|---------|---------|
| 0   | 0.00000 | 0.00000 |
| 1   | 0.80000 | 0.80000 |
| 2   | 0.92800 | 0.93120 |
| 3   | 0.97283 | 0.97327 |
| 4   | 0.98937 | 0.98944 |
| 5   | 0.99578 | 0.99579 |
| 6   | 0.99832 | 0.99832 |
| 7   | 0.99933 | 0.99933 |
| 8   | 0.99973 | 0.99973 |
| 9   | 0.99989 | 0.99989 |
| 10  | 0.99996 | 0.99996 |
| 11  | 0.99998 | 0.99998 |
| 12  | 0.99999 | 0.99999 |
| 13  | 1.00000 | 1.00000 |

Los cálculos pueden hacerse con Matlab o con la TI-92 Plus



```

X0=0; y0=0;
fprintf(' k x(k) y(k)\n')
fprintf(' %2d %10.5f %10.5f\n', 0, x0, y0)
for k=1:13
x1=(x0^2+y0^2+8)/10;
y1=(x0*y0^2+x0+8)/10;
fprintf(' %2d %10.5f %10.5f\n',k,x1,y1)
x0=x1; y0=y1;
end

```





```

Plus e4_1( )
Prgm
Define g1(x, y) = (x^2+y^2+8)/10
Define g2(x, y) = (x*y^2+x+8)/10
0→x0 : 0→y0 : 0→k : ClrIO
Disp "k  x(k)    y(k)"
string(k) &"    "&format(x0, "f5")→d
d&"    "&format(y0, "f5")→d : Disp d
For k, 1, 13
  g1(x0, y0)→x : g2(x0, y0)→y
  string(k) &"    "&format(x, "f5")→d
  d&"    "&format(y, "f5")→d : Disp d
  x→x0 : y→y0
EndFor
EndPrgm

```

Para observar la convergencia del proceso iterativo, se pudieron usar los criterios del capítulo anterior, como distancia entre dos vectores consecutivos, o bien las distancias componente a componente de dos vectores consecutivos. También existe un criterio de convergencia equivalente al de las ecuaciones 2.10 y 3.99 que puede aplicarse antes de iniciar el proceso iterativo mencionado, y que dice:

Una condición suficiente aunque no necesaria, para asegurar la convergencia es que

$$\left| \frac{\partial g_1}{\partial x} \right| + \left| \frac{\partial g_2}{\partial x} \right| \leq M < 1 ; \left| \frac{\partial g_1}{\partial y} \right| + \left| \frac{\partial g_2}{\partial y} \right| \leq M < 1 \quad (4.6)$$

para todos los puntos  $(x, y)$  de la región del plano que contiene todos los valores  $(x^k, y^k)$  y la raíz buscada  $(\bar{x}, \bar{y})$ .

Por otro lado, si  $M$  es muy pequeña en una región de interés, la iteración converge rápidamente; si  $M$  es cercana a 1 en magnitud, entonces la iteración puede converger lentamente. Este comportamiento es similar al del caso de una función univariable discutido en el capítulo 2.

Por lo general es muy difícil encontrar el sistema 4.4 a partir de la ecuación 4.3, de modo que satisfaga la condición 4.5.

De todas maneras, cualquiera que sea el sistema (4.4) a que se haya llegado y que se vaya a resolver con este método, puede aumentarse la velocidad de convergencia usando desplazamientos sucesivos en lugar de los desplazamientos simultáneos del esquema 4.5. Es decir, se iteraría mediante

$$\begin{aligned} x^{k+1} &= g^1(x^k, y^k) \\ y^{k+1} &= g^2(x^{k+1}, y^k) \end{aligned} \quad (4.7)$$

Como en el caso lineal (Jacobi y Gauss-Seidel), si la iteración por desplazamientos simultáneos diverge, generalmente el método por desplazamientos sucesivos divergería más rápido; es decir, se detecta más rápido la divergencia, por lo que se recomienda en general el uso de desplazamientos sucesivos en lugar de desplazamientos simultáneos.

**Ejemplo 4.2**

Resuelva el sistema del ejemplo 4.1 utilizando el método de punto fijo multivariable con desplazamientos sucesivos

$$f_1(x, y) = x^2 - 10x + y^2 + 8 = 0$$

$$f_2(x, y) = xy^2 + x - 10y + 8 = 0$$

**SUGERENCIA:** Se pueden seguir los cálculos con un pizarrón electrónico o se programa una calculadora.

**Solución** Al despejar  $x$  del término  $(-10x)$  y  $y$  del término  $(-10y)$  de la primera y segunda ecuaciones, respectivamente, resulta

$$x^{k+1} = g_1(x^k, y^k) = \frac{(x^k)^2 + (y^k)^2 + 8}{10}$$

$$y^k = g_2(x^{k+1}, y^k) = \frac{x^{k+1} (y^k)^2 + x^{k+1} + 8}{10}$$

Al derivar parcialmente, se obtiene

$$\frac{\partial g_1}{\partial x} = \frac{2x^k}{10}$$

$$\frac{\partial g_1}{\partial y} = \frac{2y^k}{10}$$

$$\frac{\partial g_2}{\partial x} = \frac{(y^k)^2 + 1}{10}$$

$$\frac{\partial g_2}{\partial y} = \frac{2x^{k+1} y^k}{10}$$

y evaluadas en  $x^0 = 0$  y en  $y^0 = 0$

$$\left. \frac{\partial g_1}{\partial x} \right|_{x^0, y^0} = 0$$

$$\left. \frac{\partial g_1}{\partial y} \right|_{x^0, y^0} = 0$$

$$\left. \frac{\partial g_2}{\partial x} \right|_{x^0, y^0} = 1/10$$

$$\left. \frac{\partial g_2}{\partial y} \right|_{x^0, y^0} = 0$$

con lo que se puede aplicar la condición 4.6

$$\frac{\partial g_1}{\partial x} + \frac{\partial g_2}{\partial x} = 0 + 1/10 = 1/10 < 1$$

$$\frac{\partial g_1}{\partial y} + \frac{\partial g_2}{\partial y} = 0 + 0 = 0 < 1$$

la cual se satisface; si los valores sucesivos de la iteración:  $x^1, y^1; x^2, y^2; x^3, y^3; \dots$  la satisfacen también, se llega entonces a  $\bar{x}, \bar{y}$ .

**Primera iteración**

$$x^1 = \frac{0^2 + 0^2 + 8}{10} = 0.8$$

$$y^1 = \frac{0.8(0)^2 + 0.8 + 8}{10} = 0.88$$

Cálculo de la distancia entre el vector inicial y el vector  $[x^1, y^1]^T$

$$|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}| = \sqrt{(0.8 - 0.0)^2 + (0.88 - 0.00)^2} = 1.18929$$

**Segunda iteración**

$$x^2 = \frac{(0.8)^2 + (0.88)^2 + 8}{10} = 0.94144$$

$$y^2 = \frac{0.94144(0.88)^2 + 0.94144 + 8}{10} = 0.96704$$

Cálculo de la distancia entre  $[x^2, y^2]^T$  y  $[x^1, y^1]^T$ :

$$|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}| = \sqrt{(0.94144 - 0.8)^2 + (0.96704 - 0.88)^2} = 0.16608$$

A continuación se muestran los resultados de las iteraciones.

| $k$ | $x^k$   | $y^k$   | $ \mathbf{x}^{(k+1)} - \mathbf{x}^{(k)} $ |
|-----|---------|---------|-------------------------------------------|
| 0   | 0.00000 | 0.00000 | —                                         |
| 1   | 0.80000 | 0.88000 | 1.18929                                   |
| 2   | 0.94144 | 0.96705 | 0.16608                                   |
| 3   | 0.98215 | 0.99006 | 0.04677                                   |
| 4   | 0.99448 | 0.99693 | 0.01411                                   |
| 5   | 0.99829 | 0.99905 | 0.00436                                   |
| 6   | 0.99947 | 0.99970 | 0.00135                                   |
| 7   | 0.99983 | 0.99991 | 0.00042                                   |
| 8   | 0.99995 | 0.99997 | 0.00013                                   |
| 9   | 0.99998 | 0.99999 | 0.00004                                   |
| 10  | 0.99999 | 1.00000 | 0.00001                                   |
| 11  | 1.00000 | 1.00000 | 0.00001                                   |

Hay que observar que se requirieron once iteraciones para llegar al vector solución (1,1) contra 13 del ejemplo 4.1, donde se usaron desplazamientos simultáneos.

Los cálculos pueden hacerse con Matlab o con la TI-92 Plus.



```

x0=0; y0=0;
fprintf(' k   x(k)   y(k)   Dist   \n')
fprintf(' %2d %10.5f %10.5f\n', 0, x0, y0)
for k=1:11
    X1=(x0^2+y0^2+8)/10;
    Y1=(x1*y0^2+x1+8)/10;
    Dist=((x1-x0)^2+(y1-y0)^2)^0.5;
    fprintf('%2d %10.5f %10.5f %10.5f\n', k, x1, y1, Dist)
    X0=x1; y0=y1;
End

```





```

E4_2 ( )
Prgm
Define g1(x,y) = (x^2+y^2+8)/10
Define g2(x,y) = (x*y^2+x+8)/10
0→x0 : 0→y0 : 0→k : ClrIO
Disp "k  x(k)  y(k)"
string(k)&" "&format(x0,"f5")→d
d&" "&format(y0,"f5")→d : Disp d
for k, 1, 11
  g1(x0,y0)→x " g2(x,y0)→y
  string(k)&" "&format(x,"f5")→d
  d&" "&format(y,"f5")→d: Disp d
  x→x0 : y→y0
EndFor
EndPrgm

```

A continuación se presenta un algoritmo para el método de punto fijo multivariable en sus versiones de desplazamientos simultáneos y desplazamientos sucesivos.

#### ALGORITMO 4.1 Método de punto fijo multivariable

Para encontrar una solución aproximada de un sistema de ecuaciones no lineales  $\mathbf{g}(\mathbf{x}) = \mathbf{x}$ , proporcionar las funciones  $G(I, \mathbf{x})$ ,  $I=1, 2, \dots, N$  y los

**DATOS:** El número de ecuaciones  $N$ , el vector de valores iniciales  $\mathbf{x}$ , el criterio de convergencia  $EPS$ , el número máximo de iteraciones  $MAXIT$  y  $M = 0$  para desplazamientos sucesivos o  $M = 1$  para desplazamientos simultáneos.

**RESULTADOS:** Una solución aproximada  $\mathbf{x}$  o mensaje "NO HUBO CONVERGENCIA".

- PASO 1. Hacer  $K = 1$ .
- PASO 2. Mientras  $K \leq MAXIT$ , repetir los pasos 3 a 14.
- PASO 3. Si  $M = 0$ , hacer  $\mathbf{x}_{aux} = \mathbf{x}$ . De otro modo continuar.
- PASO 4. Hacer  $I = 1$ .
- PASO 5. Mientras  $I \leq N$ , repetir los pasos 6 y 7.
- PASO 6. Si  $M = 0$ , hacer  $X(I) = G(I, \mathbf{x})$ . De otro modo hacer  $X_{AUX}(I) = G(I, \mathbf{x})$ .
- PASO 7. Hacer  $I = I + 1$ .
- PASO 8. Hacer  $I = 1$ .
- PASO 9. Mientras  $I \leq N$ , repetir los pasos 10 y 11.
- PASO 10. Si  $ABS(X_{AUX}(I) - X(I)) > EPS$  ir al paso 13. De otro modo continuar.
- PASO 11. Hacer  $I = I + 1$ .
- PASO 12. IMPRIMIR  $\mathbf{x}$  Y TERMINAR.
- PASO 13. Si  $M = 1$  hacer  $\mathbf{x} = \mathbf{x}_{aux}$ . De otro modo continuar.
- PASO 14. Hacer  $K = K + 1$ .
- PASO 15. IMPRIMIR mensaje "NO HUBO CONVERGENCIA" y TERMINAR.

**SUGERENCIA:** Desarrolle este algoritmo con Mathcad o un software equivalente.

### 4.3 Método de Newton-Raphson

El método iterativo para sistemas de ecuaciones converge linealmente. Como en el método de una incógnita, puede crearse un método de convergencia cuadrática, es decir, el método de Newton-Raphson multivariabe, a continuación se obtendrá este procedimiento para dos variables; la extensión a tres o más variables es viable generalizando los resultados.

Supóngase que se está resolviendo el sistema.

$$\begin{aligned} f_1(x, y) &= 0 \\ f_2(x, y) &= 0, \end{aligned}$$

Donde ambas funciones son continuas y diferenciables, de modo que puedan expandirse en serie de Taylor. Esto es:

$$\begin{aligned} f(x, y) = f(a, b) + \frac{\partial f}{\partial x}(x-a) + \frac{\partial f}{\partial y}(y-b) + \frac{1}{2!} \left[ \frac{\partial^2 f}{\partial x^2}(x-a)^2 + \right. \\ \left. 2 \frac{\partial^2 f}{\partial x \partial y}(x-a)(y-b) + \frac{\partial^2 f}{\partial y^2}(y-b)^2 \right] + \dots \end{aligned}$$

donde  $f(x, y)$  se ha expandido alrededor del punto  $(a, b)$  y todas las derivadas parciales están evaluadas en  $(a, b)$ .

Expandiendo  $f_1$  alrededor de  $(x^k, y^k)$

$$\begin{aligned} f_1(x^{k+1}, y^{k+1}) = f_1(x^k, y^k) + \frac{\partial f_1}{\partial x}(x^{k+1} - x^k) + \frac{\partial f_1}{\partial y}(y^{k+1} - y^k) + \\ \frac{1}{2!} \left[ \frac{\partial^2 f_1}{\partial x^2}(x^{k+1} - x^k)^2 + 2 \frac{\partial^2 f_1}{\partial x \partial y}(x^{k+1} - x^k)(y^{k+1} - y^k) + \right. \\ \left. \frac{\partial^2 f_1}{\partial y^2}(y^{k+1} - y^k)^2 \right] + \dots \end{aligned} \quad (4.8)$$

donde todas las derivadas parciales están evaluadas en  $(x^k, y^k)$ . De la misma forma puede expandirse  $f_2$  como sigue:

$$\begin{aligned} f_2(x^{k+1}, y^{k+1}) = f_2(x^k, y^k) + \frac{\partial f_2}{\partial x}(x^{k+1} - x^k) + \frac{\partial f_2}{\partial y}(y^{k+1} - y^k) + \\ \frac{1}{2!} \left[ \frac{\partial^2 f_2}{\partial x^2}(x^{k+1} - x^k)^2 + 2 \frac{\partial^2 f_2}{\partial x \partial y}(x^{k+1} - x^k)(y^{k+1} - y^k) + \right. \\ \left. \frac{\partial^2 f_2}{\partial y^2}(y^{k+1} - y^k)^2 \right] + \dots \end{aligned} \quad (4.9)$$

De igual manera que en la ecuación 4.8, todas las derivadas parciales de 4.9 están evaluadas en  $(x^k, y^k)$ .

Ahora supóngase que  $x^{k+1}$  y  $y^{k+1}$  están cerca de la raíz buscada  $(\bar{x}, \bar{y})$  que los lados izquierdos de las dos últimas ecuaciones son casi cero; además, asúmase que  $x^k$  y  $y^k$  están

tan próximos de  $x^{k+1}$  que pueden omitirse los términos a partir de los que se encuentran agrupados en paréntesis rectangulares. Con esto las ecuaciones 4.8 y 4.9 se simplifican a

$$\begin{aligned} 0 &\approx f_1(x^k, y^k) + \frac{\partial f_1}{\partial x}(x^{k+1} - x^k) + \frac{\partial f_1}{\partial y}(y^{k+1} - y^k) \\ 0 &\approx f_2(x^k, y^k) + \frac{\partial f_2}{\partial x}(x^{k+1} - x^k) + \frac{\partial f_2}{\partial y}(y^{k+1} - y^k) \end{aligned} \quad (4.10)$$

Para simplificar aún más, se cambia la notación con

$$\begin{aligned} x^{k+1} - x^k &= h \\ y^{k+1} - y^k &= j \end{aligned} \quad (4.11)$$

y así queda la  $(k + 1)$ -ésima iteración en términos de la  $k$ -ésima, como se ve a continuación

$$\begin{aligned} x^{k+1} &= x^k + h \\ y^{k+1} &= y^k + j \end{aligned} \quad (4.12)$$

La sustitución de la ecuación 4.11 en la 4.10 y el rearrreglo dan como resultado:

$$\begin{aligned} \frac{\partial f_1}{\partial x} h + \frac{\partial f_1}{\partial y} j &= -f_1(x^k, y^k) \\ \frac{\partial f_2}{\partial x} h + \frac{\partial f_2}{\partial y} j &= -f_2(x^k, y^k) \end{aligned} \quad (4.13)$$

el cual es un sistema de ecuaciones lineales en las incógnitas  $h$  y  $j$  (recuérdese que las derivadas parciales de la ecuación 4.13, así como  $f_1$  y  $f_2$  están evaluadas en  $(x^k, y^k)$  y, por tanto, son números reales).

Este sistema de ecuaciones lineales resultante tiene solución única, siempre que el determinante de la matriz de coeficientes o matriz jacobiana  $J$  no sea cero; es decir, si

$$|J| = \begin{vmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{vmatrix} \neq 0$$

Precisando: el método de Newton-Raphson consiste fundamentalmente en formar y resolver el sistema 4.13, esto último por alguno de los métodos vistos en el capítulo 3. Con la solución y la ecuación 4.12 se obtiene la siguiente aproximación.

Este procedimiento se repite hasta satisfacer algún criterio de convergencia establecido. Cuando converge este método, lo hace con orden 2, y requiere que el vector  $(x_0, y_0)$  esté muy cerca de la raíz buscada  $(\bar{x}, \bar{y})$ .



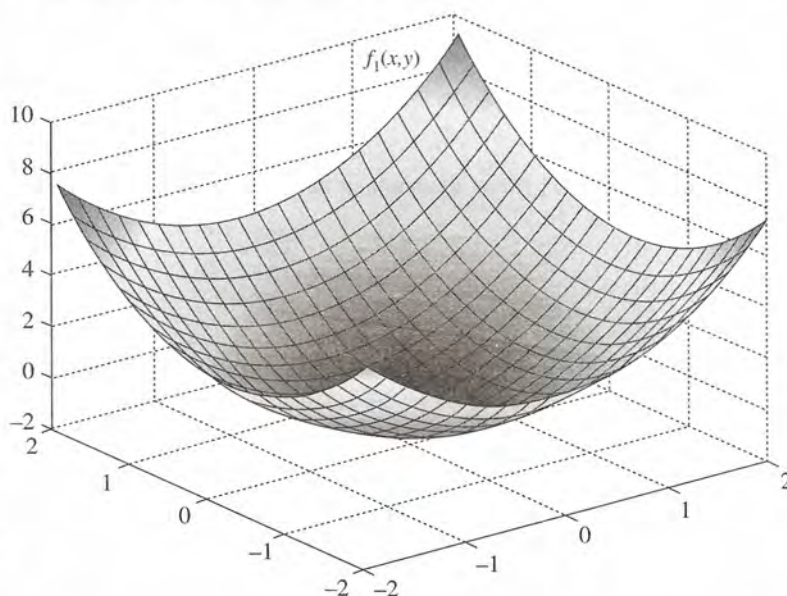
### INTERPRETACIÓN GEOMÉTRICA DEL MÉTODO DE NEWTON-RAPHSON

Desarrollemos en etapas esta interpretación para un sistema de dos ecuaciones. Sea el sistema

$$f_1(x, y) = x^2 + y^2 - 1$$

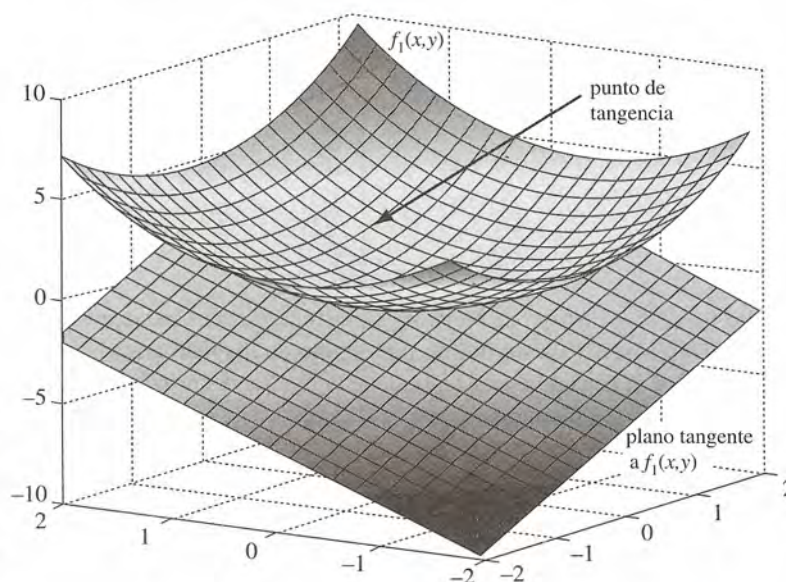
$$f_2(x, y) = x^2 - y^2 - 1$$

La gráfica de  $f_1(x, y)$  se muestra en la figura 4.4.



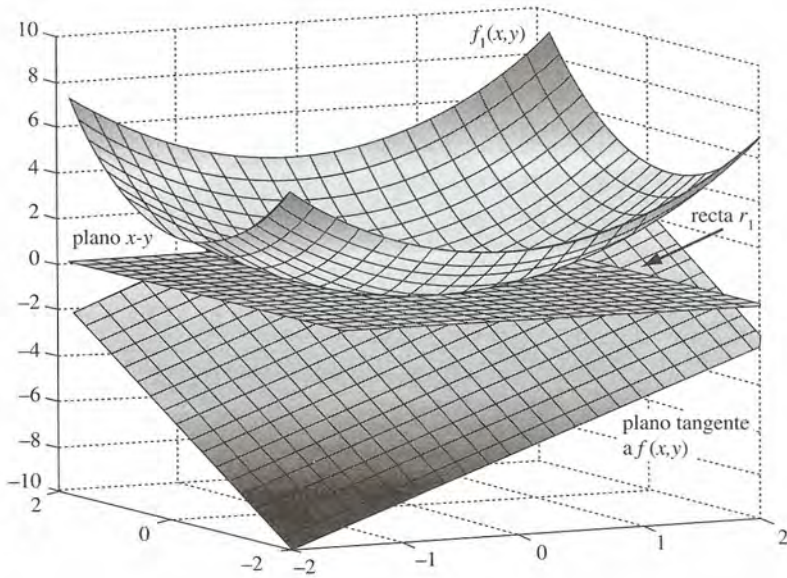
**Figura 4.4**  
Gráfica de la  
superficie  
 $f_1(x, y)$ .

Si el punto de inicio es  $(x_0, y_0) = (1, 1)$ , el plano tangente a la superficie  $f_1(x, y)$  en el punto  $(1, 1, f_1(1, 1))$  se muestra en la figura 4.5.



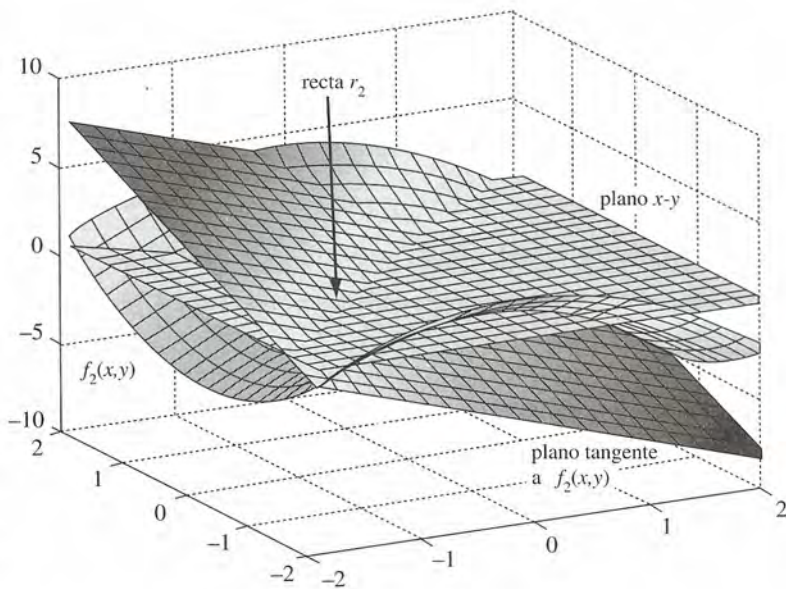
**Figura 4.5**  
Plano tangente  
a la superficie  
 $f_1(x, y)$  en el  
punto  $(1, 1, 1)$ .

La intersección del plano tangente con el plano  $x-y$ , en caso de existir, es una línea recta  $r_1$ , como puede apreciarse en la figura 4.6.



**Figura 4.6**  
Intersección del  
plano tangente  
y el plano  $x-y$ .

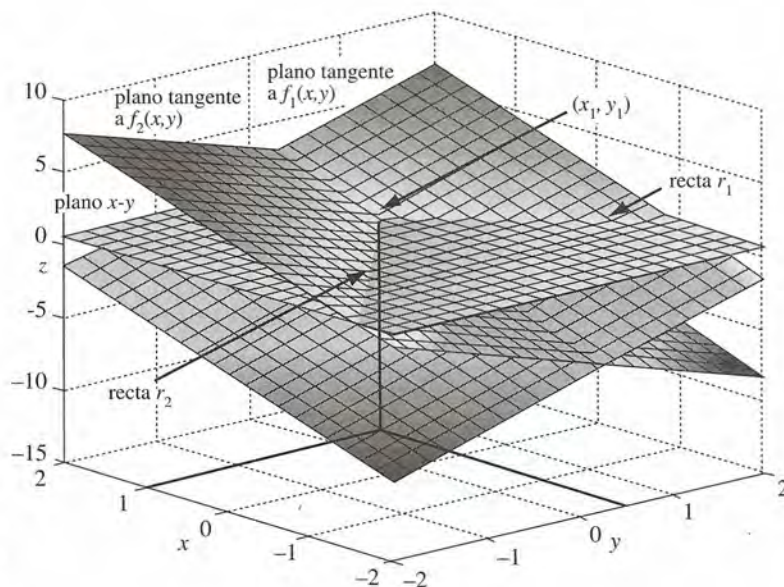
Si repetimos el procedimiento con la superficie  $f_2(x, y)$ , obtenemos la recta  $r_2$ , como puede apreciarse en la figura 4.7.



**Figura 4.7**  
Intersección del  
plano tangente  
y el plano  $x-y$   
para la función  
 $f_2(x, y)$ .



Finalmente, la intersección de los dos planos tangentes con el plano  $x$ - $y$ , en caso de existir, es la intersección de la recta  $r_1$  con la recta  $r_2$ , punto  $(x_1, y_1)$ , como puede verse en la figura 4.8, donde se han omitido las superficies  $f_1(x, y)$  y  $f_2(x, y)$  a fin de mostrar más claramente a  $(x_1, y_1)$ . Obsérvese que la intersección de las rectas  $r_1$  y  $r_2$  corresponde a la siguiente aproximación de una solución del sistema de ecuaciones no lineales.



**Figura 4.8**  
Intersección de  
los planos  
tangentes y el  
plano  $x$ - $y$ .

Podemos apreciar que este punto tiene aproximadamente las coordenadas  $(x_1, y_1) = (1, 0.5)$  y servirán como punto de partida para la siguiente iteración. El lector encontrará a continuación un inserto a color donde podrá observar mejor las intersecciones de las superficies y la raíz.

Un ejercicio interesante para el lector consistiría en identificar los pasos correspondientes entre la interpretación gráfica dada y la del método de Newton-Raphson univariable; por ejemplo, la recta tangente a la gráfica de  $f(x)$  en  $(x_0, f(x_0))$  corresponde a la superficie tangente a una de las gráficas del sistema en  $(x_0, y_0, f(x_0, y_0))$ , etcétera.

Use el método de Newton-Raphson para encontrar una solución aproximada del sistema.

#### Ejemplo 4.3

$$f_1(x, y) = x^2 - 10x + y^2 + 8 = 0$$

$$f_2(x, y) = xy^2 + x - 10y + 8 = 0$$

con el vector inicial:  $[x^0, y^0]^T = [0, 0]^T$ .

**Solución** Primero se forma la matriz coeficiente del sistema 4.13, también conocida como matriz de derivadas parciales

$$\begin{bmatrix} \frac{\partial f_1}{\partial x} = 2x - 10 & \frac{\partial f_1}{\partial y} = 2y \\ \frac{\partial f_2}{\partial x} = y^2 + 1 & \frac{\partial f_2}{\partial y} = 2xy \end{bmatrix} \quad , \quad 10$$



que aumentada en el vector de funciones resulta en:

$$\left[ \begin{array}{cc|c} 2x - 10 & 2y & -x^2 + 10x - y^2 - 8 \\ y^2 + 1 & 2xy - 10 & -xy^2 - x + 10y - 8 \end{array} \right]$$

### Primera iteración

Al evaluar la matriz en  $[x^0, y^0]^T$  se obtiene:

$$\left[ \begin{array}{cc|c} -10 & 0 & -8 \\ 1 & -10 & -8 \end{array} \right]$$

que al resolverse por eliminación de Gauss da

$$h = 0.8, \quad j = 0.88$$

al sustituir en la ecuación 4.12 se obtiene:

$$x^1 = x^0 + h = 0 + 0.8 = 0.8$$

$$y^1 = y^0 + j = 0 + 0.88 = 0.88$$

Cálculo de la distancia entre  $\mathbf{x}^{(0)}$  y  $\mathbf{x}^{(1)}$

$$|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}| = \sqrt{(0.8 - 0)^2 + (0.88 - 0)^2} = 1.18929$$

### Segunda iteración

Al evaluar la matriz en  $[x^1, y^1]^T$  resulta

$$\left[ \begin{array}{cc|c} -8.4 & 1.76 & -1.41440 \\ 1.7744 & -8.592 & -0.61952 \end{array} \right]$$

que por eliminación gaussiana da como nuevos resultados de  $h$  y  $j$

$$h = 0.19179, \quad j = 0.11171$$

de donde:

$$x^2 = x^1 + h = 0.8 + 0.19179 = 0.99179$$

$$y^2 = y^1 + j = 0.88 + 0.11171 = 0.99171$$

Cálculo\* de la distancia entre  $x^{(1)}$  y  $x^{(2)}$ :

$$|\mathbf{x}^{(2)} - \mathbf{x}^{(1)}| = \sqrt{(0.99179 - 0.8)^2 + (0.99171 - 0.88)^2} = 0.22190$$

Con la continuación de este proceso iterativo se obtienen los resultados siguientes:

| $k$ | $x^k$   | $y^k$   | $ \mathbf{x}^{k+1} - \mathbf{x}^k $ |
|-----|---------|---------|-------------------------------------|
| 0   | 0.00000 | 0.00000 | —                                   |
| 1   | 0.80000 | 0.88000 | 1.18929                             |
| 2   | 0.99179 | 0.99171 | 0.22195                             |
| 3   | 0.99998 | 0.99997 | 0.01163                             |
| 4   | 1.00000 | 1.00000 | 0.00004                             |

\* Nótese que  $|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}| = \sqrt{h^2 + j^2}$

Hay que observar que se requirieron cuatro iteraciones para llegar al vector solución (1, 1) contra once del ejemplo 4.2, donde se usó el método de punto fijo con desplazamientos sucesivos. Sin embargo, esta convergencia cuadrática implica mayor número de cálculos, ya que, como se puede observar, en cada iteración se requiere:

- a) La evaluación de  $2 \times 2$  derivadas parciales.
- b) La evaluación de 2 funciones.
- c) La solución de un sistema de ecuaciones lineales de orden 2.

Estos cálculos pueden realizarse con Matlab o con la TI-92 Plus.

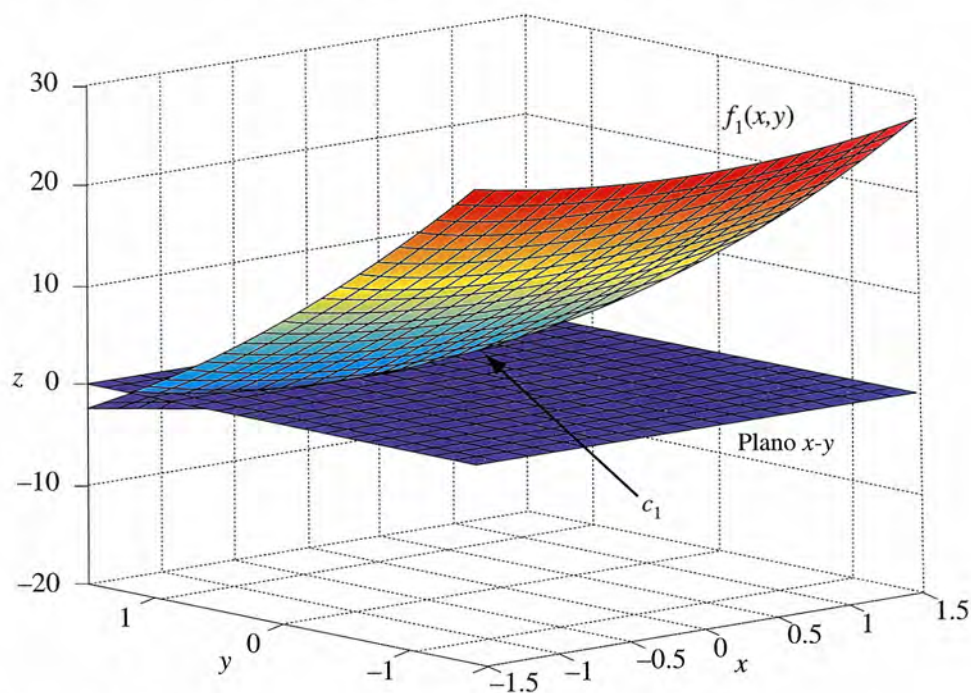


```
x0=0; y0=0;
fprintf('    k    x(k)    y(k)    |x(k+1)-x(k)| \n')
fprintf('    %2d    %10.5f    %10.5f \n', 0, x0, y0)
for k=1 : 4
    df1x=2*x0-10;    df1y=2*y0;
    df2x=y0^2+1;    df2y=2*x0*y0-10;
    f1=x0^2-10*x0+y0^2+8;
    f2=x0*y0^2+x0-10*y0+8;
    A=[df1x df1y; df2x df2y];
    b=[-f1; -f2];
    hj=inv(A)*b;
    x1=x0+hj(1); y1=y0+hj(2);
    Dist=((x1-x0)^2+(y1-y0)^2)^0.5;
    fprintf('    %2d    %10.5f    %10.5f %10.5f \n', k, x1, y1, Dist)
    x0=x1; y0=y1;
end
```

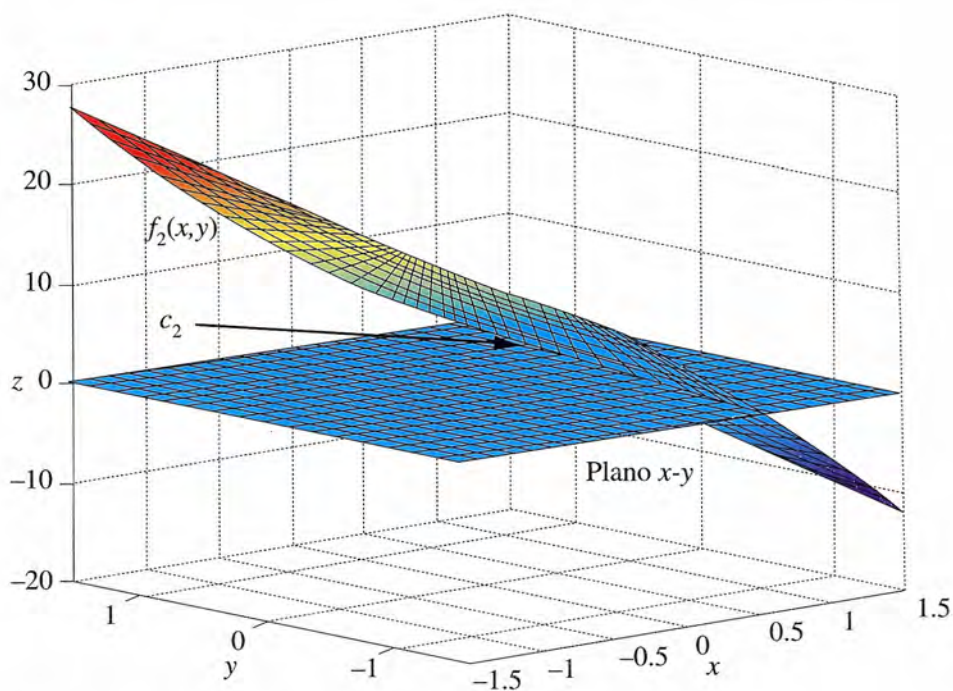


```
e4_3 ( )
Prgm
Define f1(x, y) = x^2-10*x+y^2+8
Define f2(x, y) = x*y^2+x-10*y+8
Define df1x(x, y) = 2*x-10
Define df1y(x, y) = 2*y
Define df2x(x, y) = y^2+1
Define df2y(x, y) = 2*x*y-10;
0→x0 : 0→y0 : 0→k : ClrIO
Disp "k x(k) y(k) |x(k+1)-x(k)|"
string(k) &" "&format(x0, "f5")→d
d&" "&format(y0, "f5")→d : Disp d
For k, 1, 4
    [df1x(x0,y0),df1y(x0,y0); df2x(x0,y0),df2y(x0,y0)]→a
    [-f1(x0,y0); -f2(x0,y0)]→b
    simult(a,b)→dx : x0+dx[1]→x : y0+dx[2]→y
    norm(dx)→dist : norm(x)→x : norm(y)→y
    string(k) &" "&format(x, "f5")→d
    d&" "&format(y1, "f5")→d : Disp d
    x→x0 : y→y0
EndFor
EndPrgm
```

Con el fin de ayudar al lector a visualizar la idea de raíz y el método de Newton-Raphson, se presentan a continuación las gráficas de las Fig. 4.1 a la Fig. 4.8 a color.

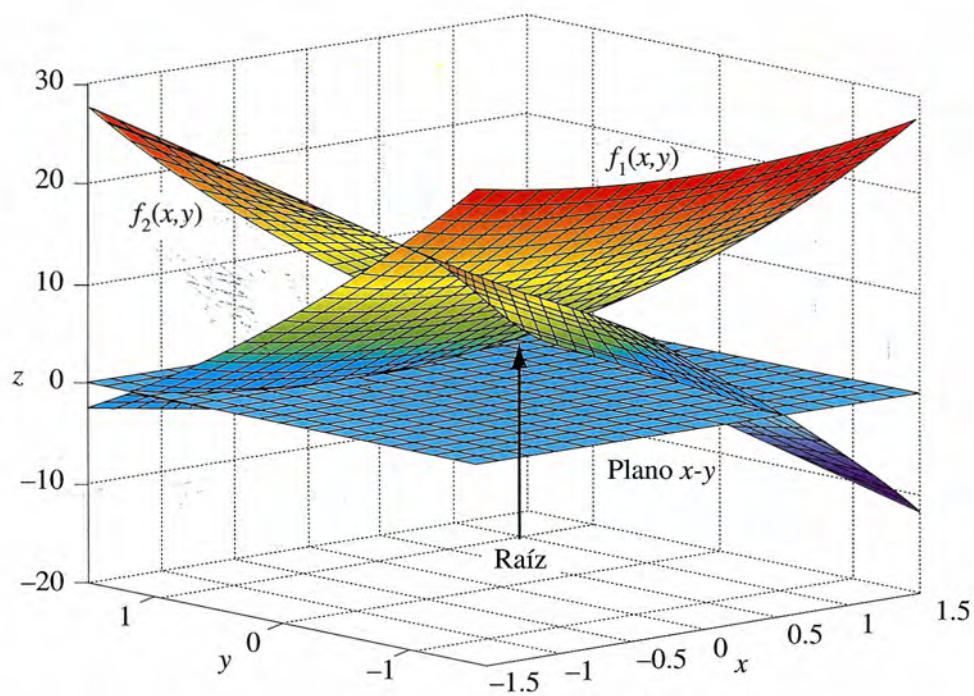


**Figura 4.1** Intersección de la superficie  $f_1(x, y)$  con el plano  $x-y$ .

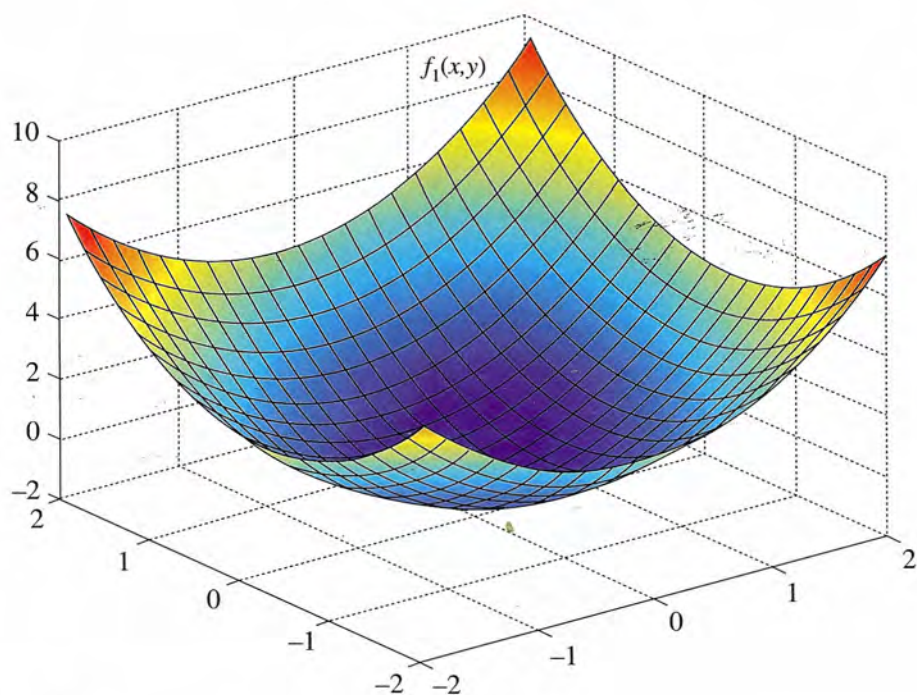


**Figura 4.2** Intersección de la superficie  $f_2(x, y)$  con el plano  $x-y$ .

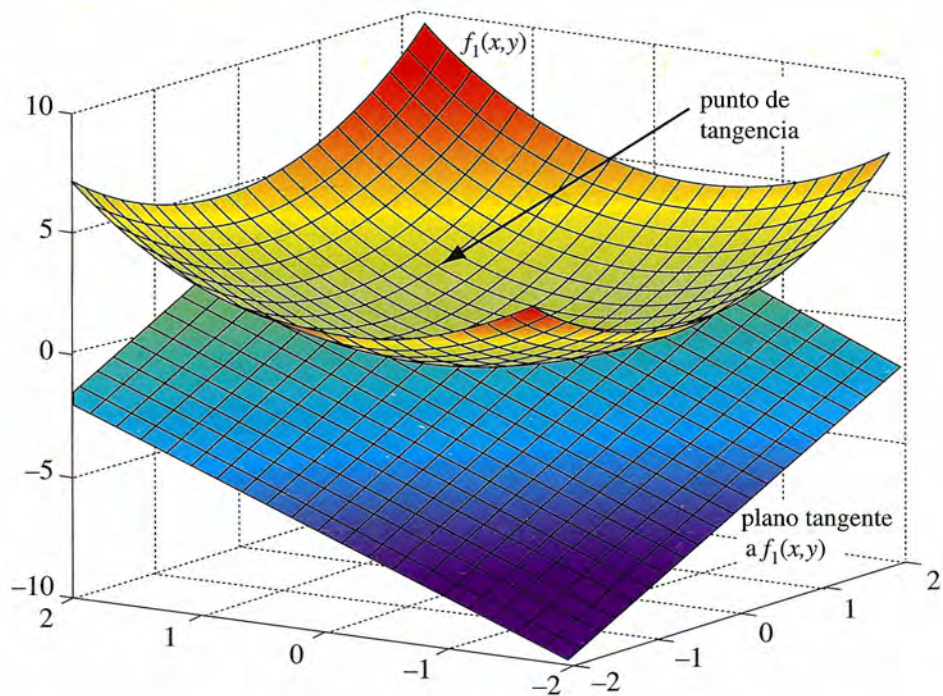




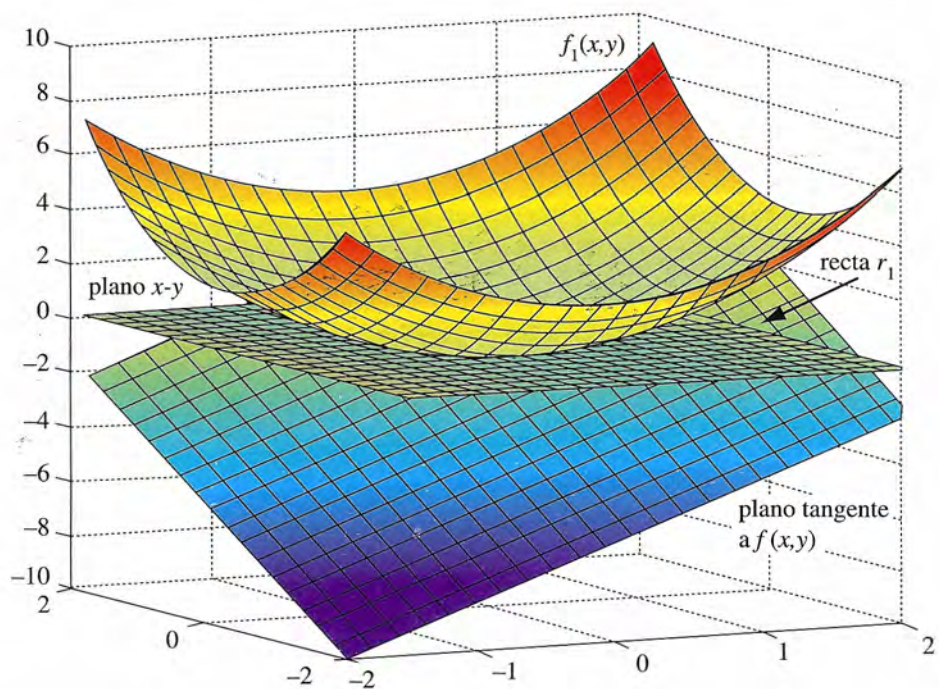
**Figura 4.3** Intersección de las superficies  $f_1(x, y)$  y  $f_2(x, y)$  con el plano  $x-y$ .



**Figura 4.4** Gráfica de la superficie  $f_1(x, y)$ .

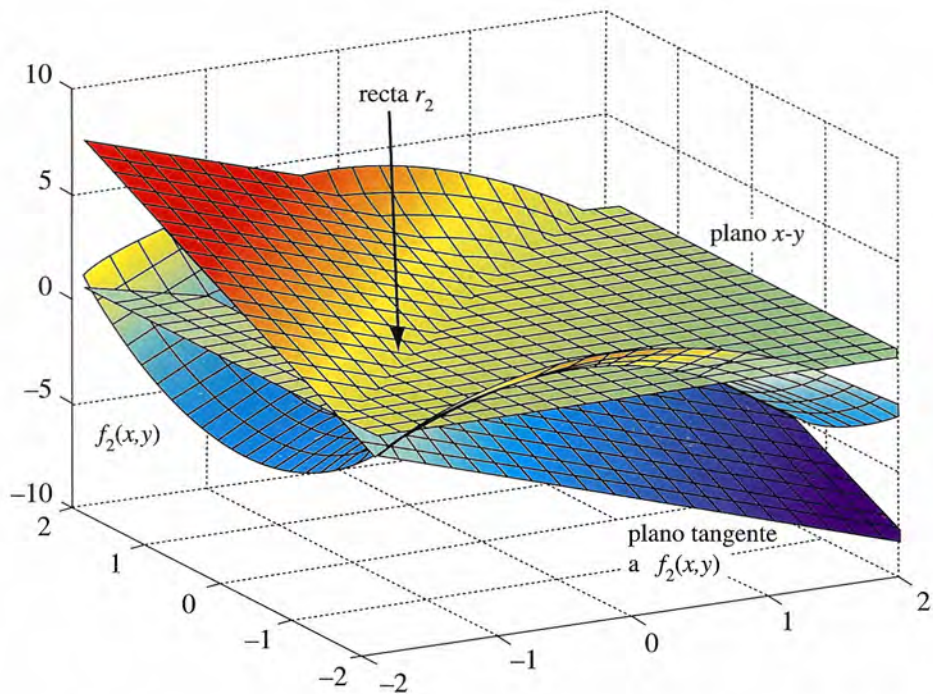


**Figura 4.5** Plano tangente a la superficie  $f_1(x, y)$  en el punto  $(1, 1, 1)$ .

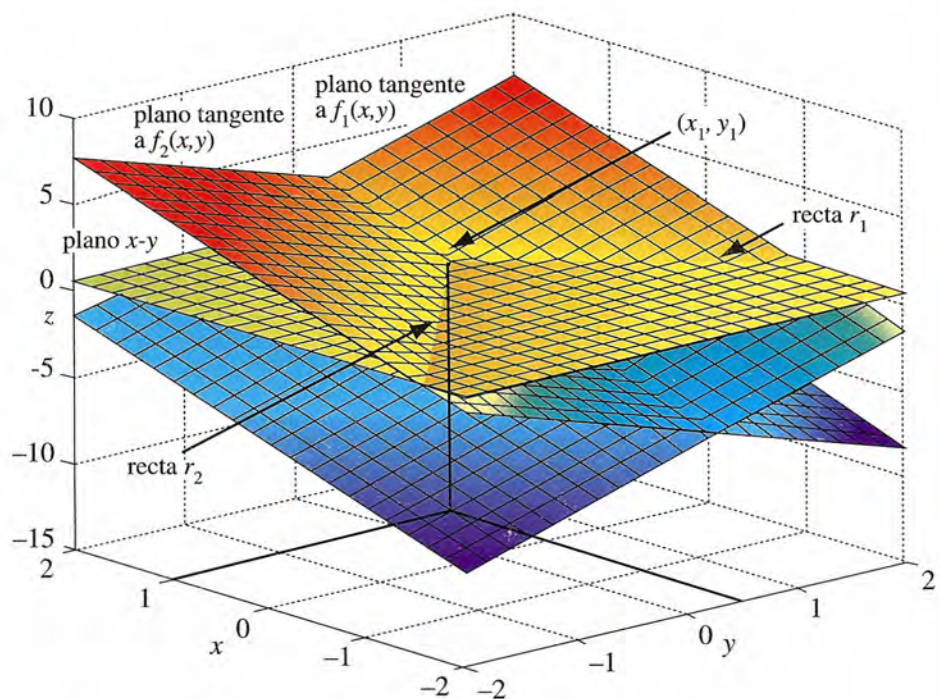


**Figura 4.6** Intersección del plano tangente y el plano x-y.





**Figura 4.7** Intersección del plano tangente y el plano  $x$ - $y$  para la función  $f_2(x, y)$ .



**Figura 4.8** Intersección de los planos tangentes y el plano  $x$ - $y$ .



**GENERALIZACIÓN**

Para un sistema de  $n$  ecuaciones no lineales con  $n$  incógnitas (véase Ec. 4.1) y retomando la notación vectorial y matricial, las ecuaciones 4.13 quedan:

$$\begin{array}{ccccccc} \frac{\partial f_1}{\partial x_1} & h_1 & + & \frac{\partial f_1}{\partial x_2} & h_2 & + & \dots + \frac{\partial f_1}{\partial x_n} h_n = -f_1 \\ \frac{\partial f_2}{\partial x_1} & h_1 & + & \frac{\partial f_2}{\partial x_2} & h_2 & + & \dots + \frac{\partial f_2}{\partial x_n} h_n = -f_2 \\ & \vdots & & & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & h_1 & + & \frac{\partial f_n}{\partial x_2} & h_2 & + & \dots + \frac{\partial f_n}{\partial x_n} h_n = -f_n \end{array} \quad (4.14)$$

o

$$J \mathbf{h} = -\mathbf{f}$$

donde las funciones  $f_i$  y las derivadas parciales  $\partial f_i / \partial x_j$ ,  $i = 1, 2, \dots, n$ ;  $j = 1, 2, \dots, n$  están evaluadas en el vector  $\mathbf{x}^{(k)}$  y

$$h_i = x_i^{k+1} - x_i^k \quad 1 \leq i \leq n \quad (4.15)$$

De donde:

$$x_i^{k+1} = x_i^k + h_i \quad 1 \leq i \leq n \quad (4.16)$$

o

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{h}^{(k)}$$

y la matriz de derivadas parciales (matriz jacobiana), ampliada en el vector de funciones queda:

$$\left[ \begin{array}{cccc|c} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \dots & \frac{\partial f_1}{\partial x_n} & -f_1 \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \dots & \frac{\partial f_2}{\partial x_n} & -f_2 \\ \vdots & \vdots & & \vdots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \dots & \frac{\partial f_n}{\partial x_n} & -f_n \end{array} \right] \quad (4.17)$$

o bien:

$$[J | -\mathbf{f}]$$

Se presenta a continuación un algoritmo para este método.

**ALGORITMO 4.2** Método de Newton-Raphson multivariable

Para encontrar una solución aproximada de un sistema de ecuaciones no lineales  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ , proporcionar la matriz jacobiana ampliada con el vector de funciones (véase Ec. 4.17) y los

DATOS: El número de ecuaciones  $N$ , el vector de valores iniciales  $\mathbf{x}$ , el número máximo de iteraciones MAXIT y el criterio de convergencia EPS.

RESULTADOS: El vector solución  $\mathbf{xn}$  o mensaje "NO CONVERGE".

PASO 1. Hacer  $K = 1$ .

PASO 2. Mientras  $K \leq \text{MAXIT}$ , repetir los pasos 3 a 9.

PASO 3. Evaluar la matriz jacobiana aumentada (4.17).

PASO 4. Resolver el sistema lineal (4.14).

PASO 5. Hacer\*  $\mathbf{xn} = \mathbf{x} + \mathbf{h}$ .

PASO 6. Si  $\|\mathbf{xn} - \mathbf{x}\| > \text{EPS}$  ir al paso 8. De otro modo continuar.

PASO 7. IMPRIMIR  $\mathbf{xn}$  y TERMINAR.

PASO 8. Hacer  $\mathbf{x} = \mathbf{xn}$ .

PASO 9. Hacer  $K = K + 1$ .

PASO 10. IMPRIMIR "NO CONVERGE" Y TERMINAR.

### Ejemplo 4.4

Con algoritmo 4.2, elabore un programa de propósito general para resolver sistemas de ecuaciones no lineales. Luego úselo para resolver el sistema.

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2x_3) - 0.5 = 0$$

$$f_2(x_1, x_2, x_3) = x_1^2 - 625x_2^2 = 0$$

$$f_3(x_1, x_2, x_3) = e^{-x_1x_2} + 20x_3 + \frac{(10\pi - 3)}{3} = 0$$

### Solución



En el disco se presenta el programa 4.1, que consta de los subprogramas GAUSSJORDAN y PIVOTEO, de propósito general; es decir, no dependen del sistema de ecuaciones para resolver.

El usuario deberá escribir el programa principal que llama al subprograma FUNCIONES, donde proporcionará la matriz jacobiana ampliada (Ec. 4.17).

La matriz jacobiana ampliada para el sistema es:

$$\begin{bmatrix} 3 & x_3 \sin(x_2x_3) & x_2 \sin(x_2x_3) & -3x_1 + \cos(x_2x_3) + 0.5 \\ 2x_1 & -1250x_2 & 0 & -x_1^2 + 625x_2^2 \\ -x_2e^{-x_1x_2} & -x_1e^{-x_1x_2} & 20 & -e^{-x_1x_2} - 20x_3 - \frac{10\pi - 3}{3} \end{bmatrix}$$

El programa queda finalmente como se muestra en el disco (programa 4.1). Su ejecución con el vector inicial  $[1 \ 1 \ 1]^T$  produce los siguientes resultados.

| $k$ | $x_1$   | $x_2$   | $x_3$    | Distancia   |
|-----|---------|---------|----------|-------------|
| 0   | 1.00000 | 1.00000 | 1.00000  | —           |
| 1   | 0.90837 | 0.50065 | -0.50286 | 1.5863      |
| 2   | 0.49927 | 0.25046 | -0.51904 | 0.47982     |
| 3   | 0.49996 | 0.12603 | -0.52045 | 0.12444     |
| 4   | 0.49998 | 0.06460 | -0.52199 | 0.61446E-01 |
| 5   | 0.49998 | 0.03540 | -0.52272 | 0.29214E-01 |
| 6   | 0.49998 | 0.02335 | -0.52302 | 0.12052E-01 |
| 7   | 0.49998 | 0.02024 | -0.52309 | 0.31095E-02 |
| 8   | 0.49998 | 0.02000 | -0.52310 | 0.23879E-03 |
| 9   | 0.49998 | 0.02000 | -0.52310 | 0.14280E-05 |

\* Operaciones vectoriales.

La solución del sistema es:

$$\begin{aligned}x_1 &= 0.49998176 \\x_2 &= 0.19999269\text{E}-01 \\x_3 &= -0.52310085\end{aligned}$$

Los cálculos también pueden realizarse usando el guión de Matlab dado en el ejemplo 4.3, con los cambios correspondientes.

Nótese que en cada iteración se requiere:

- La evaluación de  $n^2$  derivadas parciales,
- La evaluación de  $n$  funciones,
- La solución de un sistema de ecuaciones lineales de orden  $n$ ,

lo que representa una inmensa cantidad de cálculos. Debido a esto, se han elaborado métodos donde los cálculos no son tan numerosos y cuya convergencia es, en general, superior a la del método de punto fijo (superlineal). A continuación se presentan dos de estos métodos, el de Newton-Raphson modificado y el método de Broyden, siendo este último también una modificación del método de Newton-Raphson.

## 4.4 Método de Newton-Raphson modificado

El método de Newton Raphson modificado que se describe a continuación consiste en aplicar el método de Newton-Raphson univariable dos veces (para el caso de un sistema de  $n$  ecuaciones no lineales con  $n$  incógnitas, se aplicará  $n$  veces), una para cada variable. Cada vez que se hace esto, se consideran las otras variables fijas.

Considérese de nuevo el sistema

$$\begin{aligned}f_1(x, y) &= 0 \\f_2(x, y) &= 0\end{aligned}$$

Tomando los valores iniciales  $x^0, y^0$ , se calcula a partir del método de Newton-Raphson univariable un nuevo valor  $x^1$  así:

$$x^1 = x^0 - \frac{f_1(x^0, y^0)}{\partial f_1 / \partial x},$$

$\partial f_1 / \partial x$  evaluada en  $x^0, y^0$ .

Hay que observar que se ha obtenido  $x^1$  a partir de  $f_1$  y los valores más recientes de  $x$  y  $y$ :  $x^0, y^0$ .

Ahora emplearemos  $f_2$  y los valores más recientes de  $x$  y  $y$ :  $x^1, y^0$  para calcular  $y^1$

$$y^1 = y^0 - \frac{f_2(x^1, y^0)}{\partial f_2 / \partial y},$$

donde  $\partial f_2 / \partial y$  se evalúa en  $x^1, y^0$ . Se tiene ahora  $x^1$  y  $y^1$ . Con estos valores se calcula  $x^2$ , después  $y^2$ , y así sucesivamente.

Este método converge a menudo si  $x^0, y^0$  está muy cerca de  $\bar{x}, \bar{y}$ , y requiere la evaluación de sólo  $2n$  funciones por paso (cuatro para el caso de dos ecuaciones que se está manejando).

Hay que observar que se han empleado desplazamientos sucesivos, pero los desplazamientos simultáneos también son aplicables.



**Ejemplo 4.5**

Resuelva el sistema

$$f_1(x, y) = x^2 - 10x + y^2 + 8 = 0$$

$$f_2(x, y) = xy^2 + x - 10y + 8 = 0$$

Con el método de Newton-Raphson modificado, usando los valores iniciales  $x^0 = 0$ ,  $y^0 = 0$ . Puede seguir los cálculos con un pizarrón electrónico.

**Solución**

Primero se obtiene:



$$\frac{\partial f_1}{\partial x} = 2x - 10 \quad y \quad \frac{\partial f_2}{\partial y} = 2xy - 10$$

**Primera iteración**Se evalúan  $f_1$  y  $\partial f_1 / \partial x$  en  $[0, 0]^T$ :

$$f_1(0, 0) = 8$$

y

$$\left. \frac{\partial f_1}{\partial x} \right|_{\substack{x^0 \\ y^0}} = -10$$

se sustituye

$$x^1 = 0 - \frac{8}{-10} = 0.8$$

Para el cálculo de  $y^1$  se necesita evaluar  $f_2$  y  $\partial f_2 / \partial y$  en  $x^1, y^0$ 

$$f_2(0.8, 0) = 0.8(0) + 0.8 - 10(0) + 8 = 8.8$$

$$\left. \frac{\partial f_2}{\partial y} \right|_{\substack{x^1 \\ y^0}} = 2(0.8)(0) - 10 = -10$$

se sustituye:

$$y^1 = 0 - \frac{8.8}{-10} = 0.88$$

**Segunda iteración**

$$f_1(0.8, 0.88) = 1.4144 \quad y \quad \left. \frac{\partial f_1}{\partial x} \right|_{\substack{x^1 \\ y^1}} = -8.4$$

$$x^2 = 0.8 - \frac{1.4144}{-8.4} = 0.96838$$

Ahora se evalúan  $f_2$  y  $\partial f_2 / \partial y$  en  $(x^2, y^1)$ :

$$f_2(0.96838, 0.88) = 0.91830 \quad y \quad \left. \frac{\partial f_2}{\partial y} \right|_{\substack{x^2 \\ y^1}} = -8.29565$$

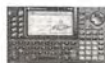
de donde:

$$y^2 = 0.88 - \frac{0.91830}{-8.59565} = 0.99070$$

Los cálculos pueden continuarse y observarse con Matlab o con la TI-92 Plus



```
x0=0; y0=0; Eps=1e-5;
fprintf(' k x(k) y(k) Dist\n')
fprintf(' %2d %10.5f %10.5f\n', 0, x0, y0)
for k=1 : 10
    f1=x0^2-10*x0+y0^2+8;
    df1x=2*x0-10;
    x1=x0-f1/df1x;
    f2=x1*y0^2+x1-10*y0+8;
    df2y=2*x1*y0-10;
    y1=y0-f2/df2y;
    fprintf(' %2d%10.5f%10.5f%10.5f\n', k, x1, y1, Dist)
    Dist=((x1-x0)^2+(y1-y0)^2)^0.5;
    If Dist < Eps
        Break
    end
    x0=x1; y0=y1;
end
```



```
e4_5 ( )
Prgm
Define f1(x,y) = x^2-10*x+y^2+8
Define f2(x,y) = x*y^2+x-10*y+8
Define df1x(x,y) = 2*x-10
Define df2y(x,y) = 2*x*y-10
0→x0 : 0→y0 : 0→k : 1E-5→eps : ClrIO
Disp "k x(k) y(k) |x(k+1)-x(k)|"
string(k) &" "&format(x0, "f5")→d
d&" "&format(y0, "f5")→d : Disp d
for k, 1, 6
    x0-f1(x0,y0)/df1x(x0,y0)→x
    y0-f2(x0,y0)/df2y(x0,y0)→y
    √((x-x0)^2+(y-y0)^2)→dist
    string(k) &" "&format(x, "f5")→d
    d&" "&format(y, "f5")→d
    d&" "&format(dist, "f5")→d : Disp d
    x→x0 : y→y0
EndFor
EndPrgm
```

Se sugiere al lector continuar las iteraciones y calcular las distancias entre cada dos vectores consecutivos. Continúe hasta que  $x^k \approx 1$  y  $y^k \approx 1$ . Compare además la velocidad de convergencia de este método con la velocidad de convergencia del método de Newton-Raphson y el de punto fijo para este sistema particular.

En la aplicación de este método se pudo tomar  $f_2$  para evaluar  $x^1$  y  $f_1$ , a fin de evaluar  $y^1$ , así:

$$\begin{aligned} x^1 &= x^0 - \frac{f_2(x^0, y^0)}{\partial f_2 / \partial x}, \\ y^1 &= y^0 - \frac{f_1(x^1, y^0)}{\partial f_1 / \partial y}, \end{aligned}$$

Esto puede producir convergencia en alguno de los arreglos y divergencia en el otro. Es posible saber de antemano si la primera o la segunda forma convergirán\* para el caso de sistemas de dos ecuaciones, pero cuando  $3 \leq n$  las posibilidades son varias ( $n!$ ) y es imposible conocer cuál de estos arreglos tiene viabilidad de convergencia, por lo cual la elección se convierte en un proceso aleatorio. Esta aleatoriedad es la mayor desventaja de este método.

En general, para un sistema de  $n$  ecuaciones con  $n$  incógnitas:  $x_1, x_2, \dots, x_n$ , el algoritmo toma la forma:

$$x_i^{k+1} = x_i^k - \frac{f_i(x_1^{k+1}, x_2^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_n^k)}{\left. \frac{\partial f_i}{\partial x_i} \right| (x_1^{k+1}, x_2^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_n^k)} \quad 1 \leq i \leq n \quad (4.18)$$

#### ALGORITMO 4.3 Método de Newton-Raphson modificado

Para encontrar una solución aproximada de un sistema de ecuaciones no lineales  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ , proporcionar las funciones  $F(I, \mathbf{x})$  y las derivadas parciales  $D(I, \mathbf{x})$  y los

**DATOS:** El número de ecuaciones  $N$ , el vector de valores iniciales  $\mathbf{x}$ , el número máximo de iteraciones  $MAXIT$ , el criterio de convergencia  $EPS$  y  $M=0$  para desplazamientos sucesivos o  $M=1$  para desplazamientos simultáneos.

**RESULTADOS:** El vector solución  $\mathbf{x}_n$  o mensaje "NO CONVERGE".

PASO 1. Hacer  $K=1$ .

PASO 2. Mientras  $K \leq MAXIT$ , repetir los pasos 3 a 11.

PASO 3. Si  $M = 0$  hacer\*  $\mathbf{x}_{aux} = \mathbf{x}$ .

PASO 4. Hacer  $I = 1$ .

PASO 5. Mientras  $I \leq N$ , repetir los pasos 6 y 7.

PASO 6. Si  $M = 0$  hacer.

$$X(I) = X(I) - F(I, \mathbf{x}) / D(I, \mathbf{x}).$$

De otro modo hacer:

$$X_{aux}(I) = X(I) - F(I, \mathbf{x}) / D(I, \mathbf{x}).$$

PASO 7. Hacer  $I = I + 1$ .

PASO 8. Si  $\|\mathbf{x}_{aux} - \mathbf{x}\| > EPS$  ir al paso 10.

De otro modo continuar.

PASO 9. IMPRIMIR  $\mathbf{x}$  y TERMINAR.

PASO 10. Si  $M=1$  hacer  $\mathbf{x} = \mathbf{x}_{aux}$ .

PASO 11. Hacer  $K = K + 1$ .

PASO 12. IMPRIMIR "NO CONVERGE" Y TERMINAR.

\* Operaciones vectoriales.

El método siguiente puede saltarse sin pérdida de continuidad.

\* Peter A. Stark. *Introduction to Numeral Methods*. Ed. McMillan.



## 4.5 Método de Broyden

Considérese ahora la generalización del método de la secante a sistemas multivariantes, conocido como el método de Broyden. Según se vio en el capítulo 2, el método de la secante consiste en remplazar  $f'(x_k)$  del método de Newton-Raphson

$$x_{k+1} = x_k - [f'(x_k)]^{-1} f(x_k) \quad (4.19)$$

por el cociente:

$$\frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}} \approx f'(x_k),$$

obtenido con los resultados de dos iteraciones previas:  $x_k$  y  $x_{k+1}$ .

Para ver la modificación o aproximación correspondiente del método de Newton-Raphson multivariable, conviene expresarlo primero en forma congruente con la ecuación 4.19, lo que se logra substituyendo en la ecuación vectorial (véase Ec. 4.16)

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{h}^{(k)} \quad (4.20)$$

el vector  $\mathbf{h}^{(k)}$  que, como se sabe, es la solución del sistema

$$J^{(k)} \mathbf{h}^{(k)} = -\mathbf{f}^{(k)}$$

Al multiplicar esta última ecuación por  $(J^{(k)})^{-1}$  se obtiene:

$$\mathbf{h}^{(k)} = -(J^{(k)})^{-1} \mathbf{f}^{(k)} \quad (4.21)$$

y al remplazar la ecuación 4.21 en la 4.20 se llega a:

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - (J^{(k)})^{-1} \mathbf{f}^{(k)} \quad (4.22)$$

la ecuación correspondiente a la 4.14 para  $n > 1$ .

El método de la secante para sistemas de ecuaciones no lineales consiste en substituir  $J^{(k)}$  en la ecuación 4.22 con una matriz  $A^{(k)}$ , cuyos componentes se obtienen con los resultados de dos iteraciones previas  $\mathbf{x}^{(k)}$  y  $\mathbf{x}^{(k-1)}$ , de la siguiente manera\*:

$$A^{(k)} = A^{(k-1)} + \frac{[\mathbf{f}(\mathbf{x}^{(k)}) - \mathbf{f}(\mathbf{x}^{(k-1)}) - A^{(k-1)} (\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})] (\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)})^T}{\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\|^2} \quad (4.23)$$

o bien:

$$A^{(k)} = A^{(k-1)} + \frac{[\Delta \mathbf{f}^{(k)} - A^{(k-1)} \Delta \mathbf{x}^{(k)}] (\Delta \mathbf{x}^{(k)})^T}{\|\Delta \mathbf{x}^{(k)}\|^2} \quad (4.24)$$

con la notación

$$\begin{aligned} \Delta \mathbf{f}^{(k)} &= \mathbf{f}(\mathbf{x}^{(k)}) - \mathbf{f}(\mathbf{x}^{(k-1)}) \\ \Delta \mathbf{x}^{(k)} &= \mathbf{x}^{(k)} - \mathbf{x}^{(k-1)} \end{aligned}$$

Para la primera aplicación de la ecuación 4.24 se requieren dos vectores iniciales:  $\mathbf{x}^{(0)}$  y  $\mathbf{x}^{(1)}$ . Este último puede obtenerse de una aplicación del método de Newton-Raphson multivariable

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} - (J^{(0)})^{-1} \mathbf{f}^{(0)},$$

\* Dennis, J.E. Jr. y J.J. More (1977), "Quasi-Newton Methods, Motivation and Theory". *SIAM Review*, 19, No. 1 (46-89).

cuya  $J^{(0)}$  a su vez puede emplearse en 4.24, con lo cual ésta queda:

$$A^{(1)} = J^{(0)} + \frac{(\Delta \mathbf{f}^{(1)} - J^{(0)} \Delta \mathbf{x}^{(1)}) (\Delta \mathbf{x}^{(1)})^T}{|\Delta \mathbf{x}^{(1)}|^2} \quad (4.25)$$

La inversión de  $A^{(k)}$  en cada iteración significa un esfuerzo computacional grande (del orden de  $n^3$ ) que, sin embargo, puede reducirse empleando una fórmula de inversión matricial de Sherman y Morrison.\* Esta fórmula establece que si  $A$  es una matriz no singular y  $\mathbf{x}$  y  $\mathbf{y}$  son vectores, entonces  $A + \mathbf{x}\mathbf{y}^T$  es no singular, siempre que  $\mathbf{y}^T A^{-1} \mathbf{x} \neq 1$ . Además, en este caso,

$$(A + \mathbf{x}\mathbf{y}^T)^{-1} = A^{-1} - \frac{A^{-1} \mathbf{x}\mathbf{y}^T A^{-1}}{1 + \mathbf{y}^T A^{-1} \mathbf{x}} \quad (4.26)$$

Esta fórmula permite calcular  $(A^{(k)})^{-1}$  a partir de  $(A^{(k-1)})^{-1}$ , eliminando la necesidad invertir una matriz en cada iteración. Para esto, primero se obtiene la inversa de la ecuación 4.24.

$$(A^{(k)})^{-1} = (A^{(k-1)})^{-1} + \frac{(\Delta \mathbf{f}^{(k)} - A^{(k-1)} \Delta \mathbf{x}^{(k)}) (\Delta \mathbf{x}^{(k)})^T}{|\Delta \mathbf{x}^{(k)}|^2}$$

Después se hace:

$$\begin{aligned} A &= A^{(k-1)} \\ \mathbf{x} &= \frac{(\Delta \mathbf{f}^{(k)} - A^{(k-1)} \Delta \mathbf{x}^{(k)})}{|\Delta \mathbf{x}^{(k)}|^2} \end{aligned}$$

y

$$\mathbf{y} = \Delta \mathbf{x}^{(k)},$$

con lo que la última ecuación queda:

$$(A^{(k)})^{-1} = (A + \mathbf{x}\mathbf{y}^T)^{-1}$$

y sustituyendo la ecuación 4.26

$$\begin{aligned} (A^{(k)})^{-1} &= (A^{(k-1)})^{-1} - \frac{(A^{(k-1)})^{-1} \left[ \frac{\Delta \mathbf{f}^{(k)} - A^{(k-1)} \Delta \mathbf{x}^{(k)}}{|\Delta \mathbf{x}^{(k)}|^2} (\Delta \mathbf{x}^{(k)})^T \right] (A^{(k-1)})^{-1}}{1 + (\Delta \mathbf{x}^{(k)})^T (A^{(k-1)})^{-1} \frac{\Delta \mathbf{f}^{(k)} - A^{(k-1)} \Delta \mathbf{x}^{(k)}}{|\Delta \mathbf{x}^{(k)}|^2}} \\ &= (A^{(k-1)})^{-1} - \frac{[(A^{(k-1)})^{-1} \Delta \mathbf{f}^{(k)} - \Delta \mathbf{x}^{(k)}] (\Delta \mathbf{x}^{(k)})^T (A^{(k-1)})^{-1}}{|\Delta \mathbf{x}^{(k)}|^2 + (\Delta \mathbf{x}^{(k)})^T (A^{(k-1)})^{-1} \Delta \mathbf{f}^{(k)} - |\Delta \mathbf{x}^{(k)}|^2} \\ &\boxed{(A^{(k)})^{-1} = (A^{(k-1)})^{-1} + \frac{[\Delta \mathbf{x}^{(k)} - (A^{(k-1)})^{-1} \Delta \mathbf{f}^{(k)}] (\Delta \mathbf{x}^{(k)})^T (A^{(k-1)})^{-1}}{(\Delta \mathbf{x}^{(k)})^T (A^{(k-1)})^{-1} \Delta \mathbf{f}^{(k)}}} \quad (4.27) \end{aligned}$$

Esta fórmula permite calcular la inversa de una matriz con sumas y multiplicaciones de matrices solamente, con lo que se reduce el esfuerzo computacional al orden  $n^2$ .

\* *ibid.*

**Ejemplo 4.6**

Use el método de Broyden para encontrar una solución aproximada del sistema

$$f_1(x, y) = x^2 - 10x + y^2 + 8 = 0$$

$$f_2(x, y) = xy^2 + x - 10y + 8 = 0,$$

tome como vector inicial:  $[x^0, y^0]^T = [0, 0]^T$ . Se recomienda especialmente emplear un pizarrón electrónico para llevar los cálculos, y así poner la atención en el algoritmo y en el análisis de los resultados.

**Solución**

En el ejemplo 4.3 se encontró una solución aproximada de este sistema, empleando el método de Newton-Raphson y el vector cero como vector inicial.

Con los resultados de la primera iteración del ejemplo 4.3

$$J^{(0)} = \begin{bmatrix} -10 & 0 \\ 1 & -10 \end{bmatrix}, \quad (J^{(0)})^{-1} = \begin{bmatrix} -0.1 & 0 \\ -0.01 & -0.1 \end{bmatrix}, \quad \mathbf{x}^{(1)} = \begin{bmatrix} 0.8 \\ 0.88 \end{bmatrix}$$

se calcula  $(A^{(1)})^{-1}$  con la ecuación 4.23

$$\begin{aligned} (A^{(1)})^{-1} &= (J^{(0)})^{-1} + \frac{(\Delta \mathbf{x}^{(1)} - (J^{(0)})^{-1} \Delta \mathbf{f}^{(1)}) (\Delta \mathbf{x}^{(1)})^T (J^{(0)})^{-1}}{(\Delta \mathbf{x}^{(1)})^T (J^{(0)})^{-1} \Delta \mathbf{f}^{(1)}} \\ (A^{(1)})^{-1} &= \begin{bmatrix} -0.1 & 0 \\ -0.01 & -0.1 \end{bmatrix} + \frac{\begin{bmatrix} .8 \\ .88 \end{bmatrix} - \begin{bmatrix} -0.1 & 0 \\ -0.01 & -0.1 \end{bmatrix} \begin{bmatrix} -6.5856 \\ -7.38048 \end{bmatrix} \begin{bmatrix} .8 \\ .88 \end{bmatrix}^T \begin{bmatrix} -0.1 & 0 \\ -0.01 & -0.1 \end{bmatrix}}{\begin{bmatrix} .8 \\ .88 \end{bmatrix}^T \begin{bmatrix} -0.1 & 0 \\ -0.01 & -0.1 \end{bmatrix} \begin{bmatrix} -6.5856 \\ -7.38048 \end{bmatrix}} \\ &= \begin{bmatrix} -0.11015 & -0.010079 \\ -0.01546 & -0.105404 \end{bmatrix} \end{aligned}$$

Se calcula ahora  $\mathbf{x}^{(2)}$  empleando la ecuación:

$$\begin{aligned} \mathbf{x}^{(2)} &= \mathbf{x}^{(1)} - (A^{(1)})^{-1} \mathbf{f}^{(1)} \\ &= \begin{bmatrix} .8 \\ .88 \end{bmatrix} - \begin{bmatrix} -0.11015 & -0.010079 \\ -0.01546 & -0.105404 \end{bmatrix} \begin{bmatrix} 1.4144 \\ 0.61952 \end{bmatrix} \\ &= \begin{bmatrix} 0.96208 \\ 0.96720 \end{bmatrix} \end{aligned}$$

Para la segunda iteración se utilizarán las ecuaciones:

$$(A^{(2)})^{-1} = (A^{(1)})^{-1} + \frac{[\Delta \mathbf{x}^{(2)} - (A^{(1)})^{-1} \Delta \mathbf{f}^{(2)}] (\Delta \mathbf{x}^{(2)})^T (A^{(1)})^{-1}}{(\Delta \mathbf{x}^{(2)})^T (A^{(1)})^{-1} \Delta \mathbf{f}^{(2)}}$$

y

$$\mathbf{x}^{(3)} = \mathbf{x}^{(2)} - (A^{(2)})^{-1} \mathbf{f}^{(2)}$$

Al sustituir valores se obtiene:

$$\mathbf{x}^{(3)} = \begin{bmatrix} 0.997433 \\ 0.996786 \end{bmatrix}$$



La continuación de las iteraciones da

$$\mathbf{x}^{(4)} = \begin{bmatrix} 0.9999037 \\ 0.9998448 \end{bmatrix}, \quad \mathbf{x}^{(5)} = \begin{bmatrix} 0.999998157 \\ 0.999996667 \end{bmatrix}$$

$$\mathbf{x}^{(6)} = \begin{bmatrix} 0.9999999849 \\ 0.9999999722 \end{bmatrix}, \quad \mathbf{x}^{(7)} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

que es la solución del sistema, tal como se obtuvo en los ejemplos 4.2 y 4.3.

Los cálculos pueden realizarse con Matlab o con la TI-92 Plus.



```
x=[0 0]; Eps=1e-8;
fprintf(' k x(k) y(k)\n')
fprintf(' %2d %10.6f %10.6f\n',0,x(1),x(2))
f1=x(1)^2-10*x(1)+x(2)^2+8;
f2=x(1)*x(2)^2+x(1)-10*x(2)+8;
dflx=2*x(1)-10; dfly=2*x(2);
df2x=x(2)^2+1; df2y=2*x(1)*x(2)-10;
J=[dflx dfly; df2x df2y];
F0=[f1; f2]; J1=inv(J); dx=-J1*F0; x1=x+dx';
for k=1:25
    f1=x1(1)^2-10*x1(1)+x1(2)^2+8;
    f2=x1(1)*x1(2)^2+x1(1)-10*x1(2)+8;
    f=[f1; f2]; df=f-F0;
    A1=J1+(dx-J1*df)*dx'*J1/(dx'*J1*df);
    dx=-A1*f; x2=x1+dx'; Dist=norm(x2-x1);
    fprintf(' %2d %10.6f %10.6f %10.5e\n',...
        k,x1(1),x1(2),Dist)
    x1=x2; J1=A1; F0=f;
    if Dist < Eps; break; end
end
```



```
e4_6 ( )
Prgm
Define f1(x)=x[1,1]^2-10*x[1,1]+x[1,2]^2+8
Define f2(x)=x[1,1]*x[1,2]^2+x[1,1]-10*x[1,2]+8
Define dflx(x)=2*x[1,1]-10 : Define dfly(x)=2*x[1,2]
Define df2x(x)=x[1,2]^2+1
Define df2y(x)=2*x[1,1]*x[1,2]-10
[0,0]→x : 1E-5→eps : 0→k : ClrIO
Disp "k      x(k)      y(k)      |x(k+1)-x(k)| "
string(k)&format(x[1,1],"f7")→d
d&" "&format(x[1,2],"f7")→d : Disp d
[dflx(x),dfly(x);df2x(x),df2y(x)]→j
[f1(x);f2(x)]→f0 : j^-1→j1 : -j1*f0→dx : x+dx^T→x1
For k,1,25
    [f1(x1);f2(x1)]→f : f-f0→dff
    j1+(dx-j1*dff)*dx^T*j1/norm(dx^T*j1*dff)→a1
    -a1*f→dx : x1+dx^T→x2 : norm(x2-x1)→dist
    string(k)&format(x2[1,1],"f7")→d
    d&" "&format(x2[1,2],"f7")→d
    d&" "&format(dist,"f5")→d : Disp d
    x2→x1 : a1→j1 : f→f0
    if dist < Eps
        exit
    EndFor
EndPrgm
```

A continuación se presenta el algoritmo para este método.

#### ALGORITMO 4.4 Método de Broyden

Para encontrar una solución aproximada de un sistema de ecuaciones no lineales  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ , proporcionar la matriz jacobiana ampliada con el vector de funciones (véase Ec. 4.17) y los

DATOS: Número de ecuaciones  $N$ , dos vectores de valores iniciales:  $\mathbf{x}_0$  y  $\mathbf{x}_1$ , el número máximo de iteraciones  $\text{MAXIT}$  y el criterio de convergencia  $\text{EPS}$ .

RESULTADOS: Una aproximación a una solución:  $\mathbf{x}_n$  o el mensaje "NO CONVERGE"

PASO 1. Calcular  $AK$ , la matriz inversa de la matriz jacobiana evaluada en  $\mathbf{x}_0$ .

PASO 2. Hacer  $K = 1$ .

PASO 3. Mientras  $K \leq \text{MAXIT}$ , repetir los pasos 4 a 10.

PASO 4. Calcular  $\mathbf{f}_0$  y  $\mathbf{f}_1$ , el vector de funciones evaluado en  $\mathbf{x}_0$  y  $\mathbf{x}_1$ , respectivamente.

PASO 5. Calcular  $(*)\mathbf{dx} = \mathbf{x}_1 - \mathbf{x}_0$ ;  $\mathbf{df} = \mathbf{f}_1 - \mathbf{f}_0$ .

PASO 6. Calcular  $AKI$ , la matriz que aproxima a la inversa de la matriz jacobiana (4.22), con la ecuación (4.27), usando como  $(A^{(k-1)})^{-1}$  a  $AK$ .

PASO 7. Calcular  $(*)\mathbf{x}_n = \mathbf{x}_1 - AKI * \mathbf{f}_1$ .

PASO 8.  $(*)$  Si  $\|\mathbf{x}_n - \mathbf{x}_1\| \leq \text{EPS}$  ir al paso 11. De otro modo continuar.

PASO 9. Hacer  $(*)\mathbf{x}_0 = \mathbf{x}_1$ ;  $\mathbf{x}_1 = \mathbf{x}_n$ ;  $AK = AKI$  (actualización de  $\mathbf{x}_0$ ,  $\mathbf{x}_1$  y  $AK$ ).

PASO 10. Hacer  $K = K + 1$

PASO 11. Si  $K \leq \text{MAXIT}$ , IMPRIMIR el vector  $\mathbf{x}_n$  y TERMINAR.

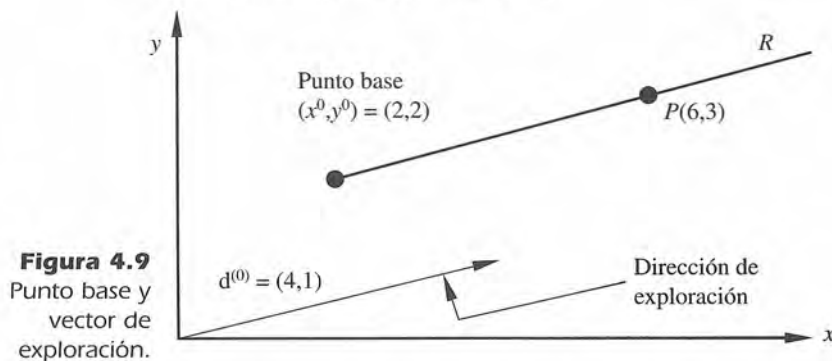
De otro modo IMPRIMIR "NO CONVERGE" y TERMINAR.

\* Operaciones matriciales.

## 4.6 Aceleración de convergencia

Al igual que en los capítulos anteriores, una vez que se tienen métodos de solución funcionales, se mejorarán o crearán nuevos algoritmos usando dicho conocimiento. También, como ya se ha visto, esto se logra con un proceso de generalización y abstracción. Se procederá en esa dirección enseguida.

En cada iteración de los algoritmos vistos se parte de un vector  $\mathbf{x}^{(k)}$ , que ahora se llamará punto base; desde ese punto se camina en una dirección, dada por un vector, que se denominará **dirección de exploración**. Considérese la figura 4.9 y el punto base  $(x^0, y^0)^* = (2, 2)$ . Si desde el punto base se camina en la dirección del vector  $\mathbf{d}^{(0)} = [4, 1]^T$ , se terminará pasando por el punto  $P(6, 3)$ .



**Figura 4.9**  
Punto base y  
vector de  
exploración.

\* De aquí en adelante se usará indistintamente  $n$ -adas ordenadas  $(x_1, x_2, \dots, x_n)$  para representar un vector de  $n$  elementos y un punto en el espacio  $n$ -dimensional.

Al avanzar en cierta dirección de exploración a partir de un punto base, se llega a un nuevo punto que va a ser base para la siguiente iteración, pudiera ser el punto  $P(6, 3)$  o cualquier otro punto de  $R$  que dará la ecuación vectorial.

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + t \mathbf{d}^{(0)}$$

o en forma más general

$$\boxed{\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + t \mathbf{d}^{(k)}} \quad (4.28)$$

donde  $t$  es el factor de tamaño de la etapa y determina la distancia del desplazamiento en la dirección especificada. Esta ecuación se obtiene fácilmente por la suma de vectores en el plano, como se muestra en la figura 4.10.

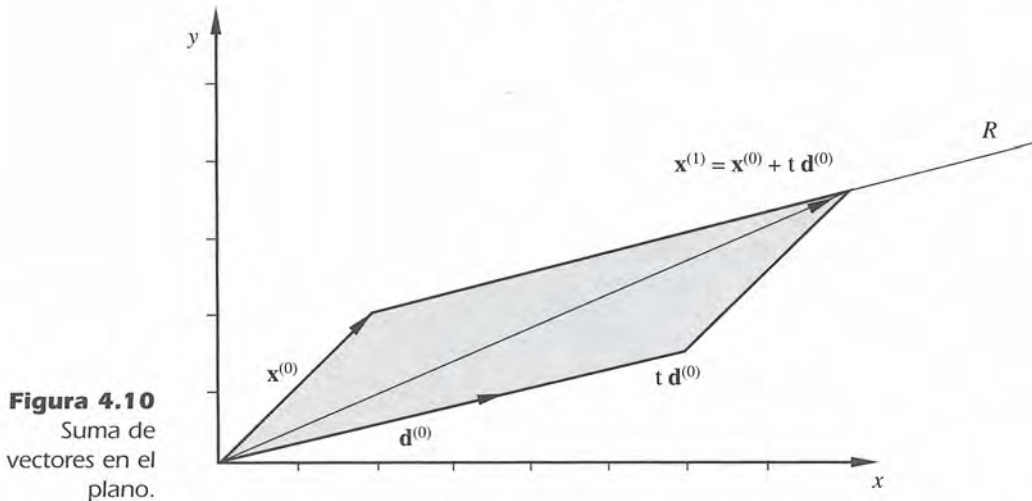
Para aclarar esta generalización, se identifica el algoritmo de Newton-Rapshon para sistemas de dos ecuaciones no lineales con la ecuación 4.28.

Primero se reescribe la ecuación 4.13.

$$\begin{aligned} \frac{\partial f_1}{\partial x} (x^{k+1} - x^k) + \frac{\partial f_1}{\partial y} (y^{k+1} - y^k) &= -f_1(x^k, y^k) \\ \frac{\partial f_2}{\partial x} (x^{k+1} - x^k) + \frac{\partial f_2}{\partial y} (y^{k+1} - y^k) &= -f_2(x^k, y^k) \end{aligned}$$

para pasarla a notación matricial como sigue:

$$\begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{bmatrix} = - \begin{bmatrix} f_1(x^k, y^k) \\ f_2(x^k, y^k) \end{bmatrix}$$



**Figura 4.10**  
Suma de  
vectores en el  
plano.

que ahora, multiplicada por la inversa de la matriz jacobiana, llega a la forma

$$- \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}^{-1} = \begin{bmatrix} f_1(x^k, y^k) \\ f_2(x^k, y^k) \end{bmatrix} \begin{bmatrix} x^{k+1} - x^k \\ y^{k+1} - y^k \end{bmatrix}$$



o también:

$$\begin{bmatrix} x^{k+1} \\ y^{k+1} \end{bmatrix} = \begin{bmatrix} x^k \\ y^k \end{bmatrix} - \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}^{-1} \begin{bmatrix} f_1(x^k, y^k) \\ f_2(x^k, y^k) \end{bmatrix} \quad (4.29)$$

y en esa última forma, ya como ecuación vectorial, se tiene la identificación total con la ecuación 4.28, con:

$$\mathbf{x}^{(k+1)} = \begin{bmatrix} x^{k+1} \\ y^{k+1} \end{bmatrix}; \quad \mathbf{x}^{(k)} = \begin{bmatrix} x^k \\ y^k \end{bmatrix}; \quad t = -1; \quad \mathbf{d}^{(k)} = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix}^{-1} \begin{bmatrix} f_1(x^k, y^k) \\ f_2(x^k, y^k) \end{bmatrix}$$

Hay que observar que en el método de Newton-Raphson, el factor de tamaño de la etapa es constante en todos los pasos iterativos del proceso y que  $\mathbf{d}^{(k)}$  el vector de exploración, es el resultado de multiplicar la inversa de la matriz jacobiana por el vector de funciones.

#### MÉTODO DE NEWTON RAPHSON CON OPTIMIZACIÓN DE $t$

Con la ecuación 4.28 puede estudiarse cómo mejorar los métodos disponibles; por ejemplo, se puede ver que en el algoritmo de Newton-Raphson tomar distintos valores de  $t$  llevaría a distintos vectores  $\mathbf{x}^{(k+1)}$ , alguno más cercano a la raíz  $\mathbf{x}$  que los demás (véase Fig. 4.11). La mejora es optimizar el valor de  $t$  en el método de Newton-Raphson.

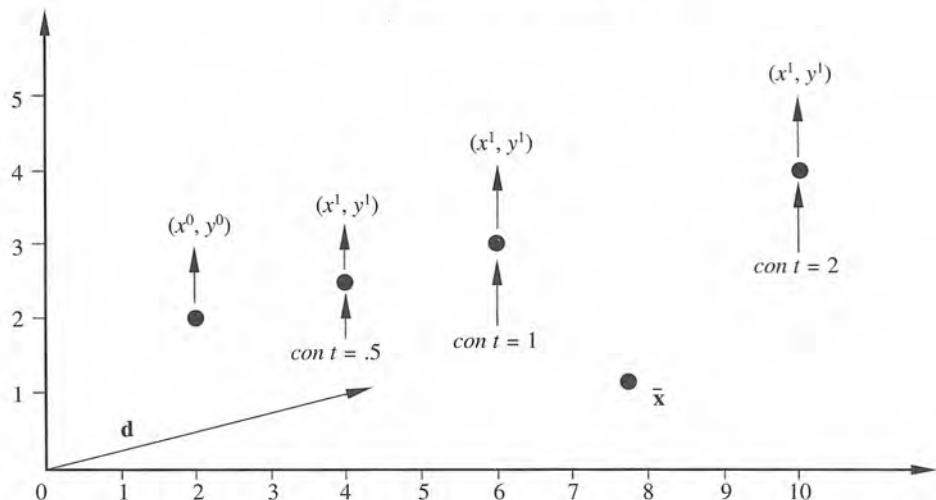
Para ejemplificar, tómense los valores de la primera iteración del ejemplo 4.3:

$$k = 0; \quad x^k = 0; \quad y^k = 0; \quad h = 0.8 \quad j = 0.88$$

de aquí  $\mathbf{d}^{(k)} = [-0.8 \quad 0.88]^T$

y la ecuación 4.29 queda:

$$\begin{aligned} x^{(k+1)} &= x^k + t d_1^k \\ y^{(k+1)} &= y^k + t d_2^k \end{aligned}$$



**Figura 4.11**  
Influencia de  $t$   
en el vector  
 $\mathbf{x}^{(k+1)}$ .

Se lista ahora una serie de valores de  $t$  y los correspondientes valores de  $\mathbf{x}^{(k+1)}$ :

| $t$   | $\mathbf{x}^{(k+1)}$ | $y^{(k+1)}$ |
|-------|----------------------|-------------|
| -0.50 | 0.4                  | 0.44        |
| -0.75 | 0.6                  | 0.66        |
| -1.00 | 0.8                  | 0.88        |
| -1.25 | 1.0                  | 1.10        |
| -1.50 | 1.2                  | 1.32        |

Para determinar cuál de las  $\mathbf{x}^{(k+1)}$  está más cerca de la raíz  $\mathbf{x}$ , se desarrolla un nuevo criterio de convergencia o avance sustentado en la definición de residuo de una función  $f(x, y)$ , dada esta última así:

El residuo de una función  $f(x, y)$  en un punto  $(x^k, y^k)$  es el valor de  $f$  en  $(x^k, y^k)$ .

Así, en el sistema

$$f_1(x, y) = x^2 + y^2 - 4 = 0$$

$$f_2(x, y) = y - x^2 = 0$$

en el punto  $(1, 1)$  los residuos son:

$$f_1(1, 1) = 1^2 + 1^2 - 4 = -2$$

y

$$f_2(1, 1) = 1 - 1^2 = 0$$

En general, el valor de la **función suma de residuos al cuadrado**

$$z_k = f_1^2(x^k, y^k) + f_2^2(x^k, y^k) \quad (4.30)$$

será indicativa de la cercanía de  $\mathbf{x}^{(k)}$  con la raíz  $\mathbf{x}$ .

Con la aplicación de este concepto a los distintos vectores  $\mathbf{x}^{(k+1)}$  obtenidos arriba, se tiene  
Para  $t = -0.5$

$$z_{k+1} = [0.4^2 - 10(0.4) + 0.44^2 + 8]^2 + [0.4(44)^2 + 0.4 - 10(0.44) + 8]^2 = 35.57$$

$$\text{Para } t = -0.75 : \quad z_{k+1} = 12.93$$

$$\text{Para } t = -1.0 : \quad z_{k+1} = 2.38$$

$$\text{Para } t = -1.25 : \quad z_{k+1} = 0.67$$

$$\text{Para } t = -1.5 : \quad z_{k+1} = 4.31$$

De donde  $\mathbf{x}^{(k+1)}$  correspondiente a  $t = -1.25$  resulta ser el más cercano a la raíz  $\mathbf{x} = [1, 1]^T$ .

Los valores propuestos de  $t$  anteriormente, se eligieron de manera arbitraria alrededor de  $-1$  y aunque el valor de  $-1.25$  es el mejor de ellos, no es el óptimo de todos los valores posibles para la primera iteración.

A continuación se da una forma de seleccionar los valores de  $t$ .

Se selecciona un intervalo de búsqueda  $[a, b]$ , dentro de ese intervalo se calculan valores de  $t$  de la siguiente manera:

$$t = a + \frac{b-a}{F} \quad \text{y} \quad t = b - \frac{b-a}{F}$$

donde  $F$  son los términos de la serie de Fibonacci.

$$F = 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, \dots$$

Para cada valor de  $t$  se calcula su correspondiente  $z_{k+1}$ , y el valor mínimo  $z_{k+1}$  proporcionará el valor óptimo de  $t$ .

Así, seleccionando el intervalo  $[-1.2, -1]$ , el valor mínimo de  $z_{k+1}$  ( $= 0.4578$ ) corresponde al valor óptimo de  $t$  ( $= -1.184$ ) en la primera iteración de la solución del ejemplo 4.3. Una vez encontrado el valor óptimo de  $t$  se toma el vector  $\mathbf{x}^{(1)}$  correspondiente y se calcula  $\mathbf{d}^{(1)}$  para proceder a optimizar el valor de  $t$  en la segunda iteración:

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + t \mathbf{d}^{(1)}$$

#### Ejemplo 4.7

Modifique el programa del ejemplo 4.4 para incluir la optimización de  $t$ .

Utilizando el programa resultante, resuelva el sistema del ejemplo 4.4.

#### Solución

Las modificaciones consisten en:



- Elaborar un subprograma para encontrar el valor de  $t$  que minimice la función  $z_k$ , utilizando la búsqueda de Fibonacci.
- Modificar el subprograma NEWTON del ejemplo 4.4 para utilizar ahora como criterio de convergencia o avance la función de  $z_k$  y la llamada al subprograma de búsqueda de Fibonacci.

En el CD (PROGRAMA 4.2) se muestran los subprogramas NEWOPT y BUSCA resultantes. El programa principal y los subprogramas SIMULT y PIVOTEO no sufren cambio alguno.

Con el programa resultante y con los valores iniciales

$$\mathbf{x}^{(0)} = [1 \ 1 \ 1]^T$$

se obtienen los siguientes resultados:

|      |   |            |            |           |
|------|---|------------|------------|-----------|
| VARI | 1 | 1.00000    | 1.00000    | 1.00000   |
| FUNC | 1 | 1.95970    | -624.00000 | 29.83985  |
| SUMA |   | .39027E+06 | TOPT=      | 1.833     |
| VARI | 2 | .83201     | .08453     | -1.75525  |
| FUNC | 2 | 1.00701    | -3.77371   | -24.70092 |
| SUMA |   | .62539E+03 | TOPT=      | .9000     |
| VARI | 3 | .53770     | .04775     | -.64629   |
| FUNC | 3 | .11359     | -1.13613   | -2.47923  |
| SUMA |   | .74503E+01 | TOPT=      | .9000     |
| VARI | 4 | .50380     | .03001     | -.53527   |
| FUNC | 4 | .01153     | -.30917    | -.24846   |
| SUMA |   | .15745E+00 | TOPT=      | 1.167     |
| VARI | 5 | .49935     | .02028     | -.52103   |
| FUNC | 5 | -.00190    | -.00767    | .04138    |
| SUMA |   | .17748E-02 | TOPT=      | .9000     |
| VARI | 6 | .49992     | .02003     | -.52289   |
| FUNC | 6 | -.00019    | -.00081    | .00414    |
| SUMA |   | .17817E-04 | TOPT=      | .9000     |
| VARI | 7 | .49998     | .02000     | -.52308   |
| FUNC | 7 | -.00002    | -.00008    | .00041    |
| SUMA |   | .17825E-06 | TOPT=      | .9000     |



La solución del sistema es:

$$X(1) = .49998116$$

$$X(2) = .19999571E-01$$

$$X(3) = -.52309883$$

Obsérvense los valores de TOPT en las diferentes iteraciones.

Los cálculos pueden realizarse con el siguiente guión de Matlab que usa la función `min4_7` que hace la búsqueda de Fibonacci descrita anteriormente. Las instrucciones que conforman la función deben guardarse en un archivo separado con el nombre `min4_7.m` y posteriormente escribir el guión que la llama, grabarlo y ejecutarlo.



```
function f=min4_7(X,Dx)
A = 0.5; B = 2.5; NP = 0; NU = 1; Menor = 1000000000; Top = 1;
for i = 1:20
    NF = NU + NP; T = A + (B - A)/NF; XX=X+T*Dx';
    Suma=(-3*XX(1)+cos(XX(2)*XX(3))+0.5)^2+...
    (-XX(1)^2+625*XX(2)^2)^2+...
    (-exp(-XX(1)*XX(2))-20*XX(3)-(10*pi-3)/3)^2;
    if Suma < Menor
        Menor = Suma;
        Topt = T;
    end
    T = B - (B - A)/NF; XX=X+T*Dx';
    Suma=(-3*xx(1)+cos(XX(2)*XX(3))+0.5)^2+...
    (-XX(1)^2+625*XX(2)^2)^2+...
    (-exp(-XX(1)*XX(2))-20*XX(3)-(10*pi-3)/3)^2;
    if Suma < Menor
        Menor = Suma;
        Topt = T;
    end
    NP = NU; NU = NF;
end
f=Topt;

n=3; x=[1 1 1]; Maxit=25; Eps=1e-5;
Dist=1;
fprintf(' k x1 x2 x3')
fprintf(' Dist Topt\n')
fprintf(' %2d %10.5f %10.5f %10.5f\n',0,x(1),x(2),x(3))
for k=1:Maxit
    J=[3 x(3)*sin(x(2)*x(3)) x'(2)*sin(x(2)*x(3));...
    2*x(1) -1250*x(2) 0;
    -x(2)*exp(-x(1)*x(2)) -x(1)*exp(-x(1)*x(2)) 20];
    b=[-3*x(1)+cos(x(2)*x(3))+0.5;...
    -x(1)^2+625*x(2)^2;...
    -exp(-x(1)*x(2))-20*x(3)-(10*pi-3)/3];
    dx=inv(J)*b;t=1; t=min4_7(x,dx); x1=x+t*dx';
    Dist=norm(x1-x);
    fprintf(' %2d %10.5f %10.5f %10.5f %10.5e %6.3f\n',...
    k,x1(1),x1(2),x1(3),Dist,t)
    if Dist < Eps
        break
    end
end
x=x1;
end
```



```

e4_7( )
Prgm
@ Inicia el subprograma para búsqueda de Fibonacci
local min4_7
Define min4_7(x,dx)=Prgm
.5→a : 2.5→b : 0→np : 1→nu : 1000→menor : 1→topt
For i,1,20
    nu+np→nf
    For j,1,2
        If j=1 Then
            a+(b-a)/nf→t
        Else
            b-(b-a)/nf→t
        EndIf
        x+t*dx→xx
        f1(xx)^2+f2(xx)^2+f3(xx)^2→suma
        If suma<menor Then
            suma→menor
            t→topt
        EndIf
    EndFor
    nu→np : nf→nu
EndFor
topt→t
EndPrgm
@ Inicia Newton-Raphson con optimización de t
Define f1(x)=3*x[1,1]-cos(x[1,2]*x[1,3])-.5
Define f2(x)=x[1,1]^2-625*x[1,2]^2
Define f3(x)=e^(-x[1,1]*x[1,2])+20*x[1,3]+(10*pi-3)/3
Define df12(x)=x[1,3]*sin(x[1,2]*x[1,3])
Define df13(x)=x[1,2]*sin(x[1,2]*x[1,3])
Define df21(x)=2*x[1,1]
Define df22(x)=-1250*x[1,2]
Define df31(x)=-x[1,2]*e^(-x[1,1]*x[1,2])
Define df32(x)=-x[1,1]*e^(-x[1,1]*x[1,2])
[1,1,1]→x : 3→n : 1E-5→eps : ClrIO
Disp "k  x1  x2  x3  Dist"
"0 "&format(x[1,1],"f4")&" "&format(x[1,2],"f4")→d
d&" "&format(x[1,3],"f4")→d : Disp d
For k,1,25
    [3,df12(x),df13(x);df21(x),df22(x),0;df31(x),df32(x),20]→j
    [-f1(x);-f2(x);-f3(x)]→b
    simult(j,b)→dx : 1→topt : min4_7(x,dx)
    x+topt*dx→x1 : norm(x1-x)→dist
    string(k)&" "&format(x1[1,1],"f5")&" "&format(x1[1,2], "f5")→d
    d&" "&format(x1[1,3],"f5")&" "&format(dist,"f5")→d: Disp d
    x1→x
    if dist<eps
        exit
    EndFor
EndPrgm

```

### MÉTODO DE NEWTON-RAPHSON MODIFICADO CON OPTIMIZACIÓN DE T (MÉTODO SOR)

En el método de Newton-Raphson modificado, la expresión general 4.18 puede identificarse con la ecuación 4.28 directamente con  $t = -1$

$$y \quad d_i^k = \frac{f_i(x_1^{k+1}, x_2^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_n^k)}{\left. \frac{\partial f_i}{\partial x_i} \right|_{(x_1^{k+1}, x_2^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_n^k)}} \quad 1 \leq i \leq n$$

Con la optimización del valor de  $t$  en cada iteración puede acelerarse la convergencia. El método así obtenido

$$x_i^{k+1} = x_i^k - t \frac{f_i(x_1^{k+1}, x_2^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_n^k)}{\left. \frac{\partial f_i}{\partial x_i} \right|_{(x_1^{k+1}, x_2^{k+1}, \dots, x_{i-1}^{k+1}, x_i^k, \dots, x_n^k)}} \quad 1 \leq i \leq n \quad (4.31)$$

se conoce como **método SOR** para sistemas no lineales. A continuación se resuelve un ejemplo con optimización de  $t$ .

#### Ejemplo 4.8

Resuelva el siguiente sistema de ecuaciones empleando el método SOR para sistemas no lineales.

$$f_1(x, y) = x^2 - 10x + y^2 + 8 = 0$$

$$f_2(x, y) = xy^2 + x - 10y + 8 = 0$$

Sean los valores iniciales  $x^0 = 0$  y  $y^0 = 0$

**SUGERENCIA:** Puede seguir los cálculos usando Mathcad o un pizarrón electrónico disponible.

**Solución** Primero se obtiene

$$\frac{\partial f_1}{\partial x} = 2x - 10 \quad y \quad \frac{\partial f_2}{\partial y} = 2xy - 10$$

**Primera iteración**

Se evalúa  $f_1$  y  $\frac{\partial f_1}{\partial x}$  en  $[0, 0]^T$

$$f_1(0, 0) = 8, \quad \left. \frac{\partial f_1}{\partial x} \right|_{x^0, y^0} = -10$$

Se elige el intervalo de búsqueda  $[-1.5, -0.5]$  y  $t = b - (b - a)/F$ , y el primer valor a prueba es:

$$t = -1.5$$

$$x^1 = x^0 + t \frac{f_1(0, 0)}{\left. \frac{\partial f_1}{\partial x} \right|_{(0, 0)}} = 0 - 1.5 \left( \frac{8}{-10} \right) = 1.2$$



$$f_2(1.2, 0) = 9.2 \left. \frac{\partial f_2}{\partial y} \right|_{x^1, y^0} = -10$$

$$y^1 = y^0 + t \frac{f_2(1.2, 0)}{\left. \frac{\partial f_2}{\partial y} \right|_{(1.2, 0)}} = 0 - 1.5 \left( \frac{9.2}{-10} \right) = 1.38$$

A partir del criterio de la suma de los residuos elevados al cuadrado, se tiene:

$$z_1 = f_1^2(1.2, 1.38) + f_2^2(1.2, 1.38)$$

$$= [1.2^2 - 10(1.2) + 1.38^2 + 8]^2 + [1.2(1.38)^2 - 1.2 - 10(1.38) + 8]^2 = 5.7877$$

El segundo valor a prueba es  $t = -1.0$ , con lo que se obtiene:

$$x^1 = 0.8 \quad y^1 = 0.88 \quad z_1 = 2.3843$$

Al continuar el proceso de búsqueda se tiene:

| $t$     | $x^1$  | $y^1$  | $z_1$   |
|---------|--------|--------|---------|
| -1.5000 | 1.2000 | 1.3800 | 5.7877  |
| -1.0000 | 0.8000 | 0.8800 | 2.3843  |
| -0.8333 | 0.6667 | 0.7222 | 8.4991  |
| -0.7000 | 0.5600 | 0.5992 | 17.1088 |

Ahora se usa  $t = a + (b - a)/F$ , y se obtiene:

| $t$     | $x^1$  | $y^1$  | $z_1$   |
|---------|--------|--------|---------|
| -0.5000 | 0.4000 | 0.4200 | 37.0420 |
| -1.0000 | 0.8000 | 0.8800 | 2.3843  |
| -1.1667 | 0.9333 | 1.0422 | 0.6151  |
| -1.3000 | 1.0400 | 1.1752 | 1.6312  |

Por tanto, el valor óptimo de  $t$  es  $-1.1666$  y los valores correspondientes de  $[x^1, y^1] = [0.9333, 1.0422]$  se toman como resultados finales de la primera iteración.

### Segunda iteración

Con el mismo intervalo de búsqueda  $[-1.5, -0.5]$  se tiene, con  $t = b - (b - a)/F$

| $t$     | $x^1$    | $y^1$    | $z_1$    |
|---------|----------|----------|----------|
| -1.5000 | 1.048416 | 0.997117 | 0.166991 |
| -1.0000 | 1.010055 | 1.002319 | 0.005733 |
| -0.8333 | 0.997268 | 1.006275 | 0.004285 |
| -0.7000 | 0.987039 | 1.010227 | 0.027122 |

y con  $t = a + (b - a)/F$

| $t$     | $x^1$    | $y^1$    | $z_1$    |
|---------|----------|----------|----------|
| -0.5000 | 0.971694 | 1.017453 | 0.107664 |
| -1.0000 | 1.010055 | 1.002319 | 0.005733 |
| -1.1667 | 1.022842 | 0.999467 | 0.036085 |
| -1.3000 | 1.033072 | 0.997987 | 0.078302 |

El valor óptimo de  $t$  es  $-0.8333$  y los valores correspondientes de  $[x^2, y^2] = [0.997268, 1.006275]$  se toman como resultados finales de la segunda iteración.

Al continuar el proceso iterativo se obtienen los siguientes valores

| $k$ | $x^k$    | $y^k$    | $z_k$    | $t_{opt}$ |
|-----|----------|----------|----------|-----------|
| 0   | 0.000000 | 0.000000 |          |           |
| 1   | 0.933333 | 1.042222 | 0.615080 | -1.1667   |
| 2   | 0.997268 | 1.006275 | 0.004285 | -0.8333   |
| 3   | 1.000854 | 1.001220 | 0.000084 | -0.8333   |
| 4   | 1.000305 | 1.000076 | 0.000005 | -1.0000   |
| 5   | 1.000019 | 1.000005 | 0.000000 | -1.0000   |

Los cálculos pueden realizarse con el siguiente guión de Matlab que usa la función `min4_8`, que hace la búsqueda de Fibonacci descrita anteriormente o con la TI92 Plus.



```
x=[0 0];      Eps=1e-6;
fprintf('  k  x(k)  y(k)')
fprintf('  z(k)  Topt\n')
fprintf('  %2d  %13.10f  %13.10f\n',0,x(1),x(2))
for k=1:10
    t=min4_8(x);    f1=x(1)^2-10*x(1)+x(2)^2+8;
    dflx=2*x(1)-10; x1(1)=x(1)+t*f1/dflx;
    f2=x1(1)*x(2)^2+x1(1)-10*x(2)+8;
    df2y=2*x1(1)*x(2)-10; x1(2)=x(2)+t*f2/df2y;
    f1=x1(1)^2-10*x1(1)+x1(2)^2+8;
    f2=x1(1)*x1(2)^2+x1(1)-10*x1(2)+8;
    Z=f1^2+f2^2;
    fprintf('          %2d  %13.10f  %13.10f',k,x1(1),x1(2))
    fprintf('          %13.10f  %8.4f\n',Z,t)
    x=x1;
    if Z < Eps; break; end
end
```

```

function f=min4_8(X)
    A = -1.5; B = -0.5; NP = 0; NU = 1;
    Menor = 1000000000; Topt = 1;
    for i = 1:4
        for j=1:2
            NF = NU + NP;
            if j == 1
                T = B-(B - A)/NF;
            else
                T=A+(B-A)/NF;
            end
            f1=X(1)^2-10*X(1)+X(2)^2+8;    dflx=2*X(1)-10;
            XX(1)=X(1)+T*f1/dflx;
            f2=XX(1)*X(2)^2+XX(1)-10*X(2)+8;
            df2y=2*XX(1)*X(2)-10; XX(2)=X(2)+T*f2/df2y;
            f1=XX(1)^2-10*XX(1)+XX(2)^2+8;
            f2=XX(1)*XX(2)^2+XX(1)-10*XX(2)+8;
            Suma=f1^2+f2^2
            if Suma < Menor
                Menor = Suma;    Topt = T;
            end
        end
        NP = NU; NU = NF;
    end
    f=Topt;

```



```

e4_8( )
Prgm
@ Inicia subprograma búsqueda de Fibonacci
Local min4_8
Define min4_8(x)=Prgm
-1.5→a : -0.5→b : 1→topt : 0→np : 1→nu:10000→menor
For i,1,20
nu+np→nf
For j,1,2
If j=1 Then
a+(b-a)/nf→t
Else
b-(b-a)/nf→t
EndIf
x→xx : x[1,1]+t*f1(x)/dflx(x) →xx[1,1]
x[1,2]+t*f2(xx)/df2y(xx) →xx[1,2]
f1(xx)^2+f2(xx)^2→suma
If suma<menor Then
suma→menor : t→topt
EndIf
EndFor
nu→np : nf→nu
EndFor
topt→t
EndPrgm
@ Inicia Newton Raphson modificado
Define f1(x)=x[1,1]^2-10*x[1,1]+x[1,2]^2+8
Define f2(x)=x[1,1]*x[1,2]^2+x[1,1]-10*x[1,2]+8
Define dflx(x)=2*x[1,1]-10
Define df2y(x)=2*x[1,1]*x[1,2]-10

```



```

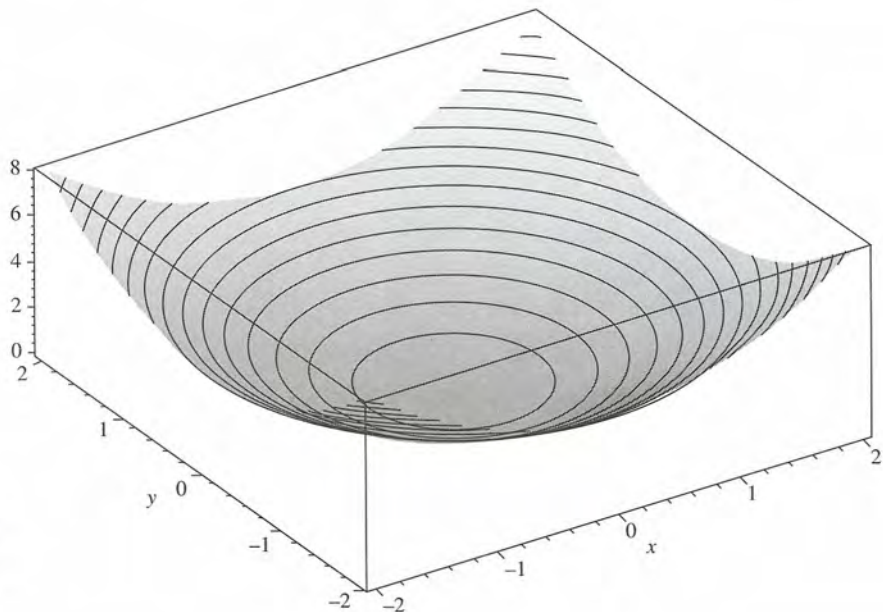
[0,0] →X : 1.E-6→eps : ClrIO
Disp "k    x1    x2    z topt"
"0" "&format(x[1,1],"f4")→d
d&" "&format(x[1,2],"f4")→d : Disp d
For k,1,10
  min4 8(x)
  x→x1 : x[1,1]+topt*f1(x)/df1x(x) →x1[1,1]
  x[1,2]+topt*f2(x1)/df2y(x1) →x1[1,2]
  f1(x1)^2+f2(x1)^2→zeta
  string(k)&" "&format(x1[1,1],"f5")→d
  d&" "&format(x1[1,2],"f5")&" "&format(zeta,"f5")→d
  d&" "&format(topt,"f5")→d: Disp d
  x1→x
  If zeta <eps
    Exit
EndFor
EndPrgm

```

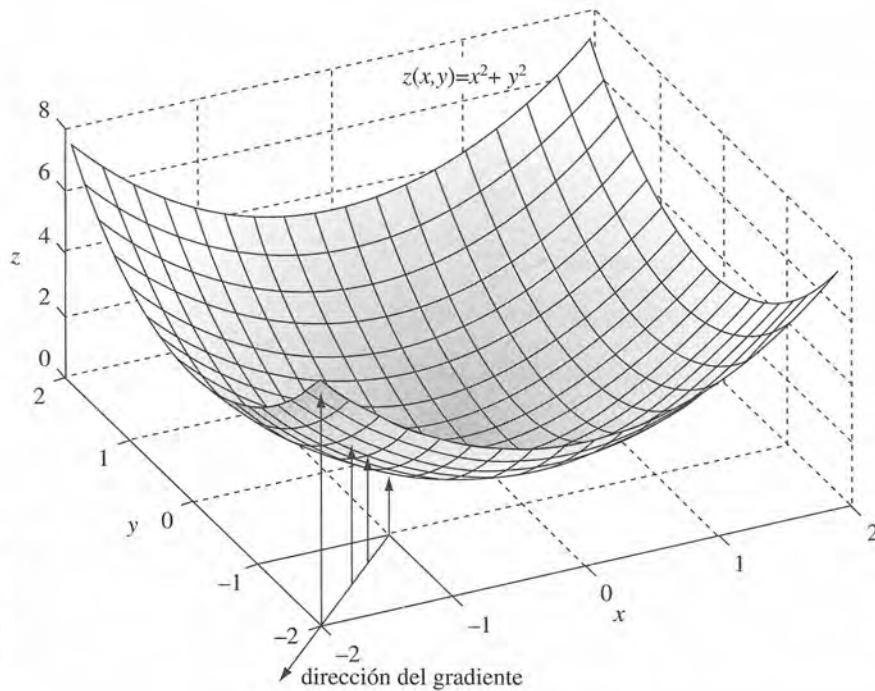
### MÉTODO DEL DESCENSO DE MÁXIMA PENDIENTE

Se ha visto cómo seguir un camino que permita ir disminuyendo  $z$  al optimizar el tamaño del paso  $t$  de un método conocido. Sin embargo, puede elaborarse un método de solución de (4.1), construyendo primero una dirección de exploración  $\mathbf{d}$  que permita disminuir el valor de  $z$  en una cantidad localmente máxima y, una vez encontrada, buscar la  $t$  óptima en esa dirección. Para el desarrollo de este algoritmo son necesarias las siguientes consideraciones.

La figura 4.12 representa la gráfica de la función  $z(x, y) = x^2 + y^2$ , y sus curvas de nivel. Si, por ejemplo, se "está" en el punto  $(x, y, z) = (-1, -1, 2)$  de la superficie (ver Fig. 4.13), el gradiente de la función  $z(x, y)$ ,



**Figura 4.12**  
Gráfica de la  
función  $z(x, y) =$   
 $x^2 + y^2$ .

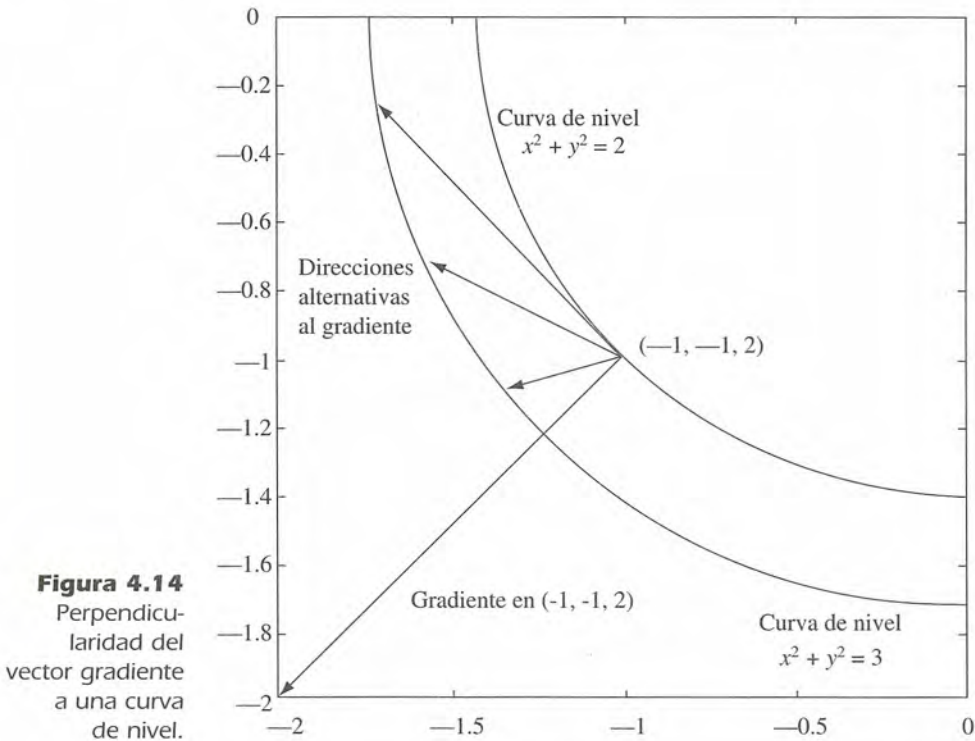


**Figura 4.13**  
Gráfica de la  
superficie  $z(x, y)$   
y dirección del  
vector  
gradiente.

$\nabla z(x, y) = \begin{bmatrix} \partial z / \partial x \\ \partial z / \partial y \end{bmatrix} = \begin{bmatrix} 2x \\ 2y \end{bmatrix}$  evaluado en  $(x, y) = (-1, -1)$  es el vector  $\begin{bmatrix} -2 \\ -2 \end{bmatrix}$  en el plano  $x - y$  cuya dirección nos indica hacia dónde avanzar en el mismo plano  $x - y$ , a fin de “ascender” **en la superficie** (a partir de  $(-1, -1, 2)$ ) lo más rápidamente posible\*. Como nuestro interés es descender lo más bruscamente posible, se toma la dirección contraria del gradiente o, matemáticamente, se va en la dirección de  $-\nabla z(x, y) = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$ . Nótese que, siguiendo la dirección opuesta del vector gradiente en la figura 4.13, se avanza hacia el punto  $(0, 0)$  del plano  $x - y$ , que es donde la función  $z(x, y)$  tiene su mínimo.

Otra propiedad del gradiente que puede ser útil para visualizar cómo encontrar un máximo o un mínimo es que es perpendicular a las curvas de nivel de la superficie. Las curvas de nivel de una superficie  $z(x, y)$  son el conjunto de puntos que satisfacen la ecuación  $z(x, y) = c$ , donde  $c$  es una constante. Así, para la superficie  $z(x, y) = x^2 + y^2$ , las curvas de nivel son la familia  $x^2 + y^2 = c$ , es decir, las circunferencias con centro en el eje  $z$  paralelas al plano  $x - y$  y a una altura  $c$  de éste, y de radio  $\sqrt{c}$  (ver Fig. 4.12). Si tomamos nuevamente el punto  $(-1, -1, 2)$ , la circunferencia  $x^2 + y^2 = 2$  con centro en  $(0, 0, 2)$  es la curva de nivel que lo contiene. Si tomamos el vector gradiente  $\begin{bmatrix} -2 \\ -2 \end{bmatrix}$  y lo llevamos paralelo al plano  $x - y$  al punto  $(-1, -1, 2)$  encontramos que es perpendicular a la curva de nivel en ese punto (ver Fig. 4.14). Aun más, al avanzar desde ese punto en la dirección que señala el gradiente, se avanza sobre la superficie hacia curvas de nivel de mayor radio por el camino más corto posible, ya que cualquier otra dirección que se tomara a partir de  $(-1, -1, 2)$  nos llevaría a otra curva de nivel; por ejemplo,  $x^2 + y^2 = 3$  por un camino más largo (ver Fig. 4.14).

\* Cabe señalar que esto sólo es cierto para un entorno del punto  $(-1, -1)$  del plano  $x - y$ .



**Figura 4.14**  
Perpendicularidad del vector gradiente a una curva de nivel.

Con esta definición de gradiente y sus propiedades, se retoma el asunto del cálculo de la dirección que asegura la disminución de  $z = f_1^2(\mathbf{x}) + f_2^2(\mathbf{x})$  (función escalar de  $x$  y  $y$ ), en una cantidad localmente máxima en un punto. Se determina el vector gradiente de  $z$  con signo negativo en dicho punto (el signo negativo se debe a que se quiere que  $z$  disminuya.). El vector gradiente de  $z$  se representa por  $\nabla z$ . Por tanto, la dirección de descenso más brusco es:

$$\mathbf{d} = -(\nabla z)$$

con cada uno de los componentes de  $\mathbf{d}$  calculados como:

$$d_1 = \frac{\partial z}{\partial x}, \quad d_2 = \frac{\partial z}{\partial y}$$

#### Ejemplo 4.9

Obtenga la dirección del descenso de máxima pendiente del sistema

$$f_1(x_1, x_2, x_3) = 3x_1 - \cos(x_2 x_3) - 0.5 = 0$$

$$f_2(x_1, x_2, x_3) = x_1^2 - 625x_2^2 = 0$$

$$f_3(x_1, x_2, x_3) = e^{-x_1 x_2} + 20x_3 + (10\pi - 3)/3 = 0$$

use como vector inicial a

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$



**Solución**

$$z = [3x_1 - \cos(x_2x_3) - 0.5]^2 + [x_1^2 - 625x_2^2]^2 + [e^{-x_1x_2} + 20x_3 + (10\pi - 3)/3]^2$$

$$d_1 = \frac{\partial z}{\partial x_1} = 6(3x_1 - \cos(x_2x_3) - 0.5) + 4x_1(x_1^2 - 625x_2^2)$$

$$-2x_2 e^{-x_1x_2} [e^{-x_1x_2} + 20x_3 + (10\pi - 3)/3]$$

$$d_2 = \frac{\partial z}{\partial x_2} = 2x_3 \sin(x_2x_3) [3x_1 - \cos(x_2x_3) - 0.5]$$

$$-2500x_2(x_1^2 - 625x_2^2) - 2x_1 e^{-x_1x_2} [e^{-x_1x_2} + 20x_3 + (10\pi - 3)/3]$$

$$d_3 = \frac{\partial z}{\partial x_3} = 2x_2 \sin(x_2x_3) [3x_1 - \cos(x_2x_3) - 0.5] +$$

$$40 [e^{-x_1x_2} + 20x_3 + (10\pi - 3)/3]$$

Al evaluar  $d_1$ ,  $d_2$ , y  $d_3$  en  $\mathbf{x}^{(0)}$  se obtiene:

$$d_1 = -9.0$$

$$d_2 = 0.0$$

$$d_3 = 418.87872$$

y entonces el vector dirección es:

$$\mathbf{d} = \begin{bmatrix} -9 \\ 0.00 \\ 418.87872 \end{bmatrix}$$

Una vez calculada la dirección, se utiliza una exploración unidimensional para localizar el mínimo en esta dirección (por ejemplo una búsqueda de Fibonacci). Ya localizado el mínimo, se calcula una nueva dirección de descenso de máxima pendiente y se repite el procedimiento. Generalmente, el método se caracteriza por cortos movimientos en zig-zag que convergen muy lentamente a la solución; sin embargo, se utiliza para acercarse a la solución y después aplicar un método de alto orden de convergencia como el de Newton-Raphson; es decir, se emplea como un método para conseguir “buenos” valores iniciales. Este método puede ejemplificarse paso a paso con el Mathcad o un software equivalente y explorar con varios valores de  $t$  para encontrar el óptimo; cabe ensayarlo con diferentes sistemas e incluso proponer vectores de exploración, en fin, llevar la matemática a nivel experimental.

#### ALGORITMO 4.5 Método del descenso de máxima pendiente

Para encontrar una solución aproximada de un sistema de ecuaciones no lineales  $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ , proporcionar las funciones  $F(\mathbf{I}, \mathbf{x})$  y las derivadas parciales de la función (ecuación (4.28))  $D(\mathbf{I}, \mathbf{x})$  y los

**DATOS:** Número de ecuaciones  $N$ , vector de valores iniciales  $\mathbf{x}$ , número máximo de iteraciones  $MAXIT$ , criterio de convergencia  $EPS$ , intervalo de búsqueda  $[A, B]$  y el número de puntos de  $[A, B]$  por ensayar  $M$ .

**RESULTADOS:** El vector solución  $\mathbf{x}$  o mensaje “NO CONVERGE”.

- PASO 1. Hacer  $K = 1$ .
- PASO 2. Mientras  $K \leq \text{MXIT}$ , repetir los pasos 3 a 27.
- PASO 3. Hacer  $Z = 0$ .
- PASO 4. Hacer  $I = 1$ .
- PASO 5. Mientras  $I \leq N$ , repetir los pasos 6 y 7.
- PASO 6. Hacer  $Z = Z + F(I, x)^2$ .
- PASO 7. Hacer  $I = I + 1$ .
- PASO 8. Si  $Z \leq \text{EPS}$  ir al paso 29. De otro modo continuar.
- PASO 9. Hacer  $\text{NP} = 0$ ,  $\text{NU} = 1$ ,  $\text{MENOR} = 1\text{E}20$ .
- PASO 10. Hacer  $J = 1$ .
- PASO 11. Mientras  $J \leq M$ , repetir los pasos 12 a 25.
- PASO 12. Hacer  $S = \text{NU} + \text{NP}$ ,  
 $T = A + (B-A)/S$ ,  $L = 1$ .
- PASO 13. Hacer  $\mathbf{x}\mathbf{a} = \mathbf{x} - T * \mathbf{dz}$ .
- PASO 14. Hacer  $Z = 0$ .
- PASO 15. Hacer  $I = 1$ .
- PASO 16. Mientras  $I \leq M$ , repetir los pasos 17 y 18.
- PASO 17. Hacer  $Z = Z + F(I, \mathbf{x}\mathbf{a})^2$ .
- PASO 18. Hacer  $I = I + 1$ .
- PASO 19. Si  $\text{MENOR} < Z$ , ir al paso 21.  
 De otro modo continuar.
- PASO 20. Hacer  $\text{MENOR} = Z$ ,  $\text{TOPT} = T$ .
- PASO 21. Si  $L = 0$  ir al paso 24. De otro modo continuar.
- PASO 22. Hacer  $T = B - (B-A)/S$ ,  $L = 0$ .
- PASO 23. Ir al paso 13.
- PASO 24. Hacer  $\text{NP} = \text{NU}$ ,  $\text{NU} = S$ .
- PASO 25. Hacer  $J = J + 1$ .
- PASO 26. Hacer  $\mathbf{x} = \mathbf{x} - \text{TOPT} * \mathbf{dz}$ .
- PASO 27. Hacer  $K = K + 1$ .
- PASO 28. IMPRIMIR "NO CONVERGE" y TERMINAR.
- PASO 29. IMPRIMIR  $\mathbf{x}$  y TERMINAR.

### Ejemplo 4.10

Con el algoritmo 4.5, elabore un programa para resolver el sistema

$$\begin{aligned} f_1(x_1, x_2, x_3) &= 3x_1 - \cos(x_2 x_3) - 0.5 = 0 \\ f_2(x_1, x_2, x_3) &= x_1^2 - 625x_2^2 = 0 \\ f_3(x_1, x_2, x_3) &= e^{-x_1 x_2} + 20x_3 + (10\pi - 3)/3 = 0 \end{aligned}$$

use como vector inicial  $\mathbf{a}$

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

### Solución



En el CD se presenta el **PROGRAMA 4.3**, basado en el método del descenso de máxima pendiente y con búsqueda de Fibonacci.

Para su empleo, el usuario proporcionará el procedimiento GRADTE, donde se forma la función  $z$  por minimizar y el gradiente de esta función  $\nabla z$ . Enseguida se anotan los resultados que se obtienen.

| $k$ | $x_1$   | $x_2$    | $x_3$    | $z$          | $t$     |
|-----|---------|----------|----------|--------------|---------|
| 0   | 0.00000 | 0.00000  | 0.00000  |              |         |
| 1   | 0.01127 | 0.00000  | -0.52458 | 1.11912e+002 | 0.00125 |
| 2   | 0.33117 | -0.00002 | -0.49597 | 2.15009e+000 | 0.03636 |
| 3   | 0.33479 | 0.00044  | -0.52365 | 5.73944e-001 | 0.00125 |
| 4   | 0.50090 | 0.00759  | -0.52085 | 2.58186e-001 | 0.05882 |
| .   | .       | .        | .        | .            | .       |
| .   | .       | .        | .        | .            | .       |
| .   | .       | .        | .        | .            | .       |
| 37  | 0.49998 | 0.02000  | -0.52310 | 1.73843e-009 | 0.03636 |
| 38  | 0.49998 | 0.02000  | -0.52310 | 4.03291e-010 | 0.00077 |

Para finalizar este capítulo estudiaremos una aplicación del método de Newton-Raphson para encontrar factores cuadráticos de una ecuación polinomial con el método de Bairstow.

## 4.7 Método de Bairstow

Este método, al igual que el método de Lin visto en el capítulo 2, permite obtener factores cuadráticos del polinomio\*

$$p(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_n, \quad (4.32)$$

aplicando el método de Newton-Raphson a un sistema relacionado con dicho polinomio. Más específicamente, la división de  $p(x)$  por el polinomio cuadrático  $x^2 - ux - v$  (factor buscado) puede expresarse como:

$$p(x) = (x^2 - ux - v)q(x) + r(x) \quad (4.33)$$

donde  $q(x)$  es un polinomio de grado  $n-2$  y  $r(x)$  el residuo lineal, dados respectivamente por

$$q(x) = b_0x^{n-2} + b_1x^{n-3} + \dots + b_{n-2} \quad (4.34)$$

$$r(x) = b_{n-1}(u, v)(x - u) + b_n(u, v) \quad (4.35)$$

donde la notación  $b_{n-1}(u, v)$  y  $b_n(u, v)$  se usa para enfatizar que  $b_{n-1}$  y  $b_n$  dependen de las  $u$  y  $v$  seleccionadas para formar el factor cuadrático.

El factor cuadrático será un factor de  $p(x)$  si podemos escoger  $u$  y  $v$  de modo que

$$b_{n-1}(u, v) = 0, \quad b_n(u, v) = 0 \quad (4.36)$$

Al desarrollar las operaciones indicadas en el lado derecho de la ecuación (4.33), se obtiene un polinomio cuyos coeficientes quedan expresados en términos de las  $b$ 's,  $u$  y  $v$ . Al igualar éstos con los coeficientes correspondientes de (4.32), se tiene a las  $b$ 's en la siguiente forma:

\* Nótese que la forma en que se manejan los subíndices de los coeficientes difieren de la forma en que se usaron anteriormente y será exclusiva para esta sección.



---


$$\begin{array}{ll}
 b_0 = a_0 & b_0 = a_0 \\
 b_1 - ub_0 = a_1 & b_1 = a_1 + ub_0 \\
 b_2 - ub_1 - vb_0 = a_2 & b_2 = a_2 + ub_1 + vb_0 \\
 \vdots & \vdots \\
 b_k - ub_{k-1} - vb_{k-2} = a_k & b_k = a_k + ub_{k-1} + vb_{k-2} \\
 \vdots & \vdots \\
 b_{n-1} - ub_{n-2} - vb_{n-3} = a_{n-1} & b_{n-1} = a_{n-1} + ub_{n-2} + vb_{n-3} \\
 b_n - ub_{n-1} - vb_{n-2} = a_n & b_n = a_n + ub_{n-1} + vb_{n-2}
 \end{array}$$


---

Si se hace artificialmente  $b_{-1} = 0$  y  $b_{-2} = 0$ , la expresión para  $b_k$  vale para  $0 \leq k \leq n$ , de modo que nuestra expresión general quedaría:

$$b_k = a_k + ub_{k-1} + vb_{k-2}; \quad 0 \leq k \leq n$$

y en particular:

$$\begin{aligned}
 b_0 &= a_0 + ub_{-1} + vb_{-2} \\
 b_1 &= a_1 + ub_0 + vb_{-1}
 \end{aligned}$$

Es interesante observar el carácter recursivo de las  $b$ 's ya que, por ejemplo,  $b_k$  está expresada en términos de  $b_{k-1}$  y  $b_{k-2}$ , y ambas a su vez se pueden expresar en  $b$ 's cuyos subíndices son  $k-2$ ,  $k-3$  y  $k-3$ ,  $k-4$ , respectivamente. Continuando de esta manera,  $b_k$  queda finalmente expresada en términos de los coeficientes de (4.32), que son conocidos, y obviamente de  $u$  y  $v$  que son propuestos. En adelante, todo se hará en forma recursiva, de modo que cualquier cálculo relacionado con el sistema (4.36) quedará sujeto a un proceso de este tipo.

La forma del factor cuadrático  $x^2 - ux - v$ , tan artificial a primera vista, tiene su razón en la facilitación del cálculo de  $p(x)$  para un argumento complejo  $x = a + bi$ . Sean las  $a_k$  reales. Haciendo  $u = 2a$  y  $v = -a^2 - b^2$ , tenemos:

$$\begin{aligned}
 x^2 - ux - v &= (a + bi)^2 - 2a(a + bi) - (-a^2 - b^2) \\
 &= a^2 + 2abi - b^2 - 2a^2 - 2abi + a^2 + b^2 = 0
 \end{aligned}$$

De esto, por la ecuación (4.32):

$$\begin{aligned}
 p(x) &= (x^2 - ux - v)q(x) + r(x) \\
 &= 0 + r(x) = b_{n-1}(a + bi - 2a) + b_n \\
 &= b_{n-1}(-a + bi) + b_n
 \end{aligned} \tag{4.37}$$

Para obtener  $b_{n-1}$  y  $b_n$  deberán evaluarse primero  $b_0, b_1, \dots, b_{n-2}$  y esto puede hacerse por aritmética real, ya que como vimos antes, se calculan en términos de los coeficientes del polinomio 4.32, que son reales y de  $u$  y  $v$  que también son reales, y sólo hasta el cálculo final se empleará aritmética compleja en la multiplicación de  $b_{n-1}$  por  $(-a + bi)$ . Si se diera el caso de que  $b_{n-1}$  y  $b_n$  fueran ceros, entonces  $p(x) = 0$ , y los complejos conjugados  $a \pm bi$  serían entonces ceros de  $p(x)$ .

El método de Bairstow consiste en usar el método de Newton para resolver el sistema (4.36).

Las derivadas parciales de  $b_{n-1}$  y  $b_n$  con respecto a  $u$  y  $v$  implican obtener primero las derivadas parciales de  $b_{n-2}, b_{n-3}, \dots, b_1$  y  $b_0$ , dada la recursividad de  $b_n$  y  $b_{n-1}$ . Por esto, sea

$$c_{-2} = \frac{\partial b_{-1}}{\partial u} = 0$$

$$\begin{aligned}
c_{-1} &= \frac{\partial b_0}{\partial u} = 0 \\
c_0 &= \frac{\partial b_1}{\partial u} = b_0 \\
c_1 &= \frac{\partial b_2}{\partial u} = \frac{\partial(a_2 + u(a_1 + ub_0) + vb_0)}{\partial u} = a_1 + 2ub_0 = b_1 + uc_0 \\
&\vdots \\
c_k &= \frac{\partial b_{k+1}}{\partial u} = b_k + uc_{k-1} + vc_{k-2} \\
&\vdots \\
c_{n-1} &= \frac{\partial b_n}{\partial u} = b_{n-1} + uc_{n-2} + vc_{n-3}
\end{aligned}$$

De este modo, las  $c_k$  se calculan a partir de las  $b_k$ , del mismo modo que las  $b_k$  se obtuvieron a partir de las  $a_k$ . Los dos resultados que necesitamos son:

$$c_{n-2} = \frac{\partial b_{n-1}}{\partial u}, \quad c_{n-1} = \frac{\partial b_n}{\partial u}$$

De igual modo, tomando derivadas respecto a  $v$  y haciendo  $d_k = \frac{\partial b_{k+2}}{\partial v}$ , encontramos:

$$\begin{aligned}
d_{-2} &= \frac{\partial b_0}{\partial v} = 0 \\
d_{-1} &= \frac{\partial b_1}{\partial v} = 0 \\
d_0 &= \frac{\partial b_2}{\partial v} = b_0 \\
d_1 &= \frac{\partial b_3}{\partial v} = \frac{\partial(a_3 + ub_2 + vb_1)}{\partial v} = \frac{\partial(a_3 + u(a_2 + ub_1 + vb_0) + vb_1)}{\partial v} \\
&= b_1 + ub_0 = b_1 + ud_0 \\
&\vdots \\
d_k &= \frac{\partial b_{k+2}}{\partial v} = b_k + ud_{k-1} + vd_{k-2} \\
&\vdots \\
d_{n-2} &= \frac{\partial b_n}{\partial v} = b_{n-2} + ud_{n-3} + vd_{n-4}
\end{aligned}$$

Como las  $c_k$  y las  $d_k$  satisfacen la misma recurrencia:

$$\begin{aligned}
c_{-2} &= d_{-2} = 0 \\
c_{-1} &= d_{-1} = 0 \\
c_0 &= d_0 = b_0
\end{aligned}$$

$$\begin{aligned}
 c_1 &= b_1 + uc_0 = b_1 + ud_0 = d_1 \\
 &\vdots \\
 c_k &= b_k + uc_{k-1} + vc_{k-2} = d_k \\
 &\vdots \\
 c_{n-2} &= b_{n-2} + uc_{n-3} + vc_{n-4} = d_{n-2}
 \end{aligned}$$

En particular

$$\frac{\partial b_{n-1}}{\partial v} = d_{n-3} = c_{n-3}, \quad \frac{\partial b_n}{\partial v} = d_{n-2} = c_{n-2}$$

y ahora se tiene todo para aplicar el método de Newton-Raphson. Supóngase que se tienen raíces aproximadas  $a \pm bi$  de  $p(x) = 0$  y con esto el factor cuadrático asociado  $x^2 - ux - v$ , de  $p(x)$ . Esto significa que tenemos raíces aproximadas de la ecuación 4.36. Aplicando el método de Newton-Raphson a (4.36) queda:

$$\begin{aligned}
 c_{n-2}h + c_{n-3}k &= -b_{n-1} \\
 c_{n-1}h + c_{n-2}k &= -b_n
 \end{aligned}$$

Dado que se trata de un sistema de dos ecuaciones lineales, se puede programar fácilmente la solución recurriendo a la regla de Cramer.

$$h = \frac{b_n c_{n-3} - b_{n-1} c_{n-2}}{c_{n-2}^2 - c_{n-1} c_{n-3}}, \quad k = \frac{b_{n-1} c_{n-1} - b_n c_{n-2}}{c_{n-2}^2 - c_{n-1} c_{n-3}}$$

### Ejemplo 4.11

Encuentre los factores cuadráticos de la ecuación polinomial de cuarto grado

$$p(x) = x^4 - 8x^3 + 39x^2 - 62x + 50 = 0$$

Utilice como valor inicial  $x = 0 + 0i$ ; esto es,  $a = 0$  y  $b = 0$

### Solución

Dado lo complejo del algoritmo, empezaremos identificando los elementos relevantes.



Grado del polinomio:  $n = 4$ .

Coefficientes del polinomio:  $a_0 = 1$ ;  $a_1 = -8$ ;  $a_2 = 39$ ;  $a_3 = -62$ ;  $a_4 = 50$ .

Factor cuadrático:  $u_0 = 2a = 2(0) = 0$ ;  $v_0 = -a^2 - b^2 = 0^2 - 0^2 = 0$ .

Cálculo de los coeficientes  $b$  de  $q(x)$

$$\begin{aligned}
 b_0 &= a_0 = 1 \\
 b_1 &= a_1 + u_0 b_0 = -8 + 0(1) = -8 \\
 b_2 &= a_2 + u_0 b_1 + v_0 b_0 = 39 + 0(-8) + 0(1) = 39 \\
 b_3 &= a_3 + u_0 b_2 + v_0 b_1 = -62 + 0(39) + 0(-8) = -62 \\
 b_4 &= a_4 + u_0 b_3 + v_0 b_2 = 50 + 0(-62) + 0(39) = 50
 \end{aligned}$$

Recuérdese que  $b_3$  y  $b_4$  deberán tender a cero, en caso de convergencia.

Cálculo de las derivadas parciales  $c$

$$\begin{aligned}
 c_0 &= b_0 = 1 \\
 c_1 &= b_1 + u_0 c_0 = -8 + 0(1) = -8 \\
 c_2 &= b_2 + u_0 c_1 + v_0 c_0 = 39 + 0(-8) + 0(1) = 39 \\
 c_3 &= b_3 + u_0 c_2 + v_0 c_1 = -62 + 0(39) + 0(-8) = -62
 \end{aligned}$$



Formando el sistema linearizado se obtiene:

$$\begin{aligned} c_{n-2}h + c_{n-3}k &= -b_{n-1} \quad \text{o bien } c_2h + c_1k = -b_3 \quad \text{o bien } 39h + 8k = -(-62) \\ c_{n-1}h + c_{n-2}k &= -b_n \quad \text{o bien } c_3h + c_2k = -b_4 \quad \text{o bien } -62h + 39k = -50 \end{aligned}$$

Al resolver este sistema por la regla de Cramer se obtiene:

$$\begin{aligned} h &= \frac{b_4c_1 - b_3c_2}{c_2^2 - c_3c_1} = \frac{50(-8) - (-62)(39)}{39^2 - (-62)(-8)} = 1.96878 \\ k &= \frac{b_3c_3 - b_4c_2}{c_2^2 - c_3c_1} = \frac{-62(-62) - 50(39)}{39^2 - (-62)(-8)} = 1.84780 \end{aligned}$$

Cálculo del nuevo factor cuadrático

$$\begin{aligned} u_1 &= u_0 + h = 0 + 1.96878 = 1.96878 \\ v_1 &= v_0 + k = 0 + 1.84780 = 1.84780 \end{aligned}$$

Las nuevas aproximaciones a las raíces son:

$$\begin{aligned} x &= a + bi; \text{ donde} \\ a &= u_1/2 = 1.96878 / 2 = 0.98439 \end{aligned}$$

y

$$\begin{aligned} b &= \pm \sqrt{-v_1 - a^2} = \pm 1.67834i \\ x &= 0.98439 \pm 1.67834i. \end{aligned}$$

Para ver si el proceso converge puede evaluarse el polinomio en las diferentes aproximaciones a las raíces y ver si  $|p(x)| \leq \varepsilon$ , en donde  $\varepsilon$ , en este caso podría tomarse como  $10^{-5}$ . La ecuación 4.28 queda entonces:

$$\begin{aligned} p(x) &= x^4 - 8x^3 + 39x^2 - 62x + 50 = -31.6831 \pm 11.3870i \\ |p(x)| &= \sqrt{(-31.6831)^2 + (11.3870)^2} = 33.6671 \end{aligned}$$

Al continuar el proceso iterativo, se obtienen los siguientes resultados:

### Segunda Iteración

| $k$      | 0       | 1        | 2        | 3         | 4        |
|----------|---------|----------|----------|-----------|----------|
| $b_k$    | 1       | -6.03122 | 28.97366 | -16.10175 | 71.83686 |
| $c_k$    | 1       | -4.06244 | 22.82341 | 21.32595  |          |
| $ p(x) $ | 13.4132 |          |          |           |          |

Con estos valores  $x = 1.04666 \pm 0.56619i$

### Tercera Iteración

| $k$      | 0        | 1        | 2        | 3        | 4        |
|----------|----------|----------|----------|----------|----------|
| $b_k$    | 1        | -5.90668 | 25.21935 | -0.84351 | 12.52183 |
| $c_k$    | 1        | -3.81336 | 15.82070 | 37.67428 |          |
| $ p(x) $ | 0.170093 |          |          |          |          |

Con estos valores  $x = 1.00299 \pm 0.99679i$

## Cuarta Iteración

| $k$      | 0                     | 1        | 2        | 3        | 4       |
|----------|-----------------------|----------|----------|----------|---------|
| $b_k$    | 1                     | -5.99401 | 24.97649 | 0.08813  | 0.23405 |
| $c_k$    | 1                     | -3.98802 | 14.97697 | 38.10619 |         |
| $ p(x) $ | $3.86 \times 10^{-4}$ |          |          |          |         |

Con estos valores  $x = 1.00000033 \pm 0.9999990063i$

Se puede ver que el proceso está convergiendo a las raíces  $x = 1 \pm li$ . En el disco se encuentra el programa 4.4 que realiza estos cálculos.

## Ejercicios

- 4.1 Uno de los problemas de ingeniería química que mejor ilustra la reducción de ecuaciones es el cálculo de la fracción de vapor  $V/F$  en una vaporización instantánea (véase ejercicio 2.6), donde se tienen las ecuaciones

$$F = L + V \quad (1)$$

$$Fz_i = Lx_i + Vy_i \quad i = 1, 2, \dots, n \quad (2)$$

provenientes del balance de materiales, y las relaciones de equilibrio líquido-vapor

$$K_i = \frac{y_i}{x_i} \quad i = 1, 2, \dots, n \quad (3)$$

donde:

$$K_i = \frac{P_i^0}{P} \quad i = 1, 2, \dots, n \quad (4)$$

y

$$P_i^0 = 10^{A_i - B_i/(C_i + T - 273.15)} \quad i = 1, 2, \dots, n \quad (5)$$

con las constantes  $A_i$ ,  $B_i$  y  $C_i$  dadas para cada componente  $i$ .

Además se tiene:

$$\sum_{i=1}^n x_i - \sum_{i=1}^n y_i = 0 \quad (6)$$

$$\sum_{i=1}^n z_i = 1 \quad (7)$$

Por otro lado, se tiene en estos problemas generalmente especificadas:  $z_i$ ,  $i = 1, 2, \dots, n - 1$ ,  $P$ ,  $T$  y  $F$ .

Para un número de componente  $n = 9$ : por ejemplo, se tiene entonces un sistema de 39 ecuaciones en las 39 incógnitas:  $L$ ,  $V$ ,  $x_i$ ,  $y_i$ ,  $K_i$ ,  $P_i$ ,  $i = 1, 2, \dots, 9$  y  $z_9$ , que puede reducirse, en general, como sigue.

Al combinar las ecuaciones (2) y (3) se eliminan las  $y_i$  y se obtiene:

$$x_i = \frac{z_i F}{(K_i V + F)} \quad (8)$$

Al combinar las ecuaciones (6) y (3):

$$\sum_{i=1}^n x_i - \sum_{i=1}^n K_i x_i = 0$$

o bien:

$$\sum_{i=1}^n x_i (1 - K_i) = 0 \quad (9)$$

con la sustitución de (8) en (9) se tiene:

$$\sum_{i=1}^n \frac{z_i F(1 - K_i)}{K_i V + L} = 0 \quad (10)$$

Pero de (1)  $L = F - V$ , con lo que queda finalmente:

$$\sum_{i=1}^n \frac{z_i (1 - K_i)}{V (K_i - 1) + F} = 0 \quad (11)$$

Hay que observar que si se conocen los valores de  $z_i$ ,  $i = 1, 2, \dots, n - 1$  (usando la ecuación (7) se obtiene  $z_n$ ), los valores de  $A_i$ ,  $B_i$ ,  $C_i$ ,  $i = 1, 2, \dots, n$  y los valores de  $P$  y  $T$  (usando (5) y (4) se obtiene  $K_i$ ,  $i = 1, 2, \dots, n$ ) y  $F$ , la ecuación (11) es ya sólo función de  $V$ , con lo que se ha reducido el sistema de 39 ecuaciones con 39 incógnitas a una sola ecuación con una incógnita ( $V$ ), cuya solución puede obtenerse con alguno de los métodos del capítulo 2.

- 4.2** La presión requerida para sumergir un objeto pesado grande en un terreno suave y homogéneo, que se encuentra sobre un terreno de base dura, puede predecirse a partir de la presión requerida para sumergir objetos más pequeños en el mismo suelo.\* En particular, la presión  $p$  requerida para sumergir una lámina circular de radio  $r$ , a una distancia  $d$  en el terreno suave, donde el terreno de base dura se encuentra a una distancia  $D > d$  debajo de la superficie, puede aproximarse mediante una ecuación de la forma:

$$p = k_1 \exp(k_2 r) + k_3 r \quad (1)$$

donde  $k_1$ ,  $k_2$  y  $k_3$  son constantes que, con  $k_2 > 0$ , dependen de  $d$  y la consistencia del terreno pero no del radio de la lámina.

- Encuentre los valores de  $k_1$ ,  $k_2$  y  $k_3$  si se supone que una lámina de radio de 1 pulgada requiere una presión de 10 lb/pulg<sup>2</sup> para sumergirse 1 pie en el terreno lodoso; una lámina de radio 2 pulgadas, requiere una presión de 12 lb/pulg<sup>2</sup> para sumergirse 1 pie; y una lámina de radio 3 pulgadas requiere una presión de 15 lb/pulg<sup>2</sup> (suponiendo que el lodo tiene una profundidad mayor que 1 pie).
- Use los cálculos de a) para predecir cuál es la lámina circular de radio mínimo que se necesitaría para sostener un peso de 500 lb en este terreno, con un hundimiento de menos de 1 pie.

### Solución Inciso a)

Al sustituir los valores de  $r$  y  $p$  en (1) para los tres casos, se tiene:

$$10 = k_1 \exp(k_2) + k_3$$

$$12 = k_1 \exp(2k_2) + 2k_3$$

$$15 = k_1 \exp(3k_2) + 3k_3$$

\* Richard L. Burden y J. Douglas Faires. *Análisis numérico*. Grupo Editorial Iberoamericana (1985).



un sistema de tres ecuaciones no lineales en las incógnitas  $k_1$ ,  $k_2$  y  $k_3$ . Se despeja  $k_3$  de la primera ecuación

$$k_3 = 10 - k_1 \exp(k_2)$$

Se sustituye  $k_3$  en las dos restantes y se tiene:

$$12 = k_1 \exp(2k_2) + 2[10 - k_1 \exp(k_2)]$$

$$15 = k_1 \exp(3k_2) + 3[10 - k_1 \exp(k_2)]$$

o bien:

$$f_1(k_1, k_2) = k_1 [\exp(2k_2) - 2\exp(k_2)] + 8 = 0 \quad (2)$$

$$f_2(k_1, k_2) = k_1 [\exp(3k_2) - 3\exp(k_2)] + 15 = 0$$

un sistema de dos ecuaciones no lineales en las incógnitas  $k_1$  y  $k_2$ .

Al dividir miembro a miembro estas dos ecuaciones

$$\frac{k_1 [\exp(2k_2) - 2\exp(k_2)]}{k_1 [\exp(3k_2) - 3\exp(k_2)]} = \frac{-8}{-15}$$

se obtiene:

$$\exp(k_2) - \frac{8}{15} \exp(2k_2) - \frac{6}{15} = 0$$

o bien:

$$f(k_2) = 15 \exp(k_2) - 8 \exp(2k_2) - 6 = 0 \quad (3)$$

una ecuación no lineal en la incógnita  $k_2$ , cuya solución con el método de Newton-Raphson visto en el capítulo 2 es:

$$k_2 = 0.259695;$$

al sustituir  $k_2$  en cualquiera de las ecuaciones (2) y despejar se tiene:

$$k_1 = \frac{-8}{\exp(2k_2) - 2\exp(k_2)} = 8.771286,$$

por último:

$$k_3 = 10 - k_1 \exp(k_2) = -1.372281$$

### Inciso b)

Un peso de 500 lb sobre un disco de radio  $r$  producirá una presión de  $500/(\pi r^2)$  lb/pulg<sup>2</sup>. Entonces:

$$p = \frac{500}{\pi r^2} = k_1 \exp(k_2 r) + k_3 r$$

o bien:

$$f'(r) = k_1 \exp(k_2 r) + k_3 r - \frac{500}{\pi r^2} = 0$$

Para obtener el valor mínimo de  $r$ , se iguala  $f'(r)$  con cero

$$f'(r) = k_1 k_2 \exp(k_2 r) + k_3 + \frac{1000r}{[\pi r^2]^2} = 0$$

lo que origina una ecuación no lineal en la incógnita  $r$ , cuya solución por medio de alguno de los métodos del capítulo 2 da

$$r = 3.18516 \text{ pulg.}$$

que corresponde a un mínimo de  $f(r)$ . El lector puede verificar esto usando alguno de los criterios del cálculo diferencial.

**4.3** Resuelva el siguiente sistema verificando primero su partición.

$$e_1: x_1 + x_4 - 10 = 0$$

$$e_2: x_2^2 x_4 x_3 - x_5 - 6 = 0$$

$$e_3: x_1 x_2^{1.7} (x_4 - 5) - 8 = 0$$

$$e_4: x_4 - 3x_1 + 6 = 0$$

$$e_5: x_1 x_3 - x_5 + 6 = 0$$

**Solución** Si bien la descomposición de un sistema en subsistemas es conocida como partición, la secuencia para resolver los subsistemas resultantes se denomina **orden de precedencia** del sistema. Existen algoritmos para partir un conjunto de ecuaciones y determinar el orden de precedencia. A continuación se seguirán las ideas de estos algoritmos a fin de partir el sistema dado.

a) Se forma una matriz de incidencia

$$\begin{array}{c} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \end{array} \begin{array}{ccccc} x_1 & x_2 & x_3 & x_4 & x_5 \\ \left[ \begin{array}{ccccc} 1 & & & 1 & \\ & 1 & 1 & 1 & 1 \\ & 1 & & 1 & \\ 1 & & & 1 & \\ 1 & & 1 & & 1 \end{array} \right] \end{array}$$

donde cada fila corresponde a una ecuación y cada columna a una variable. Un 1 aparece en la fila  $i$  y la columna  $j$ , si la variable  $x_j$  aparece en la ecuación  $e_i$ .

b) Se rearreglan las filas y columnas para apreciar mejor las particiones y el orden de precedencia. Así después de un rearreglo se llega a

$$\begin{array}{c} e_1 \\ e_4 \\ e_3 \\ e_5 \\ e_2 \end{array} \begin{array}{ccccc} x_1 & x_4 & x_2 & x_3 & x_5 \\ \left[ \begin{array}{ccccc} \left[ \begin{array}{cc} 1 & 1 \end{array} \right] & & & & \\ \left[ \begin{array}{cc} 1 & 1 \end{array} \right] & & & & \\ 1 & 1 & [1] & & \\ 1 & & & \left[ \begin{array}{cc} 1 & 1 \end{array} \right] & \\ & & & \left[ \begin{array}{cc} 1 & 1 \end{array} \right] & \end{array} \right] \end{array}$$

donde se nota de inmediato que en las ecuaciones  $e_1$  y  $e_4$  aparecen solamente las variables  $x_1$  y  $x_4$ , y constituyen entonces un subsistema que puede resolverse primero

$$e_1: x_1 + x_4 = 10$$

$$e_4: -3x_1 + x_4 = -6$$

resulta  $x_2 = 4$  y  $x_4 = 6$ .

Estos valores se sustituyen en la ecuación  $e_3$  y ésta queda en función de  $x_2$  solamente; por tanto, como una ecuación en una incógnita

$$e_3: 4x_2^{1.7} - 8 = 0$$

resulta  $x_2 = 1.5034$

Finalmente, las ecuaciones  $e_2$  y  $e_5$  pueden resolverse para  $x_3$  y  $x_5$ , lo que da

$$x_3 = 1.255$$

$$x_5 = 11.0202$$



**4.4** En un reactor se efectúan las siguientes reacciones en fase gaseosa:



A la temperatura de la reacción, las constantes de equilibrio son  $kp_1=2.6$  y  $kp_2=3.1$ . Las composiciones iniciales son 2 mol/L de A y 1 mol/L de B.

Calcule la composición a la salida del reactor, asumiendo que se alcanza el equilibrio.

**Solución** Si  $x_1$  representa los moles de A convertidos en la reacción (1), y  $x_2$  los moles de A convertidos en la reacción (2), entonces en el equilibrio tenemos:

$$\text{moles de A} = 2 - x_1 - x_2$$

$$\text{moles de B} = 1 - x_1$$

$$\text{moles de C} = x_1 - x_2$$

$$\text{moles de D} = x_1$$

$$\text{moles de E} = 2x_2$$

$$\text{moles totales} = 3$$

Con la aplicación de la ley de acción de masas se obtiene:

**Para la reacción (1)**

$$2.6 = \frac{(x_1 - x_2)(x_1)}{(2 - x_1 - x_2)(1 - x_1)}$$

**Para la reacción (2)**

$$3.1 = \frac{(2x_2)^2}{(2 - x_1 - x_2)(x_1 - x_2)}$$

que es un sistema de dos ecuaciones no lineales en dos incógnitas, cuya solución por el método de Newton-Raphson, por ejemplo, exige:

- Un vector inicial cercano a la solución, obtenible a partir de consideraciones físicas del problema.
- La matriz jacobiana, ampliada con el vector de funciones, que es relativamente fácil, puesto que las derivadas parciales son directas.

Vector inicial. En virtud de las funciones y la existencia inicial de 2 moles de A y 1 mol de B, se propone  $x_1 = 0.8$  y  $x_2 = 0.4$ .

Las derivadas parciales para la matriz jacobiana se dan a continuación



$$f_1(x_1, x_2) = \frac{(x_1 - x_2)(x_1)}{(2 - x_1 - x_2)(1 - x_1)} - 2.6 = 0$$

$$f_2(x_1, x_2) = \frac{(2x_2)^2}{(2 - x_1 - x_2)(x_1 - x_2)} - 3.1 = 0$$

$$\frac{\partial f_1(x_1, x_2)}{\partial x_1} = \frac{(2 - x_1 - x_2)(1 - x_1)(2x_1 - x_2) - (x_1 - x_2)(x_1)(-3 + 2x_1 + x_2)}{((2 - x_1 - x_2)(1 - x_1))^2}$$

$$\frac{\partial f_1(x_1, x_2)}{\partial x_2} = \frac{(2 - x_1 - x_2)(1 - x_1)(x_1) + (x_1 - x_2)(x_1)(x_1 - 1)}{((2 - x_1 - x_2)(1 - x_1))^2}$$

$$\frac{\partial f_2(x_1, x_2)}{\partial x_1} = \frac{-(2x_2)^2(2 - 2x_1)}{((2 - x_1 - x_2)(x_1 - x_2))^2}$$

$$\frac{\partial f_2(x_1, x_2)}{\partial x_2} = \frac{8(2 - x_1 - x_2)(x_1 - x_2)x_2 - 8x_2^2(-1 + x_2)}{((2 - x_1 - x_2)(x_1 - x_2))^2}$$

Con el **PROGRAMA 4.1** del CD se obtienen los siguientes resultados:

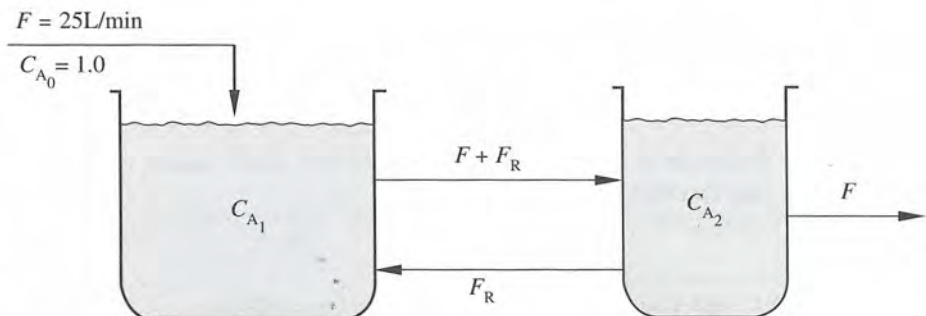
| k | X(1)    | X(2)    | Distancia    |
|---|---------|---------|--------------|
| 0 | 0.80000 | 0.40000 |              |
| 1 | 0.82175 | 0.46596 | 6.94595e-002 |
| 2 | 0.83460 | 0.45687 | 1.57387e-002 |
| 3 | 0.83176 | 0.45571 | 3.06403e-003 |
| 4 | 0.83145 | 0.45566 | 3.18286e-004 |
| 5 | 0.83144 | 0.45565 | 1.21513e-005 |
| 6 | 0.83144 | 0.45565 | 4.10222e-007 |

La solución del sistema es:

$$X(1) = 0.83144$$

$$X(2) = 0.45565$$

- 4.5** El mezclado imperfecto en un reactor continuo de tanque agitado, se puede modelar como dos o más reactores con recirculación entre ellos, como se muestra en la figura siguiente.



**Figura 4.15**  
Reactores  
químicos con  
recirculación.

En este sistema se lleva a cabo una reacción isotérmica irreversible del tipo  $A \xrightarrow{k} B$  de orden 1.8 con respecto al reactante A. Con los datos que se proporcionan abajo, calcule la concentración del reactante A en los reactores 1 y 2 ( $C_{A1}$  y  $C_{A2}$  respectivamente), una vez alcanzado el régimen permanente.

Datos

$$\begin{aligned} F &= 25 \text{ L/min} & V_1 &= 80 \text{ L} \\ C_{A0} &= 1 \text{ mol/L} & V_2 &= 20 \text{ L} \\ F_R &= 100 \text{ L/min} & k &= 0.2 \text{ (L/mol)}^{0.8} \text{ (min}^{-1}\text{)} \end{aligned}$$

**Solución** Con el balance del componente A en cada uno de los reactores se tiene

$$\text{Entra} \quad - \quad \text{Sale} \quad - \quad \text{Reacciona} \quad = \quad \text{Acumulación}$$

**Reactor 1**

$$F C_{A0} + F_R C_{A2} - (F + F_R) C_{A1} - V_1 k C_{A1}^n = 0 \quad (1)$$

**Reactor 2**

$$(F + F_R) C_{A1} - (F_R + F) C_{A2} - V_2 k C_{A2}^n = 0 \quad (2)$$

un sistema de dos ecuaciones no lineales en las incógnitas  $C_{A1}$  y  $C_{A2}$ .

No obstante, se observa que despejando a  $C_{A2}$  de la ecuación (1)

$$C_{A2} = \frac{(F + F_R) C_{A1} + V_1 k C_{A1}^n - F C_{A0}}{F_R}$$

y sustituyéndola en la ecuación (2)

$$\begin{aligned} 125 C_{A1} - 125 \frac{(F + F_R) C_{A1} + V_1 k C_{A1}^n - F C_{A0}}{F_R} \\ - k V_2 \left[ \frac{(F + F_R) C_{A1} + V_1 k C_{A1}^n - F C_{A0}}{F_R} \right]^n = 0 \end{aligned}$$

el problema se reduce a una ecuación no lineal en la incógnita  $C_{A1}$ , cuya solución se encuentra empleando alguno de los métodos del capítulo 2 y se deja al lector como ejercicio.

Resultados:

$$C_{A1} = 0.6493$$

$$C_{A2} = 0.6352$$



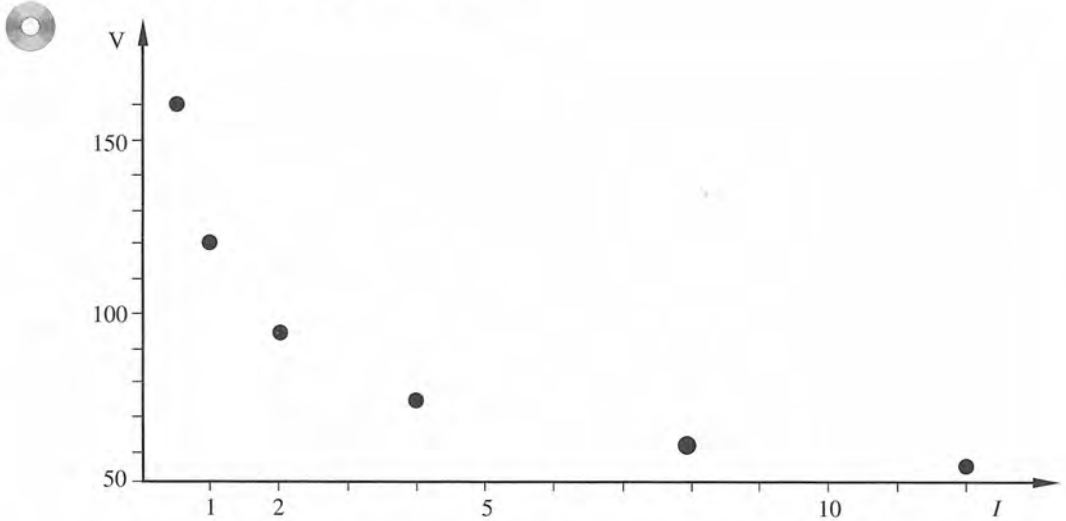
**4.6** En una lámpara de arco,\* de longitud de arco constante, se observa el voltaje  $V$  empleado por el arco para diversos valores de la corriente  $I$

|   |     |     |    |    |    |    |
|---|-----|-----|----|----|----|----|
| I | 0.5 | 1   | 2  | 4  | 8  | 12 |
| V | 160 | 120 | 94 | 75 | 62 | 56 |

Encuentre la ecuación que mejor represente estos valores, empleando el criterio de mínimos cuadrados.

\* J. Lipka. *Computaciones gráficas y mecánicas*. Lipka J. CECSA.

**Solución** Se traza el diagrama de dispersión



y se observe que la curva suave que pasa entre los puntos es hiperbólica y asíntota a alguna recta horizontal  $V = c$ . Con esto, se supone que los datos pueden quedar relacionados por la ecuación

$$V = a I^b + c \quad (1)$$

donde\*  $b < 0$ .

Los parámetros  $a$ ,  $b$  y  $c$  se determinan minimizando la función

$$f(a, b, c) = \sum_{i=1}^6 (V_i - a I_i^b - c)^2 \quad (2)$$

La ecuación (2) se deriva parcialmente con respecto a  $a$ ,  $b$  y  $c$ , y se igualan a cero dichas derivadas parciales para obtener:

$$\begin{aligned} \sum_{i=1}^6 V_i I_i^b - a \sum_{i=1}^6 I_i^{2b} - c \sum_{i=1}^6 I_i^b &= 0 \\ \sum_{i=1}^6 V_i I_i^b \ln I_i - a \sum_{i=1}^6 I_i^{2b} \ln I_i - c \sum_{i=1}^6 I_i^b \ln I_i &= 0 \\ \sum_{i=1}^6 V_i - a \sum_{i=1}^6 I_i^b - 6c &= 0 \end{aligned} \quad (3)$$

un sistema de tres ecuaciones no lineales en las incógnitas  $a$ ,  $b$  y  $c$ .

Al despejar  $c$  de la tercera ecuación:

$$c = \frac{1}{6} \sum_{i=1}^6 V_i - \frac{a}{6} \sum_{i=1}^6 I_i^b \quad (4)$$

y sustituir en las dos primeras, se tiene:

$$f_1(a, b) = \sum_{i=1}^6 V_i I_i^b - a \sum_{i=1}^6 I_i^{2b} - \frac{1}{6} \left[ \sum_{i=1}^6 V_i \right] \left[ \sum_{i=1}^6 I_i^b \right] + \left[ \frac{a}{6} \sum_{i=1}^6 I_i^b \right]^2 = 0 \quad (5)$$

$$f_2(a, b) = \sum_{i=1}^6 V_i I_i^b \ln I_i - a \sum_{i=1}^6 I_i^{2b} \ln I_i - \frac{1}{6} \left[ \sum_{i=1}^6 V_i \right] \left[ \sum_{i=1}^6 I_i^b \ln I_i \right] + \frac{a}{6} \left[ \sum_{i=1}^6 I_i^b \right] \left[ \sum_{i=1}^6 I_i^b \ln I_i \right] = 0$$

\*  $b > 0$  en el caso de una parábola, con ordenada al origen  $c$ .



un sistema de dos ecuaciones no lineales en las incógnitas  $a$  y  $b$ , cuya solución requiere valores iniciales.

Para estimar valores iniciales, en la ecuación (1) se sustituyen tres de los puntos dados

$$160 = a 0.5^b + c$$

$$75 = a 4^b + c$$

$$56 = a 12^b + c$$

al despejar  $c$  de la tercera y sustituir en las dos primeras se tiene:

$$160 = a 0.5^b + 56 - a 12^b$$

$$75 = a 4^b + 56 - a 12^b$$

o bien

$$a (0.5^b - 12^b) = 104$$

$$a (4^b - 12^b) = 19$$

Estas dos últimas ecuaciones se dividen miembro a miembro

$$\frac{0.5^b - 12^b}{4^b - 12^b} = \frac{104}{19}$$

se rearregla

$$19 (0.5)^b + 85 (12)^b - 104 (4)^b = 0$$

y se resuelve esta ecuación no lineal con alguno de los métodos del capítulo 2 para obtener:

$$b = -0.51952$$

de donde:

$$a = 89.77$$

El sistema (5) se resuelve utilizando éstos como valores iniciales y el método de Newton-Raphson multivariable, con lo que resulta

$$a = 87.78$$

$$b = -0.532$$

y al sustituir en (4) se obtiene:

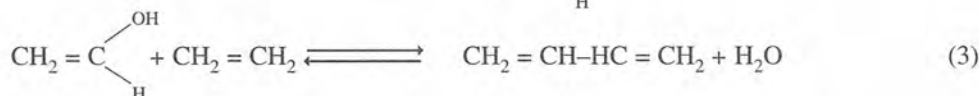
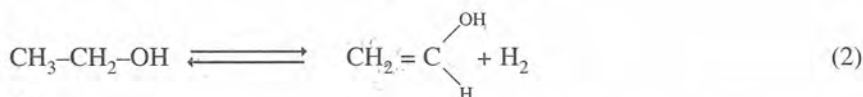
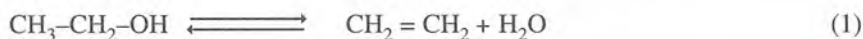
$$c = 32.86$$

De tal manera que la ecuación que mejor ajusta los datos queda:

$$V = 87.78 I^{-0.532} + 32.86$$



- 4.7 Para la obtención de butadieno a partir de etanol en fase vapor, se propone el siguiente mecanismo de reacción.



Calcule las composiciones en el equilibrio a 400 ° C y 1 atm, si las constantes de equilibrio son 5.97, 0.27 y 2.8 para las reacciones (1), (2) y (3), respectivamente.

**Solución** Base de cálculo: 1 mol de etanol.  
Si

$$\begin{aligned}x_1 &= \text{moles de etileno producidas en la reacción (1)} \\x_2 &= \text{moles de hidrógeno producidas en la reacción (2)} \\x_3 &= \text{moles de agua producidas en la reacción (3)}\end{aligned}$$

entonces en el equilibrio se tendrá:

$$\begin{aligned}\text{moles de etanol} &= 1 - x_1 - x_2 \\ \text{moles de etileno} &= x_1 - x_3 \\ \text{moles de agua} &= x_1 + x_3 \\ \text{moles de hidrógeno} &= x_2 \\ \text{moles de acetaldehído} &= x_2 - x_3 \\ \text{moles de butadieno} &= x_3 \\ \hline \text{moles totales} &= 1 + x_1 + x_2\end{aligned}$$

De acuerdo con la ley de acción de masas, se tiene:

$$\begin{aligned}5.97 &= \frac{(x_1 + x_3)(x_1 - x_3)}{(1 - x_1 - x_2)} \left[ \frac{P}{1 + x_1 + x_2} \right]^{\Delta n_1} \\ 0.27 &= \frac{(x_2 + x_3)x_2}{(1 - x_1 - x_2)} \left[ \frac{P}{1 + x_1 + x_2} \right]^{\Delta n_2} \\ 2.8 &= \frac{(x_1 + x_3)x_3}{(x_1 - x_3)(x_2 - x_3)} \left[ \frac{P}{1 + x_1 + x_2} \right]^{\Delta n_3}\end{aligned}$$

donde  $\Delta n_i$  = número de moles de los productos-número de moles de los reactantes (en la reacción  $i$ ).

Por tanto:

$$\begin{aligned}\Delta n_1 &= 2 - 1 = 1 \\ \Delta n_2 &= 2 - 1 = 1 \\ \Delta n_3 &= 2 - 2 = 0\end{aligned}$$

Por otro lado:

$$P = 1 \text{ atm.}$$

Vector inicial. Luego de observar las funciones y el hecho de que la base de cálculo es 1 mol de etanol, se propone

$$x_1 = 0.7, \quad x_2 = 0.2, \quad x_3 = 0.1$$

Si se utilizaran las ecuaciones del sistema tal como están, se tendrían serios problemas, ya que si  $x_1 + x_2 = 1$ , habría división entre cero. Un reacomodo de las ecuaciones permitiría no sólo evitar la división entre cero, sino obtener una convergencia más rápida. Por ejemplo, podría escribirse el sistema así:

$$\begin{aligned}P^{\Delta n_1} (x_1 + x_3)(x_1 - x_3) - 5.97(1 - x_1 - x_2)(1 + x_1 + x_2)^{\Delta n_1} &= 0 \\ P^{\Delta n_2} (x_2 + x_3)x_2 - 0.27(1 - x_1 - x_2)(1 + x_1 + x_2)^{\Delta n_2} &= 0 \\ P^{\Delta n_3} (x_1 + x_3)x_3 - 2.8(x_1 - x_3)(x_2 - x_3)(1 + x_1 + x_2)^{\Delta n_3} &= 0\end{aligned}$$

Luego de sustituir valores y resolver el sistema de ecuaciones no lineales resultante con el **PROGRAMA 4.1** del CD, se llega a los siguientes resultados

$$X(1) = 0.71230$$

$$X(2) = 0.24645$$

$$X(3) = 0.15792$$

- 4.8** En una columna de cinco platos, se quiere absorber tolueno contenido en una corriente de gas  $V_0$  (moles de gas sin tolueno/min), con un aceite  $L_0$  (moles de aceite sin tolueno/min). Considérese que la relación de equilibrio está dada por la ley de Henry ( $y = m x$ ), y que la columna opera a régimen permanente. Calcule la composición de tolueno en cada plato.

Datos:  $V_0 = 39.6$  moles/min

$L_0 = 6.0$  moles/min

Las moles de tolueno/min que entran a la columna con el gas y el aceite, son respectivamente:

$TV_0 = 5.4$  moles/min

$TL_0 = 0.0$  moles/min

$m = 0.155$

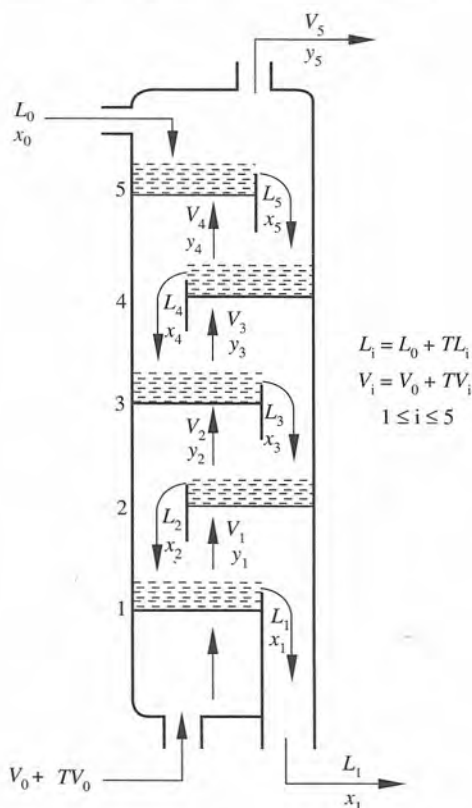
De aquí:

$$y_0 = \frac{5.4}{5.4 + 39.6} = 0.12 \quad \text{fracción mol de tolueno en el gas que entra.}$$

**Solución** Los balances de masa para el tolueno en cada plato son (véase Fig. 4.16).

Plato

Balace de tolueno



**Figura 4.16**  
Columna  
de absorción de  
cinco platos.



$$\begin{aligned}
1 \quad & (V_0 + TV_0)y_0 - (V_0 + TV_1)y_1 + (L_0 + TL_2)x_2 - (L_0 + TL_1)x_1 = 0 \\
2 \quad & (V_0 + TV_1)y_1 - (V_0 + TV_2)y_2 + (L_0 + TL_3)x_3 - (L_0 + TL_2)x_2 = 0 \\
3 \quad & (V_0 + TV_2)y_2 - (V_0 + TV_3)y_3 + (L_0 + TL_4)x_4 - (L_0 + TL_3)x_3 = 0 \\
4 \quad & (V_0 + TV_3)y_3 - (V_0 + TV_4)y_4 + (L_0 + TL_5)x_5 - (L_0 + TL_4)x_4 = 0 \\
5 \quad & (V_0 + TV_4)y_4 - (V_0 + TV_5)y_5 + (L_0 + TL_0)x_0 - (L_0 + TL_5)x_5 = 0
\end{aligned}$$

donde  $TV_i$ ,  $TL_i$ ,  $0 \leq i \leq 5$ , son los moles de tolueno/min que salen del plato  $i$  con el gas y el aceite, respectivamente.

Como

$$y_i = \frac{TV_i}{TV_i + V_0} \quad \text{y demás } y_i = mx_i,$$

se obtiene:

$$TV_i = \frac{V_0 mx_i}{1 - mx_i}$$

Por otro lado:

$$TL_i = \frac{L_0 x_i}{1 - x_i} \quad \text{para } 0 \leq i \leq 5$$

Con la sustitución de  $y_i$ ,  $TV_i$  y  $TL_i$  en los balances de masa anteriores, resulta el sistema no lineal siguiente:

$$V_0 y_0 + \frac{V_0 y_0^2}{1 - y_0} - V_0 m x_1 - \frac{V_0 m^2 x_1^2}{1 - m x_1} + L_0 x_2 + \frac{L_0 x_2^2}{1 - x_2} - L_0 x_1 - \frac{L_0 x_1^2}{1 - x_1} = 0$$

$$V_0 m x_1 + \frac{V_0 m^2 x_1^2}{1 - m x_1} - V_0 m x_2 - \frac{V_0 m^2 x_2^2}{1 - m x_2} + L_0 x_3 + \frac{L_0 x_3^2}{1 - x_3} - L_0 x_2 - \frac{L_0 x_2^2}{1 - x_2} = 0$$

$$V_0 m x_2 + \frac{V_0 m^2 x_2^2}{1 - m x_2} - V_0 m x_3 - \frac{V_0 m^2 x_3^2}{1 - m x_3} + L_0 x_4 + \frac{L_0 x_4^2}{1 - x_4} - L_0 x_3 - \frac{L_0 x_3^2}{1 - x_3} = 0$$

$$V_0 m x_3 + \frac{V_0 m^2 x_3^2}{1 - m x_3} - V_0 m x_4 - \frac{V_0 m^2 x_4^2}{1 - m x_4} + L_0 x_5 + \frac{L_0 x_5^2}{1 - x_5} - L_0 x_4 - \frac{L_0 x_4^2}{1 - x_4} = 0$$

$$V_0 m x_4 + \frac{V_0 m^2 x_4^2}{1 - m x_4} - V_0 m x_5 - \frac{V_0 m^2 x_5^2}{1 - m x_5} + L_0 x_0 + \frac{L_0 x_0^2}{1 - x_0} - L_0 x_5 - \frac{L_0 x_5^2}{1 - x_5} = 0$$

donde  $x_1, x_2, \dots, x_5$  son las incógnitas.

Este sistema se resuelve con el programa 4.2 con los siguientes valores iniciales

$$x_1 = 0.4, x_2 = 0.3, x_3 = 0.2, x_4 = 0.1, x_5 = 0.05,$$

los cuales se obtuvieron usando un perfil lineal de concentraciones a lo largo de la columna. Los resultados obtenidos son:

| k | X ( 1 ) | X ( 2 ) | X ( 3 ) | X ( 4 ) | X ( 5 ) | Distancia  |
|---|---------|---------|---------|---------|---------|------------|
| 0 | .40000  | .30000  | .20000  | .10000  | .05000  | ———        |
| 1 | .45756  | .30057  | .19940  | .12100  | .06020  | .62120E-01 |
| 2 | .45398  | .30115  | .20289  | .12717  | .06318  | .85044E-02 |
| 3 | .45432  | .30195  | .20424  | .12919  | .06416  | .27569E-02 |
| 4 | .45444  | .30222  | .20468  | .12986  | .06449  | .91471E-03 |
| 5 | .45448  | .30231  | .20483  | .13008  | .06460  | .30494E-03 |
| 6 | .45450  | .30234  | .20488  | .13016  | .06463  | .10179E-03 |
| 7 | .45450  | .30235  | .20489  | .13018  | .06465  | .34040E-04 |

La solución del sistema es:

$$X ( 1 ) = .45450091$$

$$X ( 2 ) = .30234605$$

$$X ( 3 ) = .20489225$$

$$X ( 4 ) = .13018015$$

$$X ( 5 ) = .64646289E-01$$

## Problemas

### 4.1 Resuelva el sistema

$$x_1 x_2 + x_6 x_4 = 18$$

$$x_2 + x_5 + x_6 = 12$$

$$x_1 + \ln(x_2/x_4) = 3$$

$$x_3^2 + x_3 = 2$$

$$x_2 + x_4 = 4$$

$$x_3(x_3 + 6) = 7$$

utilizando las sugerencias dadas al principio de este capítulo (reducción, partición, entre otros).

### 4.2 Resuelva el sistema

$$e_1: x_1 x_3 - x_4 = 1$$

$$e_2: x_2^2 x_3^2 + x_4 = 17$$

$$e_3: x_1 + x_2 = 6$$

$$e_4: \ln x_3 x_4^2 + x_3 x_4^2 = 1$$

mediante tanteo de ecuaciones.

### 4.3 A partir de consideraciones geométricas demuestre que el sistema no lineal

$$x^2 + y^2 - x = 0$$

$$x^2 - y^2 - y = 0$$

tiene una solución no trivial única. Además, obtenga una estimación inicial  $x^0, y^0$  y aproxíme dicha solución, empleando el método de punto fijo.

- 4.4 Dado el sistema de ecuaciones no lineales

$$x^2 + y = 37$$

$$x - y^2 = 5$$

determine un arreglo de la forma

$$g_1(x, y) = x$$

$$g_2(x, y) = y$$

y un vector inicial  $\mathbf{x}^{(0)}$  que prometa convergencia a una solución; es decir, que se satisfaga el sistema de desigualdades (Ec. 4.6).

- 4.5 Encuentre una solución del sistema de ecuaciones del problema anterior, por medio del método de Newton-Raphson y tomando como valor inicial

a)  $(x, y) = (5, 0)$

b)  $(x, y) = (5, -1)$

¿Qué criterios se pueden aplicar para saber si el proceso converge y, en tal caso, cómo se puede verificar que efectivamente se trata de una solución?

SUGERENCIA: Emplee el CD del libro.

- 4.6 Utilice el método de punto fijo multivariable para encontrar una solución de cada uno de los siguientes sistemas.

a)  $x_1(4 - 0.0003x_1 - 0.0004x_2) = 0$

$$x_2(2 - 0.0002x_1 - 0.0001x_2) = 0$$

b)  $x_1^2 + 2x_2^2 - x_2 - 2x_3 = 0$

$$x_1^2 - 8x_2^2 + 10x_3 = 0.0001$$

$$x_1^2/(7x_2x_3) - 1 = 0$$

c)  $2x_1 + x_2 + x_3 - 4\log(10x_1) = 0$

$$x_1 + 2x_2 + x_3 - 4\log(10x_2) = 0$$

$$x_1x_2x_3 - \log(10x_3) = 0$$

d)  $3x_1 \sin x_2 - \cos(x_2x_3) \sin x_2 - \sin^{-1}(-0.52356) \sin x_2 = 0$

$$x_1^2 - 625x_2^2 = 0$$

$$\exp(-x_1x_2) + 20x_3 = 9.471975$$

SUGERENCIA: Utilice el Mathcad o software equivalente.

- 4.7 Elabore un programa para resolver sistemas de ecuaciones no lineales. Utilice para ello el algoritmo 4.1.

- 4.8 Emplee el programa de problema 4.7 para resolver los sistemas del problema 4.6.

- 4.9 Mediante el PROGRAMA 4.1 del CD (véase ejemplo 4.4), resuelva los siguientes sistemas de ecuaciones no lineales.

a)  $(x_1 + \cos x_1x_2x_3 - 1)^{1/2} = 0$

$$(1 - x_1)^{1/4} + x_2 + x_3(0.05x_3 - 0.15) = 1$$

$$1 + x_1^2 + 0.1x_2^2 - 0.01x_2 - x_3 = 0$$

b)  $0.5 \sin(x_1x_2) - x_2/(4\pi) - 0.5x_1 = 0$

$$0.920423 [\exp(2x_1) - \exp(1)] + 8.65256x_2 - 2\exp(x_1) = 0$$

$$\text{Emplee } EPS = 10^{-4}.$$





- 4.10** Si en la aplicación del método de Newton–Raphson, en algún punto del proceso iterativo, por ejemplo  $\mathbf{x}^{(i)}$ , el determinante de la matriz jacobiana evaluado en ese punto es cero, o muy cercano a cero, dicho proceso no puede continuarse. ¿Qué hacer en tales casos? (véase Prob. 2.10).
- 4.11** Los métodos estudiados en este capítulo son aplicables también a sistemas de ecuaciones lineales y a ecuaciones no lineales en una variable, ya que estos dos son sólo casos particulares del caso general de sistemas de ecuaciones no lineales. Por ejemplo, si se aplicara el método de Newton–Raphson para resolver el sistema lineal

$$4x_1 - 9x_2 + 2x_3 = 5$$

$$2x_1 - 4x_2 + 6x_3 = 3$$

$$x_1 - x_2 + 3x_3 = 4$$

la matriz de derivadas parciales sería:

$$J = \begin{bmatrix} 4 & -9 & 2 \\ 2 & -4 & 6 \\ 1 & -1 & 3 \end{bmatrix}$$

Encuentre la solución utilizando el algoritmo 4.2 con un vector inicial adecuado.



- 4.12** Resuelva el problema 3.33 (considerando ahora que la reacción es de orden 0.5 con respecto a A y la constante de velocidad de reacción  $k_1$  es  $0.05 \text{ L}^{-0.5} \text{ mol}^{0.5} \text{ min}^{-1}$ . Emplee el programa del problema 4.7, o bien el **PROGRAMA 4.1** del CD.
- 4.13** Repita el problema 3.34, considerando que la reacción es de orden 0.5 y que la constante de velocidad de reacción es  $0.05 \text{ L}^{-0.5} \text{ mol}^{0.5} \text{ min}^{-1}$ . ¿La conversión de A mejora recirculando los tres tanques en lugar de recircular solamente el primero?
- 4.14** Utilice el método iterativo de punto fijo para resolver el sistema de ecuaciones no lineales del ejemplo 4.4, con el vector inicial

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

- a) con desplazamientos sucesivos  
b) con desplazamientos simultáneos

Compare la convergencia en los dos casos.

**SUGERENCIA:** Emplee el Mathcad o un software equivalente.

- 4.15** Resuelva el ejercicio 4.8, usando  $TV_0 = 9.9$
- 4.16** Resuelva los sistemas de los problemas 4.6 y 4.9 por el método de Newton–Raphson modificado.
- 4.17** Elabore un programa para resolver sistemas de ecuaciones no lineales por el método de Newton–Raphson modificado, utilizando para ello el algoritmo 4.3. Resuelva con dicho programa el sistema

$$x_1^2 + 2x_2^2 + \exp(x_1 + x_2) = 6.1718 - x_1 x_3$$

$$10x_2 = -x_2 x_3$$

$$\sin(x_1 x_3) + x_2^2 = 1.141 - x_1$$

utilizando como vector inicial a  $\mathbf{x}^{(0)} = [1, 1, 1]^T$ .

- 4.18 La siguiente tabla representa las temperaturas observadas  $T(^{\circ}\text{C})$  a diferentes tiempos  $t$  (min) del agua en un tanque de enfriamiento

|     |      |      |      |      |      |      |      |      |      |
|-----|------|------|------|------|------|------|------|------|------|
| $t$ | 0    | 1    | 2    | 3    | 5    | 7    | 10   | 15   | 20   |
| $T$ | 92.0 | 85.3 | 79.5 | 74.5 | 67.0 | 60.5 | 53.5 | 45.0 | 39.5 |

Encuentre la ecuación de enfriamiento que mejor represente estos valores, empleando el criterio de mínimos cuadrados. Véase ejercicio 4.6.



- 4.19 La relación entre el rendimiento de un cultivo y la cantidad de fertilizante  $x$ , aplicado a ese cultivo, se ha formulado así:

$$y = a - b d^x$$

donde  $0 < d < 1$

Dado los siguientes datos

|     |      |      |      |      |      |
|-----|------|------|------|------|------|
| $x$ | 0    | 1    | 2    | 3    | 4    |
| $y$ | 44.4 | 54.6 | 63.8 | 65.7 | 68.9 |

obtenga estimaciones de  $a$ ,  $b$  y  $d$  empleando el método de los mínimos cuadrados. (Véase ejercicio 4.6).

- 4.20 Resuelva los sistemas de ecuaciones no lineales del problema 4.9 con el método de Broyden. Compare el número de iteraciones requerido con el número requerido en los métodos de punto fijo y de Newton-Raphson multivariable. Emplee en la comparación  $EPS = \| \mathbf{x}^{(i)} - \mathbf{x}^{(i-1)} \| < 10^{-4}$ .

- 4.21 Elabore un programa de cómputo para resolver sistemas de ecuaciones no lineales con el método de Broyden.

Emplee para ello el algoritmo 4.4. Resuelva con dicho programa el sistema

$$\begin{aligned} x_1^2 + 2x_2^2 + \exp(x_1 + x_2) &= 6.1718 - x_1 x_3 \\ 10x_2 &= -x_2 x_3 \\ \text{sen}(x_1 x_3) + x_2^2 &= 1.141 - x_1, \end{aligned}$$

utilizando como vector inicial a  $\mathbf{x}^{(0)} = [1, 1, 1]^T$

- 4.22 El método de Broyden pertenece a una familia conocida como métodos de **Cuasi-Newton**. Otro de los miembros de dicha familia se obtiene al reemplazar a  $J^{(k)}$  de la ecuación 4.22 con una matriz  $A^{(k)}$ , cuyos componentes son las derivadas parciales numéricas; esto es, consiste en aproximar las derivadas parciales analíticas de la matriz jacobiana  $J$  por sus correspondientes derivadas parciales numéricas. Por ejemplo, para una función de dos variables  $f(x, y)$  las derivadas parciales numéricas quedan así:

$$\frac{\partial f}{\partial x} = \frac{f(x+h, y) - f(x, y)}{h}$$

y

$$\frac{\partial f}{\partial y} = \frac{f(x, y+h) - f(x, y)}{h}$$

donde  $h$  es un valor pequeño.

Con las ideas dadas, encuentre una solución aproximada del sistema de ecuaciones no lineales siguiente usando como vector inicial  $[x^0, y^0]^T = [0, 0]^T$

$$f_1(x, y) = x^2 - 10x + y^2 + 8 = 0$$

$$f_2(x, y) = xy^2 + x - 10y + 8 = 0$$

- 4.23** Resuelva los sistemas de ecuaciones no lineales de los problemas 4.6 y 4.9, mediante el método de Newton-Raphson con optimización de  $t$ .



**SUGERENCIA:** Emplee el **PROGRAMA 4.2** del CD.

- 4.24** Otra forma de seleccionar los valores del tamaño de la etapa  $t$  (véase Sec. 4.6), consiste en dividir el intervalo de búsqueda  $[a, b]$  en dos partes iguales sucesivamente. Esto es:

$$t_1 = (a + b)/2,$$

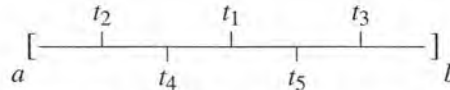
$$t_2 = (a + t_1)/2,$$

$$t_3 = (t_1 + b)/2$$

$$t_4 = (t_2 + t_1)/2,$$

$$t_5 = (t_3 + t_1)/2, \text{ etcétera}$$

Gráficamente:



Para cada valor de  $t$  se calcula el correspondiente  $z_{k+1}$ , y el valor mínimo de  $z_{k+1}$  proporcionará el valor óptimo de  $t$ . Encuentre el valor óptimo de  $t$  en la primer iteración de la solución del ejemplo 4.3 usando este método de cálculo de  $t$  y el intervalo  $[-1.2, -1]$ .



- 4.25** Modifique el **PROGRAMA 4.2** del CD de modo que se empleen los valores de  $t$  calculados en la forma indicada en el problema 4.24.
- 4.26** Resuelva los siguientes sistemas de ecuaciones algebraicas no lineales, proponiendo en cada caso vectores iniciales. Emplee en cada caso los métodos que juzgue más convenientes y el software de que disponga.

a)  $\sin x + \cos x - z = 0$

$$\exp(x + y) - x^2 \cos x - \pi/1.15 = 0$$

$$y + 3xz + x^3 = 0$$

b)  $\ln(x y) + x^2 y^2 = 8$

$$\sin x + y \exp(x) = 2$$

c)  $x_1^3 + x_2^3 - x_3^3 = 129$

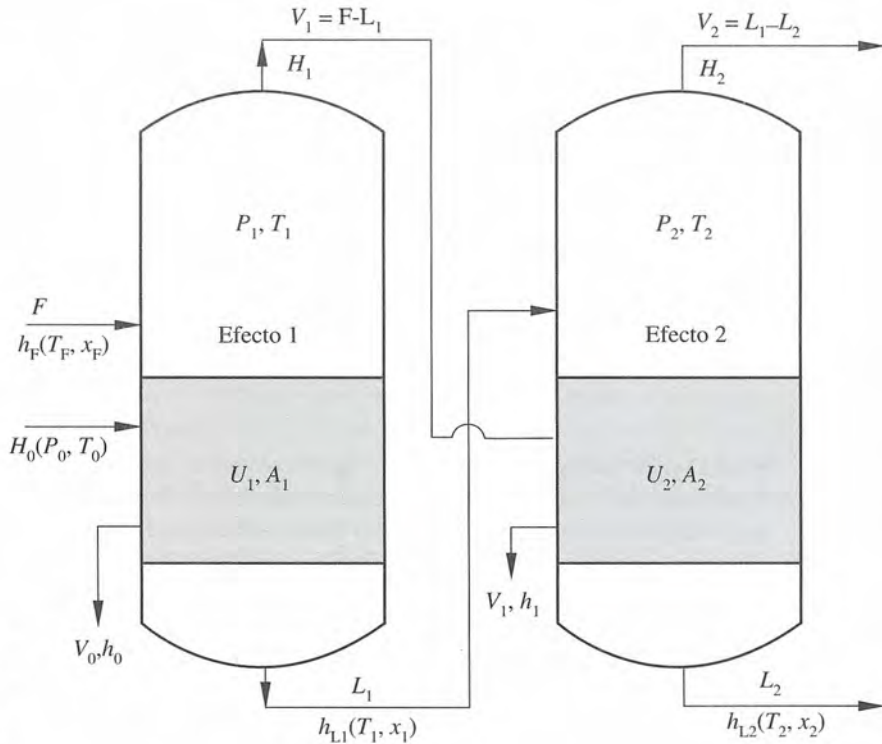
$$x_1^2 + x_2^2 - x_3^2 = 9.75$$

$$x_1 + x_2 - x_3 = 9.49$$

- 4.27** Se desea concentrar una solución con una concentración inicial de sólidos de 20% a una concentración final de 60% en un evaporador de doble efecto. Se dispone de vapor saturado a 0.68 atm (10 psig) y el segundo efecto que opera con una presión de vacío de 0.136 atm (2 psia). (Ver figura 4.17.)

Si la alimentación al sistema, 18,240.6 kg/h, entra al primer efecto a 93.3 °C, determine el área de los evaporadores,  $A_1$  y  $A_2$ , y la cantidad de vapor requerido.





**Figura 4.17**  
Sistema de  
evaporación  
de doble  
efecto.

Otros datos:

$$\begin{array}{ll} C_{p,F} = 0.9 \text{ kcal/(kg } ^\circ\text{C)} & U_1 = 3,516.5 \text{ kcal/(hm}^2\text{)} \\ C_{p,L1} = 0.8 & U_2 = 2,440.4 \\ C_{p,L2} = 0.8 \end{array}$$

**4.28** El método del eigenvalor (valor propio) dominante\* para resolver sistemas de ecuaciones no lineales, consiste en emplear el siguiente algoritmo

$$\mathbf{x}^{(k+2)} = \mathbf{x}^{(k)} + \frac{1}{1 - \lambda_1} [\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}]$$

donde  $\lambda_1$  es el eigenvalor dominante de la matriz jacobiana  $J$  (véase ecuación 4.17), evaluada en  $\mathbf{x}^{(k+1)}$  y aproximado de la siguiente manera:

$$\lambda_1 = \frac{|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}|}{|\mathbf{x}^{(k)} - \mathbf{x}^{(k+1)}|}$$

o bien:

$$\lambda_1 = \frac{\left( \sum_{i=1}^n (x_i^{k+1} - x_i^k)^2 \right)^{1/2}}{\left( \sum_{i=1}^n (x_i^k - x_i^{k-1})^2 \right)^{1/2}}$$

\* E. Kehat and M. Shacham. Chemical Processes Simulation Programs-3: Solution of systems of Non-Linear Equations. *Process Technology International*, Vol. 18, pág. 181 (1973).

Hay que observar que para la primera aplicación de este algoritmo se requieren tres vectores iniciales  $\mathbf{x}^{(0)}$ ,  $\mathbf{x}^{(1)}$  y  $\mathbf{x}^{(2)}$ , los cuales pueden obtenerse, por ejemplo, con el método de punto fijo multivariable.

Mediante este algoritmo resuelva el sistema:

$$f_1(x, y) = x^2 - 10x + y^2 + 8 = 0$$

$$f_2(x, y) = xy^2 + x - 10y + 8 = 0$$

usando como vector inicial:  $[x^0, y^0] = [0, 0]^T$  y los resultados de las dos primeras iteraciones del ejemplo 4.1.

- 4.29** La convergencia del método del eigenvalor dominante (véase Prob. 4.28), puede acelerarse usando un factor  $t$  de la siguiente manera:

$$\mathbf{x}^{(k+2)} = \mathbf{x}^{(k)} + \frac{t}{1 - \lambda_1} [\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}],$$

y ensayando varios valores de  $t$  como se hizo en los métodos de Newton-Raphson con optimización de  $t$  y del descenso de máxima pendiente. El valor de  $t$  puede calcularse también en cada iteración con una fórmula dada por Broyden,\* o sea usa un valor constante.

Obtenga una aproximación a una solución del sistema dado en el problema 4.28 utilizando un valor de  $t = 0.7$

- 4.30** Resuelva los sistema de ecuaciones no lineales de los problemas 4.6 y 4.9, empleando el método del descenso de máxima pendiente para obtener los valores iniciales; luego, con esos valores aplique el método de Newton-Raphson o el método de Broyden.
- 4.31** Hay que observar que en el método del descenso de máxima pendiente se encuentra el mínimo local de la función  $z_k = f_1^2 + f_2^2 + \dots + f_n^2$ . Este método puede emplearse para aproximar el mínimo local de una función dada analíticamente, tomando dicha función como  $z$ . Modifique el algoritmo 4.5 para aproximar los mínimos de las funciones siguientes, usando  $\text{EPS} = 10^{-5}$ .

a)  $z(x, y) = \sin(x + y) + \sin x - \cos y$

b)  $z(x_1, x_2, x_3) = x_1^2 + x_2^2 - 3x_3^2$

c)  $z(x_1, x_2, x_3) = x_1^2 + 2x_2^4 + 3x_3^3 - 1$

**SUGERENCIA:** Grafique la superficie el inciso a) usando el Mathcad o el Graphics Calculus (GC).

- 4.32** Encuentre todos los factores cuadráticos de las ecuaciones polinominales siguientes:

a)  $x^8 + 13x^6 + 35x^4 - 13x^2 - 36 = 0$

b)  $x^6 + (\pi^2 + 5)x^4 + (5\pi^2 + 6)x^2 + 6\pi^2 = 0$

- 4.33** ¿Qué pasa cuando se aplica el método de Bairstow a un polinomio que no tiene factores cuadráticos? Puede usar, por ejemplo, el polinomio

$$-2.8x^5 - 11.352x^4 + 86.468x^3 + 438.252x^2 + 32.418x - 1309.484.$$

\* C.G. Broyden. *A Class of Methods for Solving Nonlinear Simultaneous Equations*. Math Comp. 19 pág. 577 (1965).

## APROXIMACIÓN FUNCIONAL E INTERPOLACIÓN

### A dónde nos dirigimos

En este capítulo se estudiará la aproximación de funciones disponibles en forma discreta (puntos tabulados), con funciones analíticas sencillas, o bien de aproximación de funciones cuya complicada naturaleza exija su reemplazo por funciones más simples. Para lograr esto partiremos de tablas de valores dados y, utilizando la familia de los polinomios, aproximaremos una sección de la tabla por una línea recta, una parábola, etc. La elección del grado se hará analizando el fenómeno que originó los valores y, el tipo de aproximación, con base en la exactitud de éstos.

En la parte final del capítulo se estudia la aproximación utilizando el criterio de los mínimos cuadrados y se incluyen aproximaciones multilineales.

Las ideas y técnicas de interpolación-extrapolación permean el desarrollo de los métodos de los capítulos siguientes como integración, derivación, solución de ecuaciones diferenciales ordinarias y parciales e incluso se emplearon ya en la obtención de métodos para resolver ecuaciones no lineales.

### Introducción

La enorme ventaja de aproximar información discreta o funciones “complejas”, con funciones analíticas sencillas, radica en su mayor facilidad de evaluación y manipulación, situación necesaria en el campo de la ingeniería.

Las funciones de aproximación se obtienen por combinaciones lineales de elementos de familias de funciones denominadas elementales. En general tendrán la forma

$$a_0 g_0(x) + a_1 g_1(x) + \dots + a_n g_n(x), \quad (5.1)$$

donde  $a_i$ ,  $0 < i < n$ , son constantes por determinar y  $g_i(x)$ ,  $0 \leq i \leq n$  funciones de una familia particular. Los monomios en  $x$  ( $x^0, x, x^2, \dots$ ) constituyen la familia o grupo más empleado; sus combinaciones generan aproximaciones del tipo polinomial

$$a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n \quad (5.2)$$

El grupo conocido como funciones de Fourier

$$1, \sin x, \cos x, \sin 2x, \cos 2x, \dots,$$

al combinarse linealmente, genera aproximaciones del tipo:

$$a_0 + \sum_{i=1}^n a_i \cos ix + \sum_{i=1}^n b_i \sin ix \quad (5.3)$$

El grupo de las funciones exponenciales

$$1, e^x, e^{2x}, \dots$$



también puede usarse del modo siguiente

$$\sum_{i=0}^n a_i e^{ix} \quad (5.4)$$

De estos tres tipos de aproximaciones funcionales, las más comunes por su facilidad de manejo en evaluaciones, integraciones, derivaciones, etc., son las aproximaciones polinomiales (5.2) y son las que se estudiarán a continuación.

Sea una función  $f(x)$  dada en forma tabular

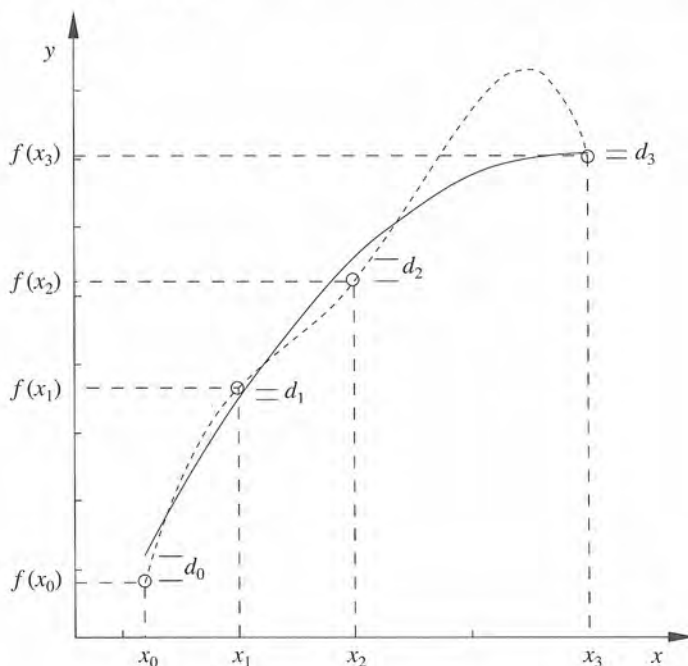
| Puntos | 0        | 1        | 2        | ... | $n$      |
|--------|----------|----------|----------|-----|----------|
| $x$    | $x_0$    | $x_1$    | $x_2$    | ... | $x_n$    |
| $f(x)$ | $f(x_0)$ | $f(x_1)$ | $f(x_2)$ | ... | $f(x_n)$ |

Para aproximar a  $f(x)$  por medio de un polinomio del tipo 5.2, se aplica alguno de los criterios siguientes: el de **ajuste exacto** o el de **mínimos cuadrados**.

La técnica del ajuste exacto consiste en encontrar una función polinomial **que pase por** los puntos dados en la tabla (véase Fig. 5.1). El método de mínimos cuadrados consiste en hallar un polinomio **que pase entre** los puntos y que satisfaga la condición de minimizar la suma de las desviaciones ( $d_i$ ) elevadas al cuadrado; es decir, que se cumpla

$$\sum_{i=0}^n (d_i)^2 = \text{mínimo}$$

Cuando la información tabular de que se dispone es aproximada hasta cierto número de cifras significativas, por ejemplo la de tablas de logaritmos o de funciones de Bessel, se recomienda usar ajuste exacto. En cambio, si la información tiene errores considerables, como en el caso de datos experimentales, no tiene sentido encontrar un polinomio que pase por esos puntos sino más bien que pase entre ellos; entonces, el método de mínimos cuadrados es aplicable.



**Figura 5.1**  
Aproximación  
polinomial con  
criterio de  
ajuste exacto  
(curva  
discontinua) y  
con mínimos  
cuadrados  
(curva llena).

Una vez que se obtiene el polinomio de aproximación, éste puede usarse para obtener puntos adicionales a los existentes en la tabla, mediante su evaluación, lo que se conoce como interpolación. También puede derivarse o integrarse a fin de obtener información adicional de la función tabular.

A continuación se describen distintas formas de aproximar con polinomios obtenidos por ajuste exacto y su uso en la interpolación. En la sección 5.8 se describe la aproximación polinomial por mínimos cuadrados, y en el capítulo 6 la derivación y la integración.

## 5.1 Aproximación polinomial simple e interpolación

La interpolación es de gran importancia en el campo de la ingeniería, ya que al consultar fuentes de información presentadas en forma tabular, es frecuente no encontrar el valor buscado como un punto en la tabla. Por ejemplo, las tablas 5.1 y 5.2 presentan la temperatura de ebullición de la acetona ( $C_3H_6O$ ) a diferentes presiones.

**Tabla 5.1** Temperatura de ebullición de la acetona a diferentes presiones.

| Puntos  | 0    | 1    | 2     | 3     | 4     | 5     | 6     |
|---------|------|------|-------|-------|-------|-------|-------|
| T (°C)  | 56.5 | 78.6 | 113.0 | 144.5 | 181.0 | 205.0 | 214.5 |
| P (atm) | 1    | 2    | 5     | 10    | 20    | 30    | 40    |

**Tabla 5.2** Temperatura de ebullición de la acetona a diferentes presiones.

| Puntos  | 0    | 1     | 2     | 3     |
|---------|------|-------|-------|-------|
| T (°C)  | 56.5 | 113.0 | 181.0 | 214.5 |
| P (atm) | 1    | 5     | 20    | 40    |

Supóngase que sólo se dispusiera de la segunda y se deseara calcular la temperatura de ebullición de la acetona a 2 atm de presión.

Una forma muy común de resolver este problema es sustituir los puntos (0) y (1) en la ecuación de la línea recta:  $p(x) = a_0 + a_1x$ , de tal modo que resultan dos ecuaciones con dos incógnitas que son  $a_0$  y  $a_1$ . Con la solución del sistema se consigue una aproximación polinomial de primer grado, lo que permite efectuar interpolaciones lineales; es decir, se sustituye el punto (0) en la ecuación de la línea recta y se obtiene:

$$56.5 = a_0 + 1 a_1$$

y al sustituir el punto (1)

$$113 = a_0 + 5 a_1,$$

sistema que al resolverse da  $a_0 = 42.375$  y  $a_1 = 14.125$

Por tanto, estos valores generan la ecuación:

$$p(x) = 42.375 + 14.125 x \quad (5.5)$$

La ecuación resultante puede emplearse para aproximar la temperatura cuando la presión es conocida. Al sustituir la presión  $x = 2$  atm se obtiene una temperatura de  $70.6^\circ\text{C}$ . A este proceso se le conoce como interpolación.

Gráficamente la tabla 5.2 puede verse como una serie de puntos (0), (1), (2) y (3) en un plano  $P$  vs  $T$  (Fig. 5.2), en donde si se unen con una línea los puntos (0) y (1), por búsqueda gráfica se obtiene  $T \approx 70.6^\circ\text{C}$ , para  $P = 2$  atm.

En realidad, esta interpolación sólo ha consistido en aproximar una función analítica desconocida [ $T = f(P)$ ] dada en forma tabular, por medio de una línea recta que pasa por los puntos (0) y (1).

Para aproximar el valor de la temperatura correspondiente a  $P = 2$  atm se pudieron tomar otros dos puntos distintos, por ejemplo (2) y (3), pero es de suponer que el resultado tendría un margen de error mayor, ya que el valor que se busca está entre los puntos (0) y (1).

Si se quisiera una aproximación mejor al valor “verdadero” de la temperatura buscada, podrían unirse más puntos de la tabla con una curva suave (sin picos), por ejemplo tres (0), (1), (2) (véase Fig. 5.3) y gráficamente obtener  $T$  correspondiente a  $P = 2$  atm.

Analíticamente, el problema se resuelve al aproximar la función desconocida [ $T = f(P)$ ] con un polinomio que pase por los tres puntos (0), (1) y (2). Este polinomio es una parábola y tiene la forma general

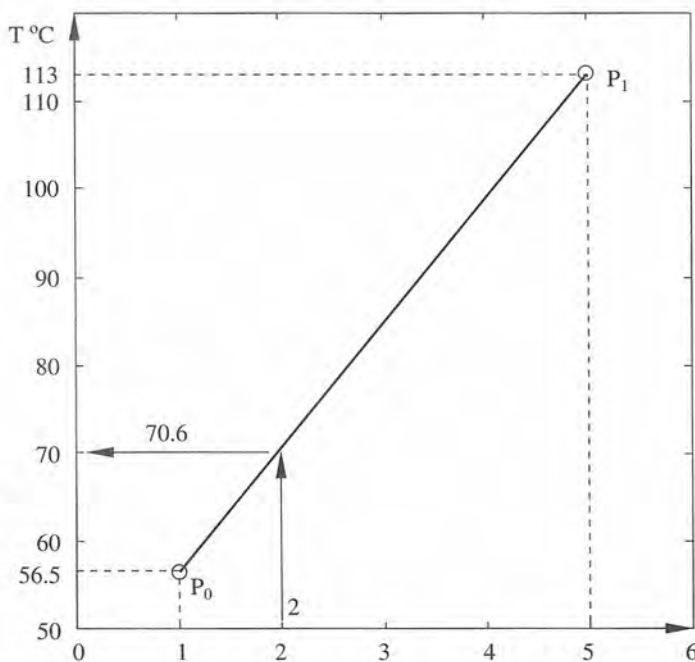
$$p_2(x) = a_0 + a_1x + a_2x^2, \quad (5.6)$$

donde los parámetros  $a_0$ ,  $a_1$  y  $a_2$  se determinan sustituyendo cada uno de los tres puntos conocidos en la ecuación 5.6; es decir:

$$\begin{aligned} 56.5 &= a_0 + a_1 \cdot 1 + a_2 \cdot 1^2 \\ 113 &= a_0 + a_1 \cdot 5 + a_2 \cdot 5^2 \\ 181 &= a_0 + a_1 \cdot 20 + a_2 \cdot 20^2 \end{aligned} \quad (5.7)$$

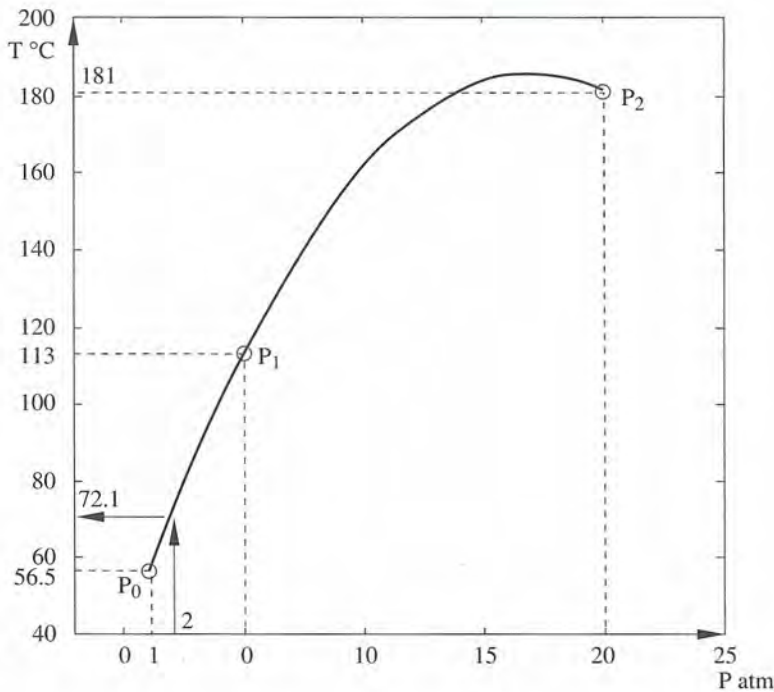
Al resolver el sistema se obtiene:

$$a_0 = 39.85, \quad a_1 = 17.15, \quad a_2 = -0.50482$$



**Figura 5.2**  
Interpolación  
gráfica de la  
temperatura de  
ebullición de la  
acetona a 2  
atm.





**Figura 5.3**  
Interpolación  
gráfica con tres  
puntos.

De tal modo que la ecuación polinomial queda:

$$p_2(x) = 39.85 + 17.15x - 0.50482x^2 \quad (5.8)$$

y puede emplearse para aproximar algún valor de la temperatura correspondiente a un valor de presión. Por ejemplo si  $x = 2 \text{ atm}$ , entonces:

$$T \approx p_2(2) = 39.85 + 17.15(2) - 0.50482(2)^2 \approx 72.1 \text{ } ^\circ\text{C}$$

La aproximación a la temperatura “correcta” es obviamente mejor en este caso. Obsérvese que ahora se ha aproximado la función desconocida  $[T = f(P)]$  con un polinomio de segundo grado (parábola) que pasa por los tres puntos más cercanos al valor buscado. En general, si se desea aproximar una función con un polinomio de grado  $n$ , se necesitan  $n + 1$  puntos, que sustituidos en la ecuación polinomial de grado  $n$ :

$$p_n(x) = a_0 + a_1x + a_2x^2 + \dots + a_nx^n \quad (5.9)$$

generan un sistema de  $n + 1$  ecuaciones lineales en las incógnitas  $a_i$ ,  $i = 0, 1, 2, \dots, n$ .

Una vez resuelto el sistema se sustituyen los valores de  $a_i$  en la ecuación (5.9), con lo cual se obtiene el polinomio de aproximación. A este método se le conoce como **aproximación polinomial simple**.

Por otro lado, como se dijo al principio de este capítulo, puede tenerse una función conocida pero muy complicada, por ejemplo:

$$f(x) = kx \ln x + \frac{1}{x} \sum_{m=0}^{\infty} C_m x^m, \quad (5.10)$$

$$f(x) = (2/x)^{1/2} \sin x \quad (5.11)$$

la cual conviene, para propósitos prácticos, aproximar con otra función más sencilla, como un polinomio. El procedimiento es generar una tabla de valores mediante la función original y a partir de dicha tabla aplicar el método descrito arriba.

**ALGORITMO 5.1** Aproximación polinomial simple

Para obtener los  $(n + 1)$  coeficientes del polinomio de grado  $n$  ( $n > 0$ ) que pasa por  $(n + 1)$  puntos, proporcionar los

DATOS: El grado del polinomio  $N$  y las  $N + 1$  parejas de valores  $(X(I), FX(I), I=0,1,\dots, N)$ .

RESULTADOS: Los coeficientes  $A(0), A(1), \dots, A(N)$  del polinomio de aproximación.

PASO 1. Hacer  $I = 0$ .

PASO 2. Mientras  $I \leq N$ , repetir los pasos 3 a 9.

PASO 3. Hacer  $B(I, 0) = 1$ .

PASO 4. Hacer  $J = 1$ .

PASO 5. Mientras  $J \leq N$ , repetir los pasos 6 y 7.

PASO 6. Hacer  $B(I, J) = B(I, J-1) * X(I)$ .

PASO 7. Hacer  $J = J + 1$ .

PASO 8. Hacer  $B(I, N+1) = FX(I)$ .

PASO 9. Hacer  $I = I + 1$ .

PASO 10. Resolver el sistema de ecuaciones lineales  $Ba = fx$  de orden  $N + 1$  con alguno de los algoritmos del capítulo 3.

PASO 11. IMPRIMIR  $A(0), A(1), \dots, A(N)$  y TERMINAR.

## 5.2 Polinomios de Lagrange

El método de aproximación polinomial dado en la sección anterior, requiere la solución de un sistema de ecuaciones algebraicas lineales que, cuando el grado del polinomio es alto, puede presentar inconvenientes. Existen otros métodos de aproximación polinomial en que no se requiere resolver un sistema de ecuaciones lineales y los cálculos se realizan directamente; entre éstos se encuentra el de aproximación polinomial de Lagrange.

Se parte nuevamente de una función desconocida  $f(x)$  dada en forma tabular y se asume que un polinomio de primer grado (ecuación de una línea recta) puede escribirse:

$$p(x) = a_0(x - x_1) + a_1(x - x_0) \quad (5.12)$$

donde  $x_1$  y  $x_0$  son los argumentos de los puntos conocidos  $[x_0, f(x_0)]$ ,  $[x_1, f(x_1)]$ , y  $a_0$  y  $a_1$  son dos coeficientes por determinar. Para encontrar el valor de  $a_0$ , se hace  $x = x_0$  en la ecuación 5.12, que al despejar da:

$$a_0 = \frac{p(x_0)}{x_0 - x_1} = \frac{f(x_0)}{x_0 - x_1} \quad (5.13)$$

y para hallar el valor de  $a_1$ , se sustituye el valor de  $x$  con el de  $x_1$ , con lo que resulta:

$$a_1 = \frac{p(x_1)}{x_1 - x_0} = \frac{f(x_1)}{x_1 - x_0} \quad (5.14)$$

de tal modo que al sustituir las ecuaciones 5.13 y 5.14 en la 5.12 queda:

$$p(x) = \frac{f(x_0)}{x_0 - x_1} (x - x_1) + \frac{f(x_1)}{x_1 - x_0} (x - x_0) \quad (5.15)$$

o en forma más compacta:

$$p(x) = L_0(x)f(x_0) + L_1(x)f(x_1) \quad (5.16)$$

donde:

$$L_0(x) = \frac{x - x_1}{x_0 - x_1} \quad y \quad L_1(x) = \frac{x - x_0}{x_1 - x_0} \quad (5.17)$$

De igual manera, un polinomio de segundo grado (ecuación de una parábola) puede escribirse:

$$p_2(x) = a_0(x - x_1)(x - x_2) + a_1(x - x_0)(x - x_2) + a_2(x - x_0)(x - x_1) \quad (5.18)$$

donde  $x_0$ ,  $x_1$  y  $x_2$  son los argumentos correspondientes a los tres puntos conocidos  $[x_0, f(x_0)]$ ,  $[x_1, f(x_1)]$ ,  $[x_2, f(x_2)]$ ; los valores de  $a_0$ ,  $a_1$  y  $a_2$  se encuentran sustituyendo  $x = x_0$ ,  $x = x_1$  y  $x = x_2$ , respectivamente, en la ecuación 5.18 para obtener:

$$a_0 = \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)}, \quad a_1 = \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)} \quad y \quad (5.19)$$

$$a_2 = \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)}$$

cuyo reemplazo en dicha ecuación genera el siguiente polinomio

$$p_2(x) = L_0(x)f(x_0) + L_1(x)f(x_1) + L_2(x)f(x_2) \quad (5.20)$$

donde:

$$L_0(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)}, \quad L_1(x) = \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} \quad (5.21)$$

$$L_2(x) = \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

Por inducción el lector puede obtener polinomios de tercer, cuarto o  $n$ -ésimo grado; este último queda como se indica a continuación

$$P_n(x) = L_0(x)f(x_0) + L_1(x)f(x_1) + \dots + L_n(x)f(x_n)$$

donde:

$$L_0(x) = \frac{(x - x_1)(x - x_2) \dots (x - x_n)}{(x_0 - x_1)(x_0 - x_2) \dots (x_0 - x_n)}$$

$$L_1(x) = \frac{(x - x_0)(x - x_2) \dots (x - x_n)}{(x_1 - x_0)(x_1 - x_2) \dots (x_1 - x_n)}$$

$$\vdots$$

$$L_n(x) = \frac{(x - x_0)(x - x_1) \dots (x - x_{n-1})}{(x_n - x_0)(x_n - x_1) \dots (x_n - x_{n-1})}$$

que en forma más compacta y útil para programarse en un lenguaje de computadora quedaría:

$$P_n(x) = \sum_{i=0}^n L_i(x)f(x_i) \quad (5.22)$$



donde\*:

$$L_i(x) = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{(x - x_j)}{(x_i - x_j)} \quad (5.23)$$

Al combinarse linealmente con  $f(x_i)$ , los polinomios  $L_i(x)$ , denominados polinomios de Lagrange, generan la aproximación polinomial de Lagrange a la información dada en forma tabular.

### Ejemplo 5.1

Para la tabla que se presenta a continuación

- Obtenga la aproximación polinomial de Lagrange con todos los puntos
- Interpole el valor de la función  $f(x)$  para  $x = 1.8$

|          |    |   |   |   |
|----------|----|---|---|---|
| $i$      | 0  | 1 | 2 | 3 |
| $f(x_i)$ | -3 | 0 | 5 | 7 |
| $x_i$    | 0  | 1 | 3 | 6 |

### Solución



- Obsérvese que hay cuatro puntos en la tabla, por lo que el polinomio será de tercer grado. Al sustituir los cuatro puntos en las ecuaciones generales 5.22 y 5.23 se obtiene:

$$\begin{aligned}
 p_3(x) = & (x-1)(x-3)(x-6) \frac{-3}{(0-1)(0-3)(0-6)} + \\
 & (x-0)(x-3)(x-6) \frac{0}{(1-0)(1-3)(1-6)} \\
 & + (x-0)(x-1)(x-6) \frac{5}{(3-0)(3-1)(3-6)} \\
 & + (x-0)(x-1)(x-3) \frac{7}{(6-0)(6-1)(6-3)}
 \end{aligned}$$

al efectuar las operaciones queda:

$$P_3(x) = (x^3 - 10x^2 + 27x - 18)(1/6) + (x^3 - 7x^2 + 6x)(-5/18) + (x^3 - 4x^2 + 3x)(7/90)$$

y finalmente resulta:

$$p_3(x) = -\frac{3}{90}x^3 - \frac{3}{90}x^2 + \frac{276}{90}x - 3$$

- El valor de  $x = 1.8$  se sustituye en la aproximación polinomial de Lagrange de tercer grado obtenida arriba y se tiene  $f(1.8) \approx 2$ .

Los cálculos pueden hacerse con Matlab o con la TI-92 Plus 1.

\*  $\prod_{i=1}^n (x - x_i) = (x - x_1)(x - x_2) \dots (x - x_n)$ .



```
x= [0 1 3 6];
y=[-3 0 5 7];
xi=1.8;
yi=interpl (x, y, xi)
```



```
e5_1()
Prgm
ClrIO
{0, 1, 3, 6}→a : {-3, 0, 5, 7}→y
4→n : 0→r : Delvar x
For i, 1, n
Y[i]→p
For j, 1, n
if i≠j
p*(x-a[j])/(a[i]-a[j])→p
EndFor
r+p→r
EndFor
Disp "Polinomio interpolante"
Disp expand(r) : Pause
FnOff : a[1]-.1*(a[n]-a[1])→xmin
a[n]+.1*(a[n]-a[1])→xmax
min(y)-.1*(max(y)-min(y)) →ymin
max(y)+.1*(max(y)-min(y)) →ymax
DrawFunc r : NewPlot 1, 1, a, y
FnOn : Pause
setMode("Split 1 App", "Home")
EndPrgm
```

Obsérvese que si se reemplaza  $x$  con cualquiera de los valores dados en la tabla, en la aproximación polinomial, se obtiene el valor de la función dado por la misma tabla.

### Ejemplo 5.2

Encuentre tanto la aproximación polinomial de Lagrange a la tabla 5.2 como el valor de la temperatura para una presión de 2 atm utilizando esta aproximación.

#### Solución

a) Aproximación polinomial de Lagrange mediante dos puntos ( $n = 1$ )

$$p(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) \quad (5.24)$$

al sustituir los primeros dos puntos de la tabla resulta:

$$p(x) = \frac{x - 5}{1 - 5} 56.5 + \frac{x - 1}{5 - 1} 113$$

Observe que la ecuación 5.24 es equivalente a la 5.5 y, por tanto, al sustituir  $x = 2$  se obtiene el mismo resultado  $T \approx 70.6^\circ\text{C}$ , como era de esperar.

b) Aproximación polinomial de Lagrange con tres puntos ( $n = 2$ )

$$p_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}f(x_0) + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}f(x_1) + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}f(x_2)$$

al sustituir los primeros tres puntos de la tabla, se obtiene:

$$p_2(x) = \frac{(x-5)(x-20)}{(1-5)(1-20)}56.5 + \frac{(x-1)(x-20)}{(5-1)(5-20)}113 + \frac{(x-1)(x-5)}{(20-1)(20-5)}181 \quad (5.25)$$

polinomio que puede servir para interpolar la temperatura de ebullición de la acetona a la presión de 2 atm; así el resultado queda  $T \approx 72.1$ . Observe que la ecuación 5.25 equivale a la 5.8.

c) La tabla 5.2 contiene cuatro puntos, por lo que la aproximación polinomial de mayor grado posible es 3. Se desarrolla la ecuación 5.22 para  $n = 3$

$$p_3(x) = \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}f(x_0) + \frac{(x-x_0)(x-x_2)(x-x_3)}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}f(x_1) + \frac{(x-x_0)(x-x_1)(x-x_3)}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}f(x_2) + \frac{(x-x_0)(x-x_1)(x-x_2)}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}f(x_3) \quad (5.26)$$

Al sustituir los puntos de la tabla, se obtiene:

$$p_3(x) = \frac{(x-5)(x-20)(x-40)}{(1-5)(1-20)(1-40)}56.5 + \frac{(x-1)(x-20)(x-40)}{(5-1)(5-20)(5-40)}113 + \frac{(x-1)(x-5)(x-40)}{(20-1)(20-5)(20-40)}181 + \frac{(x-1)(x-5)(x-20)}{(40-1)(40-5)(40-20)}214.5$$

y al simplificar queda:

$$p_3(x) = 0.01077x^3 - 0.78323x^2 + 18.4923x + 38.774$$

el cual puede emplearse para encontrar el valor de la temperatura correspondiente a la presión de 2 atm. Con la sustitución de  $x = 2$  y al evaluar  $p_3(x)$  queda:

$$T = f(2) \approx p_3(2) = 0.01077(2)^3 - 0.78323(2)^2 + 18.4923(2) + 38.774 = 72.7$$

Para realizar los cálculos puede usar Matlab o la TI92-Plus





```
P=[1 5 20 40];
T=[56.5 113 181 214.5];
xi=2;
yi=interpl (P, T, xi)
yi=interpl (P, T, xi, 'cubic')
yi=interpl (P, T, xi, 'spline')
```

Sobre 'spline' vea la sección 5.7



```
e5_2()
Prgm
ClrIO
Request "Grado del polinomio", n
expr(n) +1→n
For i, 1, n
Request "P("&string (i) &")", c
expr(c)→x[i]
Request "T("&string (i) &")", c
expr(c)→y[i]
EndFor
Request "Presion a interpolar", c
expr(c)→xint: 0→r
For i,1,n
y[i]→p
For j, 1, n
if i≠j
p*(xint-a[j])/(a[i]-a[j])→p
EndFor
r+p→r
EndFor
Disp "T("&format(xint,"f1")&")="&format(r,"f4")
EndPrgm
```

### ALGORITMO 5.2 Interpolación con polinomios de Lagrange

Para interpolar con polinomios de Lagrange de grado N, proporcionar los

**DATOS:** El grado del polinomio N, las N + 1 parejas de valores (X(I), FX(I), I=0,1,..., N) y el valor para el que se desea la interpolación XINT.

**RESULTADOS:** La aproximación FXINT, el valor de la función en XINT.

PASO 1. Hacer FXINT = 0.

PASO 2. Hacer I = 0.

PASO 3. Mientras I ≤ N, repetir los pasos 4 a 10.

PASO 4. Hacer L = 1.

PASO 5. Hacer  $J = 0$ .

PASO 6. Mientras  $J \leq N$ , repetir los pasos 7 y 8.

PASO 7. Si  $I \neq J$  Hacer  $L = L * (XINT - X(J)) / (X(I) - X(J))$ .

PASO 8. Hacer  $J = J + 1$ .

PASO 9. Hacer  $FXINT = FXINT + L * (FX(I))$ .

PASO 10. Hacer  $I = I + 1$ .

PASO 11. IMPRIMIR  $FXINT$  Y TERMINAR.

### Ejemplo 5.3

Elabore un programa para aproximar la función  $f(x) = \cos x$  en el intervalo  $[0, 8\pi]$ , con polinomios de Lagrange de grado 1, 2, 3, ..., 10. Use los puntos que se requieran, distribuidos regularmente en el intervalo.

Determine en forma práctica el error máximo que se comete al aproximar con los polinomios de los diferentes grados y compare los resultados.

### Solución



El programa se encuentra en el CD **(PROGRAMA 5.1)**. Para calcular el error máximo se dividió el intervalo  $[0, 8\pi]$  en 20 subintervalos y se calculó el valor con el polinomio interpolante y el valor verdadero con la función  $\cos x$ , determinando el error absoluto. Se obtuvieron los siguientes resultados.

| Grado | Error máximo |
|-------|--------------|
| 1     | 2.23627      |
| 2     | 2.23622      |
| 3     | 3.17025      |
| 4     | 2.23627      |
| 5     | 4.04277      |
| 6     | 4.1879       |
| 7     | 5.68560      |
| 8     | 33.74134     |
| 9     | 12.82475     |
| 10    | 35.95174     |

Se observa que al aumentar el grado del polinomio, el error absoluto máximo va aumentando.

Antes de pasar al estudio de otra forma de aproximación polinomial (de Newton), se requiere el conocimiento de las **diferencias divididas**, las cuales se presentan a continuación.

## 5.3 Diferencias divididas

Por definición de derivada en el punto  $x_0$  de una función analítica  $f(x)$  se tiene:

$$f'(x) \Big|_{x=x_0} = \lim_{x \rightarrow x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Sin embargo, cuando la función está en forma tabular

| Puntos | 0        | 1        | 2        | ... | $n$      |
|--------|----------|----------|----------|-----|----------|
| $x$    | $x_0$    | $x_1$    | $x_2$    | ... | $x_n$    |
| $f(x)$ | $f(x_0)$ | $f(x_1)$ | $f(x_2)$ | ... | $f(x_n)$ |

La derivada sólo puede obtenerse aproximadamente; por ejemplo, si se desea la derivada en el punto  $x$ , ( $x_0 < x < x_1$ ), puede estimarse como sigue:

$$f'(x) \approx \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \quad x_0 < x < x_1$$

El lado derecho de la expresión anterior se conoce como la primera\* diferencia dividida de  $f(x)$  respecto a los argumentos  $x_0$  y  $x_1$ , y se denota generalmente como  $f[x_0, x_1]$ ; así,

$$f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

La relación entre la primera diferencia dividida y la primera derivada queda establecida por el teorema del valor medio

$$\frac{f(x_1) - f(x_0)}{x_1 - x_0} = f'(\xi), \quad \xi \in (x_1, x_0)$$

siempre y cuando  $f(x)$  satisfaga las condiciones de aplicabilidad de dicho teorema. Para obtener aproximaciones de derivadas de orden más alto, se extiende el concepto de diferencias divididas a órdenes más altos como se ve en la tabla 5.3, en donde para uniformar la notación se han escrito los valores funcionales en los argumentos  $x_i$ ,  $0 \leq i \leq n$ , como  $f[x_i]$  y se les llama diferencias divididas de orden cero.

Por otro lado, de acuerdo con la tabla 5.3, la diferencia de orden  $i$  es:

$$f[x_0, x_1, x_2, \dots, x_i] = \frac{f[x_1, x_2, \dots, x_i] - f[x_0, x_1, \dots, x_{i-1}]}{x_1 - x_0}$$

En esta expresión puede observarse que:

- Para formarla se requieren  $i + 1$  puntos y
- El numerador es la resta de dos diferencias de orden  $i - 1$  y el denominador la resta de los argumentos no comunes en el numerador.

\* Se llama también diferencia dividida de primer orden.



| Información |          | Diferencias divididas                             |                                                                  |                                                                                 |
|-------------|----------|---------------------------------------------------|------------------------------------------------------------------|---------------------------------------------------------------------------------|
| $x$         | $f(x)$   | Primeras                                          | Segundas                                                         | Terceras                                                                        |
| $x_0$       | $f[x_0]$ |                                                   |                                                                  |                                                                                 |
| $x_1$       | $f[x_1]$ | $f[x_0, x_1] = \frac{f[x_1] - f[x_0]}{x_1 - x_0}$ | $f[x_0, x_1, x_2] = \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0}$ | $f[x_0, x_1, x_2, x_3] = \frac{f[x_1, x_2, x_3] - f[x_0, x_1, x_2]}{x_3 - x_0}$ |
| $x_2$       | $f[x_2]$ | $f[x_1, x_2] = \frac{f[x_2] - f[x_1]}{x_2 - x_1}$ | $f[x_1, x_2, x_3] = \frac{f[x_2, x_3] - f[x_1, x_2]}{x_3 - x_1}$ | $f[x_1, x_2, x_3, x_4] = \frac{f[x_2, x_3, x_4] - f[x_1, x_2, x_3]}{x_4 - x_1}$ |
| $x_3$       | $f[x_3]$ | $f[x_2, x_3] = \frac{f[x_3] - f[x_2]}{x_3 - x_2}$ | $f[x_2, x_3, x_4] = \frac{f[x_3, x_4] - f[x_2, x_3]}{x_4 - x_2}$ | $f[x_2, x_3, x_4, x_5] = \frac{f[x_3, x_4, x_5] - f[x_2, x_3, x_4]}{x_5 - x_2}$ |
| $x_4$       | $f[x_4]$ | $f[x_3, x_4] = \frac{f[x_4] - f[x_3]}{x_4 - x_3}$ | $f[x_3, x_4, x_5] = \frac{f[x_4, x_5] - f[x_3, x_4]}{x_5 - x_3}$ |                                                                                 |
| $x_5$       | $f[x_5]$ | $f[x_4, x_5] = \frac{f[x_5] - f[x_4]}{x_5 - x_4}$ |                                                                  |                                                                                 |

Tabla 5.3 Tabulación general de diferencias divididas

**Ejemplo 5.4**

La información de la tabla siguiente se obtuvo del polinomio

| Puntos | 0   | 1  | 2  | 3  | 4 | 5   |
|--------|-----|----|----|----|---|-----|
| $x$    | -2  | -1 | 0  | 2  | 3 | 6   |
| $f(x)$ | -18 | -5 | -2 | -2 | 7 | 142 |

A partir de ella, elabore una tabla de diferencias divididas.

**Solución** Las primeras diferencias divididas mediante los puntos (0), (1) y (1), (2), respectivamente, son:

$$f[x_0, x_1] = \frac{-5 - (-18)}{-1 - (-2)} = 13; \quad f[x_1, x_2] = \frac{-2 - (-5)}{0 - (-1)} = 3$$

La segunda diferencia dividida mediante los puntos (0), (1) y (2) es:

$$f[x_0, x_1, x_2] = \frac{3 - 13}{0 - (-2)} = -5$$

De igual manera se calculan las demás diferencias divididas, que se resumen en la siguiente tabla

| Puntos | $x$ | $f(x)$ | 1er orden | 2do orden | 3er orden | 4° orden |
|--------|-----|--------|-----------|-----------|-----------|----------|
| 0      | -2  | -18    |           |           |           |          |
|        |     |        | 13        |           |           |          |
| 1      | -1  | -5     |           | -5        |           |          |
|        |     |        | 3         |           | 1         |          |
| 2      | 0   | -2     |           | -1        |           | 0        |
|        |     |        | 0         |           | 1         |          |
| 3      | 2   | -2     |           | 3         |           | 0        |
|        |     |        | 9         |           | 1         |          |
| 4      | 3   | 7      |           | 9         |           |          |
|        |     |        | 45        |           |           |          |
| 5      | 6   | 142    |           |           |           |          |

Debe notarse que todas las diferencias divididas de tercer orden tienen el mismo valor, independientemente de los argumentos que se usen para su cálculo. Obsérvese también que las diferencias divididas de cuarto orden son todas cero, lo cual concuerda con que la tercera y cuarta derivada de un polinomio de tercer grado son —respectivamente— una constante y cero, sea cual sea el valor del argumento  $x$ . El razonamiento inverso también es válido: si al construir una tabla de diferencias divididas en alguna de las columnas el valor es constante (y en la siguiente columna es cero), la información proviene de un polinomio de grado igual al orden de las diferencias que tengan valores constantes.

Para realizar los cálculos puede usar Matlab.



```

x= [-2 -1 0 2 3 6];
fx= [-18 -5 -2 -2 7 142];
M=6;N=M-1;
for i=1: N
    T(i,1) = (fx(i+1) -fx(i))/(x(i+1)-x(i));
end
for j=2 :N
    for i=j :N
        T (i,j) = (T (i,j-1) -T(i-1,j -1))/(x(i+1)-
x(i-j+1));
    end
end
T

```

### ALGORITMO 5.3 Tabla de diferencias divididas

Para obtener la tabla de diferencias divididas de una función dada en forma tabular, proporcionar los

DATOS: El número de parejas M de la función tabular y las parejas de valores (X (I),FX(I), I= 0, 1, 2,... , M-1).

RESULTADOS: La tabla de diferencias divididas T.

PASO 1. Hacer N = M-1.

PASO 2. Hacer I = 0.

PASO 3. Mientras I ≤ N-1, repetir los pasos 4 y 5.

PASO 4. Hacer  $T(I,0) = (FX(I+1)-FX(I))/(X(I+1)-X(I))$ .

PASO 5. Hacer I = I+1.

PASO 6. Hacer J = 1.

PASO 7. Mientras J ≤ N-1, repetir los pasos 8 a 12.

PASO 8. Hacer I = J.

PASO 9. Mientras I ≤ N-1, repetir los pasos 10 y 11.

PASO 10. Hacer.

$$T(I,J) = (T(I,J-1) - T(I-1,J-1))/(X(I+1)-X(I-J)).$$

PASO 11. Hacer I = I + 1.

PASO 12. Hacer J = J + 1.

PASO 13. IMPRIMIR T y TERMINAR.

## 5.4 Aproximación polinomial de Newton

Supóngase que se tiene una función dada en forma tabular como se presenta a continuación

| Puntos | 0        | 1        | 2        | 3        | ... | n        |
|--------|----------|----------|----------|----------|-----|----------|
| x      | $x_0$    | $x_1$    | $x_2$    | $x_3$    | ... | $x_n$    |
| f(x)   | $f[x_0]$ | $f[x_1]$ | $f[x_2]$ | $f[x_3]$ | ... | $f[x_n]$ |



y que se desea aproximarla preliminarmente con un polinomio de primer grado que pasa, por ejemplo, por los puntos (0) y (1). Sea además dicho polinomio de la forma:

$$p(x) = a_0 + a_1(x - x_0), \quad (5.27)$$

donde  $x_0$  es la abscisa del punto (0) y  $a_0, a_1$  son constantes por determinar. Para encontrar el valor de  $a_0$  se hace  $x = x_0$ , de donde  $a_0 = p(x_0) = f[x_0]$ , y a fin de encontrar el valor de  $a_1$  se hace  $x = x_1$ , de donde  $a_1 = (f[x_1] - f[x_0]) / (x_1 - x_0)$ , o sea la primera diferencia dividida  $f[x_1, x_0]$ .

Al sustituir los valores de estas constantes en la ecuación 5.27 ésta queda

$$p(x) = f[x_0] + (x - x_0)f[x_0, x_1]$$

o sea un polinomio de primer grado en términos de diferencias divididas.

Y si ahora se desea aproximar la función tabular con un polinomio de segundo grado que pase por los puntos (0), (1) y (2) y que tenga la forma:

$$p_2(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) \quad (5.28)$$

donde  $x_0$  y  $x_1$  vuelven a ser las abscisas de los puntos (0) y (1) y  $a_0, a_1$  y  $a_2$  son constantes por determinar, se procede como en la forma anterior para encontrar estas constantes; o sea,

$$\text{si } x = x_0, a_0 = p_2(x_0) = f[x_0]$$

$$\text{si } x = x_1, a_1 = \frac{f[x_1] - f[x_0]}{x_1 - x_0} = f[x_0, x_1]$$

$$\text{si } x = x_2, a_2 = \frac{f[x_2] - f[x_0] - (x_2 - x_0) \frac{f[x_1] - f[x_0]}{x_1 - x_0}}{(x_2 - x_0)(x_2 - x_1)}$$

Al desarrollar algebraicamente el numerador y el denominador de  $a_2$  se llega a\*

$$a_2 = \frac{\frac{f[x_2] - f[x_1]}{x_2 - x_1} - \frac{f[x_1] - f[x_0]}{x_1 - x_0}}{x_2 - x_0} = f[x_0, x_1, x_2]$$

que es la segunda diferencia dividida respecto a  $x_0, x_1$  y  $x_2$ .

Con la sustitución de estos coeficientes en la ecuación 5.28 se obtiene:

$$p_2(x) = f[x_0] + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2]$$

que es un polinomio de segundo grado en términos de diferencias divididas.

Por inducción se puede establecer que, en general, para un polinomio de grado  $n$  escrito en la forma

$$p_n(x) = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + \dots + a_n(x - x_0)(x - x_1) \dots (x - x_{n-1}) \quad (5.29)$$

y que pasa por los puntos (0), (1), (2), ..., (n); los coeficientes  $a_0, a_1, \dots, a_n$  están dados por

$$\begin{aligned} a_0 &= f[x_0] \\ a_1 &= f[x_0, x_1] \\ a_2 &= f[x_0, x_1, x_2] \\ &\vdots \\ a_n &= f[x_0, x_1, x_2, \dots, x_n] \end{aligned}$$

\* Véase el problema 5.11.

Esta aproximación polinomial de Newton, la cual se puede expresar sintéticamente como:

$$p_n(x) = \sum_{k=0}^n a_k \prod_{i=0}^{k-1} (x - x_i) \quad (5.30)$$

Para realizar los cálculos puede usar Matlab.

### Ejemplo 5.5

Elabore una aproximación polinomial de Newton para la información tabular de las presiones de vapor de la acetona (tabla 5.2) e interpole la temperatura para una presión de 2 atm.

### Solución

Para el cálculo de los coeficientes del polinomio de Newton, se construye la tabla de diferencias divididas.

| Puntos | P  | T     | Diferencias divididas |          |         |
|--------|----|-------|-----------------------|----------|---------|
|        |    |       | Primera               | Segunda  | Tercera |
| 0      | 1  | 56.5  |                       |          |         |
|        |    |       | 14.125                |          |         |
| 1      | 5  | 113   |                       | -0.50482 |         |
|        |    |       | 4.533                 |          | 0.01085 |
| 2      | 20 | 181   |                       | -0.08167 |         |
|        |    |       | 1.675                 |          |         |
| 3      | 40 | 214.5 |                       |          |         |

a) Para  $n = 1$

$$p(x) = a_0 + a_1(x - x_0) = f[x_0] + f[x_0, x_1](x - x_0)$$

de la tabla se tiene  $f[x_0] = 56.5$  y  $f[x_0, x_1] = 14.125$ , de donde:

$$p(x) = 56.5 + 14.125(x - 1)$$

ecuación que equivale a las obtenidas anteriormente (5.5 y 5.24).

$$\text{Si } x = 2, f(2) \approx p(2) = 56.5 + 14.125(2 - 1) = 70.6 \text{ } ^\circ\text{C}$$

b) Para  $n = 2$

$$\begin{aligned} p_2(x) &= a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) \\ &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) \end{aligned}$$

de la tabla se obtienen  $a_0 = f[x_0] = 56.5$ ,  $a_1 = f[x_0, x_1] = 14.125$ ,  $a_2 = f[x_0, x_1, x_2] = -0.50482$ , que al sustituirse en la ecuación de arriba dan:

$$p_2(x) = 56.5 + 14.125(x - 1) - 0.50482(x - 1)(x - 5)$$

ecuación que equivale a 5.8 y 5.25

$$\text{Si } x = 2, f(2) \approx p_2(2) = 56.5 + 14.125(2 - 1) - 0.50482(2 - 1)(2 - 5) = 72.1 \text{ } ^\circ\text{C}$$

c) Para  $n = 3$

$$\begin{aligned}
 p_3(x) &= a_0 + a_1(x-x_0) + a_2(x-x_0)(x-x_1) + a_3(x-x_0)(x-x_1)(x-x_2) \\
 &= f[x_0] + f[x_0, x_1](x-x_0) + f[x_0, x_1, x_2](x-x_0)(x-x_1) + \\
 &\quad f[x_0, x_1, x_2, x_3](x-x_0)(x-x_1)(x-x_2)
 \end{aligned}$$

de la tabla se obtienen  $a_0 = f[x_0] = 56.5$ ,  $a_1 = f[x_0, x_1] = 14.125$ ,

$$a_2 = f[x_0, x_1, x_2] = -0.50842, \quad a_3 = f[x_0, x_1, x_2, x_3] = 0.01085.$$

que sustituidas generan el polinomio de aproximación

$$p_3(x) = 56.5 + 14.125(x-1) - 0.50482(x-1)(x-5) + 0.01085(x-1)(x-5)(x-20)$$

y que es esencialmente el polinomio obtenido con el método de Lagrange (ecuación 5.26).

$$\begin{aligned}
 \text{Si } x = 2, f(2) \approx p_3(2) &= 56.5 + 14.125(2-1) - 0.50482(2-1)(2-5) + \\
 &\quad 0.01085(2-1)(2-5)(2-20) = 70.7^\circ\text{C}
 \end{aligned}$$

Para realizar los cálculos puede usar Matlab.



```

clear
x=[1 5 20 40];
fx=[56.5 113 181 214.5];
M=4;N=M-1;
for i=1 : N
    T(i,1)=(fx(i+1)-fx(i))/(x(i+1)-x(i));
end
for j=2:N
    for i=j:N
        T(i,j)=(T(i,j-1)-T(i-1,j-1))/...
            (x(i+1)-x(i-j+1));
    end
end
T
Xint=2;
fprintf(' N Fxint \n')
px1=fx(1)+T(1,1)*(Xint-x(1));
fprintf(' %d %6.1f \n',1,px1)
px2=fx(1)+T(1,1)*(Xint-x(1))+...
    T(2,2)*(Xint-x(1))*(Xint-x(2));
fprintf(' %d %6.1f \n',2,px2)
px3=fx(1)+T(1,1)*(Xint-x(1))+...
    T(2,2)*(Xint-x(1))*(Xint-x(2))+...
    T(3,3)*(Xint-x(1))*(Xint-x(2))*(Xint-x(3));
fprintf(' %d %6.1f \n',3,px3)

```

### Ejemplo 5.6

Aproxime la temperatura de ebullición de la acetona a una presión de 30 atm usando aproximación polinomial de Newton de grado dos (véase Ej. 5.5). 36.5

### Solución

Se hace pasar el polinomio de Newton por los puntos (1), (2) y (3), con lo que toma la forma

$$p_2(x) = a_0 + a_1(x-x_1) + a_2(x-x_1)(x-x_2),$$



con los coeficientes dados ahora de la siguiente manera:

$$\begin{aligned}a_0 &= f[x_1] \\a_1 &= f[x_1, x_2] \\a_2 &= f[x_1, x_2, x_3]\end{aligned}$$

Al sustituir:

$$\begin{aligned}p_2(x) &= f[x_1] + f[x_1, x_2](x-x_1) + f[x_1, x_2, x_3](x-x_1)(x-x_2) \\&= 113 + 4.533(x-x_1) - 0.08167(x-x_1)(x-x_2)\end{aligned}$$

y al evaluar dicho polinomio en  $x = 30$ , se obtiene la aproximación buscada

$$\begin{aligned}T &= p_2(30) = 113 + 4.533(30 - 5) - 0.08167(30 - 5)(30 - 20) \\&= 205.9\end{aligned}$$

El valor reportado en la tabla 5.1 es 205, por lo que la aproximación es buena. Para realizar los cálculos puede usar Matlab.



```
clear
x= [5 20 40 ];
fx=[113 181 214.5];
M=3; N=M-1;
for i=1 : N
    T(i,1) = (fx(i+1) -fx(i))/(x(i+1)-x(i))
end
for j=2:N
    for i=j:N
        T(i,j)=(T(i,j-1)-T(i-1,j-1))/...
            (x(i+1)-x(i-j+1));
    end
end
T
Xint=30;
px2=fx(1) +T(1,1) * (Xint-x(1))+ ...
    T(2,2) * (Xint-x(1))*(Xint-x(2));
fprintf('T(%2d)=%.1f\n',Xint, px2)
```

#### ALGORITMO 5.4 Interpolación polinomial de Newton

Para interpolar con polinomios de Newton en diferencias divididas de grado  $N$ , proporcionar los

**DATOS:** El grado del polinomio  $N$ , las  $N+1$  parejas de valores  $(X(I), FX(I), I=0, 1, 2, \dots, N)$  y el valor para el que se desea interpolar  $XINT$ .

**RESULTADOS:** La aproximación  $FXINT$  al valor de la función en  $XINT$ .

PASO 1. Realizar los pasos 2 a 12 del algoritmo 5.3.

PASO 2. Hacer  $FXINT = FX(0)$ .

PASO 3. Hacer  $I = 0$ .

PASO 4. Mientras  $I \leq N-1$ , repetir los pasos 5 a 11.

- PASO 5. Hacer  $P = 1$ .  
 PASO 6. Hacer  $J = 0$ .  
 PASO 7. Mientras  $J \leq I$ , repetir los pasos 8 y 9.  
     PASO 8. Hacer  $P = P * (XINT - X(J))$ .  
     PASO 9. Hacer  $J = J + 1$ .  
 PASO 10. Hacer  $FXINT = FXINT + T(I,I)*P$ .  
 PASO 11. Hacer  $I = I + 1$ .

PASO 12. IMPRIMIR  $FXINT$  y TERMINAR.

## 5.5 Polinomio de Newton en diferencias finitas

Cuando la distancia  $h$  entre dos argumentos consecutivos cualesquiera, es la misma a lo largo de la tabla, el polinomio de Newton en diferencias divididas puede expresarse con más sencillez. Para este propósito se introduce un nuevo parámetro  $s$ , definido en  $x = x_0 + sh$ , con el cual se expresa el factor productoria

$$\prod_{i=0}^{k-1} (x - x_i),$$

de la ecuación 5.30 en términos de  $s$  y  $h$ . Para esto obsérvese que  $x_1 - x_0 = h$ ,  $x_2 - x_0 = 2h, \dots$ ,  $x_i - x_0 = ih$  y que restando  $x_i$  ( $0 \leq i \leq n$ ) en ambos miembros de  $x = x_0 + sh$ , se obtiene:

$$x - x_i = x_0 - x_i + sh = -ih + sh = h(s - i) \quad \text{para } (0 \leq i \leq n)$$

Por ejemplo si  $i = 1$ ,

$$x - x_1 = h(s - 1)$$

si  $i = 2$ ,

$$x - x_2 = h(s - 2)$$

Al sustituir cada una de las diferencias  $(x - x_i)$  con  $h(s - i)$ , en la ecuación 5.29, se llega a:

$$\begin{aligned} p_n(x) = p_n(x_0 + sh) &= f[x_0] + hs f[x_0, x_1] + h^2 s(s-1) f[x_0, x_1, x_2] \\ &+ h^3 s(s-1)(s-2) f[x_0, x_1, x_2, x_3] + \dots \\ &+ h^n s(s-1)(s-2) \dots (s-(n-1)) f[x_0, x_1, \dots, x_n] \end{aligned} \quad (5.31)$$

o en forma compacta

$$p_n(x) = \sum_{k=0}^n a_k h^k \prod_{i=0}^{k-1} (s - i) \quad (5.32)$$

Esta última ecuación puede simplificarse aún más si se introduce el operador lineal  $\Delta$ , conocido como **operador lineal en diferencias hacia delante** y definido sobre  $f(x)$  como.

$$\Delta f(x) = f(x + h) - f(x)$$

La segunda diferencia hacia delante puede obtenerse como sigue:

$$\begin{aligned} \Delta(\Delta f(x)) &= \Delta^2 f(x) = \Delta(f(x + h) - f(x)) \\ &= \Delta f(x + h) - \Delta f(x) \\ &= f(x + h + h) - f(x + h) - f(x + h) + f(x) \\ &= f(x + 2h) - 2f(x + h) + f(x) \end{aligned}$$

A su vez, las diferencias hacia delante de orden superior se generan como sigue:

$$\Delta^i f(x) = \Delta (\Delta^{i-1} f(x))$$

Estas diferencias se conocen como diferencias finitas hacia delante. Análogamente cabe definir  $\nabla$  como **operador lineal de diferencias hacia atrás**; así, la primera diferencia hacia atrás se expresa como:

$$\nabla f(x) = f(x) - f(x-h)$$

La segunda diferencia hacia atrás queda:

$$\nabla^2 f(x) = \nabla(\nabla f(x)) = \nabla(f(x) - f(x-h))$$

$$\nabla^2 f(x) = f(x) - f(x-h) - f(x-h) + f(x-2h)$$

$$\nabla^2 f(x) = f(x) - 2f(x-h) + f(x-2h)$$

de tal modo que las diferencias hacia atrás de orden superior se expresan en términos generales como:

$$\nabla^i f(x) = \nabla(\nabla^{i-1} f(x)).$$

Estas diferencias se conocen como diferencias finitas hacia atrás.

Al aplicar  $\Delta$  al primer valor funcional  $f[x_0]$  de una tabla se tiene:

$$\Delta f(x_0) = f[x_1] - f[x_0] = hf[x_0, x_1],$$

de manera que:

$$f[x_0, x_1] = \frac{1}{h} \Delta f(x_0)$$

Del mismo modo:

$$f[x_0, x_1, x_2] = \frac{\frac{f[x_2] - f[x_1]}{x_2 - x_1} - \frac{f[x_1] - f[x_0]}{x_1 - x_0}}{x_2 - x_0} = \frac{f[x_2] - 2f[x_1] + f[x_0]}{2h^2}$$

por lo que:

$$f[x_0, x_1, x_2] = \frac{1}{2h^2} \Delta^2 f(x_0)$$

En general:

$$f[x_0, x_1, \dots, x_n] = \frac{1}{n! h^n} \Delta^n f(x_0) \quad (5.33)$$

De igual manera, las diferencias divididas en función de las diferencias hacia atrás quedan:

$$f[x_n, x_{n-1}, \dots, x_0] = \frac{1}{n! h^n} \nabla^n f(x_n) \quad (5.34)$$

Consecuentemente, al sustituir  $f[x_0, x_1, \dots, x_i]$ ,  $(0 \leq i \leq n)$  en términos de diferencias finitas, la ecuación 5.31 queda:

$$p_n(x) = p_n(x_0 + sh) = f[x_0] + s\Delta f[x_0] + \frac{s(s-1)}{2!} \Delta^2 f[x_0] +$$



$$\begin{aligned}
 & + \frac{s(s-1)(s-2)}{3!} \Delta^3 f[x_0] + \dots \\
 & + \frac{s(s-1)(s-2) \dots (s-(n-1))}{n!} \Delta^n f[x_0]
 \end{aligned} \tag{5.35}$$

conocido como el **polinomio de Newton en diferencias finitas hacia delante**.

Existe una expresión equivalente a la 5.35 para diferencias hacia atrás (**polinomio de Newton en diferencias finitas hacia atrás**), cuya obtención se motiva al final del ejemplo siguiente.

### Ejemplo 5.7

La siguiente tabla proporciona las presiones de vapor en lb/plg<sup>2</sup> a diferentes temperaturas para el 1-3 butadieno

| Puntos                | 0     | 1     | 2     | 3     | 4     | 5     |
|-----------------------|-------|-------|-------|-------|-------|-------|
| T °F                  | 50    | 60    | 70    | 80    | 90    | 100   |
| P lb/plg <sup>2</sup> | 24.94 | 30.11 | 36.05 | 42.84 | 50.57 | 59.30 |

Aproxime la función tabulada por el polinomio de Newton en diferencias hacia delante e interpole la presión a la temperatura de 64 °F.

**Solución** Primero se construye la tabla de diferencias hacia delante como sigue:

| Punto | $x_i$ | $f[x_i]$ | $\Delta f[x_i]$        | $\Delta^2 f[x_i]$        | $\Delta^3 f[x_i]$        | $\Delta^4 f[x_i]$         |
|-------|-------|----------|------------------------|--------------------------|--------------------------|---------------------------|
| 0     | 50    | 24.94    |                        |                          |                          |                           |
|       |       |          | $\Delta f[x_0] = 5.17$ |                          |                          |                           |
| 1     | 60    | 30.11    |                        | $\Delta^2 f[x_0] = 0.77$ |                          |                           |
|       |       |          | $\Delta f[x_1] = 5.94$ |                          | $\Delta^3 f[x_0] = 0.08$ |                           |
| 2     | 70    | 36.05    |                        | $\Delta^2 f[x_1] = 0.85$ |                          | $\Delta^4 f[x_0] = 0.01$  |
|       |       |          | $\Delta f[x_2] = 6.79$ |                          | $\Delta^3 f[x_1] = 0.09$ |                           |
| 3     | 80    | 42.84    |                        | $\Delta^2 f[x_2] = 0.94$ |                          | $\Delta^4 f[x_1] = -0.03$ |
|       |       |          | $\Delta f[x_3] = 7.73$ |                          | $\Delta^3 f[x_2] = 0.06$ |                           |
| 4     | 90    | 50.57    |                        | $\Delta^2 f[x_3] = 1.00$ |                          |                           |
|       |       |          | $\Delta f[x_4] = 8.73$ |                          |                          |                           |
| 5     | 100   | 59.30    |                        |                          |                          |                           |

Observe que en esta información  $h=10$ , el valor por interpolar es 64 y que el valor de  $s$  se obtiene de la expresión  $x = x_0 + sh$ ; esto es:

$$s = \frac{x - x_0}{h} = \frac{64 - 50}{10} = 1.4$$

Si se deseara aproximar con un polinomio de primer grado, se tomarían sólo los dos primeros términos de la ecuación 5.35; o sea,

$$p(x) = f[x_0] + s \Delta f[x_0] = 24.94 + 1.4(5.17) = 32.18$$

Hay que observar que realmente se está extrapolando, ya que el valor de  $x$  queda fuera del intervalo de los puntos que se usaron para formar el polinomio de aproximación.

Intuitivamente se piensa que se obtendría una aproximación mejor con los puntos (1) y (2). Sin embargo, la ecuación 5.35 se desarrolló usando  $x_0$  como pivote y para aplicarla con el punto (1) y (2) debe modificarse a la forma siguiente

$$p_n(x) = f[x_1 + sh] = f[x_1] + s\Delta f[x_1] + \frac{s(s-1)}{2!} \Delta^2 f[x_1] + \dots + \frac{s(s-1)\dots(s-(n-1))}{n!} \Delta^n f[x_1] \quad (5.36)$$

la cual usa como pivote  $x_1$ , y cuyos primeros dos términos dan la aproximación polinomial de primer grado:

$$p(x) = f[x_1] + s\Delta f[x_1], \text{ donde ahora } s = \frac{x - x_1}{h} = \frac{64 - 60}{10} = 0.4;$$

al sustituir valores de la tabla se tiene:

$$f(64) \approx p(64) = 30.11 + 0.4(5.94) = 32.49$$

En cambio, si se deseara aproximar con un polinomio de segundo grado, se requerirían tres puntos y sería aconsejable tomar (0), (1) y (2) en lugar de (1), (2) y (3), ya que el argumento por interpolar está más al centro de los primeros. Con esta selección y la ecuación 5.35 queda

$$p_2(x) = f[x_0] + s\Delta f[x_0] + \frac{s(s-1)}{2!} \Delta^2 f[x_0]$$

donde:

$$s = \frac{x - x_0}{h} = \frac{64 - 50}{10} = 1.4;$$

este valor se sustituye arriba y queda:

$$p_2(64) = 24.94 + 1.4(5.17) + \frac{1.4(1.4-1)}{2!} 0.77 = 32.39$$

Si se quisiera interpolar el valor de la presión a una temperatura de 98 °F, tendría que desarrollarse una ecuación de Newton en diferencias hacia delante, usando como pivote el punto (4) para un polinomio de primer grado o el punto (3) para un polinomio de segundo grado, etc. Sin embargo, esto es factible usando un solo pivote (el punto 5 en este caso), independientemente del grado del polinomio por usar, si se emplean diferencias hacia atrás.

Para esto se debe desarrollar una ecuación equivalente a la 5.35, pero en diferencias hacia atrás; este desarrollo se presenta a continuación en dos pasos —el primero es un resultado necesario.

### Primer paso

Obtención del polinomio de Newton en diferencias divididas hacia atrás de grado  $n$  apoyado en el punto  $x_n$ .

Para simplificar se inicia con  $n = 2$  y se asume que un polinomio de segundo grado en general tiene la forma:

$$p_2(x) = a_0 + a_1(x - x_n) + a_2(x - x_n)(x - x_{n-1})$$

donde  $a_0$ ,  $a_1$  y  $a_2$  son las constantes por determinar y  $x_n$  y  $x_{n-1}$  las abscisas de los puntos  $(n)$  y  $(n-1)$ , respectivamente.

$$\text{Si } x = x_n, a_0 = p_2(x_n) = f[x_n]$$

$$\text{Si } x = x_{n-1}, a_1 = \frac{p_2(x_{n-1}) - p_2(x_n)}{x_{n-1} - x_n} = f[x_n, x_{n-1}]$$

$$\text{Si } x = x_{n-2}, a_2 = \frac{p_2(x_{n-2}) - p_2(x_n) - f[x_n, x_{n-1}](x_{n-2} - x_n)}{(x_{n-2} - x_n)(x_{n-2} - x_{n-1})}$$

al desarrollar algebraicamente el numerador y el denominador de  $a_2$  se llega a:

$$a_2 = f[x_n, x_{n-1}, x_{n-2}]$$

al sustituir estas constantes en el polinomio queda:

$$p_2(x) = f[x_n] + (x - x_n)f[x_n, x_{n-1}] + (x - x_n)(x - x_{n-1})f[x_n, x_{n-1}, x_{n-2}]$$

De lo anterior se puede inducir que, en general, para un polinomio de grado  $n$  escrito en la forma

$$p_n(x) = a_0 + a_1(x - x_n) + a_2(x - x_n)(x - x_{n-1}) + \dots + a_n(x - x_n)(x - x_{n-1})\dots(x - x_1), \quad (5.37)$$

los coeficientes  $a_0, a_1, a_2, \dots, a_n$  están dados por

$$\begin{aligned} a_0 &= f[x_n] \\ a_1 &= f[x_n, x_{n-1}] \\ &\vdots \\ a_n &= f[x_n, x_{n-1}, x_{n-2}, \dots, x_0] \end{aligned}$$

### Segundo paso

Obtención del polinomio de Newton en diferencias finitas hacia atrás de grado  $n$ , apoyado en el punto  $x_n$ .

Las ecuaciones\* siguientes se pueden construir introduciendo el parámetro  $s$  definido ahora por la expresión  $x = x_n + sh$ .

$$\begin{aligned} x - x_n &= sh \\ x - x_{n-1} &= x_n - x_{n-1} + sh = h(s+1) \\ x - x_{n-2} &= x_n - x_{n-2} + sh = h(s+2) \\ &\vdots \\ x - x_0 &= x_n - x_0 + sh = h(s+n) \end{aligned}$$

Al sustituir las ecuaciones anteriores y los coeficientes  $f[x_n], f[x_n, x_{n-1}], \dots, f[x_n, x_{n-1}, \dots, x_0]$  en la ecuación 5.37 en términos de diferencias finitas (Ec. 5.34), finalmente queda

\* Recuérdese que se considera aquí que la diferencia entre dos argumentos consecutivos cualesquiera es  $h$ .



$$p_n(x_n + sh) = f[x_n] + s \nabla f[x_n] + \frac{s(s+1)}{2!} \nabla^2 f[x_n] + \dots$$

$$+ \frac{s(s+1) \dots (s+(n-1))}{n!} \nabla^n f[x_n] \quad (5.38)$$

que es la ecuación de Newton en diferencias hacia atrás,

Para realizar los cálculos puede usar Matlab o la TI-92 Plus



```
x=[50 60 70 80 90 100];
fx=[24.94 30.11 36.05 42.84 50.57 59.30];
N=6; h=10; xint=64;
for i=1:N-1
    T(i,1)=fx(i+1)-fx(i);
end
for j=2:N-1
    for i=j:N-1
        T(i,j)=T(i,j-1)-T(i-1,j-1);
    end
end
T
s=(xint-x(1))/h;
fxint=fx(1)+s*T(1,1);
fprintf('Grado 1 P(%4.0f)=%6.2f\n',xint,fxint)
fxint=fx(1)+s*T(1,1)+s*(s-1)/2*T(2,2);
fprintf('Grado 2 P(%4.0f)=%6.2f\n',xint,fxint)
```



```
e5_7()
Prgm
{50, 60, 70, 80, 90, 100}→x : ClrIO
{24.94, 30.11, 36.05, 42.84, 50.57, 59.30}→y
6→n : 10→h : 64→xint: newMat(n-1,n-1)→t
For i,1,n-1
    y[i+1]-y[i]→t[i,1]
EndFor
for j,2,n-1
    for i,j,n-1
        t[i,j-1]-t[i-1,j-1]→ t[i,j]
    EndFor
EndFor
setMode("Display Digits","FIX 2")
disp t:Pause
(xint-x[1])/h→s
y[1]+s*t[1,1] →fxint
"P("&format(xint,"f0")&")="→d
d&format(fxint,"f3")&" con grado 1"→d
disp d
y[1]+s*t[1,1]+s*(s-1)/2*t[2,2]→fxint
"P("&format(xint,"f0")&")="→d
d&format(fxint,"f3")&" con grado 2"→d
disp d
EndPrgm
```

**Ejemplo 5.8**

Interpolar el valor de la presión a una temperatura de 98 °F, utilizando la tabla de presiones de vapor del ejemplo 5.7 y el polinomio de Newton (5.38).

**Solución** Primero se construye la tabla de diferencias hacia atrás como sigue:

| Punto | $x_i$ | $f[x_i]$ | $\nabla f[x_i]$        | $\nabla^2 f[x_i]$        | $\nabla^3 f[x_i]$        | $\nabla^4 f[x_i]$         |
|-------|-------|----------|------------------------|--------------------------|--------------------------|---------------------------|
| 0     | 50    | 24.94    |                        |                          |                          |                           |
|       |       |          | $\nabla f[x_1] = 5.17$ |                          |                          |                           |
| 1     | 60    | 30.11    |                        | $\nabla^2 f[x_2] = 0.77$ |                          |                           |
|       |       |          | $\nabla f[x_2] = 5.94$ |                          | $\nabla^3 f[x_3] = 0.08$ |                           |
| 2     | 70    | 36.05    |                        | $\nabla^2 f[x_3] = 0.85$ |                          | $\nabla^4 f[x_4] = 0.01$  |
|       |       |          | $\nabla f[x_3] = 6.79$ |                          | $\nabla^3 f[x_4] = 0.09$ |                           |
| 3     | 80    | 42.84    |                        | $\nabla^2 f[x_4] = 0.94$ |                          | $\nabla^4 f[x_5] = -0.03$ |
|       |       |          | $\nabla f[x_4] = 7.73$ |                          | $\nabla^3 f[x_5] = 0.06$ |                           |
| 4     | 90    | 50.57    |                        | $\nabla^2 f[x_5] = 1.00$ |                          |                           |
|       |       |          | $\nabla f[x_5] = 8.73$ |                          |                          |                           |
| 5     | 100   | 59.30    |                        |                          |                          |                           |

Si se usa un polinomio de primer grado, se tiene de la ecuación 5.38.

$$p(98) = f[x_5] + s \nabla f[x_5]$$

donde:

$$s = \frac{x - x_n}{h} = \frac{98 - 100}{10} = -0.2;$$

y con la tabla de diferencias finitas hacia atrás

$$p_2(98) = 59.3 - 0.2(8.73) = 57.55$$

Si en cambio se usa un polinomio de segundo grado, se emplean los tres primeros términos de la ecuación 5.38, con lo cual la aproximación queda:

$$\begin{aligned} p_2(98) &= f[x_5] + s \nabla f[x_5] + \frac{s(s+1)}{2!} \nabla^2 f[x_5] \\ &= 59.3 - 0.2(8.73) + \frac{-0.2(-0.2+1)}{2!} (1) = 57.67 \end{aligned}$$

Para realizar los cálculos puede usar Matlab o la TI-92 Plus.



```

x= [50 60 70 80 90 100];
fx=[24.94 30.11 36.05 42.84 50.57 59.30];
N=6; h=10;
for i=1:N-1
    T(i,1)=fx(i+1)-fx(i);
end
for j=2:N-1
    for i=j:N-1
        T(i,j)=T(i,j-1)-T(i-1,j-1);
    end
end
T
Xint=98;
s=(Xint-x(N))/h;
px=fx(6)+s*T(5,1);
fprintf(' T(%6.2f)= %6.2f \n',Xint,px)
px=fx(6)+s*T(5,1)+s*(s-1)/2*T(4,2);
fprintf(' T(%6.2f)= %6.2f \n',Xint,px)

```



```

e5_8()
Prgm
{50, 60, 70, 80, 90, 100}→x
{24.94, 30.11, 36.05, 42.84, 50.57, 59.30}→y
6→n: 10→h: newMat(n-1,n-1)→t
For i,1,n-1
    y[i+1]-y[i]→t[i,1]
EndFor
For j,2,n-1
    For i,j,n-1
        t[i,j-1]-t[i-1,j-1]→t[i,j]
    EndFor
EndFor
Disp t
98→xint : (xint-x[n])/h→s
y[n]+s*t[n-1,1]→yint
Disp "y("&format(xint,"f0")&")="&format(yint,"f2")
y[n]+s*t[n-1,1]+s*(s-1)/2*t[n-1,2]→yint
Disp "y("&format(xint,"f0")&")="&format(yint,"f2")

```

Si se deseara interpolar el valor de la presión a una temperatura de 82 °F, tendría que usar-se la ecuación 5.38 pero apoyada en el punto  $n-1$  [punto (4) en este caso] ; esto es,

$$\begin{aligned}
 p_n(x_{n-1} + sh) = & f[x_{n-1}] + s \nabla f[x_{n-1}] + \frac{s(s+1)}{2!} \nabla^2 f[x_{n-1}] + \dots \\
 & + \frac{s(s+1) \dots (s+(n-1))}{n!} \nabla^n f[x_{n-1}]
 \end{aligned}
 \tag{5.39}$$



**NOTA:** Es importante hacer notar que las tablas de los ejemplos 5.7 (diferencias hacia delante) y 5.8 (diferencias hacia atrás) presentan los mismos valores numéricos aunque los operadores y subíndices de sus argumentos no sean los mismos. Por lo anterior, el polinomio de Newton en diferencias hacia delante y su tabla correspondiente pueden usarse a fin de interpolar en puntos del final de la tabla con sólo invertir la numeración de los puntos en dicha tabla y los subíndices de los argumentos de cada columna de diferencias finitas (se ilustra enseguida en el Ejemplo 5.9).

También es útil observar que los valores de la tabla utilizados en las ecuaciones 5.35, 5.36 o alguna modificación de éstas, son los de las diagonales trazadas de arriba hacia abajo (véase la tabla del ejemplo 5.7) y que los valores utilizados en 5.38, 5.39 o alguna modificación de éstas, son los de las diagonales trazadas de abajo hacia arriba (ver tabla del ejemplo 5.8).

Se resuelve un ejemplo para ilustrar esto.

### Ejemplo 5.9

Con la ecuación 5.35 y la tabla de diferencias hacia delante del ejemplo 5.7, interpole la presión de vapor de 1-3 butadieno a la temperatura de 98 °F, mediante un polinomio de primer y segundo grado.

### Solución

Invertidos la numeración de los puntos en la tabla mencionada y los subíndices de los argumentos de cada columna, la tabla toma el aspecto:

| Punto | $x_i$ | $f[x_i]$ | $\nabla f[x_i]$        | $\nabla^2 f[x_i]$        | $\nabla^3 f[x_i]$        | $\nabla^4 f[x_i]$         |
|-------|-------|----------|------------------------|--------------------------|--------------------------|---------------------------|
| 5     | 50    | 24.94    |                        |                          |                          |                           |
|       |       |          | $\nabla f[x_4] = 5.17$ |                          |                          |                           |
| 4     | 60    | 30.11    |                        | $\nabla^2 f[x_3] = 0.77$ |                          |                           |
|       |       |          | $\nabla f[x_3] = 5.94$ |                          | $\nabla^3 f[x_2] = 0.08$ |                           |
| 3     | 70    | 36.05    |                        | $\nabla^2 f[x_2] = 0.85$ |                          | $\nabla^4 f[x_1] = 0.01$  |
|       |       |          | $\nabla f[x_2] = 6.79$ |                          | $\nabla^3 f[x_1] = 0.09$ |                           |
| 2     | 80    | 42.84    |                        | $\nabla^2 f[x_1] = 0.94$ |                          | $\nabla^4 f[x_0] = -0.03$ |
|       |       |          | $\nabla f[x_1] = 7.73$ |                          | $\nabla^3 f[x_0] = 0.06$ |                           |
| 1     | 90    | 50.57    |                        | $\nabla^2 f[x_0] = 1.00$ |                          |                           |
|       |       |          | $\nabla f[x_0] = 8.73$ |                          |                          |                           |
| 0     | 100   | 59.30    |                        |                          |                          |                           |

Observe que todos los valores numéricos conservan su posición en la tabla.

Se emplea la ecuación 5.35 con  $x = 98$ ,  $x_0 = 100$  y  $h = 10$ , de donde:

$$s = \frac{x - x_0}{h} = \frac{98 - 100}{10} = -0.2$$

al emplear un polinomio de primer grado se tiene:

$$p(98) = 59.30 + (-0.2)(8.73) = 57.55$$

En cambio, con uno de segundo grado:

$$p_2(98) = 59.30 + (-0.2)(8.73) + \frac{(-0.2)(-0.2 + 1)}{2!} 1 = 57.63$$

Como se puede observar, son los mismos resultados que se obtuvieron en el ejemplo 5.8.



En el CD encontrará el **PROGRAMA 5.8** de Interpolación Numérica. Con este programa usted puede proporcionar la función como una tabla de puntos e interpolar para algún valor deseado. Podrá también observar gráficamente los puntos dados, la función interpolante y el valor a interpolar

## 5.6 Estimación de errores en la aproximación

En general, al aproximar una función por un polinomio de grado  $n$ , se comete un error; por ejemplo, cuando se utiliza un polinomio de primer grado, se reemplaza la función verdadera en un intervalo con una línea recta (Fig. 5.4). En términos matemáticos, la función se podría representar exactamente como:

$$f(x) = f[x_0] + (x-x_0)f[x_0, x_1] + R_1(x) = p_1(x) + R_1(x) \quad (5.40)$$

Donde  $R_1(x)$  es el error cometido al aproximar linealmente la función  $f(x)$  y  $p_1(x)$  es, por ejemplo, el polinomio de primer grado en diferencias divididas.

Al despejar  $R_1(x)$  de la ecuación 5.40 y tomando como factor común  $(x-x_0)$  queda:

$$\begin{aligned} R_1(x) &= f(x) - f[x_0] - (x-x_0)f[x_0, x_1] \\ &= (x-x_0) \left( \frac{f(x) - f[x_0]}{x-x_0} - f[x_0, x_1] \right) \\ &= (x-x_0)(f[x_0, x] - f[x_0, x_1]) \end{aligned}$$

al multiplicar y dividir por  $(x-x_1)$  se obtiene:

$$R_1(x) = (x-x_0)(x-x_1)f[x, x_0, x_1]$$

donde  $f[x, x_0, x_1]$  es la segunda diferencia dividida respecto a los argumentos  $x_0$ ,  $x_1$  y  $x$ . Resulta imposible calcular exactamente  $f[x, x_0, x_1]$ , ya que no se conoce la  $f(x)$  necesaria para su evaluación. Sin embargo, si se tiene otro valor de  $f(x)$ , sea  $f(x_2)$  (y si la segunda diferencia  $f[x, x_0, x_1]$  no varía significativamente en el intervalo donde están los puntos  $x_0$ ,  $x_1$  y  $x_2$ ), entonces  $R_1(x)$  se aproxima de la siguiente manera:

$$R_1(x) \approx (x-x_0)(x-x_1)f[x_0, x_1, x_2]$$

de tal modo que al sustituirlo en la ecuación original quede:

$$f(x) \approx f[x_0] + (x-x_0)f[x_0, x_1] + (x-x_0)(x-x_1)f[x_0, x_1, x_2]$$

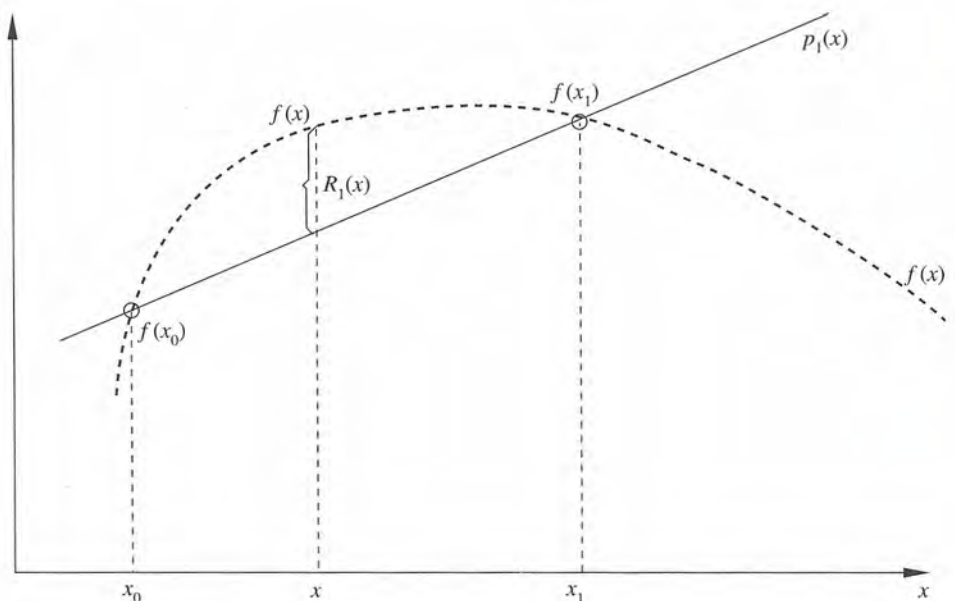


Figura 5.4

Observe que el lado derecho de esta expresión es el polinomio de segundo grado en diferencias divididas. Como se había intuido, esto confirma que —en general— se aproxima mejor la función  $f(x)$  con un polinomio de grado dos que con uno de primer grado.

Por otro lado, si se aproxima a función  $f(x)$  con un polinomio de segundo grado  $p_2(x)$ , se espera que el error  $R_2(x)$  sea en general menor. La función expresada en estos términos queda:

$$f(x) = p_2(x) + R_2(x) = f[x_0] + (x-x_0)f[x_0, x_1] + (x-x_0)(x-x_1)f[x_0, x_1, x_2] + R_2(x)$$

de donde  $R_2(x)$  puede despejarse:

$$R_2(x) = f(x) - f[x_0] - (x-x_0)f[x_0, x_1] - (x-x_0)(x-x_1)f[x_0, x_1, x_2]$$

y como en el caso de un polinomio de primer grado, se demuestra\* que el término del error para la aproximación polinomial de segundo grado es:

$$R_2(x) = (x-x_0)(x-x_1)(x-x_2)f[x_0, x_1, x_2]$$

De igual modo que  $f[x, x_0, x_1]$  en el caso lineal  $f[x, x_0, x_1, x_2]$  no se puede determinar con exactitud; sin embargo, si se tiene un punto adicional  $(x_3, f(x_3))$ , cabe aproximar  $f[x, x_0, x_1, x_2]$  con:

$$f[x, x_0, x_1, x_2] \approx f[x_0, x_1, x_2, x_3],$$

que sustituida proporciona una aproximación a  $R_2(x)$ :

$$R_2(x) \approx (x-x_0)(x-x_1)(x-x_2)f[x_0, x_1, x_2, x_3].$$

Si se continúa este proceso puede establecerse por inducción que:

$$f(x) = p_n(x) + R_n(x),$$

donde  $p_n(x)$  es el polinomio de grado  $n$  en diferencias divididas que aproxima la función tabulada, y  $R_n(x)$  es el término correspondiente del error. Esto es:

$$P_n(x) = f[x_0] + (x-x_0)f[x_0, x_1] + \dots + (x-x_0)(x-x_1)\dots(x-x_{n-1})f[x_0, \dots, x_n]$$

y

$$R_n(x) = (x-x_0)(x-x_1)\dots(x-x_n)f[x, x_0, x_1, \dots, x_n]$$

o

$$R_n(x) = \left[ \prod_{i=0}^n (x-x_i) \right] f[x, x_0, x_1, \dots, x_n] \quad (5.41)$$

en donde  $f[x, x_0, x_1, \dots, x_n]$  puede aproximarse con un punto adicional  $(x_{n+1}, f(x_{n+1}))$  así:

$$f[x, x_0, x_1, \dots, x_n] \approx f[x_0, x_1, x_2, \dots, x_n, x_{n+1}] \quad (5.42)$$

entonces  $R_n(x)$  queda como:

$$R_n(x) \approx \left[ \prod_{i=0}^n (x-x_i) \right] f[x_0, x_1, x_2, \dots, x_n, x_{n+1}]$$

La ecuación

$$f(x) = p_n(x) + R_n(x)$$

es conocida como la **fórmula fundamental de Newton en diferencias divididas**. Al analizar el factor productoria (producto acumulado)

$$\prod_{i=0}^n (x-x_i)$$

\* Véase el problema 5.24.



de  $R_n(x)$ , se observa que para disminuirlo (y, por ende, disminuir el error  $R_n(x)$ ) deben usarse argumentos  $x_i$  lo más cercanos posible al valor por interpolar  $x$  (regla que se había seguido por intuición y que ahora se confirma matemáticamente). También de esta productoria se infiere que en general en una extrapolación ( $x$  fuera del intervalo de las  $x_i$  usadas) el error es mayor que en una interpolación. Puede decirse también que si bien se espera una mejor aproximación al aumentar el grado  $n$  del polinomio  $p_n(x)$ , es cierto que el valor del factor productoria aumenta al incrementarse  $n$ , por lo que debe existir un grado óptimo para el polinomio que se usará en el proceso de interpolación. Por último, en términos generales es imposible determinar el valor exacto de  $R_n(x)$ ; a lo más que se puede llegar es determinar el intervalo en que reside el error.

Los ejemplos que se dan a continuación ilustran estos comentarios.

### Ejemplo 5.10

Suponga que tiene la tabla siguiente de la función  $\cos x$ .

| Puntos          | 0      | 1      | 2      | 3      |
|-----------------|--------|--------|--------|--------|
| $x$ (grados)    | 0      | 50     | 60     | 90     |
| $f(x) = \cos x$ | 1.0000 | 0.6400 | 0.5000 | 0.0000 |

y desea interpolar el valor de la función en  $x = 10^\circ$ .

**Solución** Al interpolar linealmente con los puntos (0) y (1) queda:

$$p(x) = f[x_0] + (x-x_0)f[x_0, x_1]$$

Al sustituir valores da  $p(10) = 0.9280$ .

La interpolación con un polinomio de segundo grado y los puntos (0), (1) y (2) da:

$$p_2(x) = f[x_0] + (x-x_0)f[x_0, x_1] + (x-x_0)(x-x_1)f[x_0, x_1, x_2]$$

Al sustituir valores resulta  $p_2(10) = 0.9845$ .

Se interpola con un polinomio de tercer grado (usando los cuatro puntos) y queda

$$p_3(x) = f[x_0] + (x-x_0)f[x_0, x_1] + (x-x_0)(x-x_1)f[x_0, x_1, x_2] + (x-x_0)(x-x_1)(x-x_2)f[x_0, x_1, x_2, x_3]$$

Al sustituir valores da  $p_3(10) = 0.9764$ .

El valor correcto de  $\cos 10^\circ$  hasta la cuarta cifra significativa es 0.9848, así que el error en por ciento para el primer grado es 5.77, para el segundo 0.03, y para el tercero 0.85

El grado óptimo del polinomio de aproximación para este caso particular es 2 (usando los puntos más cercanos al valor por interpolar: (0) (1) y (2)). Si se usaran los puntos (0), (1) y (3) el error sería 1.80%, como puede verificar el lector.

### Ejemplo 5.11

Con la ecuación 5.41 encuentre una cota inferior del error de interpolación  $R_n(x)$  para  $x = 1.5$  cuando  $f(x) = \ln x$ ,  $n=3$ ,  $x_0=1$ ,  $x_1=4/3$ ,  $x_2=5/3$  y  $x_3 = 2$ .

**Solución** La ecuación 5.41 con  $n = 3$  queda:

$$R_3(x) = f[x, x_0, x_1, x_2, x_3] \prod_{i=0}^3 (x - x_i),$$

donde el factor productoria puede evaluarse directamente como sigue

$$\prod_{i=0}^3 (x - x_i) = (1.5-1) (1.5-4/3) (1.5-5/3) (1.5-2) = 0.00694$$

En cambio, el factor  $f[x, x_0, x_1, x_2, x_3]$  es —como se ha dicho antes— imposible de determinar, pues no se cuenta con el valor de  $f(x)$  (necesario para su evaluación). Sin embargo, el valor de  $f[x, x_0, x_1, x_2, x_3]$  está estrechamente relacionado con la cuarta derivada de  $f(x)$ , como lo expresa el siguiente teorema.

### Teorema \*

Sea  $f(x)$  una función de valor real, definida en  $[a, b]$  y  $k$  veces diferenciable en  $(a, b)$ . Si  $x_0, x_1, \dots, x_k$  son  $k+1$  puntos distintos en  $[a, b]$ , entonces existe  $\xi \in (a, b)$  tal que:

$$f[x_0, x_1, \dots, x_k] = \frac{f^{(k)}(\xi)}{k!}$$

$$\text{con } \xi \in (\min x_i, \max x_i), 0 \leq i \leq n$$

Al utilizar esta información se tiene, en general,

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^n (x - x_i) \quad \text{con } \xi \in (\min x_i, \max x_i), 0 \leq i \leq n$$

y para  $n = 3$

$$f[x, x_0, x_1, x_2, x_3] = \frac{f^{IV}(\xi)}{4!}$$

Se deriva sucesivamente  $f(x)$  cuatro veces y se tiene:

$$f'(x) = 1/x; f''(x) = -1/x^2; f'''(x) = 2/x^3; f^{IV}(x) = -6/x^4$$

Como  $f^{IV}(x)$  es creciente en el intervalo de interés  $(1, 2)$  (al aumentar  $x$  en éste se incrementa  $f^{IV}(x)$ ), alcanza su valor mínimo en  $x = 1$  y, por tanto, la cota inferior buscada está dada por:

$$0.00694 \frac{f^{IV}(1)}{4!} = 0.00694 \frac{-6}{(1)^4 4!} = -0.00174,$$

es decir:

$$R_3(1.5) \geq -0.00174$$

Este valor indica que el error de interpolación cuando  $x = 1.5$  es mayor o igual que  $-0.00174$ . Sin embargo, para conocer el intervalo donde reside el error, es necesario conocer la cota superior, que se calcula en el ejemplo siguiente.

### Ejemplo 5.12

Calcule la cota superior del error  $R_3(x)$  del ejemplo anterior y confirme que al utilizar diferencias divididas para interpolar en  $x = 1.5$ , el error obtenido está en el intervalo cuyos extremos son las cotas obtenidas. Use 0.40547 como valor verdadero en  $\ln 1.5$ .

### Solución

Como se vio, la función  $-6/x^4$  es creciente en  $(1, 2)$ ; por tanto alcanza su valor máximo en  $x = 2$  y la cota superior está dada por:

\* Para su demostración véase Conte, S.D. y De Boor C. *Análisis numérico*. 2ª. Ed., MC Graw-Hill (1967), pp. 226-227.

$$0.00694 \frac{-6}{2^4 4!} = -0.00011,$$

es decir:

$$R_3(1.5) \leq -0.00011$$

Por medio de la interpolación con diferencias divididas con un polinomio de tercer grado se obtiene:

$$\begin{aligned} p_3(1.5) &= f[x_0] + (1.5-x_0)f[x_0, x_1] + (1.5-x_0)(1.5-x_1)f[x_0, x_1, x_2] \\ &\quad + (1.5-x_0)(1.5-x_1)(1.5-x_2)f[x_0, x_1, x_2, x_3] \\ &= 0.40583 \end{aligned}$$

y el error es  $ln\ 1.5 - p_3(1.5) = -0.00036$  que, efectivamente, está en el intervalo  $[-0.00174, -0.00011]$ .

## 5.7 Aproximación polinomial segmentaria

En alguno de los casos previos pudo pensarse en aproximar  $f(x)$  por medio de un polinomio de grado "alto", 10 o 20. Esto pudiera ser por diversas razones: porque se quiere mayor exactitud; por manejar un solo polinomio que sirva para interpolar en cualquier punto del intervalo  $[a, b]$ , etcétera.

Sin embargo, hay serias objeciones al empleo de la aproximación de grado "alto"; la primera es que los cálculos para obtener  $p_n(x)$  son mayores, hay que verificar más cálculos para evaluar  $p_n(x)$  y, lo peor del caso, es que los resultados son poco confiables como puede verse en el ejemplo 5.10.

Si bien lo anterior es grave, lo es más que el error de interpolación aumenta en lugar de disminuir (véase Sec. 5.6 y ejemplo 5.3). Para abundar un poco más en la discusión de la sección 5.6, se retomará el factor productoria de la ecuación 5.41.

$$\prod_{i=0}^n (x - x_i),$$

donde, si  $n$  es muy grande, los factores  $(x - x_i)$ , son numerosos y, si su magnitud es mayor de 1, evidentemente su influencia será aumentar el error  $R_n(x)$ .

Para disminuir  $R_n(x)$ , atendiendo el factor productoria exclusivamente, es menester que los factores  $(x - x_i)$  sean en su mayoría menores de 1 en magnitud, lo cual puede lograrse tomando intervalos pequeños alrededor de  $x$ . Como el intervalo sobre el cual se va a aproximar  $f(x)$  generalmente se da de antemano, lo anterior se logra dividiendo dicho intervalo en subintervalos suficientemente pequeños y aproximar  $f(x)$  en cada subintervalo por medio de un polinomio adecuado; por ejemplo, mediante una línea recta en cada subintervalo (véase Fig. 5.5)

Esto da como aproximación de  $f(x)$  una línea quebrada o segmentos de líneas rectas —que se llamarán  $g_1(x)$ — cuyos puntos de quiebre son  $x_1, x_2, \dots, x_{n-1}$ . Las funciones  $f(x)$  y  $g_1(x)$  coinciden en  $x_0, x_1, x_2, \dots, x_n$  y el error en cualquier punto  $x$  de  $[x_0, x_n]$  queda acotado, de acuerdo con el teorema del ejemplo 5.11 aplicado a cada subintervalo  $[x_i, x_{i+1}]$  con  $i = 0, 1, 2, \dots, n-1$ , por

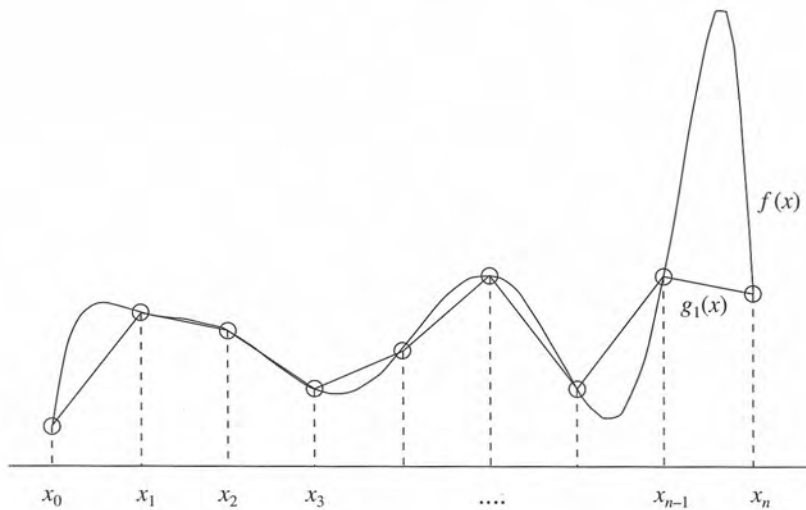
$$R_1(x) = |f(x) - g_1(x)| \leq \max_{a \leq \xi \leq b} \left| \frac{f''(\xi)}{2!} \right| \max_i |(x - x_i)(x - x_{i+1})| \quad (5.43)$$



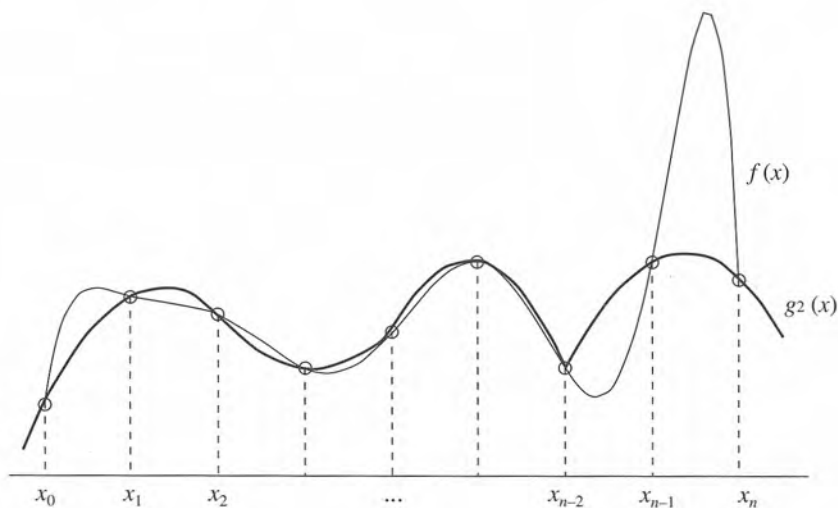
Si  $f(x)$  fuera diferenciable dos veces en  $[x_0, x_n]$ , el valor máximo de  $|(x-x_i)(x-x_{i+1})|$  para  $x \in [x_i, x_{i+1}]$  se dan en  $x = (x_i + x_{i+1})/2$ , el punto medio de  $[x_i, x_{i+1}]$ ; de modo que

$$\begin{aligned} \max_i |(x-x_i)(x-x_{i+1})| &= \max_i \left| \left( \frac{x_i + x_{i+1}}{2} - x_i \right) \left( \frac{x_i + x_{i+1}}{2} - x_{i+1} \right) \right| \\ &= \max_i \left| \frac{x_{i+1} - x_i}{2} \frac{x_i - x_{i+1}}{2} \right| \\ &= \max_i \frac{(x_{i+1} - x_i)^2}{4} = \max_i \frac{\Delta x_i^2}{4} \end{aligned}$$

**Figura 5.5**  
Aproximación  
de  $f(x)$  por una  
línea quebrada.



**Figura 5.6**  
Aproximación  
de  $f(x)$  por  
parábolas.



Al sustituir en la ecuación 5.43

$$R_1(x) = \left| f(x) - g_1(x) \right| \leq \max_{a \leq \xi \leq b} \left| \frac{f''(\xi)}{2!} \right| \max_i \frac{\Delta x_i^2}{4} \quad (5.44)$$

Donde se aprecia que el error  $R_1(x)$  puede reducirse tanto como se quiera, haciendo  $\Delta x_i$  pequeño para toda  $i$ ; por ejemplo, tomando un número suficientemente grande de subintervalos en  $[a, b]$ , o bien empleando polinomios de grado dos (véase Fig. 5.6) para cada subintervalo  $[x_i, x_{i+1}]$ ; de esta última manera se consiguen segmentos polinomiales de grado dos  $g_2(x)$  cuyo término del error (5.44) correspondiente tendrá  $\Delta x_i^3$  en lugar de  $\Delta x_i^2$ . Esto da una disminución del error respecto al empleo de líneas rectas. El empleo de polinomios de grado 3 en cada subintervalo  $[x_i, x_{i+1}]$  es de las técnicas más difundidas y se discute en detalle enseguida.

### APROXIMACIÓN CÚBICA SEGMENTARIA DE HERMITE

Se parte del hecho que se tiene una función  $f(x)$  de valor real, dada en forma tabular o analítica en el intervalo  $[a, b]$ , con

$$a = x_0 < x_1 < x_2 < \dots < x_n = b \quad (5.45)$$

Se quiere construir una función  $g_3(x)$  con segmentos de polinomios cúbicos\*  $p_i(x)$  en cada  $[x_i, x_{i+1}]$  con  $i = 0, 1, 2, \dots, n-1$ , tal que

$$g_3(x_i) = f(x_i) \text{ con } i = 0, 1, 2, \dots, n,$$

de donde:

$$p_i(x_i) = f(x_i) \text{ y } p_i(x_{i+1}) = f(x_{i+1}) \text{ para } i = 0, 1, \dots, n-1 \quad (5.46)$$

y esta última implica que:

$$p_{i-1}(x_i) = p_i(x_i) \quad i = 1, 2, \dots, n$$

de modo que  $g_3(x)$  es continua en  $[a, b]$  y tiene los puntos interiores  $x_1, x_2, \dots, x_{n-1}$  como puntos de quiebre o donde  $g_3(x)$  no es diferenciable en general.

De acuerdo con el álgebra, se sabe que para que un polinomio cúbico quede determinado en forma única se requieren cuatro puntos. Hasta ahora, cada uno de los segmentos cúbicos  $p_i(x)$  tiene que pasar por  $(x_i, f(x_i))$  y  $(x_{i+1}, f(x_{i+1}))$ , de modo que quedan dos puntos o condiciones que se pueden establecer para definir en forma única  $p_i(x)$ .

La elección de estas dos condiciones faltantes depende, por ejemplo, de la utilización que se vaya a dar a  $g_3(x)$ , de  $f(x)$  y del contexto donde se trabaje (de la ingeniería o matemático).

Por ejemplo, desde el punto de vista de la ingeniería, sería deseable que  $g_3(x)$  fuera diferenciable en los puntos interiores:  $x_1, x_2, \dots, x_{n-1}$ ; es decir, que  $g_3(x)$  fuese suave en  $[a, b]$ , en lugar de tener picos o puntos de quiebre. Esto se daría con dos condiciones como la 5.46, pero en derivadas; así

$$p'_i(x_i) = f'(x_i) \text{ y } p'_i(x_{i+1}) = f'(x_{i+1}) \quad i = 0, 1, \dots, n-1 \quad (5.47)$$

previsto que  $f'(x)$  fuese conocida o aproximada en cada uno de los puntos  $x_0, x_1, \dots, x_n$ . Con esto quedan cubiertas las dos condiciones faltantes.

\* En lo que sigue de esta sección, el subíndice indica el subintervalo, no el grado del polinomio como en otras ocasiones.

De la ecuación 5.47 se infiere

$$p'_{i-1}(x_i) = p'_i(x_i) \quad i = 1, 2, \dots, n \quad (5.48)$$

En este punto cabe empezar a hablar del cálculo de los polinomios  $p_i(x)$ ; por tanto, como paso siguiente se aproxima  $p_i(x)$ ,  $i = 1, 2, \dots, n$  con diferencias divididas así.

$$\begin{aligned} p_i(x) = f(x_i) &+ f[x_p, x_i](x-x_i) + f[x_p, x_i, x_{i+1}](x-x_i)^2 \\ &+ f[x_p, x_p, x_{i+1}, x_{i+1}](x-x_i)^2(x-x_{i+1}) \end{aligned} \quad (5.49)$$

como:

$$f[x_p, x_i] = \lim_{\Delta x \rightarrow 0} \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} = f'(x_i)$$

y al sustituir  $(x-x_{i+1})$  con  $(x-x_i) + (x_i-x_{i+1})$  y agrupar se tiene

$$\begin{aligned} p_i(x) = f(x_i) &+ f'(x_i)(x-x_i) \\ &+ (f[x_p, x_p, x_{i+1}] - f[x_p, x_p, x_{i+1}, x_{i+1}]\Delta x_i)(x-x_i)^2 + f[x_p, x_p, x_{i+1}, x_{i+1}](x-x_i)^3 \end{aligned} \quad (5.50)$$

Para facilidad de manejo en su programación, la ecuación 5.50 se escribe:

$$p_i(x) = c_{1,i} + c_{2,i}(x-x_i) + c_{3,i}(x-x_i)^2 + c_{4,i}(x-x_i)^3 \quad (5.51)$$

con

$$\begin{aligned} c_{1,i} &= f(x_i), & c_{2,i} &= f'(x_i), \\ c_{3,i} &= f[x_p, x_p, x_{i+1}] - f[x_p, x_p, x_{i+1}, x_{i+1}]\Delta x_i \\ &= \frac{f[x_p, x_{i+1}] - f[x_p, x_i]}{x_{i+1} - x_i} - c_{4,i}\Delta x_i = \frac{f[x_p, x_{i+1}] - c_{2,i}}{\Delta x_i} - c_{4,i}\Delta x_i \end{aligned}$$

y

$$\begin{aligned} c_{4,i} &= f[x_p, x_p, x_{i+1}, x_{i+1}] = \frac{f[x_p, x_{i+1}, x_{i+1}] - f[x_p, x_p, x_{i+1}]}{\Delta x_i} \\ &= \frac{\frac{f[x_{i+1}, x_{i+1}] - f[x_p, x_{i+1}]}{x_{i+1} - x_i} - \frac{f[x_p, x_{i+1}] - f[x_p, x_i]}{x_{i+1} - x_i}}{x_{i+1} - x_i} \\ &= \frac{f'(x_{i+1}) - 2f[x_p, x_{i+1}] + f'(x_i)}{(x_{i+1} - x_i)^2} = \frac{f'(x_{i+1}) - 2f[x_p, x_{i+1}] + c_{2,i}}{\Delta x_i^2} \end{aligned} \quad (5.52)$$

### Ejemplo 5.13

Resuelva el problema del ejemplo 5.3 usando aproximación segmentaria, con polinomios de grado 2, 4, 6, ..., 16 y estime, como antes, el error máximo en forma práctica.

### Solución

En el CD se encuentra el **PROGRAMA 5.2** que realiza los cálculos solicitados





## Resultados

| Número de intervalos | Error máximo |
|----------------------|--------------|
| 2                    | 2.23620      |
| 4                    | 2.23622      |
| 6                    | 0.73979      |
| 8                    | 0.04213      |
| 10                   | 0.09341      |
| 12                   | 0.06417      |
| 14                   | 0.03299      |
| 16                   | 0.01279      |

En contraste con la aproximación polinomial (véase el ejemplo 5.3), el error máximo decrece conforme  $n$  crece.

**APROXIMACIÓN CÚBICA SEGMENTARIA DE BESSEL**

La aproximación cúbica de Hermite requiere el conocimiento de  $f'(x_i)$ ,  $i = 0, 1, \dots, n$ . Esta información, como se ha visto a lo largo del capítulo, no siempre existe, aún conociendo  $f(x)$  analíticamente no siempre es fácil obtenerla.

La aproximación cúbica segmentaria de Bessel se distingue por emplear una aproximación de  $f'(x_i)$  por

$$f'(x_i) \approx f[x_{i-1}, x_{i+1}], \quad i = 0, 1, \dots, n \quad (5.53)$$

y en todo lo demás se procede tal como en la aproximación de Hermite.

La expresión 5.53 requiere dos puntos adicionales a los que se tienen y son  $x_{-1}$  y  $x_{n+1}$ , ya que

$$f'(x_0) \approx f[x_{-1}, x_1], \text{ y } f'(x_n) \approx f[x_{n-1}, x_{n+1}],$$

llamadas **derivadas frontera** de  $g_3(x)$ .

Una forma de obtenerlos es una nueva subdivisión de  $[a, b]$ , como:

$$a = x_{-1} < x_0 < x_1 < x_2 < \dots < x_{n+1} = b. \quad (5.54)$$

podría también usarse

$$f'(x_{-1}) \text{ y } f'(x_{n+1}), \quad (5.55)$$

en caso de disponer de ellas y calcular las derivadas restantes de acuerdo con la ecuación (5.53).

Otra forma sería tomar  $f'(x_0)$  y  $f'(x_n)$  de manera que  $g_3(x)$  satisfaga las condiciones de extremo libre.

$$g_3''(a) = g_3''(b) = 0 \quad (5.56)$$

Independientemente de cómo se obtengan los puntos  $x_{-1}$  y  $x_{n+1}$ , las funciones  $g_3(x)$  y  $f(x)$  coinciden en los puntos de quiebre  $x_0, x_1, x_2, \dots, x_n$ . Por esto  $g_3(x)$  es continua en  $[a, b]$ , y por la ecuación 5.48 también es continuamente diferenciable. Además, es posible, y se muestra adelante, determinar  $f'(x_0), f'(x_1), \dots, f'(x_n)$  de manera que la  $g_3(x)$  resultante sea dos veces continuamente diferenciable.

El método de determinar  $g_3(x)$  con esta característica se conoce como **aproximación cúbica de trazador**, ya que la gráfica de  $g_3(x)$  se aproxima a la forma que tomaría una varilla delgada flexible si se forzara a pasar por cada punto  $(x_0, f(x_0)), (x_1, f(x_1)), \dots, (x_n, f(x_n))$ .

El requisito de que  $g_3(x)$  sea continuamente diferenciable dos veces puede darse como:

$$p_{i-1}''(x_i) = p_i''(x_i), \quad i = 1, 2, \dots, n-1 \quad (5.57)$$

o

$$2c_{3,i-1} + 6c_{4,i-1}\Delta x_{i-1} = 2c_{3,i}, \quad i = 1, 2, \dots, n-1,$$

conforme la ecuación 5.51 (derivándola dos veces).

Al sustituir las expresiones de la ecuación 5.52 en la última ecuación, se tiene:

$$\frac{2(f[x_{i-1}, x_i] - f'(x_{i-1}))}{\Delta x_{i-1}} + 6c_{4,i-1}\Delta x_{i-1} =$$

$$\frac{2(f[x_i, x_{i+1}] - f'(x_i))}{\Delta x_i} - 2c_{4,i}\Delta x_i, \quad i = 1, 2, \dots, n-1$$

Al continuar la sustitución y simplificar, se tiene:

$$\Delta x_i f'(x_{i-1}) + 2(\Delta x_{i-1} + \Delta x_i) f'(x_i) + \Delta x_{i-1} f'(x_{i+1}) =$$

$$3(f[x_{i-1}, x_i]\Delta x_i + f[x_i, x_{i+1}]\Delta x_{i-1}), \quad i = 1, 2, \dots, n-1 \quad (5.58)$$

Un sistema de  $n-1$  ecuaciones lineales en las  $(n+1)$  incógnitas  $f'(x_0), f'(x_1), \dots, f'(x_n)$ . Al obtener  $f'(x_0)$  y  $f'(x_n)$  de alguna manera (por ejemplo mediante las ecuaciones 5.53 o 5.55) se resuelve la 5.58 para  $f'(x_1), f'(x_2), \dots, f'(x_{n-1})$  por alguno de los métodos vistos en el capítulo 3; no obstante, como el sistema 5.58 es tridiagonal, conviene utilizar el algoritmo de Thomas.

#### Ejemplo 5.14

La siguiente tabla muestra las viscosidades del isopentano a 59°F y a diferentes presiones.

| Presión (psia) | Viscosidad (micropoises) |
|----------------|--------------------------|
| 426.690        | 2468                     |
| 483.297        | 2482                     |
| 497.805        | 2483                     |
| 568.920        | 2498                     |
| 995.610        | 2584                     |
| 1422.300       | 2672                     |
| 2133.450       | 2811                     |
| 3555.750       | 3094                     |
| 4266.900       | 3236                     |
| 7111.500       | 3807                     |

Elabore un programa para aproximar el valor de la viscosidad a las presiones de 355.575, 711.150, 2844.600, 5689.200 y 8533.801 psia, utilizando la aproximación cúbica segmentaria de Bessel.

**Solución** En el CD se encuentra el **PROGRAMA 5.3**, el cual proporciona los siguientes resultados.



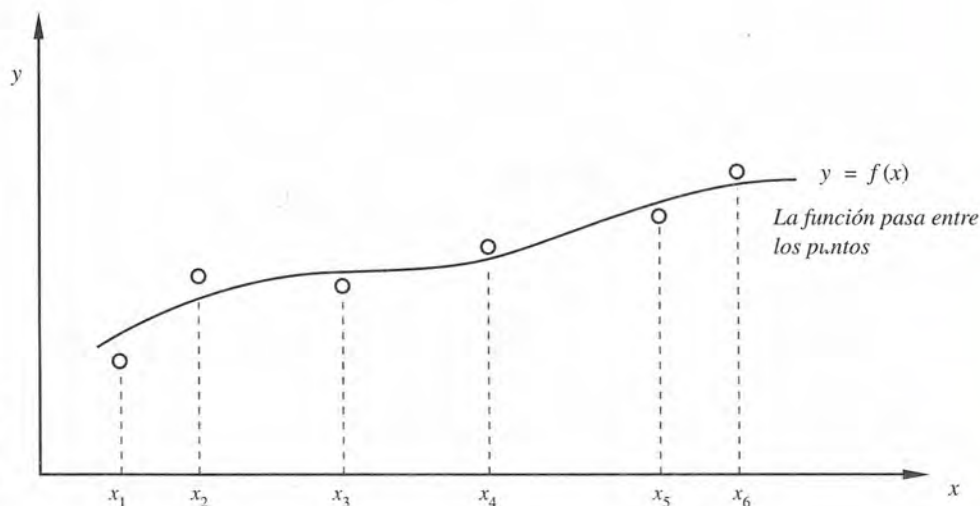
| Presión (psia) | Viscosidad (micropoises) |
|----------------|--------------------------|
| 355.575        | 2453.56                  |
| 711.150        | 2531.32                  |
| 2844.600       | 2950.92                  |
| 5689.200       | 3520.79                  |
| 8533.801       | 4093.21                  |

## 5.8 Aproximación polinomial con mínimos cuadrados

Hasta ahora el texto se ha enfocado en encontrar un polinomio de aproximación que pase **por los puntos** dados en forma tabular. Sin embargo, a veces la información (dada en la tabla) tiene errores significativos; por ejemplo, cuando proviene de medidas físicas. En estas circunstancias no tiene sentido pasar un polinomio de aproximación **por** los puntos dados, sino sólo cerca de ellos (véase Fig. 5.7).

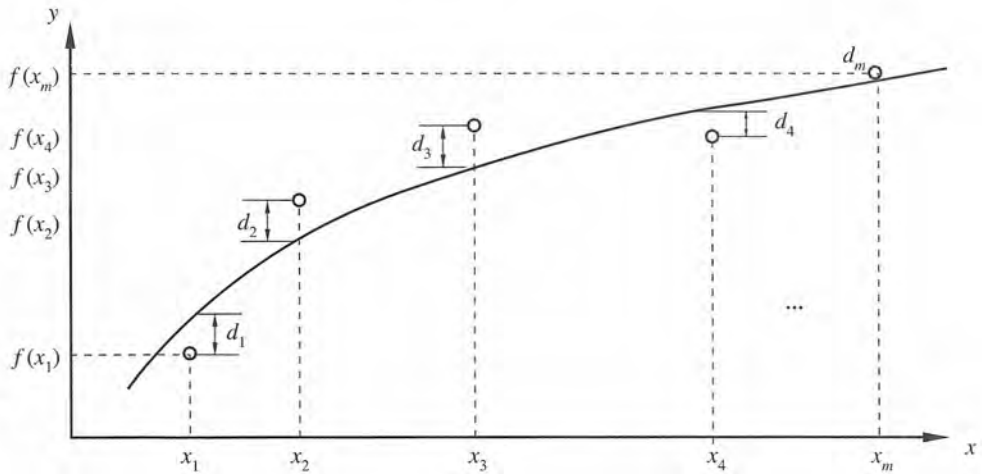
No obstante, esto crea un problema, ya que se puede pasar un número infinito de curvas **entre los puntos**. Para determinar **la mejor** curva se establece un criterio que la fije y una metodología que la determine. El criterio más común consiste en pedir que la suma de las distancias calculadas entre el valor de la función que aproxima  $p(x_i)$  y el valor de la función  $f(x_i)$  dada en la tabla, sea mínima (véase Fig. 5.8); es decir, que

$$\sum_{i=1}^m |p(x_i) - f(x_i)| = \sum_{i=1}^m d_i = \text{mínimo}$$



**Figura 5.7**  
Aproximación  
polinomial que  
pasa por entre  
los puntos.





**Figura 5.8**  
Ilustración de  
las distancias  $d_i$   
a minimizar.

Para evitar problemas de derivabilidad más adelante, se acostumbra utilizar las distancias  $d_i$ , elevadas al cuadrado:

$$\sum_{i=1}^m [p(x_i) - f(x_i)]^2 = \sum_{i=1}^m d_i^2 = \text{mínimo}$$

En la figura 5.8 se observan los puntos tabulados, la aproximación polinomial  $p(x)$  y las distancias  $d_i$  entre los puntos correspondientes, cuya suma hay que minimizar.

Si se utiliza

$$p(x) = a_0 + a_1x \quad (5.59)$$

para aproximar la función dada por la tabla, el problema queda como el de minimizar

$$\sum_{i=1}^m [a_0 + a_1x_i - f(x_i)]^2 \quad (5.60)$$

Hay que observar que del número infinito de polinomios que pasan entre los puntos, se selecciona aquél cuyos coeficientes  $a_0$  y  $a_1$  minimicen (5.60).

En el cálculo de funciones de una variable, el lector ha aprendido que para encontrar el mínimo o máximo de una función, se deriva y se iguala con cero esa derivada. Después se resuelve la ecuación resultante para obtener los valores de la variable que pudieran minimizar o maximizar la función. En el caso en estudio, donde se tiene una función por minimizar de dos variables ( $a_0$  y  $a_1$ ), el procedimiento es derivar parcialmente con respecto a cada una de las variables e igualar a cero cada derivada, con lo cual se obtiene un sistema de dos ecuaciones algebraicas en las incógnitas  $a_0$  y  $a_1$ ; o sea,

$$\begin{aligned} \frac{\partial}{\partial a_0} \left[ \sum_{i=1}^m (a_0 + a_1x_i - f(x_i))^2 \right] \\ \frac{\partial}{\partial a_1} \left[ \sum_{i=1}^m (a_0 + a_1x_i - f(x_i))^2 \right] \end{aligned} \quad (5.61)$$

Se deriva dentro del signo de sumatoria

$$\begin{aligned} \sum_{i=1}^m \frac{\partial}{\partial a_0} [a_0 + a_1x_i - f(x_i)]^2 &= \sum_{i=1}^m 2 [a_0 + a_1x_i - f(x_i)] \cdot 1 = 0 \\ \sum_{i=1}^m \frac{\partial}{\partial a_1} [a_0 + a_1x_i - f(x_i)]^2 &= \sum_{i=1}^m 2 [a_0 + a_1x_i - f(x_i)] x_i = 0 \end{aligned}$$

al desarrollar las sumatorias se tiene:

$$[a_0 + a_1 x_1 - f(x_1)] + [a_0 + a_1 x_2 - f(x_2)] + \dots + [a_0 + a_1 x_m - f(x_m)] = 0$$

$$[a_0 x_1 + a_1 x_1^2 - f(x_1)x_1] + [a_0 x_2 + a_1 x_2^2 - f(x_2)x_2] + \dots +$$

$$+ [a_0 x_m + a_1 x_m^2 - f(x_m)x_m] = 0$$

que simplificadas quedan:

$$m a_0 + a_1 \sum_{i=1}^m x_i = \sum_{i=1}^m f(x_i)$$

$$a_0 \sum_{i=1}^m x_i + a_1 \sum_{i=1}^m x_i^2 = \sum_{i=1}^m f(x_i) x_i$$

El sistema se resuelve por la regla de Cramer y se tiene

$$a_0 = \frac{[\sum_{i=1}^m f(x_i)] [\sum_{i=1}^m x_i^2] - [\sum_{i=1}^m x_i] [\sum_{i=1}^m f(x_i) x_i]}{m \sum_{i=1}^m x_i^2 - [\sum_{i=1}^m x_i]^2} \quad (5.62)$$

$$a_1 = \frac{m \sum_{i=1}^m f(x_i) x_i - [\sum_{i=1}^m f(x_i)] [\sum_{i=1}^m x_i]}{m \sum_{i=1}^m x_i^2 - [\sum_{i=1}^m x_i]^2}$$

que sustituidos en la ecuación 5.59 dan la aproximación polinomial de primer grado que **mejor ajusta** la información tabulada. Este polinomio puede usarse a fin de aproximar valores de la función para argumentos no conocidos en la tabla.

### Ejemplo 5.15

En la tabla siguiente se presentan los alargamientos de un resorte correspondientes a fuerzas de diferente magnitud que lo deforman.

| Puntos                        | 1     | 2     | 3     | 4     | 5     |
|-------------------------------|-------|-------|-------|-------|-------|
| Fuerza (kgf): $x$             | 0     | 2     | 3     | 6     | 7     |
| Longitud del resorte (m): $y$ | 0.120 | 0.153 | 0.170 | 0.225 | 0.260 |

Determine por mínimo cuadrados el mejor polinomio de primer grado (recta) que represente la función dada.

### Solución

Para facilitar los cálculos y evitar errores en los mismos, primero se construye la siguiente tabla.

| Puntos | Fuerza $x_i$      | Longitud $y_i$       | $x_i^2$             | $x_i y_i$                |
|--------|-------------------|----------------------|---------------------|--------------------------|
| 1      | 0                 | 0.120                | 0                   | 0.000                    |
| 2      | 2                 | 0.153                | 4                   | 0.306                    |
| 3      | 3                 | 0.170                | 9                   | 0.510                    |
| 4      | 6                 | 0.225                | 36                  | 1.350                    |
| 5      | 7                 | 0.260                | 49                  | 1.820                    |
|        | $\Sigma x_i = 18$ | $\Sigma y_i = 0.928$ | $\Sigma x_i^2 = 98$ | $\Sigma x_i y_i = 3.986$ |

Los valores de las sumatorias de la última fila se sustituyen en el sistema de ecuaciones 5.62 y se obtiene:

$$a_0 = 0.11564 \text{ y } a_1 = 0.019434, \text{ de donde}$$

$$p(x) = 0.11564 + 0.019434 x.$$

Los cálculos pueden realizarse con Matlab o con la TI-92 Plus.

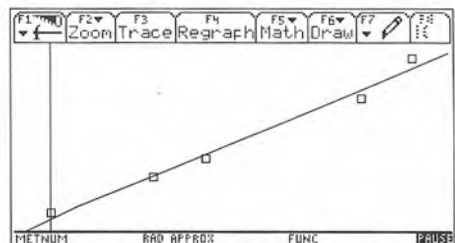


```
format long
x=[0 2 3 6 7];
y=[0.120 0.153 0.170 0.225 0.260];
a=polyfit(x,y,1)
fprintf('a0=%8.5f a1=%9.6f\n',a(2),a(1))
```



```
e5_15()
Prgm
{0,2,3,6,7}→1x : 5→n
{.12,.153,.17,.225,.26}→1y
LinReg 1x,1y
ShowStat
1x[1]-.1*(max(1x)-min(1x))→xmin
1x[n]+.1*(max(1x)-min(1x))→xmax
min(1y)-.1*(max(1y)-min(1y))→ymin
max(1y)+.1*(max(1y)-min(1y))→ymax
regeq(x)→y1(x)
NewPlot 1,1,1x,1y
setMode("Split 1 App","Graph")
Pause
setMode("Split 1 App","Home")
EndPrgm
```

Este programa genera las dos pantallas siguientes en la TI-92 Plus



El grado del polinomio no tiene relación con el número de puntos usados y debe seleccionarse de antemano con base en consideraciones teóricas que apoyan el fenómeno estudiado, el diagrama de dispersión (puntos graficados en el plano  $x$ - $y$ ) o ambos.

El hecho de tener la **mejor recta** que aproxima la información, no significa que la información esté bien aproximada; quizá convenga aproximarla con una parábola o una cúbica.



Para encontrar el polinomio de segundo grado  $p_2(x) = a_0 + a_1x + a_2x^2$  que mejor aproxime la tabla, se minimiza

$$\sum_{i=1}^m [a_0 + a_1x_i + a_2x_i^2 - f(x_i)]^2 \quad (5.63)$$

donde los parámetros  $a_0$ ,  $a_1$  y  $a_2$  se obtienen al resolver el sistema de ecuaciones lineales que resulta de derivar parcialmente e igualar a cero la función por minimizar con respecto a cada uno. Dicho sistema queda:

$$\begin{aligned} m a_0 + a_1 \sum_{i=1}^m x_i + a_2 \sum_{i=1}^m x_i^2 &= \sum_{i=1}^m f(x_i) \\ a_0 \sum_{i=1}^m x_i + a_1 \sum_{i=1}^m x_i^2 + a_2 \sum_{i=1}^m x_i^3 &= \sum_{i=1}^m f(x_i)x_i \\ a_0 \sum_{i=1}^m x_i^2 + a_1 \sum_{i=1}^m x_i^3 + a_2 \sum_{i=1}^m x_i^4 &= \sum_{i=1}^m f(x_i)x_i^2, \end{aligned} \quad (5.64)$$

Cuya solución puede obtenerse por alguno de los métodos vistos en el capítulo 3.

### Ejemplo 5.16

El calor específico  $C_p$  (cal/k gmo1) del  $Mn_3O_4$  varía con la temperatura de acuerdo con la siguiente tabla.

| Punto              | 1    | 2    | 3     | 4    | 5    | 6    |
|--------------------|------|------|-------|------|------|------|
| T (K)              | 280  | 650  | 1000  | 1200 | 1500 | 1700 |
| $C_p$ (cal/k gmo1) | 32.7 | 45.4 | 52.15 | 53.7 | 52.9 | 50.3 |

Aproxime esta información con un polinomio por el método de mínimos cuadrados.

### Solución



El calor específico aumenta con la temperatura hasta el valor tabulado de 1200 K, para disminuir posteriormente en valores más altos de temperatura. Esto sugiere utilizar un polinomio con curvatura en vez de una recta, por ejemplo, uno de segundo grado, que es el más simple.

Para facilitar el cálculo de los coeficientes del sistema de ecuaciones 5.64, se construye la siguiente tabla.

| Puntos<br>$i$    | T<br>$x_i$ | $C_p$<br>$y_i$ | $x_i^2$            | $x_i^3$             | $x_i^4$                | $y_i x_i$ | $y_i x_i^2$          |
|------------------|------------|----------------|--------------------|---------------------|------------------------|-----------|----------------------|
| 1                | 280        | 32.7           | $0.78 \times 10^5$ | $0.022 \times 10^9$ | $0.062 \times 10^{11}$ | 9156      | $2.56 \times 10^6$   |
| 2                | 650        | 45.4           | $0.42 \times 10^6$ | $0.275 \times 10^9$ | $1.785 \times 10^{11}$ | 29510     | $19.18 \times 10^6$  |
| 3                | 1000       | 52.15          | $1.00 \times 10^6$ | $1.000 \times 10^9$ | $1.000 \times 10^{12}$ | 52150     | $52.15 \times 10^6$  |
| 4                | 1200       | 53.7           | $1.44 \times 10^6$ | $1.728 \times 10^9$ | $2.074 \times 10^{12}$ | 64440     | $77.33 \times 10^6$  |
| 5                | 1500       | 52.9           | $2.25 \times 10^6$ | $3.375 \times 10^9$ | $5.063 \times 10^{12}$ | 79350     | $119.03 \times 10^6$ |
| 6                | 1700       | 50.3           | $2.89 \times 10^6$ | $4.900 \times 10^9$ | $8.350 \times 10^{12}$ | 85510     | $145.37 \times 10^6$ |
| $\Sigma$ Totales | 6330       | 287.15         | $8.08 \times 10^6$ | $11.3 \times 10^9$  | $166.7 \times 10^{11}$ | 320116    | $415.62 \times 10^6$ |

Los coeficientes se sustituyen en el sistema de ecuaciones 5.64 y se obtiene:

$$6 a_0 + 6330 a_1 + 8.08 \times 10^6 a_2 = 287.15$$

$$6330 a_0 + 8.08 \times 10^6 a_1 + 11.30 \times 10^9 a_2 = 320116$$

$$8.08 \times 10^6 a_0 + 11.30 \times 10^9 a_1 + 166.70 \times 10^{11} a_2 = 415.62 \times 10^6$$

cuya solución por el método de eliminación Gaussiana arroja

$$a_0 = 19.29544, a_1 = 0.053728, a_2 = -2.08787 \times 10^{-5},$$

que forma la aproximación polinomial siguiente:

$$C_p(T) \approx p_2(T) = 19.29544 + 0.053728 T - 2.08787 \times 10^{-5} T^2$$

Los valores de las sumas no se escribieron con todas sus cifras significativas, pero el polinomio de regresión se calculó usando todas las cifras que conserva la computadora.

Los cálculos pueden realizarse con Matlab o con la TI-92 Plus.



```
format long
T=[280 650 1000 1200 1500 1700];
Cp=[32.7 45.4 52.15 53.7 52.9 50.3];
a=polyfit(T,Cp,2);
fprintf('a0=%8.5f a1=%9.6f a2=%9.6f\n',a(3),a(2),a(1))
Tint=800;
Cpint=a(3)+a(2)*Tint+a(1)*Tint^2;
fprintf(' Cp(%4.0f)=%6.1f\n',Tint,Cpint)
```

Otra forma, usando las ecuaciones normales para la regresión:

```
T=[280 650 1000 1200 1500 1700];
Cp=[32.7 45.4 52.15 53.7 52.9 50.3];
A=[length(T) sum(T) sum(T.^2);...
    sum(T) sum(T.^2) sum(T.^3);...
    sum(T.^2) sum(T.^3) sum(T.^4)]
b=[sum(Cp); sum(Cp.*T); sum(Cp.*T.^2)]
a=A\b
for i=1:length(a)
    aa(i)=a(length(a)+1-i);
end
polyval(aa,800)
```

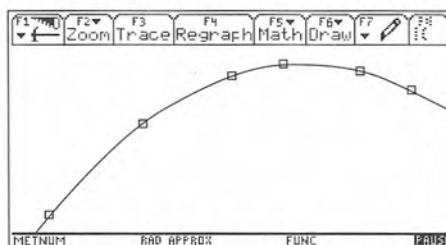
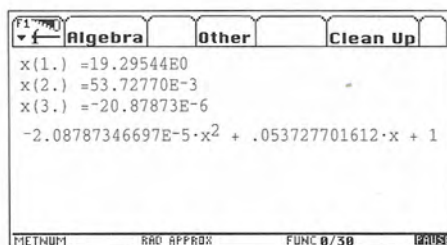


```

e5_16()
Prgm
Define minimos(lx,ly,np,ng)=Prgm
ng+1→nec:2*nec-1→nn : newList(nn)→s:s→ss
For i,1,np
  l→xx
  For j,1,nn
    If j≤nec
      ss[j]+xx*ly[i]→ss[j]
      xx*lx[i]→xx : s[j]+xx→s[j]
    EndFor
  EndFor
  newMat(nec,nec)→a : newMat(nec,1)→b : np→a[1,1]
  For i,1,nec
    For j,1,nec
      If not(i=1 and j=1)
        s[j-2+i]→a[i,j]
      Endfor
      ss[i]→b[i,1]
    EndFor
  EndPrgm
  {280,650,1000,1200,1500,1700}→lx :
  {32.7,45.4,52.15,53.7,52.9,50.3}→ly : 6→np: 2→ng: ClrIO
  minimos(lx,ly,np,ng) : simult(a,b)→c : 0→p : DelVar x
  For i,1,ng+1
    Disp "x("&string(i)&")="&format(c[i,1],"e6") :
    p+c[i,1]*x(i-1)→p
  EndFor
  Disp p : Pause : FnOff
  lx[1]-.1*(max~lx)-min(lx)→xmin
  lx[np]+.1*(max(lx)-min~lx)→xmax
  min(ly)-.1*(max(ly)-min(ly)→ymin
  max(ly)+.1*(max(ly)-min(ly)→ymax
  2→xres:0→yscl:NewPlot 1,1,lx,ly:DrawFunc p:FnOn:Pause
  setMode("Split 1 App","Home")
EndPrgm

```

Con este programa se obtienen los siguientes resultados:



**NOTA:** Muchas de las calculadoras de mano cuentan con un programa interno para obtener esta aproximación; por otro lado, puede usarse un pizarrón electrónico para los cálculos (sumatorias, solución de ecuaciones, etcétera).



**Ejemplo 5.17**

Use aproximación polinomial de segundo grado obtenida en el ejemplo anterior para aproximar el calor específico del  $\text{Mn}_3\text{O}_4$  a una temperatura de 800 K.

**Solución**

Con la sustitución de  $T = 800$  K en el polinomio de aproximación se tiene:

$$\begin{aligned} C_p(800) &\approx p_2(800) = 19.29544 + 0.053728(800) - 2.08787 \times 10^{-5} (800)^2 \\ &= 48.9 \text{ cal/K g mol.} \end{aligned}$$

En caso de querer aproximar una función dada en forma tabular con un polinomio de grado más alto,  $n$  por ejemplo, el procedimiento es el mismo; esto es, minimizar la función

$$\sum_{i=1}^m [a_0 + a_1 x_i + a_2 x_i^2 + \dots + a_n x_i^n - f(x_i)]^2,$$

lo cual se obtiene derivándola parcialmente con respecto a cada coeficiente  $a_j$ , ( $0 < j < n$ ) e igualando a cero cada una de estas derivadas. Con esto se llega al sistema lineal

$$\begin{array}{cccccccl} ma_0 & + & a_1 \sum x & + & a_2 \sum x^2 & + & \dots & + & a_n \sum x^n & = & \sum y \\ a_0 \sum x & + & a_1 \sum x^2 & + & a_2 \sum x^3 & + & \dots & + & a_n \sum x^{n+1} & = & \sum xy \\ a_0 \sum x^2 & + & a_1 \sum x^3 & + & a_2 \sum x^4 & + & \dots & + & a_n \sum x^{n+2} & = & \sum x^2 y \\ \vdots & & & & & & & & \vdots & & \\ a_0 \sum x^n & + & a_1 \sum x^{n+1} & + & a_2 \sum x^{n+2} & + & \dots & + & a_n \sum x^{n+n} & = & \sum x^n y \end{array}$$

donde se han omitido, los subíndices  $i$ , de  $x$  y  $y$ , así como los límites de las sumatorias que van de 1 hasta  $m$  para simplificar su escritura.

**ALGORITMO 5.5** Aproximación con mínimos cuadrados

Para obtener los  $N+1$  coeficientes del polinomio óptimo de grado  $N$  que pasa entre  $M$  parejas de puntos, proporcionar los

**DATOS:** El grado del polinomio de aproximación  $N$ , el número de parejas de valores  $(X(I), F_X(I), I = 1, 2, \dots, M)$ .

**RESULTADOS:** Los coeficientes  $A(0), A(1), \dots, A(N)$  del polinomio de aproximación.

PASO 1. Hacer  $J = 0$ .

PASO 2. Mientras  $J \leq (2*N-1)$ , repetir los pasos 3 a 5.

PASO 3. Si  $J \leq N$  Hacer  $SS(J) = 0$ . De otro modo continuar.

PASO 4. Hacer  $S(J) = 0$ .

PASO 5. Hacer  $J = J + 1$ .

PASO 6. Hacer  $I = 1$ .

PASO 7. Mientras  $I \leq M$ , repetir los pasos 8 a 15.

PASO 8. Hacer  $XX = 1$ .

PASO 9. Hacer  $J = 0$ .

PASO 10. Mientras  $J \leq (2*N-1)$ , repetir los pasos 11 a 14.

PASO 11. Si  $J \leq N$  hacer  $SS(J) = SS(J) + XX*F_X(I)$ .

De otro modo continuar.

PASO 12. Hacer  $XX = XX*X(I)$ .

PASO 13. Hacer  $S(J) = S(J) + XX$ .

PASO 14. Hacer  $J = J + 1$ .

PASO 15. Hacer  $I = I + 1$ .

PASO 16. Hacer  $B(0,0) = M$ .

PASO 17. Hacer  $I = 0$ .

PASO 18. Mientras  $I \leq N$ , repetir los pasos 19 a 24.

PASO 19. Hacer  $J = 0$ .

PASO 20. Mientras  $J \leq N$ , repetir los pasos 21 y 22.

PASO 21. Si  $I \neq 0$  y  $J \neq 0$ .

Hacer  $B(I, J) = S(J-1+I)$ .

PASO 22. Hacer  $J = J + 1$ .

PASO 23. Hacer  $B(I, N+1) = SS(I)$ .

PASO 24. Hacer  $I = I + 1$ .

PASO 25. Resolver el sistema de ecuaciones lineales  $B \mathbf{a} = \mathbf{ss}$  de orden  $N+1$  con alguno de los algoritmos del capítulo 3.

PASO 26. IMPRIMIR  $A(0), A(1), \dots, A(N)$  y TERMINAR.



En el CD encontrará el **PROGRAMA 5.9** de Regresión. Con este programa usted puede proporcionar la función como una tabla de puntos, aproximar con el método de mínimos cuadrados utilizando un polinomio de grado seleccionado. Podrá además observar gráficamente los puntos dados, el polinomio de ajuste y el valor a interpolar.

## 5.9 Aproximación multilinear con mínimos cuadrados

Con frecuencia se tienen funciones de más de una variable; esto es,  $f(u, v, z)$ . Si se sospecha una funcionalidad lineal en las distintas variables; es decir, si se piensa que la función

$$y = a_0 + a_1 u + a_2 v + a_3 z$$

Puede ajustar los datos de la tabla siguiente:

| Puntos | $u$   | $v$   | $z$   | $y$                |
|--------|-------|-------|-------|--------------------|
| 1      | $u_1$ | $v_1$ | $z_1$ | $f(u_1, v_1, z_1)$ |
| 2      | $u_2$ | $v_2$ | $z_2$ | $f(u_2, v_2, z_2)$ |
| 3      | $u_3$ | $v_3$ | $z_3$ | $f(u_3, v_3, z_3)$ |
| .      | .     | .     | .     | .                  |
| .      | .     | .     | .     | .                  |
| .      | .     | .     | .     | .                  |
| $m$    | $u_m$ | $v_m$ | $z_m$ | $f(u_m, v_m, z_m)$ |

Se puede aplicar el método de los mínimos cuadrados para determinar los coeficientes  $a_0, a_1, a_2$  y  $a_3$  que mejor aproximen la función de varias variables tabulada. El procedimiento es análogo al descrito anteriormente y consiste en minimizar la función.

$$\sum_{i=1}^m [(a_0 + a_1 u_i + a_2 v_i + a_3 z_i) - y_i]^2$$

que derivada parcialmente con respecto de cada coeficiente por determinar:  $a_0, a_1, a_2, a_3$  e igualada a cero cada una, queda

$$\frac{\partial}{\partial a_0} \sum_{i=1}^m [(a_0 + a_1 u_i + a_2 v_i + a_3 z_i) - y_i]^2 = 2 \sum_{i=1}^m (a_0 + a_1 u_i + a_2 v_i + a_3 z_i - y_i) = 0$$

$$\frac{\partial}{\partial a_1} \sum_{i=1}^m [(a_0 + a_1 u_i + a_2 v_i + a_3 z_i) - y_i]^2 = 2 \sum_{i=1}^m (a_0 + a_1 u_i + a_2 v_i + a_3 z_i - y_i) u_i = 0$$

$$\frac{\partial}{\partial a_2} \sum_{i=1}^m [(a_0 + a_1 u_i + a_2 v_i + a_3 z_i) - y_i]^2 = 2 \sum_{i=1}^m (a_0 + a_1 u_i + a_2 v_i + a_3 z_i - y_i) v_i = 0$$

$$\frac{\partial}{\partial a_3} \sum_{i=1}^m [(a_0 + a_1 u_i + a_2 v_i + a_3 z_i) - y_i]^2 = 2 \sum_{i=1}^m (a_0 + a_1 u_i + a_2 v_i + a_3 z_i - y_i) z_i = 0$$

ecuaciones que arregladas generan el sistema algebraico lineal siguiente:

$$\begin{aligned}
 m a_0 + a_1 \sum u + a_2 \sum v + a_3 \sum z &= \sum y \\
 a_0 \sum u + a_1 \sum u^2 + a_2 \sum uv + a_3 \sum uz &= \sum uy \\
 a_0 \sum v + a_1 \sum vu + a_2 \sum v^2 + a_3 \sum vz &= \sum vy \\
 a_0 \sum z + a_1 \sum zu + a_2 \sum zv + a_3 \sum z^2 &= \sum zy
 \end{aligned} \tag{5.65}$$

en las incógnitas  $a_0, a_1, a_2$  y  $a_3$ . Para simplificar la escritura se han omitido los índices  $i$ , de  $u, v$ , y  $z$  y los límites de las sumatorias, que van de 1 hasta  $m$ .

### Ejemplo 5.18

A partir de un estudio experimental acerca de la estabilización de arcilla muy plástica, se observó que el contenido de agua para moldear con densidad óptima dependía linealmente de los porcentajes de cal y puzolana mezclados con la arcilla. Se tuvieron así los resultados que se dan abajo. Ajuste una ecuación de la forma:

$$y = a_0 + a_1 u + a_2 v$$

a los datos de dicha tabla.

| Agua ( % ) | Cal ( % ) | Puzolana ( % ) |
|------------|-----------|----------------|
| $y$        | $u$       | $v$            |
| 27.5       | 2.0       | 18.0           |
| 28.0       | 3.5       | 16.5           |
| 28.8       | 4.5       | 10.5           |
| 29.1       | 2.5       | 2.5            |
| 30.0       | 8.5       | 9.0            |
| 31.0       | 10.5      | 4.5            |
| 32.0       | 13.5      | 1.5            |

**Solución** El sistema por resolver es una modificación del sistema de ecuaciones 5.65 para una función  $y$  de dos variables  $u$  y  $v$

$$\begin{aligned}
 n a_0 + a_1 \sum u + a_2 \sum v &= \sum y \\
 a_0 \sum u + a_1 \sum u^2 + a_2 \sum uv &= \sum uy \\
 a_0 \sum v + a_1 \sum vu + a_2 \sum v^2 &= \sum vy
 \end{aligned}$$

Con el objeto de facilitar el cálculo del sistema anterior se construye la siguiente tabla.

| $i$              | $u_i$ | $v_i$ | $y_i$ | $u_i^2$ | $u_i v_i$ | $v_i^2$ | $u_i y_i$ | $v_i y_i$ |
|------------------|-------|-------|-------|---------|-----------|---------|-----------|-----------|
| 1                | 2.0   | 18.0  | 27.5  | 4.00    | 36.00     | 324.00  | 55.00     | 495.00    |
| 2                | 3.5   | 16.5  | 28.0  | 12.25   | 57.75     | 272.25  | 98.00     | 462.00    |
| 3                | 4.5   | 10.5  | 28.8  | 20.25   | 47.25     | 110.25  | 129.60    | 302.40    |
| 4                | 2.5   | 2.5   | 29.1  | 6.25    | 6.25      | 6.25    | 72.75     | 72.75     |
| 5                | 8.5   | 9.0   | 30.0  | 72.25   | 76.50     | 81.00   | 255.00    | 270.00    |
| 6                | 10.5  | 4.5   | 31.0  | 110.25  | 47.25     | 20.25   | 325.50    | 139.50    |
| 7                | 13.5  | 1.5   | 32.0  | 182.25  | 20.25     | 2.25    | 432.00    | 48.00     |
| $\Sigma$ Totales | 45.0  | 62.5  | 206.4 | 407.5   | 291.25    | 816.25  | 1367.85   | 1789.65   |



Los coeficientes se sustituyen en el sistema de ecuaciones y al aplicar alguno de los métodos del capítulo 3, se obtiene:

$$a_0 = 28.69, \quad a_1 = 0.2569, \quad a_2 = -0.09607$$

al sustituir estos valores se tiene:

$$y = 28.69 + 0.2569 u - 0.09607 v$$

Los cálculos pueden realizarse con Matlab o con la TI-92 Plus

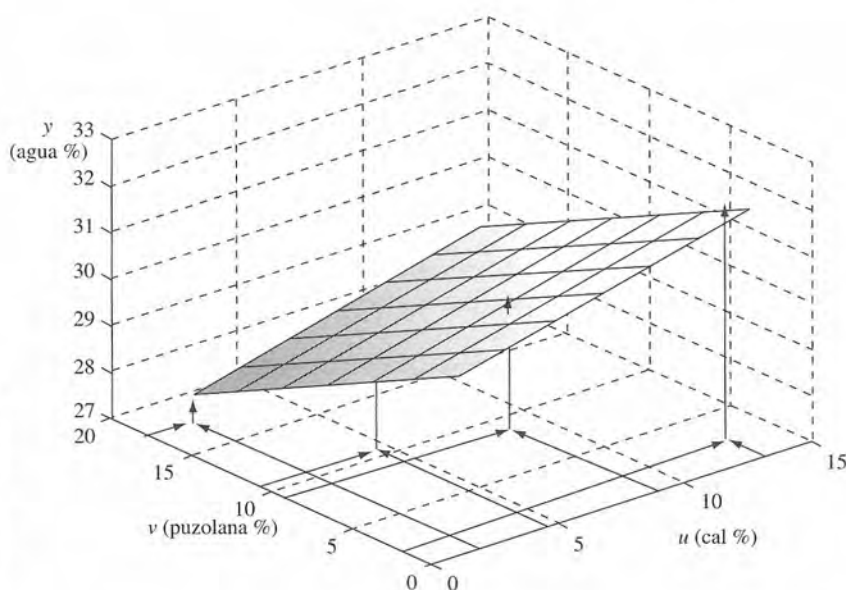


```
u=[2; 3.5; 4.5; 2.5; 8.5; 10.5; 13.5];
v=[18; 16.5; 10.5; 2.5; 9; 4.5; 1.5];
y=[27.5; 28; 28.8; 29.1; 30; 31; 32];
A=[size(u,1) sum(u) sum(v);...
   sum(u) sum(u.^2) sum(u.*v);...
   sum(v) sum(v.*u) sum(v.^2)];
b=[sum(y);sum(u.*y);sum(v.*y)]
a=A\b
```



```
e5_18()
Prgm
{2,3.5,4.5,2.5,8.5,10.5,13.5}→u
{18,16.5,10.5,2.5,9,4.5,1.5}→v
{27.5,28,28.8,29.1,30,31,32}→y
@Nota: Las siguientes dos líneas son una sola instruccion
[dim(u),sum(u),sum(v);sum(u),sum(u.^2),sum(u.*v);
sum(v),sum(v.*u),sum(v.^2)]→a
[sum(y);sum(y.*u);sum(y.*v)]→b
simult(a,b)→c
Disp c
EndPrgm
```

Al graficar en el espacio la ecuación  $y = 28.69 + 0.2569 u - 0.09607 v$ , resulta un plano que pasa por entre los puntos experimentales, quedando algunos de ellos abajo y otros arriba y los demás en la superficie, pero la suma de los cuadrados de las distancias de estos puntos a la superficie es mínima, respecto a cualquier otro plano que pase entre dichos puntos (ver Fig. 5.9).



**Figura 5.9**  
Gráfica del  
plano  $y = 28.69$   
 $+ 0.2569 u$   
 $- 0.09607 v$  y  
algunos datos  
experimentales.

## Ejercicios

**5.1** A continuación se presentan las presiones de vapor del cloruro de magnesio.

| Puntos   | 0   | 1   | 2    | 3    | 4    | 5    | 6    | 7    |
|----------|-----|-----|------|------|------|------|------|------|
| P (mmHg) | 10  | 20  | 40   | 60   | 100  | 200  | 400  | 760  |
| T (°C)   | 930 | 988 | 1050 | 1088 | 1142 | 1316 | 1223 | 1418 |

Calcule la presión de vapor correspondiente a  $T = 1000$  °C.

### Solución



Como la información no está regularmente espaciada en los argumentos (T), pueden usarse diferencias divididas o polinomios de Lagrange para la interpolación. (Se sugiere ver los PROGRAMAS 5.4 Y 5.5 del CD).

Con los polinomios de Lagrange de segundo grado se tiene:

$$p_2(x) = f(x_0) \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f(x_1) \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} + f(x_2) \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)}$$

Al tomar las presiones como valores de la función  $f(x)$ , las temperaturas como los argumentos  $x$ , seleccionar los puntos (0), (1) y (2), y sustituir los valores, se obtiene:

$$p_2(1000) = 10 \frac{(1000 - 988)(1000 - 1050)}{(930 - 988)(930 - 1050)} + 20 \frac{(1000 - 930)(1000 - 1050)}{(988 - 930)(988 - 1050)} + 40 \frac{(1000 - 930)(1000 - 988)}{(1050 - 930)(1050 - 988)} = 23.12 \text{ mmHg} \approx 23 \text{ mmHg}$$

## 5.2 Dada la tabla

| Puntos   | 0       | 1       | 2       | 3       | 4       |
|----------|---------|---------|---------|---------|---------|
| $x_i$    | 1.00    | 1.35    | 1.70    | 1.90    | 3.00    |
| $f(x_i)$ | 0.00000 | 0.30010 | 0.53063 | 0.64185 | 1.09861 |

Construya una tabla de diferencias divididas para aproximar  $f(x)$  en  $x = 1.50$ ; utilice un polinomio de Newton de segundo grado.

**Solución** A continuación se da la tabla de diferencias divididas.



| Puntos | $x_i$ | $f(x_i)$ | Diferencias divididas |          |          |          |
|--------|-------|----------|-----------------------|----------|----------|----------|
|        |       |          | Primeras              | Segundas | Terceras | Cuartas  |
| 0      | 1.00  | 0.00000  |                       |          |          |          |
|        |       |          | 0.85743               |          |          |          |
| 1      | 1.35  | 0.30010  |                       | -0.28396 |          |          |
|        |       |          | 0.65866               |          | 0.10832  |          |
| 2      | 1.70  | 0.53063  |                       | -0.18647 |          | -0.03049 |
|        |       |          | 0.55610               |          | -0.04735 |          |
| 3      | 1.90  | 0.64185  |                       | -0.10835 |          |          |
|        |       |          | 0.41524               |          |          |          |
| 4      | 3.00  | 1.09861  |                       |          |          |          |

Se pueden seleccionar los puntos (0), (1) y (2) para el polinomio de interpolación o bien (1), (2) y (3). Se escoge el segundo conjunto de puntos, ya que están más cerca de 1.5 que el primero; sin embargo, al querer emplear la fórmula

$$p_2(x) = f[x_0] + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2]$$

y la tabla construida se deberá tener cuidado, ya que el valor  $x_0$  de la fórmula en realidad corresponderá a  $x_1$  de la tabla;  $x_1$  de la fórmula, a  $x_2$  de la tabla, etc. Consecuentemente  $f[x_0]$ ,  $f[x_0, x_1, x_2]$  de la fórmula corresponderá a  $f[x_1]$ ,  $f[x_1, x_2]$  y  $f[x_1, x_2, x_3]$  de la tabla respectivamente (véase la línea diagonal de la tabla).

Con la sustitución de valores queda

$$\begin{aligned} p_2(1.5) &= 0.30010 + (1.5 - 1.35)(0.65866) + (1.5 - 1.35)(1.5 - 1.7)(-0.18647) \\ &= 0.40449 \end{aligned}$$

Una solución alterna es construir la tabla de diferencias de modo que quede como punto (0) el más cercano a 1.5 (1.35 en este caso), entre los puntos restantes se elige como punto (1) el más cercano a 1.5 (1.70 en este caso), etc. Adelante se muestra cómo queda esta tabla.



| Puntos | $x_i$ | $f(x_i)$ | Diferencias divididas |          |          |           |
|--------|-------|----------|-----------------------|----------|----------|-----------|
|        |       |          | Primeras              | Segundas | Terceras | Cuartas   |
| 0      | 1.35  | 0.30010  |                       |          |          |           |
|        |       |          | 0.65866               |          |          |           |
| 1      | 1.70  | 0.53063  |                       | -0.18647 |          |           |
|        |       |          | 0.55610               |          | 0.10846  |           |
| 2      | 1.90  | 0.64185  |                       | -0.22443 |          | -0.030567 |
|        |       |          | 0.71320               |          | 0.05802  |           |
| 3      | 1.00  | 0.00000  |                       | -0.14900 |          |           |
|        |       |          | 0.54930               |          |          |           |
| 4      | 3.00  | 1.09861  |                       |          |          |           |

con lo cual puede usarse el polinomio

$$p_2(x) = f[x_0] + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2]$$

directamente, ya que ahora los subíndices de los argumentos de la fórmula y de la tabla se corresponden. Sustituyendo valores se tiene:

$$\begin{aligned} p_2(1.5) &= 0.30010 + (1.5 - 1.35)(0.65866) + (1.5 - 1.35)(1.5 - 1.7)(-0.18647) \\ &= 0.40449 \end{aligned}$$

Obsérvese que el valor interpolado es el mismo que se tuvo anteriormente.

- 5.3** Las densidades de las soluciones acuosas del ácido sulfúrico varían con la temperatura y la concentración de acuerdo con la tabla.

| C (%) | T (°C) |        |        |        |
|-------|--------|--------|--------|--------|
|       | 10     | 30     | 60     | 100    |
| 5     | 1.0344 | 1.0281 | 1.0140 | 0.9888 |
| 20    | 1.1453 | 1.1335 | 1.1153 | 1.0885 |
| 40    | 1.3103 | 1.2953 | 1.2732 | 1.2446 |
| 70    | 1.6923 | 1.6014 | 1.5753 | 1.5417 |

- Calcule la densidad a una concentración de 40% y una temperatura de 15 °C.
- Calcule la densidad a 30 °C y concentración de 50%.
- Calcule la densidad a 50 °C y 60% de concentración.
- Calcule la temperatura a la cual una solución al 30% tiene una densidad de 1.215.

### Solución

- La temperatura se toma como el argumento  $x$  y las densidades (a 40%) como el valor de la función  $f(x)$ .

Con una interpolación lineal entre las densidades a 10 °C y 30 °C se tiene:

$$p(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1)$$

$$d(15) \approx \frac{15 - 30}{10 - 30} 1.3103 + \frac{15 - 10}{30 - 10} 1.2953 = 1.3066$$

- b) Se toman ahora las concentraciones como argumentos  $x$  y las densidades (a 30 °C) como los valores funcionales; luego, mediante una interpolación lineal entre las concentraciones a 40% y 70% queda:

$$d(50) = \frac{50 - 70}{40 - 70} 1.2953 + \frac{50 - 40}{70 - 40} 1.6014 = 1.3973$$

- c) La densidad se aproxima a 50 °C, utilizando primero la fila de 40% de concentración y después la fila de 70% de concentración. Con estas densidades obtenidas a 50 °C se aproxima la densidad a 60% de concentración.

### Primer paso

Aproximación de la densidad a 40% y 50 °C.

$$d \approx \frac{50 - 60}{30 - 60} 1.2953 + \frac{50 - 30}{60 - 30} 1.2732 = 1.2806$$

### Segundo paso

Aproximación de la densidad a 70% y 50 °C.

$$d \approx \frac{50 - 60}{30 - 60} 1.6014 + \frac{50 - 30}{60 - 30} 1.5753 = 1.5840$$

### Tercer paso

Aproximación de la densidad a 60% y 50 °C usando los valores obtenidos en los pasos anteriores

$$d \approx \frac{60 - 70}{40 - 70} 1.2806 + \frac{60 - 40}{70 - 40} 1.5840 = 1.4829$$

- d) En este caso es necesario interpolar los valores de la densidad a 30% de concentración a diferentes temperaturas, para después interpolar la temperatura que corresponde a una densidad de 1.215.

### Primer paso

Aproximación de la densidad a 30% y 10 °C.

$$d \approx \frac{30 - 20}{40 - 20} 1.1453 + \frac{30 - 40}{20 - 40} 1.3103 = 1.2278$$

Aproximación de la densidad a 30% y 30 °C.

$$d \approx \frac{30 - 20}{40 - 20} 1.1335 + \frac{30 - 40}{20 - 40} 1.2953 = 1.2144$$

Como la densidad dato (1.215) está entre estos dos valores obtenidos, la temperatura estará también entre 10 °C y 30 °C; por lo que interpolando linealmente entre estos dos valores de densidad (que ahora es el argumento  $x$ ) se tiene:

**Segundo paso**

Aproximación de la temperatura a la que una solución con 30% de concentración tiene una densidad de 1.215

$$T = \frac{1.215 - 1.2144}{1.2278 - 1.2144} 10 + \frac{1.215 - 1.2278}{1.2144 - 1.2278} 30 \approx 29.1 \text{ } ^\circ\text{C}$$

- 5.4** Elabore un programa para leer una tabla de  $m$  pares de valores e interpolar o extrapolar, utilizando el polinomio de Newton de grado  $n$  en diferencias divididas. Pruebe este programa con los datos del ejercicio 5.1.

**Solución**

En el disco se encuentra el programa 5.3 que lee  $a$ ) el número de pares de valores ( $M$ );  $b$ ) el grado del polinomio interpolante ( $N$ );  $c$ ) el argumento que se desea interpolar ( $XINT$ ), y  $d$ ) los pares de valores ( $X(1), FX(1)$ ), ( $X(2), FX(2)$ ), ..., ( $X(M), FX(M)$ ). Con esta información primero llama al subprograma TABLA que elabora la tabla de diferencias divididas. Con los valores resultantes y el argumento donde se quiere aproximar el valor de la función y el grado del polinomio interpolante, llama al subprograma INTERPOLA que realiza los cálculos de interpolación.

El resultado es

$$\text{PARA } XINT = 1000.0000 \quad FXINT = 23.1201$$

- 5.5** Elabore un programa que lea una tabla de  $m$  (seleccionado por el usuario) pares de valores, y qué interpole o extrapole con el polinomio de Lagrange de orden  $m-1$ .

**Solución**

En el disco se encuentra el programa 5.4, donde se leen  $M$  pares de valores  $X(I)$  y  $FX(I)$  de una tabla y el valor por interpolar  $XINT$ .

Resultado:

$$\text{PARA } XINT = 1000.0000 \quad FXINT = 23.1201$$

- 5.6** Con el programa 5.4 y la tabla de valores del ejercicio 5.1, calcule la presión de vapor del cloruro de magnesio a las siguientes temperaturas:

- |                              |                              |
|------------------------------|------------------------------|
| $a)$ 800 °C (extrapolación)  | $b)$ 950 °C (interpolación)  |
| $c)$ 1098 °C (interpolación) | $d)$ 1500 °C (extrapolación) |

**Solución**

PARA  $XINT = 800.0000$        $FXINT = 18.1702$  con los puntos (0), (1) y (2).  
 PARA  $XINT = 950.0000$        $FXINT = 12.4972$  con los puntos (0), (1) y (2).  
 PARA  $XINT = 1098.0000$        $FXINT = 65.5236$  con los puntos (2), (3) y (4).  
 PARA  $XINT = 1500.0000$        $FXINT = 1156.1016$  con los puntos (5), (6) y (7).

- 5.7** Con la información del ejercicio 5.2 estime el error cometido  $R_2(1.5)$ , aproxime  $f(x)$  en  $x = 1.5$  con un polinomio de tercer grado y estime el error correspondiente  $R_3(1.5)$ .

**Solución**

El valor obtenido con un polinomio de segundo grado (Ejer. 5.2) es:

$$p_2(1.5) = 0.40449$$

Al usar la ecuación 5.41 y los valores de la segunda tabla de diferencias divididas (Ejer. 5.2), se tiene:

$$\begin{aligned} R_2(x) &\approx (x-x_0)(x-x_1)(x-x_2)f[x_0, x_1, x_2, x_3] \\ &\approx (1.5-1.35)(1.5-1.7)(1.5-1.9)(0.10846) = 0.00130 \end{aligned}$$



Para aproximar  $f(x)$  en  $x = 1.5$  con un polinomio de tercer grado se adiciona  $R_2(1.5)$  al valor  $p_2(1.5)$ , se obtiene:

$$p_3(1.5) = 0.40449 + 0.00130 = 0.40579$$

y la estimación del error en esta interpolación es:

$$\begin{aligned} R_3(x) &= (x-x_0)(x-x_1)(x-x_2)(x-x_3)f[x_0, x_1, x_2, x_3, x_4] \\ &= (1.5-1.35)(1.5-1.7)(1.5-1.9)(1.5-1.0)(-0.030567) \\ &= 0.00018 \end{aligned}$$

Hay que observarse que  $R_3(1.5)$  es menor que  $R_2(1.5)$ , por lo que el polinomio de tercer grado da mejor aproximación a esta interpolación que el de segundo grado.

- 5.8** Para calibrar un medidor de orificio se miden la velocidad  $v$  de un fluido, y la caída de presión  $\Delta P$ . Los datos experimentales se dan a continuación y se buscan los mejores parámetros  $a$  y  $b$  de la ecuación que represente estos datos:

$$v = a(\Delta P)^b \quad (1)$$

donde:  $v$  = velocidad promedio (pies/s)

$\Delta P$  = caída de presión (mm Hg)

| $i$          | 1     | 2    | 3    | 4    | 5    | 6     | 7     | 8     | 9     | 10    | 11    |
|--------------|-------|------|------|------|------|-------|-------|-------|-------|-------|-------|
| $v_i$        | 3.83  | 4.17 | 4.97 | 6.06 | 6.71 | 7.17  | 7.51  | 7.98  | 8.67  | 9.39  | 9.89  |
| $\Delta P_i$ | 30.00 | 35.5 | 50.5 | 75.0 | 92.0 | 105.0 | 115.0 | 130.0 | 153.5 | 180.0 | 199.5 |

### Solución



Este problema puede resolverse mediante el método de mínimos cuadrados de la siguiente manera:

Se aplican logaritmos a la ecuación 1 y se tiene:

$$\ln v = \ln a + b \ln (\Delta P) \quad (2)$$

al definir  $y = \ln v$ ;  $a_0 = \ln a$ ;  $a_1 = b$ ;  $x = \ln (\Delta P)$  y sustituir en la ecuación 2 queda:

$$y = a_0 + a_1 x \quad (3)$$

ecuación de una línea recta.

Si se calculan los parámetros  $a_0$  y  $a_1$  de la recta (Ec.3) con el método de mínimos cuadrados, se obtienen (indirectamente) los mejores valores  $a$  y  $b$  que representan los datos experimentales.

Para calcular  $a_0$  y  $a_1$  se construye la siguiente tabla para que los cálculos sean más eficientes (puede usarse una hoja de cálculo electrónica o un pizarrón electrónico).

| Puntos<br>$i$ | $v_i$ | $\Delta P_i$ | $y_i$<br>$\ln v_i$ | $x_i$<br>$\ln \Delta P_i$ | $x_i^2$<br>$(\ln \Delta P_i)^2$ | $y_i x_i$<br>$\ln v_i \ln \Delta P_i$ |
|---------------|-------|--------------|--------------------|---------------------------|---------------------------------|---------------------------------------|
| 1             | 3.83  | 30.0         | 1.34286            | 3.40120                   | 11.56816                        | 4.56734                               |
| 2             | 4.17  | 35.5         | 1.42792            | 3.56953                   | 12.74154                        | 5.09700                               |
| 3             | 4.97  | 50.5         | 1.60342            | 3.92197                   | 15.38185                        | 6.28857                               |
| 4             | 6.06  | 75.0         | 1.80171            | 4.31749                   | 18.64072                        | 7.77886                               |
| 5             | 6.71  | 92.0         | 1.90360            | 4.52179                   | 20.44658                        | 8.60768                               |
| 6             | 7.17  | 105.0        | 1.96991            | 4.65396                   | 21.65934                        | 9.16788                               |
| 7             | 7.51  | 115.0        | 2.01624            | 4.74493                   | 22.51436                        | 9.56692                               |
| 8             | 7.98  | 130.0        | 2.07694            | 4.86753                   | 23.69285                        | 10.10957                              |
| 9             | 8.67  | 153.5        | 2.15987            | 5.03370                   | 25.33814                        | 10.87214                              |
| 10            | 9.39  | 180.0        | 2.23965            | 5.19296                   | 26.96683                        | 11.63041                              |
| 11            | 9.89  | 199.5        | 2.29581            | 5.29581                   | 28.04560                        | 12.13545                              |
| Totales       |       |              | 20.83364           | 49.52087                  | 226.99598                       | 95.82182                              |

Los valores de las sumatorias se sustituyen en el sistema de ecuaciones 5.62 y se tiene:

$$a_0 = \frac{\begin{vmatrix} 20.83364 & 49.52087 \\ 95.82182 & 226.99598 \end{vmatrix}}{\begin{vmatrix} 11.0 & 49.52087 \\ 49.52087 & 226.99598 \end{vmatrix}} = -0.35904$$

$$a_1 = \frac{\begin{vmatrix} 11.0 & 20.83364 \\ 49.52087 & 95.82182 \end{vmatrix}}{\begin{vmatrix} 11.0 & 49.52087 \\ 49.52087 & 226.99598 \end{vmatrix}} = 0.50046$$

Ecuación resultante:

$$y = -0.35904 + 0.50046 x$$

De donde:

$$\ln a = -0.35904 \quad y \quad a = 0.69835$$

$$b = 0.50046$$

Con estos valores, la ecuación que representa los datos experimentales queda:

$$v = 0.69835 (\Delta P)^{0.50046}$$

Para este ejercicio recomendamos ver el **PROGRAMA 5.6** del CD.

- 5.9** Al medir la velocidad (con un tubo de Pitot) en una tubería circular de diámetro interior de 20 cm, se encontró la siguiente información:

|            |     |     |     |     |     |
|------------|-----|-----|-----|-----|-----|
| $v$ (cm/s) | 600 | 550 | 450 | 312 | 240 |
| $r$ (cm)   | 0   | 3   | 5   | 7   | 8   |

donde  $r$  es la distancia en cm medida a partir del centro del tubo.

- Obtenga la curva  $v = f(r)$  que aproxima estos datos experimentales.
- Calcule la velocidad en el punto  $r = 4$  cm.

## Solución



- a) Se asume que en la experimentación hay errores, de tal modo que se justifica usar una aproximación por mínimos cuadrados. Por otro lado se sabe que el perfil de velocidades en una tubería generalmente es de tipo parabólico, por lo que se ensayará un polinomio de segundo grado

$$v(r) = a_0 + a_1 r + a_2 r^2$$

Al construir la tabla que proporcione los coeficientes del sistema de ecuaciones 5.64 se tiene

| Puntos  | $v$  | $r$ | $r^2$ | $r^3$ | $r^4$ | $vr$ | $vr^2$ |
|---------|------|-----|-------|-------|-------|------|--------|
| 1       | 600  | 0   | 0     | 0     | 0     | 0    | 0      |
| 2       | 550  | 3   | 9     | 27    | 81    | 1650 | 4950   |
| 3       | 450  | 5   | 25    | 125   | 625   | 2250 | 11250  |
| 4       | 312  | 7   | 49    | 343   | 2401  | 2184 | 15288  |
| 5       | 240  | 8   | 64    | 512   | 4096  | 1920 | 15360  |
| Totales | 2152 | 23  | 147   | 1007  | 7203  | 8004 | 46848  |

Estos valores se sustituyen en el sistema de ecuaciones 5.64 particularizado para un polinomio de segundo grado y se tiene:

$$\begin{aligned} 5 a_0 + 23 a_1 + 147 a_2 &= 2152 \\ 23 a_0 + 147 a_1 + 1007 a_2 &= 8004 \\ 147 a_0 + 1007 a_1 + 7203 a_2 &= 46848 \end{aligned}$$

Al resolver para los parámetros  $a_0$ ,  $a_1$  y  $a_2$  y sustituirlos en el polinomio propuesto queda:

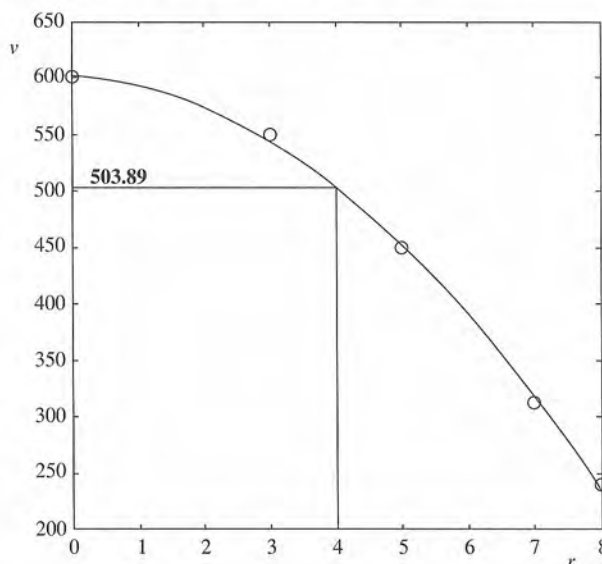
$$v(r) = 601.714 - 3.0667 r - 5.347 r^2$$

- b) Con la sustitución  $r = 4$ , se obtiene:

$$v(4) = 503.89 \text{ cm/s}$$

Hay que observar que la distribución de velocidades sólo se presenta del centro a la pared del tubo, ya que es simétrica.

SUGERENCIA: Vea el PROGRAMA 5.7 del CD.



Distribución de la velocidad del centro a la pared de tubo.





Se continúa adicionando las filas necesarias:  $x_3, x_4, x_5, x_1^2, x_2^2, x_1, x_2, \dots$  y sumando los totales de cada una para conseguir los coeficientes y el vector de términos independientes del sistema. Dichos cálculos dan como resultado el siguiente sistema de ecuaciones:

$$\begin{bmatrix} 18 & 1800 & 360 & 187500 & 7500 & 36000 \\ 1800 & 187500 & 36000 & 20250000 & 750000 & 3750000 \\ 360 & 36000 & 7500 & 3750000 & 162000 & 750000 \\ 187500 & 20250000 & 3750000 & 2254687500 & 78125000 & 405000000 \\ 7500 & 750000 & 162000 & 78125000 & 3607500 & 16200000 \\ 36000 & 3750000 & 750000 & 405000000 & 16200000 & 78125000 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{bmatrix} = \begin{bmatrix} 251.94 \\ 24170.0 \\ 5287.9 \\ 2420850.5 \\ 115143.0 \\ 508702.5 \end{bmatrix}$$

cuya solución por medio de alguno de los métodos del capítulo 3 es:

$$\begin{aligned} a_0 &= 56.4264, & a_1 &= -0.362597, & a_2 &= -2.74767 \\ a_3 &= 0.00081632, & a_4 &= 0.0816, & a_5 &= 0.00314 \end{aligned}$$

se sustituyen en el modelo y resulta:

$$y = 56.4264 - 0.362597x_1 - 2.74767x_2 + 0.00081632x_1^2 + 0.0816x_2^2 + 0.00314x_1x_2$$

Una vez obtenidos los coeficientes, puede estimarse el porcentaje de impurezas correspondiente a un tiempo de esterilización y una temperatura dados; por ejemplo, a un tiempo de 19 min y una temperatura de 80 °C se tiene un porcentaje de impurezas de:

$$\begin{aligned} y &= 56.4264 - 0.362597(80) - 2.74767(19) + 0.00081632(80)^2 + \\ &\quad 0.0816(19)^2 + 0.00314(80)(19) = 14.67 \end{aligned}$$

## Problemas

- 5.1 La densidad del carbonato neutro de potasio en solución acuosa varía con la temperatura y la concentración de acuerdo con la tabla siguiente.

| c (%) | T (°C) |        |        |        |
|-------|--------|--------|--------|--------|
|       | 0      | 40     | 80     | 100    |
| 4     | 1.0381 | 1.0276 | 1.0063 | 0.9931 |
| 12    | 1.1160 | 1.1013 | 1.0786 | 1.0663 |
| 20    | 1.1977 | 1.1801 | 1.1570 | 1.1451 |
| 28    | 1.2846 | 1.2652 | 1.2418 | 1.2301 |

- Calcule la densidad a 40°C y 15% de concentración.
- Calcule la densidad a 50°C y 28% de concentración.
- Calcule la densidad a 90°C y 25% de concentración.
- Calcule la concentración que tiene una solución de densidad 1.129 a una temperatura de 60 °C.

Utilice interpolaciones cuadráticas en todos los incisos.

- 5.2 Los datos de presión-temperatura-volumen para el etano se muestran en la tabla siguiente, donde la temperatura ( $T$ ) está en °C, la presión ( $P$ ) en atmósferas y el volumen específico ( $1/V$ ) en moles/litro.

| $T$ | $P$   |       |        |        |        |        |        |
|-----|-------|-------|--------|--------|--------|--------|--------|
|     | 1     | 2     | 4      | 6      | 8      | 9      | 10     |
| 25  | 20.14 | 32.84 | —      | —      | —      | —      | —      |
| 75  | 24.95 | 43.80 | 68.89  | 85.95  | 104.38 | 118.32 | 139.23 |
| 150 | 31.89 | 59.31 | 106.06 | 151.38 | 207.66 | 246.57 | 298.02 |
| 200 | 36.44 | 69.38 | 130.18 | 194.53 | 276.76 | 332.56 | —      |
| 250 | 40.87 | 79.16 | 153.59 | 237.38 | 345.38 | —      | —      |

Calcule el volumen específico en moles/litro para una presión de 7 atmósferas y una temperatura de 175 °C.

5.3 Datos:

| Puntos | 0        | 1        | 2        |
|--------|----------|----------|----------|
| $x$    | $x_0$    | $x_1$    | $x_2$    |
| $f(x)$ | $f(x_0)$ | $f(x_1)$ | $f(x_2)$ |

- Encuentre los coeficientes  $a_0, a_1, a_2$ , del polinomio de segundo grado que pasa por estos tres puntos, por el métodos de Lagrange.
- Realice el mismo proceso que en a) pero ahora empleando el método de aproximación polinomial simple.
- Demuestre que los polinomios en los incisos a) y b) son el mismo, pero escrito en diferente forma.

5.4 Dada una función  $y = f(x)$  en forma tabular, a menudo se desea encontrar un valor  $x$  correspondiente a un valor dado de  $y$ ; este proceso, llamado interpolación inversa, se lleva a cabo en la forma ya vista, pero intercambiando los papeles de  $x$  y  $y$ . Dada la siguiente tabla.

| Puntos | 0     | 1    | 2    | 3    | 4    | 5    | 6    |
|--------|-------|------|------|------|------|------|------|
| $x$    | 0.0   | 2.5  | 5.0  | 7.5  | 10.0 | 12.5 | 15.0 |
| $y$    | 10.00 | 4.97 | 2.47 | 1.22 | 0.61 | 0.30 | 0.14 |

donde  $y$  es la amplitud de la oscilación de un péndulo largo, en cm y  $x$  es el tiempo medido en min desde que empezó la oscilación.

Encuentre el polinomio de aproximación de Lagrange de segundo grado que pasa por los puntos (1), (2) y (3) y el valor de  $x$  correspondiente a  $y = 2$  cm.

5.5 Sea  $z(x) = \prod_{j=0}^n (x - x_j)$ . Demuestre que el polinomio 5.22 puede escribirse en la forma

$$p_n(x) = z(x) \sum_{i=0}^n \frac{f(x_i)}{(x - x_j) z'(x_i)}$$

5.6 Use las ideas dadas en el problema anterior para demostrar que

$$\sum_{i=0}^n L_i(x) \approx 1 \text{ para toda } x.$$



**SUGERENCIA:** Considere que la expresión dada corresponde al polinomio de aproximación por polinomios de Lagrange de  $f(x) = 1$  (un polinomio de grado cero).

- 5.7 Demuestre que el polinomio de aproximación de Lagrange de primer grado puede escribirse en notación de determinantes así

$$p_{0,1}(x) = \frac{1}{x_1 - x_0} \begin{vmatrix} p_0(x) & (x_0 - x) \\ p_1(x) & (x_1 - x) \end{vmatrix}$$

donde  $p_0(x) = f(x_0)$  y  $p_1(x) = f(x_1)$  y los subíndices 0 y 1 de  $p(x)$  se refieren a los puntos (0) y (1) por donde pasa el polinomio de aproximación.

Demuestre también que para el caso del polinomio de aproximación de Lagrange de segundo grado que pasa por los puntos (0), (1) y (2).

$$p_{0,1,2}(x) = \frac{1}{x_2 - x_1} \begin{vmatrix} p_{0,1}(x) & (x_1 - x) \\ p_{0,2}(x) & (x_2 - x) \end{vmatrix}$$

- 5.8 Lo demostrado en el problema anterior es válido, en general, para aproximaciones de tercero, cuarto, ...,  $n$  grado. Aitken desarrolló un método para interpolar con este tipo de polinomios y consiste en construir la tabla siguiente:

|       |       |           |             |               |                 |  |             |
|-------|-------|-----------|-------------|---------------|-----------------|--|-------------|
| $x_0$ | $p_0$ |           |             |               |                 |  | $(x_0 - x)$ |
| $x_1$ | $p_1$ | $p_{0,1}$ |             |               |                 |  | $(x_1 - x)$ |
| $x_2$ | $p_2$ | $p_{0,2}$ | $p_{0,1,2}$ |               |                 |  | $(x_2 - x)$ |
| $x_3$ | $p_3$ | $p_{0,3}$ | $p_{0,1,3}$ | $p_{0,1,2,3}$ |                 |  | $(x_3 - x)$ |
| $x_4$ | $p_4$ | $p_{0,4}$ | $p_{0,1,4}$ | $p_{0,1,2,4}$ | $p_{0,1,2,3,4}$ |  | $(x_4 - x)$ |

donde  $p_i = f(x_i)$  y  $x$  el valor donde se desea interpolar.

Para el cálculo de

$$p_{0,i}(x) = \frac{1}{x_i - x_0} \begin{vmatrix} p_0 & (x_0 - x) \\ p_i & (x_i - x) \end{vmatrix}$$

donde el denominador resulta ser  $(x_i - x) - (x_0 - x)$ .

En cambio para  $p_{0,1,i}$  se usa

$$p_{0,1,i}(x) = \frac{1}{x_i - x_1} \begin{vmatrix} p_{0,1} & (x_1 - x) \\ p_{0,i} & (x_i - x) \end{vmatrix}$$

cuyo denominador es  $(x_i - x) - (x_1 - x)$ .

Se aconseja denotar la abscisa más cercana a  $x$  como  $x_0$ , la segunda más próxima a  $x$  como  $x_1$ , y así sucesivamente.

Con ese ordenamiento los valores  $p_{0,1}$ ,  $p_{0,1,2}$ ,  $p_{0,1,2,3}$ , etc., representan la mejor aproximación al valor buscado  $f(x)$  con polinomios de primero, segundo, tercero, ...,  $n$  grado.

Con el método descrito, aproxime el valor de la función de Bessel ( $J_0$ ) dada abajo en  $x = 0.8$

| Puntos   | 0      | 1      | 2      | 3      |
|----------|--------|--------|--------|--------|
| $x$      | 0.5    | 0.7    | 0.9    | 1.0    |
| $J_0(x)$ | 0.9385 | 0.8812 | 0.8075 | 0.7652 |

- 5.9 En el método de posición falsa (capítulo 2) se realiza una interpolación inversa: dados los puntos  $(x_P, f(x_P))$  y  $(x_D, f(x_D))$  se encuentra el polinomio  $p(x)$  que pasa por esos puntos y luego el valor de  $x$  correspondiente a  $p(x) = 0$ . Discuta la interpolación inversa para encontrar raíces de ecuaciones no lineales empleando tres puntos.
- 5.10 Demuestre que si la función  $f(x)$  dada en forma tabular corresponde a un polinomio de grado  $n$ , entonces el polinomio de aproximación  $p(x)$  de grado mayor o igual a  $n$  que pasa por los puntos de la tabla es  $f(x)$  misma.

**SUGERENCIA:** Con el polinomio  $y = 2x + 3$  forme un tabla de valores y tomando dos de esos valores encuentre  $p(x)$  y observe que  $p(x) = y$ ; después tomando 3 valores cualesquiera observe que el  $p(x)$  obtenido es nuevamente  $y = 2x + 3$ . Sólo resta generalizar estos resultados.

- 5.11 Desarrolle algebraicamente el numerador y el denominador de

$$a_2 = \frac{f[x_2] - f[x_0] - (x_2 - x_0) \frac{f[x_1] - f[x_0]}{x_1 - x_0}}{(x_2 - x_0)(x_2 - x_1)}$$

para llegar a

$$a_2 = \frac{\frac{f[x_2] - f[x_1]}{x_2 - x_1} - \frac{f[x_1] - f[x_0]}{x_1 - x_0}}{x_2 - x_0} = f[x_0, x_1, x_2]$$

- 5.12 Verifique que para tres puntos distintos cualesquiera de abscisas  $x_0, x_1$  y  $x_2$  se cumple que

$$f[x_0, x_1, x_2] = f[x_2, x_0, x_1] = f[x_1, x_2, x_0]$$

así como con cualquier otra permutación de  $x_1, x_2, x_0$ . Esta propiedad de las diferencias de segundo orden es conocida como **simetría respecto a los argumentos** y la cumplen también las diferencias de primer orden (trivial), las de orden 3, etcétera.

- 5.13 Elabore un subprograma de propósito general para construir la tabla de diferencias divididas de una función tabulada.

**SUGERENCIA:** Vea el algoritmo 5.3. Puede usar una hoja de cálculo electrónica.

- 5.14 Para los valores siguientes

| Puntos | 0    | 1    | 2    | 3    | 4    | 5    | 6    |
|--------|------|------|------|------|------|------|------|
| $e$    | 40   | 60   | 80   | 100  | 120  | 140  | 160  |
| $p$    | 0.63 | 1.36 | 2.18 | 3.00 | 3.93 | 6.22 | 8.59 |

donde  $e$  son los volts y  $p$  los kilowatts en una curva de pérdida en el núcleo para un motor eléctrico:

- Elabore una tabla de diferencias divididas.
- Con el polinomio de Newton en diferencias divididas de segundo grado, aproxime el valor de  $p$  correspondiente a  $e = 90$  volts.

- 5.15 En la tabla siguiente:

| $i$ | 1   | 2  | 3  | 4  |
|-----|-----|----|----|----|
| $v$ | 120 | 94 | 75 | 62 |

donde  $i$  es la corriente y  $v$  el voltaje consumido por un arco magnético, aproxime el valor de  $v$  para  $i = 3.5$  por un polinomio de Newton en diferencias divididas y compare con el valor dado por la fórmula empírica.

$$v = 30.4 + 90.4 i^{-0.507}$$

- 5.16** Corrobore que el polinomio de Newton en diferencias divididas puede escribirse en términos de  $\prod$ , así:

$$p_n(x) = p_n(x_0 + sh) = \sum_{k=0}^n \Delta^k f(x_0) \prod_{i=0}^{k-1} \frac{s-i}{i+1} \quad (1)$$

**Nota:** Considere que  $\Delta^0 f(x_0) = f(x_0)$ . Esta notación es generalmente más útil para programar este algoritmo.

- 5.17** Con los resultados del problema anterior y con la definición de función binomial siguiente, exprese la ecuación (1) en términos de  $\binom{s}{k}$ .

$$\binom{s}{k} = \begin{cases} 1 & k = 0 \\ \prod_{i=0}^{k-1} \frac{s-i}{i+1} = \frac{s(s-1)(s-2)\dots(s-(k-1))}{1(2)(3)\dots(k)} & k > 0 \end{cases}$$

- 5.18** Con los siguientes valores.

| Puntos | 0      | 1     | 2     | 3     |
|--------|--------|-------|-------|-------|
| $l/r$  | 140    | 180   | 220   | 240   |
| $p/a$  | 12,800 | 7,500 | 5,000 | 3,800 |

donde  $p/a$  es la carga en lb/pulg<sup>2</sup> que causa la ruptura de una columna de hierro dulce con extremos redondeados y  $l/r$  es la razón de la longitud de la columna al mínimo radio de giro de su sección transversal.

Encuentre el polinomio de tercer grado que pasa por estos puntos en sus distintas formas:

- $p_3(x) = a_0 + a_1x + a_2x^2 + a_3x^3$  (aproximación polinomial simple).
  - Forma de Lagrange.
  - Aproximación de Newton (en diferencias divididas).
  - Aproximación de Newton en diferencias finitas (hacia delante y hacia atrás).
- 5.19** En una reacción química, la concentración del producto  $C_B$  cambia con el tiempo como se indica en la tabla de abajo. Calcule la concentración  $C_B$  cuando  $t = 0.82$ , usando un polinomio de Newton en diferencias finitas.

|       |      |      |      |      |      |      |
|-------|------|------|------|------|------|------|
| $C_B$ | 0.00 | 0.30 | 0.55 | .80  | 1.10 | 1.15 |
| $t$   | 0.00 | 0.10 | 0.40 | 0.60 | 0.80 | 1.00 |

- 5.20** Resuelve el problema 5.13, empleando diferencias finitas; compare los cálculos realizados y los resultados obtenidos en ambos problemas.
- 5.21** Elabore un diagrama de flujo y codifíquelo para leer  $n$  pares de valores  $x$  y  $f(x)$ , calcular e imprimir la tabla de diferencias finitas hacia atrás.



| Puntos | $x_i$     | $f[x_i]$     | $\nabla f[x_i]$ | $\nabla^2 f[x_i]$   | $\nabla^3 f[x_i]$     | ... $\nabla^{n-2} f[x_i]$     |
|--------|-----------|--------------|-----------------|---------------------|-----------------------|-------------------------------|
| 0      | $x_0$     | $f[x_0]$     |                 |                     |                       |                               |
|        |           |              | $\nabla f[x_1]$ |                     |                       |                               |
| 1      | $x_1$     | $f[x_1]$     |                 | $\nabla^2 f[x_1]$   |                       |                               |
|        |           |              | $\nabla f[x_2]$ |                     | $\nabla^3 f[x_1]$     |                               |
| 2      | $x_2$     | $f[x_2]$     |                 | $\nabla^2 f[x_2]$   |                       |                               |
|        |           |              | $\nabla f[x_3]$ |                     | $\nabla^3 f[x_2]$     |                               |
| 3      | $x_3$     | $f[x_3]$     |                 | $\nabla^2 f[x_3]$   |                       |                               |
|        |           |              | .               | .                   | .                     | ... $\nabla^{n-2} f[x_{n-1}]$ |
|        | .         | .            | .               | .                   | .                     | .                             |
|        | .         | .            | .               | .                   | $\nabla^2 f[x_{n-1}]$ |                               |
|        | .         | .            |                 | $\nabla f[x_{n-1}]$ |                       |                               |
| $n-1$  | $x_{n-1}$ | $f[x_{n-1}]$ |                 |                     |                       |                               |

$$\nabla f[x] = f[x] - f[x-h]$$

$$\nabla^m f[x] = \nabla^{m-1} f[x] - \nabla^{m-1} f[x-h]$$

- 5.22** En el caso en que la distancia  $h$  entre dos argumentos consecutivos cualesquiera es la misma a lo largo de la tabla, puede usarse la ecuación 5.35 para interpolar en puntos cercanos a  $x_0$  o bien 5.38 cuando se quiere interpolar en puntos al final de la tabla (véase Sec. 5.5). Si hay que interpolar en puntos centrales de la tabla, resulta conveniente denotar alguno de dichos puntos centrales como  $x_0$ , como  $x_1, x_2, x_3, \dots$ , las abscisas mayores que  $x_0$  y como  $x_{-1}, x_{-2}, x_{-3}, \dots$ , las abscisas menores que  $x_0$ . En estas condiciones e introduciendo el operador lineal  $\delta$ , conocido como **operador en diferencias centrales** y definido sobre  $f(x)$  como:

$$\delta f(x) = f(x + h/2) - f(x - h/2) \quad (1)$$

y cuya aplicación sucesiva conduce a:

$$\delta(\delta f(x)) = \delta^2 f(x) = f(x + h) - 2f(x) + f(x - h)$$

y en general a:

$$\delta^i f(x) = \delta(\delta^{i-1} f(x)) \quad (2)$$

Nótese que  $\delta f(x_0)$  no emplea, en general, los valores de la tabla, la cual constituye una dificultad para su uso. En cambio, la segunda diferencia central

$$\delta^2 f(x_k) = f(x_k + h) - 2f(x_k) + f(x_k - h)$$

incluye sólo valores funcionales tabulados; esto es cierto para todas las diferencias centrales de orden par. A fin de evitar que se requieran valores funcionales no tabulados en la primera diferencia central, puede aplicarse  $\delta$  a puntos no tabulados; por ejemplo,  $f(x_k + h/2)$  con la cual queda:

$$\delta f(x_k + h/2) = f(x_k + h) - f(x_k) = f(x_{k+1}) - f(x_k),$$

donde ya sólo aparecen valores funcionales de la tabla.

En general  $\delta^{2i+1} f(x_k + h/2)$  (orden impar) queda en función de ordenadas presentes en la tabla.

Con la notación de diferencias divididas se tiene que:

$$\begin{aligned}
\delta f(x_0 + h/2) &= f(x_1) - f(x_0) = h f[x_0, x_1] \\
\delta f(x_0 - h/2) &= f(x_0) - f(x_{-1}) = h f[x_0, x_{-1}] \\
\delta^2 f(x_1) &= \delta f(x_1 + h/2) - \delta f(x_1 - h/2) = h f[x_1, x_2] - h f[x_0, x_1] \\
&= 2! h^2 f[x_0, x_1, x_2]
\end{aligned}$$

y en general:

$$\delta^{2i+1} f(x_k + h/2) = h^{2i+1} (2i+1)! f[x_{k-i}, \dots, x_k, \dots, x_{k+i}, x_{k+i+1}] \quad (3)$$

$$\delta^{2i+1} f(x_k - h/2) = h^{2i+1} (2i+1)! f[x_{k-i-1}, x_{k-i}, \dots, x_k, \dots, x_{k+i}] \quad (4)$$

para orden impar y

$$\delta^{2i} f(x_k) = h^{2i} (2i)! f[x_{k-i}, \dots, x_k, \dots, x_{k+i}] \quad (5)$$

para orden par.

La tabla de diferencias centrales queda entonces

|          |             |                          |                      |                            |         |
|----------|-------------|--------------------------|----------------------|----------------------------|---------|
|          |             |                          |                      |                            |         |
|          |             |                          |                      |                            |         |
|          |             |                          |                      |                            |         |
| $x_{-2}$ | $f(x_{-2})$ |                          |                      |                            |         |
|          |             | $\delta f(x_{-2} + h/2)$ |                      |                            |         |
| $x_{-1}$ | $f(x_{-1})$ |                          | $\delta^2 f(x_{-1})$ |                            |         |
|          |             | $\delta f(x_{-1} + h/2)$ |                      | $\delta^3 f(x_{-1} + h/2)$ |         |
| $x_0$    | $f(x_0)$    |                          | $\delta^2 f(x_0)$    |                            | $\dots$ |
|          |             | $\delta f(x_0 + h/2)$    |                      | $\delta^3 f(x_0 + h/2)$    |         |
| $x_1$    | $f(x_1)$    |                          | $\delta^2 f(x_1)$    |                            |         |
|          |             | $\delta f(x_1 + h/2)$    |                      |                            |         |
| $x_2$    | $f(x_2)$    |                          |                      |                            |         |
|          |             |                          |                      |                            |         |
|          |             |                          |                      |                            |         |
|          |             |                          |                      |                            |         |

Note que el argumento permanece constante en cualquier línea horizontal de la tabla. Con esta notación y la aplicación sucesiva de la ecuaciones 3 y 5 con  $k=0$ , la 5.29 se transforma en:

$$\begin{aligned}
f(x) &= f(x_0) + (x - x_0) \frac{\delta f(x_0 + h/2)}{1!h} + (x - x_0)(x - x_1) \frac{\delta^2 f(x_0)}{2!h^2} + \\
&\quad (x - x_0)(x - x_1)(x - x_{-1}) \frac{\delta^3 f(x_0 + h/2)}{3!h^3} + \dots
\end{aligned} \quad (6)$$

Al emplear el cambio de variable:

$$x = x_0 + sh,$$

en donde:

$$s = \frac{x - x_0}{h}$$

el polinomio (6) queda

$$\begin{aligned}
 p_n(x_0 + sh) = & f(x_0) + s \delta f(x_0 + h/2) + \frac{s(s-1)}{2!} \delta^2 f(x_0) + \\
 & \frac{s(s^2-1^2)}{3!} \delta^3 f(x_0 + h/2) + \frac{s(s^2-1^2)(s-2)}{4!} \delta^4 f(x_0) + \\
 & \dots + \frac{s(s^2-1^2) \dots (s^2-(i-1)^2)}{(2i)!} \delta^{2i} f(x_0) (s-i)
 \end{aligned} \quad (7)$$

cuando el grado del polinomio es par; si es impar, el último término de la ecuación 7 queda como:

$$\dots + \frac{s(s^2-1^2) \dots (s^2-i^2)}{(2i+1)!} \delta^{2i+1} f(x_0 + h/2)$$

Este polinomio se conoce como la **fórmula hacia delante de Gauss**.

Con la tabla del ejemplo 5.7 construya una tabla de diferencias centrales y mediante la ecuación 7 encuentre por interpolación la presión correspondiente a una temperatura de 76 °F.

- 5.23** Si aproxima la función dada abajo por un polinomio de segundo grado y con éste interpola en  $x = 10$ , estime el error cometido en esta interpolación.

| Puntos | 0     | 1     | 2     | 3     | 4     | 5     | 6     |
|--------|-------|-------|-------|-------|-------|-------|-------|
| $x$    | 0     | 1     | 6     | 8     | 11.5  | 15    | 19    |
| $f(x)$ | 38000 | 38500 | 35500 | 27500 | 19000 | 15700 | 11000 |

- 5.24** Demuestre que el término del error para la aproximación polinomial de segundo grado es:

$$R_2(x) = (x-x_0)(x-x_1)(x-x_2)f[x, x_0, x_1, x_2]$$

- 5.25** Encuentre una cota inferior y una cota superior del error de interpolación  $R_3(x)$  en  $x = 6.3$  para la función  $f(x) = e^x$  dada en los puntos  $x_0 = 5$ ,  $x_1 = 6$ ,  $x_2 = 7$ ,  $x_3 = 8$  (véase ejemplo 5.11)
- 5.26** Demuestre que la función dada por  $z(x) = |(x-x_0)(x-x_1)|$  con  $x_0 \leq x \leq x_1$  alcanza su valor máximo en  $(x_0 + x_1)/2$  y está dado por  $(x_1 - x_0)^2/4$ .
- 5.27** Con lo resultados del problema anterior y la fórmula

$$R_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^n (x - x_i),$$

demuestre que el error  $R_1(x)$  con  $x_0 \leq x \leq x_1$  correspondiente a una aproximación lineal de  $f(x)$  usando como argumento  $x_0$  y  $x_1$  es menor en magnitud (valor absoluto) que  $M(x_1 - x_0)^2/8$ , donde  $M$  es el valor máximo de  $|f''(x)|$  en  $[x_0, x_1]$ .

- 5.28** Los siguientes valores fueron obtenidos de una tabla de distribución binomial.

| $b$ |     | $p$    |        |        |        |        |
|-----|-----|--------|--------|--------|--------|--------|
| $n$ | $x$ | 0.0500 | 0.1000 | 0.1500 | 0.2000 | 0.2500 |
| 3   | 0   | 0.8574 | 0.7290 | 0.6141 | 0.5120 | 0.4219 |
|     | 1   | 0.1354 | 0.2430 | 0.3251 | 0.3840 | 0.4219 |
|     | 2   | 0.0071 | 0.0270 | 0.0574 | 0.0960 | 0.1406 |
|     | 3   | 0.0001 | 0.0010 | 0.0034 | 0.0080 | 0.0156 |



Al pie de dicha tabal se lee “la interpolación lineal dará valores exactos de  $b$  a lo más en dos cifras decimales”.

Encuentre una aproximación de  $f(x; n, p) = f(1; 3, 0.13)$  exacta en tres cifras decimales. Recuerde que:

$$b(x; n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

¿Cree usted que si los valores de la tabla son exactos en las cuatro cifras decimales dadas, pueda obtenerse exactitud con cuatro cifras decimales aplicando el método de interpolación?

- 5.29** En la siguiente tabla,  $r$  es la resistencia de una bobina en ohms y  $T$  la temperatura de la bobina en  $^{\circ}\text{C}$ . Por mínimos cuadrados determine el mejor polinomio lineal que representa la función dada.

|     |        |        |        |        |        |        |
|-----|--------|--------|--------|--------|--------|--------|
| $r$ | 10.421 | 10.939 | 11.321 | 11.794 | 12.242 | 12.668 |
| $T$ | 10.50  | 29.49  | 42.70  | 60.01  | 75.51  | 91.05  |

- 5.30** En la tabla

|        |       |       |       |       |       |       |       |       |       |
|--------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| Puntos | 0     | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
| $v$    | 26.43 | 22.40 | 19.08 | 16.32 | 14.04 | 12.12 | 10.51 | 9.15  | 8.00  |
| $P$    | 14.70 | 17.53 | 20.80 | 24.54 | 28.83 | 33.71 | 39.25 | 45.49 | 52.52 |

$v$  es el volumen en  $\text{pie}^3$  de una lb de vapor y  $P$  es la presión en psia. Encuentre los parámetros  $a$  y  $b$  de la ecuación

$$P = a v^b$$

aplicando el método de mínimos cuadrados.

- 5.31** Se sabe que el número de pulgadas de una estructura recién construida que se hunde en el suelo está dada por:

$$y = 3 - 3 e^{-ax}$$

donde  $x$  es el número de meses que lleva construida la estructura. Con los valores

|     |      |      |      |      |      |      |
|-----|------|------|------|------|------|------|
| $x$ | 2    | 4    | 6    | 12   | 18   | 24   |
| $y$ | 1.07 | 1.88 | 2.26 | 2.78 | 2.97 | 2.99 |

estime  $a$ , usando el criterio de los mínimos cuadrados (véase ejercicio 5.8).

- 5.32** En el estudio de la constante de velocidad  $k$  de una reacción química a diferentes temperaturas, se obtuvieron los datos:

|         |                       |                       |                       |      |        |        |        |
|---------|-----------------------|-----------------------|-----------------------|------|--------|--------|--------|
| $T$ (K) | 293                   | 300                   | 320                   | 340  | 360    | 380    | 400    |
| $k$     | $8.53 \times 10^{-5}$ | $19.1 \times 10^{-5}$ | $1.56 \times 10^{-3}$ | 0.01 | 0.0522 | 0.2284 | 0.8631 |

Calcule el factor de frecuencia  $z$  y la energía de activación  $E$ , asumiendo que los datos experimentales siguen la ley de Arrhenius:

$$k = z e^{-E/1.98T}$$

- 5.33** Sieder y Tate\* encontraron que una ecuación que relaciona la transferencia de calor de líquidos por dentro de tubos en cambiadores de calor, se puede representar con números adimensionales.

$$Nu = a (Re)^b (Pr)^c \left(\frac{\mu}{\mu_w}\right)^d$$

Donde  $Nu$  es el número de Nusselt,  $Re$  es el número de Reynolds,  $Pr$  el número de Prandtl y  $\mu$  y  $\mu_w$  las viscosidades del líquido a la temperatura promedio de éste y a la temperatura de la pared del tubo, respectivamente.

Encuentre los valores de  $a$ ,  $b$ ,  $c$  y  $d$  asumiendo que la tabla siguiente representa datos experimentales para un grupo de hidrocarburos a diferentes condiciones de operación.

|             |       |        |        |        |        |        |        |        |
|-------------|-------|--------|--------|--------|--------|--------|--------|--------|
| $Nu$        | 97.45 | 109.50 | 129.90 | 147.76 | 153.44 | 168.90 | 177.65 | 175.16 |
| $Re$        | 10500 | 12345  | 15220  | 18300  | 21050  | 25310  | 28560  | 31500  |
| $Pr$        | 18.2  | 17.1   | 16.8   | 15.3   | 12.1   | 10.1   | 8.7    | 6.5    |
| $\mu/\mu_w$ | 0.85  | 0.90   | 0.96   | 1.05   | 1.08   | 1.15   | 1.18   | 1.22   |

- 5.34** Elabore un programa de propósito general, para aproximar una función dada en forma tabular por un polinomio de grado  $n$  usando el método de mínimos cuadrados.
- 5.35** En una reacción gaseosa de expansión a volumen constante, se observa que la presión del reactor (*batch*) aumenta con el tiempo de reacción según se muestra en la tabla de abajo ¿Qué grado de polinomio (con el criterio de ajuste exacto) aproxima mejor la función  $P = f(t)$ ?

|           |        |        |        |        |        |        |        |
|-----------|--------|--------|--------|--------|--------|--------|--------|
| $P$ (atm) | 1.0000 | 1.0631 | 1.2097 | 1.3875 | 1.7232 | 2.0000 | 2.9100 |
| $t$ (min) | 0.0    | 0.1    | 0.3    | 0.5    | 0.8    | 1.0    | 1.5    |

- 5.36** La aparición de una corriente inducida en un circuito que tiene la constante de tiempo  $\tau$  está dada por:

$$I = 1 - e^{-t/\tau}$$

donde  $t$  es el tiempo medio desde el instante en que el interruptor se cierra e  $I$  la razón de la corriente en tiempo  $t$  al valor total de la corriente dado por la ley de Ohm. Con las mediciones siguientes, estime la constante de tiempo  $\tau$  de este circuito (consulte el Ej. 5.8).

|           |       |       |       |       |       |       |       |
|-----------|-------|-------|-------|-------|-------|-------|-------|
| $I$       | 0.073 | 0.220 | 0.301 | 0.370 | 0.418 | 0.467 | 0.517 |
| $t$ (seg) | 0.1   | 0.2   | 0.3   | 0.4   | 0.5   | 0.6   | 0.7   |

\* Sieder y Tate. Ind. and Eng. Chem. 28, 1429 (1936).

## 5.37 Los valores

|     |      |      |      |      |
|-----|------|------|------|------|
| $t$ | 0.0  | 10.0 | 27.4 | 42.1 |
| $s$ | 61.5 | 62.1 | 66.3 | 70.3 |

representan la cantidad  $s$  en  $gr$  de dicromato de potasio disueltos en 100 partes de agua a la temperatura  $t$  indicada en  $^{\circ}C$ . La relación entre estas variables es:

$$\log_{10}s = a + b t + c t^2$$

Calcule los parámetros  $a$ ,  $b$  y  $c$  por el método de mínimos cuadrados.

5.38 Para la tabla de datos que se da abajo, encuentre los parámetros  $a$  y  $b$  de la ecuación

$$y = a + (0.49 - a) e^{-b(x-8)}$$

|     |      |      |      |      |
|-----|------|------|------|------|
| $x$ | 10   | 20   | 30   | 40   |
| $y$ | 0.48 | 0.42 | 0.40 | 0.39 |

## 5.39 Veinte tipos de hojas de acero procesadas en frío tienen diferentes composiciones de cobre y temperaturas de templado. Al medir su dureza resultante se obtuvieron los siguientes valores:

| Dureza Rockwell 30-T | $u$<br>Contenido de cobre<br>% | $v$<br>Temp. de Templado<br>$^{\circ}F$ |
|----------------------|--------------------------------|-----------------------------------------|
| 78.9                 | 0.02                           | 1000                                    |
| 65.1                 | 0.02                           | 1100                                    |
| 55.2                 | 0.02                           | 1200                                    |
| 56.4                 | 0.02                           | 1300                                    |
| 80.9                 | 0.10                           | 1000                                    |
| 69.7                 | 0.10                           | 1100                                    |
| 57.4                 | 0.10                           | 1200                                    |
| 55.4                 | 0.10                           | 1300                                    |
| 85.3                 | 0.18                           | 1000                                    |
| 71.8                 | 0.18                           | 1100                                    |
| 60.7                 | 0.18                           | 1200                                    |
| 58.9                 | 0.18                           | 1300                                    |

Se sabe que la dureza depende en forma lineal del contenido  $u$  de cobre en % y de la temperatura de templado  $v$

$$y = a_0 + a_1 u + a_2 v$$

Determine los parámetros  $a_0$ ,  $a_1$  y  $a_2$ , siguiendo el criterio de los mínimos cuadrados.



# INTEGRACIÓN Y DIFERENCIACIÓN NUMÉRICA

## A dónde nos dirigimos

En este capítulo se abordarán los temas clásicos de integración definida y de evaluación de derivadas en algún punto, por medio de técnicas numéricas. Para ello se utilizarán procesos finitos, en los que —a diferencia de los métodos analíticos, donde el concepto de límite es central y por tanto los procesos infinitos— se manejan conjuntos de puntos discretos y haremos pasar por ellos o entre ellos un polinomio, para después integrar o derivar dicho polinomio. Con estas técnicas podremos integrar y derivar funciones dadas tabularmente o bien funciones analíticas muy complejas e, incluso, integrar aquellas cuya integral “no existe”\*, como es el caso de  $e^{-x^2}$ ,  $\frac{\sin x}{x}$ ,  $\frac{\cos x}{x}$ ,  $\sqrt{1-k^2 \sin^2 x}$ , etc. Aun más, cualquiera de estas técnicas es extendible a la aproximación de integrales dobles y triples.

Siendo la derivada la medida de cambio puntual o instantáneo y la integral la suma o acumulación de tales cambios, resulta fundamental en cualquier actividad de ingeniería o ciencias, conocer las técnicas aquí estudiadas y, no menos importante, darle sentido físico a los resultados.

## Introducción

Una vez que se ha determinado un polinomio  $p_n(x)$ \*\* de manera que aproxime satisfactoriamente una función dada  $f(x)$  sobre un intervalo de interés, puede esperarse que al diferenciar  $p_n(x)$  o integrarla en forma definida, también aproxime satisfactoriamente la derivada o integral definida correspondientes a  $f(x)$ . Sin embargo, si se observa la figura 6.1 —donde aparece la gráfica de un polinomio  $p_n(x)$  que aproxima la curva que representa la función  $f(x)$ — puede anticiparse que aunque la desviación de  $p_n(x)$  y  $f(x)$  en el intervalo  $[x_0, x_n]$  es pequeña, las pendientes de las curvas que las representan pueden diferir considerablemente; esto es, la diferenciación numérica tiende a ampliar pequeñas discrepancias o errores del polinomio de aproximación.

Por otro lado, en el proceso de integración (véase Fig. 6.2), el valor de

$$\int_{x_0}^{x_n} f(x) dx,$$

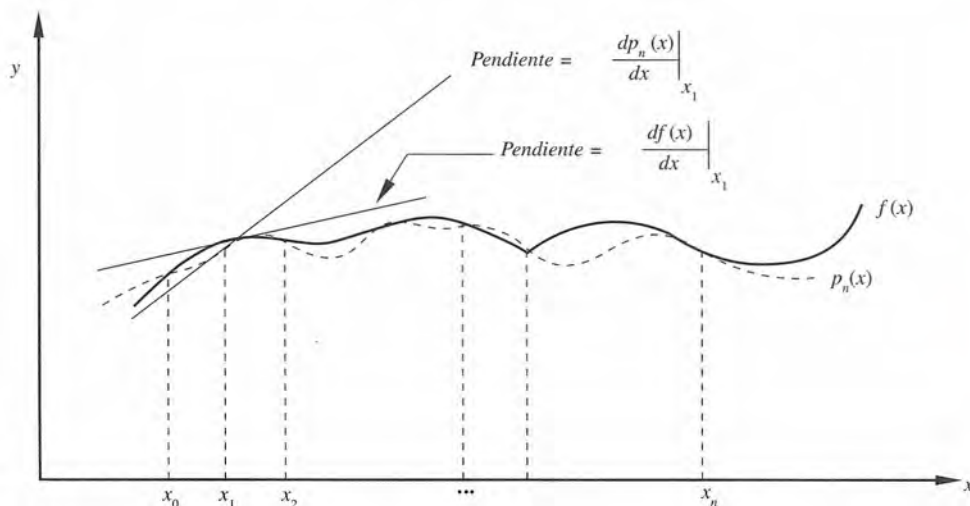
está dado por el área bajo la curva de  $f(x)$ , mientras que la aproximación

$$\int_{x_0}^{x_n} p_n(x) dx,$$

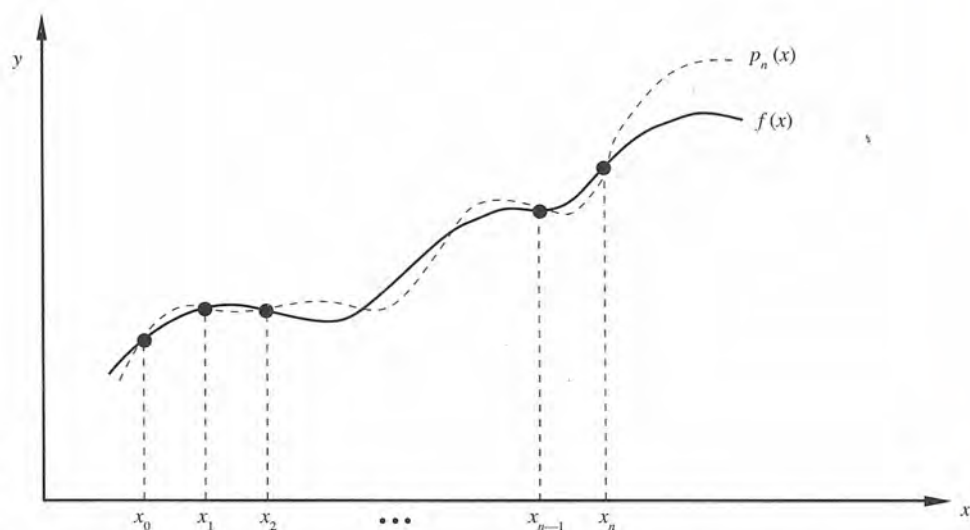
está dada por el área bajo la curva de  $p_n(x)$  y los errores que se cometen en diferentes segmentos del intervalo tienden a cancelarse entre sí o a reducirse. Por esto, el error total al

\* Se dice que la integral de una función  $f(x)$  no existe, cuando no hay una función elemental (polinomial, racional, trascendente y sus combinaciones finitas) cuya derivada sea  $f(x)$ .

\*\* Ya sea por el criterio del ajuste exacto o el de mínimos cuadrados.



**Figura 6.1**  
Diferenciación  
del polinomio de  
aproximación.



**Figura 6.2**  
Integración del  
polinomio de  
interpolación.

integrar  $p_n(x)$  entre  $x_0$  y  $x_n$  puede ser muy pequeño, aun cuando  $p_n(x)$  no sea una buena aproximación de  $f(x)$ .

En resumen: si la aproximación polinomial  $p_n(x)$  es buena, la integral

$$\int_{x_0}^{x_n} p_n(x) dx,$$

puede dar una aproximación excelente de  $\int_{x_0}^{x_n} f(x) dx$ . Por otro lado,  $\frac{d}{dx} [p_n(x)]$ , que da la pendiente de la línea tangente a  $p_n(x)$  puede variar en magnitud respecto a  $\frac{d}{dx} [f(x)]$  significativamente, aunque  $p_n(x)$  sea una buena aproximación a  $f(x)$ . Por tanto, la diferenciación numérica debe tomarse con el cuidado y reservas que lo amerita; particularmente cuando los datos obtenidos experimentalmente puedan tener errores significativos.

Los métodos de integración comúnmente usados pueden clasificarse en dos grupos: los que emplean valores dados de la función  $f(x)$  en abscisas equidistantes y que se conocen como **fórmulas de Newton-Cotes**, y aquellos que utilizan valores de  $f(x)$  en abscisas desigualmente espaciadas, determinadas por ciertas propiedades de familias de polinomios ortogonales, conocidas como **fórmulas de cuadratura gaussiana**.

## 6.1 Métodos de Newton-Cotes

Para estimar  $I = \int_a^b f(x) dx$ , los métodos de Newton-Cotes funcionan en general en dos pasos:

1. Se divide el intervalo  $[a, b]$  en  $n$  intervalos de igual amplitud, cuyos valores extremos son sucesivamente

$$x_i = x_0 + i \left( \frac{b-a}{n} \right), \quad i = 0, 1, 2, \dots, n \quad (6.1)$$

Para quedar en la nueva notación  $x_0 = a$  y  $x_n = b$ .

2. Se aproxima  $f(x)$  por un polinomio de grado  $n$ ,  $p_n(x)$  y se integra para obtener la aproximación de  $I$ .

Es evidente que se obtendrán valores diferentes de  $I$  para distintos valores de  $n$ , como se muestra a continuación.

### MÉTODO TRAPEZOIDAL

En el caso de  $n = 1$ , el intervalo de integración  $[a, b]$  queda tal cual y  $x_0 = a$ ,  $x_1 = b$ ; la aproximación polinomial de  $f(x)$  es una línea recta (un polinomio de primer grado  $p_1(x)$ ) y la aproximación a la integral es el área de trapecioide bajo esta línea recta, como se ve en la figura 6.3. Este método de integración se llama **regla trapezoidal**.

Para llevar a cabo la integración  $\int_{x_0}^{x_1} p_1(x) dx$ , es preciso seleccionar una de las formas de representación del polinomio  $p_1(x)$ , y como  $f(x)$  está dada para valores equidistantes de  $x$  con distancia  $h$ , la elección lógica es una de las fórmulas en diferencias finitas (hacia delante, hacia atrás o centrales).<sup>\*</sup> Si se eligen las diferencias finitas hacia delante, se tendrá entonces que:

$$f(x) \approx p_1(x)$$

donde  $p_1(x)$  es, según la ecuación 5.35

$$p_1(x) = p_1(x_0 + sh) = f(x_0) + s \Delta f(x_0)$$

Se reemplaza  $p_1(x)$  en la integral y se tiene:

$$\int_a^b f(x) dx \approx \int_{x_0}^{x_1} [f(x_0) + s \Delta f(x_0)] dx \quad (6.2)$$

Para realizar la integración del lado derecho de la ecuación 6.2 es necesario tener a toda la integral en términos de la nueva variable  $s$  que, como se sabe, está dada por la expresión:

$$x = x_0 + sh,$$

de ésta, la diferencial de  $x$  queda en términos de la diferencial de  $s$

$$dx = h ds,$$

ya que  $x_0$  y  $h$  son constantes.

Para que los límites de integración  $x_0$  y  $x_1$  queden a su vez en términos de  $s$ , se sustituyen por  $x$  en  $x = x_0 + sh$  y se despeja  $s$ , lo que da, respectivamente,

$$\begin{aligned} x_0 &= x_0 + sh \text{ de donde } s = 0 \\ x_1 &= x_0 + sh \text{ de donde } s = 1, \end{aligned}$$

<sup>\*</sup> Consúltase el capítulo 5.



y resulta

$$\int_{x_0}^{x_1} [f(x_0) + s \Delta f(x_0)] dx = \int_0^1 h [f(x_0) + s \Delta f(x_0)] ds$$

Al integrar se tiene:

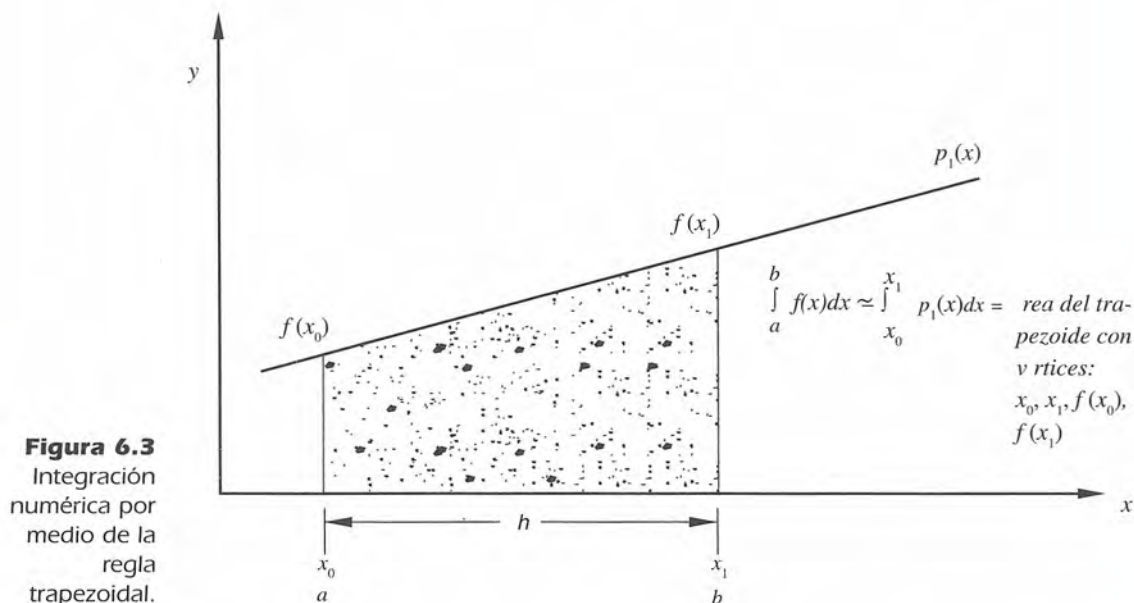
$$h \int_0^1 [f(x_0) + s \Delta f(x_0)] ds = h \left[ sf(x_0) + \frac{s^2}{2} \Delta f(x_0) \right] \Big|_0^1 = h \left[ f(x_0) + \Delta \frac{f(x_0)}{2} \right]$$

como  $\Delta f(x_0) = f(x_0+h) - f(x_0)$ , se llega finalmente a

$$\int_a^b f(x) dx \approx \frac{h}{2} [f(x_0) + f(x_1)] \quad (6.3)$$

el algoritmo del método trapezoidal.

Hay que observar que el lado derecho de la ecuación 6.3 es el área de un trapecioide de altura  $h$  y lados paralelos de longitud  $f(x_0)$  y  $f(x_1)$ , (véase Fig. 6.3).



**Figura 6.3**  
Integración  
numérica por  
medio de la  
regla  
trapezoidal.

Antes de empezar a resolver ejercicios, es conveniente observar que los métodos vistos y los siguientes sirven también cuando la función  $f(x)$  está dada analíticamente y las técnicas estudiadas en el cálculo integral no dan resultado, o bien cuando esta función es imposible de integrar analíticamente. En estos casos, la tabla de puntos se elabora evaluando la función del integrando en abscisas seleccionadas adecuadamente.

### Ejemplo 6.1

Uso del algoritmo trapezoidal.

- a) Aproxime el área  $A_1$  bajo la curva de la función dada por la tabla siguiente, en el intervalo  $a = 500$ ,  $b = 1800$ .

| Puntos | 0   | 1    | 2    | 3    | 4    | 5    |
|--------|-----|------|------|------|------|------|
| $f(x)$ | 9.0 | 13.4 | 18.7 | 23.0 | 25.1 | 27.2 |
| $x$    | 500 | 900  | 1400 | 1800 | 2000 | 2200 |

b) Aproxime  $A_2 = \int_0^5 (2 + 3x) dx$

c) Aproxime  $A_3 = \int_{-2}^4 (1 + 2x + 3x^2) dx$

d) Aproxime  $A_4 = \int_0^{\pi/2} \sin x dx$

**Solución**

Con la ecuación 6.3 se tiene

a)  $x_0 = 500$ ,  $x_1 = 1800$ , por tanto  $h = 1800 - 500 = 1300$

$$A_1 \approx \frac{1300}{2} (9 + 23) = 20800$$

b)  $x_0 = 0$ ,  $x_1 = 5$ , por tanto  $h = 5 - 0 = 5$

$$A_2 \approx \frac{5}{2} ([2+3(0)] + [2+3(5)]) = 47.5$$

La TI-92 Plus permite graficar la función del integrando y realizar la integración. Para ello ejecuta las siguientes instrucciones.



En la línea de edición de la pantalla Home escriba:

$2+3*x \rightarrow y1(x)$

$\{0, 5\} \rightarrow 1x: \{2, 17\} \rightarrow 1y$

Newplot 1, 2, 1x, 1y

En la pantalla Window (♦E) establezca los siguientes parámetros:

$xmin = -0.5$

$xmax = 5.5$

$xscl = 2$

$ymin = -2$

$ymax = 18$

$yscl = 0$

$xres = 2$

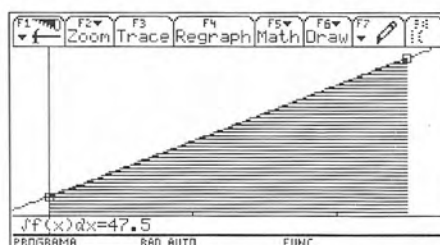
Invoque la pantalla Graph con ♦R

Oprima la tecla F5 y luego 7

Solicita Lower Limit? Escriba 0 y oprima Enter

Solicita Upper Limit? Escriba 5 y oprima Enter

Obtendrá el resultado siguiente:



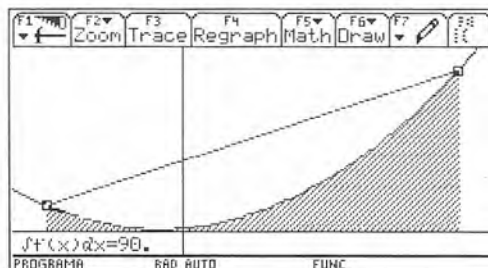
$$c) \quad x_0 = -2, \quad x_1 = 4, \quad \text{por tanto } h = 4 - (-2) = 6$$

$$A_3 \approx \frac{6}{2} ([1 + 2(-2) + 3(-2)^2] + [1 + 2(4) + 3(4)^2]) = 198$$

Para este caso, haga los cambios correspondientes a las instrucciones dadas en el inciso b).



```
1+2*x+3*x^2→y1 (x)
{-2, 4}→1x
{9, 57}→1y
xmin= -2.5
xmax= 4.5
ymin= -8
ymax= 64
Lower Limit? -2
Upper Limit? 4
```



$$d) \quad x_0 = 0, \quad x_1 = \pi/2, \quad \text{por tanto } h = \pi/2 - 0 = \pi/2$$

$$A_4 \approx \frac{\pi/2}{2} (\sin(0) + \sin(\pi/2)) = \pi/4$$

Puede el lector hacer modificaciones a las instrucciones dadas en el inciso a) y obtener los valores y gráficas correspondientes a este caso.

Se deja al lector la comparación y discusión de los resultados obtenidos en estos cálculos y los dados por la calculadora, así como las gráficas de los trapezoides y de las funciones del integrando.

## MÉTODO DE SIMPSON

Si  $n = 2$ ; esto es, el intervalo de integración  $[a, b]$  se divide en dos subintervalos, se tendrán tres abscisas dadas por la ecuación 6.1 como:

$$\begin{aligned} x_0 &= a \\ x_1 &= x_0 + 1 \frac{(b-a)}{2} = a + \frac{b}{2} - \frac{a}{2} = \frac{1}{2}(b+a), \\ x_2 &= b \end{aligned}$$

Se aproxima  $f(x)$  con una parábola [un polinomio de segundo grado  $p_2(x)$ ], y la aproximación a la integral será el área bajo el segmento de parábola comprendida entre  $f(x_0)$  y  $f(x_2)$  como muestra la figura 6.4. Esto es:

$$\int_a^b f(x) dx \approx \int_{x_0}^{x_2} p_2(x) dx$$

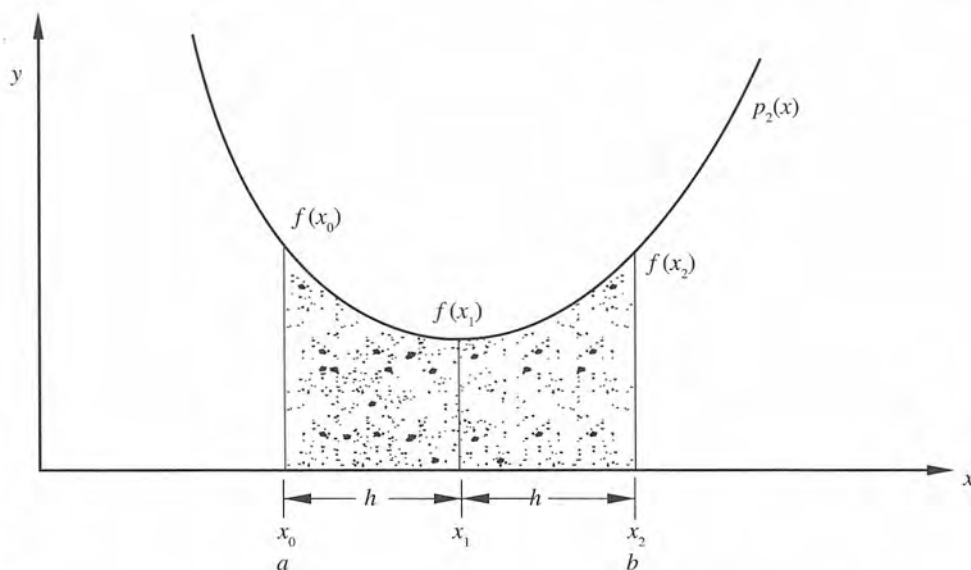
para realizar la integración  $\int_{x_0}^{x_2} p_2(x) dx$ , se usa la fórmula de Newton en diferencias finitas hacia delante para expresar  $p_2(x)$  (Ec. 5.35)

$$p_2(x) = p_2(x_0 + sh) = f(x_0) + s \Delta f(x_0) + \frac{s(s-1)}{2!} \Delta^2 f(x_0)$$

al sustituir  $p_2(x)$  y expresar toda la integral en términos de la nueva variable  $s$ , queda

$$\int_a^b f(x) dx \approx \int_{x_0}^{x_2} p_2(x) dx = h \int_0^2 p_2(x_0 + sh) ds$$





**Figura 6.4**  
Integración  
numérica  
mediante la  
regla de  
Simpson.

$$\begin{aligned} \int_0^2 p_2(x_0 + sh) ds &= h \int_0^2 \left[ f(x_0) + s \Delta f(x_0) + \frac{s(s-1)}{2!} \Delta^2 f(x_0) \right] ds \\ &= h \left[ s f(x_0) + \frac{s^2}{2} \Delta f(x_0) + \frac{s^3}{3!} \Delta^2 f(x_0) - \frac{s^2}{4} \Delta^2 f(x_0) \right] \Big|_0^2 \\ &= h \left[ 2 f(x_0) + 2 \Delta f(x_0) + \frac{1}{3} \Delta^2 f(x_0) \right] \end{aligned}$$

De la definición de la primera y segunda diferencia hacia adelante se tiene:

$$\Delta f(x_0) = f(x_0 + h) - f(x_0) = f(x_1) - f(x_0)$$

y

$$\Delta^2 f(x_0) = f(x_0 + 2h) - 2f(x_0 + h) + f(x_0) = f(x_2) - 2f(x_1) + f(x_0)$$

que sustituidas en la última ecuación dan lugar a

$$\int_a^b f(x) dx \approx \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)] \quad (6.4)$$

el algoritmo de Simpson.

### Ejemplo 6.2

Con el algoritmo de Simpson aproxime las integrales del ejemplo 6.1

#### Solución

Con la ecuación 6.4 se tiene

$$a) \quad h = \frac{1800-500}{2} = 650, \quad x_0 = 500, \quad x_1 = x_0 + h = 500 + 650 = 1150, \quad x_2 = 1800$$

$$f(x_0) = 9, \quad f(x_1) = 16.08, \quad f(x_2) = 23$$

[ $f(x_1)$  se obtiene interpolando con un polinomio de segundo grado en diferencias finitas]

$$A_1 \approx \frac{650}{3} [9 + 4(16.08) + 23] = 20869.33$$

$$b) \quad h = \frac{5-0}{3} = 2.5, x_0 = 0, x_1 = 0 + 2.5 = 2.5, x_2 = 5$$

$$A_2 \approx \frac{2.5}{3} [2 + 3(0) + 4(2 + 3(2.5)) + 2 + 3(5)] = 47.5$$

$$c) \quad h = \frac{4 - (-2)}{2} = 3, x_0 = -2, x_1 = 2 + 3 = 1, x_2 = 4$$

$$A_3 \approx \frac{3}{3} [1 + 2(-2) + 3(-2)^2 + 4(1 + 2(1) + 3(1)^2) + 1 + 2(4) + 3(4)^2] = 90$$

$$d) \quad h = \frac{\frac{\pi}{2} - 0}{2} = \pi/4, x_0 = 0, x_1 = 0 + \pi/4, x_2 = \frac{\pi}{2}$$

$$A_4 \approx \frac{\pi/4}{3} (\sin 0 + 4 \sin \pi/4 + \sin \pi/2) = 1.0023$$

Se deja al lector la comparación y discusión de los resultados obtenidos [casos de los incisos b), c) y d)] con los obtenidos en el ejemplo 6.1.

### CASO GENERAL

A continuación se verá el caso más general, donde el intervalo de integración  $[a, b]$  se divide en  $n$  subintervalos y da lugar a  $n + 1$  abscisas equidistantes  $x_0, x_1, \dots, x_n$ , con  $x_0 = a$  y  $x_n = b$  (véase Fig. 6.2). Esta vez el polinomio de interpolación es de  $n$ -ésimo grado  $p_n(x)$  y se utilizará la representación 5.35 para éste.

La aproximación a la integral  $\int_a^b f(x) dx$  está dada por

$$\begin{aligned} \int_a^b f(x) dx &\approx \int_{x_0}^{x_n} p_n(x) dx = h \int_0^n p_n(x_0 + sh) ds \\ &= h \int_0^n [f(x_0) + s \Delta f(x_0) + \frac{s(s-1)}{2!} \Delta^2 f(x_0) + \frac{s(s-1)(s-2)}{3!} \Delta^3 f(x_0) \\ &\quad + \dots + \frac{s(s-1)(s-2) \dots (s-(n-1))}{n!} \Delta^n f(x_0)] ds \end{aligned}$$

Con la integración de los cinco primeros términos se tiene:

$$\begin{aligned} h \int_0^n p_n(x_0 + sh) ds &= h \left[ sf(x_0) + \frac{s^2}{2} \Delta f(x_0) + \left(\frac{s^3}{6} - \frac{s^2}{4}\right) \Delta^2 f(x_0) \right. \\ &\quad + \left(\frac{s^4}{24} - \frac{s^3}{6} + \frac{s^2}{6}\right) \Delta^3 f(x_0) \\ &\quad \left. + \left(\frac{s^5}{120} - \frac{s^4}{16} + \frac{11s^3}{72} - \frac{s^2}{8}\right) \Delta^4 f(x_0) + \text{términos faltantes} \right] \Big|_0^n \end{aligned}$$

Todos los términos son cero en el límite inferior, por lo que

$$\begin{aligned}
 \int_a^b f(x) dx \approx h \left[ n f(x_0) + \frac{n^2}{2} \Delta f(x_0) + \left( \frac{n^3}{6} - \frac{n^2}{4} \right) \Delta^2 f(x_0) \right. \\
 + \left( \frac{n^4}{24} - \frac{n^3}{6} + \frac{n^2}{6} \right) \Delta^3 f(x_0) \\
 \left. + \left( \frac{n^5}{120} - \frac{n^4}{16} + \frac{11n^3}{72} - \frac{n^2}{8} \right) \Delta^4 f(x_0) + \text{términos faltantes} \right]
 \end{aligned} \quad (6.5)$$

A continuación se dan las fórmulas de Newton-Cotes para integrar cuando  $n = 1, 2, 3, 4, 5$  y  $6$ . El lector puede verificarlas sustituyendo el valor seleccionado de  $n$  y las diferencias correspondientes en términos de sus valores funcionales en la ecuación 6.5

$n = 1$

$$\int_{x_0}^{x_1} f(x) dx \approx \frac{h}{2} [f(x_0) + f(x_1)] \quad \text{trapezoidal}$$

$n = 2$

$$\int_{x_0}^{x_2} f(x) dx \approx \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)] \quad \text{Simpson 1/3}$$

$n = 3$

$$\int_{x_0}^{x_3} f(x) dx \approx \frac{3h}{8} [f(x_0) + 3f(x_1) + 3f(x_2) + f(x_3)] \quad \text{Simpson 3/8}$$

$n = 4$

$$\int_{x_0}^{x_4} f(x) dx \approx \frac{2h}{45} [7f(x_0) + 32f(x_1) + 12f(x_2) + 32f(x_3) + 7f(x_4)] \quad (6.6)$$

$n = 5$

$$\int_{x_0}^{x_5} f(x) dx \approx \frac{5h}{288} [19f(x_0) + 75f(x_1) + 50f(x_2) + 50f(x_3) + 75f(x_4) + 19f(x_5)]$$

$n = 6$

$$\int_{x_0}^{x_6} f(x) dx \approx \frac{h}{140} [41f(x_0) + 216f(x_1) + 27f(x_2) + 272f(x_3) + 27f(x_4) + 216f(x_5) + 41f(x_6)]$$

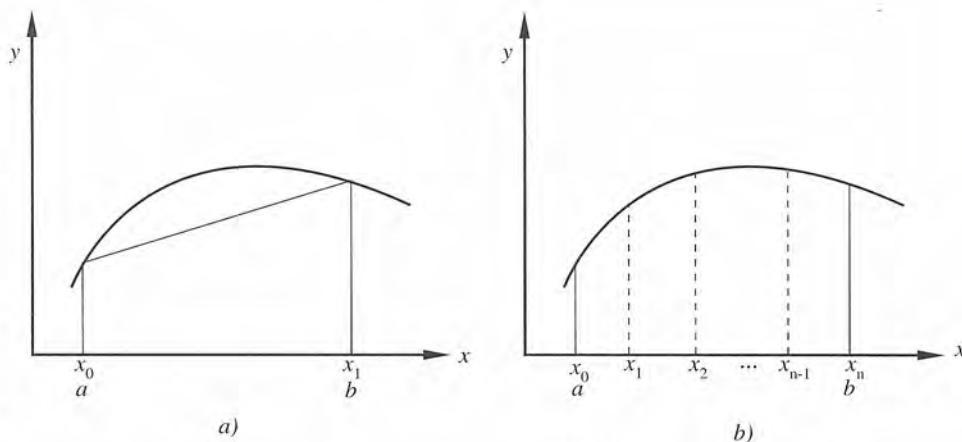
## MÉTODOS COMPUESTOS DE INTEGRACIÓN

Algunas veces el intervalo de integración es tan amplio, que resulta conveniente dividirlo en subintervalos y aproximar cada uno por medio de un polinomio.

### MÉTODO TRAPEZOIDAL COMPUESTO

Por ejemplo, en vez de aproximar la integral de  $f(x)$  en  $[a, b]$  por una recta (véase Fig. 6.5.a), conviene dividir  $[a, b]$  en  $n$  subintervalos y aproximar cada uno por un polinomio de primer grado (véase Fig. 6.5 b). Una vez hecho esto, se aplica la fórmula trapezoidal a cada subintervalo y se obtiene el área de cada trapezoide, de tal modo que la suma de todas ellas da la aproximación al área bajo la curva  $f(x)$ .





**Figura 6.5**  
Integración por  
el método  
trapezoidal  
compuesto.

Esto es

$$I = \int_a^b f(x) dx \approx \int_{x_0}^{x_1} p_1(x) dx + \int_{x_1}^{x_2} p_2(x) dx + \dots + \int_{x_{n-1}}^{x_n=b} p_n(x) dx$$

donde  $p_i(x)$  es la ecuación de la recta que pasa por los puntos  $(x_{i-1}, f(x_{i-1}))$ ,  $(x_i, f(x_i))$ . Con la ecuación 6.3 se tiene

$$I = \frac{x_1 - x_0}{2} [f(x_0) + f(x_1)] + \frac{x_2 - x_1}{2} [f(x_1) + f(x_2)] + \dots + \frac{x_n - x_{n-1}}{2} [f(x_{n-1}) + f(x_n)] \quad (6.7)$$

Si todos los subintervalos son del mismo tamaño  $h$ , esto es, si  $x_{i+1} - x_i = h$ , para  $i = 0, 1, \dots, (n-1)$ , entonces la ecuación 6.7 puede anotarse

$$I \approx \frac{h}{2} [f(x_0) + 2f(x_1) + 2f(x_2) + \dots + 2f(x_{n-1}) + f(x_n)],$$

que puede escribirse con la notación de sumatoria

$$I \approx \frac{h}{2} [f(x_0) + 2 \sum_{i=1}^{n-1} f(x_i) + f(x_n)] \quad (6.8)$$

### Ejemplo 6.3

Mediante el algoritmo trapezoidal compuesto, aproxime el área bajo la curva de la siguiente función dada en forma tabular, entre  $x = -1$  y  $x = 4$ .

| Puntos | 0  | 1  | 2  | 3  | 4  | 5   |
|--------|----|----|----|----|----|-----|
| $x$    | -1 | 0  | 1  | 2  | 3  | 4   |
| $f(x)$ | 8  | 10 | 10 | 20 | 76 | 238 |

### Solución

Si se toman todos los puntos de la tabla, se puede aplicar cinco veces el método trapezoidal. Como todos los intervalos son del mismo tamaño ( $h=1$ ), se usa la ecuación 6.8 directamente.

$$A \approx \frac{1}{2} [8 + 2(10 + 10 + 20 + 76) + 238] = 239$$

Compárese este resultado con la solución analítica (los datos de la tabla corresponden a función  $f(x) = x^4 - 2x^2 + x + 10$ ).

**ALGORITMO 6.1** Método trapezoidal compuesto

Para aproximar el área bajo la curva de una función analítica  $f(x)$  en el intervalo  $[a, b]$ , proporcionar la función por integrar  $F(x)$  y los

DATOS: El número de trapezoides  $N$ , el límite inferior  $A$  y límite superior  $B$ .

RESULTADOS: El área aproximada  $\hat{A}$ REA.

PASO 1. Hacer  $X = A$ .

PASO 2. Hacer  $S = 0$ .

PASO 3. Hacer  $H = (B - A)/N$ .

PASO 4. SI  $N = 1$ , ir al paso 10. De otro modo continuar.

PASO 5. Hacer  $I = 1$ .

PASO 6. Mientras  $I \leq N - 1$ , repetir los pasos 7 a 9.

PASO 7. Hacer  $X = X + H$ .

PASO 8. Hacer  $S = S + F(X)$ .

PASO 9. Hacer  $I = I + 1$ .

PASO 10. Hacer  $\hat{A}$ REA =  $H/2 * (F(A) + 2 * S + F(B))$ .

PASO 11. IMPRIMIR  $\hat{A}$ REA Y TERMINAR.

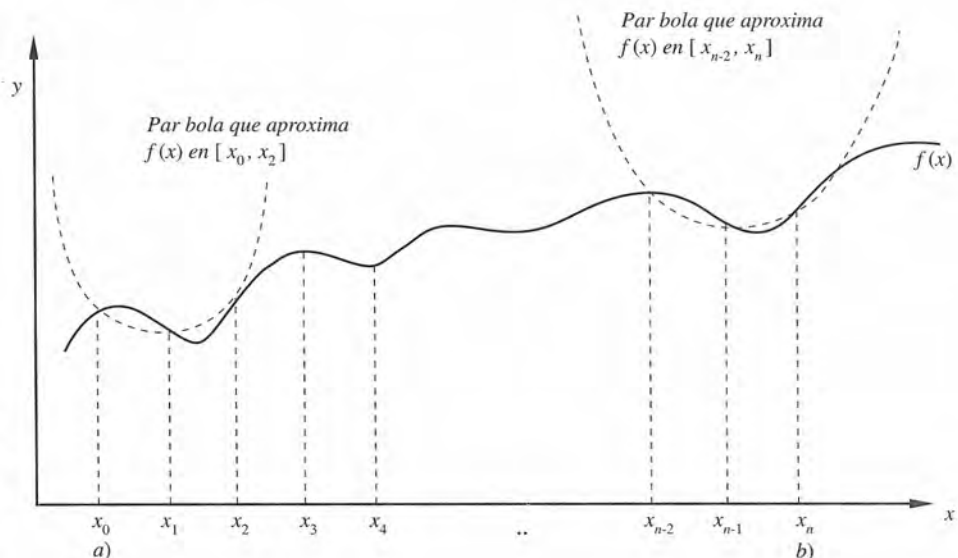
**MÉTODO DE SIMPSON COMPUESTO**

Como para cada aplicación de la regla de Simpson se requieren dos subintervalos, a fin de aplicarla  $n$  número de veces, deberá dividirse el intervalo  $[a, b]$  en un número de subintervalos igual a  $2n$  (véase Fig. 6.6).

Cada par de subintervalos sucesivos se aproxima por un polinomio de segundo grado (parábola) y se integra usando la ecuación 6.4, de tal manera que la suma de las áreas bajo cada segmento de parábola sea la aproximación a la integración deseada. Esto es:

$$I = \int_a^b f(x) dx \approx \int_{a=x_0}^{x_2} p_1(x) dx + \int_{x_2}^{x_4} p_2(x) dx + \dots + \int_{x_{n-2}}^{x_n=b} p_n(x) dx$$

donde  $p_i(x)$ ,  $i = 1, 2, \dots, n$ , es el polinomio de segundo grado que pasa por tres puntos consecutivos.



**Figura 6.6**  
Integración por  
método de  
Simpson  
compuesto.

Al sustituir la ecuación 6.4 en cada uno de los sumandos se tiene:

$$I \approx \frac{h_1}{3} [f(x_0) + 4f(x_1) + f(x_2)] + \frac{h_2}{3} [f(x_2) + 4f(x_3) + f(x_4)] + \dots + \frac{h_n}{3} [f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)] \quad (6.9)$$

donde:

$$\begin{aligned} h_1 &= x_1 - x_0 = x_2 - x_1 \\ h_2 &= x_3 - x_2 = x_4 - x_3 \\ &\vdots \\ h_n &= x_{n-1} - x_{n-2} = x_n - x_{n-1} \end{aligned}$$

Si  $h_1 = h_2 = \dots = h_n$ , la ecuación (6.9) queda como sigue:

$$I \approx \frac{h}{3} [f(x_0) + 4f(x_1) + f(x_2)] + \frac{h}{3} [f(x_2) + 4f(x_3) + f(x_4)] + \dots + \frac{h}{3} [f(x_{n-2}) + 4f(x_{n-1}) + f(x_n)]$$

que usando la notación de sumatoria queda de la siguiente manera:

$$I \approx \frac{h}{3} \left[ f(x_0) + 4 \sum_{\substack{i=1 \\ \Delta i=2}}^{n-1} f(x_i) + 2 \sum_{\substack{i=2 \\ \Delta i=2}}^{n-2} f(x_i) + f(x_n) \right] \quad (6.10)$$

donde  $\Delta i$  significa el incremento de  $i$ .

### Ejemplo 6.4

Mediante el algoritmo de Simpson de integración, aproxime el área bajo la curva del ejemplo 6.3.

#### Solución

Con los puntos dados de la tabla, se puede aplicar la regla de Simpson en dos ocasiones; por ejemplo, una vez con los puntos (0), (1) y (2) y otra con los puntos (2), (3) y (4). Como la integración debe hacerse de  $x = -1$  a  $x = 4$ , se integra entre los puntos (4) y (5) con el método trapezoidal y la suma será la aproximación buscada:

- a) Método de Simpson aplicado dos veces:  $h_1 = h_2 = h_3 = h_4 = 1$ , entonces  
Puede usarse la ecuación 6.10

$$A_1 \approx \frac{1}{3} [8 + 4(10 + 20) + 2(10) + 76] = 74.666$$

- b) Método trapezoidal aplicado a los puntos (4) y (5)

$$A_2 \approx \frac{1}{2} (76 + 238) = 157$$

por tanto, la aproximación al área es:

$$A \approx 74.666 + 157 = 231.666$$

Compare este resultado con el obtenido en el ejemplo 6.3 y el resultado de la solución analítica (la función tabulada es  $f(x) = x^4 - 2x^2 + x + 10$ ).



**Ejemplo 6.5**

Encuentre la integral aproximada de la función

$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

que da lugar a la curva normal tipificada, entre los límites  $-1$  y  $1$ .

- a) Utilice la regla trapezoidal con varios trapezoides y compare con el resultado (0.682) obtenido en tablas.  
 b) Use la regla de Simpson varias veces y compare con el resultado (0.682) obtenido de tablas.

**Solución**

a) Con  $n = 1$ ,  $h = \frac{1 - (-1)}{1} = 2$

$$I \approx \frac{2}{2\sqrt{2\pi}} [f(x_0) + f(x_1)] = \frac{1}{\sqrt{2\pi}} [0.606 + 0.606] = 0.484$$

El error relativo, tomando el valor de tablas como verdadero, es:

$$E_{n=1} = \frac{|0.484 - 0.682|}{0.682} = 0.29 \quad \text{o} \quad 29\%$$

Con  $n = 2$ ,  $h = \frac{1 - (-1)}{2} = 1$

$$I \approx \frac{1}{2\sqrt{2\pi}} [f(x_0) + 2f(x_1) + f(x_2)] = \frac{1}{2\sqrt{2\pi}} [0.606 + 2(1) + 0.606] = 0.64$$

$$E_{n=2} = \frac{|0.64 - 0.682|}{0.682} = 0.0587 \quad \text{o} \quad 5.87\%$$

Con  $n = 4$ ,  $h = \frac{1 - (-1)}{4} = 0.5$

$$\begin{aligned} I &\approx \frac{0.5}{2\sqrt{2\pi}} [f(x_0) + 2f(x_1) + 2f(x_2) + 2f(x_3) + f(x_4)] \\ &= \frac{0.5}{2\sqrt{2\pi}} [0.606 + 2(0.882) + 2(1) + 2(0.882) + 0.606] = 0.672 \end{aligned}$$

$$E_{n=4} = \frac{|0.672 - 0.682|}{0.682} = 0.0147 \quad \text{o} \quad 1.47\%$$

b) Con  $n = 2$ ,  $h = \frac{1 - (-1)}{2} = 1$

$$I \approx \frac{1}{3\sqrt{2\pi}} [f(x_0) + 4f(x_1) + f(x_2)] = \frac{1}{3\sqrt{2\pi}} [0.606 + 4(1) + 0.606] = 0.693$$

$$E_{n=2} = \frac{|0.693 - 0.682|}{0.682} = 0.0162 \quad \text{o} \quad 1.62\%$$

Con  $n = 4$ ,  $h = \frac{1 - (-1)}{4} = 0.5$

$$I \approx \frac{0.5}{3\sqrt{2\pi}} [f(x_0) + 4f(x_1) + 2f(x_2) + 4f(x_3) + f(x_4)]$$

$$= \frac{0.5}{3\sqrt{2\pi}} [0.606 + 4(0.882) + 2(1) + 4(0.882) + 0.606] = 0.683$$

$$E_{n=4} = \frac{|0.683 - 0.682|}{0.682} = 0.0015 \quad \text{o} \quad 0.15\%$$

### ALGORITMO 6.2 Método de Simpson compuesto

Para aproximar el área bajo la curva de una función analítica  $f(x)$  en el intervalo  $[a, b]$ , proporcionar la función por integrar  $F(X)$  y los

DATOS: El número (par) de subintervalos  $N$ , el límite inferior  $A$  y el límite superior  $B$ .

RESULTADOS: El área aproximada  $AREA$ .

- PASO 1. Hacer  $S1=0$ .  
 PASO 2. Hacer  $S2=0$ .  
 PASO 3. Hacer  $X=A$ .  
 PASO 4. Hacer  $H=(B-A)/N$ .  
 PASO 5. Si  $N=2$ , ir al paso 13. De otro modo continuar.  
 PASO 6. Hacer  $I=1$ .  
 PASO 7. Mientras  $I \leq N/2-1$ , repetir los pasos 8 a 12.  
     PASO 8. Hacer  $X = X + H$ .  
     PASO 9. Hacer  $S1 = S1 + F(X)$ .  
     PASO 10. Hacer  $X = X + H$ .  
     PASO 11. Hacer  $S2 = S2 + F(X)$ .  
     PASO 12. Hacer  $I = I + 1$ .  
 PASO 13. Hacer  $X = X + H$ .  
 PASO 14. Hacer  $S1 = S1 + F(X)$ .  
 PASO 15. Hacer  $AREA = H/3 * (F(A) + 4*S1 + 2*S2 + F(B))$ .  
 PASO 16. IMPRIMIR  $AREA$  Y TERMINAR.

### Ejemplo 6.6



Elabore un subprograma para integrar la función del ejemplo 6.5 con el método trapecoidal compuesto, usando sucesivamente 1, 2, 4, 8, 16, 32, 64, ..., 1024 subintervalos y calcule sus correspondientes errores relativos en por ciento. Con los resultados obtenidos elabore una gráfica, error porcentual contra número de subintervalos y discútala.

El programa 6.1 que se encuentra en el disco fue diseñado para usar el subprograma **TRAPECIOS** en la integración de

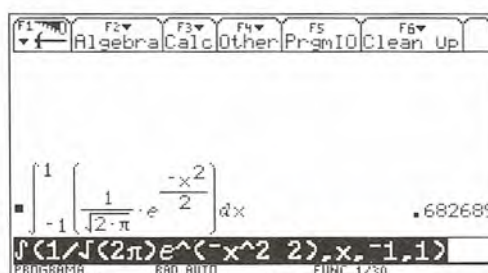
$$\frac{1}{\sqrt{2\pi}} e^{-x^2/2},$$

en el intervalo  $[-1, 1]$ , usando  $N=1, 2, 4, \dots, 1024$  subintervalos sucesivamente.

Los resultados obtenidos para cada valor de  $N$  son:

| N    | APROXIMACIÓN AL ÁREA | ERROR EN % |
|------|----------------------|------------|
| 1    | 0.483941             | 29.112458  |
| 2    | 0.640913             | 6.119330   |
| 4    | 0.672522             | 1.489283   |
| 8    | 0.680164             | 0.369906   |
| 16   | 0.682059             | 0.092278   |
| 32   | 0.682532             | 0.023006   |
| 64   | 0.682650             | 0.005697   |
| 128  | 0.682680             | 0.001370   |
| 256  | 0.682687             | 0.000288   |
| 512  | 0.682689             | 0.000018   |
| 1024 | 0.682689             | 0.000050   |

El error relativo porcentual se calculó utilizando como valor verdadero el reportado por la TI-92 Plus a seis cifras decimales, que es 0.682689\*.



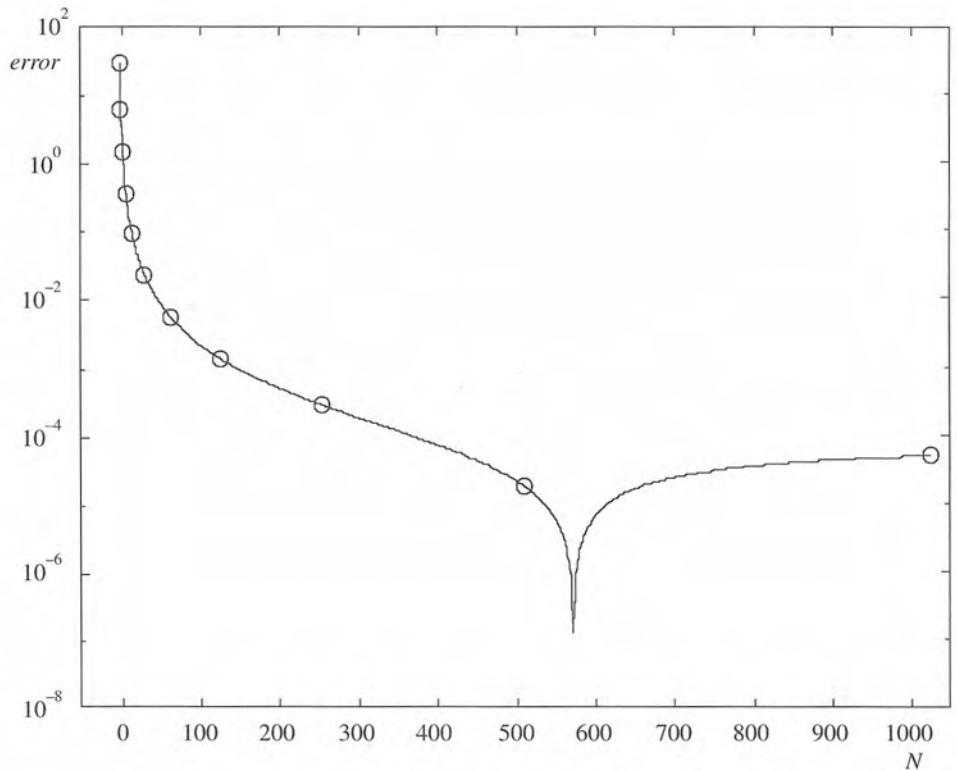
Se traza la gráfica y se hacen los comentarios correspondientes

#### Comentarios:

- La gráfica se elaboró en escalas semilogarítmicas.
- El error obtenido por el programa es básicamente la suma de dos tipos de errores, el de truncamiento (debido a la aproximación de la función en cada subintervalo por una línea recta) y el de redondeo (por el tamaño de la palabra de memoria de la computadora en que se realizan los cálculos).

\* Los valores de la tabla y como consecuencia la gráfica se modificarán dependiendo de la máquina que se utilice para hacer los cálculos y del número de cifras que se tomen como "valor verdadero", aunque los comentarios se cumplen en lo general.





El error de truncamiento disminuye al aumentar el número de subintervalos y teóricamente tiende a cero cuando  $N$  tiende a infinito. Por otro lado, el error de redondeo crece al aumentar el número de subintervalos (debido al aumento del número de cálculos). En la gráfica se ve que el error global disminuye al incrementar el número de subintervalos hasta llegar a un mínimo entre 500 y 600 subintervalos, para luego aumentar debido a que el peso del error de redondeo empieza a dominar.

Para realizar los cálculos puede usarse Matlab o la TI-92 Plus.



```
a=-1; b=1
for i=1:11
    n=2^(i-1); h=(b-a)/n; x=a; s=0;
    if n>1
        for j=1:n-1
            x=x+h;
            f=1/sqrt(2*pi)*exp(-x^2/2) ;
            s=f+s;
        end
    end
    fa=1/sqrt(2*pi)*exp(-a^2/2);
    fb=1/sqrt(2*pi)*exp(-b^2/2);
    s=h/2*(fa+2*s + fb);
    e=abs(0.682689-s)/s*100;
    fprintf ('%4d %8.6f %8.6f\n',n,s,e)
end
```



```

e6_6( )
Prgm
Define f(x)=1./((2*pi)))*e^(-x^2/2)
ClrIO: ClrHome:-1.→a: 1.→b
For i, 1, 10
2^(i-1)→n: (b-a)/n→h: a→x: 0→s
If n>1 Then
For j, 1, n-1
x + h→x:f(x)+s→s
EndFor
EndIf
h/2*(f(a) +2*s+f(b))→s
abs(0.682689-s)/s*100→e
format(n,"f0")&" "&format(s,"f3")&" "&format(e,"f3")→d
Disp d
EndFor
EndPrgm

```

En general, cuando se recurre a una integración numérica, no se tiene el resultado verdadero y resulta conveniente integrar con un número  $n$  de subintervalos y luego con el doble. Si los resultados no difieren considerablemente, puede aceptarse como bueno cualquiera de los dos.

### Ejemplo 6.7

Elabore un subprograma para integrar una función analítica por el método de Simpson compuesto, usando sucesivamente 2, 4, 8, 16, ..., 2048 subintervalos. Compruébela con la función del ejemplo 6.5.

### Solución

Ver PROGRAMA 6.2 en el CD.



En el CD encontrará el PROGRAMA 6.5 de Integración Numérica con los métodos de trapecios, el de Simpson 1/3 y el de cuadratura de Gauss (el cual se estudia en la sección 6.2). Con este programa usted puede proporcionar la función a integrar simbólicamente, los límites de integración y el número de intervalos. Podrá apreciar simultáneamente la representación gráfica de los distintos métodos y valores numéricos para diferentes secciones del área.

### ANÁLISIS DEL ERROR DE TRUNCAMIENTO EN LA APROXIMACIÓN TRAPEZOIDAL

Considérese el  $i$ -ésimo trapecio de una integración trapezoidal compuesta o sucesiva, con abscisas  $x_{i-1}$  y  $x_i$ . La distancia entre estas abscisas es  $h = (b - a)/n$ . Sea además  $F(x)$  la primitiva del integrando  $f(x)$ , es decir  $dF(x)/dx = f(x)$ . Con esto la integral de la función  $f(x)$  en el intervalo  $[x_{i-1}, x_i]$  queda dada por:

$$I_i = \int_{x_{i-1}}^{x_i} f(x) dx = F(x_i) - F(x_{i-1}) \quad (6.11)$$

Por otro lado, la aproximación de  $I_i$  usando el método trapezoidal es

$$T_i = \frac{h}{2} [f(x_{i-1}) + f(x_i)] \quad (6.12)$$

En ausencia de errores de redondeo puede definirse el error de truncamiento para este trapecioide particular así:

$$E_i = T_i - I_i \quad (6.13)$$

Para continuar con este análisis,  $f(x)$  se expande en serie de Taylor alrededor de  $x = x_i$ , de modo de obtener  $f(x_{i-1})$

$$f(x_{i-1}) = f(x_i) + (x_{i-1} - x_i) f'(x_i) + \frac{(x_{i-1} - x_i)^2}{2!} f''(x_i) + \dots$$

como  $h = x_i - x_{i-1}$

$$f(x_{i-1}) = f(x_i) - h f'(x_i) + \frac{h^2}{2!} f''(x_i) - \dots \quad (6.14)$$

se sustituye la ecuación 6.14 en la 6.12

$$\begin{aligned} T_i &= \frac{h}{2} [2f(x_i) - h f'(x_i) + \frac{h^2}{2!} f''(x_i) + \dots] \\ T_i &= h f(x_i) - \frac{h^2}{2} f'(x_i) + \frac{h^3}{2(2!)} f''(x_i) + \dots \end{aligned} \quad (6.15)$$

En forma análoga puede obtenerse

$$F(x_{i-1}) = F(x_i) - h F'(x_i) + \frac{h^2}{2!} F''(x_i) - \frac{h^3}{3!} F'''(x_i) + \dots$$

Cuya sustitución en la ecuación 6.11 produce

$$I_i = h F'(x_i) - \frac{h^2}{2!} F''(x_i) + \frac{h^3}{3!} F'''(x_i) - \dots$$

Como:

$$\begin{aligned} f(x) &= F'(x) \\ f'(x) &= F''(x) \\ f''(x) &= F'''(x) \\ &\vdots \\ &\vdots \end{aligned}$$

y al sustituir se obtiene:

$$I_i = h f(x_i) - \frac{h^2}{2!} f'(x_i) + \frac{h^3}{3!} f''(x_i) - \dots \quad (6.16)$$

Al reemplazar las ecuaciones 6.15 y 6.16 en la (6.13)

$$\begin{aligned} E_i &= [h f(x_i) - \frac{h^2}{2!} f'(x_i) + \frac{h^3}{2(2!)} f''(x_i) + \dots] \\ &\quad - [h f(x_i) - \frac{h^2}{2!} f'(x_i) + \frac{h^3}{3!} f''(x_i) - \dots] \\ E_i &= (\frac{1}{4} - \frac{1}{6}) h^3 f''(x_i) + \text{términos en } h^4, h^5, \text{ etcétera.} \end{aligned}$$

Considerando que  $h$  es pequeña ( $h \ll 1$ ), los términos en  $h^4$ ,  $h^5$ , etc., pueden despreciarse, de modo que el error de truncamiento del  $i$ -ésimo trapecioide queda dado aproximadamente así:

$$E_i \approx \frac{h^3}{12} f''(x_i) \quad (6.17)$$

Si además  $|f''(x)| \leq M$  para  $a \leq x \leq b$ , entonces:

$$|E_i| \leq \frac{h^3}{12} M,$$



de donde el error de truncamiento usando  $n$  trapecios en la integración de  $f(x)$  en  $[a, b]$  queda dado por:

$$|E_T| < \frac{nh^3}{12} M = nh \frac{h^2}{12} M = (b-a) \frac{h^2}{12} M \quad (6.18)$$

Por tanto, el error de truncamiento en el método trapezoidal es proporcional a  $h^2$  lo cual, para fines de análisis, suele expresarse así:

$$\int_a^b f(x) dx = \frac{h}{2} [f(x_0) + 2 \sum_{i=1}^{n-1} f(x_i) + f(x_n)] + O(h^2) \quad (6.19)$$

y se dice que es una fórmula que genera aproximaciones del orden  $O(h^2)$  (véase Ec. 6.8).

### EXTRAPOLACIÓN DE RICHARDSON. INTEGRACIÓN DE ROMBERG

Con el nombre de **extrapolación de Richardson** se conoce un conjunto de técnicas que generan mejores aproximaciones a los resultados buscados o aproximaciones equivalentes a métodos de alto orden, a partir de las aproximaciones obtenidas por medio de algún método de bajo orden, y pocos cálculos. Estas técnicas están basadas en el análisis del error de truncamiento, cuya aplicación a la integración numérica se presenta a continuación.

Supóngase que el error de truncamiento de cierto algoritmo de aproximación de

$$I = \int_a^b f(x) dx$$

se expresa:

$$E = c h^r f^{(r)}(\xi)$$

donde  $c$  es independiente de  $h$ ,  $r$  es un entero positivo y  $\xi$  un punto desconocido de  $(a, b)$ . Luego de obtener dos aproximaciones de  $I$  con tamaños de paso distintos:  $h_1$  y  $h_2$ , de llamar a dichas aproximaciones  $I_1$  y  $I_2$ , respectivamente, y despreciar errores de redondeo, se puede escribir:

$$I - I_1 = c h_1^r f^{(r)}(\xi_1)$$

$$I - I_2 = c h_2^r f^{(r)}(\xi_2)$$

Estas dos últimas ecuaciones se dividen miembro a miembro y como  $f^{(r)}(\xi_1)$  y  $f^{(r)}(\xi_2)$  son prácticamente iguales, se tiene:

$$\frac{I - I_1}{I - I_2} = \frac{c h_1^r f^{(r)}(\xi_1)}{c h_2^r f^{(r)}(\xi_2)}$$

de donde:

$$I = \frac{h_1^r I_2 - h_2^r I_1}{h_1^r - h_2^r} \quad (6.20)$$

Si en particular  $h_2 = h_1/2$ , la ecuación 6.20 se simplifica a:

$$I \approx \frac{2^r I_2 - I_1}{2^r - 1} \quad (6.21)$$

Este proceso, conocido como integración de Romberg, es efectivo cuando  $f^{(r)}(x)$  no varía bruscamente en  $(a, b)$  y no cambia de signo en dicho intervalo. En estos casos, las ecuaciones 6.20 y 6.21 permiten obtener una mejor aproximación a  $I$  a partir de  $I_1$  y  $I_2$  sin repetir el proceso de integración y con cálculos breves.

En el método trapezoidal (véase Ec. 6.18); por ejemplo,  $r = 2$  y la ecuación 6.21 toma la forma

$$I \approx \frac{2^2 I_2 - I_1}{2^2 - 1} = \frac{4 I_2 - I_1}{3}$$

Para sistematizar la integración de Romberg en la aproximación trapezoidal, denótese por  $I_k^{(0)}$  las aproximaciones de  $I$  obtenidas empleando  $2^k$  trapezoides (véase tabla 6.1). Ahora, para obtener mejores aproximaciones de  $I$  mediante  $I_k^{(0)}$  y  $I_{k+1}^{(0)}$ , se aplica la extrapolación de Richardson

$$I \approx \frac{2^2 I_{k+1}^{(0)} - I_k^{(0)}}{2^2 - 1}$$

Este resultado se denota como  $I_k^{(1)}$  y se genera la cuarta columna de la tabla 6.1. Estos valores sirven para producir una segunda extrapolación y obtener una mejor aproximación de  $I$ . Con el empleo de  $I_k^{(1)}$  y  $I_{k+1}^{(1)}$  se llega a:

$$I \approx \frac{2^4 I_{k+1}^{(1)} - I_k^{(1)}}{2^4 - 1}$$

que se denota como  $I_k^{(2)}$ , con lo que se genera la quinta columna de la tabla 6.1. Este proceso puede continuar en tanto cada iteración responda al algoritmo:

$$I_k^{(m)} = \frac{4^m I_{k+1}^{(m-1)} - I_k^{(m-1)}}{4^m - 1}; m = 1, 2, 3, \dots$$

(6.22)

Cuando los valores de  $I_k^{(0)} \rightarrow I$  al crecer  $k$ , los valores de la diagonal superior de la tabla convergen\* a  $I$ .

Para entender mejor esto, a continuación se resuelve y analiza un ejemplo.

**Tabla 6.1** Aplicación del método de Romberg.

| $k$ | Número de trapezoides $2^k$ | Aproximación trapezoidal | Primera Extrapolación | Segunda Extrapolación | ... |
|-----|-----------------------------|--------------------------|-----------------------|-----------------------|-----|
| 0   | 1                           | $I_0^{(0)}$              |                       |                       |     |
| 1   | 2                           | $I_1^{(0)}$              | $I_0^{(1)}$           |                       |     |
| 2   | 4                           | $I_2^{(0)}$              | $I_1^{(1)}$           | $I_0^{(2)}$           |     |
| 3   | 8                           | $I_3^{(0)}$              | $I_2^{(1)}$           | $I_1^{(2)}$           | ... |
| 4   | 16                          | $I_4^{(0)}$              | $I_3^{(1)}$           | $I_2^{(2)}$           |     |
| .   | .                           | .                        | .                     | .                     |     |
| .   | .                           | .                        | .                     | .                     |     |
| .   | .                           | .                        | .                     | .                     |     |

\* Ralston, A. *Introducción al análisis numérico*. Limusa-Wiley, S.A. [1970], pp. 149-152.

**Ejemplo 6.8**

Encuentre una aproximación de la integral

$$\int_0^1 \sin \pi x \, dx,$$

empleando 1, 2, 4, 8 y 16 trapezoides.

Con los resultados obtenidos y la ecuación 6.22, obtenga mejores aproximaciones. Compare los valores obtenidos con el valor calculado analíticamente: 0.6366197.

**Solución**

Con el programa del ejemplo 6.6 se obtienen los valores



| $k$ | $2^k$ | $I_k^{(0)}$ |
|-----|-------|-------------|
| 0   | 1     | 0.0         |
| 1   | 2     | 0.5         |
| 2   | 4     | 0.6035534   |
| 3   | 8     | 0.6284174   |
| 4   | 16    | 0.6345731   |

Nótese que  $I_k^{(0)}$  converge al valor analítico al aumentar  $k$ ; sin embargo, emplear aún más subintervalos implica aumentar los errores de redondeo y un considerable incremento en el número de cálculos.

En cambio, si se aplica la ecuación 6.22 con  $m = 1$ , se obtiene sucesivamente

$$I_0^{(1)} = \frac{4^1 (0.5) - 0}{4^1 - 1} = 0.6666667$$

$$I_1^{(1)} = \frac{4^1 (0.6035534) - 0.5}{4^1 - 1} = 0.6380712$$

$$I_2^{(1)} = \frac{4^1 (0.6284174) - 0.6035534}{4^1 - 1} = 0.6367054$$

$$I_3^{(1)} = \frac{4^1 (0.6345731) - 0.6284174}{4^1 - 1} = 0.6366250$$

Hay que observar que con estos breves cálculos se obtienen mejores aproximaciones de la integral.

Al aplicar la ecuación 6.22 con  $m = 2$  y los valores de arriba

$$I_0^{(2)} = \frac{4^2 (0.6380712) - 0.6666667}{4^2 - 1} = 0.6361648$$

$$I_1^{(2)} = \frac{4^2 (0.6367054) - 0.6380712}{4^2 - 1} = 0.6366143$$

$$I_2^{(2)} = \frac{4^2 (0.6366250) - 0.63667054}{4^2 - 1} = 0.6366196$$

Al continuar con  $m = 3$  y  $m = 4$  se obtienen la sexta y séptima columnas de la tabla de resultados.



| $k$ | $2^k$ | $I_k^{(0)}$ | $I_k^{(1)}$ | $I_k^{(2)}$ | $I_k^{(3)}$ | $I_k^{(4)}$ |
|-----|-------|-------------|-------------|-------------|-------------|-------------|
| 0   | 1     | 0.0000000   | —           | —           | —           | —           |
| 1   | 2     | 0.5000000   | 0.6666667   | —           | —           | —           |
| 2   | 4     | 0.6035534   | 0.6380712   | 0.6361648   | —           | —           |
| 3   | 8     | 0.6284174   | 0.6367054   | 0.6366143   | 0.6366214   | —           |
| 4   | 16    | 0.6345731   | 0.6366250   | 0.6366196   | 0.6366197   | 0.6366197   |

El valor  $I_4^{(4)}$  es el valor analítico de la integral.

Los cálculos pueden realizarse con Matlab o con la TI-92 Plus.



```

a=0; b=1 ;
for k=0 : 4
    n=2^k; h= (b-a)/n; x=a; s=0 ;
    if n>1
        for j=1 : n-1
            x=x+h;
            f=sin(pi*x) ;
            s=f+s ;
        end
    end
    fa=sin(pi*a) ;
    fb=sin(pi*b) ;
    s=h/2*(fa+2*s+fb) ;
    I(k+1, 1) =s ;
    fprintf ( ' %4d  %10.7f\n', n, s)
end
for m=2:5
    for k=1 : 5-m+1
        I (k, m)=(4^(m-1)*I(k+1,m-1)-I(k,m-1))/(4^(m-1)-1);
    end
end
I

```



```

e6_8( )
Pgrm
Define f(x) =sin(n*x)
ClrIO : newMat(5, 5)→I
0→a : 1→b
For k, 0, 4
    2^k→n: (b-a)/n→h : a→x:0→s
    If n>1 Then
        For j, 1, n-1
            x+h→x : s+f(x)→s
        EndFor
    EndIf
    h/2* (f(a) +2 * s + f(b))→I [k+1,1]
    Disp format (n,"f0")&" "&format(I[k + 1, 1],"f7" )
EndFor
For m, 2, 5

```

```

For k, 1, 5-m+1
  (4^(m-1)*I[k+1,m-1]-I[k,m-1]) / (4^(m-1)-1) → I[k, m]
EndFor
Disp I
EndPrgm

```

El método de Romberg puede emplearse sucesivamente hasta que dos elementos consecutivos de una fila  $I_k^{(m)}$ ,  $I_k^{(m+1)}$  coincidan hasta cierta cifra decimal; esto es,

$$| I_k^{(m)} - I_k^{(m+1)} | \leq \text{EPS}$$

Además puede generarse otra columna y ver si

$$| I_{k+1}^{(m+2)} - I_k^{(m+2)} | \leq \text{EPS}$$

con lo que se evita la posibilidad de que dos elementos consecutivos de una fila coincidan entre sí, pero no con el valor de la integral que se está aproximando.

Utilice estos criterios para resolver el ejemplo 6.8 con  $\text{EPS} = 10^{-6}$ .

## 6.2 Cuadratura de Gauss

Gauss investigó y encontró que es factible disminuir el error en la integración cambiando la localización de los puntos sobre la curva de integración  $f(x)$ . El investigador desarrolló su propio método, conocido como **cuadratura de Gauss**, el cual se describe a continuación.

En la figura 6.7 se tiene la curva de la función  $f(x)$  que se desea integrar entre los límites  $a$  y  $b$ . La parte (a) de la figura muestra cómo se integraría usando un trapecioide: uniendo el punto A de coordenadas  $(a, f(a))$  con el punto B  $(b, f(b))$  mediante un segmento de recta  $p_1(x)$ . Esto forma un trapecioide de altura  $h = (b - a)$ , cuya área es:

$$T = \frac{h}{2} [f(a) + f(b)],$$

y que podría escribirse como\*

$$T = w_1 f(a) + w_2 f(b), \quad (6.23)$$

donde  $w_1 = w_2 = \frac{h}{2}$ .

El área del trapecioide calculada  $T$ , aproxima el área bajo la curva  $f(x)$ .

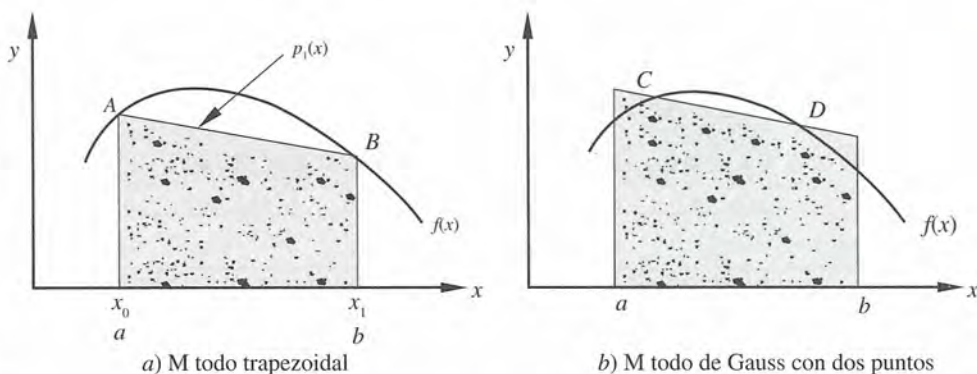
Por otro lado, en la aplicación de la cuadratura de Gauss, en lugar de tomar los dos puntos A y B en los extremos del intervalo, se escogen dos puntos interiores C y D (véase la parte b de la Fig. 6.7).

\* De hecho, cualquiera de las fórmulas de integración desarrolladas en las secciones anteriores puede ponerse en la forma.

$$\int_a^b f(x) dx \approx \sum_{i=1}^n w_i f(x_i)$$

donde, por ejemplo, la regla de Simpson aplicada una vez tendría  $w_1 = w_3 = h/3$  y  $w_2 = 4h/3$  (véase Ec. 6.4).

**Figura 6.7**  
Desarrollo del  
método de  
integración de  
Gauss usando  
dos puntos a  
partir del  
método  
trapezoidal.



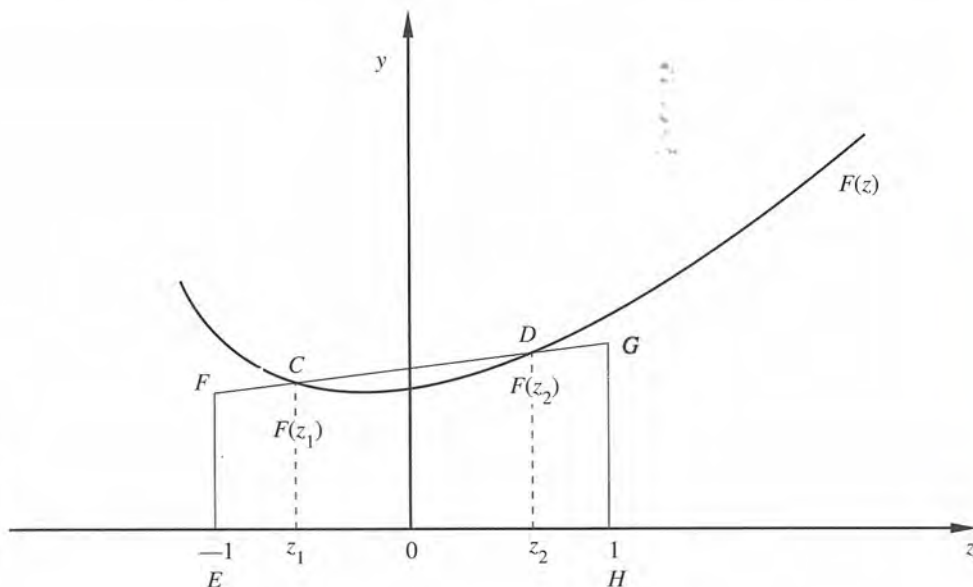
Se traza una línea recta por estos dos puntos, se extiende hasta los extremos del intervalo y se forma el trapecioide sombreado. Parte del trapecioide queda por encima de la curva y parte por abajo. Si se escogen adecuadamente los puntos C y D, cabe igualar las dos zonas de modo que el área del trapecioide sea igual al área bajo la curva; el cálculo del área del trapecioide resultante da la integral *exacta*. El método de Gauss consiste esencialmente en seleccionar los puntos C y D **adecuados**. La técnica se deduce a continuación.

Considérese primero, sin que esto implique perder generalidad, que se desea integrar la función mostrada en la figura 6.8 entre los límites  $-1$  y  $+1$ .<sup>\*</sup> Los puntos C y D se escogen sobre la curva y se forma el trapecioide con vértices E, F, G, H.

Sean las coordenadas del punto C( $z_1, f(z_1)$ ) y las del punto D( $z_2, f(z_2)$ ). Motivado por la fórmula trapezoidal (Ec. 6.3), Gauss se propuso desarrollar una fórmula del tipo

$$A = w_1 F(z_1) + w_2 F(z_2) \quad (6.24)$$

ya que esto simplificaría relativamente el cálculo del área. El problema planteado de esta manera, consiste en encontrar los valores de  $z_1, z_2, w_1$  y  $w_2$ . Entonces hay cuatro parámetros por



**Figura 6.8**  
Derivación del  
método de  
integración de  
Gauss.

<sup>\*</sup> Si los límites son distintos, se hace un cambio de variable para pasarlos a  $-1$  y  $+1$ .



determinar y, por tanto, cuatro condiciones que se pueden imponer. Éstas se eligen de manera que el método dé resultados exactos cuando la función por integrar sea alguna de las cuatro siguientes o combinaciones lineales de ellas.

$$F(z) = 1$$

$$F(z) = z$$

$$F(z) = z^2$$

$$F(z) = z^3$$

Los valores exactos de integrar estas cuatro funciones entre  $-1$  y  $+1$  son:

$$I_1 = \int_{-1}^1 1 \, dz = z \Big|_{-1}^1 = 1 - 1(-1) = 2$$

$$I_2 = \int_{-1}^1 z \, dz = \frac{z^2}{2} \Big|_{-1}^1 = \frac{1^2}{2} - \frac{(-1)^2}{2} = 0$$

$$I_3 = \int_{-1}^1 z^2 \, dz = \frac{z^3}{3} \Big|_{-1}^1 = \frac{1^3}{3} - \frac{(-1)^3}{3} = 2$$

$$I_4 = \int_{-1}^1 z^3 \, dz = \frac{z^4}{4} \Big|_{-1}^1 = \frac{1^4}{4} - \frac{(-1)^4}{4} = 0$$

Suponiendo que una ecuación de la forma 6.24 funciona exactamente, se tendría el siguiente sistema de ecuaciones

$$I_1 = w_1(1) + w_2(1) = 2$$

$$I_2 = w_1 z_1 + w_2 z_2 = 0$$

$$I_3 = w_1 z_1^2 + w_2 z_2^2 = 2/3$$

$$I_4 = w_1 z_1^3 + w_2 z_2^3 = 0$$

De la primera ecuación se tiene que  $w_1 + w_2 = 2$ ; nótese que también que si

$$w_1 = w_2$$

y

$$z_1 = -z_2,$$

se satisfacen la segunda y la cuarta ecuaciones. Entonces se elige

$$w_1 = w_2 = 1$$

y

$$z_1 = -z_2$$

y al sustituir en la tercera ecuación se obtiene

$$z_1^2 + (-z_1)^2 = 2/3$$

o bien

$$z_1^2 = 1/3$$

de donde

$$z_1 = \pm \frac{1}{\sqrt{3}} = \pm 0.57735...$$

y queda entonces:

$$z_1 = -0.57735\dots$$

$$z_2 = 0.57735\dots$$

Con lo que se tiene la fórmula

$$\int_{-1}^1 F(z) dz = w_1 F(z_1) + w_2 F(z_2) = F(-0.57735\dots) + F(+0.57735\dots) \quad (6.26)$$

que, salvo porque se tiene que calcular el valor de la función en un valor irracional de  $z$ , es tan simple como la regla trapezoidal; además, trabaja perfectamente para funciones cúbicas, mientras que la regla trapezoidal lo hace sólo para líneas rectas.

En páginas anteriores se comentó que para integrar en un intervalo distinto de  $[-1, 1]$ , se requiere un cambio de variable a fin de pasar del intervalo de integración general  $[a, b]$  a  $[-1, 1]$  y así aplicar la ecuación 6.26; por ejemplo, si se desea obtener

$$\int_0^5 e^{-x} dx$$

se puede cambiar a  $z = \frac{2}{5}x - 1$ , de modo que si  $x = 0$ ,  $z = -1$  y si  $x = 5$ ,  $z = 1$ .

El resto de la integral se pone en términos de la nueva variable  $z$  y se encuentra que

$$e^{-x} = e^{-5(z+1)/2}$$

y

$$dx = d\left(\frac{5}{2}(z+1)\right) = \frac{5}{2} dz$$

entonces la integral queda:

$$\int_0^5 e^{-x} dx = \frac{5}{2} \int_{-1}^1 e^{-5(z+1)/2} dz,$$

de modo que las condiciones de aplicación del método de Gauss quedan satisfechas.

Al resolver se tiene

$$\begin{aligned} \frac{5}{2} \int_{-1}^1 e^{-5(z+1)/2} dz &\approx \frac{5}{2} [w_1 F(-0.57735\dots) + w_2 F(+0.57735\dots)] \\ &= \frac{5}{2} [(1) e^{-5(-0.57735+1)/2} + (1) e^{-5(-0.57735+1)/2}] = 0.91752 \end{aligned}$$

Esto es:

$$\int_0^5 e^{-x} dx \approx 0.91752$$

El valor exacto de esta integral es 0.99326.

En general, si se desea calcular  $\int_a^b f(x) dx$  aplicando la ecuación 6.26, se cambia el intervalo de integración con la siguiente fórmula\*:

$$z = \frac{2x - (a+b)}{b-a} \quad (6.27)$$

ya que si  $x = a$ ,  $z = -1$ ; y si  $x = b$ ,  $z = 1$

El integrando  $f(x) dx$  en términos de la nueva variable queda:

$$f(x) = F\left(\frac{b-a}{2}z + \frac{a+b}{2}\right)$$

y

$$dx = d\left(\frac{b-a}{2}z + \frac{a+b}{2}\right) = \frac{b-a}{2} dz$$

\* Sólo es aplicable cuando los límites de integración  $a$  y  $b$  son finitos.

Por lo que la integral queda finalmente como:

$$\begin{aligned} \int_a^b f(x) dx &= \frac{b-a}{2} \int_{-1}^1 F\left(\frac{b-a}{2}z + \frac{a+b}{2}\right) dz \\ &\approx \frac{b-a}{2} \left[ F\left(\frac{b-a}{2}(-0.57735) + \frac{a+b}{2}\right) + F\left(\frac{b-a}{2}(+0.57735) + \frac{a+b}{2}\right) \right] \end{aligned} \quad (6.28)$$

Una cuestión importante es que el método de Gauss puede extenderse a tres o más puntos; por ejemplo, si se escogen tres puntos *no equidistantes* en el segmento de la curva  $f(z)$  comprendida entre  $-1$  y  $1$ , se podría pasar una parábola por los tres como en la regla de Simpson, excepto en que dichos puntos se escogerían de modo que minimicen o anulen el error. Similarmente es factible elegir cuatro puntos y una curva cuadrática, cinco puntos y una curva cúbica, etc. En general, el algoritmo tiene la forma.

$$\int_{-1}^1 F(z) dz = A \approx w_1 F(z_1) + w_2 F(z_2) + w_3 F(z_3) + \dots + w_n F(z_n) \quad (6.29)$$

donde se han calculado los valores  $w_i$  y  $z_i$  por usar y la tabla 6.2 da valores hasta para seis puntos.

Con dos puntos, el método de Gauss está diseñado para obtener exactitud en polinomios cúbicos; con tres, se tendrá exactitud en polinomios de cuarto grado y así sucesivamente.

Los coeficientes y abscisas dadas en la tabla 6.2 sirven para integrar sobre todo el intervalo de interés, o bien puede dividirse el intervalo en varios subintervalos (como en los métodos compuestos de integración) y aplicar el método de Gauss a cada uno de ellos.

**Tabla 6.2** Coeficientes y abscisas en el método de la cuadratura de Gauss Legendre.

| Número de puntos | Coeficientes $w_i$         | Abscisas $z_i$              |
|------------------|----------------------------|-----------------------------|
| 2                | $w_1 = w_2 = 1.0$          | $-z_1 = z_2 = 0.5773502692$ |
| 3                | $w_2 = 0.88888\dots$       | $-z_1 = z_3 = 0.7745966692$ |
|                  | $w_1 = w_3 = 0.55555\dots$ | $z_2 = 0.0$                 |
| 4                | $w_1 = w_4 = 0.3478548451$ | $-z_1 = z_4 = 0.8611363116$ |
|                  | $w_2 = w_3 = 0.6521451549$ | $-z_2 = z_3 = 0.3399810436$ |
| 5                | $w_1 = w_5 = 0.2369268851$ | $-z_1 = z_5 = 0.9061798459$ |
|                  | $w_2 = w_4 = 0.4786286705$ | $-z_2 = z_4 = 0.5384693101$ |
|                  | $w_3 = 0.56888\dots$       | $z_3 = 0.0$                 |
| 6                | $w_1 = w_6 = 0.1713244924$ | $-z_1 = z_6 = 0.9324695142$ |
|                  | $w_2 = w_5 = 0.3607615730$ | $-z_2 = z_5 = 0.6612093865$ |
|                  | $w_3 = w_4 = 0.4679139346$ | $-z_3 = z_4 = 0.2386191861$ |



**Ejemplo 6.9**

Integre la función  $\frac{1}{\sqrt{2\pi}} e^{-x^2/2}$  en el intervalo  $(-0.8, 1.5)$  por cuadratura de Gauss.

**Solución**

a) Con dos puntos

Cambio de límites de la integral con la ecuación

$$z = \frac{2x - (a + b)}{b - a} = \frac{2x - 0.7}{2.3}$$

Si  $x = -0.8$ ,  $z = -1$ ; si  $x = 1.5$ ,  $z = 1$ .

Con el cambio de la función en términos de la nueva variable  $z$  queda:

$$\begin{aligned} I &= \frac{1}{\sqrt{2\pi}} \int_{-0.8}^{1.5} e^{-x^2/2} dx \\ &= \frac{1}{\sqrt{2\pi}} \int_{-1}^1 \left[ \frac{1.5 - (-0.8)}{2} \right] e^{-\left[ \frac{1.5 - (-0.8)}{2} z + \frac{-0.8 + 1.5}{2} \right]^2 / 2} dz \\ &= \frac{2.3}{\sqrt{2\pi}} \int_{-1}^1 e^{-(2.3z + 0.7)^2 / 8} dz \end{aligned}$$

De la tabla 6.2  $w_1 = w_2 = 1.0$ ;  $-z_1 = z_2 = 0.5773502692$ .

Al evaluar la función del integrando en  $z_1, z_2$

$$F(0.5773502692) = e^{-[2.3(-0.5773502692) + 0.7]^2/8} = 0.5980684$$

$$F(-0.5773502692) = e^{-[2.3(0.5773502692) + 0.7]^2/8} = 0.9519115$$

Se aplica la ecuación 6.29

$$I = \frac{2.3}{2\sqrt{2\pi}} [1(+0.5980684) + 1(0.9519115)] = 0.711105$$

b) Con tres puntos

De la tabla 6.2

$$w_1 = w_3 = 0.55555..., \quad w_2 = 0.88888...$$

$$-z_1 = z_3 = 0.7745966692, \quad z_2 = 0.0$$

Al evaluar la función del integrando en  $z_1, z_2$  y  $z_3$  y emplear la ecuación 6.29 se tiene:

$$\begin{aligned} I &\approx \frac{2.3}{2\sqrt{2\pi}} [(0.55555...) (0.4631...) + (0.88888...) (0.9405...) \\ &\quad + (0.55555...) (0.8639...)] = 0.721825 \end{aligned}$$

**Ejemplo 6.10**

Halle  $\int_0^{2\pi} \sin x \, dx$ , por el método de la cuadratura de Gauss utilizando tres puntos.

**Solución**

Se cambian variables y límites de integración con la expresión

$$z = \frac{2x - (a + b)}{b - a}$$

como  $a = 0$ ,  $b = 2\pi$ , entonces

$$z = \frac{2x - 2\pi}{2\pi} = \frac{x - 2\pi}{\pi}$$

se despeja  $x : x = \pi z + \pi$ , de donde  $dx = \pi dz$

Se sustituye en la integral

$$\int_0^{2\pi} \sin x \, dx = \int_{-1}^1 \sin(\pi z + \pi) \pi \, dz = \pi \int_{-1}^1 \sin(\pi z + \pi) \, dz$$

Con el empleo de la ecuación 6.29 con  $n = 3$  y los valores de la tabla 6.2 queda:

$$\begin{aligned} A \approx \pi \{ & (0.55555\dots)[\sin(\pi(-0.7745966692) + \pi)] \\ & + (0.88888\dots)[\sin(\pi(0) + \pi)] \\ & + (0.55555\dots)[\sin(\pi(0.7745966692) + \pi)] \} \end{aligned}$$

Se deja al lector la comparación de este resultado con la solución analítica.

La expresión 6.29 puede ponerse en forma más general y adecuada para programarla, así:

$$\int_a^b f(x) \, dx = \frac{b-a}{2} \sum_{i=1}^n w_i F\left[\frac{(b-a)z_i + b + a}{2}\right] \quad (6.30)$$

la cual puede deducirse de los ejemplos resueltos (ver problema 6.21).

A continuación se presenta un algoritmo para la cuadratura de Gauss-Legendre.

### ALGORITMO 6.3 Cuadratura de Gauss-Legendre

Para aproximar el área bajo la curva de una función analítica  $f(x)$  en el intervalo  $[a, b]$ , proporcionar la función a integrar  $F(X)$  y los

DATOS: El número de puntos (2, 3, 4, 5 o 6) por utilizar: N, el límite inferior A y el límite superior B.

RESULTADOS: El área aproximada AREA.

- PASO 1. Hacer  $(NP(I), I = 1, 2, \dots, 5) = (2, 3, 4, 5, 6)$ .
- PASO 2. Hacer  $(IAUX(I), I = 1, 2, \dots, 6) = (1, 2, 4, 6, 9, 12)$ .
- PASO 3. Hacer  $(Z(I), I = 1, 2, \dots, 11) = (0.577350269, 0.0, 0.774596669, 0.339981044, 0.861136312, 0.0, 0.538469310, 0.906179846, 0.238619186, 0.661209387, 0.932469514)$ .
- PASO 4. Hacer  $(W(I), I = 1, 2, \dots, 11) = (1.0, 0.888888888, 0.555555555, 0.652145155, 0.347854845, 0.568888888, 0.478628671, 0.236926885, 0.467913935, 0.360761573, 0.171324493)$ .
- PASO 5. Hacer  $I = 1$ .
- PASO 6. Mientras  $I \leq 5$ , repetir los pasos 7 y 8.
- PASO 7. Si  $N = NP(I)$ , ir al paso 10. De otro modo continuar.
- PASO 8. Hacer  $I = I + 1$ .
- PASO 9. IMPRIMIR "N NO ES 2, 3, 4, 5 o 6" y TERMINAR.
- PASO 10. Hacer  $S = 0$ .
- PASO 11. Hacer  $J = IAUX(I)$ .
- PASO 12. Mientras  $J \leq IAUX(I+1) - 1$ , repetir los pasos 13 a 17.
- PASO 13. Hacer  $ZAUX = (Z(J) * (B - A) + B + A) / 2$ .
- PASO 14. Hacer  $S = S + F(ZAUX) * W(J)$ .
- PASO 15. Hacer  $ZAUX = (-Z(J) * (B - A) + B + A) / 2$ .
- PASO 16. Hacer  $S = S + F(ZAUX) * W(J)$ .
- PASO 17. Hacer  $J = J + 1$ .
- PASO 18. Hacer  $ÁREA = (B - A) / 2 * S$ .
- PASO 19. IMPRIMIR AREA Y TERMINAR.

**Ejemplo 6.11**

Elabore un programa que integre funciones analíticas con la cuadratura de Gauss-Legendre usando 2, 3, 4, 5, o 6 puntos, mismos que usted elegirá. Pruebe el programa con la función del ejemplo 6.8.

**Solución** La expresión general de Gauss-Legendre para integrar es:



$$\int_a^b f(x) dx \approx \frac{b-a}{2} \sum_{i=1}^n w_i F\left[\frac{(b-a)z_i + b + a}{2}\right]$$

donde  $w_i, z_i, i = 1, 2, \dots, n$ , están dados en la tabla 6.2

En el disco se encuentra el programa 6.3 solicitado.

## 6.3 Integrales múltiples

Cualquiera de las técnicas de integración estudiadas en este capítulo es modificable, de modo que se puede aplicar en la aproximación de integrales dobles o triples. A continuación se ilustra el método de Simpson 1/3 en la solución de integrales dobles

$$a) \int_0^\pi \int_0^3 y \sin x \, dx \, dy \quad y \quad b) \int_1^3 \int_0^4 e^{x+y} \, dx \, dy$$

Para la integral del inciso a), se divide el intervalo  $[a, b] = [0, 3]$  en  $n = 6$  subintervalos iguales, con lo que la amplitud de cada subintervalo es igual a:

$$h_1 = \frac{3-0}{6} = 0.5$$

y se aplica la regla de Simpson compuesta a la integral interna, manteniendo constante la variable  $y$  (nótese que se está integrando en el eje  $x$ ).

$$\begin{aligned} \int_0^\pi \int_0^3 y \sin x \, dx \, dy &\approx \int_0^\pi \frac{h_1}{3} [y \sin 0 + 4y (\sin 0.5 + \sin 1.5 + \sin 2.5) + \\ &\quad 2y (\sin 1 + \sin 2) + y \sin 3] \, dy \\ &\approx \int_0^\pi 1.9907 y \, dy \end{aligned}$$

Ahora se integra en el eje  $y$ . El intervalo  $[c, d] = [0, \pi]$  se divide en  $m = 8$  subintervalos, por ejemplo, y queda:

$$h_2 = \frac{\pi-0}{8} = \frac{\pi}{8}$$

y

$$\begin{aligned} 1.9907 \int_0^\pi y \, dy &\approx 1.9907 \frac{h_2}{3} \left[ 0 + 4 \left( \frac{\pi}{8} + \frac{3\pi}{8} + \frac{5\pi}{8} + \frac{7\pi}{8} \right) + \right. \\ &\quad \left. 2 \left( \frac{2\pi}{8} + \frac{4\pi}{8} + \frac{6\pi}{8} \right) + \frac{8\pi}{8} \right] \\ &\approx 9.82373 \end{aligned}$$

entonces se tiene:

$$\int_0^\pi \int_0^3 y \sin x \, dx \, dy \approx 9.82373$$

Hay que observar que se ha efectuado una integración repetida; esto es, se ha integrado siguiendo el proceso

$$\int_0^\pi \int_0^3 y \sin x \, dx \, dy = \int_0^\pi \left( \int_0^3 y \sin x \, dx \right) dy$$



en la primera integración y se mantuvo constante. Es importante recordar que la integración iterada puede llevarse a cabo con respecto a  $y$  primero y después respecto a  $x$ , pero intercambiando los límites de integración. Esto se indica

$$\int_0^3 \int_0^\pi y \sin x \, dy \, dx,$$

y el resultado es el mismo (véase Prob. 6.30).

Para la integral (b), el intervalo  $[a, b] = [0, 4]$  se divide en  $n = 4$  subintervalos, de donde

$$h_1 = \frac{4 - 0}{4} = 1$$

$$\int_1^3 \int_0^4 e^{x+y} \, dx \, dy = \int_1^3 \frac{1}{3} [e^{0+y} + 4(e^{1+y} + e^{3+y}) + 2e^{2+y} + e^{4+y}] \, dy$$

cuya integración por Simpson 1/3 con  $m = 6$  en el eje  $y$  da con:

$$h_2 = \frac{3 - 1}{6} = \frac{1}{3}$$

$$\begin{aligned} \int_1^3 \int_0^4 e^{x+y} \, dx \, dy &= \frac{1}{3} \frac{1}{3(3)} [e^1 + 4(e^{4/3} + e^2 + e^{6/3}) + 2(e^{5/3} + e^{7/3}) + e^3] + \\ &\frac{1}{3} \frac{1}{3(3)} [e^{0.5+1} + 4(e^{0.5+4/3} + e^{0.5+2} + e^{0.5+8/3}) + \\ &2(e^{0.5+5/3} + e^{0.5+7/3}) + e^{0.5+3}] + \\ &\frac{1}{3} \frac{1}{3(3)} [e^{1.5+1} + 4(e^{1.5+4/3} + e^{1.5+2} + e^{1.5+8/3}) + \\ &2(e^{1.5+5/3} + e^{1.5+7/3}) + e^{1.5+3}] + \\ &\frac{1}{3} \left( \frac{1}{3} [e^{1+1} + 4(e^{1+4/3} + e^{1+2} + e^{1+8/3}) + \right. \\ &\quad \left. 2(e^{1+5/3} + e^{1+7/3}) + e^{1+3}] \right) + \\ &\frac{1}{3} \left( \frac{1(0.5)}{3} [e^{2+1} + 4(e^{2+4/3} + e^{2+2} + e^{2+8/3}) + \right. \\ &\quad \left. 2(e^{2+5/3} + e^{2+7/3}) + e^{2+3}] \right) \end{aligned}$$

$$\approx 935.53 \text{ (el resultado analítico es } 930.853)$$

En general, la integración de una función  $f(x, y)$  sobre una región  $R$  del plano  $x - y$  dada así:  $\{(x, y): a \leq x \leq b, c \leq y \leq d\}$  por el método de Simpson 1/3 es:

$$\begin{aligned} \int_c^d \int_a^b f(x, y) \, dx \, dy &= \int_c^d \left[ \int_a^b f(x, y) \, dx \right] dy \\ &\approx \int_c^d \left( \frac{h_1}{3} [f(x_0, y) + 4 \sum_{\substack{i=1 \\ \Delta i=2}}^{n-1} f(x_i, y) + 2 \sum_{\substack{i=2 \\ \Delta i=2}}^{n-2} f(x_i, y) \right. \\ &\quad \left. + f(x_n, y)] \right) dy \end{aligned}$$

donde  $h_1 = \frac{b-a}{n}$ . Desarrollando se tiene:

$$\int_c^d \int_a^b f(x, y) \, dx \, dy \approx \frac{h_1}{3} \int_c^d f(x_0, y) \, dy + \frac{4}{3} h_1 \sum_{\substack{i=1 \\ \Delta i=2}}^{n-1} \int_c^d f(x_i, y) \, dy$$

$$+ \frac{2h_1}{3} \sum_{i=2}^{n-2} \int_c^d f(x_i, y) dy + \frac{h_1}{3} \int_c^d f(x_n, y) dy,$$

e integrando nuevamente por Simpson 1/3 con  $h_2 = \frac{d-c}{m}$

$$\begin{aligned} \int_c^d \int_a^b f(x, y) dx dy &\approx \\ &\frac{h_1}{3} \frac{h_2}{3} \left[ f(x_0, y_0) + 4 \sum_{j=1}^{m-1} f(x_0, y_j) + 2 \sum_{j=2}^{m-2} f(x_0, y_j) + f(x_0, y_m) \right] \\ &+ \frac{4h_1}{3} \frac{h_2}{3} \sum_{i=1}^{n-1} \left[ f(x_i, y_0) + 4 \sum_{j=1}^{m-1} f(x_i, y_j) + 2 \sum_{j=2}^{m-2} f(x_i, y_j) + f(x_i, y_m) \right] \\ &+ \frac{2h_1}{3} \frac{h_2}{3} \sum_{i=2}^{n-2} \left[ f(x_i, y_0) + 4 \sum_{j=1}^{m-1} f(x_i, y_j) + 2 \sum_{j=2}^{m-2} f(x_i, y_j) + f(x_i, y_m) \right] \\ &+ \frac{h_1}{3} \frac{h_2}{3} \left[ f(x_n, y_0) + 4 \sum_{j=1}^{m-1} f(x_n, y_j) + 2 \sum_{j=2}^{m-2} f(x_n, y_j) + f(x_n, y_m) \right] \\ &\approx h_1 \frac{h_2}{9} \left[ f(x_0, y_0) + f(x_0, y_m) + f(x_n, y_0) + f(x_n, y_m) + \right. \\ &\quad + 4 \sum_{j=1}^{m-1} (f(x_0, y_j) + f(x_n, y_j)) + 2 \sum_{j=2}^{m-2} (f(x_0, y_j) + f(x_n, y_j)) \\ &\quad + 4 \sum_{i=1}^{n-1} (f(x_i, y_0) + 4 \sum_{j=1}^{m-1} f(x_i, y_j) + 2 \sum_{j=2}^{m-2} f(x_i, y_j) + f(x_i, y_m)) \\ &\quad \left. + 2 \sum_{i=2}^{n-2} (f(x_i, y_0) + 4 \sum_{j=1}^{m-1} f(x_i, y_j) + 2 \sum_{j=2}^{m-2} f(x_i, y_j) + f(x_i, y_m)) \right] \quad (6.31) \end{aligned}$$

Igualmente puede emplearse la cuadratura de Gauss para integrales dobles o triples. Así, en el caso general

$$\int_c^d \int_a^b f(x, y) dx dy$$

primero se cambian los intervalos de  $x$  y  $y$  a  $[-1, 1]$  con las fórmulas

$$t = \frac{2x - (a + b)}{b - a} \quad y \quad u = \frac{2y - (c + d)}{d - c}$$

y asimismo se cambian  $dx$ ,  $dy$  y  $f(x, y)$  a términos de las nuevas variables  $t$  y  $u$ . Para esto, se despeja  $x$

$$x = \frac{(b-a)}{2} t + \frac{(b+a)}{2} \quad \text{de donde} \quad dx = \frac{(b-a)}{2} dt$$

y después  $y$

$$y = \frac{(d-c)u}{2} + \frac{(c+d)}{2} \quad \text{de donde} \quad dy = \frac{(d-c)}{2} du$$

se sustituye

$$\int_c^d \int_a^b f(xy) dx dy = \frac{(b-a)(d-c)}{4} \int_{-1}^1 \int_{-1}^1 f \left[ \frac{(b-a)}{2} t + \frac{(a+b)}{2}, \frac{(d-c)}{2} u + \frac{(c+d)}{2} \right] dt du \quad (6.32)$$

a la cual cabe aplicar la fórmula 6.30. Para ilustrar esto, a continuación se resuelve la integral

$$\int_2^3 \int_0^4 e^{x+y} dx dy$$

empleando dos puntos.

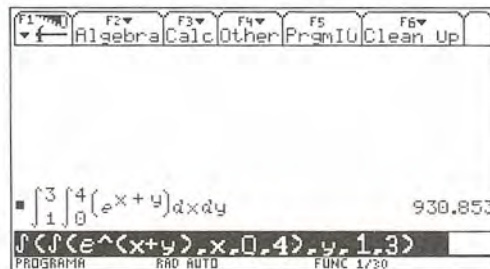
Primero se sustituye e integra respecto a  $t$

$$\begin{aligned} \int_1^3 \int_0^4 e^{x+y} dx dy &\approx \frac{(4-0)(3-1)}{4} \int_{-1}^1 \int_{-1}^1 e^{(2t+2)+(u+2)} dt du \\ &\approx 2 \int_{-1}^1 [e^{2(-0.57735)+4+u} + e^{2(0.57735)+4+u}] du \end{aligned}$$

Ahora se integra respecto a  $u$

$$\begin{aligned} \int_1^3 \int_0^4 e^{x+y} dx dy &\approx 2 [e^{2.84530-0.57735} + e^{2.84530+0.57735} \\ &\quad + e^{5.15470-0.57735} + e^{5.15470+0.57735}] = 892.335 \end{aligned}$$

La TI-92 Plus permite llevar a cabo integrales dobles o triples. Se muestra enseguida el cálculo de la integral que nos ocupa.



Para mayor sencillez y facilidad de programación, es conveniente emplear la fórmula

$$\int_c^d \int_a^b f(x, y) dx dy \approx \frac{(b-a)(d-c)}{4} \sum_{j=1}^m \sum_{i=1}^n w_j w_i F \left[ \frac{b-a}{2} t_i + \frac{b+a}{2}, \frac{d-c}{2} u_j + \frac{c+d}{2} \right] \quad (6.33)$$

donde  $n$  y  $m$  son los números de puntos por usar en los ejes  $x$  y  $y$ , respectivamente. Su aplicación a la integral del inciso a) empleando tres puntos en ambos ejes conduce a

$$\int_1^3 \int_0^4 e^{x+y} dx dy \approx \frac{(4-0)(3-1)}{4} \sum_{j=1}^3 \sum_{i=1}^3 w_j w_i F(2t_i + 2, u_j + 2)$$

donde  $w_1, w_2$  y  $w_3$  y  $t_1 = u_1 = z_1$ ,  $t_2 = u_2 = z_2$  y  $t_3 = u_3 = z_3$  están dados en la tabla 6.2.

Al sustituir valores se tiene:

$$\int_1^3 \int_0^4 e^{x+y} dx dy \approx 934.39$$

La solución analítica de esta integral es 930.85.



Hasta ahora sólo se han visto integraciones dobles sobre regiones  $R$  rectangulares. No obstante, también pueden resolverse integrales del tipo

$$\int_c^d \int_{a(y)}^{b(y)} f(x, y) dx dy$$

o del tipo

$$\int_a^b \int_{c(x)}^{d(x)} f(x, y) dy dx$$

cuyas regiones  $R_1$  y  $R_2$  quedan dadas como se muestra en la figura 6.9 a y b.

A continuación se resuelve por el método de Simpson 1/3 la integral

$$\int_0^2 \int_{x^2}^{2x} (x^3 + 4y) dy dx$$

que representa el área sombreada de la figura 6.10.

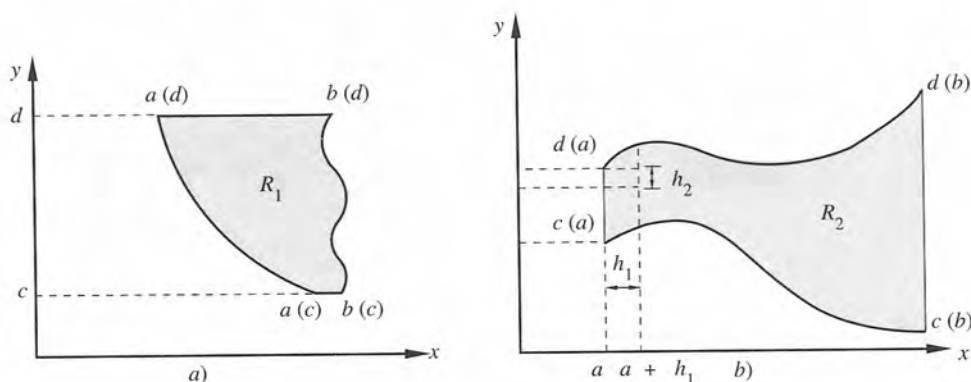
El intervalo  $[a, b] = [0, 2]$  se divide en, por ejemplo, dos subintervalos y queda  $h_1 = (2 - 0)/2 = 1$ ; el tamaño de paso en el eje  $y$  varía con  $x$  de acuerdo con la expresión

$$h_2(x) = \frac{d(x) - c(x)}{m}$$

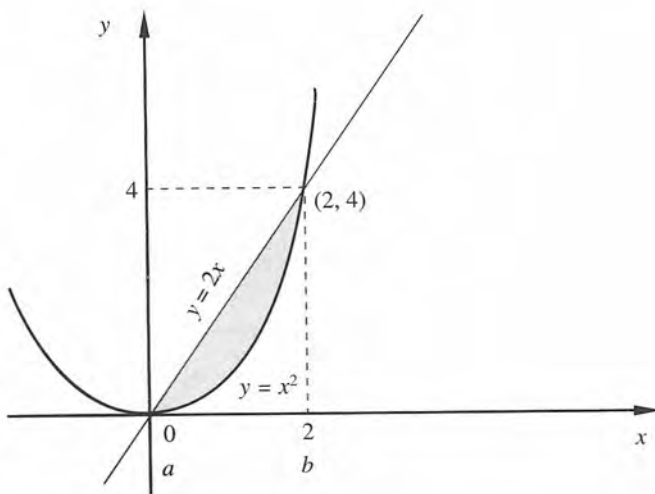
donde  $m$  es el número de subintervalos en que se divide el eje  $y$ .

Si se hace  $m = 2$ , se tiene

**Figura 6.9**  
Regiones no  
rectangulares  
de integración.



**Figura 6.10**  
Región de  
integración  
delimitada por  
una recta y una  
parábola.



$$\begin{aligned}
\int_0^2 \int_x^{2x} (x^3 + 4y) dy dx &\approx \int_0^2 \left( \frac{h_2(x)}{3} [x^3 + 4x^2 + 4(x^3 + 4(x^2 + h_2(x))) + x^3 + 4(2x)] \right) dx \\
&\approx \int_0^2 \frac{h_2(x)}{3} [6x^3 + 20x^2 + 8x + 16h_2(x)] dx \\
&= \frac{h_1}{3} \left( \frac{h_2(0)}{3} [6(0)^3 + 20(0)^2 + 8(0) + 16h_2(0)] \right. \\
&\quad + \frac{4h_2(0+1)}{3} [6(1)^3 + 20(1)^2 + 8(1) + 16h_2(1)] \\
&\quad \left. + \frac{h_2(2)}{3} [6(2)^3 + 20(2)^2 + 8(2) + 16h_2(2)] \right)
\end{aligned}$$

ya que

$$h_2(0) = \frac{2(0) - 0^2}{2} = 0, \quad h_2(1) = \frac{2(1) - 1^2}{2} = 0.5, \text{ y}$$

$$h_2(2) = \frac{2(2) - 2^2}{2} = 0,$$

$$\int_0^2 \int_x^{2x} (x^3 + 4y) dy dx \approx 9.33$$

Si se divide el intervalo  $[a, b]$  en cuatro subintervalos y se mantiene  $m = 2$ , se tiene  $h_1 = (2 - 0)/4 = 0.5$ . Entonces, la integración queda como sigue

$$\begin{aligned}
\int_0^2 \int_x^{2x} (x^3 + 4y) dy dx &\approx \frac{0.5}{3} \left( \frac{4h_2(0.5)}{3} [6(0.5)^2 + 20(0.5)^3 + 8(0.5) + 16h_2(0.5)] \right. \\
&\quad + \frac{2h_2(1)}{3} [6(1)^3 + 20(1)^2 + 8(1) + 16h_2(1)] \\
&\quad \left. + \frac{4h_2(1.5)}{3} [6(1.5)^3 + 20(1.5)^2 + 8(1.5) + 16h_2(1.5)] \right)
\end{aligned}$$

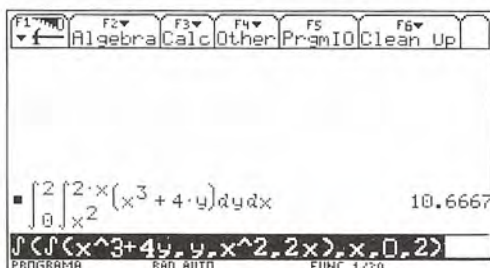
ya que  $h_2(0) = 0, h_2(2) = 0, h_2(0.5) = \frac{2(0.5) - 0.5^2}{2} = 0.375, \text{ y}$

$$h_2(1) = 0.5, h_2(1.5) = \frac{2(1.5) - 1.5^2}{2} = 0.375$$

Al sustituir valores se obtiene

$$\int_0^2 \int_x^{2x} (x^3 + 4y) dy dx \approx 10.583$$

Se muestra enseguida el cálculo de esta integral con la TI92 Plus.



**ALGORITMO 6.4** Integración doble por Simpson 1/3

Para aproximar  $\int_a^b \int_{c(x)}^{d(x)} f(x, y) dy dx$ , proporcionar las funciones  $C(X)$ ,  $D(X)$  y  $F(X, Y)$  y los

**DATOS:** El número  $N$  de subintervalos a usar en el eje  $x$ , el número  $M$  de subintervalos por emplear en el eje  $y$ , el límite inferior  $A$  y el límite superior  $B$ .

**RESULTADOS:** El área aproximada  $AREA$ .

- PASO 1. Hacer  $H1 = (B - A)/N$
- PASO 2. Hacer  $S = (F(A, C(A)) + F(A, D(A))) * (D(A) - C(A))/M$   
 $+ (F(B, C(B)) + F(B, D(B))) * (D(B) - C(B))/M$
- PASO 3. Hacer  $S1 = 0$ ;  $S2 = 0$ ;  $Y1 = C(A)$ ;  $Y2 = C(B)$
- PASO 4. Hacer  $J = 1$
- PASO 5. Mientras  $J \leq M - 1$ , repetir los pasos 6 a 9.
- PASO 6. Hacer  $H2A = (D(A) - C(A))/M$ ;  
 $Y1 = Y1 + H2A$ ;  
 $S1 = S1 + H2A * F(A, Y1)$ ;  
 $H2B = (D(B) - C(B))/M$ ;  
 $Y2 = Y2 + H2B$ ;  $S1 = S1 + H2B * F(B, Y2)$
- PASO 7. SI  $J = M - 1$ , ir al paso 9. De otro modo continuar.
- PASO 8. Hacer  $Y1 = Y1 + H2A$ ;  
 $S2 = S2 + H2A * F(A, Y1)$ ;  
 $Y2 = Y2 + H2B$ ;  $S2 = S2 + H2B * F(B, Y2)$
- PASO 9. Hacer  $J = J + 2$
- PASO 10. Hacer  $S3 = 0$ ;  $S6 = 0$ ;  $S7 = 0$ ;  $X = A$
- PASO 11. Hacer  $I = 1$
- PASO 12. Mientras  $I \leq N - 1$ , repetir los pasos 13 a 16.
- PASO 13. Hacer  $X = X + H1$ ;  $H2 = (D(X) - C(X))/M$ ;  
 $S3 = S3 + H2 * F(X, C(X))$ ;  
 $S6 = S6 + H2 * F(X, D(X))$
- PASO 14. SI  $I = N - 1$ , ir al paso 16. De otro modo continuar.
- PASO 15. Hacer  $X = X + H1$ ;  $H2 = (D(X) - C(X))/M$ ;  
 $S7 = S7 + 2 * (F(X, C(X)) + F(X, D(X)))$
- PASO 16. Hacer  $I = I + 2$
- PASO 17. Hacer  $S4 = 0$ ;  $S5 = 0$ ;  $S8 = 0$ ;  $S9 = 0$ ;  $X = A - H1$
- PASO 18. Hacer  $I = 1$
- PASO 19. Mientras  $I \leq N - 1$ , repetir los pasos 20 a 31.
- PASO 20. Hacer  $X = X + 2 * H1$ ;  
 $Y1 = C(X)$ ;  $Y2 = C(X + H1)$ ;  
 $HA = (D(X) - C(X))/M$ ;  
 $HB = (D(X + H1) - C(X + H1))/M$ ;
- PASO 21. Hacer  $J = 1$ .
- PASO 22. Mientras  $J \leq M - 1$ , repetir los pasos 23 a 30.
- PASO 23. Hacer  $Y1 = Y1 + HA$ ;  
 $S4 = S4 + HA * F(X, Y1)$
- PASO 24. SI  $I = N - 1$ , ir al paso 26. De otro modo continuar.
- PASO 25. Hacer  $Y2 = Y2 + HB$ ;  
 $S8 = S8 + HB * F(X + H1, Y2)$
- PASO 26. SI  $J = M - 1$ , ir al paso 30. De otro modo continuar.
- PASO 27. Hacer  $Y1 = Y1 + HA$ ;  
 $S5 = S5 + HA * F(X, Y1)$
- PASO 28. SI  $I = N - 1$ , ir al paso 30. De otro modo continuar.
- PASO 29. Hacer  $Y2 = Y2 + HB$ ;  $S9 = S9 + HA * F(X + H1, Y2)$
- PASO 30. Hacer  $J = J + 2$
- PASO 31. Hacer  $I = I + 2$
- PASO 32. Hacer  $AREA = H1/9 * (S + 4 * (S1 + S3 + S6 + S9) + 2 * (S2 + S7) + 16 * S4 + 8 * (S5 + S8))$
- PASO 33. IMPRIMIR  $AREA$  y TERMINAR.



## 6.4 Diferenciación numérica

En la introducción del capítulo 5 se comentó que cuando se va a practicar una operación en una función tabulada, el camino es aproximar la tabla por alguna función y efectuar la operación en la función aproximante. Así se procedió en la integración numérica y así se procederá en la diferenciación numérica; esto es, se aproximará la función tabulada  $f(x)$  y se diferenciará la aproximación  $p_n(x)$ .

Si la aproximación es polinomial y con el criterio de **ajuste exacto**<sup>\*</sup>, la diferenciación numérica consiste simplemente en diferenciar la fórmula del polinomio interpolante que se utilizó. Sea en general.

$$f(x) = p_n(x) + R_n(x)$$

y la aproximación de la primera derivada queda entonces

$$\frac{df(x)}{dx} \approx \frac{dp_n(x)}{dx}$$

o en general

$$\frac{d^n f(x)}{dx^n} \approx \frac{d^n p_n(x)}{dx^n} \quad (6.36)$$

Al diferenciar la fórmula fundamental de Newton dada arriba se tiene

$$\frac{d^n f(x)}{dx^n} = \frac{d^n p_n(x)}{dx^n} + \frac{d^n R_n(x)}{dx^n} \quad (6.37)$$

donde  $\frac{d^n R_n(x)}{dx^n}$  es el error que se comete al aproximar  $\frac{d^n f(x)}{dx^n}$  por  $\frac{d^n p_n(x)}{dx^n}$ .

Si las abscisas dadas  $x_0, x_1, \dots, x_n$  están espaciadas regularmente por intervalos de longitud  $h$ , entonces  $p_n(x)$  puede escribirse en términos de diferencias finitas. Al sustituir  $f[x_0]$ ,  $f[x_0, x_1]$ , etcétera en la ecuación 5.29 en términos de diferencias finitas (véase Ec.5.35), se obtiene

$$\begin{aligned} p_n(x) = & f[x_0] + (x - x_0) \frac{\Delta f[x_0]}{h} + (x - x_0)(x - x_1) \frac{\Delta^2 f[x_0]}{2! h^2} + \dots \\ & + (x - x_0)(x - x_1) \dots (x - x_{n-1}) \frac{\Delta^n f[x_0]}{n! h^n} \end{aligned}$$

y se tendrá

$$\begin{aligned} \frac{df(x)}{dx} \approx \frac{dp_n(x)}{dx} = & \frac{df[x_0]}{dx} + \frac{d \left[ (x - x_0) \frac{\Delta f[x_0]}{h} \right]}{dx} + \frac{d \left[ (x - x_0)(x - x_1) \frac{\Delta^2 f[x_0]}{2! h^2} \right]}{dx} \\ & + \dots + \frac{d \left[ (x - x_0)(x - x_1) \dots (x - x_{n-1}) \frac{\Delta^n f[x_0]}{n! h^n} \right]}{dx} \end{aligned} \quad (6.38)$$

<sup>\*</sup> Si la aproximación es por mínimos cuadrados, la diferenciación numérica consistirá en diferenciar el polinomio que mejor ajuste la información tabulada.

Se desarrollan algunos de los primeros términos y se tiene

$$\begin{aligned} \frac{df(x)}{dx} &\approx \frac{dp_n(x)}{dx} = \frac{\Delta f[x_0]}{h} + (2x - x_0 - x_1) \frac{\Delta^2 f[x_0]}{2! h^2} \\ &+ [3x^2 - 2(x_0 + x_1 + x_2)x + (x_0x_1 + x_0x_2 + x_1x_2)] \frac{\Delta^3 f[x_0]}{3! h^3} \end{aligned} \quad (6.39)$$

Seleccíonese ahora un valor particular para  $n$ ; por ejemplo, tómesese  $n = 1$ , es decir que se aproxime la función tabulada  $f(x)$  por una línea recta. Entonces

$$p_n(x) = p_1(x) = f[x_0] + (x - x_0) \frac{\Delta f[x_0]}{h},$$

y la primera derivada de  $f(x)$  queda aproximada por

$$\frac{df(x)}{dx} \approx \frac{dp_1(x)}{dx} = \frac{\Delta f[x_0]}{h} = \frac{f(x_1) - f(x_0)}{x_1 - x_0} = f[x_0, x_1]$$

$$\boxed{\frac{df(x)}{dx} \approx \frac{f(x_1) - f(x_0)}{h}} \quad (6.40)$$

y, como es de esperarse

$$\frac{d^2f(x)}{dx^2} \approx \frac{d^2p_1(x)}{dx^2} = 0$$

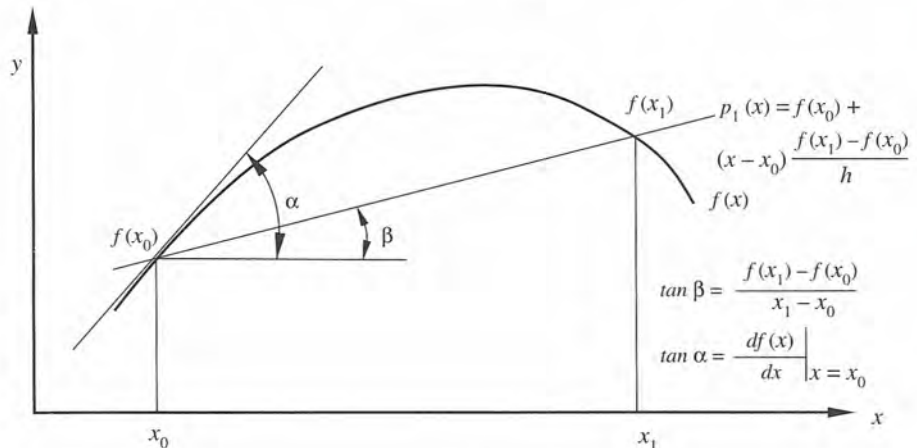
y así cualquier otra derivada superior de  $f(x)$  quedará aproximada por cero.

Geoméricamente esto equivale a tomar como primera derivada la pendiente de la recta que une los dos puntos de la curva  $f(x)$  de abscisas  $x_0$  y  $x_1$  (véase Fig. 6.11).

La primera derivada de  $f(x)$  en todo el intervalo  $[x_0, x_1]$  queda aproximada por el valor constante  $(f(x_1) - f(x_0))/h$ , el cual es muy diferente del valor verdadero  $df(x)/dx$  en general.

Si ahora  $n = 2$ ; es decir, aproximando la función tabulada  $f(x)$  por un polinomio de segundo grado, se tiene

$$p_n(x) \approx p_2(x) = f[x_0] + (x - x_0) \frac{\Delta f[x_0]}{h} + (x - x_0)(x - x_1) \frac{\Delta^2 f[x_0]}{2! h^2}$$



**Figura 6.11**  
Aproximación  
lineal de la  
primera  
derivada.

y la primera derivada de  $f(x)$  queda aproximada por

$$\frac{df(x)}{dx} \approx \frac{dp_2(x)}{dx} = \frac{\Delta f[x_0]}{h} + (2x - x_0 - x_1) \frac{\Delta^2 f[x_0]}{2! h^2}$$

Se desarrollan las diferencias hacia delante y se tiene

$$\frac{df(x)}{dx} \approx \left( \frac{2x - x_0 - x_1 - 2h}{2h^2} \right) f(x_0) + \left( \frac{2x_0 - 4x + 2x_1 + 2h}{2h^2} \right) f(x_1) + \left( \frac{2x - x_0 - x_1}{2h^2} \right) f(x_2) \quad (6.41)$$

La segunda derivada puede calcularse derivando una vez más con respecto a  $x$ , o sea

$$\frac{d^2f(x)}{dx^2} \approx \frac{d^2p_2(x)}{dx^2} = \frac{\Delta^2 f[x_0]}{h^2} = 2f[x_0, x_1, x_2]$$

$$\frac{d^2f(x)}{dx^2} \approx \frac{1}{h^2} f(x_0) - \frac{2}{h^2} f(x_1) + \frac{1}{h^2} f(x_2) \quad (6.42)$$

De igual modo se obtienen las distintas derivadas para  $n > 2$ .

Como se dijo al inicio de esta sección, el error cometido al aproximar  $\frac{d^n f(x)}{dx^n}$  por  $\frac{d^n p_n(x)}{dx^n}$  está dado por  $\frac{d^n R_n(x)}{dx^n}$ , donde a su vez  $R_n(x)$  está dado por la ecuación 5.41

$$R_n(x) = \left( \prod_{i=0}^n (x - x_i) \right) f[x, x_0, x_1, \dots, x_n]$$

que quedaría más compacta si se denota por  $\psi(x)$  a  $\prod_{i=0}^n (x - x_i)$ , es decir

$$R_n(x) = \psi(x) f[x, x_0, x_1, \dots, x_n] \quad (6.43)$$

En este punto es importante recordar que hay una estrecha relación entre las **diferencias divididas y las derivadas**. En general, esta relación está dada así

$$f[x, x_0, x_1, \dots, x_n] = \frac{d^n f(\xi)}{n! dx^n}, \text{ con } \xi \in (\text{mín } x_i, \text{ máx } x_i) \quad 0 \leq i \leq n$$

esto es,  $\xi$  es un valor de  $x$  desconocido, del cual sólo se sabe que está entre los valores menor y mayor de los argumentos. Se sustituye en la ecuación 6.43

$$R_n(x) = \psi(x) \frac{d^{n+1} f(\xi_1(x))}{(n+1)! dx^{n+1}}, \text{ con } \xi_1(x) \in (\text{mín } x, x_i, \text{ máx } x, x_i) \quad 0 \leq i \leq n$$

donde se ha escrito  $\xi_1$  como una función de  $x$ , ya que su valor depende del argumento  $x$  donde se desee evaluar la derivada.

Su primera derivada es

$$\frac{dR_n(x)}{dx} = \psi(x) \frac{d \left( \frac{d^{n+1} f(\xi_1(x))}{(n+1)! dx^{n+1}} \right)}{dx} + \frac{d^{n+1} f(\xi_1(x))}{(n+1)! dx^{n+1}} \frac{d\psi(x)}{dx} \quad (6.44)$$



Puede encontrarse que\*

$$d \left( \frac{d^{n+1}f(\xi_1(x))}{(n+1)! dx^{n+1}} \right) = \frac{d^{n+2}f(\xi_2(x))}{(n+2)! dx^{n+2}} \quad (6.45)$$

con  $\xi_1(x), \xi_2(x) \in (\min x, x_i, \max x, x_i)$   $0 \leq i \leq n$ , donde  $\xi_2$  es una función de  $x$  distinta de  $\xi_1$ .

Por esto, la ecuación 6.44 puede reescribirse como

$$\frac{dR_n(x)}{dx} = \psi(x) \frac{d^{n+2}f(\xi_2(x))}{(n+2)! dx^{n+2}} + \frac{d^{n+1}f(\xi_1(x))}{(n+1)! dx^{n+1}} \frac{d\psi(x)}{dx} \quad (6.46)$$

con  $\xi_1(x), \xi_2(x) \in (\min x, x_i, \max x, x_i)$   $0 \leq i \leq n$ .

En particular, para  $x = x_i$  [cuando  $x$  es una de las abscisas de la tabla de  $f(x)$ ] el **error de truncamiento** dado por la ecuación 6.46 se simplifica, ya que  $\psi(x_i) = (x_i - x_0)(x_i - x_1) \dots (x_i - x_i) \dots (x_i - x_n) = 0$ . Entonces

$$\begin{aligned} \left. \frac{dR_n(x)}{dx} \right|_{x_i} &= \frac{d^{n+1}f(\xi_1(x_i))}{(n+1)! dx^{n+1}} \left. \frac{d\psi(x)}{dx} \right|_{x_i} \\ &= \frac{d^{n+1}f(\xi_1(x_i))}{(n+1)! dx^{n+1}} \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j) \quad \xi_1(x_i) \in (\min x_i, \max x_i) \quad 0 \leq i \leq n \end{aligned} \quad (6.47)$$

En los ejercicios (al final de este capítulo) se demuestra que

$$\left. \frac{d\psi(x)}{dx} \right|_{x_i} = \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j)$$

Por ejemplo, la ecuación 6.41 puede escribirse en términos del error como sigue

$$\begin{aligned} \left. \frac{df(x)}{dx} \right|_{x_0} &= \left( \frac{2x_0 - x_0 - x_1 - 2h}{2h^2} \right) f(x_0) + \left( \frac{2x_0 - 4x_0 + 2x_1 + 2h}{2h^2} \right) f(x_1) \\ &\quad + \left( \frac{2x_0 - x_0 - x_1}{2h^2} \right) f(x_2) + (x_0 - x_1)(x_0 - x_2) \frac{d^3f(x)}{3! dx^3} \Big|_{\xi} \end{aligned}$$

o

$$\left. \frac{df(x)}{dx} \right|_{x_0} = \frac{1}{2h} [-3f(x_0) + 4f(x_1) - f(x_2)] + \frac{h^2}{3} \frac{d^3f(x)}{dx^3} \Big|_{\xi} \quad (6.48)$$

con  $\xi \in (\min x_i, \max x_i)$ ,  $i = 0, 1, 2$

y en la misma forma

$$\left. \frac{df(x)}{dx} \right|_{x_1} = \frac{1}{2h} [f(x_2) - f(x_0)] + \frac{h^2}{6} \frac{d^3f(x)}{dx^3} \Big|_{\xi} \quad (6.49)$$

con  $\xi \in (\min x_i, \max x_i)$ ,  $i = 0, 1, 2$

y

$$\left. \frac{df(x)}{dx} \right|_{x_2} = \frac{1}{2h} [f(x_0) - 4f(x_1) + 3f(x_2)] + \frac{h^2}{3} \frac{d^3f(x)}{dx^3} \Big|_{\xi} \quad (6.50)$$

con  $\xi \in (\min x_i, \max x_i)$ ,  $i = 0, 1, 2$ .

\* Pizer, M.S. *Numerical Computing and Mathematical Analysis*. S.R.A. (1975), pp. 315-317.

Hay que observar que el término del error para la derivada en  $x_1$  es la mitad del término del error para la derivada en  $x_0$  y  $x_2$ . Esto es así porque en la primera derivada en  $x_1$  se utilizan valores de la función a ambos lados de  $x_1$ .

En la diferenciación numérica, el error de truncamiento puede ser muy grande. Si por ejemplo  $f^{(n+2)}(x) / (n+2)!$  y  $f^{(n+1)}(x) / (n+1)!$  son de la misma magnitud, lo cual es común, el primer término de la ecuación 6.46 tiene aproximadamente la misma magnitud, que el error de interpolación\*; entonces puede decirse que el error de la aproximación de la derivada es, generalmente, mayor que el error de interpolación en

$$\frac{d^{n+1} f(\xi_1(x))}{(n+1)! dx^{n+1}} \quad \frac{d\psi(x)}{dx}$$

que es el segundo término de la ecuación 6.46. Además, cuando  $x = x_i$ , la ecuación 6.47 muestra que el error en la derivada en  $x_i$  tiene la misma forma que el error de interpolación (ver nota al pie de esta página), salvo que los polinomios factores son distintos.

Se ha discutido solamente el error de la aproximación de la primera derivada; pero todo lo visto también es aplicable a derivadas de mayor orden.

### Ejemplo 6.12

La ecuación de Van der Waals para un gmol de  $\text{CO}_2$  es

$$\left(P + \frac{a}{v^2}\right)(v - b) = RT$$

donde

$$a = 3.6 \times 10^{-6} \text{ atm cm}^6/\text{gmol}^2$$

$$b = 42.8 \text{ cm}^3/\text{gmol}$$

$$R = 82.1 \text{ atm cm}^3/(\text{gmol K})$$

Si  $T = 350 \text{ K}$ , se obtiene la siguiente tabla de valores.

| Puntos                    | 0      | 1      | 2      | 3      |
|---------------------------|--------|--------|--------|--------|
| $P \text{ (atm)}$         | 13.782 | 12.577 | 11.565 | 10.704 |
| $v \text{ (cm}^3\text{)}$ | 2000   | 2200   | 2400   | 2600   |

Calcule  $\partial P/\partial v$  cuando  $v = 2300 \text{ cm}^3$  y compárelo con el valor de la derivada analítica

**Solución** Al usar la ecuación 6.41 con los puntos (0), (1) y (2) se obtiene

$$\begin{aligned} \frac{\partial P}{\partial v} &= \frac{2v - v_0 - v_1 - 2h}{2h^2} P_0 + \frac{2v_0 - 4v + 2v_1 + 2h}{2h^2} P_1 + \frac{2v - v_0 - v_1}{2h^2} P_2; \text{ con } h = 200 \\ &= \frac{2(2300) - 2000 - 2200 - 2(200)}{2(200)^2} 13.782 \\ &\quad + \frac{2(2000) - 4(2300) + 2(2200) + 2(200)}{2(200)^2} 12.577 \\ &\quad + \frac{2(2300) - 2000 - 2200}{2(200)^2} 11.565 = -0.00506 \end{aligned}$$

\* Recuérdese que el error de interpolación es  $\prod_{i=0}^n (x - x_i) \frac{f^{(n+1)}(\xi)}{(n+1)!}$  donde  $\xi$  depende de  $x$  de un modo desconocido.

La derivada analítica es

$$\frac{\partial P}{\partial v} = \frac{-RT}{(v-b)^2} + \frac{2a}{v^3} = \frac{-82.1(350)}{(2300-42.8)^2} + \frac{2(3.6 \times 10^{-6})}{2300^3} = -0.005048$$

Hay que observar que la aproximación es muy buena (error relativo = -0.24%) a pesar de haber aplicado un polinomio de segundo grado para aproximar la ecuación de Van der Waals que, como se sabe, es un polinomio de tercer grado en  $v$ .

### Ejemplo 6.13

Obtenga la primera derivada del polinomio general de Langrange (ecuaciones 5.22 y 5.23).

#### Solución

De la ecuación  $p_n(x) = \sum_{i=0}^n f(x_i) \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j}$

Se deriva con respecto a  $x$

$$\frac{dp_n(x)}{dx} = \sum_{i=0}^n f(x_i) \frac{d}{dx} \left[ \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} \right]$$

Se hace

$$y = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j}$$

y se toman logaritmos en ambos lados, con lo que se tiene

$$\ln y = \ln \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} = \sum_{\substack{j=0 \\ j \neq i}}^n \ln \frac{x - x_j}{x_i - x_j}$$

ya que el logaritmo de un producto es igual a la suma de logaritmos de los factores.  
Ambos miembros se derivan con respecto a  $x$

$$\frac{d}{dx} (\ln y) = \frac{1}{y} \frac{dy}{dx} = \sum_{\substack{j=0 \\ j \neq i}}^n \frac{d}{dx} \left( \ln \frac{x - x_j}{x_i - x_j} \right) = \sum_{\substack{j=0 \\ j \neq i}}^n \frac{1}{x - x_j}$$

se despeja  $dy/dx$

$$\frac{dy}{dx} = y \sum_{\substack{j=0 \\ j \neq i}}^n \frac{1}{x - x_j}$$

Se sustituye  $y$  en lado derecho

$$\frac{dy}{dx} = \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} \sum_{\substack{j=0 \\ j \neq i}}^n \frac{1}{x - x_j}$$

y finalmente

$$\frac{dp_n(x)}{dx} = \sum_{i=0}^n f(x_i) \left[ \prod_{\substack{j=0 \\ j \neq i}}^n \frac{x - x_j}{x_i - x_j} \sum_{\substack{j=0 \\ j \neq i}}^n \frac{1}{x - x_j} \right]$$



Hay que observar que esta ecuación no sirve para evaluar la derivada en una de las abscisas de la tabla, ya que significaría dividir entre cero en la sumatoria dentro del paréntesis. Sin embargo, manipulando algebraicamente, el lado derecho puede escribirse en la forma

$$\frac{dp_n(x)}{dx} = \sum_{i=0}^n \frac{f(x_i)}{\prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j)} \sum_{\substack{k=0 \\ k \neq i}}^n \prod_{\substack{j=0 \\ j \neq k, i}}^n (x - x_j) \quad (6.51)$$

La cual ya no tiene la limitante mencionada.

### Ejemplo 6.14

En una reacción química  $A + B \xrightarrow{k}$  Productos, la concentración del reactante  $A$  es una función de la presión  $P$  y la temperatura  $T$ . La siguiente tabla presenta la concentración de  $A$  en  $\text{gmol/L}$  como función de estas dos variables

| P (kg/cm <sup>2</sup> ) | T (K) |      |      |      |
|-------------------------|-------|------|------|------|
|                         | 273   | 300  | 325  | 360  |
| 1                       | 0.99  | 0.97 | 0.96 | 0.93 |
| 2                       | 0.88  | 0.82 | 0.79 | 0.77 |
| 8                       | 0.62  | 0.51 | 0.48 | 0.45 |
| 15                      | 0.56  | 0.49 | 0.46 | 0.42 |
| 20                      | 0.52  | 0.44 | 0.41 | 0.37 |

Calcule la variación de la concentración de  $A$  con la temperatura a  $P = 8 \text{ kg/cm}^2$  y  $T = 300 \text{ K}$ , usando un polinomio de segundo grado.

**Solución** Lo que se busca es en sí  $\left. \frac{\partial C_A}{\partial T} \right|_{T=300, P=8}$  que se puede evaluar con la ecuación 6.51.



Al desarrollarla para  $n = 2$  se tiene

$$\frac{dp_2(x)}{dx} = \frac{(2x - x_1 - x_2)f(x_0)}{(x_0 - x_1)(x_0 - x_2)} + \frac{(2x - x_0 - x_2)f(x_1)}{(x_1 - x_0)(x_1 - x_2)} + \frac{(2x - x_0 - x_1)f(x_2)}{(x_2 - x_0)(x_2 - x_1)}$$

donde  $f(x)$  representa a  $C_A$  y  $x$  a  $T$ ; de tal modo que sustituyendo los tres puntos enmarcados de la tabla queda

$$\begin{aligned} \left. \frac{dp_2(x)}{dx} = \frac{\partial C_A}{\partial T} \right|_{T=300, P=8} &= \frac{(2(300) - 300 - 325)(0.62)}{(273 - 300)(273 - 325)} + \frac{(2(300) - 273 - 325)(0.51)}{(300 - 273)(300 - 325)} \\ &+ \frac{(2(300) - 273 - 300)(0.48)}{(325 - 273)(325 - 300)} = -0.0026 \frac{\text{gmol}}{\text{L K}} \end{aligned}$$

**Ejemplo 6.15**

Obtenga la primera y segunda derivadas evaluadas en  $x = 1$  para la siguiente función tabulada

| Puntos | 0  | 1 | 2  | 3   | 4   |
|--------|----|---|----|-----|-----|
| $x$    | -1 | 0 | 2  | 5   | 10  |
| $f(x)$ | 11 | 3 | 23 | 143 | 583 |

**Solución** Al construir la tabla de diferencias divididas se tiene

| Puntos | $x$ | $f(x)$ | Diferencias divididas |          |
|--------|-----|--------|-----------------------|----------|
|        |     |        | Primeras              | Segundas |
| 0      | -1  | 11     |                       |          |
|        |     |        | -8                    |          |
| 1      | 0   | 3      |                       | 6        |
|        |     |        | 10                    |          |
| 2      | 2   | 23     |                       | 6        |
|        |     |        | 40                    |          |
| 3      | 5   | 143    |                       | 6        |
|        |     |        | 88                    |          |
| 4      | 10  | 583    |                       |          |

Hay que observar que un polinomio de segundo grado puede representar exactamente la función (ya que la segunda diferencia dividida es constante).

El polinomio de Newton de segundo grado en diferencias divididas es

$$p_2(x) = f[x_0] + (x - x_0)f[x_0, x_1] + (x - x_0)(x - x_1)f[x_0, x_1, x_2]$$

que al derivarse da

$$\frac{dp_2(x)}{dx} = f[x_0, x_1] + (2x - x_0 - x_1)f[x_0, x_1, x_2]$$

y al derivarlo nuevamente se obtiene

$$\frac{d^2p_2(x)}{dx^2} = 2f[x_0, x_1, x_2]$$

con la sustitución de valores finalmente resulta

$$\frac{dp_2(1)}{dx} = -8 + (2(1) - (-1) - 0)(6) = 10 \quad \text{y} \quad \frac{d^2p_2(1)}{dx^2} = 12$$

**ALGORITMO 6.5** Derivación con polinomios de Lagrange

Para obtener una aproximación a la primera derivada de una función tabular  $f(x)$  en un punto  $x$ , proporcionar los

**DATOS:** El grado  $N$  del polinomio de Lagrange por usar, las  $(N+1)$  parejas de valores  $(X(I), FX(I), I=0, 1, 2, \dots, N)$  y el punto  $XD$  en que se desea la evaluación.

**RESULTADOS:** Aproximación a la primera derivada en  $XD$ :  $DP$ .

PASO 1. Hacer  $DP = 0$   
 PASO 2. Hacer  $I = 0$   
 PASO 3. Mientras  $I \leq N$ , repetir los pasos 4 a 21.  
     PASO 4. Hacer  $P = 1$   
     PASO 5. Hacer  $J = 0$   
     PASO 6. Mientras  $J \leq N$ , repetir los pasos 7 a 8.  
         PASO 7. SI  $I \neq J$  Hacer  $P = P * (X(I) - X(J))$   
         PASO 8. Hacer  $J = J + 1$   
     PASO 9. Hacer  $S = 0$   
     PASO 10. Hacer  $K = 0$   
     PASO 11. Mientras  $K \leq N$ , repetir los pasos 12 a 19.  
         PASO 12. SI  $I \neq K$ , realizar los pasos 13 a 18.  
         PASO 13. Hacer  $P1 = 1$   
         PASO 14. Hacer  $J = 0$   
         PASO 15. Mientras  $J \leq N$ , repetir los pasos 16 a 17  
             PASO 16. SI  $J \neq I$  y  $J \neq K$   
                 Hacer  $P1 = P1 * (XD - X(J))$   
             PASO 17. Hacer  $J = J + 1$   
         PASO 18. Hacer  $S = S + P1$   
         PASO 19. Hacer  $K = K + 1$   
     PASO 20. Hacer  $DP = DP + FX(I)/P * S$   
     PASO 21. Hacer  $I = I + 1$   
 PASO 22. IMPRIMIR  $DP$  Y TERMINAR.

## Ejercicios

- 6.1** Tras haber analizado durante varios años una gran cantidad de datos empíricos, el astrónomo alemán Johannes Kepler (1571-1630) formuló tres leyes que describen el movimiento de los planetas alrededor del sol. Estas leyes pueden enunciarse como sigue:

**Primera Ley.** La órbita de cada planeta es una elipse que tiene al sol en uno de sus focos

**Segunda Ley.** El vector que va del centro del Sol al del planeta en movimiento describe áreas iguales en tiempos iguales.

**Tercera Ley.** Si el tiempo que requiere un planeta para recorrer una vez su órbita elíptica es  $T$  y el eje mayor de tal elipse es  $2a$ , entonces  $T^2 = ka^2$  para una constante  $k$ .

Estas leyes ponen en relevancia a la matemática de la elipse, por ejemplo la medición de la longitud de una elipse (la trayectoria de un planeta que gira alrededor del sol). Este cálculo no resulta tan trivial como pudiera pensarse y tiene que recurrirse a los métodos numéricos para obtener una aproximación. Por ejemplo si se tiene que

$$\mathbf{R}(\theta) = a \sin \theta \mathbf{i} + b \cos \theta \mathbf{j}$$



donde  $\mathbf{r}$  es el vector de posición de un punto de la elipse y  $\theta$  es el ángulo descrito por dicho punto,  $a$  es el semieje mayor y  $b$  es el semieje menor de la elipse, el cálculo de la longitud de arco ( $s$ ) de una curva se sabe que está dado por

$$s = \int_0^\theta \sqrt{\mathbf{r}' \cdot \mathbf{r}'} dt$$

Como  $\mathbf{r}'(\theta) = a \cos \theta \mathbf{i} - b \sin \theta \mathbf{j}$

$$\mathbf{r}' \cdot \mathbf{r}' = a^2 \cos^2 \theta + b^2 \sin^2 \theta$$

sustituyendo la integral

$$\begin{aligned} s &= \int_0^\theta \sqrt{a^2 \cos^2 t + b^2 \sin^2 t} dt = \int_0^\theta \sqrt{a^2 - a^2 \sin^2 t + b^2 \sin^2 t} dt \\ &= \int_0^\theta \sqrt{a^2 - (a^2 - b^2) \sin^2 t} dt \\ s &= \int_0^\theta a \sqrt{\frac{a^2 - (a^2 - b^2) \sin^2 t}{a^2}} dt = \int_0^\theta a \sqrt{1 - \frac{(a^2 - b^2)}{a^2} \sin^2 t} dt \\ &= \int_0^\theta a \sqrt{1 - k^2 \sin^2 t} dt \end{aligned}$$

donde  $k$  es la excentricidad y sus valores van de cero a 1.

Si la excentricidad de la órbita de Mercurio es 0.206, ¿cuál sería la longitud del recorrido de una órbita completa de este planeta?

**Solución** Se sabe que el semieje mayor de la trayectoria de Mercurio es  $a = 0.387 \text{ UA}$  ( $1 \text{ UA} = 150\,000\,000 \text{ Km}$ ), de modo que

$$s = \int_0^\theta a \sqrt{1 - k^2 \sin^2 t} dt$$

$$s = 4 \int_0^{\pi/2} 0.387 \sqrt{1 - 0.206^2 \sin^2 t} dt$$

Utilizando el método de Simpson 1/3 con 10 subintervalos, se obtiene:  $s = 2.40558697 \text{ UA}$

**6.2** La siguiente tabla representa el gasto instantáneo del petróleo crudo en un oleoducto (en miles de libras por hora). El flujo se mide a intervalos de 12 minutos.

|       |      |      |      |      |      |      |      |      |
|-------|------|------|------|------|------|------|------|------|
| Hora  | 6:00 | 6:12 | 6:24 | 6:36 | 6:48 | 7:00 | 7:12 | 7:24 |
| Gasto | 6.2  | 6.0  | 5.9  | 5.9  | 6.2  | 6.4  | 6.5  | 6.8  |

|       |      |      |      |      |
|-------|------|------|------|------|
| Hora  | 7:36 | 7:48 | 8:00 | 8:12 |
| Gasto | 6.9  | 7.1  | 7.3  | 6.9  |

¿Cuál es la cantidad de petróleo bombeado en 2 horas y 12 minutos?

Calcule el gasto promedio en ese periodo.

**Solución** El petróleo bombeado se calcula multiplicando el gasto por el tiempo; pero como el gasto es variable, se aplica la integral siguiente



$$W = \int_0^{2.2} G dt \quad \text{lb de petróleo}$$

Integral que se puede aproximar por la regla del trapecioide (véase la Ec. 6.8).

$$I = \frac{h}{2} (f(x_0) + 2 \sum_{i=1}^{n-1} f(x_i) + f(x_n))$$

en donde

$$h = \frac{2.2}{11} = 0.2$$

$f(x_i)$  = gastos en lb/hr a cada intervalo.

Al sustituir valores queda

$$W = \frac{0.2}{2} [6.2 + 2(6.0 + 5.9 + 5.9 + 6.2 + 6.4 + 6.5 + 6.8 + 6.9 + 7.1 + 7.3) + 6.9]$$

$$= 14.31$$

Este valor se multiplica por 1000, ya que la tabla muestra los valores del gasto en miles de libras por hora.

El gasto promedio se calcula directamente

$$W_{\text{prom}} = \frac{W}{t} = \frac{14310}{2.2} = 6500 \text{ lb/hr}$$

- 6.3** En el interior de un cilindro de aluminio se tiene una resistencia eléctrica que genera una temperatura  $T_1 = 1200^\circ\text{F}$ . En la superficie exterior del cilindro circula un fluido que mantiene su temperatura a  $T_2 = 300^\circ\text{F}$ . Calcule la cantidad de calor transferido al fluido por unidad de tiempo.

**Datos adicionales**

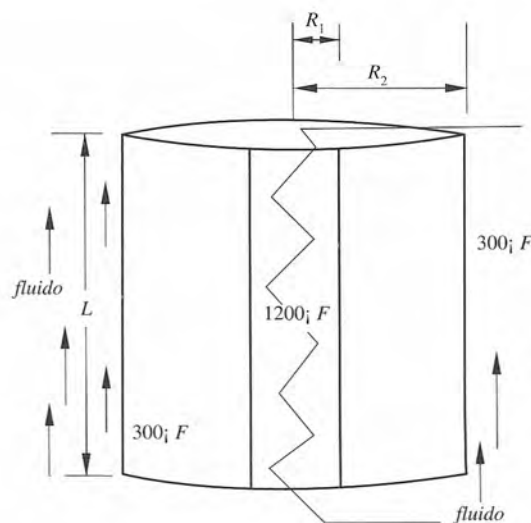
$R_1 = 2$  pulg,  $R_2 = 12$  pulg  $L = 12$  pulg

La conductividad térmica del aluminio varía con la temperatura según la tabla

| $k$ BTU/ ( hr pie <sup>2</sup> (°F/pie) ) | 165  | 150 | 130 | 108 |
|-------------------------------------------|------|-----|-----|-----|
| $T$ °F                                    | 1200 | 900 | 600 | 300 |

**Solución** Se asume un régimen permanente y se modela el proceso con la ecuación de Fourier

$$q = -k A \frac{dT}{dr}$$



donde:

$q$  = calor transferido al fluido en BTU/hr.

$k$  = conductividad térmica del aluminio en  $\frac{\text{BTU}}{\text{hr pie}^2 (^\circ\text{F/pie})}$ .

$A$  = área de transferencia de calor en  $\text{pie}^2$

$T$  = temperatura en  $^\circ\text{F}$ .

$r$  = distancia radial a partir del centro del cilindro en pies.

Al separar variables, integrar y aplicar límites, la ecuación de Fourier queda

$$\int_{R_1}^{R_2} \frac{dr}{A} = - \frac{1}{q} \int_{T_1}^{T_2} k dT$$

Al sustituir el área de transmisión de calor  $A$  en función de la distancia radial  $r$ :  $A = 2\pi rL$  e integrar analíticamente el lado izquierdo se tiene

$$\frac{1}{2\pi L} \ln\left(\frac{R_2}{R_1}\right) = - \frac{1}{q} \int_{T_1}^{T_2} k dT$$

Sin embargo, hay que integrar numéricamente el lado derecho, ya que  $k = f(T)$  está dada en forma discreta (tabulada). Así que, despejando  $q$ , se tiene

$$q = - \frac{\int_{T_1}^{T_2} k dT}{\frac{1}{2\pi L} \ln\left(\frac{R_2}{R_1}\right)},$$

y al integrar con la regla trapezoidal el numerador y sustituir valores se obtiene

$$q = - \frac{-124950}{\frac{1}{2\pi (12/12)} \ln\left(\frac{12}{2}\right)} = 438163.7 \text{ BTU/hr.}$$

- 6.4** Evalúe el coeficiente de fugacidad  $\phi$  del butano a 40 atm y 200  $^\circ\text{C}$  con la cuadratura de Gauss-Legendre con dos puntos. El coeficiente de fugacidad está dado por la ecuación:

$$\ln \phi = \int_0^P \frac{z-1}{P} dP$$

y la relación de la presión con el factor de compresibilidad  $z$  se determinó experimentalmente y se da en la tabla

| Puntos  | 1     | 2     | 3     | 4     | 5     | 6     | 7     | 8     |
|---------|-------|-------|-------|-------|-------|-------|-------|-------|
| P (atm) | 5     | 8     | 15    | 19    | 25    | 30    | 35    | 40    |
| $z$     | 0.937 | 0.887 | 0.832 | 0.800 | 0.781 | 0.754 | 0.729 | 0.697 |

Se sabe también que  $\lim_{P \rightarrow 0} \frac{z-1}{P} = -0.006 \text{ atm}^{-1}$

**Solución** La expresión de Gauss-Legendre para dos puntos queda\*

$$\int_a^b f(t) dt \approx \frac{b-a}{2} \left[ w_1 f\left(\frac{x_1(b-a) + b + a}{2}\right) + w_2 f\left(\frac{x_2(b-a) + b + a}{2}\right) \right]$$

\* Véase el problema 6.21 al final de este capítulo.



donde  $w_1 = w_2 = 1$ ;  $x_1 = 0.5773502692$ ;  $x_2 = -0.5773502692$

Con el cálculo de los argumentos de la función  $f$  se tiene

$$\frac{x_1 (b-a) + b + a}{2} = \frac{0.5773502692(40-0) + 40 + 0}{2} = 31.547$$

$$\frac{x_2 (b-a) + b + a}{2} = \frac{-0.5773502692(40-0) + 40 + 0}{2} = 8.453$$

El cálculo del factor de compresibilidad  $z$  a los valores de  $P = 31.547$  y  $P = 8.453$  se realiza por interpolación.

Con los puntos (6), (7) y (8) de la tabla y alguno de los métodos del capítulo 5 se obtiene  $z(31.547) = 0.746$ , y con los puntos (1), (2) y (3) se obtiene  $z(8.453) = 0.881$ .

Con los valores de  $z$  y  $P$  en los dos puntos, se calcula el valor de la función por integrar

$$\frac{z-1}{P} = \frac{0.746-1}{31.457} = -0.00805$$

$$\frac{z-1}{P} = \frac{0.881-1}{8.453} = -0.01408$$

Se sustituyen valores en la ecuación de Gauss-Legendre y se tiene:

$$\ln \phi = \int_0^{40} \frac{z-1}{P} dP = \frac{40-0}{2} [1(-0.00805) + 1(-0.01408)] = -0.4426$$

de donde  $\phi = 0.6424$

Hay que observar que basta tener el valor experimental de  $z$  a las presiones de 8.453 y 31.547, que en este ejemplo se determinaron por interpolación. Es importante señalar que procediendo a la inversa; es decir, calculando los valores de las presiones a las que se requiere el valor de  $z$  y después determinando experimentalmente dichos valores, se ahorra un considerable número de experimentos (2 contra 8 en este caso). Esto constituye una de las ventajas más importantes del método de la cuadratura de Gauss-Legendre.

- 6.5** Encuentre el centro de masa de una lámina rectangular de  $2\pi \times \pi$ , suponiendo que la densidad en un punto  $P(x, y)$  de la lámina está dado por

$$\rho(x, y) = e^{-(x^2 + y^2)/2}.$$

### Solución



Por definición, los momentos de inercia con respecto al eje  $x$  y  $y$ , respectivamente, están dados por:

$$M_x = \int_R \int x \rho(x, y) dx dy, \quad M_y = \int_R \int y \rho(x, y) dx dy$$

y el centro de masa de la lámina es el punto  $(\bar{x}, \bar{y})$  tal que

$$\bar{x} = \frac{M_x}{M}, \quad \bar{y} = \frac{M_y}{M}$$

donde  $M = \int_R \int \rho(x, y) dx dy$ .

Para facilitar las integraciones, la lámina se pone como se muestra en la figura 6.12, con lo que

$$R = \{ (x, y): 0 \leq x \leq 2\pi, 0 \leq y \leq \pi \}$$

Primero se obtiene  $M$  con el método de cuadratura de Guss empleando tres puntos

$$M = \int_0^\pi \int_0^{2\pi} e^{-(x^2 + y^2)/2} dx dy = 1.56814$$

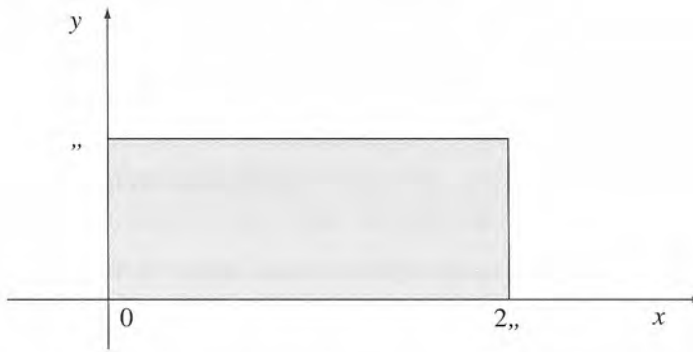


Figura 6.12

Después se calculan  $M_x$  y  $M_y$ , donde

$$M_x = \int_0^\pi \int_0^{2\pi} x e^{-(x^2 + y^2)/2} dx dy \approx 1.2556$$

$$M_y = \int_0^\pi \int_0^{2\pi} y e^{-(x^2 + y^2)/2} dx dy \approx 1.24449$$

Finalmente

$$\bar{y} \approx \frac{1.24449}{1.56814} = 0.7936,$$

$$\bar{x} \approx \frac{1.2556}{1.56814} = 0.8007,$$

el centro de masa es el punto del primer cuadrante (0.8007, 0.7936).

- 6.6** Las integrales del tipo  $\int_0^\infty e^{-x} f(x) dx$  se conocen como integrales impropias y se pueden aproximar, si su límite existe, por la cuadratura de Gauss-Laguerre

$$\int_0^\infty e^{-x} f(x) dx \approx \sum_{i=1}^n H_i f(x_i) \quad (1)$$

donde  $x_i$  es la  $i$ -ésima raíz del polinomio de Laguerre  $L_n(x)$  y

$$H_i = -\frac{[(n-1)!]^2}{L'_n(x_i) L_{n-1}(x_i)}; \quad i = 1, 2, \dots, n$$

se dan a continuación los primeros polinomios de Laguerre

$$L_0(x) = 1; L_1(x) = 1 - x; L_2(x) = 2 - 4x + x^2$$

$$L_3(x) = 6 - 18x + 9x^2 - x^3; L_4(x) = 24 - 96x + 72x^2 - 16x^3 + x^4$$

$$L_5(x) = 120 - 600x + 600x^2 - 200x^3 + 25x^4 - x^5$$

y la ecuación

$$L_{i+1}(x) = (1 + 2i - x) L_i(x) - i^2 L_{i-1}(x)$$

que permite obtener el polinomio de Laguerre de grado  $i + 1$  en términos de los polinomios de Laguerre de grado  $i$  e  $i-1$ .

Aproxime  $\int_0^\infty e^{-x} \sin x dx$  con  $n=2$

**Solución** Como  $n = 2$ :  $L_2(x) = 2 - 4x + x^2$ ;  $L'_2(x) = -4 + 2x$

Las raíces de  $L_2(x)$  son:  $x_1 = 2 - \sqrt{2}$ ,  $x_2 = 2 + \sqrt{2}$

Con la sustitución en  $H_i$  se tiene :

$$H_1 = - \frac{[(2-1)!]^2}{(-4+2(2-\sqrt{2}))(1-(2-\sqrt{2}))} = \frac{2+\sqrt{2}}{4}$$

$$H_2 = - \frac{[(2-1)!]^2}{(-4+2(2+\sqrt{2}))(1-(2+\sqrt{2}))} = \frac{2-\sqrt{2}}{4}$$

y la integral queda entonces

$$\int_0^\infty e^{-x} \sin x \, dx \approx \frac{1}{4} [(2+\sqrt{2}) \sin(2-\sqrt{2}) + (2-\sqrt{2}) \sin(2+\sqrt{2})] = 0.43246$$

En general, este proceso de integración puede programarse con una expresión del tipo 6.30

$$\int_0^\infty e^{-ax} f(x) \, dx \approx \frac{1}{a} \sum_{i=1}^n w_i f(x_i/a)$$

donde los pesos  $w_i$  y las abscisas  $x_i$  para  $2 \leq n \leq 5$  están dadas en la tabla 6.1

**Tabla 6.1** Coeficientes y abscisas para la integración por cuadratura de Gauss-Laguerre.

| Número de puntos<br>$n$ | Abscisas<br>$x_i$ | Coeficientes<br>$w_i$ |
|-------------------------|-------------------|-----------------------|
| 2                       | 0.585786          | 0.853553              |
|                         | 3.414214          | 0.146447              |
| 3                       | 0.415775          | 0.711093              |
|                         | 2.294280          | 0.278518              |
|                         | 6.289945          | 0.0103893             |
| 4                       | 0.322548          | 0.603154              |
|                         | 1.745761          | 0.357419              |
|                         | 4.536620          | 0.0388879             |
|                         | 9.395071          | 0.000539295           |
| 5                       | 0.263560          | 0.521756              |
|                         | 1.413403          | 0.398667              |
|                         | 3.596426          | 0.0759424             |
|                         | 7.085810          | 0.00361176            |
|                         | 12.640801         | 0.00002337            |



**6.7** De la gráfica de un diagrama de Mollier del amoníaco se obtienen los siguientes datos de temperatura (T) contra presión (P) a entalpía constante ( $H = 700$  BTU/Lb).

| Puntos   | 0   | 1   | 2   | 3   | 4   |
|----------|-----|-----|-----|-----|-----|
| T (°F)   | 175 | 200 | 225 | 250 | 275 |
| P (psia) | 100 | 270 | 450 | 640 | 850 |



Calcule el coeficiente de Joule-Thompson a una presión de 270 psia

- Mediante la derivada de la fórmula generalizada del polinomio de Lagrange del ejemplo 6.13 con los primeros cuatro puntos.
- Mediante la derivada analítica de una curva empírica polinomial de segundo grado calculada con mínimos cuadrados usando todos los puntos.

**Solución** El coeficiente de Joule-Thompson está definido como la derivada parcial de la temperatura con respecto a la presión a entalpía constante, o sea

$$\mu = \left( \frac{\partial T}{\partial P} \right)_H$$

- a) La fórmula del ejemplo 6.13 se desarrolla para  $n = 3$  y se obtiene

$$\begin{aligned} \frac{dp_3(x)}{dx} = & [3x^2 - 2(x_1 + x_2 + x_3)x + (x_1x_2 + x_1x_3 + x_2x_3)] \frac{f(x_0)}{(x_0 - x_1)(x_0 - x_2)(x_0 - x_3)} \\ & + [3x^2 - 2(x_0 + x_2 + x_3)x + (x_0x_2 + x_0x_3 + x_2x_3)] \frac{f(x_1)}{(x_1 - x_0)(x_1 - x_2)(x_1 - x_3)} \\ & + [3x^2 - 2(x_0 + x_1 + x_3)x + (x_0x_1 + x_0x_3 + x_1x_3)] \frac{f(x_2)}{(x_2 - x_0)(x_2 - x_1)(x_2 - x_3)} \\ & + [3x^2 - 2(x_0 + x_1 + x_2)x + (x_0x_1 + x_0x_2 + x_1x_2)] \frac{f(x_3)}{(x_3 - x_0)(x_3 - x_1)(x_3 - x_2)} \end{aligned}$$

donde  $x$  representa la presión y  $p_3(x)$  la temperatura. Al sustituir  $x$  por 270 psia así como los valores de los puntos dados, se obtiene

$$\mu = \left( \frac{\partial T}{\partial P} \right)_H \approx 0.1429 \text{ } ^\circ\text{F} / \text{psia}$$

Los cálculos se pueden realizar con Matlab o la TI92-Plus



```
N = 3;    Xd = 270 ;
X=[100 270 450 640] ;
Fx=[175 200 225 250] ;
Dp = 0;
for I = 1 : N + 1
    P = 1;
    for J = 1 : N+1
        if I ~= J
            P = P * (X(I) - X(J));
        end
    end
    S = 0;
    for K = 1: N + 1
        if I ~=K
            Pl = 1
            for J = 1: N+1
                if J == I | J == K
                    else
```

```

        P1 = P1 * (Xd - X (J) ) ;
    end
end
S = S + P1 ;
end
end
Dp = Dp + Fx (I) / P * S ;
end
Dp

```



```

ejer6_7( )
Prgm
4→n: 270→xd
{100, 270, 450, 640}→x
{175, 200, 225, 250}→fx
0→dp
For I, 1, n
    1→p
    For j, 1, n,
        If i ≠ j
            p*(x[i]-x [j])→p
        End for
    0→s
    For k,1,n
        If i≠k Then
            1→p1
            For j,1,n
                If j≠i and j≠k
                    p1*(xd-x[j])→p1
                EndFor
                s+p1→s
            EndIf
        EndFor
        dp+fx [I] / p*s→dp
    EndFor
    Disp dp
EndPrgm.

```

- b) El sistema de ecuaciones 5.64 se resuelve usando los cinco puntos de la tabla a fin de obtener los coeficientes del polinomio de segundo grado que mejor aproxima la función tabulada

$$a_0 = 159.5134, \quad a_1 = 0.156799, \quad a_2 = -0.2453 \times 10^{-4}$$

que sustituidos dan

$$T(P) \approx 159.5134 + 0.156799P - 0.2453 \times 10^{-4} P^2$$

cuya derivada es

$$\left( \frac{\partial T}{\partial P} \right)_H \approx 0.156799 - 2 (0.2453 \times 10^{-4}) P$$

que evaluada en  $P = 270$  resulta

$$\left( \frac{\partial T}{\partial P} \right)_H \approx 0.1436 \text{ } ^\circ\text{F} / \text{psia}$$

- 6.8** Encuentre la primera derivada numérica de  $xe^x$  en el punto  $x = 1$ , usando un polinomio de aproximación de segundo grado. Estime el error cometido.

**Solución** Lo más conveniente es emplear la ecuación 6.49

$$\left. \frac{df(x)}{dx} \right|_{x_1} = \frac{1}{2h} [f(x_2) - f(x_0)] - \frac{h^2}{6} \left. \frac{d^3f(x)}{dx^3} \right|_{\xi}; \quad \xi \in (x_0, x_2)$$

También se observa que el término del error en general es proporcional al valor de  $h$ , por lo que, si es posible, deberá tomarse muy pequeño. Se tomará en este caso  $h = 0.5$ . Con esto

$$\begin{aligned} x_0 &= 0.5, & x_1 &= 1, & x_2 &= 1.5 \\ f(x_0) &= 0.82436, & f(x_1) &= 2.71828, & f(x_2) &= 6.72253 \end{aligned}$$

Se sustituyen valores

$$\begin{aligned} \left. \frac{d(xe^x)}{dx} \right|_{x_1=1} &= \frac{1}{2(0.5)} [6.72253 - 0.82436] - \frac{0.5^2}{6} (xe^x + 3e^x) \Big|_{\xi} \\ &= 5.89817 - 0.04166 (xe^x + 3e^x) \Big|_{\xi} \end{aligned}$$

donde 5.89819 es la aproximación a la primera derivada y el factor  $xe^x + 3e^x$  es la derivada de tercer orden de  $xe^x$ . Como se desconoce  $\xi$  y  $xe^x + 3e^x$  es una función creciente, puede calcularse en  $x = 1.5$ . De esta manera se obtiene el valor máximo posible y, por ende, el valor máximo para el término del error resulta

$$0.04166(1.5 e^{1.5} + 3 e^{1.5}) = 0.84032$$

- 6.9** Demuestre que

$$\left. \frac{d}{dx} \left[ \prod_{j=0}^n (x - x_j) \right] \right|_{x=x_i} = \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j)$$

**Solución** Si  $n = 1$  se tiene

$$\left. \frac{d}{dx} [(x - x_0)(x - x_1)] \right|_{x=x_i} = [(x - x_0) + (x - x_1)] \Big|_{x=x_i} = \prod_{\substack{j=0 \\ j \neq i}}^1 (x_i - x_j)$$

Si  $n = 2$  se tiene

$$\begin{aligned} \left. \frac{d}{dx} [(x - x_0)(x - x_1)(x - x_2)] \right|_{x=x_i} &= [(x - x_2) \frac{d}{dx} [(x - x_0)(x - x_1)] \\ &\quad + (x - x_0)(x - x_1) \frac{d(x - x_2)}{dx}] \Big|_{x=x_i} \\ &= [(x - x_2)(x - x_0) + (x - x_2)(x - x_1) + (x - x_0)(x - x_1)] \Big|_{x=x_i} = \prod_{\substack{j=0 \\ j \neq i}}^2 (x_i - x_j) \end{aligned}$$

y por inducción puede llegarse a

$$\begin{aligned} \left. \frac{d}{dx} [(x - x_0)(x - x_1) \dots (x - x_n)] \right|_{x=x_i} &= [(x - x_n) \frac{d}{dx} [(x - x_0)(x - x_1) \dots (x - x_{n-1})] \\ &\quad + (x - x_0)(x - x_1) \dots (x - x_{n-1}) \frac{d(x - x_n)}{dx}] \Big|_{x=x_i} = \prod_{\substack{j=0 \\ j \neq i}}^n (x_i - x_j) \end{aligned}$$



**6.10** En la siguiente tabla

| $t$ | 0    | 1    | 2    | 3    | 4    | 5    |
|-----|------|------|------|------|------|------|
| T   | 93.1 | 85.9 | 78.8 | 75.1 | 69.8 | 66.7 |

T representa la temperatura ( $^{\circ}\text{C}$ ) de una salmuera utilizada como refrigerante y  $t$  (mín) es el tiempo. Encuentre la velocidad de enfriamiento en los tiempos  $t = 2.5$  y  $t = 4$  mín.

**Solución** Al emplear la ecuación 6.41 con  $n = 2$  se tiene

$$\frac{dp_2(t)}{dt} = \frac{(2t - t_0 - t_1 - 2h) T_0}{2h^2} + \frac{(2t_0 - 4t + 2t_1 + 2h) T_1}{2h^2} + \frac{(2t - t_0 - t_1) T_2}{2h^2}$$

Se toman  $t_0 = 1$ ,  $t_1 = 2$ ,  $t_2 = 3$  y  $h = 1$  y se tiene, para  $t = 2.5$

$$\begin{aligned} \frac{dp_2(t)}{dt} &= \frac{[2(2.5) - 1 - 2 - 2(1)]}{2(1)^2} (85.9) \\ &+ \frac{[2(1) - 4(2.5) + 2(2) + 2(1)]}{2(1)^2} (78.8) + \frac{(2(2.5) - 1 - 2)(75.1)}{2(1)^2} = -3.7 \end{aligned}$$

Al tomar  $t_0 = 2$ ,  $t_1 = 3$  y  $t_2 = 4$  se tiene, para  $t = 2.5$

$$\begin{aligned} \frac{dp_2(t)}{dt} &= \frac{[2(2.5) - 2 - 3 - 2(1)]}{2(1)^2} (78.8) \\ &+ \frac{[2(2) - 4(2.5) + 2(3) + 2(1)]}{2(1)^2} (75.1) + \frac{(2(2.5) - 2 - 3)(69.8)}{2(1)^2} = -3.7 \end{aligned}$$

Estos valores confirman que la función tabular se comporta como una parábola ( $n=2$ ); por lo tanto, el grado seleccionado es adecuado

Se deja al lector repetir estos cálculos para  $t=4$  mín.

**6.11** La siguiente tabla muestra las medidas observadas en una curva de imantación del hierro.

| $\beta$ | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   |
|---------|------|------|------|------|------|------|------|------|
| $\mu$   | 1090 | 1175 | 1245 | 1295 | 1330 | 1340 | 1320 | 1250 |

En ella  $\beta$  es el número de kilolíneas por  $\text{cm}^2$  y  $\mu$  la permeabilidad. Encuentre la permeabilidad máxima.

**Solución** Como la permeabilidad máxima registrada en la tabla es 1340, correspondiente a  $\beta = 10$ , se utilizan los puntos de abscisas  $\beta_0 = 9$ ,  $\beta_1 = 10$ ,  $\beta_2 = 11$  para obtener un polinomio de segundo grado, por el método de aproximación polinomial simple



$$a_0 + a_1(9) + a_2(9)^2 = 1330$$

$$a_0 + a_1(10) + a_2(10)^2 = 1340$$

$$a_0 + a_1(11) + a_2(11)^2 = 1320$$

Al resolver se tiene  $a_0 = -110$ ,  $a_1 = 295$  y  $a_2 = -15$ , por lo que el polinomio es

$$\mu = -110 + 295\beta - 15\beta^2$$

Para obtener la permeabilidad máxima, se deriva e iguala con cero este polinomio

$$\frac{d\mu}{d\beta} = 295 - 30\beta = 0$$

$$\text{Al despejar } \beta = \frac{295}{30} = 9.83333$$

de donde

$$\mu_{\max} = -110 + 295(9.83333) - 15(9.83333)^2 = 1340.416$$

## Problemas

- 6.1** Emplee la ecuación 6.5 con  $n = 3$  para obtener la ecuación de Simpson 3/8 (véase Ecs. 6.6)  
**6.2** Mediante el método de Simpson 3/8 aproxime las integrales del ejemplo 6.1 Compare los resultados con los obtenidos en los ejemplos 6.1 y 6.2.  
**6.3** Siguiendo las ideas que llevaron a las ecuaciones 6.8 y 6.10, encuentre la ecuación correspondiente a usar la fórmula de Simpson 3/8 sucesivamente.  
**6.4** Con el algoritmo obtenido en el problema anterior, integre la función

$$\frac{1}{2\pi} e^{-x^2/2}$$

entre los límites  $a = -1$  y  $b = 1$ . Compare el resultado con los valores obtenidos en el ejemplo 6.5.

- 6.5** En el gasoducto Cactus, Tab. a Reynosa, Tamps. se determina el gasto  $W$  (kg/min) y su contenido de azufre  $S$  (en por ciento) periódicamente durante el día. Los resultados se presentan en la tabla.

| $t$ (hr)     | 0    | 4    | 8    | 12   | 15   | 20   | 22   | 24   |
|--------------|------|------|------|------|------|------|------|------|
| $W$ (kg/min) | 20   | 22   | 19.5 | 23   | 21   | 20   | 20.5 | 20.8 |
| $S$ (%)      | 0.30 | 0.45 | 0.38 | 0.35 | 0.30 | 0.43 | 0.41 | 0.40 |

- a) ¿Cuál es el gasto promedio?  
 b) ¿Qué cantidad de gas se bombea en 24 horas?  
 c) ¿Cuál es el contenido de azufre (%) promedio diario?  
 d) ¿Qué cantidad de azufre se bombea en 24 horas?
- 6.6** Integre la función de Bessel de primera especie y orden 1

$$J_1(x) = \sum_{k=0}^{\infty} (-1)^k \frac{(x/2)^{2k+1}}{k! (1+k+1)!}$$

con el método de Simpson compuesto (aplicado siete veces) en el intervalo  $[0,7]$ ; esto es,

$$\int_0^7 J_1(x) dx$$

**SUGERENCIA:** Consulte las tablas de funciones de Bessel.

6.7 Obtenga

$$\int_1^2 \frac{h^3 dh}{(1 + h^{1/2})}$$

6.8 Obtenga

$$\int_1^3 x e^{e^x} dx$$

6.9 Elabore un subprograma para integrar una función analítica por el método de Simpson 3/8 compuesto, usando sucesivamente 3, 6, 12, 24, 48, ..., 3072 subintervalos. Compruébelo con la función del ejemplo 6.5

6.10. De acuerdo con las ideas acerca del análisis del error de truncamiento en la aproximación trapezoidal, analice dicho error en la aproximación de Simpson 1/3.

SUGERENCIA: La expresión a que debe llegar es

$$|E_T| \leq \frac{n}{2} \frac{h^5}{90} M = n h \frac{h^4}{180} M = (b - a) \frac{h^4}{180} M,$$

donde  $|f^{IV}(x_i)| \leq M$  y  $n$  es el número de subintervalos en que se divide  $[a, b]$ ; de donde, el error de truncamiento en el método de Simpson 1/3 es proporcional a  $h^4$ , lo cual conjuntamente con la ecuación 6.10 se expresa

$$\int_a^b f(x) dx = \frac{h}{3} [f(x_0) + 4 \sum_{i=1}^{n-1} f(x_i) + 2 \sum_{i=2}^{n-2} f(x_i) + f(x_n)] + O(h^4)$$

6.11 Mediante la ecuación 6.18 encuentre una cota para el error de truncamiento al integrar la función  $e^{-x^2/2}$  entre los límites  $[-1, 1]$ , usando 2, 4, 8, 16, ..., 1024 subintervalos.

6.12 Emplee la integración de Romberg a fin de evaluar  $I_2^{(2)}$  para las siguientes integrales definidas

$$\begin{array}{ll} a) \int_{-1}^1 e^{-x^2/2} dx & b) \int_0^1 x^3 e^x dx \\ c) \int_2^3 \frac{dx}{x} & d) \int_1^3 \ln x dx \end{array}$$

6.13 Con el criterio de convergencia siguiente

$$|I_k^{(m)} - I_{k-1}^{(m+1)}| \leq 10^{-3},$$

aproxime las integrales que se dan a continuación

$$\begin{array}{ll} a) \int_0^1 e^{\cos 2\pi x} dx & b) \int_1^2 \frac{\cos x}{\sqrt{x}} dx \\ c) \int_0^2 e^{x/\pi} \cos kx dx & \text{con } k = 1, 2 \end{array}$$

En el inciso c) compare con la solución analítica.

6.14 Pruebe que en la integración de Romberg con  $h_2 = h_1/2$  [Ec.6.21],  $I_k^{(1)} = s$ , donde  $s$  es la aproximación de Simpson compuesta [Ec. 6.10] empleando  $2^k$  subintervalos.

6.15 Estime las siguientes integrales

$$\begin{array}{l} a) \int_0^1 \sin(101 \pi x) dx \\ b) \int_0^1 f(x) dx \text{ con } f(x) = \begin{cases} \frac{\sin x}{x} & \text{en } x \neq 0 \\ 1 & \text{en } x = 0 \end{cases} \end{array}$$

con una aproximación de  $10^{-5}$



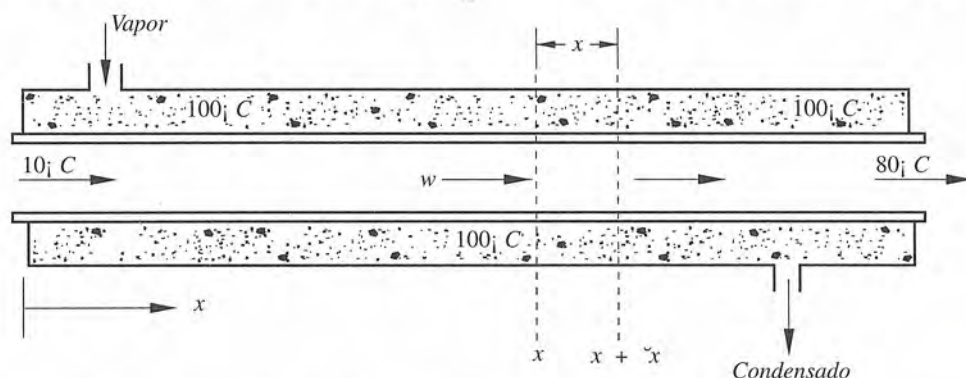
SUGERENCIA: Emplee como criterio de convergencia

$$|I_k^{(m)} - I_k^{(m+1)}| \leq 10^{-6}$$

con

$$|I_{k+1}^{(m+1)} - I_{k+1}^{(m+2)}| \leq 10^{-6}$$

- 6.16** En un cambiador de calor de tubos concéntricos se calienta nitrobenceno con un gasto  $w = 3000$  lb/hr en el tubo interior con diámetro  $D_i = 3$  pulg. Entre los tubos se circula vapor saturado que mantiene la temperatura de los tubos a temperatura constante  $T_s = 100^\circ\text{C}$ . ¿Cuál será la longitud necesaria de los tubos para calentar el nitrobenceno desde  $10$  hasta  $80^\circ\text{C}$ ?



El modelo del proceso se obtiene mediante un balance de calor a régimen permanente en un elemento diferencial de longitud  $\Delta x$

Acumulación = entrada - salida + transmisión

$$0 = w C_p T|_x - w C_p T|_{x+\Delta x} + D_i \pi \Delta x h (T_s - T)$$

al dividir entre  $\Delta x$  y hacer  $\Delta x \rightarrow 0$ , se obtiene en el límite

$$w C_p \frac{dT}{dx} = D_i \pi h (T_s - T)$$

y al separar variables queda

$$\frac{w}{D_i \pi} \int_{10}^{80} \frac{C_p dt}{h (T_s - T)} = \int_0^L dx = L - 0 = L$$

Donde  $h$  es el coeficiente de transmisión de calor en BTU/(hr pie  $^\circ\text{F}$ ) y se calcula con la expresión

$$h = 0.023 \text{ Re}^{0.8} \text{ Pr}^{0.33}$$

Hay que observar que  $\text{Re}$ ,  $C_p$  y  $\text{Pr}$  son funciones de  $T$ .

- 6.17** A principios del siglo, Lord Rayleigh resolvió el problema de la **destilación binaria simple** (una etapa) **por lotes**, con la ecuación que ahora lleva su nombre

$$\int_{L_i}^{L_f} \frac{dL}{L} = \int_{x_i}^{x_f} \frac{dx}{y - x}$$

donde  $L$  son los moles de la mezcla líquida en el hervidor,  $x$  las fracciones mol del componente más volátil en la mezcla líquida y  $y$  las fracciones mol de su vapor en equilibrio. Los subíndices  $i$  y  $f$  se refieren al estado inicial y final.

Calcule qué fracción de un lote es necesario destilar en una mezcla binaria para que  $x$  cambie de  $x_i = 0.7$  a  $x_f = 0.4$ . La relación de equilibrio está dada por la ecuación

$$y = \frac{\alpha x}{1 + (\alpha - 1)x}$$

donde  $\alpha$  es la volatilidad relativa de los componentes y es una función de  $x$  según la siguiente tabla (para una mezcla dada)

| $x$      | 0.70 | 0.65 | 0.60 | 0.55 | 0.50 | 0.45 | 0.40 |
|----------|------|------|------|------|------|------|------|
| $\alpha$ | 2.20 | 2.17 | 2.13 | 2.09 | 2.04 | 1.99 | 1.94 |

- 6.18** La integral  $\int_{-\pi}^{\pi} \frac{\sin x}{x} dx$  puede presentar serias dificultades.

Estudie cuidadosamente el integrando y aproxime dicha integral empleando alguno de los métodos vistos

- 6.19** Ensaye varios métodos de integración numérica para aproximar

$$\int_{-1}^1 \frac{x^2 dx}{\sqrt{1-x^2}}$$

- 6.20** Sea la función  $f(x)$  definida en  $(0,1)$  como sigue

$$f(x) = \begin{cases} x & 0 \leq x \leq 0.5 \\ 1-x & 0.5 < x \leq 1 \end{cases}$$

aproxime numéricamente  $\int_0^1 f(x) dx$  utilizando

- El método trapezoidal aplicado una vez en  $(0, 1)$
- El método trapezoidal aplicado una vez en  $(0, 0.5)$  y otra en  $(0.5, 1)$
- El método de Simpson 1/3 aplicado una vez en  $(0, 1)$ .

Compare los resultados con el valor analítico y explique las diferencias.

- 6.21** Demuestre que la expresión general para integrar por Gauss-Legendre puede ponerse

$$\int_a^b f(t) dt \approx \frac{b-a}{2} \sum_{i=1}^n w_i f\left[\frac{x_i(b-a) + b+a}{2}\right]$$

donde  $w_i, x_i, i = 1, 2, \dots, n$  dependen de  $n$  y están dados en la tabla 6.2

- 6.22** Use la cuadratura de Gauss con  $n = 3$  para aproximar las integrales de los problemas 6.18 y 6.19.

- 6.23** Dada la función  $f(x)$  en forma tabular

| $x$    | 0 | 41   | 56   | 95   | 145  | 180  | 212  | 320  |
|--------|---|------|------|------|------|------|------|------|
| $f(x)$ | 0 | 1.18 | 1.65 | 2.70 | 3.75 | 4.10 | 4.46 | 5.10 |

encuentre:

$$\int_0^{320} x^2 f(x) dx$$

usando la cuadratura de Gauss con varios puntos

- 6.24** Calcule el cambio de entropía  $\Delta S$  que sufre un gas ideal a presión constante al cambiar su temperatura de 300 a 380 K. Utilice la cuadratura gaussiana de tres puntos.

$$\int_{T_1}^{T_2} \frac{C_p dT}{T}$$

| T ( K )                 | 280  | 310  | 340  | 370  | 400  |
|-------------------------|------|------|------|------|------|
| $C_p(\text{cal/mol K})$ | 4.87 | 5.02 | 5.16 | 5.25 | 5.30 |



- 6.25** Modifique el programa del ejemplo 6.11 de modo que se puedan integrar funciones dadas en forma discreta o tabular.

**SUGERENCIA:** Vea el programa de interpolación de Lagrange en el ejercicio 5.5.

- 6.26** Una partícula de masa  $m$  se mueve a través de un fluido sujeta a una resistencia  $R$  que es función de la velocidad  $v$  de  $m$ . La relación entre la resistencia  $R$ , la velocidad  $v$ , y el tiempo  $t$  está dada por la ecuación

$$t = \int_{v_0}^{v_f} \frac{m}{R(v)} dv$$

Supóngase que  $R(v) = -v\sqrt{v} + 0.0001$  para un fluido particular. Si  $m = 10$  kg y  $v_0 = 10$  m/s, aproxime el tiempo requerido para que la partícula reduzca su velocidad a  $v_f = 5$  m/s, usando el método de cuadratura de Gauss con dos y tres puntos.

- 6.27** Aproxime las siguientes integrales usando la cuadratura de Gauss-Laguerre. Consulte el ejercicio 6.6.

$$\begin{array}{ll} a) \int_0^\infty e^{-3x} \ln x \, dx & b) \int_0^\infty e^{-2x} (\tan x + \sen x) \, dx \\ c) \int_0^\infty e^{-x} x^3 \, dx & d) \int_0^\infty e^{-x} 3x \, dx \\ e) \int_0^\infty e^{-3x} \, dx & \end{array}$$

- 6.28** Las integrales del tipo  $\int_a^\infty f(x) \, dx$ , con  $a > 0$  se conocen como integrales impropias y se pueden aproximar, si su límite existe, por los métodos de cuadratura haciendo el cambio de variable  $t = x^{-1}$ . Con este cambio el integrando y los límites pasan a ser

Como  $t = x^{-1}$ ,  $x = t^{-1}$  y  $dx = -1/t^2 \, dt$ .

El integrando queda  $f(x)dx = (-1/t^2)f(t^{-1}) \, dt$ .

Los límites pasan a

Como  $x = a$ ,  $t = 1/a$

y

$$x = \infty, \quad t = 1/\infty = 0$$

Al sustituir queda

$$\int_a^\infty f(x) \, dx = - \int_{1/a}^0 t^{-2} f(t^{-1}) \, dt = \int_0^{1/a} t^{-2} f(t^{-1}) \, dt$$

que ya puede aproximarse por los métodos vistos en este capítulo.

Aplique estas ideas para aproximar las siguientes integrales

$$\begin{array}{ll} a) \int_5^\infty x^{-3} \, dx & b) \int_1^\infty x^{-4} \sen(1/x) \, dx \\ c) \int_{10}^\infty x^{-2} e^{-3x} \cos(4/x) \, dx & \end{array}$$



Utilice el método numérico que considere más conveniente.

- 6.29 Elabore un algoritmo correspondiente al algoritmo 6.3 usando la cuadratura de Gauss-Laguerre (véase Ejer. 6.6).

- 6.30 Aproxime las integrales siguientes

a)  $\int_0^1 \int_1^2 x e^{xy} dy dx$       b)  $\int_1^2 \int_0^1 x e^{-xy} dx dy$

c)  $\int_1^3 \int_4^8 \sin x \cos y dx dy$       d)  $\int_1^1 \int_0^1 x y dx dy$

empleando el método de Simpson 1/3 dividiendo el intervalo  $(a, b)$  del eje  $x$  en  $n$  (par) subintervalos y el intervalo  $(c, d)$  del eje  $y$  en  $m$  (par) subintervalos. La ecuación por utilizar es la 6.31

- 6.31 Aproxime las integrales del problema 6.30 empleando la cuadratura de Gauss-Legendre

$$\int_c^d \int_a^b f(x, y) dx dy \approx \frac{(b-a)(d-c)}{4} \sum_{j=1}^m \sum_{i=1}^n w_j w_i f\left[\frac{b-a}{2} t_i + \frac{b+a}{2}, \frac{d-c}{2} u_j + \frac{c+d}{2}\right]$$

donde  $w_i$  o  $w_j$  son los coeficientes  $w_i$  dados en la tabla 6.2,  $t_i$  o  $u_j$  son las abscisas  $z_i$  dadas en la tabla 6.2 y  $n$  y  $m$  son los números de puntos por usar en los ejes  $x$  y  $y$ , respectivamente.

- 6.32 Mediante el método de Simpson 1/3 aproxime las integrales

a)  $\int_0^\pi \int_0^y y \sin x dx dy$       b)  $\int_0^1 \int_{\sqrt{x}}^1 dy dx$

c)  $\int_1^2 \int_x^{x^2} dy dx$       d)  $\int_0^2 \int_1^{e^y} dx dy$

- 6.33 En el estudio de integrales dobles, un problema típico es demostrar que

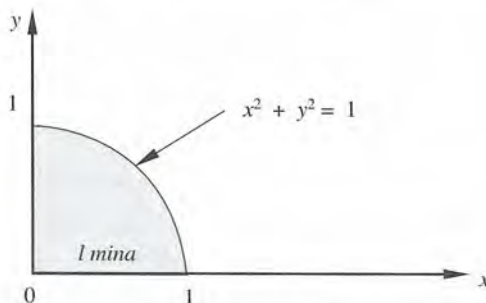
$$\int_0^R e^{-x^2} dx = \int_0^R \int_0^R e^{-x^2 - y^2} dy dx$$

Demuéstrelo numéricamente con  $R = 1$ . Utilice el método de Simpson 1/3.

- 6.34 Elabore un algoritmo para aproximar integrales dobles empleando el método de cuadratura de Gauss-Legendre.

- 6.35 Encuentre el centro de masa de una lámina cuya forma se encuentra en la figura adjunta, suponiendo que la densidad en un punto  $\rho(x, y)$  de la lámina está dada por

$$\rho(x, y) = x \sin y$$



- 6.36 La expresión  $\int_c^d \int_a^b dx dy$  representa el área del rectángulo cuyos vértices son  $a, b, c$  y  $d$  (corrobórela), de modo que  $\int_c^d \int_a^b f(x, y) dx dy$  representa el volumen del cuerpo cuya base es el rectángulo  $a, b, c, d$ , y cuya altura para cualquier punto  $(x, y)$  dentro de dicho rectángulo es  $f(x, y)$ . Aproximar el volumen de los siguientes cuerpos

a)  $f(x, y) = \sin x + e^{xy}$ ;  $(a, b, c, d) = (0, 4, 1, 3)$

b)  $f(x, y) = \sin \pi x \cos \pi y$ ,  $(a, b, c, d) = (0, \pi/4, 0, \pi/4)$

- 6.37** Emplee las ideas que llevaron a las ecuaciones 6.41 y 6.42 para obtener la aproximación de una función tabulada por un polinomio de tercer grado y su primera y segunda derivadas.
- 6.38** La ecuación de estado de Redlich-Kwong es

$$\left[ P + \frac{a}{T^{0.5} V(V+b)} \right] (V - b) = RT$$

donde  $a = 17.19344$  y  $b = 0.02211413$  para el oxígeno molecular.  
Si  $T = 373.15$  K, se obtiene la siguiente tabla de valores

| Puntos     | 0        | 1        | 2        | 3        |
|------------|----------|----------|----------|----------|
| P (atm)    | 30.43853 | 27.68355 | 25.38623 | 23.44122 |
| V (L/gmol) | 1.0      | 1.1      | 1.2      | 1.3      |

- a) Calcule la  $dP/dV$  cuando  $V = 1.05$  L utilizando las ecuaciones 6.40 y 6.41 y compárelo con el valor de la derivada analítica.
- b) Proceda como en el inciso anterior, pero ahora aplique la ecuación 6.51 con  $n = 1$  y  $n = 2$ .
- 6.39** Calcular  $\left. \frac{\partial CA}{\partial P} \right|_{T=325, P=10}$  utilizando la información del ejemplo 6.14.
- 6.40** Obtenga la segunda derivada evaluada en  $x = 3.7$  para la función que se da en seguida

| Puntos | 0 | 1       | 2       | 3       | 4        | 5        |
|--------|---|---------|---------|---------|----------|----------|
| $x$    | 1 | 1.8     | 3       | 4.2     | 5        | 6.5      |
| $f(x)$ | 3 | 4.34536 | 6.57735 | 8.88725 | 10.44721 | 13.39223 |

- Utilice un polinomio de Newton en diferencias divididas para aproximar  $f(x)$ .
- 6.41** Dada la función  $f(x) = x e^x + e^x$  aproxime  $f'(x), f''(x)$  en  $x = 0.6$ , empleando los valores de  $h = 0.4, 0.1, 0.0002$  con  $n = 1, 2, 3$ , para cada  $h$ . Compare los resultados con los valores analíticos.
- 6.42** Elabore un programa que aproxime la primera derivada de una función dada en forma tabular, usando el algoritmo 6.5
- 6.43** Encuentre la primera derivada numérica de  $x \ln x$  en el punto  $x = 2$ , usando un polinomio de aproximación de tercer grado. Estime el error cometido.
- 6.44** Dada la tabla:

| $x$     | 0.2     | 0.3     | 0.4     | 0.5     | 0.6     |
|---------|---------|---------|---------|---------|---------|
| $f(x)$  | 0.24428 | 0.40496 | 0.59673 | 0.82436 | 1.09327 |
| $f'(x)$ |         | 1.75482 | 2.08855 | 2.47308 |         |

- calcule  $f'(x)$  para  $x = 0.3, 0.4$  y  $0.5$  con  $n = 2$  (Ec. 6.41) y compare con los valores analíticos dados en la tabla
- 6.45** En la tabla siguiente,  $x$  es la distancia en metros que recorre una bala a lo largo de un cañón en  $t$  segundos. Encuentre la velocidad de la bala en  $x = 3$

|     |   |        |        |        |        |        |
|-----|---|--------|--------|--------|--------|--------|
| $x$ | 0 | 1      | 2      | 3      | 4      | 5      |
| $t$ | 0 | 0.0359 | 0.0493 | 0.0596 | 0.0700 | 0.0786 |

- 6.46** Dado un circuito con un voltaje  $E(t)$  y una inductancia  $L$ , la primera ley de Kirchhoff que lo modela es

$$E = L \, di/dt + R \, i$$

donde  $i$  es la corriente en amperes y  $R$  la resistencia en ohms. La tabla de abajo da los valores experimentales de  $i$  correspondientes a varios tiempos  $t$  dados en segundos. Si la inductancia  $L$  es constante e igual a 0.97 henries y la resistencia es de 0.14 ohms, aproxime el voltaje  $E$  en los valores de  $t$  dados en la tabla usando la ecuación 6.41.

|     |      |      |      |      |      |      |
|-----|------|------|------|------|------|------|
| $t$ | 0.95 | 0.96 | 0.97 | 0.98 | 0.99 | 1.0  |
| $i$ | 0.90 | 1.92 | 2.54 | 2.88 | 3.04 | 3.10 |

- 6.47** La reacción en fase líquida entre trimetilamina y bromuro de propilo en benceno, se llevó a cabo introduciendo cinco ampollas con una mezcla de reactantes en un baño a temperatura constante.

Las ampollas se sacan a varios tiempos, se enfrían para detener la reacción y se analiza su contenido. El análisis se basa en que la sal cuaternaria de amoniaco está ionizada, de aquí que la concentración de los iones bromuro se pueda obtener por titulación. Los resultados obtenidos son

|                |    |    |    |    |     |
|----------------|----|----|----|----|-----|
| Tiempo (min)   | 10 | 35 | 60 | 85 | 110 |
| Conversión (%) | 12 | 28 | 40 | 46 | 52  |

Calcule la variación de la conversión con respecto al tiempo de los distintos puntos de la tabla.

- 6.48** La densidad de soluciones de cloruro de calcio ( $\text{CaCl}_2$ ) a diferentes temperaturas y concentraciones se presenta en la siguiente tabla

| c% peso | T°C    |        |        |        |        |        |
|---------|--------|--------|--------|--------|--------|--------|
|         | -5     | 0      | 20     | 40     | 80     | 100    |
| 2       | —      | 1.0171 | 1.0148 | 1.0084 | 0.9881 | 0.9748 |
| 8       | 1.0708 | 1.0703 | 1.0659 | 1.0586 | 1.0382 | 1.0257 |
| 16      | 1.1471 | 1.1454 | 1.1386 | 1.1301 | 1.1092 | 1.0973 |
| 30      | —      | 1.2922 | 1.2816 | 1.2709 | 1.2478 | 1.2359 |
| 40      | —      | —      | 1.3957 | 1.3826 | 1.3571 | 1.3450 |

Calcule

- La variación de la densidad con respecto a la temperatura a  $T = 0^\circ\text{C}$  y  $c = 40\%$
- La variación de la densidad con respecto a la concentración a  $T = 0^\circ\text{C}$  y  $c = 40\%$
- La variación de la densidad con respecto a la temperatura a  $T = 30^\circ\text{C}$  y  $c = 10\%$
- La variación de la densidad con respecto a la concentración a  $T = 30^\circ\text{C}$  y  $c = 10\%$



# CAPÍTULO 7

## ECUACIONES DIFERENCIALES ORDINARIAS

### A dónde nos dirigimos

En este capítulo se estudiarán las técnicas numéricas de solución de ecuaciones diferenciales con condiciones iniciales o de frontera, denominados problemas de valor inicial o de frontera, respectivamente. Para esto se inicia formulando tales problemas, y luego, a partir de las ideas de extrapolación, se plantean métodos como el de Euler y los de Taylor. Más adelante, en un proceso de ponderación de pendientes, se obtienen métodos con diferentes órdenes de exactitud donde no se requiere de derivaciones complicadas de funciones, pagándose como precio de ello un mayor número de cálculos. Tales métodos son conocidos como métodos de Runge-Kutta. Basándose en el proceso de integración implicado en la solución de las ecuaciones diferenciales y en la aproximación de funciones vista en el capítulo 5, se plantean familias de métodos denominados de predicción-corrección.

Al final del capítulo se extienden las técnicas vistas a ecuaciones diferenciales de orden superior al primero, transformándolas a sistemas de ecuaciones diferenciales de primer orden y resolviéndolas como tales.

Dado que las ecuaciones diferenciales ordinarias permiten modelar procesos dinámicos: vaciado de recipientes, reactores químicos, movimientos amortiguados, desarrollos poblacionales e incluso situaciones estáticas como la deflexión de vigas y problemas geométricos y de que las técnicas analíticas son válidas sólo para ciertas ecuaciones muy particulares, las técnicas de este capítulo resultan no sólo complementarias sino necesarias.

### Introducción

Se llama **ecuación diferencial** aquella ecuación que contiene una variable dependiente y sus derivadas con respecto a una o más variables independientes. Muchas de la leyes generales de la naturaleza se expresan en el lenguaje de las ecuaciones diferenciales; abundan también las aplicaciones en ingeniería, economía, en las mismas matemáticas y en muchos otros campos de la ciencia aplicada.

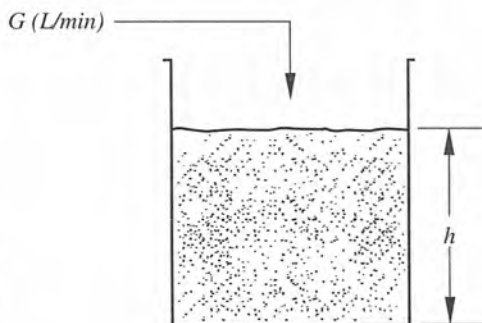
Esta gran utilidad de las ecuaciones diferenciales es fácil de explicar; recuérdese que si se tiene la función  $y = f(x)$ , su derivada  $dy/dx$  puede interpretarse como la velocidad de cambio de  $y$  con respecto a  $x$ . En cualquier proceso natural, las variables incluidas y sus velocidades de cambio se relacionan entre sí mediante los principios científicos que gobiernan el proceso. El resultado de expresar en símbolos matemáticos estas relaciones, a menudo es una ecuación diferencial.

Se tratará de ilustrar estos comentarios con el siguiente ejemplo.

Supóngase que se quiere conocer cómo varía la altura  $h$  del nivel en un tanque cilíndrico de área seccional  $A$  cuando se llena con un líquido de densidad  $\rho$  a razón de  $G$  (L/min) como se muestra en la figura 7.1.

La ecuación diferencial se obtiene mediante un balance de materia (principio universal de continuidad) en el tanque:

**Figura 7.1**  
Llenado de un  
tanque  
cilíndrico.



|                         |   |                     |   |                    |
|-------------------------|---|---------------------|---|--------------------|
| Acumulación<br>(Kg/min) | = | Entrada<br>(Kg/min) | - | Salida<br>(Kg/min) |
|-------------------------|---|---------------------|---|--------------------|

donde la acumulación significa la variación de la masa de líquido en el tanque con respecto al tiempo, la cual se expresa matemáticamente como una derivada:

$d(V\rho)/dt$ . Lo que entra es  $\frac{G}{\rho}$  (Kg/min) y el término de salida es nulo, con lo cual la ecuación de continuidad queda como sigue:

$$\frac{d(V\rho)}{dt} = G/\rho$$

Por otro lado, el volumen de líquido  $V$  que contiene el tanque a una altura  $h$  es\*  $V = A h$ . Al sustituir  $V$  en la ecuación diferencial de arriba y considerando que la densidad  $\rho$  es constante, se llega a:

$$A \frac{dh}{dt} = G \quad (7.1)$$

ecuación diferencial cuya solución describe cómo cambia la altura  $h$  del líquido dentro del tanque con respecto al tiempo  $t$ . A continuación se enumeran ejemplos de ecuaciones diferenciales.

$$\frac{dy}{dt} = -ky \quad (7.2)$$

$$m \frac{d^2y}{dt^2} = ky \quad (7.3)$$

$$\frac{dy}{dx} + 2xy = e^{-x^2} \quad (7.4)$$

$$\frac{d^2y}{dx^2} - 5 \frac{dy}{dx} + 6y = 0 \quad (7.5)$$

$$(1 - x^2) \frac{d^2y}{dx^2} - 2x \frac{dy}{dx} + p(p+1)y = 0 \quad (7.6)$$

$$x^2 \frac{d^2y}{dx^2} + x \frac{dy}{dx} + (x^2 - p^2)y = 0 \quad (7.7)$$

\* El fondo del tanque es plano.

La variable dependiente en cada una de estas ecuaciones es  $y$ , y la variable independiente es  $x$  o  $t$ . Las letras  $k$ ,  $m$  y  $p$  representan constantes. Una ecuación diferencial es **ordinaria** si sólo tiene una variable independiente, por lo que todas las derivadas que tiene son ordinarias o totales. Las ecuaciones 7.1 a 7.7 son ordinarias. El **orden** de una ecuación diferencial es el orden de la derivada de más alto orden en ella. Las ecuaciones 7.1, 7.2 y 7.4 son de primer orden, y las demás de segundo.

## 7.1 Formulación del problema de valor inicial

La ecuación diferencial ordinaria (EDO) general de primer orden es:

$$\frac{dy}{dx} = f(x, y) \quad (7.8)$$

En la teoría de las EDO se establece que su solución general debe contener una constante arbitraria  $c$ , de tal modo que la solución general de la ecuación 7.8 es:

$$F(x, y, c) = 0 \quad (7.9)$$

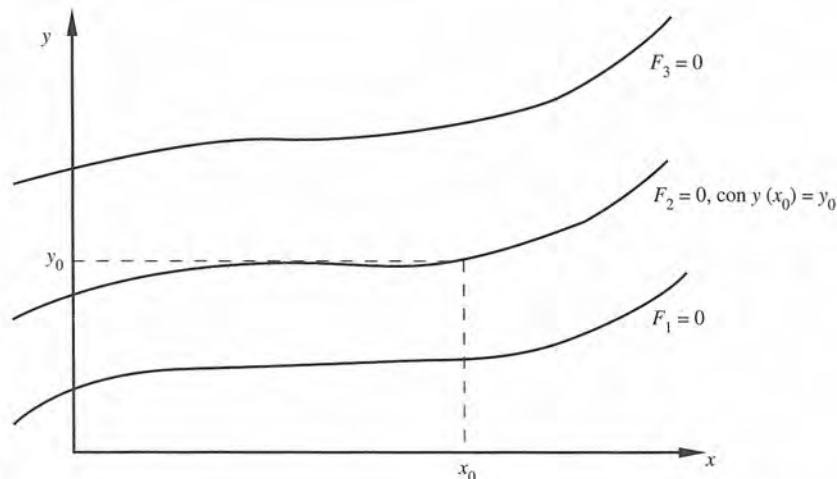
La ecuación 7.9 representa una familia de curvas en el plano  $x$ - $y$ , obtenida cada una de ellas para un valor particular de  $c$  como se muestra en la figura 7.2. Cada una de estas curvas corresponde a una solución particular de la EDO 7.8, y analíticamente dichas constantes se obtienen exigiendo que la solución de esa ecuación pase por algún punto  $(x_0, y_0)$ ; esto es, que

$$y(x_0) = y_0 \quad (7.10)$$

lo cual significa que la variable dependiente  $y$  vale  $y_0$  cuando la variable independiente  $x$  vale  $x_0$  (véase la curva  $F_2$  de la Fig. 7.2).

En los cursos regulares de cálculo y ecuaciones diferenciales se estudian técnicas **analíticas** para encontrar soluciones del tipo de la ecuación 7.9 a problemas como el de la ecuación 7.8 o mejor aún, a problemas de valor inicial —ecuación 7.8 y condición 7.10, simultáneamente.

En la práctica la mayoría de las ecuaciones no pueden resolverse utilizando estas técnicas, y se deberá recurrir a los **métodos numéricos**.



**Figura 7.2**  
Representación  
gráfica de la  
solución  
general,  
ecuación 7.9.



Cuando se usan métodos numéricos no se encuentran soluciones de la forma  $F(x, y, c) = 0$ , ya que trabajan con números y dan por resultado números. Sin embargo, el propósito usual de encontrar una solución es determinar valores de  $y$  (números) correspondientes a valores específicos de  $x$ , lo cual es factible con los mencionados métodos numéricos sin tener que encontrar  $F(x, y, c) = 0$ .

El problema de valor inicial (PVI) por resolver numéricamente queda formulado como sigue:

- a) Una ecuación diferencial de primer orden (del tipo 7.8)
- b) El valor de  $y$  en un punto conocido  $x_0$  (condición inicial)
- c) El valor  $x_f$  donde se quiere conocer el valor de  $y$  ( $y_f$ )

que en lenguaje matemático quedará así

$$\text{PVI} \begin{cases} \frac{dy}{dx} = f(x, y) \\ y(x_0) = y_0 \\ y(x_f) = ? \end{cases} \quad (7.11)$$

Formulado el problema de valor inicial, a continuación se describen una serie de técnicas numéricas para resolverlo.

## 7.2 Método de Euler

El método de Euler es el más simple de los métodos numéricos para resolver un problema de valor inicial del tipo 7.11. Consiste en dividir el intervalo que va de  $x_0$  a  $x_f$  en  $n$  subintervalos de ancho  $h$  (véase Fig. 7.3); o sea,

$$h = \frac{x_f - x_0}{n} \quad (7.12)$$

de manera que se obtiene un conjunto discreto de  $(n + 1)$  puntos\*:  $x_0, x_1, x_2, \dots, x_n$  del intervalo de interés  $[x_0, x_f]$ . Para cualquiera de estos puntos se cumple que

$$x_i = x_0 + ih, \quad 0 \leq i \leq n \quad (7.13)$$

Nótese la similitud de este desarrollo con el primer paso de la integración numérica.

La condición inicial  $y(x_0) = y_0$  representa el punto  $P_0 = (x_0, y_0)$  por donde pasa la curva solución de la ecuación 7.11, la cual por simplicidad se denotará como  $F(x) = y$ , en lugar de  $F(x, y, c) = 0$ .

Con el punto  $P_0$  se puede evaluar la primera derivada de  $F(x)$  en ese punto; a saber

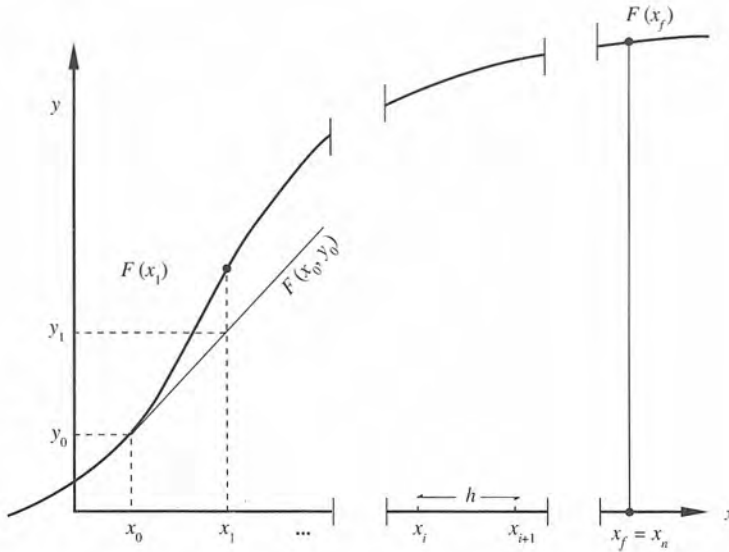
$$F'(x) = \frac{dy}{dx} \Big|_{P_0} = f(x_0, y_0) \quad (7.14)$$

Con esta información se traza una recta, aquella que pasa por  $P_0$  y de pendiente  $f(x_0, y_0)$ . Esta recta aproxima  $F(x)$  en una vecindad de  $x_0$ . Tómesela la recta como remplazo de  $F(x)$  y localícese en ella (la recta) el valor de  $y$  correspondiente a  $x_1$ . Entonces, de la figura 7.3

$$\frac{y_1 - y_0}{x_1 - x_0} = f(x_0, y_0) \quad (7.15)$$

\*  $x_f$  se convierte en  $x_n$ .

**Figura 7.3**  
Deducción  
gráfica del  
método de  
Euler.



Se resuelve para  $y_1$

$$y_1 = y_0 + (x_1 - x_0) f(x_0, y_0) = y_0 + h f(x_0, y_0) \quad (7.16)$$

Es evidente que la ordenada  $y_1$  calculada de esta manera no es igual a  $F(x_1)$ , pues existe un pequeño error. No obstante, el valor  $y_1$  sirve para aproximar  $F'(x)$  en el punto  $P = (x_1, y_1)$  y repetir el procedimiento anterior a fin de generar la sucesión de aproximaciones siguiente:

$$\begin{aligned} y_1 &= y_0 + h f(x_0, y_0) \\ y_2 &= y_1 + h f(x_1, y_1) \\ &\vdots \\ y_{i+1} &= y_i + h f(x_i, y_i) \\ &\vdots \\ y_n &= y_{n-1} + h f(x_{n-1}, y_{n-1}) \end{aligned} \quad (7.17)$$

Como se muestra en la figura 7.4, en esencia se trata de aproximar la curva  $y = F(x)$  por medio de una serie de segmentos de línea recta.

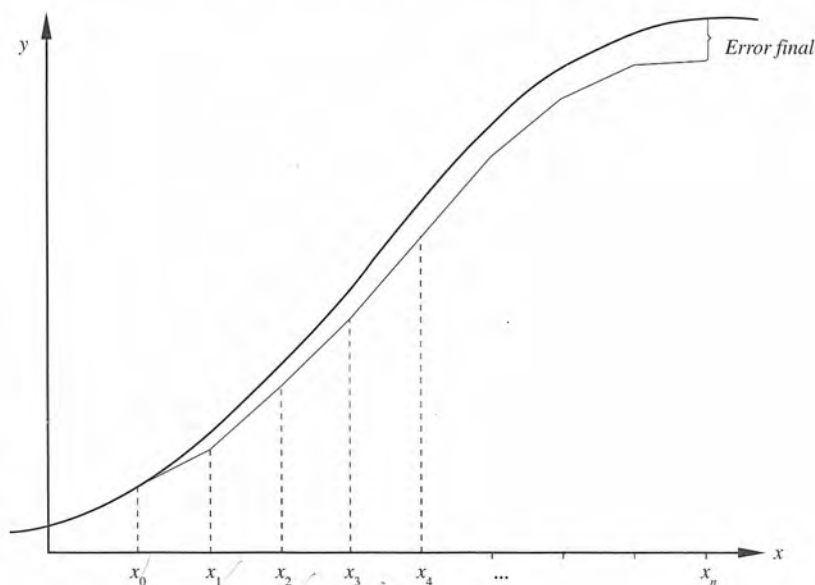
Como la aproximación a una curva mediante una línea recta no es exacta, se comete un error propio del método mismo. De modo similar a otros capítulos, éste se denominará **error de truncamiento**. Dicho error puede disminuirse tanto como se quiera (al menos teóricamente) reduciendo el valor de  $h$ , pero a cambio de un mayor número de cálculos y tiempo de máquina y, por consiguiente, de un **error de redondeo** más alto.

### Ejemplo 7.1

Resuelva el siguiente

$$\text{PVI} \begin{cases} \frac{dy}{dx} = (x - y) \\ y(0) = 2 \\ y(1) = ? \end{cases}$$

mediante el método de Euler.



**Figura 7.4**  
Aplicación  
repetida del  
método de  
Euler.

### Solución

**SUGERENCIA:** Puede usar un pizarrón electrónico o el GC para seguir los cálculos.



El intervalo de interés para este ejemplo es  $[0, 1]$  y al dividirlo en cinco subintervalos se tiene:

$$h = \frac{1 - 0}{5} = 0.2$$

con lo cual se generan los argumentos

$$x_0 = 0.0, x_1 = x_0 + h = 0.0 + 0.2 = 0.2$$

$$x_2 = x_1 + h = 0.2 + 0.2 = 0.4$$

⋮

$$x_5 = x_4 + h = 0.8 + 0.2 = 1.0$$

Con  $x_0 = 0.0$  y  $y_0 = 2$  y las ecuaciones 7.17 se obtienen los valores

$$y_1 = y(0.2) = 2 + 0.2[0.0 - 2] = 1.6$$

$$y_2 = y(0.4) = 1.6 + 0.2[0.2 - 1.6] = 1.32$$

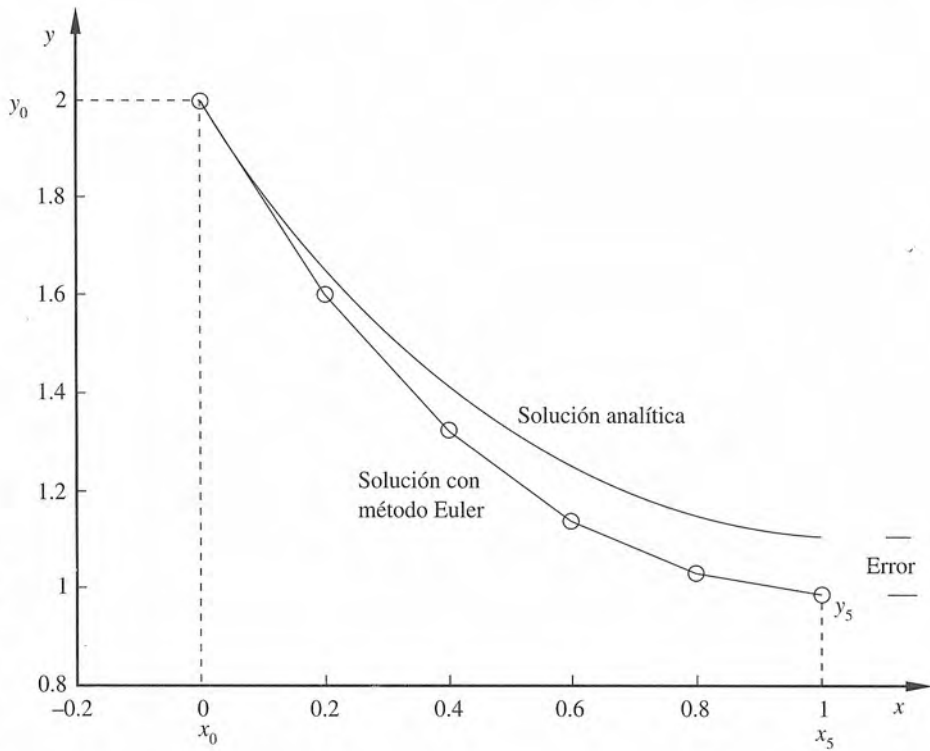
$$y_3 = y(0.6) = 1.32 + 0.2[0.4 - 1.32] = 1.136$$

$$y_4 = y(0.8) = 1.136 + 0.2[0.6 - 1.136] = 1.0288$$

$$y_5 = y(1.0) = 1.0288 + 0.2[0.8 - 1.0288] = 0.98304$$

Por otro lado, la solución analítica es 1.10364 (el lector puede verificarla resolviendo analíticamente el PVI); el error cometido es 0.1206 en valor absoluto y 10.92 en por ciento (ver Fig. 7.5).





**Figura 7.5**  
Solución  
analítica en  
contraste con el  
método de  
Euler aplicado  
cinco veces.

### ALGORITMO 7.1 Método de Euler

Para obtener la aproximación  $Y_F$  a la solución de un problema de valor inicial o PVI (Ec. 7.11), proporcionar la función  $F(X,Y)$  y los

**DATOS:** La condición inicial  $X_0, Y_0$ , el valor  $X_F$  donde se desea conocer el valor de  $Y_F$  y el número  $N$  de subintervalos por emplear.

**RESULTADOS:** Aproximación a  $Y_F$ :  $Y_0$ .

PASO 1. Hacer  $H = (X_F - X_0)/N$

PASO 2. Hacer  $I = 1$

PASO 3. Mientras  $I \leq N$ , repetir los pasos 4 a 6.

PASO 4. Hacer  $Y_0 = Y_0 + H * F(X_0, Y_0)$

PASO 5. Hacer  $X_0 = X_0 + H$

PASO 6. Hacer  $I = I + 1$

PASO 7. IMPRIMIR  $Y_0$  y TERMINAR.

## 7.3 Métodos de Taylor

Antes de explicar estos métodos, conviene hacer una acotación al método de Euler.

Puede decirse que el método de Euler utiliza los primeros dos términos de la serie de Taylor para su primera iteración; o sea,

$$F(x_1) \approx y_1 = F(x_0) + F'(x_0)(x_1 - x_0) \quad (7.18)$$

donde se señala que  $y_1$  no es igual a  $F(x_1)$ .

Esto pudo hacer pensar que para encontrar  $y_2$ , se expandió de nuevo  $F(x)$  en serie de Taylor, como sigue:

$$F(x_2) \approx y_2 = F(x_1) + F'(x_1) (x_2 - x_1), \quad (7.19)$$

sin embargo, no se dispone de los valores exactos de  $F(x_1)$  y  $F'(x_1)$  y, rigurosamente hablando, son los que deben usarse en una expansión de Taylor de  $F(x)$  —en este caso alrededor de  $x_1$ —; por tanto, el lado derecho de la ecuación 7.19 no es evaluable. Por ello, sólo en la primera iteración, para encontrar  $y_1$ , se usa realmente una expansión en serie de Taylor de  $F(x)$ , aceptando desde luego que se tienen valores exactos en la condición inicial  $y_0 = F(x_0)$ . Después de eso, se emplea la ecuación

$$\begin{aligned} y_{i+1} &= y_i + f(x_i, y_i) (x_{i+1} - x_i) \\ &= F(x_i) + F'(x_i) (x_{i+1} - x_i) \end{aligned} \quad (7.20)$$

que guarda similitud con una expansión en serie de Taylor.

Aclarado este punto, a continuación se aplicará la información acerca de las series de Taylor para mejorar la exactitud del método de Euler y obtener extensiones que constituyen la familia de métodos llamados **algoritmos de Taylor**.

Si se usan tres términos en lugar de dos en la expansión de  $F(x_1)$ , entonces:

$$F(x_1) \approx y_1 = F(x_0) + F'(x_0) (x_1 - x_0) + F''(x_0) \frac{(x_1 - x_0)^2}{2!} \quad (7.21)$$

Como

$$F''(x) = \frac{dF'(x)}{dx} = \frac{df(x, y)}{dx},$$

y

$$h = x_1 - x_0,$$

la primera iteración (Ec. 7.21) tomaría la forma\*

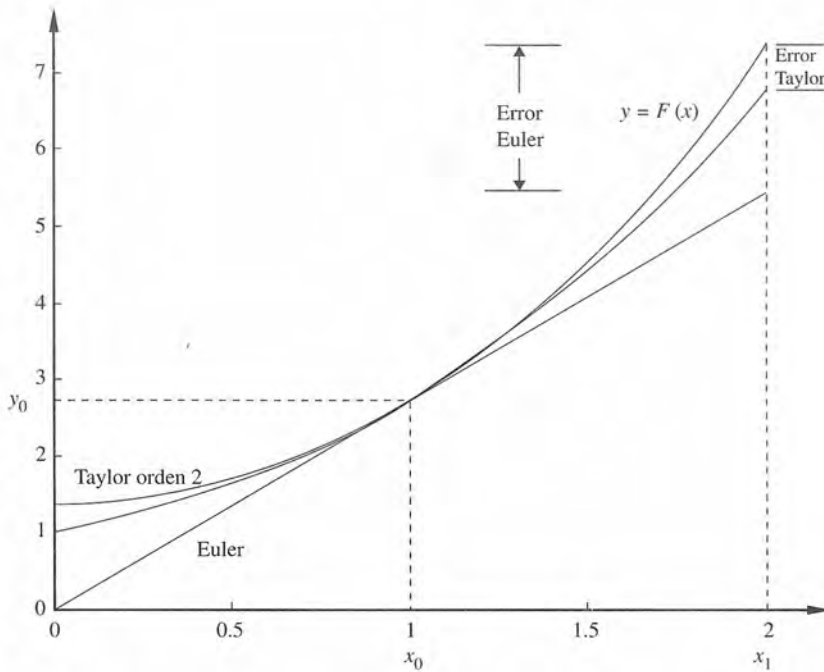
$$y_1 = y_0 + hf(x_0, y_0) + \frac{h^2}{2!} \left. \frac{df(x, y)}{dx} \right|_{x_0, y_0} \quad (7.22)$$

Ahora cabe pensar que usando una fórmula de iteración basada en la ecuación 7.22 para obtener  $y_2, y_3, \dots, y_n$  mejoraría la exactitud obtenida con la 7.18. Se propone entonces la fórmula

$$\boxed{y_{i+1} = y_i + hf(x_i, y_i) + \frac{h^2}{2!} \left. \frac{df(x, y)}{dx} \right|_{x_i, y_i}} \quad (7.23)$$

que equivaldría a usar una curva que pasa por el punto  $(x_0, y_0)$ , cuya pendiente y segunda derivada sean iguales que las de la función desconocida  $F(x)$  en el punto  $(x_0, y_0)$ . Como puede verse en la figura 7.6, se obtiene en general una mejor aproximación que con el método de Euler, aunque con un mayor número de cálculos.

\* La notación  $\left. \frac{df(x, y)}{dx} \right|_{x_0, y_0}$  significa la evaluación de la derivada de  $f(x, y)$  con respecto a  $x$  en el punto  $(x_0, y_0)$ .



**Figura 7.6**  
Comparación  
gráfica de  
los errores  
del método de  
Euler y el  
de Taylor de  
orden 2.

La utilidad de esta ecuación depende de cuán fácil sea la diferenciación de  $f(x, y)$ . Si  $f(x, y)$  es una función sólo de  $x$ , la diferenciación con respecto a  $x$  es relativamente fácil y la fórmula propuesta es muy práctica.

Si, como es el caso general,  $f(x, y)$  es una función de  $x$  y  $y$ , habrá que usar derivadas totales. La derivada total de  $f(x, y)$  con respecto a  $x$  está dada por:

$$\frac{df(x, y)}{dx} = \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} \frac{dy}{dx}$$

Si se aplican las ideas vistas en el método de Euler, pero empleando como fórmula la ecuación 7.23, se obtiene el método de Taylor de **segundo orden**. Esto último es indicativo de la derivada de mayor orden que se emplea y de cierta exactitud. Con esta terminología, al método de Euler le correspondería el nombre de **método de Taylor de primer orden**.

### Ejemplo 7.2

Resuelva el PVI del ejemplo 7.1 por el método de Taylor de segundo orden. Puede usar un pizarrón electrónico para seguir los cálculos.

**Solución** Al utilizar de nuevo cinco intervalos se tiene

$$\begin{array}{llll} h = 0.2, & x_0 = 0.0, & x_1 = 0.2, & x_2 = 0.4, \\ & x_3 = 0.6, & x_4 = 0.8, & x_5 = 1.0 \end{array}$$

Se aplica la ecuación 7.23 con  $y_0 = 2$  y con

$$\frac{df(x, y)}{dx} = \frac{\partial f(x, y)}{\partial x} + \frac{\partial f(x, y)}{\partial y} (x - y) = 1 - x + y$$

ya que  $\frac{dy}{dx} = x - y$

$$y_1 = y(0.2) = y_0 + h(x_0 - y_0) + \frac{h^2}{2!}(1 - x_0 + y_0)$$



$$\begin{aligned}
&= 2 + 0.2(0 - 2) + \frac{0.2^2}{2}(1 - 0 + 2) = 1.66 \\
y_2 &= y(0.4) = y_1 + h(x_1 - y_1) + \frac{h^2}{2!}(1 - x_1 + y_1) \\
&= 1.66 + 0.2(0.2 - 1.66) + \frac{0.2^2}{2}(1 - 0.2 + 1.66) = 1.4172
\end{aligned}$$

al continuar este procedimiento se llega a

$$y_5 = y(1.0) = 1.11222$$

que da un error absoluto de 0.00858 y un error porcentual de 0.78. Nótese la mayor exactitud y el mayor número de cálculos.

La extensión de esta idea a cuatro, cinco o más términos de la serie de Taylor significaría obtener métodos con mayor exactitud, pero menos prácticos, ya que incluirían diferenciaciones complicadas de  $f(x, y)$ ; por ejemplo, si se quisieran usar cuatro términos de la serie, se necesitaría la segunda derivada de  $f(x, y)$ , la cual está dada por:

$$\begin{aligned}
\frac{d^2f(x, y)}{dx^2} &= \frac{\partial^2f(x, y)}{\partial x^2} + 2 \frac{dy}{dx} \frac{\partial^2f(x, y)}{\partial x \partial y} + \left(\frac{dy}{dx}\right)^2 \frac{\partial^2f(x, y)}{\partial y^2} \\
&\quad + \frac{\partial f(x, y)}{\partial x} \frac{\partial f(x, y)}{\partial y} + \left(\frac{\partial f(x, y)}{\partial y}\right)^2 \frac{dy}{dx}
\end{aligned}$$

Las derivadas totales de orden superior al segundo de  $f(x, y)$  son aún más largas y complicadas.

Ya que el uso de varios términos de la serie de Taylor presenta serias dificultades, los investigadores han buscado métodos comparables con ellos en exactitud pero más fáciles. De hecho, el patrón para evaluarlos son los métodos derivados de la serie de Taylor; por ejemplo, dado un método, se compara con el derivado de la serie de Taylor que proporcione la misma exactitud. La derivada de más alto orden en este último confiere el orden del primero. Un método que diera una exactitud comparable al método de Euler sería de **primer orden**; si proporcionara una exactitud comparativamente igual a usar tres términos de la serie de Taylor, sería de **segundo orden**, y así sucesivamente.

A continuación se estudian métodos de orden dos, tres, etc., en los que no se requieren diferenciaciones de  $f(x, y)$ .

## 7.4 Método de Euler modificado

En el método de Euler se tomó como válida para todo el primer subintervalo la derivada encontrada en un extremo de éste (véase Fig. 7.3). Para obtener una exactitud razonable se utiliza un intervalo muy pequeño, a cambio de un error de redondeo mayor (ya que se realizarán más cálculos).

El método de Euler modificado trata de evitar este problema utilizando un valor promedio de la derivada tomada en los dos extremos del intervalo, en lugar de la derivada tomada en un solo extremo.

El método de Euler modificado consta de dos pasos básicos\*:

\* Se omitió la subdivisión de  $[x_0, x_f]$  en  $n$  subintervalos para dar énfasis a los pasos fundamentales de **predicción y corrección**.

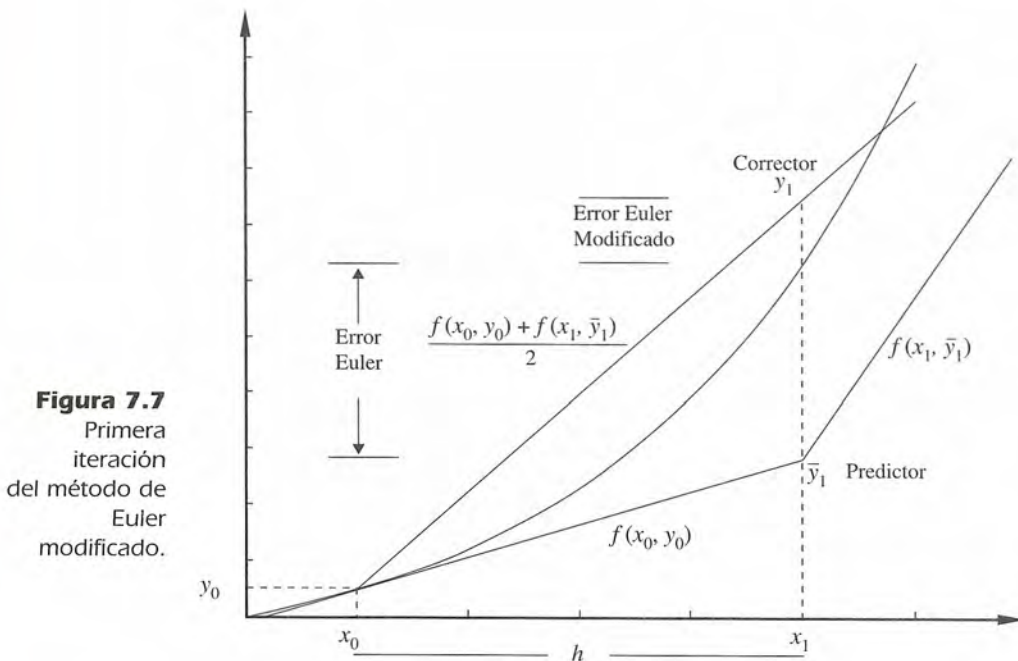
1. Se parte de  $(x_0, y_0)$  y se utiliza el método de Euler a fin de calcular el valor de  $y$  correspondiente a  $x_1$ . Este valor de  $y$  se denotará aquí como  $\bar{y}_1$ , ya que sólo es un valor transitorio para  $y_1$ . Esta parte del proceso se conoce como paso **predictor**.
2. El segundo paso se llama **corrector**, pues trata de corregir la predicción. En el nuevo punto obtenido  $(x_1, y_1)$  se evalúa la derivada  $f(x_1, y_1)$  usando la ecuación diferencial ordinaria del PVI que se esté resolviendo; se obtiene la media aritmética de esta derivada y la derivada en el punto inicial  $(x_0, y_0)$ .

$$\frac{1}{2} [f(x_0, y_0) + f(x_1, \bar{y}_1)] = \text{derivada promedio}$$

Se usa la derivada promedio para calcular un nuevo valor de  $y_1$ , con la ecuación 7.17 que deberá ser más exacto que  $y_1$

$$y_1 = y_0 + \frac{(x_1 - x_0)}{2} [f(x_0, y_0) + f(x_1, \bar{y}_1)]$$

y que se tomará como valor definitivo de  $y_1$  (ver Fig. 7.7). Este procedimiento se repite hasta llegar a  $y_n$ .



**Figura 7.7**  
Primera  
iteración  
del método de  
Euler  
modificado.

El esquema iterativo para este método quedaría en general así.

Primero, usando el paso de predicción resulta:

$$\bar{y}_{i+1} = y_i + hf(x_i, y_i). \quad (7.24a)$$

Una vez obtenida  $\bar{y}_{i+1}$  se calcula  $f(x_{i+1}, \bar{y}_{i+1})$ , la derivada en el punto  $(x_{i+1}, \bar{y}_{i+1})$ , y se promedia con la derivada previa  $f(x_i, y_i)$  para encontrar la derivada promedio

$$\frac{1}{2} [f(x_i, y_i) + f(x_{i+1}, \bar{y}_{i+1})]$$

Se sustituye  $f(x_i, y_i)$  con este valor promedio en la ecuación de iteración de Euler y se obtiene:

$$y_{i+1} = y_i + \frac{h}{2} [f(x_i, y_i) + f(x_{i+1}, \bar{y}_{i+1})] \quad (7.24b)$$

### Ejemplo 7.3

Resuelva el PVI del ejemplo 7.1 por el método de Euler modificado.

#### Solución



Al utilizar nuevamente cinco intervalos para que la comparación de los resultados obtenidos sea consistente con los anteriores se tiene:

#### Primera iteración

$$\text{Primer paso: } \bar{y}_1 = y_0 + hf(x_0, y_0) = 2 + 0.2(0 - 2) = 1.6$$

$$\text{Segundo paso: } \frac{1}{2} [f(x_0, y_0) + f(x_1, \bar{y}_1)] = \frac{1}{2} [(0 - 2) + (0.2 - 1.6)] = -1.7$$

$$y(0.2) = y_1 = 2 + 0.2(-1.7) = 1.66$$

#### Segunda iteración

$$\text{Primer paso: } \bar{y}_2 = y_1 + hf(x_1, y_1) = 1.66 + 0.2(0.2 - 1.66) = 1.368$$

$$\text{Segundo paso: } \frac{1}{2} [f(x_1, y_1) + f(x_2, \bar{y}_2)] = \frac{1}{2} [(0.2 - 1.66) + (0.4 - 1.368)] = -1.214$$

$$y(0.4) = y_2 = 1.66 + 0.2(-1.214) = 1.4172$$

Al continuar los cálculos se llega a

$$\bar{y}_5 = 1.08509$$

$$y_5 = 1.11222$$

Los resultados obtenidos en este caso son idénticos a los del ejemplo 7.2 en que se utilizó el método de Taylor de segundo orden; por tanto, presumiblemente el método de Euler modificado es de segundo orden. Esto se demuestra en la siguiente sección.

### ALGORITMO 7.2 Método de Euler modificado

Para obtener la aproximación YF a la solución de un PVI, proporcionar la función F(X, Y) y los

DATOS: La condición inicial X0, Y0, el valor XF donde se desea conocer el valor de YF y el número N de subintervalo por emplear.

RESULTADOS: Aproximación a YF: Y0.

PASO 1. Hacer  $H = (XF - X0) / N$

PASO 2. Hacer  $I = 1$

PASO 3. Mientras  $I \leq N$ , repetir los pasos 4 a 7.

PASO 4. Hacer  $Y1 = Y0 + H * F(X0, Y0)$

PASO 5. Hacer  $Y0 = Y0 + H/2 * (F(X0, Y0) + F(X0+H, Y1))$

PASO 6. Hacer  $X0 = X0 + H$

PASO 7. Hacer  $I = I + 1$

PASO 8. IMPRIMIR Y0 y TERMINAR



## 7.5 Métodos de Runge-Kutta

Los métodos asociados con los nombres de Runge (1885), Kutta (1901), Heun (1900) y otros para resolver el PVI (Ec. 7.11) consisten en obtener un resultado que se obtendría al utilizar un número finito de términos de una serie de Taylor de la forma

$$y_{i+1} = y_i + hf(x_i, y_i) + \frac{h^2}{2!} f'(x_i, y_i) + \frac{h^3}{3!} f''(x_i, y_i) + \dots \quad (7.25)$$

con una aproximación en la cual se calcula  $y_{i+1}$  de una fórmula del tipo\*

$$y_{i+1} = y_i + h [\alpha_0 f(x_i, y_i) + \alpha_1 f(x_i + \mu_1 h, y_i + b_1 h) + \alpha_2 f(x_i + \mu_2 h, y_i + b_2 h) + \dots + \alpha_p f(x_i + \mu_p h, y_i + b_p h)] \quad (7.26)$$

donde las  $\alpha$ ,  $\mu$  y  $b$  se determinan de modo que si se expandiera  $f(x_i + \mu_j h, y_i + b_j h)$ , con  $1 \leq j \leq p$  en series de Taylor alrededor de  $(x_i, y_i)$ , se observaría que los coeficientes de  $h$ ,  $h^2$ ,  $h^3$ , etc., coincidirían con los coeficientes correspondientes de la ecuación 7.25.

A continuación se derivará sólo el caso más simple, cuando  $p = 1$ , para ilustrar el procedimiento del caso general, ya que los lineamientos son los mismos.

A fin de simplificar y sistematizar la derivación, conviene expresar la ecuación 7.26 con  $p = 1$  en la forma

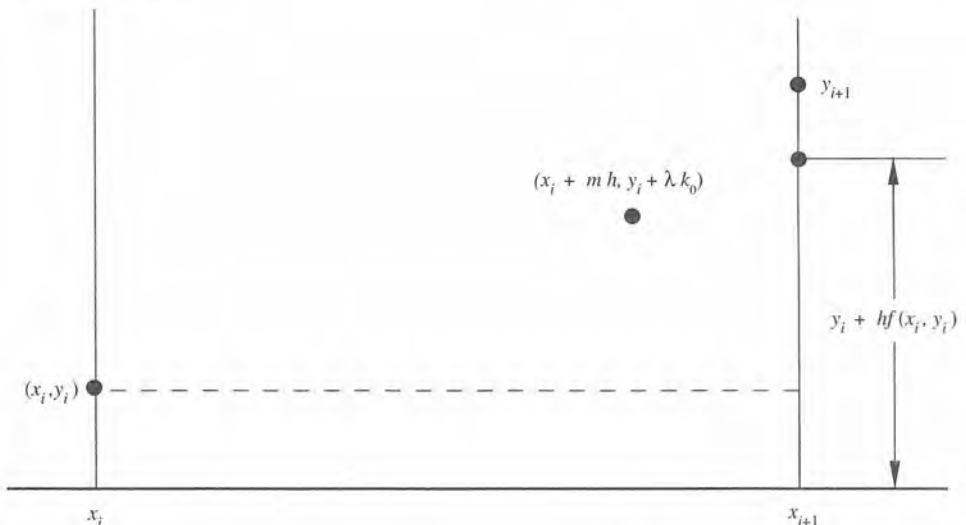
$$y_{i+1} = y_i + h[\alpha_0 f(x_i, y_i) + \alpha_1 f(x_i + \mu h, y_i + b h)] \quad (7.27)$$

Obsérvese que en esta expresión se evalúa  $f$  en  $(x_i, y_i)$  y  $(x_i + \mu h, y_i + b h)$ . El valor  $x_i + \mu h$  es tal que  $x_i < x_i + \mu h \leq x_{i+1}$  para mantener la abscisa del segundo punto dentro del intervalo de interés (véase Fig. 7.8), con lo que  $0 < \mu \leq 1$ .

Por otro lado,  $b$  puede manejarse más libremente y expresarse  $y_i + b h$ , sin pérdida de generalidad, como una ordenada arriba o abajo de la ordenada que da el método de Euler simple

$$y_i + b h = y_i + \lambda h f(x_i, y_i) = y_i + \lambda k_0 \quad (7.28)$$

con  $k_0 = h f(x_i, y_i)$



**Figura 7.8**  
Deducción del  
método de  
Runge-Kutta.

\* Nótese que en la ecuación 7.26 ya no aparecen diferenciaciones, sólo evaluaciones de  $f(x, y)$ .

Queda entonces por determinar  $\alpha_0$ ,  $\alpha_1$ ,  $\mu$  y  $\lambda$  tales que la ecuación 7.27 tenga una expansión en potencias de  $h$  cuyos primeros términos, tantos como sea posible, coincidan con los primeros términos de la 7.25.

Para obtener los parámetros desconocidos, se expande primero  $f(x_i + \mu h, y_i + \lambda k_0)$  en serie de Taylor (obviamente mediante el desarrollo de Taylor de funciones de dos variables).\*

$$\begin{aligned} f(x_i + \mu h, y_i + \lambda k_0) &= f(x_i, y_i) + \mu h \frac{\partial f}{\partial x} + \lambda k_0 \frac{\partial f}{\partial y} + \frac{\mu^2 h^2}{2!} \frac{\partial^2 f}{\partial x^2} + \\ &+ \mu h \lambda k_0 \frac{\partial^2 f}{\partial x \partial y} + \frac{\lambda^2 k_0^2}{2!} \frac{\partial^2 f}{\partial y^2} + O(h^3) \end{aligned} \quad (7.29)$$

Todas las derivadas parciales son evaluadas en  $(x_i, y_i)$ .

Se sustituye en la ecuación 7.27

$$\begin{aligned} y_{i+1} &= y_i + \alpha_0 h f(x_i, y_i) + \alpha_1 h [f(x_i, y_i) + \mu h \frac{\partial f}{\partial x} + \lambda k_0 \frac{\partial f}{\partial y} \\ &+ \frac{\mu^2 h^2}{2!} \frac{\partial^2 f}{\partial x^2} + \mu h \lambda k_0 \frac{\partial^2 f}{\partial x \partial y} + \frac{\lambda^2 k_0^2}{2!} \frac{\partial^2 f}{\partial y^2} + O(h^3)] \end{aligned}$$

Esta última ecuación se arregla en potencias de  $h$ , y queda:

$$\begin{aligned} y_{i+1} &= y_i + h(\alpha_0 + \alpha_1) f(x_i, y_i) + h^2 \alpha_1 \left( \mu \frac{\partial f}{\partial x} + \lambda f(x_i, y_i) \frac{\partial f}{\partial y} \right) \\ &+ \frac{h^3}{2} \alpha_1 \left( \mu^2 \frac{\partial^2 f}{\partial x^2} + 2\mu \lambda f(x_i, y_i) \frac{\partial^2 f}{\partial x \partial y} + \lambda^2 f^2(x_i, y_i) \frac{\partial^2 f}{\partial y^2} \right) + O(h^4) \end{aligned} \quad (7.30)$$

Para que los coeficientes correspondientes de  $h$  y  $h^2$  coincidan en las ecuaciones 7.25 y 7.30 se requiere:

$$\begin{aligned} \alpha_0 + \alpha_1 &= 1 \\ \mu \alpha_1 &= \frac{1}{2}, \quad \lambda \alpha_1 = \frac{1}{2} \end{aligned} \quad (7.31)$$

Hay cuatro incógnitas para sólo tres ecuaciones y, por tanto, se tiene un grado de libertad en la solución de la ecuación 7.31. Podría pensarse en usar este grado de libertad para hacer coincidir los coeficientes de  $h^3$ . Sin embargo, es obvio que esto es imposible para cualquier forma que tenga la función  $f(x, y)$ . Existe entonces un número infinito de soluciones de la ecuación 7.31, pero quizá la más simple sea:

$$\alpha_0 = \alpha_1 = \frac{1}{2}; \quad \mu = \lambda = 1$$

Esta elección conduce al sustituir en la ecuación 7.27 a:

$$y_{i+1} = y_i + \frac{h}{2} [f(x_i, y_i) + f(x_i + h, y_i + hf(x_i, y_i))]$$

o bien

\* Spiegel, M.R. *Manual de fórmulas y tablas matemáticas*, Schaum. McGraw Hill. Serie Schaum. (1970), p. 113.

$$\begin{array}{l}
 y_{i+1} = y_i + \frac{h}{2} (k_0 + k_1) \\
 \text{con} \quad k_0 = f(x_i, y_i) \quad ; \quad k_1 = f(x_i + h, y_i + hk_0)
 \end{array}
 \quad (7.32)$$

conocida como **algoritmo de Runge-Kutta de segundo orden** (lo de segundo orden por coincidir con los primeros tres términos de la serie de Taylor), y que es la fórmula del método de Euler modificado, con dos pasos sintetizados en uno.

Por ser orden superior al de Euler, este método proporciona mayor exactitud (véase Ej. 7.3); por tanto, es posible usar un valor de  $h$  no tan pequeño como en el primero. El precio es la evaluación de  $f(x, y)$  dos veces en cada subintervalo, contra una en el método de Euler.

Las fórmulas de Runge-Kutta de cualquier orden se puede derivar en la misma forma en que se llega a la ecuación 7.32.

El **método de Runge-Kutta de cuarto orden** (igual que para orden dos, existen muchos métodos de cuarto orden) es una de las fórmulas más usadas de esta familia y está dado como:

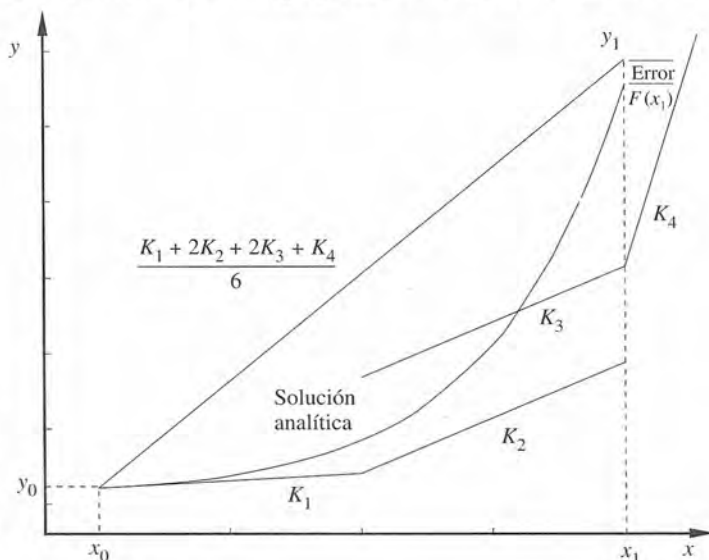
$$y_{i+1} = y_i + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4), \quad (7.33)$$

donde:

$$\begin{aligned}
 k_1 &= f(x_i, y_i) \\
 k_2 &= f(x_i + h/2, y_i + hk_1/2) \\
 k_3 &= f(x_i + h/2, y_i + hk_2/2) \\
 k_4 &= f(x_i + h, y_i + hk_3)
 \end{aligned}$$

En la ecuación 7.33 hay coincidencia con los primeros cinco términos de la serie de Taylor, lo cual significa gran exactitud sin cálculo de derivadas; pero a cambio, hay que evaluar la función  $f(x, y)$  cuatro veces en cada subintervalo.

Al igual que en el método de Euler modificado, puede verse a los métodos de Runge-Kutta como la ponderación de pendientes,  $k_1, k_2, k_3$  y  $k_4$  con pesos 1, 2, 2, 1, respectivamente para el caso de cuarto orden, dando lugar a una recta de pendiente  $(k_1 + 2k_2 + 2k_3 + k_4)/6$  y que pasa por el punto  $(x_0, y_0)$  y que es la que se usa para obtener  $y_1$  (ver Fig. 7.9).



**Figura 7.9**  
Interpretación  
gráfica del  
método de  
Runge-Kutta de  
cuarto orden.



**Ejemplo 7.4**

Resuelva el PVI del ejemplo 7. 1 por el método de Runge-Kutta de cuarto orden (RK-4). Se recomienda usar un pizarrón electrónico.

**Solución**

Al tomar nuevamente cinco subintervalos y emplear la ecuación 7.33 se tiene

**Primera iteración**

Cálculo de las constantes  $k_1, k_2, k_3, k_4$

$$\begin{aligned} k_1 &= f(x_0, y_0) = (0 - 2) = -2 \\ k_2 &= f(x_0 + h/2, y_0 + hk_1/2) = [(0 + 0.2/2) - (2 + 0.2(-2)/2)] = -1.7 \\ k_3 &= f(x_0 + h/2, y_0 + hk_2/2) = [(0 + 0.2/2) - (2 + 0.2(-1.7)/2)] = -1.73 \\ k_4 &= f(x_0 + h, y_0 + hk_3) = [(0 + 0.2) - (2 + 0.2(-1.73))] = -1.454 \end{aligned}$$

Cálculo de  $y_1$

$$\begin{aligned} y(0.2) &= y_1 = y_0 + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \\ &= 2 + (0.2/6) (-2 + 2(-1.7) + 2(-1.73) - 1.454) = 1.6562 \end{aligned}$$

**Segunda iteración**

Cálculo de las constantes  $k_1, k_2, k_3, k_4$

$$\begin{aligned} k_1 &= f(x_1, y_1) = (0.2 - 1.6562) = -1.4562 \\ k_2 &= f(x_1 + h/2, y_1 + hk_1/2) = [(0.2 + 0.2/2) - (1.6562 + 0.2(-1.4562)/2)] = -1.21058 \\ k_3 &= f(x_1 + h/2, y_1 + hk_2/2) = [(0.2 + 0.2/2) - (1.6562 + 0.2(-1.21058)/2)] = -1.235142 \\ k_4 &= f(x_1 + h, y_1 + hk_3) = [(0.2 + 0.2) - (1.6562 + 0.2(-1.235142))] = -1.0091716 \end{aligned}$$

Cálculo de  $y_2$

$$\begin{aligned} y(0.4) &= y_2 = y_1 + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \\ &= 1.6562 + (0.2/6) (-1.4562 + 2(-1.21058) + 2(-1.235142) - 1.0091716) = 1.410972813 \end{aligned}$$

Con la continuación de este procedimiento se obtiene:

$$\begin{aligned} y(0.6) &= y_3 = 1.246450474 \\ y(0.8) &= y_4 = 1.148003885 \\ y(1.0) &= y_5 = 1.103655714 \end{aligned}$$

que da un error absoluto de 0.00001 y un error porcentual de 0.0009.

Los cálculos pueden realizarse con la TI-92 Plus.



```
e7_4()
Prgm
Define f(x, y) = x-y
0→x0: 2→y0: 0.2→h: ClrIO
Disp "k x(k) y(k)"
Disp "0 "&format(x0, "f1")&" "&format(y0, "f9")
For i, 1, 5
f(x0, y0)→k1
f(x0+h/2, y0+h/2*k1)→k2
f(x0+h/2, y0+h/2*k2)→k3
```

```

f(x0+h, y0+h*k3) → k4
y0+h/6*(k1+2*k2+2*k3+k4) → y0
x0+h → x0
Disp format(i, "f0")&" "&format(x0, "f1")&" "&format(y0, "f9")
EndFor
EndPrm

```

Matlab proporciona un conjunto de funciones para resolver sistemas de ecuaciones diferenciales. A continuación se muestra cómo usar Matlab para resolver este ejemplo con una de dichas funciones.

Se escribe una función con el vector de funciones (en este caso de un solo elemento) y se graba con el nombre E74.m, por ejemplo.

```

function f=E74(x,y)
f(1) = x-y;

```

Después se usa el siguiente guión:

```

xx=0 : 0.2:1; y0=[2];
[T,Y]=ode45('e74', xx, y0);
Y

```

### ALGORITMO 7.3 Método de Runge-Kutta de cuarto orden

Para obtener la aproximación YF a la solución de un PVI, proporcionar la función F(X,Y) y los

DATOS: La condición inicial  $X_0, Y_0$ , el valor XF donde se desea conocer el valor de YF y el número N de subintervalos a emplear.

RESULTADOS: Aproximación a YF: Y0

PASO 1. Hacer  $H = (XF - X_0)/N$

PASO 2. Hacer  $I = 1$

PASO 3. Mientras  $I \leq N$ , repetir los pasos 4 a 10.

PASO 4. Hacer  $K_1 = F(X_0, Y_0)$

PASO 5. Hacer  $K_2 = F(X_0 + H/2, Y_0 + H * K_1/2)$

PASO 6. Hacer  $K_3 = F(X_0 + H/2, Y_0 + H * K_2/2)$

PASO 7. Hacer  $K_4 = F(X_0 + H, Y_0 + H * K_3)$

PASO 8. Hacer  $Y_0 = Y_0 + H/6 * (K_1 + 2*K_2 + 2*K_3 + K_4)$

PASO 9. Hacer  $X_0 = X_0 + H$

PASO 10. Hacer  $I = I + 1$

PASO 11. IMPRIMIR Y0 y TERMINAR.

Los métodos descritos hasta aquí se conocen como **métodos de un solo paso** porque se apoyan y usan el punto  $(x_i, y_i)$  para el cálculo de  $y_{i+1}$  (por ejemplo los métodos de Taylor). Los métodos de Runge-Kutta además se apoyan en puntos entre  $x_i$  y  $x_{i+1}$  pero nunca en

puntos anteriores a  $x_i$ . Sin embargo, si se usa información previa a  $x_i$  para el cálculo de  $y_{i+1}$ , es posible obtener otras familias de métodos con otras características distintas a las ya vistas. A estos métodos se les llama **métodos de múltiples pasos o métodos de predicción-corrección**.

## 7.6 Métodos de predicción-corrección

En el esquema iterativo del método de Euler modificado (Sección 7.4) se utiliza la fórmula

$$y_{i+1} = y_i + \frac{h}{2} [f(x_i, y_i) + f(x_{i+1}, \bar{y}_{i+1})]$$

El segundo término del miembro derecho de esta ecuación recuerda la integración trapezoidal compuesta del capítulo 6.

Para ver mejor esta similitud, recuérdese que la solución analítica de la ecuación diferencial del PVI (Ec. 7.11) es:

$$y = F(x)$$

y que

$$F'(x) = f(x, y),$$

e integrando ambos miembros con respecto a  $x$

$$\int F'(x) dx = F(x) = \int f(x, y) dx$$

A partir de que  $F(x)$  es la integral indefinida de  $f(x, y)$ , se integra  $f(x, y)$  entre los límites de  $x : x_i$  y  $x_{i+1}$  para obtener:

$$\begin{aligned} \int_{x_i}^{x_{i+1}} f(x, y) dx &= F(x) \Big|_{x_i}^{x_{i+1}} \\ &= F(x_{i+1}) - F(x_i) \approx y_{i+1} - y_i \end{aligned} \quad (7.34)$$

donde  $y_i$  y  $y_{i+1}$  son aproximaciones a  $F(x_i)$  y  $F(x_{i+1})$ , respectivamente.

Por otro lado, es factible realizar la misma integración, pero con una aproximación trapezoidal entre los puntos  $(x_i, y_i)$  y  $(x_{i+1}, y_{i+1})$ , donde  $y_{i+1}$  se obtuvo en el paso de predicción.

$$\int_{x_i}^{x_{i+1}} f(x, y) dx \approx \frac{h}{2} [f(x_i, y_i) + f(x_{i+1}, \bar{y}_{i+1})] \quad (7.35)$$

donde  $h$  es la altura del trapecioide

$$h = x_{i+1} - x_i$$

Al igualar las integrales 7.34 y 7.35, se tiene

$$y_{i+1} - y_i = \frac{h}{2} [f(x_i, y_i) + f(x_{i+1}, \bar{y}_{i+1})]$$

o bien

$$y_{i+1} = y_i + \frac{h}{2} [f(x_i, y_i) + f(x_{i+1}, \bar{y}_{i+1})]$$



que da la ecuación de corrección del método de Euler modificado; de esta manera se establece la identificación de este algoritmo y la integración trapezoidal. Esto sugiere a su vez la obtención de esquemas iterativos de solución del PVI por medio de la regla de Simpson u otros métodos de integración numérica que usan mayor número de puntos.

A continuación se derivará un corrector basado en el método de Simpson 1/3.

La ecuación 7.34 toma ahora la forma

$$\int_{x_{i-1}}^{x_{i+1}} f(x, y) dx = F(x_{i+1}) - F(x_{i-1}) \approx y_{i+1} - y_{i-1} \quad (7.36)$$

y la correspondiente a 7.35 queda:

$$\int_{x_{i-1}}^{x_{i+1}} f(x, y) dx \approx \frac{h}{3} [f(x_{i-1}, y_{i-1}) + 4f(x_i, y_i) + f(x_{i+1}, \bar{y}_{i+1})] \quad (7.37)$$

Nótese que se está integrando de  $x_{i-1}$  a  $x_{i+1}$  ya que se utilizan dos subintervalos para cada integración.

Al igualar 7.36 y 7.37 se llega a la fórmula de corrección

$$y_{i+1} = y_{i-1} + \frac{h}{3} [f(x_{i-1}, y_{i-1}) + 4f(x_i, y_i) + f(x_{i+1}, \bar{y}_{i+1})] \quad (7.38)$$

donde nuevamente hay que obtener  $\bar{y}_{i+1}$  con un predictor.

Al partir de  $(x_0, y_0)$ , la ecuación 7.38 tomaría la forma

$$y_2 = y_0 + \frac{h}{3} [f(x_0, y_0) + 4f(x_1, y_1) + f(x_2, \bar{y}_2)] \quad (7.39)$$

para su primera aplicación. En 7.39  $\bar{y}_2$  es estimada con un predictor, el cual a su vez requiere  $y_1$  y  $f(x_1, y_1)$ . Así pues, antes de realizar la primera predicción deben evaluarse ciertos valores iniciales [en este caso  $y_1$  y  $f(x_1, y_1)$ ].

En esta evaluación se usa alguno de los métodos ya vistos (los de Runge-Kutta, por ejemplo). Este paso se utiliza sólo una vez en el proceso iterativo y se conoce como **paso de inicialización**.

Es evidente que para la predicción también puede utilizarse un método de los ya estudiados o, como se verá más adelante, puede derivarse un predictor usando las mismas ideas que condujeron a la ecuación 7.39

### Ejemplo 7.5

Resuelva el problema de valor inicial del ejemplo 7.1 utilizando el corrector dado por la ecuación 7.38 y el método Euler modificado como inicializador y como predictor.

**Solución** El intervalo se divide otra vez en cinco subintervalos y se tiene

#### Primera iteración

Inicialización (se toma el valor de  $y_1$  del ejemplo 7.3):

$$y_1 = 1.66$$

Predicción: (se toma el valor de  $y_2$  del ejemplo 7.3):

$$\bar{y}_2 = 1.4172$$

Corrección: se utiliza la ecuación 7.39 (puede usar un pizarrón electrónico)

$$\begin{aligned} y(0.4) &= y_2 = 2 + \frac{0.2}{3} [(0 - 2) + 4(0.2 - 1.66) + (0.4 - 1.4172)] \\ &= 1.40952 \end{aligned}$$

**Segunda iteración**

Predicción

$$\begin{aligned}\bar{y}_3 &= y_2 + \frac{h}{2} [f(x_2, y_2) + f(x_2 + h, y_2 + hf(x_2, y_2))] \\ &= 1.40952 + \frac{0.2}{2} [(0.4 - 1.40952) + [(0.4 + 0.2) - (1.40952) \\ &\quad + 0.2 (0.4 - 1.40952)]] = 1.2478064\end{aligned}$$

Corrección (con la ecuación 7.38)

$$\begin{aligned}y(0.6) &= y_3 = y_1 + \frac{h}{3} [f(x_1, y_1) + 4f(x_2, y_2) + f(x_3, \bar{y}_3)] \\ &= 1.66 + \frac{0.2}{3} [(0.2 - 1.66) + 4(0.4 - 1.40952) \\ &\quad + (0.6 - 1.2478064)] = 1.25027424\end{aligned}$$

**Tercera iteración**

Predicción

$$\begin{aligned}\bar{y}_4 &= y_3 + \frac{h}{2} [f(x_3, y_3) + f(x_3 + h, y_3 + hf(x_3, y_3))] \\ &= 1.25027424 + \frac{0.2}{2} [(0.6 - 1.25027424) + [(0.6 + 0.2) \\ &\quad - (1.25027424 + 0.2 (0.6 - 1.25027424))]] = 1.153224877\end{aligned}$$

Corrección (con la ecuación 7.38)

$$\begin{aligned}y(0.8) &= y_4 = y_2 + \frac{h}{3} [f(x_2, y_2) + 4f(x_3, y_3) + f(x_4, \bar{y}_4)] \\ &= 1.40952 + \frac{0.2}{3} [(0.4 - 1.40952) + 4(0.6 - 1.25027424) \\ &\quad + (0.8 - 1.153224877)] = 1.145263878\end{aligned}$$

**Cuarta iteración**

Predicción

$$\begin{aligned}\bar{y}_5 &= y_4 + \frac{h}{2} [f(x_4, y_4) + f(x_4 + h, y_4 + hf(x_4, y_4))] \\ &= 1.145263878 + \frac{0.2}{2} [(0.8 - 1.145263878) + [(0.8 + 0.2) \\ &\quad - (1.145263878 + 0.2 (0.8 - 1.145263878))]] = 1.10311638\end{aligned}$$

Corrección (con la ecuación 7.38)

$$\begin{aligned}y(1) &= y_5 = y_3 + \frac{h}{3} [f(x_3, y_3) + 4f(x_4, y_4) + f(x_5, \bar{y}_5)] \\ &= 1.25027424 + \frac{0.2}{3} [(0.6 - 1.25027424) + 4(0.8 - 1.145263878) \\ &\quad + (1 - 1.10311638)] = 1.107977831\end{aligned}$$

que da un error absoluto de 0.00434 y 0.0393 en porcentaje.

En general, puede obtenerse un corrector de cualquier orden utilizando la fórmula

$$y_{i+1} = y_{i-k} + \int_{x_{i-k}}^{x_{i+1}} f(x, y) dx, \quad k = 0, 1, 2, \dots \quad (7.40)$$

donde la integración se realiza sustituyendo  $f(x, y)$  con un polinomio de grado  $k + 1$  que pasa por  $(x_{i+1}, \bar{y}_{i+1})$ ,  $(x_i, y_i)$ ,  $\dots$ ,  $(x_{i-k}, y_{i-k})$ .

En virtud de que se está utilizando  $x_{i+1}$  y las abscisas previas a ésta y a sus espaciamentos regulares, lo más indicado para interpolar  $f(x, y)$  es el polinomio de interpolación en su forma de diferencias hacia atrás, dado por la ecuación 5.38 del capítulo 5. La ecuación 7.40 queda entonces:

$$y_{i+1} = y_{i-k} + \int_{x_{i-k}}^{x_{i+1}} p(x_i + sh) dx, \quad (7.41)$$

Para la obtención de  $p(x_i + sh)$ , dada por la ecuación 5.38, se empleó el cambio de variable

$$x = x_i + sh,$$

que permite escribir la ecuación 7.41 en términos de la nueva variable  $s$ , ya que

$$\begin{aligned} dx &= h ds \\ x_{i+1} &= x_i + sh && \text{de donde } s = 1 \\ x_{i-k} &= x_i + sh && \text{de donde } s = -k \end{aligned} \quad (7.42)$$

Al sustituir se llega a

$$y_{i+1} = y_{i-k} + h \int_{-k}^1 p(x_i + sh) ds$$

o bien

$$\begin{aligned} y_{i+1} &= y_{i-k} + h \int_{-k}^1 [f(x_{i+1}, \bar{y}_{i+1}) + (s-1) \nabla f(x_{i+1}, \bar{y}_{i+1}) + \\ &+ \frac{(s-1)s}{2!} \nabla^2 f(x_{i+1}, \bar{y}_{i+1}) + \frac{(s-1)s(s+1)}{3!} \nabla^3 f(x_{i+1}, \bar{y}_{i+1}) \\ &+ \dots + \frac{(s-1)s(s+1)\dots(s+r-2)}{r!} \nabla^r f(x_{i+1}, \bar{y}_{i+1})] ds \end{aligned}$$

La disimilitud de los coeficientes de las diferencias hacia atrás con los de la ecuación 5.38 se debe a que se está utilizando  $x_{i+1}$  como punto base. Si se denota por  $f_j = f(x_j, \bar{y}_j)$  para  $j = i-k, i-k+1, \dots, i+1$ , la última ecuación queda:

$$\begin{aligned} y_{i+1} &= y_{i-k} + h \int_{-k}^1 [f_{i+1} + (s-1) \nabla f_{i+1} + \frac{(s-1)s}{2!} \nabla^2 f_{i+1} + \\ &+ \frac{(s-1)s(s+1)}{3!} (s-1)s(s+1) \nabla^3 f_{i+1} + \dots + \frac{(s-1)s\dots(s+r-2)}{r!} \nabla^r f_{i+1}] ds \end{aligned} \quad (7.43)$$

y al integrar se llega a

$$\begin{aligned} y_{i+1} &= y_{i-k} + h [sf_{i+1} + s(\frac{s}{2} - 1) \nabla f_{i+1} + \frac{s^2(\frac{s}{3} - \frac{1}{2})}{2!} \nabla^2 f_{i+1} \\ &+ \frac{s^2(\frac{s^2}{4} - \frac{1}{2})}{3!} \nabla^3 f_{i+1} + \frac{(\frac{s^5}{5} + \frac{s^4}{2} - \frac{s^3}{3} - s^2)}{4!} \nabla^4 f_{i+1} + \text{términos restantes}] \Big|_{-k}^1 \end{aligned} \quad (7.44)$$



para  $k = 0, 1, 3$ , y  $5$ , la ecuación 7.44 da

$$k = 0$$

$$y_{i+1} = y_{i+h} [f_{i+1} - \frac{1}{2} \nabla f_{i+1} - \frac{1}{12} \nabla^2 f_{i+1} - \frac{1}{24} \nabla^3 f_{i+1} + \text{términos restantes}] \quad (7.44a)$$

$$k = 1$$

$$y_{i+1} = y_{i-1} + h [2f_{i+1} - 2\nabla f_{i+1} + \frac{1}{3} \nabla^2 f_{i+1} + 0 \nabla^3 f_{i+1} - \frac{1}{90} \nabla^4 f_{i+1} + \text{términos restantes}] \quad (7.44b)$$

$$k = 3$$

$$y_{i+1} = y_{i-3} + h [4f_{i+1} - 8\nabla f_{i+1} + \frac{20}{3} \nabla^2 f_{i+1} - \frac{8}{3} \nabla^3 f_{i+1} + \frac{14}{45} \nabla^4 f_{i+1} + 0 \nabla^5 f_{i+1} + \text{términos restantes}] \quad (7.44c)$$

$$k = 5$$

$$y_{i+1} = y_{i-5} + h [6f_{i+1} - 18\nabla f_{i+1} + 27 \nabla^2 f_{i+1} - 24 \nabla^3 f_{i+1} + \frac{123}{10} \nabla^4 f_{i+1} - \frac{33}{10} \nabla^5 f_{i+1} + \text{términos restantes}] \quad (7.44d)$$

Independientemente del valor que se elija para  $k$ , se debe seleccionar también el orden del corrector, el cual está dado en estas fórmulas por el orden  $r$  más uno de la diferencia hacia atrás de más alto orden que se utilice. Por ejemplo, para correctores de cuarto orden cabe emplear, entre otras, las combinaciones:

$$k = 0, \quad r = 3$$

$$y_{i+1} = y_i + h [f_{i+1} - \frac{1}{2} \nabla f_{i+1} - \frac{1}{12} \nabla^2 f_{i+1} - \frac{1}{24} \nabla^3 f_{i+1}] \quad (7.45a)$$

$$k = 1, \quad r = 3$$

$$y_{i+1} = y_{i-1} + h [2f_{i+1} - 2\nabla f_{i+1} + \frac{1}{3} \nabla^2 f_{i+1} + 0 \nabla^3 f_{i+1}] \quad (7.45b)$$

Por ejemplo, para el orden sexto se usa

$$k = 3, \quad r = 5$$

$$y_{i+1} = y_{i-3} + h [4f_{i+1} - 8\nabla f_{i+1} + \frac{20}{3} \nabla^2 f_{i+1} - \frac{8}{3} \nabla^3 f_{i+1} + \frac{14}{45} \nabla^4 f_{i+1}] \quad (7.45c)$$

Si se desarrollan las diferencias hacia atrás en estas fórmulas, se obtienen versiones de 7.45a, 7.45b y 7.45c más útiles para programar; es decir

$$k = 0, \quad r = 3$$

$$y_{i+1} = y_i + \frac{h}{24} [9f_{i+1} + 19f_i - 5f_{i-1} + f_{i-2}] \quad (7.46a)$$

$$k = 1, \quad r = 3$$

$$y_{i+1} = y_{i-1} + \frac{h}{3} [f_{i+1} + 4f_i + f_{i-1}] \quad (7.46b)$$

$$k = 3, \quad r = 5$$

$$y_{i+1} = y_{i-3} + \frac{2h}{45} [7f_{i+1} + 32f_i + 12f_{i-1} + 32f_{i-2} + 7f_{i-3}] \quad (7.46c)$$

Esta familia de correctores se conoce como **correctores de Adams-Moulton** y uno de los más usados es la ecuación 7.46a, la cual toma la forma

$$y_3 = y_2 + \frac{h}{24} [9f_3 + 19f_2 - 5f_1 + f_0]$$

para su primera aplicación o, regresando a la notación original

$$y_3 = y_2 + \frac{h}{24} [9f(x_3, \bar{y}_3) + 19f(x_2, y_2) - 5f(x_1, y_1) + f(x_0, y_0)] \quad (7.47)$$

donde  $y_1, f(x_1, y_1); y_2, f(x_2, y_2)$  deben calcularse previamente por un inicializador y  $\bar{y}_3$  por un predictor. No podría emplearse este corrector para calcular, por ejemplo,  $y_2$ , ya que tomaría la forma

$$y_2 = y_1 + \frac{h}{24} [9f(x_2, \bar{y}_2) + 19f(x_1, y_1) - 5f(x_0, y_0) + f(x_{-1}, y_{-1})]$$

que requiere información en la abscisa  $x_{-1}$  que está fuera del intervalo de interés.

### Ejemplo 7.6

Resuelva el PVI del ejemplo 7.1 con el corrector de la ecuación 7.46a.

#### Solución

El intervalo de interés  $[0, 1]$  se vuelve a dividir en cinco subintervalos y se usa el método de Runge Kutta de cuarto orden tanto de inicializador como de predictor. Es conveniente utilizar un inicializador y un predictor del mismo orden que el corrector.

#### Primera iteración

Inicialización con RK-4 (se toman los valores del ejemplo 7.4)

$$y(0.2) = 1.656200000 = y_1$$

$$y(0.4) = 1.410972813 = y_2$$

Predicción con RK-4 (se toma el valor del ejemplo 7.4)

$$y(0.6) = 1.246450474 = \bar{y}_3$$

Corrección con la ecuación 7.47

$$y_3 = 1.410972813 + \frac{02}{24} [9(0.6 - 1.246450474) +$$

$$19(0.4 - 1.410972813) - 5(0.2 - 1.6562) + (0 - 2)] = 1.246426665$$

#### Segunda iteración

Predicción con RK -4

Cálculo de las constantes  $k_1, k_2, k_3$  y  $k_4$

$$\begin{aligned}k_1 &= f(x_3, y_3) = (0.6 - 1.246426665) = -0.646426665 \\k_2 &= f(x_3 + h/2, y_3 + hk_1/2) = [(0.6 + 0.2/2) - (1.246426665 \\&\quad + 0.2 (-0.646426665)/2)] = -0.481783999 \\k_3 &= f(x_3 + h/2, y_3 + hk_2/2) = [(0.6 + 0.2/2) - (1.246426665 \\&\quad + 0.2 (-0.481783999)/2)] = -0.498248265 \\k_4 &= f(x_3 + h, y_3 + hk_3) = [(0.6 + 0.2) - (1.246426665 \\&\quad + 0.2 (-0.498248265))] = -0.346777012\end{aligned}$$

Cálculo de  $\bar{y}_4$

$$\begin{aligned}\bar{y}_4 &= y_3 + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \\&= 1.246426665 + \frac{0.2}{6} (-0.646426665 + 2 (-0.481783999) \\&\quad + 2 (-0.498248265) - 0.346777012) = 1.147984392\end{aligned}$$

Corrección con la ecuación 7.46a

$$\begin{aligned}y_4 &= y_3 + \frac{h}{24} [9f(x_4, \bar{y}_4) + 19f(x_3, y_3) - 5f(x_2, y_2) + f(x_1, y_1)] \\&= 1.246426665 + \frac{0.2}{24} [9(0.8 - 1.147984392) + 19(0.6 - 1.246426665)] \\&\quad - 5(0.4 - 1.410972813) + (0.2 - 1.6562)] = 1.147965814\end{aligned}$$

### Tercera iteración

Predicción con RK-4  $\bar{y}_5 = 1.103624544$

Corrección con (7.46a)  $y_5 = 1.103609057$

con un error absoluto de 0.0000292 y un error porcentual de 0.00265.

Los cálculos pueden hacerse con la TI-92 Plus



```
e7_6()
Prgm
Define f(x,y)= x-y
Define rk4(h,i) = Prgm
f(x[i],y[i])→k1
f(x[i]+h/2,y[i]+h/2*k1)→k2
f(x[i]+h/2,y[i]+h/2*k2)→k3
f(x[i]+h,y[i]+h*k3)→k4
y[i]+h/6*(k1+2*k2+2*k3+k4)→y[i+1]
x[i]+h→x[i+1]
EndPrgm
0→x[1]: 2→y[1] : 0.2→h: ClrIO
Disp "k x(k) y(k)"
Disp "0 "&format(x[1], "f1")&" "&format(y[1], "f9")
```



```

For i, 1, 2
rk4(h,i)
Disp format(i,"f0")&" "&format(x[i+1],"f1")&" "&format(y[i+1],"f9")
End For
For i, 3, 5
rk4(h,i)
y[i] +h/24*(9*f(x[i+1],y[i+1]) + 19*f(x[i],y[i])-5*f(x[i-1],y[i-1]) +
f(x[i-2],y[i-2]))->y[i+1]
disp format(i,"f0")&" "&format(x[i+1],"f1")&" "&format(y[i+1],"f9")
EndFor
EndPrgm

```

## MÉTODOS DE PREDICCIÓN

Ya antes se habló de una familia de predictores obtenida a partir del mismo principio de integración que se empleó para lo métodos de Adams-Moulton. A esta familia, que se deduce a continuación, se le llama **métodos de Adams-Bashforth**.

En general, para obtener un predictor de cualquier orden se utiliza la fórmula 7.40

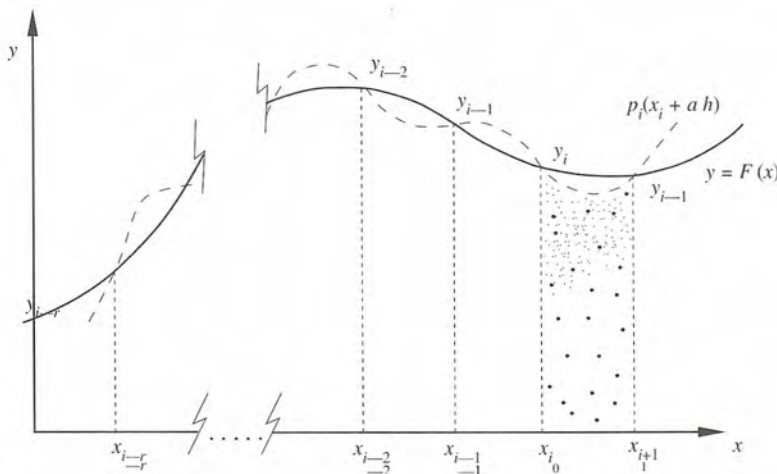
$$y_{i+1} = y_{i-k} + \int_{x_{i-k}}^{x_{i+1}} f(x, y) dx,$$

pero ahora la integración se realiza sustituyendo  $f(x, y)$  con un polinomio de grado  $k$  que pasa por  $(x_i, y_i), \dots, (x_{i-k}, y_{i-k})$ ; (véase Fig. 7.10). Obviamente, se utiliza el polinomio de interpolación en su forma de diferencias hacia atrás, pues  $x_i, \dots, x_{i-k}$  están regularmente espaciadas. Entonces, al aplicar la ecuación 5.38 se obtiene

$$y_{i+1} = y_{i-k} + \int_{x_{i-k}}^{x_{i+1}} p(x_i + sh) ds,$$

donde los límites de integración y  $dx$  en términos de la nueva variable  $s$  quedan como en la ecuación 7.42. Por tanto

$$\begin{aligned}
 y_{i+1} &= y_{i-k} + h \int_{-k}^1 p(x_i + sh) ds, \\
 y_{i+1} &= y_{i-k} + h \int_{-k}^1 [f_i + s \nabla f_i + s(s+1) \frac{\nabla^2 f_i}{2!} \\
 &\quad + s(s+1)(s+2) \frac{\nabla^3 f_i}{3!} + \dots + s(s+1)(s+2)\dots(s+k-1) \frac{\nabla^r f_i}{r!}] ds
 \end{aligned} \tag{7.48}$$



**Figura 7.10**  
Métodos de  
Adams-  
Bashforth.

Nótese que ahora el integrando es exactamente la ecuación 5.38, ya que en esta ocasión se está utilizando  $x_i$  como punto base. Al integrar la ecuación 7.48 se obtiene:

$$y_{i+1} = y_{i-k} + h \left[ s f_i + \frac{s^2}{2} \nabla f_i + s^2 \left( \frac{s}{3} + \frac{1}{2} \right) \frac{\nabla^2 f_i}{2!} + s^2 \left( \frac{s^2}{4} + s + 1 \right) \frac{\nabla^3 f_i}{3!} + s^2 \left( \frac{s^3}{5} + \frac{3s^2}{2} + \frac{11s}{3} + 3 \right) \frac{\nabla^4 f_i}{4!} + \text{términos faltantes} \right] \Bigg|_{-k}^1 \quad (7.49)$$

La ecuación 7.49 para  $k = 0, 1, 2$  y  $3$  toma las formas

$k = 0$

$$y_{i+1} = y_i + h \left[ f_i + \frac{1}{2} \nabla f_i + \frac{5}{12} \nabla^2 f_i + \frac{3}{8} \nabla^3 f_i + \frac{251}{720} \nabla^4 f_i + \text{términos faltantes} \right] \quad (7.50a)$$

$k = 1$

$$y_{i+1} = y_{i-1} + h \left[ 2f_i + 0 \nabla f_i + \frac{1}{3} \nabla^2 f_i + \frac{1}{3} \nabla^3 f_i + \frac{29}{90} \nabla^4 f_i + \text{términos faltantes} \right] \quad (7.50b)$$

$k = 2$

$$y_{i+1} = y_{i-2} + h \left[ 3f_i - \frac{3}{2} \nabla f_i + \frac{3}{4} \nabla^2 f_i + \frac{3}{8} \nabla^3 f_i + \frac{27}{80} \nabla^4 f_i + \text{términos faltantes} \right] \quad (7.50c)$$

$k = 3$

$$y_{i+1} = y_{i-3} + h \left[ 4f_i - 4 \nabla f_i + \frac{8}{3} \nabla^2 f_i + 0 \nabla^3 f_i + \frac{14}{45} \nabla^4 f_i + \text{términos faltantes} \right] \quad (7.50d)$$

La 7.50a significa la integración aproximada de una función que pasa por los puntos  $(x_{i-r}, y_{i-r})$ ,  $(x_{i-r+1}, y_{i-r+1})$ , ...,  $(x_i, y_i)$ , donde el subíndice  $r$  representa el grado del polinomio que se toma y  $r+1$  da el orden del predictor. El intervalo de integración es  $x_i, x_{i+1}$  (véase Fig. 7.10).

La ecuación 7.50b usa los mismos puntos que la ecuación 7.50a, pero con intervalo de integración  $[x_{i-1}, x_{i+1}]$ .

Las fórmulas más usadas de esta familia son:

$k = 0, r = 3$

$$y_{i+1} = y_i + h \left[ f_i + \frac{1}{2} \nabla f_i + \frac{5}{12} \nabla^2 f_i + \frac{3}{8} \nabla^3 f_i \right] \quad (7.51)$$

$k = 1, r = 1$

$$y_{i+1} = y_{i-1} + h \left[ 2f_i + 0 \nabla f_i \right] \quad (7.52)$$

$k = 3, r = 3$

$$y_{i+1} = y_{i-3} + h \left[ 4f_{i-4} \nabla f_i + \frac{8}{3} \nabla^2 f_i + 0 \nabla^3 f_i \right] \quad (7.53)$$

$$k = 5, r = 5$$

$$y_{i+1} = y_{i-5} + h [6f_i - 12 \nabla f_i + 15 \nabla^2 f_i - 9 \nabla^3 f_i + \frac{33}{10} \nabla^4 f_i + 0 \nabla^5 f_i] \quad (7.54)$$

cuya apariencia al desarrollar los operadores en diferencias hacia atrás resulta:

$$k = 0, r = 3$$

$$y_{i+1} = y_i + \frac{h}{24} [55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}] \quad (7.55)$$

$$k = 1, r = 1$$

$$y_{i+1} = y_{i-1} + 2hf_i \quad (7.56)$$

$$k = 3, r = 3$$

$$y_{i+1} = y_{i-3} + \frac{4h}{3} [2f_i - f_{i-1} + 2f_{i-2}] \quad (7.57)$$

$$k = 5, r = 5$$

$$y_{i+1} = y_{i-5} + \frac{3h}{10} [11f_i - 14f_{i-1} + 26f_{i-2} - 14f_{i-3} + 11f_{i-4}] \quad (7.58)$$

Es importante hacer notar que estas fórmulas son métodos para resolver el PVI (Ec. 7.11).

La ecuación 7.55 toma la forma

$$y_4 = y_3 + \frac{h}{24} [55f(x_3, y_3) - 59f(x_2, y_2) + 37f(x_1, y_1) - 9f(x_0, y_0)] \quad (7.59)$$

para su primera aplicación, y no sería posible determinar con ella un valor de  $y$  menor de  $y_4$  ( $y_3$ , por ejemplo). Por otro lado,  $y_1, f(x_1, y_1)$ ;  $y_2, f(x_2, y_2)$  y  $y_3, f(x_3, y_3)$  deberán determinarse con un inicializador.

Con estos métodos y la familia de los Adams-Moulton pueden integrarse esquemas iterativos conocidos como **métodos de predicción-corrección**, que en general funcionan como sigue:

1. Inicialización\* (se sugiere uno de la familia de Runge Kutta).
2. Predictor (para corresponder con el inicializador se sugiere usar un predictor del mismo orden).
3. Corrección (se emplea un corrector del mismo orden que el predictor y el inicializador).

### Ejemplo 7.7

Resuelva el PVI del ejemplo 7.1 usando como inicializador un RK-4, como predictor la ecuación 7.55 y como corrector la 7.46a

### Solución

El intervalo de interés  $[0, 1]$  se divide nuevamente en cinco subintervalos y se tiene

#### Primera iteración

Inicialización (tómense nuevamente los valores del ejemplo 7.4)

$$y_1 = 1.656200000$$

\* Recuérdese que este paso sólo se da en la primera iteración.



$$y_2 = 1.410972813$$

$$y_3 = 1.246450474$$

Predicción

$$\begin{aligned}\bar{y}_4 &= 1.246450474 + \frac{0.2}{24} [55(0.6 - 1.246450474) - 59(0.4 \\ &\quad - 1.410972813) + 37(0.2 - 1.6562) - 9(0 - 2)] = 1.148227306\end{aligned}$$

Corrección (con la ecuación 7.46a)

$$\begin{aligned}y_4 &= y_3 + \frac{h}{24} [9f(x_4, \bar{y}_4) + 19f(x_3, y_3) - 5f(x_2, y_2) + f(x_1, y_1)] \\ &= 1.246450474 + \frac{0.2}{24} [9(0.8 - 1.148227306) + 19(0.6 - 1.246450474) \\ &\quad - 5(0.4 - 1.410972813) + (0.2 - 1.6562)] = 1.147967635\end{aligned}$$

**Segunda iteración**

Predicción

$$\begin{aligned}\bar{y}_5 &= y_4 + \frac{h}{24} [55f(x_4, y_4) - 59f(x_3, y_3) + 37f(x_2, y_2) - 9f(x_1, y_1)] \\ &= 1.147967635 + \frac{0.2}{24} [55(0.8 - 1.147967635) - 59(0.6 - 1.246450474) \\ &\quad + 37(0.4 - 1.410972813) - 9(0.2 - 1.6562)] = 1.103819001\end{aligned}$$

Corrección (con la ecuación 7.46a)

$$\begin{aligned}y_5 &= y_4 + \frac{h}{24} [9f(x_5, \bar{y}_5) + 19f(x_4, y_4) - 5f(x_3, y_3) + f(x_2, y_2)] \\ &= 1.147967635 + \frac{0.2}{24} [9(1 - 1.103819001) + 19(0.8 - 1.147967635) \\ &\quad - 5(0.6 - 1.246450474) + (0.4 - 1.410972813)] = 1.103596997\end{aligned}$$

con un error absoluto de 0.00004 y un error porcentual de 0.0037

Nótese que aunque el corrector puede emplearse para mejorar  $y_3$  en su primera aplicación (véase Ej. 7.6), el predictor estima a  $y_4$  en su primera aplicación y a partir de ahí se comienza a corregir. Ésta es sólo una de las muchas formas en que se utilizan estos métodos de predicción-corrección.

#### ALGORITMO 7.4 Método predictor-corrector

(Inicialización con Runge-Kutta de cuarto orden, predicción con la ecuación 7.55 y corrección con la 7.46a).  
Para obtener la aproximación YF a la solución de un PVI, proporcionar la función F(X, Y) y los

DATOS: La condición inicial  $X_0, Y_0$ ; el valor XF donde se desea conocer el valor de YF y el número N de subintervalo por emplear.  
RESULTADOS: Aproximación YF: Y(4).

- PASO 1. Hacer  $H = (XF - X0)/N$   
 PASO 2. Hacer  $X(0) = X0$   
 PASO 3. Hacer  $Y(0) = Y0$   
 PASO 4. Hacer  $J = 1$   
 PASO 5. Mientras  $J \leq 3$ , repetir los pasos 6 a 9.  
     PASO 6. Realizar los pasos 4 a 9 del algoritmo 7.3.  
     PASO 7. Hacer  $X(J) = X0$   
     PASO 8. Hacer  $Y(J) = Y0$   
     PASO 9. Hacer  $J = J + 1$   
 PASO 10. Hacer  $I = 4$   
 PASO 11. Mientras  $I \leq N$ , repetir los pasos 12 a 20.  
     PASO 12. Hacer  $Y(4) = Y(3) + H/24 * (F(X(3), Y(3)) - 59 * F(X(2), Y(2)) + 37 * F(X(1), Y(1)) - 9 * F(X(0), Y(0)))$   
     PASO 13. Hacer  $X(4) = X(3) + H$   
     PASO 14. Hacer  $Y(4) = Y(3) + H/24 * (9 * F(X(4), Y(4)) + 19 * F(X(3), Y(3)) - 5 * F(X(2), Y(2)) + F(X(1), Y(1)))$   
     PASO 15. Hacer  $J = 0$   
     PASO 16. Mientras  $J \leq 3$ , repetir los pasos 17 a 19.  
     PASO 17. Hacer  $X(J) = X(J + 1)$   
     PASO 18. Hacer  $Y(J) = Y(J + 1)$   
     PASO 19. Hacer  $J = J + 1$   
     PASO 20. Hacer  $I = I + 1$   
 PASO 21. IMPRIMIR  $Y(4)$  y TERMINAR

## 7.7 Ecuaciones diferenciales ordinarias de orden superior y sistemas de ecuaciones diferenciales ordinarias

Cuando en el problema de valor inicial aparecen una ecuación diferencial de orden  $n$ ,  $n$  condiciones especificadas en un punto  $x_0$  y un punto  $x_f$  donde hay que encontrar el valor de  $y(x_f)$ , se tiene el problema de valor inicial general (PVIG)

$$\text{PVIG} \begin{cases} \frac{d^n y}{dx^n} = f(x, y, y', y'', \dots, y^{(n-1)}) \\ y(x_0) = y_0, y'(x_0) = y_0', \dots, y^{(n-1)}(x_0) = y_0^{(n-1)} \\ y(x_f) = ? \end{cases} \quad (7.60)$$

Para resolver la ecuación anterior no se desarrollan nuevos métodos, sino que se emplea una extensión de los estudiados en este capítulo. Para ello necesitaremos primero pasar la ecuación diferencial ordinaria o EDO de 7.60 a un **sistema de  $n$  ecuaciones diferenciales simultáneas de primer orden** cada una. Esto se logra de la siguiente manera:

Sea dada

$$\frac{d^n y}{dx^n} = f(x, y, y', y'', \dots, y^{(n-1)})$$

Se realiza el siguiente cambio de variables

$$\begin{aligned} y_1 &= y \\ y_2 &= y' \\ y_3 &= y'' \\ y_4 &= y''' \\ &\vdots \\ y_n &= y^{(n-1)} \end{aligned}$$

Se deriva miembro a miembro la primera y se sustituye en la segunda, con lo que se obtiene

$$y'_1 = y_2$$

Al derivar la segunda y sustituir en la tercera resulta

$$y'_2 = y_3$$

El procedimiento se repite hasta llegar al sistema de  $n$  ecuaciones de primer orden siguiente

$$y'_1 = y_2$$

$$y'_2 = y_3$$

$$y'_3 = y_4$$

$$\vdots$$

$$\vdots$$

$$y'_{n-1} = y_n$$

$$y'_n = \frac{d^n y}{dx^n} = f(x, y, y', y'', \dots, y^{(n-1)}) = f(x, y_1, y_2, y_3, \dots, y_n)$$

### Ejemplo 7.8

Pase la ecuación diferencial ordinaria

$$\frac{d^2 y}{dx^2} + \frac{dy}{dx} = x^2 + y^2$$

a un sistema de dos ecuaciones diferenciales ordinarias simultáneas de primer orden.

**Solución** Con el “despeje” de la derivada de segundo orden se tiene

$$\frac{d^2 y}{dx^2} = -y' + x^2 + y^2$$

El cambio de variables es:

$$y_1 = y; y_2 = y'$$

Al derivar la primera y sustituir en la segunda queda:

$$y'_1 = y_2$$

Se deriva la segunda

$$y'_2 = y''$$

y las nuevas variables se sustituyen en la ecuación diferencial, con lo cual resulta

$$\begin{aligned} y'_1 &= y_2 \\ y'_2 &= -y_2 + x^2 + y_1^2 \end{aligned}$$

el sistema pedido.

### Ejemplo 7.9

Una de las ecuaciones diferenciales ordinarias más empleadas en la matemática física es la ecuación de Bessel

$$x^2 y'' + x y' + (x^2 - n^2) y = 0$$

donde  $n$  puede tener cualquier valor, pero generalmente toma un valor entero.



Escriba esta ecuación como un sistema de ecuaciones diferenciales ordinarias de primer orden.

**Solución** La ecuación se pone en la forma normal

$$y'' = -\frac{1}{x} y' + \left(\frac{n^2}{x^2} - 1\right) y$$

Algunas veces es más conveniente para los cálculos computacionales emplear

$$\begin{aligned} y &= y \\ y' &= z \end{aligned}$$

como nuevas variables. Se deriva la segunda y se tiene

$$y'' = z'$$

El sistema queda:

$$\begin{aligned} y' &= z \\ z' &= -\frac{1}{x} z + \left(\frac{n^2}{x^2} - 1\right) y \end{aligned}$$

sistema que sólo podrá resolverse para valores de  $x$  distintos de cero.

En general, una ecuación diferencial ordinaria de  $n$ -ésimo orden queda convertida en un sistema de  $n$  ecuaciones diferenciales ordinarias simultáneas de la forma general

$$\begin{aligned} y'_1 &= f_1(x, y_1, y_2, \dots, y_n) \\ y'_2 &= f_2(x, y_1, y_2, \dots, y_n) \\ &\vdots \\ y'_n &= f_n(x, y_1, y_2, \dots, y_n) \end{aligned}$$

que puede resolverse aplicando, por ejemplo, alguno de los métodos de Runge-Kutta a cada ecuación, e iterando cada ecuación en turno, tal como en los sistemas de ecuaciones no lineales del capítulo 4, a los métodos de predicción-corrección.

Si se aplica, por ejemplo, el método de Runge-Kutta de cuarto orden a dos ecuaciones simultáneas de la forma

$$\begin{aligned} y' &= f_1(x, y, z) \\ z' &= f_2(x, y, z) \end{aligned}$$

donde sólo se emplea  $z$  como nueva variable a fin de no usar subíndices dobles en las ecuaciones

$$\begin{aligned} y_{i+1} &= y_i + \frac{h}{6} (k_1 + 2k_2 + 2k_3 + k_4) \\ z_{i+1} &= z_i + \frac{h}{6} (c_1 + 2c_2 + 2c_3 + c_4) \end{aligned} \tag{7.61a}$$

las cuales se calculan alternadamente, y las  $k$  y  $c$  se obtienen de

$$\begin{aligned} k_1 &= f_1(x_i, y_i, z_i) \\ c_1 &= f_2(x_i, y_i, z_i) \\ k_2 &= f_1(x_i + h/2, y_i + hk_1/2, z_i + hc_1/2) \\ c_2 &= f_2(x_i + h/2, y_i + hk_1/2, z_i + hc_1/2) \end{aligned} \tag{7.71b}$$

$$k_3 = f_1(x_i + h/2, y_i + hk_2/2, z_i + hc_2/2)$$

$$c_3 = f_2(x_i + h/2, y_i + hk_2/2, z_i + hc_2/2)$$

$$k_4 = f_1(x_i + h, y_i + hk_3, z_i + hc_3)$$

$$c_4 = f_2(x_i + h, y_i + hk_3, z_i + hc_3)$$

calculadas en ese orden.

### Ejemplo 7.10

Resuelva el siguiente problema de valor inicial por el método de Runge-Kutta de cuarto orden. (Puede usar el CD, el GC o un pizarrón electrónico).

$$\text{PVI} \begin{cases} y'' = -\frac{1}{x} y' + \left(\frac{1}{x^2} - 1\right) y \\ y(1) = 1 \\ y'(1) = 2 \\ y(3) = ? \end{cases}$$

Nótese que la EDO es la ecuación de Bessel con  $n = 1$  (véase Ejem. 7.9).

Al escribir la EDO como un sistema, el PVI queda:

$$\text{PVI} \begin{cases} y' = z \\ z' = -\frac{1}{x} z + \left(\frac{1}{x^2} - 1\right) y \\ y(1) = 1 \\ z(1) = 2 \\ y(3) = ? \end{cases}$$

### Solución



Al dividir el intervalo de interés  $[1, 3]$  en ocho subintervalos, el tamaño del paso de integración  $h$  es igual a 0.25.

**Primera iteración** (usando la ecuación 7.61a)

Cálculo de las constantes  $k$  y  $c$  con 7.61b

$$k_1 = f_1(x_0, y_0, z_0) = z_0 = z(1) = 2$$

$$c_1 = f_2(x_0, y_0, z_0) = -\frac{1}{x_0} z_0 + \left(\frac{1}{x_0^2} - 1\right) y_0$$

$$= -\frac{1}{1} (2) + \left(\frac{1}{1^2} - 1\right)(1) = -2$$

$$k_2 = f_1(x_0 + h/2, y_0 + hk_1/2, z_0 + hc_1/2)$$

$$= z_0 + hc_1/2 = 2 + 0.25(-2)/2 = 1.75$$

$$c_2 = f_2(x_0 + h/2, y_0 + hk_1/2, z_0 + hc_1/2)$$

$$= -\frac{1}{x_0 + h/2} (z_0 + hc_1/2) + \left[\frac{1}{(x_0 + h/2)^2} - 1\right] (y_0 + hk_1/2)$$

$$= -\frac{1}{1 + 0.25/2} (2 + 0.25(-2)/2) + \left[\frac{1}{(1 + 0.25/2)^2} - 1\right] (1 + 0.25(2)/2)$$

$$= -1.817901235$$

$$\begin{aligned}
k_3 &= f_1(x_0 + h/2, y_0 + hk_2/2, z_0 + hc_2/2) = z_0 + hc_2/2 \\
&= 2 + 0.25(-1.817901235)/2 = 1.772762346 \\
c_3 &= f_2(x_0 + h/2, y_0 + hk_2/2, z_0 + hc_2/2) \\
&= -\frac{1}{x_0 + h/2}(z_0 + hc_2/2) + \left[\frac{1}{(x_0 + h/2)^2} - 1\right](y_0 + hk_2/2) \\
&= -\frac{1}{1 + 0.25/2}(2 + 0.25(-1.817901235)/2) + \\
&\quad \left[\frac{1}{(1 + 0.25/2)^2} - 1\right](1 + 0.25(1.75)/2) = -1.831575789 \\
k_4 &= f_1(x_0 + h, y_0 + hk_3, z_0 + hc_3) = z_0 + hc_3 \\
&= 2 + 0.25(-1.831575789) = 1.542106053 \\
c_4 &= f_2(x_0 + h, y_0 + hk_3, z_0 + hc_3) \\
&= -\frac{1}{x_0 + h}(z_0 + hc_3) + \left[\frac{1}{(x_0 + h)^2} - 1\right](y_0 + hk_3) \\
&= -\frac{1}{1 + 0.25}(2 + 0.25(-1.831575789)) \\
&\quad + \left[\frac{1}{(1 + 0.25)^2} - 1\right](1 + 0.25(1.772762346)) = -1.753233454
\end{aligned}$$

Cálculo de  $y_1 = y(1.25)$  y  $z_1 = z(1.25)$  con la ecuación 7.61a

$$\begin{aligned}
y_1 &= y_0 + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\
&= 1 + \frac{0.25}{6}[2 + 2(1.75) + 2(1.772762346) + 1.542106053] \\
&= 1.441151281 \\
z_1 &= z_0 + \frac{h}{6}(c_1 + 2c_2 + 2c_3 + c_4) \\
&= 2 + \frac{0.25}{6}[-2 + 2(-1.817901235) + 2(-1.831575789) \\
&\quad - 1.753233454] = 1.539492187
\end{aligned}$$

### Segunda iteración

Cálculo de las  $k$  y  $c$  con la ecuación 7.61b

$$\begin{aligned}
k_1 &= f_1(x_1, y_1, z_1) = z_1 = 1.539492187 \\
c_1 &= f_2(x_1, y_1, z_1) = -\frac{1}{x_1^2}z_1 + \left(\frac{1}{x_1^2} - 1\right)y_1 \\
&= -\frac{1}{1.25}(1.539492187) + \left(\frac{1}{(1.25)^2} - 1\right)(1.441151281) \\
&= -1.750408211 \\
k_2 &= f_1(x_1 + h/2, y_1 + hk_1/2, z_1 + hc_1/2) = z_1 + hc_1/2 \\
&= 1.539492187 + 0.25(-1.750408211)/2 = 1.320691161
\end{aligned}$$



$$\begin{aligned}
c_2 &= f_2(x_1 + h/2, y_1 + hk_1/2, z_1 + hc_1/2) \\
&= -\frac{1}{x_1 + h/2}(z_1 + hc_1/2) + \left[\frac{1}{(x_1 + h/2)^2} - 1\right](y_1 + hk_1/2) \\
&= -\frac{1}{1.25 + 0.25/2}(1.539492187 + 0.25(-1.750408211)/2) \\
&\quad + \left[\frac{1}{(1.25 + 0.25/2)^2} - 1\right](1.441151281 + 0.25(1.539492187)/2) \\
&= -1.730044025 \\
k_3 &= f_1(x_1 + h/2, y_1 + hk_2/2, z_1 + hc_2/2) = z_1 + hc_2/2 \\
&= 1.539492187 + 0.25(-1.730044025)/2 = 1.323236684 \\
c_3 &= f_2(x_1 + h/2, y_1 + hk_2/2, z_1 + hc_2/2) \\
&= -\frac{1}{x_1 + h/2}(z_1 + hc_2/2) + \left[\frac{1}{(x_1 + h/2)^2} - 1\right](y_1 + hk_2/2) \\
&= -\frac{1}{1.25 + 0.25/2}(1.539492187 + 0.25(-1.730044025)/2) \\
&\quad + \left[\frac{1}{(1.25 + 0.25/2)^2} - 1\right](1.441151281 + 0.25(1.320691161)/2) \\
&= -1.71901137 \\
k_4 &= f_1(x_1 + h, y_1 + hk_3, z_1 + hc_3) = z_1 + hc_3 \\
&= 1.539492187 + 0.25(-1.71901137) = 1.109739345 \\
c_4 &= f_2(x_1 + h, y_1 + hk_3, z_1 + hc_3) \\
&= -\frac{1}{x_1 + h}(z_1 + hc_3) + \left[\frac{1}{(x_1 + h)^2} - 1\right](y_1 + hk_3) \\
&= -\frac{1}{1.25 + 0.25}(1.539492187 + 0.25(-1.71901137)) \\
&\quad + \left[\frac{1}{(1.25 + 0.25)^2} - 1\right](1.441151281 + 0.25(1.323236684)) \\
&= -1.724248703
\end{aligned}$$

Cálculo de  $y_2 = y(1.5)$  y  $z_2 = z(1.5)$  con la ecuación 7.61a

$$\begin{aligned}
y_2 &= y_1 + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\
&= 1.441151281 + \frac{0.25}{6}[1.539492187 + 2(1.320691161) + \\
&\quad 2(1.323236684) + 1.109739345] = 1.771863249 \\
z_2 &= z_1 + \frac{h}{6}(c_1 + 2c_2 + 2c_3 + c_4) \\
&= 1.539492187 + \frac{0.25}{6}[-1.750408211 + 2(-1.730044025) \\
&\quad + 2(-1.71901137) - 1.724248703] = 1.107293533
\end{aligned}$$

Se continúa calculando en la misma forma y se obtiene:

|                         |                          |
|-------------------------|--------------------------|
| $y(1.75) = 1.994766280$ | $z(1.75) = 0.675599895$  |
| $y(2.00) = 2.109754328$ | $z(2.00) = 0.245291635$  |
| $y(2.25) = 2.118486566$ | $z(2.25) = -0.172076357$ |
| $y(2.50) = 2.026089844$ | $z(2.50) = -0.561053191$ |
| $y(2.75) = 1.841680320$ | $z(2.75) = -0.905578495$ |
| $y(3.00) = 1.578253875$ | $z(3.00) = -1.190934201$ |

El valor buscado es  $y(3) = 1.578253875$

Este ejemplo se puede resolver también con Matlab (ver Ejem. 7.4).

```
function f=E7_10(x,y)
f1=y(2);
f2=-1/x.*y(2)+(1/x.^2-1).*y(1);
f=[f1;f2];
```

```
xx=1 : 0.25 : 3;
[T,Y]=ode45('e7_10', xx, [1;2]);
Y
```

Si el problema es de inicio un sistema de EDO's de primer orden con sus correspondientes condiciones iniciales, el procedimiento es el mismo visto hasta ahora pero ahorrándose el paso de convertir la EDO de orden  $n$  a un sistema de  $n$  ecuaciones diferenciales de primer orden (consúltese los ejercicios).

A continuación se presenta un algoritmo para el método de Runge-Kutta de cuarto orden con objeto de resolver un sistema de dos ecuaciones diferenciales ordinarias.

### ALGORITMO 7.5 Método de Runge-Kutta de cuarto orden para un sistema de dos ecuaciones diferenciales ordinarias

Para aproximar la solución al PVI

$$\begin{aligned} y' &= f_1(x, y, z) \\ z' &= f_2(x, y, z) \\ y(x_0) &= y_0; y(x_f) = ? \\ z(x_0) &= z_0; z(x_f) = ?, \end{aligned}$$

proporcionar las funciones  $F1(X, Y, Z)$  y  $F2(X, Y, Z)$  y los

DATOS: La condición inicial  $X_0, Y_0, Z_0$ , el valor  $X_F$  y el número de  $N$  de subintervalos por emplear,

RESULTADOS: La aproximación a los valores  $Y(X_F)$  y  $Z(X_F)$ :  $Y_0$  y  $Z_0$ .

PASO 1. Hacer  $H = (X_F - X_0) / N$

PASO 2. Hacer  $I = 1$

PASO 3. Mientras  $I \leq N$ , repetir los pasos 4 al 15.

PASO 4. Hacer  $K1 = F1(X_0, Y_0, Z_0)$

PASO 5. Hacer  $C1 = F2(X0, Y0, Z0)$   
 PASO 6. Hacer  $K2 = F1(X0 + H/2, Y0 + H/2 * K1, Z0 + H/2 * C1)$   
 PASO 7. Hacer  $C2 = F2(X0 + H/2, Y0 + H/2 * K1, Z0 + H/2 * C1)$   
 PASO 8. Hacer  $K3 = F1(X0 + H/2, Y0 + H/2 * K2, Z0 + H/2 * C2)$   
 PASO 9. Hacer  $C3 = F2(X0 + H/2, Y0 + H/2 * K2, Z0 + H/2 * C2)$   
 PASO 10. Hacer  $K4 = F1(X0 + H, Y0 + H * K3, Z0 + H * C3)$   
 PASO 11. Hacer  $C4 = F2(X0 + H, Y0 + H * K3, Z0 + H * C3)$   
 PASO 12. Hacer  $Y0 = Y0 + H/6 * (K1 + 2 * K2 + 2 * K3 + K4)$   
 PASO 13. Hacer  $Z0 = Z0 + H/6 * (C1 + 2 * C2 + 2 * C3 + C4)$   
 PASO 14. Hacer  $X0 = X0 + H$   
 PASO 15. Hacer  $I = I + 1$

PASO 16. IMPRIMIR  $Y0, Z0$  y TERMINAR.

## 7.8 Formulación del problema de valores en la frontera

Un problema de valores en la frontera (PVF), para ecuaciones diferenciales ordinarias, puede estar dado, por ejemplo, por una EDO de segundo orden y dos condiciones de frontera: CF1 y CF2

$$\text{PVF} \begin{cases} \text{EDO} & \frac{d^2y}{dx^2} = f(x, y, y') \\ \text{CF1} & y(x_0) = y_0 \\ \text{CF2} & y(x_f) = y_f \\ & y(x) = ? \text{ para } x_0 < x < x_f \end{cases} \quad (7.62)$$

Obsérvese que ahora se da como información dos puntos distintos por donde pasa la curva desconocida  $y$ , solución de la EDO; es decir, conocemos el valor de  $y$  correspondiente a dos abscisas distintas\*:  $x_0$  y  $x_f$  y queremos conocer el valor de  $y$  en el intervalo  $(x_0, x_f)$ . Esto se ilustra gráficamente en la figura 7.11.

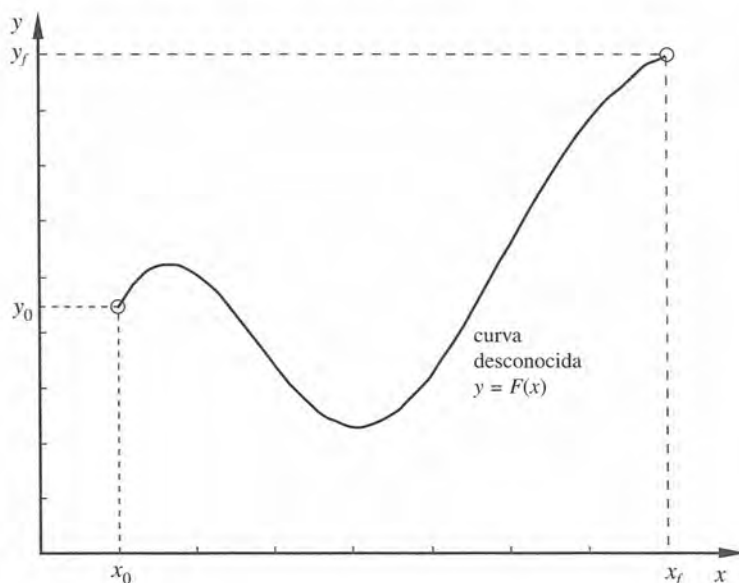


Figura 7.11

\* A diferencia del PVI donde la información está dada en un solo punto  $x_0$ .



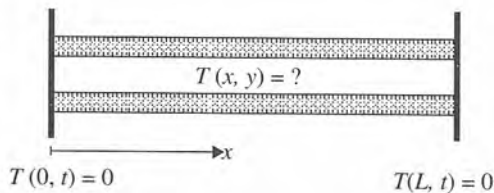
Desde luego también contamos para encontrar a  $y$  con su segunda derivada, esto es  $f(x, y, y')$ .

Este tipo de problemas surgen, por ejemplo, cuando se resuelven ecuaciones diferenciales parciales analíticamente. Así, si se tiene el problema

$$\begin{cases} \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \\ T(0, t) = 0 \\ T(L, t) = 0 \\ T(x, 0) = f(x) \\ T(x, t) = ? \end{cases} \quad \text{para } 0 < x < L \text{ y } t > 0 \quad (7.63)$$

que describe la conducción de calor en una barra aislada longitudinalmente\* (véase Fig. 7.12);  $T(0, t)$  y  $T(L, t)$  representan la temperatura  $T$  de la barra en los extremos izquierdo y derecho, respectivamente, sostenidos constantes e iguales a cero (en general son funciones del tiempo  $t$ ).

**Figura 7.12**  
Barra aislada longitudinalmente con extremos sujetos a temperaturas establecidas.



La aplicación del método de separación de variables a la ecuación 7.63 transforma el problema en un PVI y en el PVF siguiente:

$$\begin{cases} \frac{\partial^2 \phi}{\partial x^2} = -\lambda \phi \\ \phi(0) = 0 \\ \phi(L) = 0 \\ \phi(x) = ? \end{cases} \quad \text{para } 0 < x < L \quad (7.64)$$

cuya solución conjuntamente con la del PVI mencionado permitirán resolver 7.63.

A continuación se da un método para resolver problemas del tipo 7.62 conocido como método del “disparo”, por analogía con el tiro o disparo contra un blanco fijo.

### MÉTODO DEL DISPARO

Consideremos el problema siguiente

$$\begin{cases} y'(x) = y \\ y(0) = 0 \\ y(1) = 2 \\ y(x) = ? \end{cases} \quad \text{para } 0 < x < 1 \quad (7.65)$$

Para resolverlo podemos usar uno de los métodos de valor inicial discutidos en las secciones anteriores, para lo cual tendríamos que proponer, de consideraciones físicas o de otro

\* Ver capítulo 8.

tipo, una condición inicial, por ejemplo  $y'(0) = \alpha_0$ . Siguiendo la metáfora del disparo, esto representaría una medida del ángulo que forma el cañón con el piso. Contando con esta condición inicial, se puede formar a partir de 7.65 el siguiente PVI

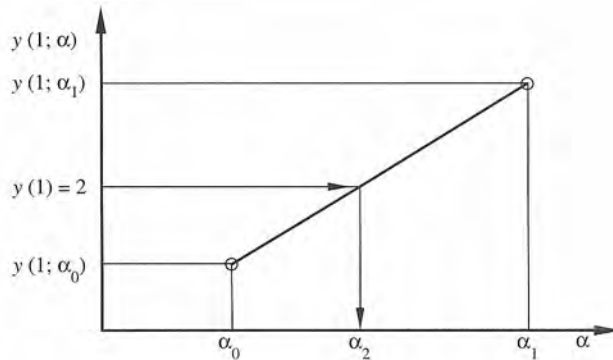
$$\begin{cases} y''(x) = y \\ y(0) = 0 \\ y'(0) = \alpha_0 \\ y(1) = ? \end{cases} \quad \text{que convertido a sistema queda:} \quad \begin{cases} y' = z \\ z' = y \\ y(0) = 0 \\ z(0) = \alpha_0 \\ y(1) = ? \end{cases}$$

Al resolver este PVI se obtiene un valor de  $y(1)$  correspondiente a  $\alpha_0$  o más fácilmente  $y(1; \alpha_0)$  que podremos comparar con el valor  $y(1) = 2$  dado en el problema original y así estimar la bondad de la  $\alpha_0$  propuesta. Con esta información podremos proponer una “mejor”  $\alpha$  (un nuevo ángulo de disparo),  $\alpha_1$ , con lo que se obtendría un nuevo PVI

$$\begin{cases} y' = z \\ z' = y \\ y(0) = 0 \\ z(0) = \alpha_1 \\ y(1) = ? \end{cases}$$

Al resolver obtenemos  $y(1; \alpha_1)$ .

En estas condiciones podemos plantear una nueva aproximación de  $y'(0)$ , pero considerando a  $y(1; \alpha)$  como una función de  $\alpha$  y de la cual se tienen ya dos puntos  $(\alpha_0, y(1; \alpha_0))$  y  $(\alpha_1, y(1; \alpha_1))$ , como se ve en la figura 7.13.



**Figura 7.13**

Si unimos  $(\alpha_0, y(1; \alpha_0))$  y  $(\alpha_1, y(1; \alpha_1))$  con una línea recta podremos con una interpolación (extrapolación) lineal inversa, obtener una nueva aproximación a  $\alpha$ ,  $\alpha_2$  dada algebraicamente por:

$$\alpha_2 = \alpha_1 - (\alpha_1 - \alpha_0) \frac{y(1; \alpha_1) - y(1; \alpha_2)}{y(1; \alpha_1) - y(1; \alpha_0)}$$

y con ella formular el PVI con  $y'(0) = \alpha_2$ .

El proceso puede continuarse usando las últimas dos alfas  $\alpha_{i-1}$  y  $\alpha_i$  para la interpolación (extrapolación) lineal inversa hasta que  $|y(1; \alpha_{i+1}) - y(1)| < \varepsilon$  o hasta que se hayan realizado un número máximo de iteraciones.

**Ejemplo 7.11**

Resolver el PVF 7.65 con una  $\varepsilon = 10^{-5}$  y MAXIT = 10 iteraciones, con el método del disparo.

**Solución**

$x_0 = 0$ ,  $x_f = 1$ ,  $y_0 = 0$ ,  $y'(0) = \alpha_0 = 1.5$  (valor inicial propuesto)

Al resolver el PVI

$$\begin{cases} y' = z \\ z' = y \\ y(0) = 0 \\ z(0) = \alpha_0 = 1.5 \\ y(1) = ? \end{cases}$$

con el método de Runge-Kutta de cuarto orden y un tamaño de paso  $h = 0.1$ , se obtiene  $y(1; \alpha_0) = 1.76279998$ . Se propone ahora un valor de  $\alpha_1 = 2.5$  y se resuelve nuevamente el PVI. Se obtiene  $y(1; \alpha_1) = 2.93799996$ . Con estos valores se interpola para obtener  $\alpha_2$ :

$$\begin{aligned} \alpha_2 &= \alpha_1 - (\alpha_1 - \alpha_0) \frac{y(1; \alpha_1) - y(2)}{y(1; \alpha_1) - y(1; \alpha_0)} \\ &= 2.5 - (2.5 - 1.5) \left( \frac{2.93799996 - 2}{2.93799996 - 1.76279998} \right) = 1.701838 \end{aligned}$$

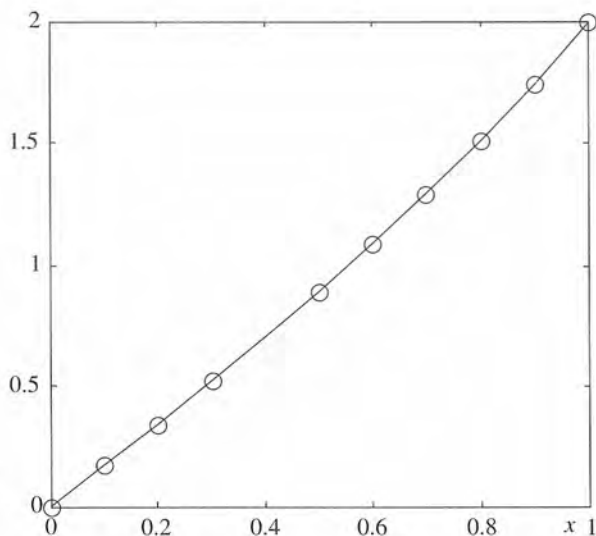
Se resuelve el PVI con  $\alpha_2 = 1.701838$  y se obtiene  $y(1; \alpha_2) = 1.99999999183644$ . El proceso se detiene puesto que  $|y(1; \alpha_1) - y(2)| < \varepsilon$ , tomándose entonces como valor “verdadero” de  $y'(0)$  a  $\alpha_2 = 1.701838$ . Los valores de  $y$  en el intervalo  $(0, 1)$  son los que generó el método de Runge-Kutta de cuarto orden en la última iteración:

| $x$ | $y$      |
|-----|----------|
| 0.0 | 0.00000  |
| 0.1 | 0.170467 |
| 0.2 | 0.342641 |
| 0.3 | 0.518244 |
| 0.4 | 0.699033 |
| 0.5 | 0.886819 |
| 0.6 | 1.083480 |
| 0.7 | 1.290985 |
| 0.8 | 1.511411 |
| 0.9 | 1.746963 |
| 1.0 | 1.999999 |

Al analizar la tabla se encuentra que, por ejemplo, las diferencias finitas son crecientes en sus diferentes órdenes, lo cual sugiere que la solución  $y = F(x)$  tiene un término exponencial.

Los cálculos pueden realizarse con Matlab adaptando el guión del ejercicio 7.12.





**Figura 7.14**  
Solución gráfica  
del PVF.

## Ejercicios

- 7.1** Un ataque cilíndrico de fondo plano con un diámetro de 1.5 m (Fig. 7.15), contiene un líquido de densidad  $\rho = 1.5 \text{ kg/L}$  a una altura  $a$  de 3. m. Se desea saber la altura del líquido dentro del tanque tres minutos después de que se abre completamente la válvula de salida, la cual da un gasto de  $0.6A \sqrt{2ga} \text{ m}^3/\text{s}$ , donde  $A$  es el área seccional del tubo de salida y es  $78.5 \times 10^{-4} \text{ m}^2$  y  $g = 9.81 \text{ m/s}^2$ .

### Solución



Como se vio en el ejemplo de la sección 7.1, el vaciado o llenado de un tanque cilíndrico se modela haciendo un balance de materia con la siguiente expresión

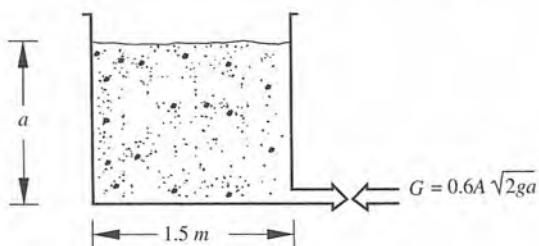
$$\begin{aligned} \text{Acumulación} &= \text{Entrada} && - \text{Salida} \\ \frac{dV\rho}{dt} &= 0 && - 0.6A \sqrt{2ga} \end{aligned}$$

donde:

$$V = \frac{\pi}{4} (1.5)^2 a$$

entonces:

$$\frac{\pi}{4} (1.5)^2 \frac{da}{dt} = -0.6A \sqrt{2ga}$$



**Figura 7.15**  
Vaciado de un  
tanque  
cilíndrico.

de donde:

$$\frac{da}{dt} = -\frac{2.4A\sqrt{2g}a}{\pi(1.5)^2} = -0.0026653\sqrt{2g}a$$

Al considerar como tiempo cero el momento de abrir la válvula y además la altura buscada a un tiempo de 180 s, se llega a

$$\text{PVI} \begin{cases} \frac{da}{dt} = -0.0026653\sqrt{2g}a \\ a(0) = 3 \text{ m} \\ a(180) = ? \end{cases}$$

En virtud de que la exactitud de los resultados que se esperan no es grande, se usa el método de Euler para resolver este PVI.

Los resultados que se obtiene con  $h = 30$  s son

| tiempo (s) | 0    | 30   | 60   | 90   | 120  | 150  | 180  |
|------------|------|------|------|------|------|------|------|
| a (m)      | 3.00 | 2.39 | 1.84 | 1.36 | 0.95 | 0.60 | 0.33 |

- 7.2 Calcule el tiempo necesario para que el nivel del líquido dentro del tanque esférico con radio  $r = 5$  m mostrado en la figura 7.16 pase de 4 m a 3 m. La velocidad de salida por el orificio del fondo es  $v = 4.895\sqrt{a}$  m/s, el diámetro de dicho orificio es de 10 cm.

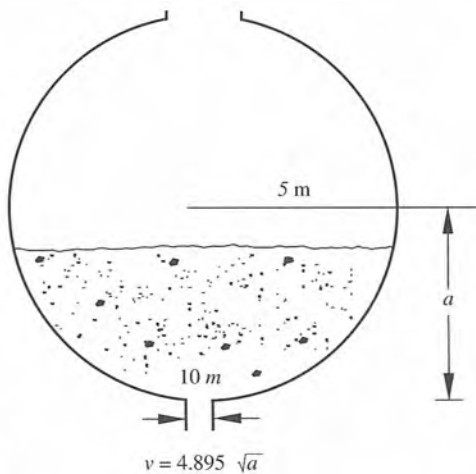
**Solución** Balance de materia en el tanque

$$\text{Acumulación} = \text{Entrada} - \text{Salida}$$

$$\rho \frac{dV}{dt} = 0 - A v \rho$$

donde  $V$  es el volumen del líquido en el tanque, que en función de la altura está dado por

$$V = \pi \left( 5a^2 - \frac{a^3}{3} \right) \text{ m}^3$$



**Figura 7.16**  
Vaciado de un  
tanque esférico.

A es el área del orificio de salida

$$A = \frac{\pi}{4} (0.1)^2 \text{ m}^2$$

y

$$v = 4.895 \sqrt{a} \text{ m/s}$$

Estas cantidades se sustituyen en la primera ecuación y se tiene

$$\pi \frac{d \left( 5 a^2 - \frac{a^3}{3} \right)}{dt} = - \frac{\pi}{4} (0.1)^2 4.895 \sqrt{a}$$

Se deriva

$$10 a \frac{da}{dt} - \frac{3a^2}{3} \frac{da}{dt} = - \frac{(0.1)^2}{4} 4.895 \sqrt{a}$$

y al despejar se tiene

$$\frac{da}{dt} = \frac{-4.895 (0.1)^2 \sqrt{a}}{4 (10 a - a^2)}$$

que con la condición inicial y la pregunta forman el siguiente

$$\text{PVI} \begin{cases} \frac{da}{dt} = - \frac{0.012375 \sqrt{a}}{(10 a - a^2)} \\ a(0) = 4 \text{ m} \\ a(?) = 3 \text{ m} \end{cases}$$

Con el método de Euler modificado y un paso de integración h de 10 segundo, se tiene

| tiempo<br>(s) | altura a<br>(m) |
|---------------|-----------------|
| 0             | 4.0000          |
| 10            | 3.8982          |
| 20            | 3.7968          |
| 30            | 3.6957          |
| 40            | 3.5948          |
| 50            | 3.4941          |
| 60            | 3.3935          |
| 70            | 3.2939          |
| 80            | 3.1924          |
| 90            | 3.0917          |
| 100           | 2.9908          |

Este último valor de altura puede considerarse como 3 m, por lo que el tiempo necesario para que el nivel del líquido dentro del tanque esférico pase de 4 a 3 m es aproximadamente 100 segundos.



- 7.3 En un tanque perfectamente agitado se tiene 400 L de una salmuera en la cual están disueltos 25 kg de sal común (NaCl), en cierto momento se hace llegar al tanque un gasto de 80 L/min de una salmuera que contiene 0.5 kg de sal común por litro. Si se tiene un gasto de salida de 80 L/min determine.

- a) ¿Qué cantidad de sal hay en el tanque transcurridos 10 minutos?  
b) ¿Qué cantidad de sal hay en el tanque transcurrido un tiempo muy grande?

**Solución**

- a) Si se llaman  $x$  los kg de sal en el tanque después de  $t$  minutos, la acumulación de sal en el tanque está dada por  $dx/dt$  y por la expresión

$$\frac{dx}{dt} = \text{masa de sal que entra} - \text{masa de sal que sale}$$

los valores conocidos se sustituyen y se llega a la ecuación:

$$\frac{dx}{dt} = 80 (0.5) - 80 \left( \frac{x}{400} \right)$$

o

$$\frac{dx}{dt} = 40 - 0.2x$$

que con la condición inicial de que hay 25 kg de sal al tiempo cero, da el siguiente

$$\text{PVI} \begin{cases} \frac{dx}{dt} = 40 - 0.2x \\ x(0) = 25 \\ x(10) = ? \end{cases}$$

Como vía de ilustración se utilizará un método de Runge-Kutta de tercer orden cuyo algoritmo está dado por:

$$y_{i+1} = y_i + \frac{h}{6} (k_1 + 4k_2 + k_3)$$

con

$$k_1 = f(x_i, y_i)$$

$$k_2 = f(x_i + h/2, y_i + hk_1/2)$$

$$k_3 = f(x_i + h, y_i + 2hk_2 - hk_1)$$

En el CD se encuentra el **PROGRAMA 7.1** para resolver este problema de valor inicial con el algoritmo anotado arriba. El resultado obtenido es

$$x(10) = 176.3 \text{ con un paso de integración } h \text{ de 1 min.}$$

- b) La solución se obtiene dejando correr el programa hasta que la cantidad de sal en el tanque no cambie con el tiempo; esto es, hasta que se alcance régimen permanente. Al dejar correr el programa se obtuvieron los siguientes resultados:

CONDICIÓN INICIAL:  $Y(.00) = 25.0000$

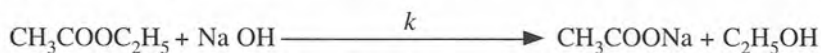
PASO DE INTEGRACIÓN  $H = 1.000$

VALOR FINAL DE  $X = 50.000$

SE IMPRIME CADA 2 ITERACIONES

| X       | Y        |
|---------|----------|
| 2.0000  | 82.7124  |
| 4.0000  | 121.3920 |
| 6.0000  | 147.3158 |
| 8.0000  | 164.6902 |
| 10.0000 | 176.3348 |
| 12.0000 | 184.1393 |
| 14.0000 | 189.3699 |
| 16.0000 | 192.8755 |
| 18.0000 | 195.2251 |
| 20.0000 | 196.7998 |
| 22.0000 | 197.8552 |
| 24.0000 | 198.5625 |
| 26.0000 | 199.0366 |
| 28.0000 | 199.3543 |
| 30.0000 | 199.5673 |
| 32.0000 | 199.7100 |
| 34.0000 | 199.8056 |
| 36.0000 | 199.8698 |
| 38.0000 | 199.9127 |
| 40.0000 | 199.9415 |
| 42.0000 | 199.9608 |
| 44.0000 | 199.9738 |
| 46.0000 | 199.9824 |
| 48.0000 | 199.9883 |
| 50.0000 | 199.9921 |

- 7.4 Se hacen reaccionar isotérmicamente 260 g de acetato de etilo ( $\text{CH}_3 \text{COOC}_2 \text{H}_5$ ) con 175 g de hidróxido de sodio ( $\text{NaOH}$ ) en solución acuosa (ajustando el volumen total a 5 litros) para dar acetato de sodio ( $\text{CH}_3\text{COONa}$ ) y alcohol etílico ( $\text{C}_2\text{H}_5\text{OH}$ ), de acuerdo con la siguiente ecuación estequiométrica



Si la constante de velocidad de reacción  $k$  está dada por:

$$k = 1.44 \times 10^{-2} \frac{\text{L}}{\text{mol min}}$$

determine la cantidad de acetato de sodio y alcohol etílico presentes 30 minutos después de iniciada la reacción.

**Solución**

Si  $x$  denota el número de moles por litro de acetato de etilo que han reaccionado al tiempo  $t$ , entonces la velocidad de reacción  $dx/dt$  viene dada por la ley de acción de masas así:

$$\frac{dx}{dt} = k C_A^1 C_B^1$$

donde  $C_A$  y  $C_B$  denotan las concentraciones molares de los reactantes acetato de etilo e hidróxido de sodio, respectivamente, al tiempo  $t$  y los exponentes son sus coeficientes estequiométricos en la reacción. Entonces:

$$C_A = \frac{260 \text{ g}}{\text{PM}_{\text{CH}_3\text{COOC}_2\text{H}_5} 5\text{L}} - x \frac{\text{mol}}{\text{L}}$$

$$C_B = \frac{175 \text{ g}}{\text{PM}_{\text{NaOH}} 5\text{L}} - x \frac{\text{mol}}{\text{L}}$$

Al substituir valores y añadir la condición inicial y la pregunta a la ecuación diferencial resultante, se tiene:

$$\text{PVI} \begin{cases} \frac{dx}{dt} = 1.44 \times 10^{-2} (0.59 - x)(0.875 - x) \\ x(0) = 0.0 \\ x(30) = ? \end{cases}$$

Al correr el programa 7.2 se obtiene:

$$x(30) = 0.169 \text{ con un paso de integración } h \text{ de } 1 \text{ min.}$$

de donde:

$$\text{Cantidad de acetato de sodio} = 0.169 \times 5 \times 62 = 52.39\text{g}$$

- 7.5** Se conecta un inductor (inductancia) de 0.4 henries en serie con una resistencia de 8 ohms, un capacitor de 0.015 farads y un generador de corriente alterna dada por la función  $30 \sin 5t$  volts para  $t \geq 0$  (véase Fig. 7.17).

- Establezca una ecuación diferencial para la carga instantánea en el capacitor.
- Encuentre la carga a distintos tiempos.

**Solución**

- La caída de voltaje en la resistencia es  $8I$ , en la inductancia es  $0.4 dI/dt$  y en la capacitancia  $Q/0.015 = 66.6666 Q$   
Según las leyes de Kirchhoff

$$8I + 0.4 \frac{dI}{dt} + 66.6666 Q = 30 \sin 5t$$

o

$$0.4 \frac{d^2Q}{dt^2} + 8 \frac{dQ}{dt} + 66.6666 Q = 30 \sin 5t$$

ya que

$$\frac{dQ}{dt} = I$$



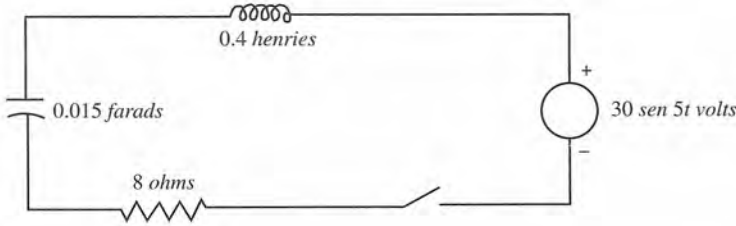
y finalmente

$$\frac{d^2Q}{dt^2} + 20 \frac{dQ}{dt} + 166.6666 Q = 75 \text{ sen } 5t$$

con las condiciones

$$Q = 0, I = \frac{dQ}{dt} = 0 \text{ a } t = 0$$

**Figura 7.17**



Al pasar a un sistema con el cambio de variable  $\frac{dQ}{dt} = z$

$$\text{PVI} \begin{cases} \frac{dQ}{dt} = z \\ \frac{dz}{dt} = 75 \text{ sen } 5t - 20z - 166.6666 Q \\ Q(0) = 0 \\ z(0) = 0 \end{cases}$$

b) Al resolver por el método de Runge-Kutta de cuarto orden y usando  $h = 0.1$  se tiene:

| t   | Q        | $\frac{dQ}{dt}$ | t   | Q        | $\frac{dQ}{dt}$ |
|-----|----------|-----------------|-----|----------|-----------------|
| 0.1 | 0.03093  | 0.96008         | 1.1 | -0.43060 | -0.43375        |
| 0.2 | 0.16949  | 1.67198         | 1.2 | -0.34174 | 1.40254         |
| 0.3 | 0.33066  | 1.33585         | 1.3 | -0.16921 | 2.02794         |
| 0.4 | 0.42549  | 0.38455         | 1.4 | -0.04475 | 2.15684         |
| 0.5 | 0.41473  | -0.71114        | 1.5 | 0.24775  | 1.75767         |
| 0.6 | 0.29996  | -1.62002        | 1.6 | 0.39010  | 0.92817         |
| 0.7 | 0.11080  | -2.11960        | 1.7 | 0.43693  | -0.12859        |
| 0.8 | -0.10561 | -2.09630        | 1.8 | 0.37679  | -1.15386        |
| 0.9 | -0.29609 | -1.55950        | 1.9 | 0.22440  | -1.89662        |
| 1.0 | -0.41404 | -0.64128        | 2.0 | 0.01706  | -2.17503        |

**7.6** Un proyectil de masa  $m = 0.11$  kg se lanza verticalmente hacia arriba con una velocidad inicial  $v_0 = 80$  m/s y se va frenando debido a la fuerza de gravedad  $F_g = -mg$  y a la resistencia del aire  $F_r = -kv^2$ , donde  $g = 9.8$  m/s y  $k = 0.002$  kg/m. La ecuación diferencial para la velocidad  $v$  está dada por

$$mv' = -mg - kv^2$$

Encuentre la velocidad del proyectil a diferentes tiempos en su ascenso y el tiempo que tarda en llegar a su altura máxima.

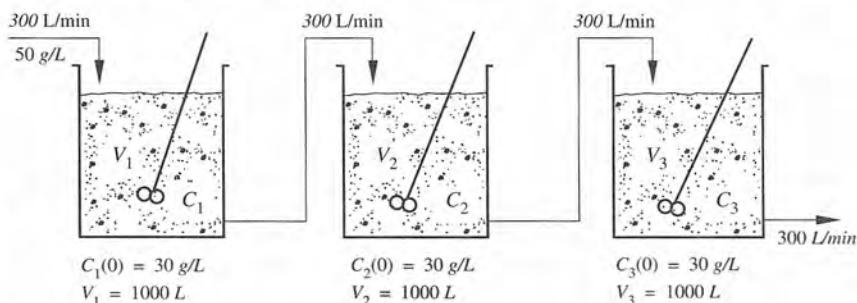
**Solución** Al emplear el método de Runge-Kutta de cuarto orden con  $h = 0.01$  se tiene

| $t \text{ (s)}$ | $v \text{ (m/s)}$ |
|-----------------|-------------------|
| 0               | 80                |
| 0.3             | 53.55             |
| 0.6             | 39.11             |
| 0.9             | 29.76             |
| 1.2             | 23.04             |
| 1.5             | 17.83             |
| 1.8             | 13.55             |
| 2.1             | 9.86              |
| 2.4             | 6.54              |
| 2.7             | 3.46              |
| 3.00            | 0.49              |
| 3.01            | 0.39              |
| 3.02            | 0.30              |
| 3.03            | 0.20              |
| 3.04            | 0.10              |
| 3.05            | 0.002             |
| 3.06            | -0.10             |

Dado que al llegar a  $t = 3.06$  s, la velocidad es negativa, se toma 3.05 como el lapso que tarda en llegar a su altura máxima.

- 7.7 Se tiene tres tanques de 1000 litros de capacidad cada uno, perfectamente agitados (véase Fig. 7.18). Los tres recipientes están completamente llenos con una solución cuya concentración es 30 g/L. A partir de cierto momento se alimenta al primer tanque una solución que contiene 50 g/L con un gasto de 300 L/min (hay un arreglo entre los tres recipientes tal que al haber un gasto al primero, la misma cantidad fluye de éste al segundo, del segundo al tercero y de éste afuera del sistema, con lo cual se mantiene constante el volumen en todos ellos).

Calcule la concentración en cada tanque después de 10 minutos de haber empezado a agregar solución al primero.



**Figura 7.18**  
Arreglo de tres  
tanques  
interconectados.

**Solución**

Balance de soluto en el primer tanque

$$\begin{array}{lcl} \text{Acumulación} & = & \text{Entrada} \quad \quad \quad - \text{Salida} \\ \frac{dC_1 V_1}{dt} & = & 300 (50) \quad \quad \quad - 300 C_1 \end{array}$$

como  $V_1 = 1000$  L y permanece constante

$$\frac{dC_1}{dt} = 15 - 0.3 C_1 \text{ y } C_1(0) = 30 \quad (1)$$

Balance de soluto en el segundo tanque

$$V_2 \frac{dC_2}{dt} = 300 C_1 - 300 C_2$$

como

$$V_2 = 1000 \text{ L}$$

$$\frac{dC_2}{dt} = 0.3 (C_1 - C_2) \text{ y } C_2(0) = 30 \quad (2)$$

Balance de soluto en el tercer tanque

$$V_3 \frac{dC_3}{dt} = 300 C_2 - 300 C_3$$

como

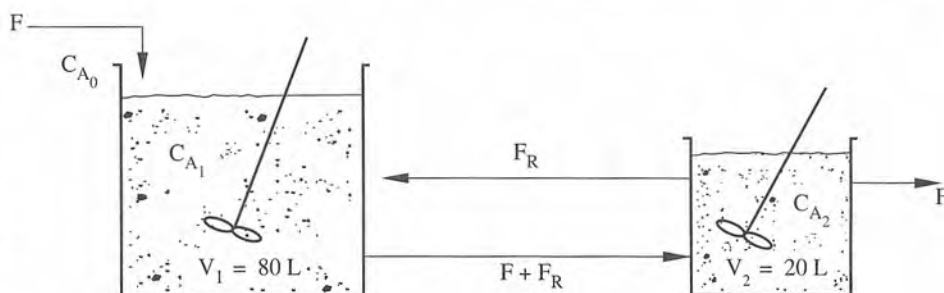
$$V_3 = 1000 \text{ L}$$

$$\frac{dC_3}{dt} = 0.3 (C_2 - C_3) \text{ y } C_3(0) = 30 \quad (3)$$

Las ecuaciones 1 a 3, con sus respectivas condiciones iniciales, constituyen un sistema cuya solución representa los valores buscados, esto es:

$$\text{PVI} \left\{ \begin{array}{l} \frac{dC_1}{dt} = 15 - 0.3C_1 \\ \frac{dC_2}{dt} = 0.3 (C_1 - C_2) \\ \frac{dC_3}{dt} = 0.3 (C_2 - C_3) \\ C_1(0) = 30 \\ C_2(0) = 30 \\ C_3(0) = 30 \\ C_1(10) = ? \\ C_2(10) = ? \\ C_3(10) = ? \end{array} \right.$$

Se utiliza un paso de integración de 1 min y el algoritmo de RK-4 para sistemas y se tiene



**Figura 7.19**  
Modelación de  
un reactor con  
mezclado  
imperfecto.

| tiempo ( min ) | C1    | C2    | C3    |
|----------------|-------|-------|-------|
| 0              | 30.00 | 30.00 | 30.00 |
| 1              | 35.18 | 30.74 | 30.07 |
| 2              | 39.02 | 32.44 | 30.46 |
| 3              | 41.87 | 34.55 | 31.25 |
| 4              | 43.98 | 36.75 | 32.41 |
| 5              | 45.54 | 38.85 | 33.82 |
| 6              | 46.69 | 40.74 | 35.89 |
| 7              | 47.55 | 42.41 | 37.01 |
| 8              | 48.19 | 43.83 | 38.61 |
| 9              | 48.66 | 45.03 | 40.13 |
| 10             | 49.00 | 46.02 | 41.54 |

- 7.8** El mezclado imperfecto en un reactor continuo de tanque agitado se puede modelar como dos o más reactores con recirculación entre ellos, como se muestra en la figura 7.19.

En este sistema se lleva a cabo una reacción isotérmica irreversible del tipo  $A \xrightarrow{k} B$  de orden 1.8 con respecto al reactante A. Con los datos que se dan abajo, calcule la concentración del reactante A en los reactores (1) y (2) ( $C_{A1}$  y  $C_{A2}$ , respectivamente) durante el tiempo necesario para alcanzar el régimen permanente. Ensaye varios tamaños de paso de integración y compare los resultados obtenidos en el ejercicio 4.5.

Datos:

$$\begin{aligned}
 F &= 25 \text{ L/min} & C_{A0} &= 1 \text{ mol/L} \\
 F_R &= 100 \text{ L/min} & C_{A1}(0) &= 0.0 \text{ mol/L} \\
 C_{A2}(0) &= 0.0 \text{ mol/L} & k &= 0.2 \left( \frac{\text{L}}{\text{mol}} \right)^{0.8} \text{ min}^{-1}
 \end{aligned}$$

**Solución** Un balance del componente A en cada uno de los reactores da



Acumulación = Entrada - Salida - Reacciona

Reactor 1

$$\frac{dV_1 C_{A1}}{dt} = F C_{A0} + F_R C_{A2} - (F + F_R) C_{A1} - V_1 k C_{A1}^{1.8}$$

Reactor 2

$$\frac{dV_2 C_{A2}}{dt} = (F + F_R) C_{A1} - (F + F_R) C_{A2} - V_2 k C_{A2}^{1.8}$$



Como  $V_1$  y  $V_2$  son constantes, mediante la sustitución de valores y con las condiciones de operación a tiempo cero, se llega a

$$\text{PVI} \left\{ \begin{array}{l} \frac{dC_{A1}}{dt} = 1.25 C_{A2} + \frac{25}{80} - \frac{125}{80} C_{A1} - 0.2 C_{A1}^{1.8} \\ \frac{dC_{A2}}{dt} = \frac{125}{20} (C_{A1} - C_{A2}) - 0.2 C_{A2}^{1.8} \\ C_{A1}(0) = 0.0 \\ C_{A2}(0) = 0.0 \\ C_{A1}(0 \text{ a } r p) = ? \\ C_{A2}(0 \text{ a } r p) = ? \end{array} \right.$$

donde  $C_{A1}(0 \text{ a } r p)$  significa la concentración del reactante A en el reactor 1 desde el tiempo 0 hasta alcanzar el régimen permanente.

Con el programa 7.3 y con un paso de integración de 0.4 minutos, el valor de  $C_{A2}$  en la primera iteración resulta negativo (lo cual es imposible) y al efectuar la segunda iteración e intentar calcular el término  $C_{A2}^{1.8}$  (véase segunda Ec. del PVI) el programa aborta. Se ensaya ahora un tamaño de paso menor, ya que la constante de velocidad de reacción es alta y es de esperarse que la reacción sea muy rápida y que un paso de 0.4 minutos resulte muy grande. A continuación se dan los resultados para  $h = 0.3$  minutos.

#### CONDICIONES INICIALES :

Y1 ( .00) = .000

Y2 ( .00) = .000

PASO DE INTEGRACIÓN H = .300

VALOR FINAL DE X = 20.000

SE IMPRIME CADA 5 INTERACCIONES

| X       | Y1    | Y2    |
|---------|-------|-------|
| 1.5000  | .3143 | .2796 |
| 3.0000  | .4839 | .4635 |
| 4.5000  | .5706 | .5528 |
| 6.0000  | .6123 | .5966 |
| 7.5000  | .6321 | .6172 |
| 9.0000  | .6413 | .6268 |
| 10.5000 | .6456 | .6313 |
| 12.0000 | .6476 | .6334 |
| 13.5000 | .6485 | .6343 |
| 15.0000 | .6489 | .6348 |
| 16.5000 | .6491 | .6350 |
| 18.0000 | .6492 | .6351 |
| 19.5000 | .6493 | .6351 |

Puede observarse que el régimen permanente se alcanza a los 18 minutos. Los valores de las concentraciones a régimen permanente coinciden con los obtenidos en el ejercicio 4.5.

Se probaron además tamaños de paso 0.25, 0.2 y 0.1 minutos; en cada caso los mismos resultados se obtuvieron que para 0.3 minutos.

- 7.9** En un reactor de laboratorio continuo tipo tanque perfectamente agitado, se lleva a cabo una reacción química exotérmica cuya temperatura se controla por medio de un líquido que circula por una chaqueta que se mantiene a una temperatura uniforme  $T_j$ . Calcule la temperatura  $T$  y la concentración  $C_A$  de la corriente de salida cuando el reactor trabaja a régimen transitorio y hasta alcanzar el régimen permanente para el caso de una reacción de primer orden.

Aplice la siguiente información referida a la figura 7.20.

Condiciones iniciales:  $C_A(0) = 5 \text{ gmol/L}$  y  $T(0) = 300 \text{ K}$

$F$  = Gasto de alimentación al reactor = 10 ml/s

$V$  = Volumen de reactor = 2000 ml

$C_{A0}$  = Concentración del reactante A en el flujo de alimentación =  $5 \frac{\text{gmol}}{\text{L}}$

$T_0$  = Temperatura del flujo de alimentación = 300 K

$\Delta H$  = Calor de reacción =  $-10000 \text{ cal/gmol}$

$U$  = Coeficiente global de transmisión de calor =  $100 \frac{\text{cal}}{^\circ\text{C s m}^2}$

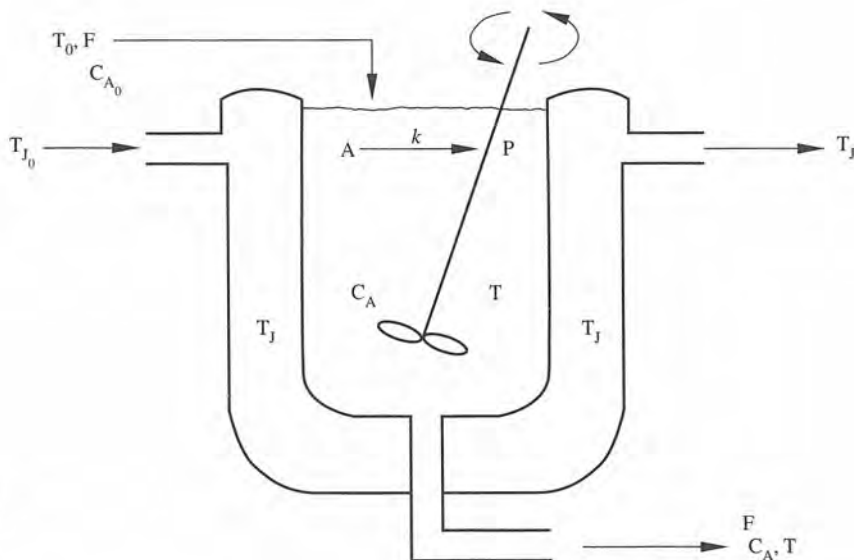
$A$  = Área de transmisión de calor =  $0.02 \text{ m}^2$

$k$  = Constante de velocidad de reacción =  $8 \times 10^{12} \exp(-22500/1.987 T) \text{ s}^{-1}$

$T_j$  = Temperatura del líquido que circula por la chaqueta = 330 K

$C_p$  = Calor específico de la masa reaccionante =  $1 \text{ Kcal/kg}^\circ\text{C}$

$\rho$  = Peso específico de la masa reaccionante =  $1 \text{ kg/L}$



**Figura 7.20**  
Reactor tipo  
tanque agitado  
con chaqueta.

**Solución** Balance de materia para el reactante A

$$\begin{array}{rclcl} \text{Acumulación} & = & \text{Entrada} & - & \text{Salida} & - & \text{Reacciona} \\ \frac{dVC_A}{dt} & = & FC_{A0} & - & FC_A & - & k VC_A^n \end{array}$$

Balance de calor

$$\begin{array}{rclcl} \text{Acumulación} & = & \text{Entrada} & - & \text{Salida-generado} & - & \text{eliminado} \\ \frac{dV\rho CpT}{dt} & = & F\rho Cp(T_0 - T) & - & \Delta HkVC_A^n & - & UA(T - T_j) \end{array}$$

Como  $V$ ,  $\rho$  y  $C_p$  se consideran constantes, al sustituir valores se tiene

$$\text{PVI} \left\{ \begin{array}{l} \frac{dC_A}{dt} = 0.005(5 - C_A) - 8 \times 10^{12} \exp(-22500/1.98T) C_A \\ \frac{dT}{dt} = 0.005(300 - T) + 8 \times 10^{13} \exp(-22500/1.98T) C_A - 0.001(T - 330) \\ C_A(0) = 5 \text{ gmol/L} \\ T(0) = 300 \text{ K} \end{array} \right.$$

Al resolver con el **PROGRAMA 7.3**, que utiliza el método de Runge-Kutta de tercer orden para un sistema de ecuaciones, se obtienen los resultados:

SOLUCIÓN DE UN PVI CON UN SISTEMA DE N  
ECUACIONES DIFERENCIALES ORDINARIAS DE PRIMER ORDEN  
POR EL MÉTODO DE RUNGE-KUTTA DE TERCER ORDEN

CONDICIONES INICIALES:

$$Y1(.00) = 5.000$$

$$Y2(.00) = 300.000$$

$$\text{PASO DE INTEGRACIÓN } H = 20.000$$

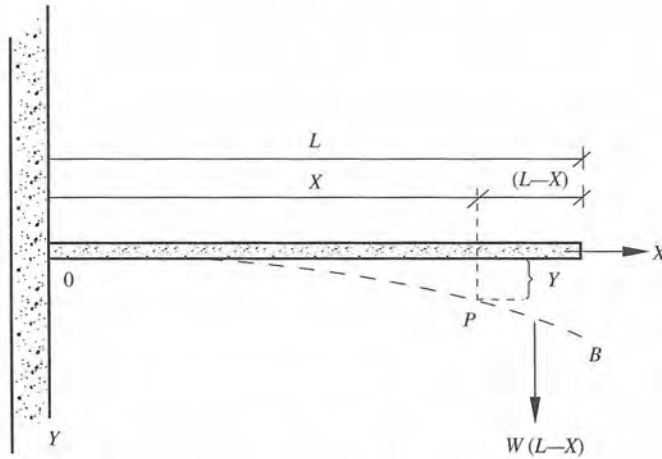
$$\text{VALOR FINAL DE } X = 3000.000$$

SE IMPRIME CADA 10 ITERACIONES

| X         | Y1     | Y2       |
|-----------|--------|----------|
| .0000     | 5.0000 | 300.0000 |
| 200.0000  | 4.6623 | 306.6382 |
| 400.0000  | 4.3180 | 310.6624 |
| 600.0000  | 3.9803 | 313.8187 |
| 800.0000  | 3.6243 | 316.9112 |
| 1000.0000 | 2.1727 | 320.8165 |
| 1200.0000 | 2.3743 | 327.8928 |
| 1400.0000 | .7730  | 342.0851 |
| 1600.0000 | .6438  | 341.9108 |
| 1800.0000 | .7104  | 340.8714 |
| 2000.0000 | .7314  | 340.5911 |
| 2200.0000 | .7359  | 340.5366 |
| 2400.0000 | .7366  | 340.5287 |
| 2600.0000 | .7367  | 340.5278 |
| 2800.0000 | .7367  | 340.5278 |
| 3000.0000 | .7367  | 340.5278 |

- 7.10** Encuentre la curva elástica de una viga uniforme con un extremo libre, de longitud  $L = 5$  m y peso constante de  $w = 300$  kg. Determine también la deflexión del extremo libre. Tome  $EI = 150\,000$ .

**Figura 7.21**  
Viga empotrada  
con un extremo  
libre.



**Solución**



La figura 7.21 muestra la viga y su curva elástica (línea punteada). Se toma el origen  $O$  de un sistema coordenado en el extremo empotrado de la viga y la dirección positiva del eje  $y$  hacia abajo.

Sea  $x$  un punto cualquiera de la viga. Para calcular el momento de flexión en el punto  $x$ ,  $M(x)$ , considere la parte de la viga a la derecha de  $P$  y que sólo una fuerza hacia abajo actúa en esa porción,  $w(L-x)$ , produciendo el momento positivo

$$M(x) = w(L-x) \left[ \frac{(L-x)}{2} \right] = w(L-x)^2/2$$

En la teoría de vigas, se demuestra que  $M(x)$  está relacionado con el radio de curvatura de la curva elástica calculado en  $x$  así

$$EI \frac{y'''}{[1 + (y')^2]^{1/2}} = M(x) \quad (1)$$

donde  $E$  es el módulo de elasticidad de Young y depende del material con el que se construyó la viga e  $I$  es el momento de inercia de la sección transversal de la viga en  $x$ .

Si se asume que la viga se flexiona muy poco, que es el caso general, la pendiente  $y'$  de la curva elástica es tan pequeña que

$$1 + (y')^2 \approx 1$$

y la ecuación 4 puede aproximarse por

$$EI y'' = M(x) = w(L-x)^2/2$$

Al cambiar de variable en la forma  $y' = dy/dx = z$ , se obtiene el siguiente

$$\text{PVI} \begin{cases} \frac{dy}{dx} = z \\ \frac{dz}{dx} = \frac{w(L-x)^2}{2EI} \\ y(0) = 0 \\ z(0) = 0 \\ y(5) = ? \end{cases}$$



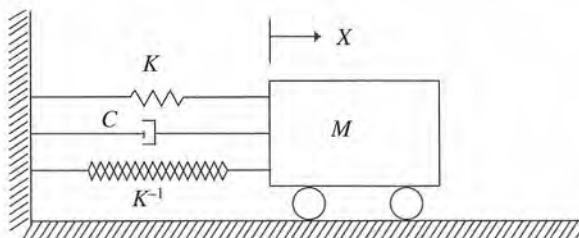
Con el PROGRAMA 7.3 y con  $h = 0.5 \text{ m}$  se obtiene

| $x(\text{m})$ | $y(\text{m})$ |
|---------------|---------------|
| 0             | 0             |
| 0.5           | 0.003         |
| 1.0           | 0.011         |
| 1.5           | 0.023         |
| 2.0           | 0.038         |
| 2.5           | 0.055         |
| 3.0           | 0.074         |
| 3.5           | 0.094         |
| 4.0           | 0.115         |
| 4.5           | 0.135         |
| 5.0           | 0.156         |

- 7.11** En la industria del transporte terrestre y aéreo surgen problemas de choque y vibración a partir de muy diferentes tipos de fuentes de excitación. La eliminación del choque y de la vibración es de importancia vital para aislar instrumentos y controles, y en la protección de los ocupantes de los vehículos. La solución usual a los problemas que involucran transmisión de vibraciones excesivas incluye el uso de soportes flexibles levemente amortiguados. Estos soportes suaves causan que la frecuencia natural de un sistema de suspensión quede por debajo de la frecuencia de disturbio. Esta solución es efectiva para aislar la vibración en estado estacionario; sin embargo, cuando estas suspensiones encuentran situaciones de choque, su suavidad a menudo lleva a deflexiones grandes dañinas. Se ha señalado que estas características no deseables están ausentes en los sistemas de suspensión que utilizan resortes simétricamente no lineales que se rigidizan. Estos resortes son progresivamente más rígidos al sujetarse a deflexiones grandes a partir del “punto de operación”. El dispositivo mostrado en la figura 7.22 consiste de un objeto de masa  $m$  conectado a una pared por medio de un resorte lineal con coeficiente  $k$ , un amortiguador con coeficiente  $c$ , y un resorte no lineal que ejerce una fuerza de recuperación proporcional a una constante  $k'$  veces la tercera potencia del desplazamiento. Este resorte “cúbico” provee un comportamiento no lineal simétrico que satisface la necesidad de aislar el choque y la vibración.

Debido a que la ecuación diferencial que describe el movimiento de este sistema es no lineal:

$$m \frac{d^2x}{dt^2} + c \frac{dx}{dt} + kx + k'x^3 = 0$$



**Figura 7.22**  
Un objeto  
amortiguado  
conectado a  
una pared.

el desplazamiento de  $x$  en función del tiempo no puede encontrarse con los métodos analíticos tradicionales. Por esta razón se usa una solución numérica a esta ecuación diferencial.

Si los parámetros físicos del sistema de suspensión son:

$$k = 2.0 \text{ N/cm}; k' = 0.2 \text{ N/cm}^3; c = 0.15 \text{ Ns/cm}; m = 0.01 \text{ kg}$$

y las condiciones iniciales son:

$$x(0) = 10 \text{ cm desplazamiento del objeto en la dirección positiva del eje } x,$$

$$x'(0) = 0 \text{ cm/s velocidad inicial que se imprime al objeto,}$$

elabore y ejecute un programa que simule el movimiento de este sistema desde un tiempo cero hasta un segundo.

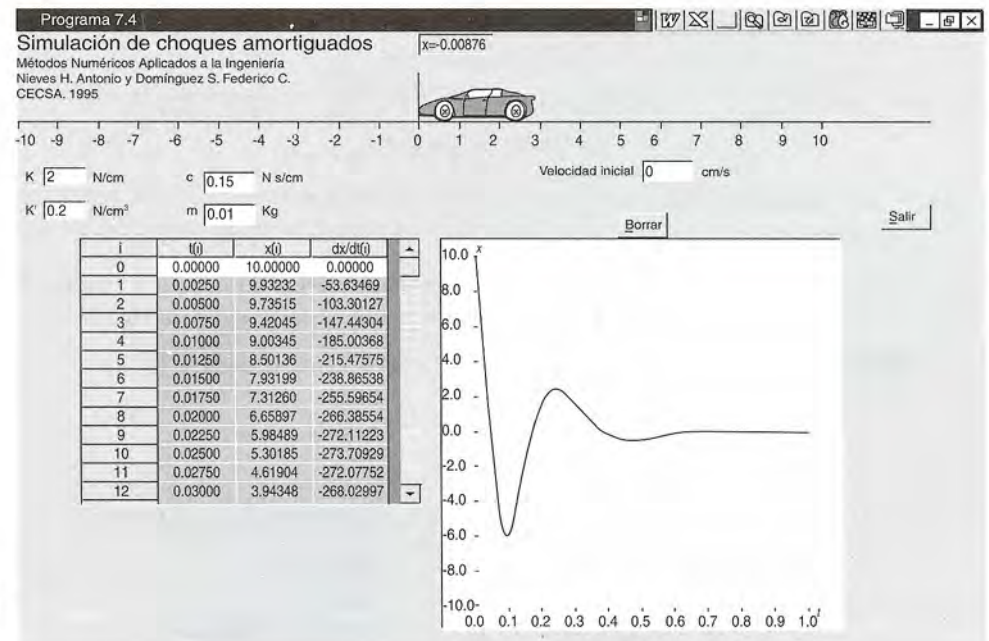
### Solución



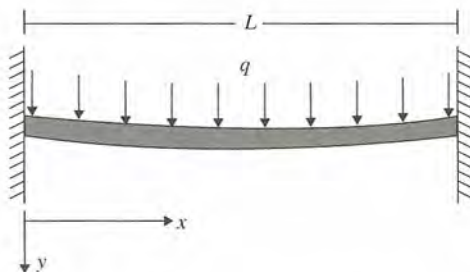
Pasando la ecuación diferencial a un sistema de ecuaciones diferenciales de primer orden y las condiciones iniciales a términos de las nuevas variables, se tiene:

$$\text{PVI} \left\{ \begin{array}{l} \frac{dx}{dt} = z \\ \frac{dz}{dt} = \frac{-cz - kx - k'x^3}{m} \\ x(0) = 10 \\ z(0) = 0 \\ x(t) = ? \quad \text{para } 0 < t < 1 \end{array} \right.$$

En el CD se encuentra el **PROGRAMA 7.4** que resuelve este PVI con el método de Runge-Kutta de cuarto orden y  $h = 0.0025$ . El programa simula el movimiento amortiguado de un carrito y muestra una tabla y la gráfica correspondiente de los valores de la posición del carrito a diferentes tiempos. En la figura 7.23 se muestra una impresión de la interfase, donde por ejemplo puede observarse que la variabilidad de la magnitud de la velocidad  $dx/dt$  refleja el hecho de que hay una aceleración (ver tabla de la figura 7.23). El signo negativo de la velocidad puede interpretarse como que el carrito avanza en dirección opuesta al eje  $x$ . En la gráfica de la figura 7.23 también puede observarse que el carrito se desplaza hasta



**Figura 7.23**  
Interfase del  
programa 7.4.



**Figura 7.24**  
Viga  
rectangular con  
extremos fijos.

el valor  $x = -6$  cm transcurridos alrededor de 0.1 segundos para luego avanzar en la dirección positiva del eje  $x$  hasta llegar a 2.4 cm y después regresarse; también puede verse que transcurridos 0.7 segundos, el carrito prácticamente se ha detenido. El programa permite cambiar los valores de  $k$ ,  $k'$ ,  $c$  y  $m$  para que el lector pueda simular el fenómeno con diferentes parámetros y sacar sus propias conclusiones (para iniciar la simulación, arrastre el carrito con el ratón, tomándolo de la argolla y suéltelo,  $x'(0) = 0$ ).

- 7.12** Un problema común en ingeniería civil es el cálculo de la deflexión de una viga rectangular sujeta a carga uniforme, cuando los extremos de la viga están fijos y, por tanto, no experimentan deflexión. La ecuación diferencial que aproxima este fenómeno físico tiene la forma (véase ejercicios 2.10 y 7.10).

$$\frac{d^2y}{dx^2} = \frac{S}{EI}y + \frac{qx}{2EI}(x - L) \quad (1)$$

En la ecuación (1),  $y$  es la deflexión de la viga a una distancia  $x$ , medida a partir del extremo izquierdo (ver Fig. 7.24),  $L$  la longitud,  $q$  es la intensidad de la carga uniforme,  $S$  el esfuerzo o tensión en los extremos,  $I$  es el momento de inercia que depende de la forma de la sección transversal de la viga y  $E$  el módulo de elasticidad.

Dado que los extremos de la viga están fijos, se tiene:

$$y(0) = y(L) = 0.$$

La ecuación (1) conjuntamente con estas condiciones constituye un problema de valores en la frontera, esto es:

$$\text{PVF} \left\{ \begin{array}{l} \frac{d^2y}{dx^2} = \frac{S}{EI}y + \frac{qx}{2EI}(x - L) \\ y(0) = 0 \\ y(L) = 0 \\ y(x) = ? \quad \text{para } 0 < x < L \end{array} \right.$$

Suponga que se tienen los siguientes datos:  $L = 350$  cm,  $q = 1$  kg/cm,  $E = 2 \times 10^6$  kg/cm<sup>2</sup>,  $S = 400$  kg,  $I = 2.5 \times 10^4$  cm<sup>4</sup>. Encuentre la deflexión de la viga cada 10 cm, usando  $\varepsilon = 10^{-8}$ .

**Solución** Pasando la ecuación diferencial a un sistema se tiene:

$$\text{PVF} \left\{ \begin{array}{l} \frac{dy}{dx} = z \\ \frac{dz}{dx} = \frac{S}{EI}y + \frac{qx}{2EI}(x - L) \\ y(0) = 0 \\ y(L) = 0 \\ y(x) = ? \quad \text{para } 0 < x < L \end{array} \right.$$



Se inicia el método del disparo con  $\alpha_0 = 0.01$  y  $\alpha_1 = 0.02$ , ya que la deflexión en general es muy pequeña (en la figura 7.24 se ha exagerado con fines ilustrativos). Los valores positivos de  $\alpha$  (recuerde que representan la pendiente de la tangente a la curva y en el extremo izquierdo de la viga) se deben a que la dirección positiva del eje y es hacia abajo. Con estos valores de  $\alpha$  se plantean los dos PVI's siguientes:

$$\text{PVI}_0 \left\{ \begin{array}{l} \frac{dy}{dx} = z \\ \frac{dz}{dx} = \frac{S}{EI} y + \frac{qx}{2EI} (x - L) \\ y(0) = 0 \\ z(0) = \alpha_0 = 0.01 \\ y(x) = ? \quad \text{para } 0 < x < L \end{array} \right. \quad \text{PVI}_1 \left\{ \begin{array}{l} \frac{dy}{dx} = z \\ \frac{dz}{dx} = \frac{S}{EI} y + \frac{qx}{2EI} (x - L) \\ y(0) = 0 \\ z(0) = \alpha_1 = 0.02 \\ y(x) = ? \quad \text{para } 0 < x < L \end{array} \right.$$

Al resolverlos con el método de Runge-Kutta de cuarto orden y  $h = 5$  cm, se obtiene

$$y(L; \alpha_0) = 3.4880656693$$

$$y(L; \alpha_1) = 6.988637364$$

La interpolación inversa (en este caso realmente es una extrapolación) da

$$\begin{aligned} \alpha_2 &= \alpha_1 - (\alpha_1 - \alpha_0) \frac{y(L; \alpha_1) - y(2)}{y(L; \alpha_1) - y(L; \alpha_0)} \\ &= 0.2 - (0.2 - 0.1) \frac{6.988637364 - 0}{6.988637364 - 3.4880656693} = 0.000035726 \end{aligned}$$

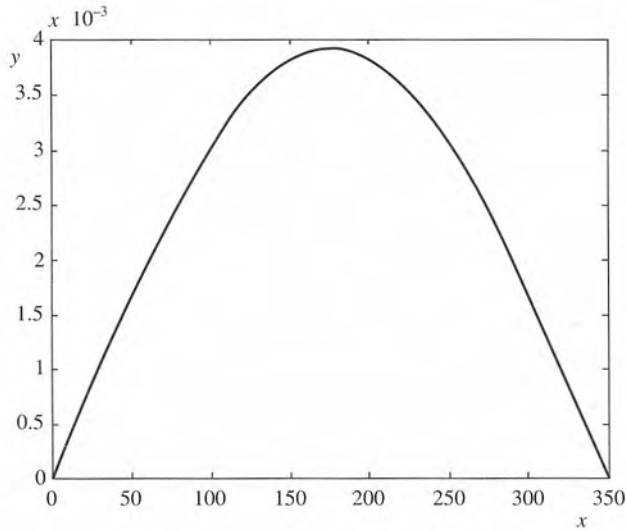
Con este valor se resuelve el nuevo PVI

$$\text{PVI}_2 \left\{ \begin{array}{l} \frac{dy}{dx} = z \\ \frac{dz}{dx} = \frac{S}{EI} y + \frac{qx}{2EI} (x - L) \\ y(0) = 0 \\ z(0) = \alpha_2 = 0.000035726 \\ y(x) = ? \quad \text{para } 0 < x < L \end{array} \right.$$

con lo que se obtiene  $y(L; \alpha_2) = 0.00000000$ , y el problema queda terminado. Los resultados son (se muestran cada 25 cm y hasta la mitad de la viga ya que son simétricos):

| $x(\text{cm})$ | $y(\text{cm})$ |
|----------------|----------------|
| 0.0            | 0.0000000000   |
| 25.0           | 0.0008843533   |
| 50.0           | 0.0017185808   |
| 75.0           | 0.0024597179   |
| 100.0          | 0.0030726121   |
| 125.0          | 0.0035299226   |
| 150.0          | 0.0038121204   |
| 175.0          | 0.0039074883   |





**Figura 7.25**  
Gráfica de la  
deflexión de la  
viga.

Los valores de la tabla revelan que la deflexión y de la viga es imperceptible para el ojo; en la figura 7.25 puede verse que la deformación de la viga es simétrica y que la máxima deflexión se da en el centro, como era de esperar.

Los cálculos pueden realizarse con Matlab:



```
h=10; Eps=1e-8; L=350; xf=L; yf=0;
alfa(1)=.1; alfa(2)=.2;
for k=1:20
x0=0; y0=0; z0=alfa(k); x(1)=x0; y(1)=y0;
fprintf(' %12.9f\n',z0);
for i=1:35
k1=z0; c1=Ejer7_12f(x0,y0,z0);
k2=z0+h/2*c1; c2=Ejer7_12f(x0+h/2,y0+h/2*c1,z0+h/2*c1);
k3=z0+h/2*c2; c3=Ejer7_12f(x0+h/2,y0+h/2*c2,z0+h/2*c2);
k4=z0+h*c3; c4=Ejer7_12f(x0+h,y0+h*c3,z0+h*c3);
y0=y0+h/6*(k1+2*k2+2*k3+k4); z0=z0+h/6*(c1+2*c2+2*c3+c4);
x0=x0+h; x(i+1)=x0; y(i+1)=y0;
end
yt(k)=y0;
if k>= 2
alfa(k+1)=alfa(k)-(alfa(k)-alfa(k-1))*...
(yt(k)-yf)/(yt(k)-yt(k-1));
end
Error=abs(y0-yf);
for i=1:36
fprintf('%6.1f %12.10f\n',x(i),y(i))
end
if Error<Eps
break
end
end
plot(x,y,'k')
```

```
function f=Ejer7_12f(x,y,z)
L=350; q=1; E=2e6; S=400; I=2.5e4; EI=E*I;
f=S*y/EI+q*x/(2*EI)*(x-L);
```



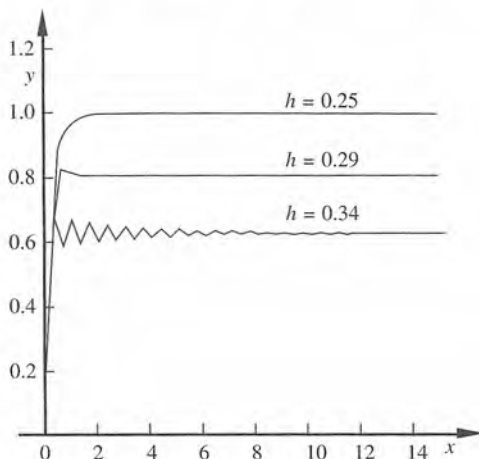
**7.13** El empleo de los métodos numéricos proporciona solamente soluciones aproximadas; no obstante debemos diferenciar éstas de soluciones incorrectas. Una forma práctica y rápida de probar si los resultados son aproximados a la solución sería correr el programa reduciendo el tamaño de paso a la mitad y comparar resultados en los mismos puntos. Se resuelve enseguida un problema\* donde el tamaño de paso es clave para obtener la solución correcta, pero donde las soluciones incorrectas pueden resultar de mayor interés aún, por el aparente caos que presentan. El problema de valor inicial al que nos referimos es:

$$\text{PVI} \begin{cases} \frac{dy}{dx} = 10y(1-y) \\ y(0) = 0.1 \\ y(x) = ? \end{cases} \quad (1)$$

cuya ecuación diferencial es conocida como Ecuación Logística. La solución analítica, obtenida por variables separables, es  $y = e^{10x}/(9 + e^{10x})$ , donde al tender  $x$  hacia valores grandes,  $y$  tiende hacia 1.

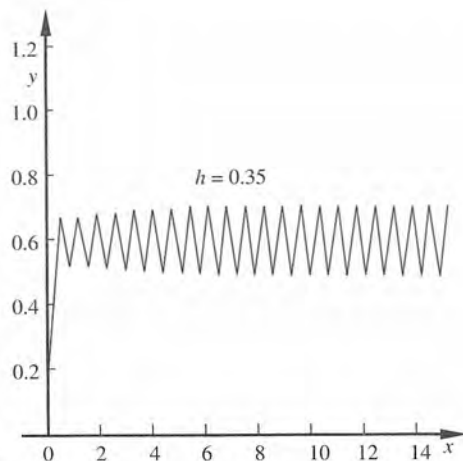
Las figuras 7.26 a 7.29 muestran algunas soluciones numéricas empleando el método de Runge-Kutta de cuarto orden con diferentes tamaños de paso  $h$ . En la figura 7.26 vemos que con  $h = 0.25$ , la solución numérica se aproxima a  $y(x) = 1$ , como debería hacerlo. Con  $h = 0.29$  y  $h = 0.34$ , la solución numérica se aproxima a  $y(x) = 0.76186$  y  $y(x) = 0.63153$ , respectivamente, oscilando esta última al principio. La solución numérica en la figura 7.27 (con  $h = 0.35$ ) no se aproxima a un solo valor, pero eventualmente oscila entre los dos valores 0.71982 y 0.50541. Decimos que esta solución numérica tiene periodo 2. Del mismo modo se dice que las soluciones numéricas anteriores que se aproximan a un valor, a saber, 1, 0.76186 y 0.63153 tienen periodo 1. La solución numérica en la figura 7.28 (con  $h = 0.365$ ) no oscila entre dos valores, pero eventualmente oscila entre cuatro: 0.39189, 0.82492, 0.41291 y 0.79970 y así la solución numérica tiene periodo cuatro. La

**Figura 7.26**  
Soluciones  
numéricas del  
PVI (1) con  
diferentes  
tamaños de  
paso.

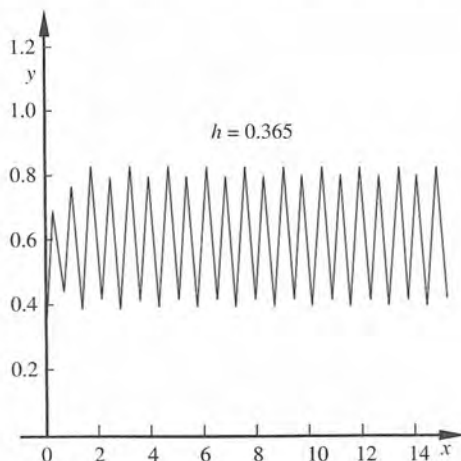


\* Tomado de: Lomen, David y Lovelock, David, Ecuaciones Diferenciales a través de Gráficas, Modelos y Datos, CECSA (2000) pp. 106-109.

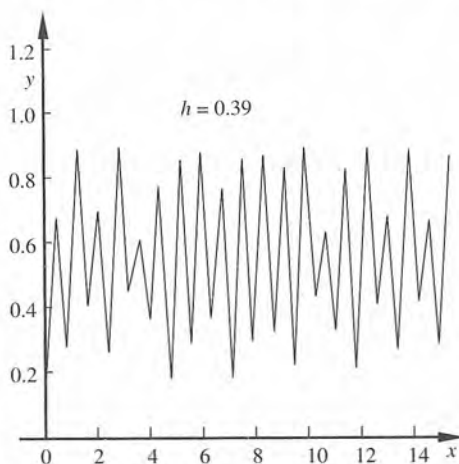
**Figura 7.27**  
Soluciones  
numéricas del  
PVI (1) con  
tamaño de paso  
igual a 0.35.



**Figura 7.28**  
Soluciones  
numéricas del  
PVI (1) con  
tamaño de paso  
igual a 0.365.



**Figura 7.29**  
Soluciones  
numéricas del  
PVI (1) con  
tamaño de paso  
igual a 0.39.



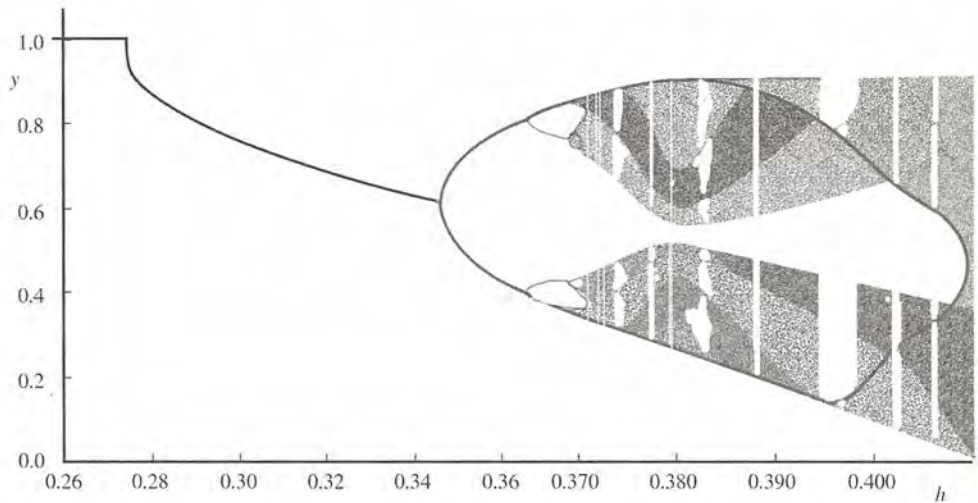
solución numérica en la figura 7.29 (con  $h = 0.39$ ) es bastante caótica; no muestra un comportamiento periódico. Todas estas soluciones numéricas, exceptuando la correspondiente a  $h = 0.25$ , están equivocadas y no predicen el comportamiento correcto de la solución.

Podemos llegar a comprender este proceso si resolvemos la ecuación diferencial numéricamente para valores particulares del tamaño de paso  $h$ , y posteriormente examina-

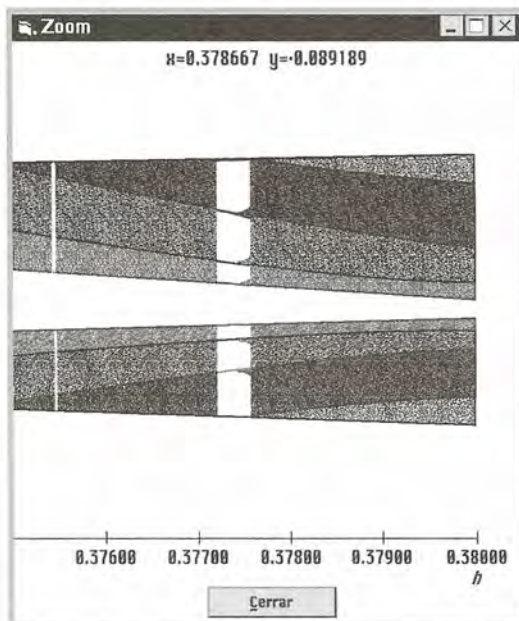


**Figura 7.30**

Diagrama de órbita para la solución numérica del PVI (1).



mos los últimos valores generados por el método numérico, para ver si eventualmente siguen un patrón. Una manera de ver esto gráficamente es comenzar con una  $h$  específica, calcular la solución numérica para, digamos, los primeros 100 pasos, y después graficar los siguientes 100 valores de  $y$  correspondientes a ese valor de  $h$ ; luego incrementar  $h$  y repetir el proceso. La figura 7.30 presenta una gráfica de esta clase, conocida a menudo como un diagrama de órbita, donde se manejan escalas diferentes en el eje  $h$ , a fin de tener una mayor "resolución" en el diagrama para valores de  $h$  mayores de 0.364. En el disco del libro de encuentra el **PROGRAMA 7.5** donde puede obtenerse cualquiera de las figuras 7.26 a 7.30; en el último caso se puede incluso obtener un acercamiento (zoom) seleccionando un intervalo en el eje  $h$  a partir de 0.364 (donde resulta relevante este tipo de tratamiento). Para ello, el lector puede utilizar el sistema de ejes coordenados y señalar a través de los botones del *mouse* o escribiendo en los cuadros correspondientes el valor inicial y final del intervalo de interés. La figura 7.31 muestra un acercamiento en el intervalo (0.375, 0.380).

**Figura 7.31**

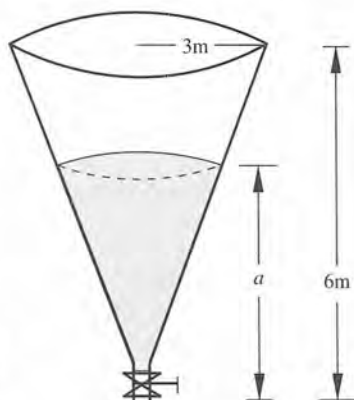
Acercamiento con el programa 7.5 en el intervalo (0.375, 0.380).



Para asegurarnos de haber compendido la figura 7.30, examinemos el punto en el diagrama que corresponde a un tamaño de paso de 0.25, observamos que los últimos 100 valores numéricos de  $y$  son aproximadamente el mismo número, a saber, 1. Esto concuerda con la curva  $h = 0.25$  en la figura 7.26. De la misma manera, con  $h = 0.30$ , la figura 7.30 nos dice que los últimos 100 valores numéricos de  $y$  son aproximadamente el mismo número, a saber, 0.76186, otra vez de acuerdo con la figura 7.26. Con  $h = 0.35$ , la figura 7.30 nos dice que los últimos 100 valores numéricos de  $y$  caen aproximadamente en dos números diferentes, a saber, 0.71982 y 0.50541, lo que concuerda con la figura 7.27. Así, cuando  $h$  se encuentra aproximadamente entre 0.346 y 0.364 esperamos soluciones de periodo 2. Podemos ver las soluciones de periodo 4 ( $h$  entre 0.365 y 0.368) y las soluciones caóticas (con  $h = 0.39$ ) en la figura 7.30. ¿Puede apreciar algunas soluciones de periodo 8? ¿Dónde buscaría una solución de periodo 16? ¿Esperaría el mismo diagrama de órbita usando otros métodos de solución de este PVI?

## Problemas

- 7.1 Si al tanque de la figura 7.32, al momento de llegar el nivel del líquido a 0.5 m se hace llegar un gasto de alimentación de  $0.04 \text{ m}^3/\text{s}$ , el nivel del líquido aumentará. Determine el tiempo necesario para que el nivel se recupere nuevamente a 3 m. El flujo de salida por el orificio del fondo es  $15.55 \sqrt{a} \text{ L/s}$ .
- 7.2 El tiempo que requiere el tanque del ejercicio 7.1 para recuperar su nivel de 0.5 a 3 m con un gasto de alimentación de  $0.04 \text{ m}^3/\text{s}$  es de aproximadamente 432 s. Calcule el gasto de alimentación que se requiere para reducir este tiempo a la mitad.
- 7.3 Calcule el tiempo necesario para que el nivel del líquido del tanque de la figura 7.32 pase de 6 m a 1 m. El flujo de salida por el orificio del fondo es  $3.457 \sqrt{a} \text{ L/s}$ .
- 7.4 Se hace llegar un gasto de alimentación de 7 L/s al tanque de la figura 7.32 cuando la altura del fluido en él es de 5 m. Treinta minutos después, este gasto es interrumpido por falla de la bomba, que se repara y arranca una hora después. Determine el gasto necesario para que el nivel se recupere y se mantenga en 5 m, así como el tiempo necesario para alcanzar ese nivel (régimen permanente). El flujo de salida es  $3.457 \sqrt{a} \text{ L/s}$  ininterrumpidamente.
- 7.5 Un tanque perfectamente agitado contiene 400 litros de salmuera en la cual están disueltos 10 kg de sal. Si se hace llegar 1.0 L/min de una salmuera que contiene 3 kg de sal en cada 5 litros y por el fondo se sacan 8 L/min de una salmuera, determine la concentración de sal en el tanque a distintos tiempos.



**Figura 7.32**  
Vaciado de un tanque cónico.

- 7.6 Se ha encontrado experimentalmente que la constante de velocidad de reacción a volumen constante y a 30 °C de la ecuación estequiométrica



es  $0.4967 \text{ (mol/L)}^{-1} \text{ min}^{-1}$ . Determine el tiempo necesario para alcanzar 90% de conversión del reactivo limitante en cada uno de los casos que se dan abajo, si se mantiene todo el tiempo la mezcla reaccionante a 30 °C.

| Concentraciones<br>(mol / L) |          |
|------------------------------|----------|
| $C_{A0}$                     | $C_{B0}$ |
| 0.5                          | 1.0      |
| 1.0                          | 1.5      |
| 1.5                          | 2.0      |
| 1.0                          | 1.0      |
| 2.0                          | 0.5      |

- 7.7 La aplicación de las leyes de Kirchhoff en un circuito cerrado da lugar a sistemas de ecuaciones diferenciales del tipo

$$\frac{dI_1}{dt} = -4 I_1 + 3I_2 + 6$$

$$\frac{dI_2}{dt} = -2.4 I_1 + 1.6 I_2 + 3.6$$

Si se tienen las condiciones iniciales

$$I_1(0) = 0, I_2(0) = 0$$

Calcule  $I_1(3)$  e  $I_2(3)$  con pasos de tiempo 0.05, 0.1, 0.5 y 1.0

- 7.8 Un capacitor de 0.001 farads está conectado en serie con una fem de 20 volts y una inductancia de 0.4 henries. Si  $t = 0$ ,  $Q = 0$ , e  $I = 0$ , encuentre una ecuación para modelar este circuito y use el método de Runge-Kutta de tercer orden para hallar el valor de  $Q$  a distintos tiempos (véase ejercicio 7.5).
- 7.9 Repita el ejercicio 7.8 para  $k = 0.0002$ . ¿Qué sucede si  $k \rightarrow 0$ ?
- 7.10 Un objeto que pesa 500 kg se coloca en la superficie de un tanque lleno de agua y se suelta ( $v_0 = 0$ ). Las fuerzas que actúan sobre el objeto son la de empuje hacia arriba de 100 kg y la resistencia del agua que es de  $30 v$ , donde  $v$  está en m/s. ¿Qué distancia recorre el cuerpo en 5 segundos?
- 7.11 Las ecuaciones

$$\frac{d^2x}{dt^2} = 0$$

y

$$\frac{d^2y}{dt^2} = -g$$

$$\text{con } x = 0, y = 0, \frac{dx}{dt} = v_0 \cos \theta_0, \frac{dy}{dt} = v_0 \sin \theta_0 \text{ at } t = 0$$

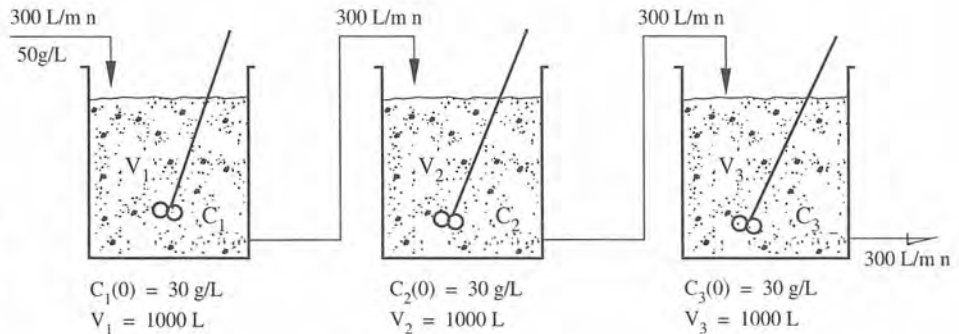
describen la trayectoria de un proyectil disparado con una velocidad inicial  $v_0$  y un ángulo de inclinación  $\theta_0$ . Aquí  $x$  y  $y$  son las distancias horizontal y vertical que recorre el proyectil.

Si  $\theta_0 = 60$  y  $v_0 = 50$  m/s, calcule

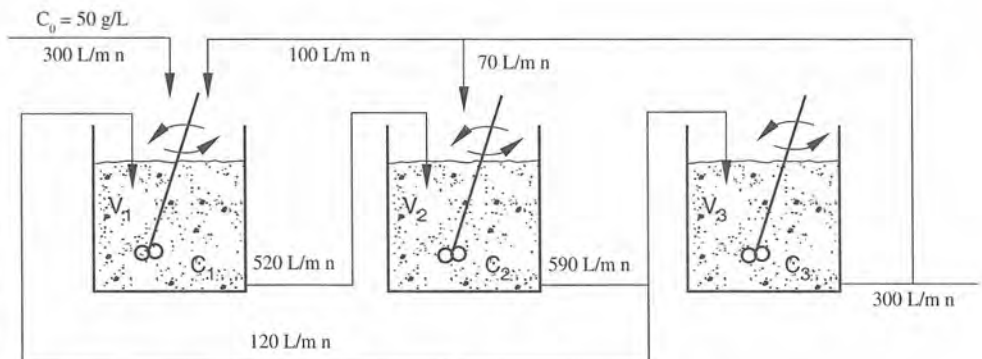
- El tiempo de vuelo del proyectil
- La distancia que recorre
- La altura máxima que alcanza

- 7.12 Si en el ejercicio 7.7 se cambian las condiciones de concentración en los tanques a  $C_2(0) = C_1(0) = C_3(0) = 0$ . ¿Cuál es el tiempo necesario para alcanzar el régimen permanente?
- 7.13 Si en el diagrama de la figura 7.33 se toma una corriente de recirculación de 150 L/min a la salida del tanque 3 y se lleva al tanque 2, en tanto el volumen se conserva constante en cada tanque e igual a 1000 litros, determine la concentración en cada tanque 10 minutos después de iniciado el proceso.
- 7.14 Si en el diagrama del problema anterior se adicionan las corrientes de recirculación mostradas en la figura 7.34, pero conservando la característica de que el volumen en los tres tanques permanece constante, las concentraciones  $C_1$ ,  $C_2$  y  $C_3$  variarán de manera distinta a como se vio en el ejercicio 7.7.

Con los datos de la figura 7.34 determine las concentraciones en cada uno de los tanques a los diez minutos de haber empezado a agregar solución al primero.



**Figura 7.33**  
Arreglo de tres tanques interconectados.



**Figura 7.34**  
Arreglo de tres tanques interconectados con recirculación.



**7.15** Repita el ejercicio 7.8 con lo siguientes cambios

- a)  $V_1 = 80, V_2 = 20, F_R = 10$
- b)  $V_1 = 80, V_2 = 20, F_R = 0.1$
- c)  $V_1 = 50, V_2 = 50, F_R = 10$
- d)  $V_1 = 20, V_2 = 80, F_R = 10$
- e)  $V_1 = 50, V_2 = 50, F_R = 200$

**7.16** Si en el ejercicio 7.9 la reacción es de segundo orden, calcule la temperatura  $T$  y la concentración  $C_A$  de la corriente de salida cuando el reactor trabaja en régimen transitorio y hasta alcanzar el régimen permanente. Utilice

$$k = 1 \times 10^{13} \exp \left( \frac{-23200}{1.987 T} \right) \frac{\text{L}}{\text{gmol s}}$$

y la información presentada en el ejercicio 7.9

**7.17** Repita el ejercicio 7.9 utilizando la misma información, con los siguientes cambios

$$T_j = 310, 320, 340 \text{ y } 350$$

Analice los resultados.

**7.18** El término  $EI$  del ejercicio 7.10 depende del material de que está construida la viga. Repita el ejercicio para otros materiales, en los que

- a)  $EI = 117187$
- b)  $EI = 100000$

las demás condiciones se conservan.

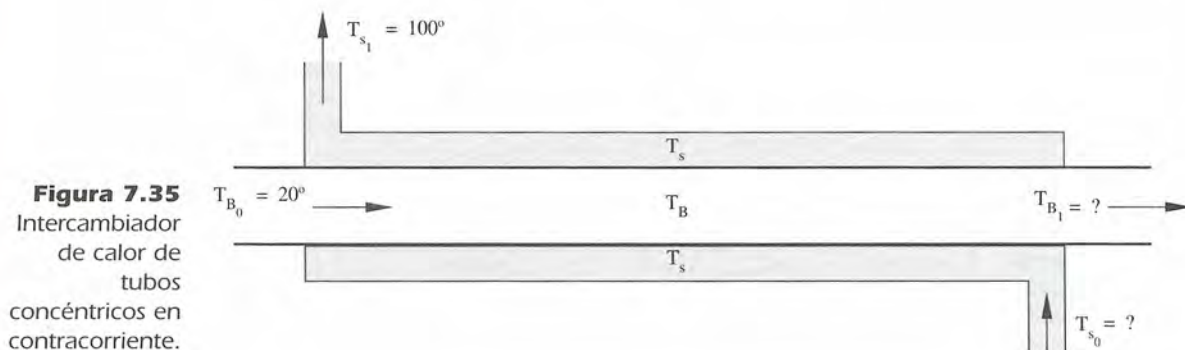
**7.19** Si en la viga del ejercicio 7.10 se aplica además una carga concentrada de 500 kg en el extremo libre, determine el perfil de flexión a lo largo de la viga.

**7.20** Se tiene un intercambiador de calor de tubos concéntricos en contracorriente y sin cambio de fase (véase Fig. 7.35). Las ecuaciones que describen el intercambiador de calor en ciertas condiciones de operación son:

$$\frac{dT_B}{dx} = 0.03 (T_s - T_B)$$

$$\frac{dT_s}{dx} = 0.04 (T_s - T_B)$$

Elabore un programa para calcular  $T_{B1}$  y  $T_{s0}$  si el intercambiador de calor tiene una longitud de 3 m; use el método de Runge-Kutta de cuarto orden.



**Figura 7.35**  
Intercambiador  
de calor de  
tubos  
concéntricos en  
contracorriente.



- 7.21 Un tanque cilíndrico de 5 m de diámetro y 11 m de largo aislado con asbesto se carga con un líquido que está a 220 °F y el cual se deja reposar durante cinco días. A partir de los datos de diseño del tanque, las propiedades térmicas y físicas del líquido, y el valor de la temperatura ambiente, se encuentra la ecuación

$$\frac{dT}{dt} = 0.615 + 0.175 \cos\left(\frac{\pi t}{12}\right) - 0.0114 T$$

que relaciona la temperatura  $T$  del líquido (en °C) con el tiempo  $t$  en horas. ¿Cuál es la temperatura final del líquido?

- 7.22 El radio se desintegra en razón proporcional a la cantidad presente en cada instante. La constante de proporcionalidad es  $k = 10^{-2}$  día<sup>-1</sup>. Si se tienen inicialmente 60 g de radio, calcule la cantidad que hay presente transcurridos cinco días mediante el siguiente esquema de predicción-corrección

$$\bar{y}_{i+1} = y_i + \frac{h}{24} (55f_i - 59f_{i-1} - 37f_{i-2} - 9f_{i-3})$$

$$y_{i+1} = y_i + \frac{h}{24} (9f_{i+1} + 19f_i - 5f_{i-1} + f_{i-2})$$

- 7.23 Considere un sistema ecológico simple compuesto solamente de coyotes ( $y$ ) y correacaminos ( $x$ ), donde los primeros se alimentan de los segundos (cuando los alcanzan). Los tamaños de las poblaciones cambian de acuerdo con las ecuaciones

$$\frac{dx}{dt} = k_1 x - k_2 xy$$

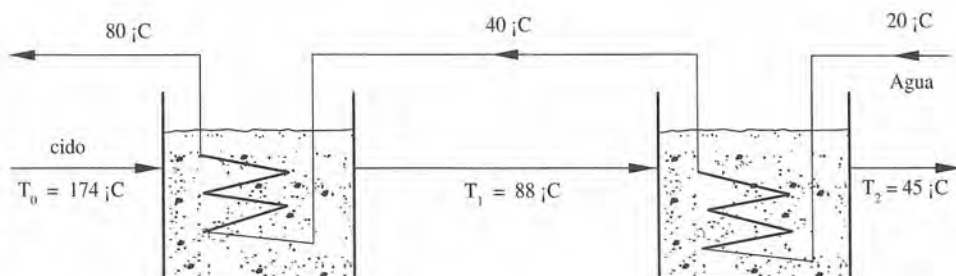
$$\frac{dy}{dt} = k_3 xy - k_4 y$$

que se pueden entender como sigue:

Si no hay coyotes ( $y$ ) los correacaminos se reproducen con una velocidad de crecimiento  $k_1 x$ ; si no hay correacaminos, la especie de coyotes desaparece con velocidad  $k_4 y$ . El término  $xy$  representa la interacción de las dos especies y las constantes  $k_2$  y  $k_3$  dependen de la habilidad de los depredadores para atrapar a los correacaminos y de la habilidad de éstos para huir.

Las poblaciones de los coyotes cambian cíclicamente. Calcule el ciclo al resolver el modelo con  $k_1 = 0.4$ ,  $k_2 = 0.02$ ,  $k_3 = 0.001$  y  $k_4 = 0.3$  Use  $x(0) = 30$  y  $y(0) = 3$  como condiciones iniciales.

- 7.24 Se utilizan dos tanques en serie y provistos de serpentín de enfriamiento por el cual circula el agua en contracorriente para enfriar 10000 lb/hr de ácido sulfúrico. Las condiciones de operación se muestran en la figura 7.36. Si en un momento dado fallara el suministro de agua de enfriamiento, ¿cuál será la temperatura del ácido sulfúrico  $T_2$  a la salida del segundo tanque después de una hora? Las ecuaciones que describen el proceso son



**Figura 7.36**  
Dos tanques  
interconectados  
y con serpentín  
de enfriamiento.

$$3600 T_0 - 3600 T_1 = 2850 \frac{dT_1}{dt}$$

$$3600 T_1 - 3600 T_2 = 2850 \frac{dT_2}{dt}$$

donde  $T_0$ ,  $T_1$  y  $T_2$  están en °C y  $t$  en horas.

- 7.25** Utilice el métodos de Taylor (elija el orden) para resolver los siguientes problemas de valor inicial (PVI) y compare con las soluciones analíticas

- a)  $dy/dx = 3x^2$ ,  $y(0) = 0$ ,  $y(1) = ?$  con  $h = 0.1$   
 b)  $dy/dx = \ln x$ ,  $y(1) = 3$ ,  $y(2) = ?$  con  $h = 0.2$   
 c)  $dy/dx = 2xy$ ,  $y(1) = 0.5$ ,  $y(2) = ?$  con  $h = 0.25$   
 d)  $dy/dx = y^2$ ,  $y(0) = 1$ ,  $y(0.5) = ?$  con  $h = 0.1$

- 7.26** Resuelva los PVI del problema anterior por el métodos de Runge-Kutta de segundo orden.

- 7.27** Resuelva los PVI del problema 7.25 por el método de Runge-Kutta de cuarto orden.

- 7.28** Resuelva los siguientes PVI con la fórmula

$$y_{i+1} = y_i + \frac{h}{24} [55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3}]$$

Y el métodos de Runge-Kutta de tercer orden como inicializador

- a)  $dy/dx + y = 0$ ,  $y(0) = 1$ ,  $y(2) = ?$ , con  $h = 0.25$   
 b)  $dy/dx + 2xy = 2x^3$ ,  $y(0) = 0$ ,  $y(2.5) = ?$ , con  $h = 0.5$   
 c)  $dy/dx + xy = g(x)$ ,  $y(0) = 1$ ,  $y(1) = ?$ , con  $h = 0.25$

donde:

| $x$    | 0.0 | 0.2     | 0.4     | 0.6     | 0.8     | 1.0     |
|--------|-----|---------|---------|---------|---------|---------|
| $g(x)$ | 0.0 | 0.19471 | 0.35868 | 0.46602 | 0.49979 | 0.45465 |

- d)  $dy/dx = -y - xy^2$ ,  $y(0) = 1$ ,  $y(-1) = ?$ , con  $h = -0.25$   
 e)  $xdy/dx = 1 - y + x^2y^2$ ,  $y(1) = 1$ ,  $y(1.5) = ?$ , con  $h = 0.125$

- 7.29** Resuelva los PVI del problema anterior con la fórmula

$$y_{i+1} = y_{i-5} + \frac{3h}{10} [11f_i - 14f_{i-1} + 26f_{i-2} - 14f_{i-3} + 11f_{i-4}]$$

con el método de Runge-Kutta de cuarto orden como inicializador.

- 7.30** Resuelva los PVI del problema 7.28 con los siguientes esquemas de solución

- a) Inicialización de Runge-Kutta de tercer orden.  
 Predicción con la fórmula dada en el ejercicio 7.28  
 Corrección con

$$y_{i+1} = y_i + \frac{h}{24} [9f_{i+1} - 19f_i + 5f_{i-1} - f_{i-2}]$$

- b) Inicialización de Runge-Kutta de cuarto orden.  
 Predicción con la fórmula del problema 7.29  
 Corrección con

$$y_{i+1} = y_{i-3} + \frac{2h}{45} [7f_{i+1} - 32f_i + 12f_{i-1} + 32f_{i-2} + 7f_{i-3}]$$

c) Inicialización con Runge-Kutta de cuarto orden.

Predicción con Adams-Bashford de cuarto orden

$$y_{i+1} = y_i + \frac{h}{720} [1901 f_i - 2774 f_{i-1} + 2616 f_{i-2} - 1274 f_{i-3} + 251 f_{i-4}]$$

Corrección con Adams-Moulton de cuarto orden

$$y_{i+1} = y_i + \frac{h}{720} [251 f_{i+1} - 646 f_i - 264 f_{i-1} + 106 f_{i-2} - 19 f_{i-3}]$$

**7.31** Resuelva el siguiente PVI con el método de Runge-Kutta de segundo orden

$$\begin{aligned} dy/dx &= z, & y(0) &= 1, & y(1) &= ? \\ dz/dx &= y, & z(0) &= -1, & z(1) &= ? \text{ con } h = 0.1 \end{aligned}$$

**7.32** Resuelva el siguiente PVI con el método de Runge-Kutta de cuarto orden con  $h = 0.1$

$$\text{PVI} \left\{ \begin{aligned} \frac{d^2y}{dx^2} + 2 \frac{dy}{dx} + 2y &= 0 \\ y(0) &= 1 \\ \frac{dy}{dx} \Big|_{x=0} &= -1 \\ y(1) &= ? \end{aligned} \right.$$

**7.33** Resuelva el PVI del problema 7.31 con el esquema de solución (a) del 7.30.

**7.34** Resuelva el PVI del problema 7.32 con el esquema de solución (c) del 7.30.

**7.35** Resuelva el siguiente PVI

$$\left\{ \begin{aligned} \frac{dy}{dx} &= z \\ \frac{dz}{dx} &= -125y - 20z \\ y(0) &= 0 \quad y(1) = ? \\ z(0) &= 1 \quad z(1) = ? \end{aligned} \right.$$

con el método de Runge-Kutta de cuarto orden usando

a)  $h = 0.5$

b)  $h = 0.1$

Compare los resultados con la solución analítica

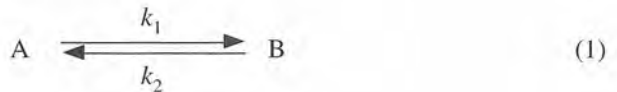
$$\begin{aligned} y &= \frac{1}{5} e^{-10x} \sin 5x \\ z &= e^{-10x} (\cos 5x - 2 \sin 5x) \end{aligned}$$

**7.36** Escriba las siguientes ecuaciones diferenciales como un sistema de ecuaciones diferenciales ordinarias de primer orden. Pase las condiciones iniciales a términos de las nuevas variables para construir un PVIG y resuélvalo con los métodos vistos usando los tamaños de paso sugeridos:



- a)  $y'' - 2y' + 2y = e^{2t} \sin t$   $0 \leq t \leq 1$   
 $y(0) = -0.4; y'(0) = -0.6$   $h = 0.01$
- b)  $y'' + 2ty' = e^t$   $0 \leq t \leq 2$   
 $y(0) = 1; y'(0) = -1$   $h = 0.1$
- c)  $y''' - 2y'' - y' - 2y = e^t$   $0 \leq t \leq 3$   
 $y(0) = 1; y'(0) = 2; y''(0) = 0$   $h = 0.2$

7.37 Considere el conjunto de reacciones reversibles



Asuma que hay una mol de A solamente al inicio, y tome  $N_A$ ,  $N_B$  y  $N_C$  como las mols de A, B, y C presentes, respectivamente.

Como la reacción se verifica a volumen constante,  $N_A$ ,  $N_B$  y  $N_C$  son proporcionales a las concentraciones. Sean  $k_1$  y  $k_2$  las constantes de velocidad de reacción a derecha e izquierda, respectivamente, de la ecuación 1; igualmente sean  $k_3$  y  $k_4$  aplicables a la 2.

La velocidad de desaparición neta de A está dada por:

$$\frac{dN_A}{dt} = -k_1 N_A + k_2 N_B$$

y para B

$$\frac{dN_B}{dt} = -(k_2 + k_3) N_B + k_1 N_A + k_4 N_C$$

Determine  $N_A$ ,  $N_B$  y  $N_C$  transcurridos 50 minutos del inicio de las reacciones mediante

$$\begin{aligned} k_1 &= 0.1 \text{ min}^{-1} \\ k_2 &= 0.01 \text{ min}^{-1} \\ k_3 &= 0.09 \text{ min}^{-1} \\ k_4 &= 0.009 \text{ min}^{-1} \end{aligned}$$



7.38 El PROGRAMA 7.4 del CD (ver ejercicio 7.11) puede servir de laboratorio para experimentar y analizar resultados. Algunas sugerencias son:

- Modificar la velocidad inicial (un valor negativo significaría un choque por atrás del carrito).
- Al aumentar la masa  $m$  se incrementaría su energía cinética ( $0.5 mv^2$ ), por lo que sería interesante ver cómo funcionan los resortes al cambiar los valores de  $m$ .
- Al hacer cero alguno de los parámetros  $c$ ,  $k$  o  $k'$  estaríamos eliminando alguno de los resortes o el amortiguador
- Procure hacer sus análisis usando los valores de la tabla, la gráfica y el movimiento del carrito.

7.39 Modifique el ejercicio 7.12 con  $q = 2$  y determine si se cumple la condición de que

$$\max y(x) < 0.00131 \text{ dada por los reglamentos de construcción.}$$

$$0 < x < L$$



- 7.40** La deflexión de una placa de acero rectangular larga ( $L \gg \text{Ancho}$ ), con carga uniformemente distribuida, sujeta a una fuerza de tensión axial, está gobernada para pequeñas deflexiones por el PVF

$$\left\{ \begin{array}{l} \frac{d^2 y}{dx^2} - \frac{S}{D} y = -\frac{qL}{2D} x + \frac{q}{2D} x^2 \\ y(0) = 0 \\ y(L) = 0 \\ y(x) = ?, \quad \text{para } 0 < x < L \end{array} \right.$$

Determine la deflexión de la placa con los siguientes valores de los parámetros:  $q = 15 \text{ kg/cm}^2$ ,  $S = 18 \text{ kg/cm}$ ,  $D = 1.02 \times 10^8 \text{ kg/cm}$  y  $L = 130 \text{ cm}$ . Construya una tabla y una gráfica con los resultados obtenidos y analícelos.

- 7.41** Modifique el guión de Matlab del ejercicio 7.12 para realizar una interpolación cuadrática inversa en lugar de la interpolación lineal inversa.
- 7.42** Resuelva el problema no lineal de valores en la frontera modificando el guión del ejercicio 7.12. Use el guión del problema 7.41 y compare el número de iteraciones.

$$\left\{ \begin{array}{l} y \frac{d^2 y}{dx^2} + \left( \frac{dy}{dx} \right)^2 = -1 \\ y(0) = 2 \\ y(3) = 5 \\ y(x) = ?, \quad \text{para } 0 < x < 3 \end{array} \right.$$

# ECUACIONES DIFERENCIALES PARCIALES

## A dónde nos dirigimos

En este capítulo se presentará una breve introducción a algunas de las técnicas para aproximar la solución de ecuaciones diferenciales parciales lineales de segundo orden y con dos variables independientes. Para esto se parte de la modelación de fenómenos físicos como la conducción de calor en una barra aislada y la vibración de una cuerda elástica flexible. Una vez que se tiene formulado el problema, se discretiza el dominio de definición de la función involucrada en las ecuaciones, es decir, se forma una malla, y se aproximan las derivadas de la ecuación diferencial con diferencias hacia delante, hacia atrás o centrales en cada nodo de la malla, generándose con esto, de acuerdo al tipo de diferencias, diversos métodos con distintas características, como por ejemplo estabilidad y convergencia. Tales aproximaciones generan problemas algebraicos como sistemas de ecuaciones lineales y no lineales que habrá que resolver con las técnicas vistas en los capítulos 2 y 3.

Una vez resuelto el problema algebraico se enfatiza en el análisis de resultados y en su interpretación física, a fin de darle significado y de tener presente el fenómeno original en todo el proceso de solución.

Ya que probablemente este sea el primer encuentro del lector con ecuaciones diferenciales parciales, se ha cuidado la claridad y sencillez del aspecto expositivo de modo de darle elementos sólidos para continuar este estudio.

## Introducción

Las ecuaciones diferenciales parciales (EDP) involucran una función de más de una variable independiente y sus derivadas parciales. La importancia de este tema radica en que prácticamente en todos los fenómenos que se estudian en ingeniería y otras ciencias, aparecen más de dos variables,\* y su modelación matemática conduce frecuentemente a EDP.

Primero se clasificarán las ecuaciones diferenciales parciales lineales atendiendo al siguiente modelo general

$$A(x, y) \frac{\partial^2 U}{\partial x^2} + B(x, y) \frac{\partial^2 U}{\partial x \partial y} + C(x, y) \frac{\partial^2 U}{\partial y^2} = F(x, y, \frac{\partial U}{\partial x}, \frac{\partial U}{\partial y}) \quad (8.1)$$

en el cual se asume que  $A(x, y)$ ,  $B(x, y)$  y  $C(x, y)$  son funciones continuas de  $x$  y  $y$ . Dependiendo de los valores de  $A(x, y)$ ,  $B(x, y)$  y  $C(x, y)$  en algún punto particular  $(x, y) = (a, b)$ , la ecuación (8.1) puede ser elíptica, parabólica o hiperbólica, de acuerdo con las condiciones

$$\begin{array}{ll} B^2(a, b) - 4 A(a, b) C(a, b) < 0 & \text{Elíptica en } (a, b) \\ B^2(a, b) - 4 A(a, b) C(a, b) = 0 & \text{Parabólica en } (a, b) \\ B^2(a, b) - 4 A(a, b) C(a, b) > 0 & \text{Hiperbólica en } (a, b) \end{array} \quad (8.2)$$

\* En el análisis del comportamiento de los gases, por ejemplo, se tiene temperatura, presión y volumen; en la transmisión de calor intervienen temperatura, tiempo y direcciones del espacio:  $x, y, z$ ; etcétera.

Una misma EDP puede ser parabólica en un punto, e hiperbólica en otro, etc. Si en cambio  $A(x, y)$ ,  $B(x, y)$  y  $C(x, y)$  son constantes, entonces es elíptica, parabólica o hiperbólica completamente (ver ejercicio 8.1).

Algunos ejemplos de estas ecuaciones son

$$\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} = 0$$

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} = f(x, y)$$

$$\frac{\partial^2 y}{\partial x^2} = \alpha \frac{\partial^2 y}{\partial t^2}$$

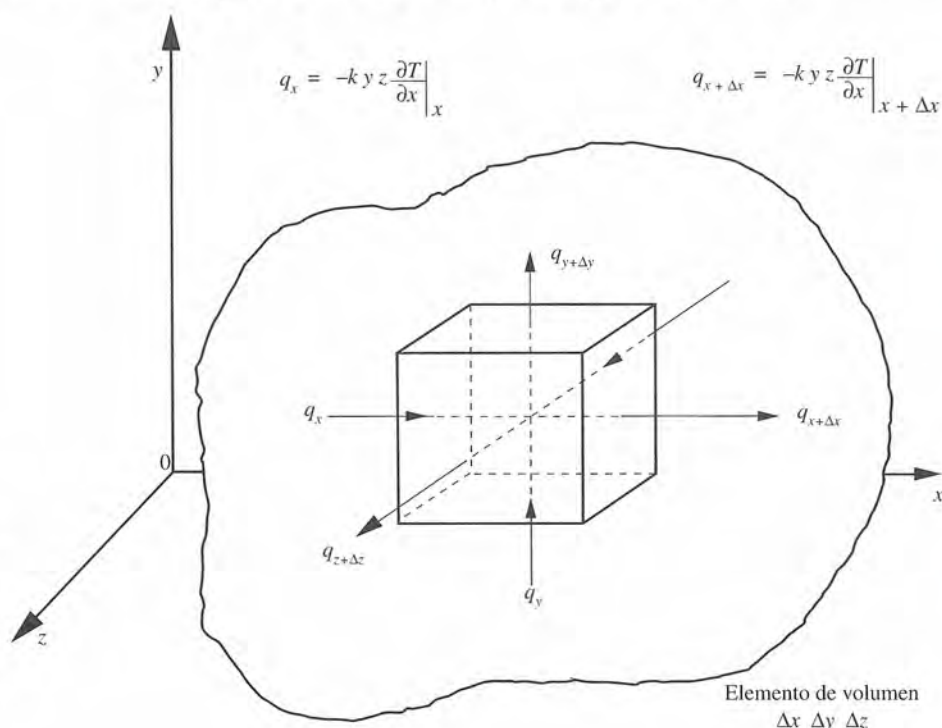
## 8.1 Obtención de algunas ecuaciones diferenciales parciales a partir de la modelación de fenómenos físicos (ecuación de calor y ecuación de onda)

A continuación se presenta la derivación de dos de las ecuaciones en estudio más comunes.

### a) Ecuación general de la conducción de calor

Supóngase un cuerpo sólido de conductividad térmica  $k$ , peso específico  $\rho$  y calor específico  $C_p$  independientes de la temperatura  $T$ , en el cual fluye calor en las tres dimensiones del espacio y puede generar o absorber calor debido a algún fenómeno, por ejemplo de reacción química.

Al efectuar un balance de calor en un elemento diferencial como el de la figura 8.1 de dimensiones  $\Delta x$ ,  $\Delta y$  y  $\Delta z$  se tiene, de acuerdo con la ley de la continuidad



**Figura 8.1**  
Balance de calor en un elemento diferencial de dimensiones  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$ .



$$\begin{array}{ccccccc}
 \text{Acumulación de} & = & \text{calor que entra al} & - & \text{calor que sale del} & + & Q\Delta x\Delta y\Delta z \\
 \text{calor} & & \text{elemento diferencial} & & \text{elemento diferencial} & & \\
 & & \text{en cada una de las} & & \text{en cada una de las} & & \\
 & & \text{tres dimensiones} & & \text{tres dimensiones} & & \\
 (\text{cal/s}) & & (\text{cal/s}) & & (\text{cal/s}) & & (\text{cal/s})
 \end{array} \quad (8.3)$$

donde  $Q$  puede ser positivo o negativo dependiendo de si el calor es generado o absorbido por unidad de volumen y por unidad de tiempo en el elemento diferencial.

También en la figura 8.1 se realiza un esquema de la entrada  $q_i$  y la salida  $q_i + \Delta_i$  de los flujos de calor (en cal/s) representados por la ley de Fourier, o sea que son proporcionales a la conductividad térmica  $k$ , el área de transmisión y el gradiente de temperatura en dirección de la transmisión. El signo negativo es para obtener flujos de calor positivos (por convención), ya que los gradientes  $dT/dx$ ,  $dT/dy$  y  $dT/dz$  son negativos.

Los flujos de calor se sustituyen en la ecuación (8.3) y se tiene:

$$\begin{aligned}
 \frac{d}{dt}(\Delta x\Delta y\Delta z\rho C_p T) = & -k\Delta y\Delta z \left. \frac{dT}{dx} \right|_x - (-k\Delta y\Delta z \left. \frac{dT}{dx} \right|_{x+\Delta x}) + \\
 & -k\Delta x\Delta z \left. \frac{dT}{dy} \right|_y - (-k\Delta x\Delta z \left. \frac{dT}{dy} \right|_{y+\Delta y}) + \\
 & -k\Delta x\Delta y \left. \frac{dT}{dz} \right|_z - (-k\Delta x\Delta y \left. \frac{dT}{dz} \right|_{z+\Delta z}) + Q\Delta x\Delta y\Delta z
 \end{aligned} \quad (8.4)$$

al dividir miembro a miembro entre  $\Delta x$ ,  $\Delta y$ ,  $\Delta z$  y hacerlos muy pequeños, o sea  $\Delta x, \Delta y, \Delta z \rightarrow 0$ , queda

$$\begin{aligned}
 \rho C_p \frac{\partial T}{\partial t} = & k \lim_{\Delta x \rightarrow 0} \left[ \frac{\left. \frac{dT}{dx} \right|_{x+\Delta x} - \left. \frac{dT}{dx} \right|_x}{\Delta x} \right] + k \lim_{\Delta y \rightarrow 0} \left[ \frac{\left. \frac{dT}{dy} \right|_{y+\Delta y} - \left. \frac{dT}{dy} \right|_y}{\Delta y} \right] \\
 & + k \lim_{\Delta z \rightarrow 0} \left[ \frac{\left. \frac{dT}{dz} \right|_{z+\Delta z} - \left. \frac{dT}{dz} \right|_z}{\Delta z} \right] + Q
 \end{aligned} \quad (8.5)$$

y al aplicar la definición de derivada se obtiene

$$\frac{\partial T}{\partial t} = \alpha \left[ \frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2} \right] + \frac{Q}{\rho C_p} \quad (8.6)$$

donde se ha sustituido a  $\frac{k}{\rho C_p}$  con  $\alpha$ , la cual se llama **coeficiente de difusividad térmica**, y sus unidades son, por ejemplo,

$$\text{m}^2/\text{s} \text{ ya que } k [ = ] \text{ cal}/(\text{s m } ^\circ\text{C}), C_p [ = ] \text{ cal}/(\text{g } ^\circ\text{C}) \text{ y } \rho [ = ] \text{ g}/\text{cm}^3.$$

La ecuación 8.6 se conoce como **ecuación de conducción de calor en régimen transitorio** en tres dimensiones (cartesianas), y es muy empleada en el campo de la ingeniería. También se conoce como **ecuación de difusión**, ya que representa la difusión molecular de masa entre fluidos, cuando la variable dependiente es la concentración  $C$  y el coeficiente  $\mathcal{D}$  representa la **difusividad**  $\mathcal{D}$ . Así:

$$\frac{\partial C}{\partial t} = \mathcal{D} \left[ \frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} + \frac{\partial^2 C}{\partial z^2} \right]$$



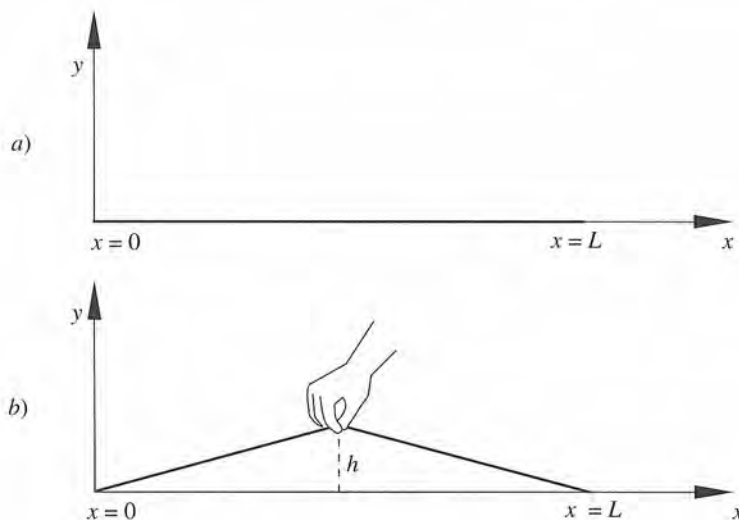


Figura 8.2

donde las unidades pueden ser:

$$C [=] \text{ moles/cm}^3, \mathcal{D} [=] \text{ cm}^2/\text{s}, t [=] \text{ s}$$

### b) Ecuación de onda en una dimensión

Considérese una cuerda (como la de una guitarra) elástica y flexible, la cual se estira y se sujeta en dos puntos fijos en  $x=0$  y  $x=L$  sobre el eje de las  $x$  (Fig. 8.2a). A un tiempo  $t=0$ , la cuerda se toma del centro y se eleva verticalmente a una altura  $y=h$  (véase Fig. 8.2b). Después se suelta. La descripción del movimiento producido constituye el problema por resolver.

Para simplificarlo, se considera que  $h$  es pequeño en comparación con  $L$  ( $h \ll L$ ).

### MODELO

Si en un instante se tomara una fotografía de la cuerda vibrando, se tendría ésta como en la figura 8.3a. El desplazamiento de un punto  $x$  de la cuerda en el tiempo  $t$  queda indicado por  $y(x, t)$ , de igual forma para un punto vecino  $x + \Delta x$  y en el mismo tiempo  $t$ , su desplazamiento queda indicado por  $y(x + \Delta x, t)$ .

En la misma figura 8.3b se muestra una ampliación del segundo segmento de cuerda  $\Delta S$ , la cual está sometido a dos tensiones que siempre actúan en la dirección de la tangente a  $\Delta S$  a izquierda y derecha de la cuerda, o sea  $T(x, t)$  y  $T(x + \Delta x, t)$ , respectivamente. Hay que observar que la tensión es función de la posición  $x$  sobre la cuerda y el tiempo  $t$ .

Al hacer una composición de fuerzas sobre el elemento de cuerda  $\Delta S$  en las direcciones vertical y horizontal se tiene:

$$\text{Fuerza vertical neta} = T(x + \Delta x, t) \sin \theta_2 - T(x, t) \sin \theta_1$$

$$\text{Fuerza horizontal neta} = T(x + \Delta x, t) \cos \theta_2 - T(x, t) \cos \theta_1$$

La fuerza horizontal neta es cero si se considera que el desplazamiento del punto  $x$  de su posición de equilibrio a la posición  $y(x, t)$  es vertical. Por otro lado, la fuerza neta vertical sobre  $\Delta S$  produce una aceleración definida por la segunda ley de Newton; o sea,

$$\text{Fuerza vertical neta} = T(x + \Delta x, t) \sin \theta_2 - T(x, t) \sin \theta_1$$

$$= \rho \Delta S \frac{\partial^2 y}{\partial t^2} \quad (8.7)$$

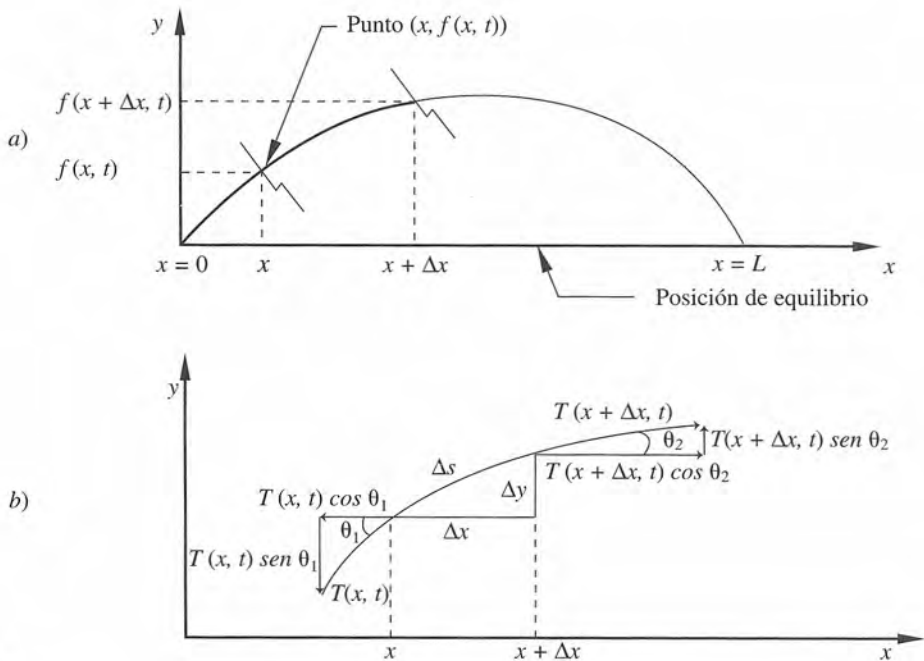


Figura 8.3

donde  $\rho$  es la densidad de la cuerda en unidades de masa/longitud y  $\partial^2 y / \partial t^2$  la aceleración de  $\Delta S$ .

Como  $\theta$  es función de la posición  $x$  y el tiempo  $t$ ,  $\theta_1 = \theta(x, t)$  y  $\theta_2 = \theta(x + \Delta x, t)$ .

Estas expresiones se sustituyen en la ecuación 8.7 y al dividir entre  $\Delta x$  queda

$$\frac{T(x + \Delta x, t) \sin \theta(x + \Delta x, t) - T(x, t) \sin \theta(x, t)}{\Delta x} = \rho \frac{\Delta S}{\Delta x} \frac{\partial^2 y}{\partial t^2}$$

Para vibraciones cortas  $\theta$  es pequeño, por lo que  $\Delta S / \Delta x \approx 1$  y  $\sin \theta \approx \text{tg } \theta$ ; de modo que la última ecuación se puede escribir:

$$\frac{T(x + \Delta x, t) \text{tg } \theta(x + \Delta x, t) - T(x, t) \text{tg } \theta(x, t)}{\Delta x} \approx \rho \frac{\partial^2 y}{\partial t^2}$$

y haciendo  $\Delta x \rightarrow 0$

$$\frac{\partial}{\partial x} [T(x, t) \text{tg } \theta(x, t)] = \rho \frac{\partial^2 y}{\partial t^2}$$

Como  $\text{tg } \theta(x, t) = \partial y / \partial x$  y si la tensión  $T$  es constante se obtiene:

$$c^2 \frac{\partial^2 y}{\partial x^2} = \frac{\partial^2 y}{\partial t^2} \quad (8.8)$$

donde  $c^2 = \frac{T}{\rho}$

Puesto que la cuerda permanece sujeta en sus extremos  $x = 0$  y  $x = L$ , el desplazamiento  $y(x, t)$  satisface las condiciones siguientes en todo el proceso

$$\begin{aligned} y(0, t) &= 0 \\ y(L, t) &= 0 \end{aligned} \quad \text{para } t \geq 0 \quad (8.9)$$

conocidas como **condiciones extremas o frontera (CF)**.

Por otro lado, la posición de la cuerda en el momento de soltarse (Fig. 8.2b) puede darse matemáticamente así:

$$y(x, 0) = f(x) \quad (8.10)$$

ecuación que se conoce como **condición inicial** (CI) por describir la condición que se tiene al inicio del proceso.

En resumen, la ecuación 8.8 y las condiciones inicial y de frontera (Ecs. 8.10 y 8.9, respectivamente) constituyen un modelo matemático denotado como **problema de valores en la frontera\*** (PVF)

|     |                                                                             |                                |
|-----|-----------------------------------------------------------------------------|--------------------------------|
|     | $\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}$ | (ecuación diferencial parcial) |
| PVF | $y(x, 0) = f(x), \quad 0 < x < L$                                           | (condición inicial)            |
|     | $y(0, t) = 0, \quad t > 0$                                                  | (condición frontera 1)         |
|     | $y(L, t) = 0, \quad t > 0$                                                  | (condición frontera 2)         |

y cuya solución  $y(x, t)$  describe la posición de cualquier punto de la cuerda en un tiempo  $t$ .

## 8.2 Aproximación de derivadas por diferencias finitas

### GENERALIDADES

La expansión de una función  $f(x)$  diferenciable en una serie de Taylor alrededor de un punto  $x_i$  se estudió en los capítulos 5 y 7 y está definida por

$$f(x_i + a) = f(x_i) + a f'(x_i) + \frac{a^2}{2!} f''(x_i) + \frac{a^3}{3!} f'''(x_i) + \dots \quad (8.11)$$

Esta vez, la utilidad de la serie de Taylor no será estimar el valor de la función  $f(x)$  en el punto  $x_i + a$ , sino aproximar la derivada de la función en  $x_i$  a partir de los valores de la función en  $x_i$  y en  $x_i + a$ . Para ello, considérese que  $a > 0$  (con esto la ecuación 8.11 sólo es válida delante del punto  $x_i$ ) y que  $a$  es tan pequeña ( $a \ll 1$ ) como para despreciar los términos tercero, cuarto, etc., del lado derecho de la expansión, con lo que la derivada  $f'(x_i)$  puede aproximarse así

$$\left. \frac{df}{dx} \right|_{x_i} = f'(x_i) \approx \frac{f(x_i + a) - f(x_i)}{a} \quad (8.12)$$

Esta ecuación quedó definida en el capítulo 5 como la aproximación de la primera derivada de  $f(x)$  en  $x_i$  con **diferencias hacia delante**.

Un resultado similar, válido a la izquierda de  $x_i$  se obtendrá restando  $a$  de  $x_i$  en la ecuación 8.11; esto es,

$$f(x_i - a) = f(x_i) - a f'(x_i) + \frac{a^2}{2!} f''(x_i) - \frac{a^3}{3!} f'''(x_i) + \dots \quad (8.13)$$

y como  $a \ll 1$ , puede llegarse a

$$\left. \frac{df}{dx} \right|_{x_i} = f'(x_i) \approx \frac{f'(x_i) - f(x_i - a)}{a} \quad (8.14)$$

la aproximación a la primera derivada de  $f(x)$  en  $x_i$  con **diferencias hacia atrás**.

\* Algunos autores lo llaman problema de valor inicial y valores en la frontera (PVIF).



Si en cambio se resta miembro a miembro la ecuación 8.13 de la 8.11 y se aplican los razonamientos anteriores, se llega a la expresión

$$\left. \frac{df}{dx} \right|_{x_i} = f'(x_i) \approx \frac{f(x_i + a) - f(x_i - a)}{2a} \quad (8.15)$$

conocida como la aproximación a la primera derivada de  $f(x)$  en  $x_i$  con diferencias centrales (nótese que se puede obtener la expresión 8.15 sumando miembro a miembro las ecuaciones 8.12 y 8.14 y luego despejando  $f'(x_i)$ ).

Si en las expansiones 8.11 y 8.13 se desprecian los términos quinto, sexto, etc., del lado derecho y se suman miembro a miembro los términos que quedan, se obtiene

$$f''(x_i) = \left. \frac{d^2 f}{dx^2} \right|_{x_i} \approx \frac{f(x_i + a) - 2f(x_i) + f(x_i - a)}{a^2} \quad (8.16)$$

que es la aproximación de la segunda derivada de  $f(x)$  en  $x_i$  por diferencias centrales.

Las aproximaciones de derivadas no están restringidas a funciones de una sola variable; cuando se tiene una función de dos variables, por ejemplo  $T(x, t)$ , sus derivadas parciales por definición son como sigue

$$\begin{aligned} \frac{\partial T}{\partial x} &= \lim_{\Delta x \rightarrow 0} \frac{T(x + \Delta x, t) - T(x, t)}{\Delta x} \\ \frac{\partial T}{\partial t} &= \lim_{\Delta t \rightarrow 0} \frac{T(x, t + \Delta t) - T(x, t)}{\Delta t} \end{aligned} \quad (8.17)$$

Por esto, sus aproximaciones con diferencias hacia delante en el punto  $(x_i, t_j)$  quedan

$$\begin{aligned} \left. \frac{\partial T}{\partial x} \right|_{(x_i, t_j)} &\approx \frac{T(x_i + a, t_j) - T(x_i, t_j)}{a} \quad \text{con } a > 0 \\ \left. \frac{\partial T}{\partial t} \right|_{(x_i, t_j)} &\approx \frac{T(x_i, t_j + b) - T(x_i, t_j)}{b} \quad \text{con } b > 0 \end{aligned} \quad (8.18)$$

La aproximación con diferencias hacia atrás queda

$$\begin{aligned} \left. \frac{\partial T}{\partial x} \right|_{(x_i, t_j)} &\approx \frac{T(x_i, t_j) - T(x_i - a, t_j)}{a} \\ \left. \frac{\partial T}{\partial t} \right|_{(x_i, t_j)} &\approx \frac{T(x_i, t_j) - T(x_i, t_j - b)}{b} \end{aligned} \quad (8.19)$$

y sumando las correspondientes de 8.18 y 8.19 se obtienen

$$\begin{aligned} \left. \frac{\partial T}{\partial x} \right|_{(x_i, t_j)} &\approx \frac{T(x_i + a, t_j) - T(x_i - a, t_j)}{2a} \\ \left. \frac{\partial T}{\partial t} \right|_{(x_i, t_j)} &\approx \frac{T(x_i, t_j + b) - T(x_i, t_j - b)}{2b} \end{aligned} \quad (8.20)$$

que son las aproximaciones con diferencias centrales a las primeras derivadas parciales de  $T(x, t)$ .

Las segundas derivadas parciales de  $T(x, t)$  quedan aproximadas con diferencias centrales así.

$$\left. \frac{\partial^2 T}{\partial x^2} \right|_{(x_i, t_j)} \approx \frac{T(x_i + a, t_j) - 2T(x_i, t_j) + T(x_i - a, t_j)}{a^2} \quad (8.21)$$

$$\left. \frac{\partial^2 T}{\partial t^2} \right|_{(x_i, t_j)} \approx \frac{T(x_i, t_j + b) - 2T(x_i, t_j) + T(x_i, t_j - b)}{b^2}$$

Finalmente, se da la aproximación de la segunda derivada parcial combinada; esto es,

$$\frac{\partial^2 T}{\partial x \partial t} \approx \frac{T(x_i + a, t_j + b) - T(x_i - a, t_j + b) - T(x_i + a, t_j - b) + T(x_i - a, t_j - b)}{4ab} \quad (8.22)$$

cuya deducción se deja al lector como ejercicio

Es importante observar que las ecuaciones 8.18 a 8.22 se pueden obtener a partir de la expansión de  $T(x, t)$  en serie de Taylor, alrededor  $(x_i, t_j)$ ; esto es, de

$$\begin{aligned} T(x_i + a, t_j + b) = & T(x_i, t_j) + a \left. \frac{\partial T}{\partial x} \right|_{(x_i, t_j)} + b \left. \frac{\partial T}{\partial t} \right|_{(x_i, t_j)} \\ & + a^2 \left. \frac{\partial^2 T}{\partial x^2} \right|_{(x_i, t_j)} + 2ab \left. \frac{\partial^2 T}{\partial x \partial t} \right|_{(x_i, t_j)} + b^2 \left. \frac{\partial^2 T}{\partial t^2} \right|_{(x_i, t_j)} + \dots \end{aligned} \quad (8.23)$$

aplicando los mismos razonamientos que condujeron a 8.12 y de 8.14 a 8.16. (véase el Prob. 8.3 al final del capítulo).

### ECUACIÓN DE CALOR UNIDIMENSIONAL EN DIFERENCIAS FINITAS

Una de las ecuaciones diferenciales parciales más estudiadas es:

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \quad (8.24)$$

que describe la conducción de calor en régimen transitorio en una dimensión, la difusión unidireccional de masa en régimen transitorio, etcétera.

Por ejemplo, puede describir la conducción de calor en una barra aislada longitudinalmente durante cierto periodo, tomado a partir de  $t = 0$ . La barra se considera suficientemente delgada y de longitud  $L$  muy grande en comparación con su grosor. Sean los extremos de la barra tomados como  $x = 0$  y  $x = L$  (véase Fig. 8.4).

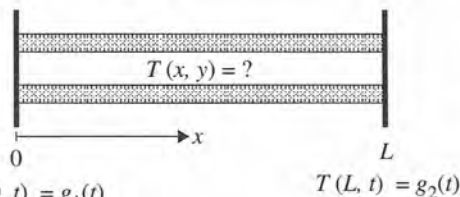
Sean además las condiciones siguientes:

$$a) \quad T(x, 0) = f(x) \quad 0 < x < L$$

Esta expresión, conocida como **condición inicial**, da el valor de la temperatura  $T$  en cualquier punto de la barra al tiempo de inicio  $t = 0$ .

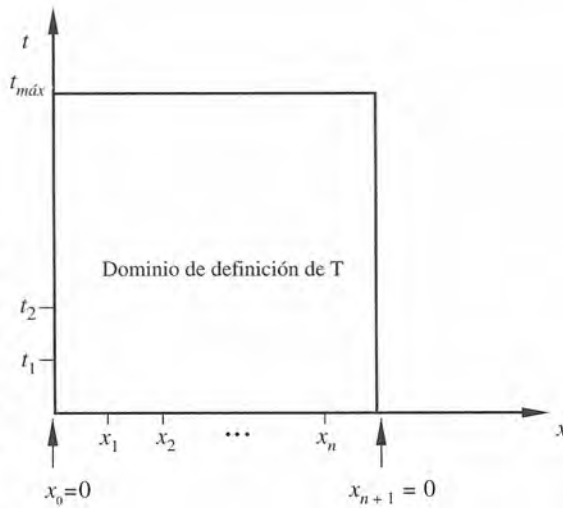
$$b) \quad T(0, t) = g_1(t) \quad \text{con } t > 0$$

$$c) \quad T(L, t) = g_2(t)$$



**Figura 8.4**  
Barra aislada longitudinalmente.

**Figura 8.5**  
Dominio de  
definición de la  
función  
solución del  
problema de  
valor en  
la frontera (PVF)  
planteado.



Estas expresiones, conocidas como **condiciones frontera**, dan los valores de la temperatura  $T$  de la barra en sus extremos a cualquier tiempo  $t$ .

La ecuación 8.24 y las condiciones a), b) y c) constituyen un problema de valores en la frontera (PVF). Resolver este problema numéricamente, significa encontrar los valores de  $T$  en puntos seleccionados en la barra:  $x_1, x_2, \dots, x_n$  a ciertos tiempos escogidos:  $t_1 < t_2 < \dots < t_{\text{máx}}$ ; esto es, calcular

$$\begin{array}{ccc} T(x_1, t_1), & T(x_2, t_1), \dots, & T(x_n, t_1) \\ T(x_1, t_2), & T(x_2, t_2), \dots, & T(x_n, t_2) \\ \vdots & & \vdots \\ T(x_1, t_{\text{máx}}), & T(x_2, t_{\text{máx}}), \dots, & T(x_n, t_{\text{máx}}) \end{array} \quad (8.25)$$

Para ver esto geométricamente, primero se representa el dominio de definición de  $T$  como el rectángulo que se ilustra en el sistema coordenado  $x-t$  de la figura 8.5, y los puntos del dominio de definición donde se aproximarán los valores de  $T$  son los puntos de cruce de las horizontales  $t = t_1, \dots, t = t_{\text{máx}}$  y las verticales  $x = x_1, \dots, x = x_n$ , que en adelante se llamarán **nodos** (véase Fig. 8.6).

La ecuación 8.24 es válida en todo el dominio de definición, por lo que evidentemente será válida en cualquier nodo, por ejemplo  $(x_i, t_j)$ ; esto es,

$$\left. \frac{\partial T}{\partial t} \right|_{(x_i, t_j)} = \alpha \left. \frac{\partial^2 T}{\partial x^2} \right|_{(x_i, t_j)}$$

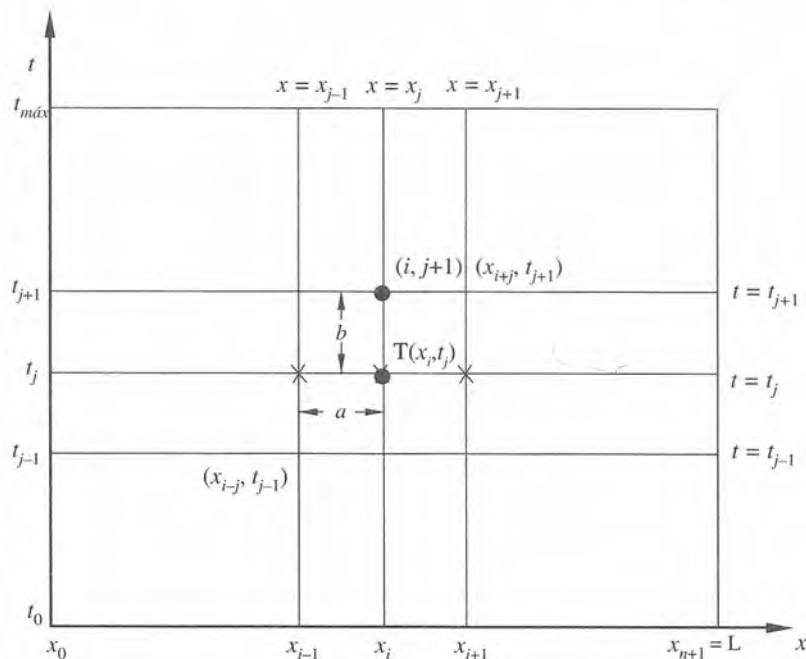
Si se sustituye ahora las derivadas parciales evaluadas en  $(x_i, t_j)$  con sus aproximaciones con diferencias finitas en esta ecuación; por ejemplo, con diferencias finitas hacia adelante a  $\partial T / \partial t$  y diferencias centrales a  $\partial^2 T / \partial x^2$ , se obtiene

$$\frac{T_{i,j+1} - T_{i,j}}{b} = \alpha \frac{T_{i-1,j} - 2T_{i,j} + T_{i+1,j}}{a^2} \quad (8.26)$$

Se ha remplazado  $T(x_i, t_j)$  con  $T_{i,j}$  para simplificar la notación.

Hay que observar además que los nodos marcados con punto negro ( $\bullet$ ) en la figura 8.6 son los nodos usados para aproximar  $\partial T / \partial t$  y los marcados con una cruz ( $\times$ ) se emplean a fin de aproximar a  $\partial^2 T / \partial x^2$ .





**Figura 8.6**  
Nodos de una  
red construida  
en el dominio  
de definición.

De manera similar se obtienen las ecuaciones para los demás nodos de la red (o mailla), lo que conduce a un conjunto de ecuaciones algebraicas que pueden ser simultáneas o no y que involucran los valores de  $T_{i,j}$  que se buscan. Su solución es la misma del problema de valor en la frontera (PVF).

Es oportuno hacer notar que la derivada parcial en el tiempo  $\partial T / \partial t$  se pudo aproximar con diferencias hacia atrás o con diferencias divididas centrales.

## 8.3 Solución de la ecuación de calor unidimensional

### MÉTODO EXPLÍCITO

Para ilustrar este método se resuelve el PVF planteado en la sección anterior con los datos siguientes:

$$\text{PVF} \left\{ \begin{array}{l} \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \\ T(x, 0) = 20^\circ \text{F}, \quad 0 < x < L \\ T(0, t) = 100^\circ \text{F}, \quad t > 0 \\ T(1, t) = 100^\circ \text{F}, \quad t > 0 \end{array} \right.$$

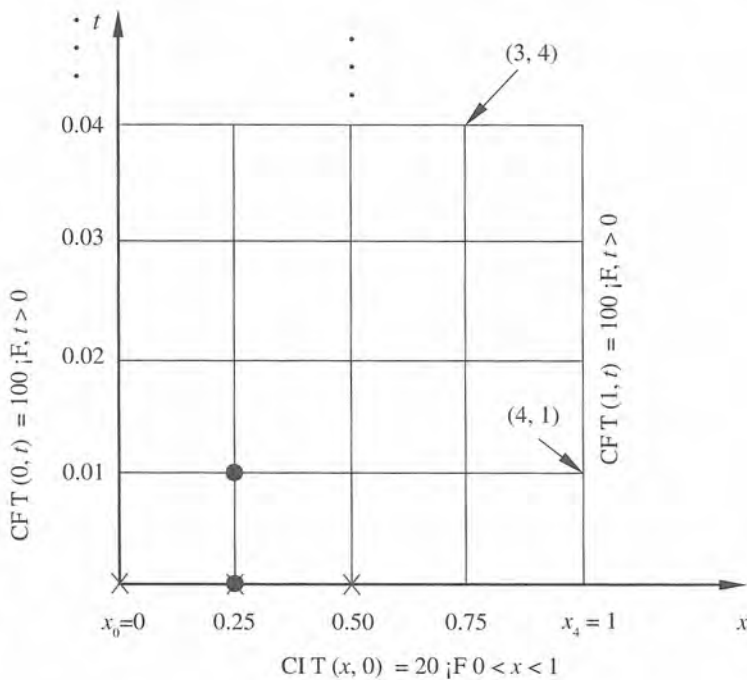
y

$$\alpha = 1 \text{ pie}^2/\text{h}$$

$$L = 1 \text{ pie}$$

$$t_{\text{máx}} = 1 \text{ h}$$

**Solución** Primero se construye la malla en el dominio de definición dividiendo la longitud de la barra (1 pie) en cuatro subintervalos y el intervalo de tiempo (1 h) en 100 subintervalos.



**Figura 8.7**  
Representación  
de la malla en  
el dominio de  
definición.

Las condiciones frontera proporcionan la temperatura en cualquier punto del eje  $t$  y de la vertical  $x = 1$  a cualquier tiempo, mientras que la condición inicial proporciona la temperatura en cualquier punto del eje horizontal  $x$  al tiempo cero.

Cada nodo de la malla queda definido por dos coordenadas  $(i, j)$ ; por ejemplo, el nodo de coordenadas  $(3, 4)$  representa la temperatura en el punto  $x = 0.75$  pies de la barra al tiempo  $t = 0.04$  horas, y el nodo  $(4, 1)$  la temperatura de la barra en  $x = 1$  pie (su frontera) y a  $t = 0.01$  horas (véase Fig. 8.7).

Hay que observar que en el nodo de coordenadas  $(0, 0)$  (esquina izquierda inferior de la malla), la temperatura debería ser  $20^\circ\text{F}$  atendiendo la condición inicial, mientras que la condición frontera  $T(0, t)$  establece que debería ser de  $100^\circ\text{F}$ .

Los puntos que presentan estas características se llaman **puntos singulares**; se acostumbra tomar en ellos un valor de temperatura igual a la media aritmética de las temperatura sugeridas por la condición inicial y la condición frontera correspondientes. La temperatura tomada para el nodo  $(0, 0)$  es  $60^\circ\text{F}$ . De igual manera se trata el punto  $(4, 0)$ , cuya temperatura también es  $60^\circ\text{F}$ .

Hechas estas consideraciones, el segundo paso consiste en aproximar la ecuación diferencial parcial del problema de valor en la frontera en el nodo  $(1, 0)$  por la ecuación 8.26, entonces queda:

$$\frac{T_{1,1} - T_{1,0}}{b} = \alpha \frac{T_{0,0} - 2T_{1,0} + T_{2,0}}{a^2}$$

Los nodos involucrados en esta ecuación están marcados por el círculo y cruces en la figura 8.7. De éstos, solamente en el nodo  $(1,1)$  la temperatura es desconocida, por lo que puede despejarse; entonces resulta

$$T_{1,1} = \alpha \frac{b}{a^2} (T_{0,0} - 2T_{1,0} + T_{2,0}) + T_{1,0}$$

al sustituir valores queda

$$T_{1,1} = 1 \frac{0.01}{(0.25)^2} (60 - 2(20) + 20) + 20 = 26.4$$

Si ahora se aproxima la EDP en el nodo  $(i, j) = (2, 0)$  mediante la 8.26, se obtiene

$$\frac{T_{2,1} - T_{2,0}}{b} = \alpha \frac{T_{1,0} - 2T_{2,0} + T_{3,0}}{a^2}$$

Ahora sólo se desconoce la temperatura del punto  $(2, 1)$ , ya que todos los demás están dados por la condición inicial; despejando se tiene

$$T_{2,1} = \alpha \frac{b}{a^2} (T_{1,0} - 2T_{2,0} + T_{3,0}) + T_{2,0}$$

se sustituyen valores

$$T_{2,1} = 1 \frac{0.01}{(0.25)^2} (20 - 2(20) + 20) + 20 = 20$$

Se repiten las mismas consideraciones y cálculos para el punto  $(3, 0)$  y se obtiene:

$$T_{3,1} = 1 \frac{0.01}{(0.25)^2} (T_{2,0} - 2T_{3,0} + T_{4,0}) + T_{3,0}$$

$$T_{3,1} = 26.4$$

De esta manera se han obtenido aproximaciones a la temperatura en los tres puntos seleccionados de la barra a un tiempo de 0.01 horas. Al momento se tiene la temperatura de todos los nodos de las dos primeras líneas horizontales (filas) de la malla y se procederá, siguiendo el razonamiento anterior, a calcular la temperatura en todos los nodos intermedios de la tercera fila  $(1, 2)$ ,  $(2, 2)$  y  $(3, 2)$ .

Se empieza con el punto  $(i, j) = (1, 1)$  y se aplica la ecuación 8.26, con lo que se obtiene:

$$\frac{T_{1,2} - T_{1,1}}{b} = \alpha \frac{T_{0,1} - 2T_{1,1} + T_{2,1}}{a^2}$$

de la que

$$T_{1,2} = \alpha \frac{b}{a^2} (T_{0,1} - 2T_{1,1} + T_{2,1}) + T_{1,1}$$

con la sustitución de valores queda:

$$T_{1,2} = 1 \frac{0.01}{(0.25)^2} (100 - 2(26.4) + 20) + 26.4 = 37.152$$

Al proceder análogamente para los otros puntos se llega a:

$$\begin{aligned} T_{2,2} &= 22.048 \\ T_{3,2} &= 37.152 \end{aligned}$$

Con esto se tiene la temperatura en los tres puntos seleccionados de la barra cuando hayan transcurrido 0.02 horas.

Este procedimiento se repite para la cuarta, quinta, etc. filas, con lo cual se obtienen las temperaturas en los puntos seleccionados de la barra a tiempo  $t = 0.03$ ,  $t = 0.04$ , etc., hasta llegar al tiempo fijado como  $t_{\text{máx}} = 1$  hora.



De los cien conjuntos de temperaturas obtenidas, en la tabla 8.1 se muestran sólo algunos para facilitar su presentación y análisis.

Este método también se conoce como **métodos de diferencias hacia delante**.

## DISCUSIÓN DE RESULTADOS

- Hay simetría en la distribución de temperaturas en la barra debido a que: *a)* la temperatura inicial es constante; *b)* la temperatura es constante e igual en las fronteras, y *c)* las propiedades físicas de la barra son independientes de  $x$  y  $t$ .

**Tabla 8.1** Resultados de la solución del PVF de conducción de calor en una barra metálica.

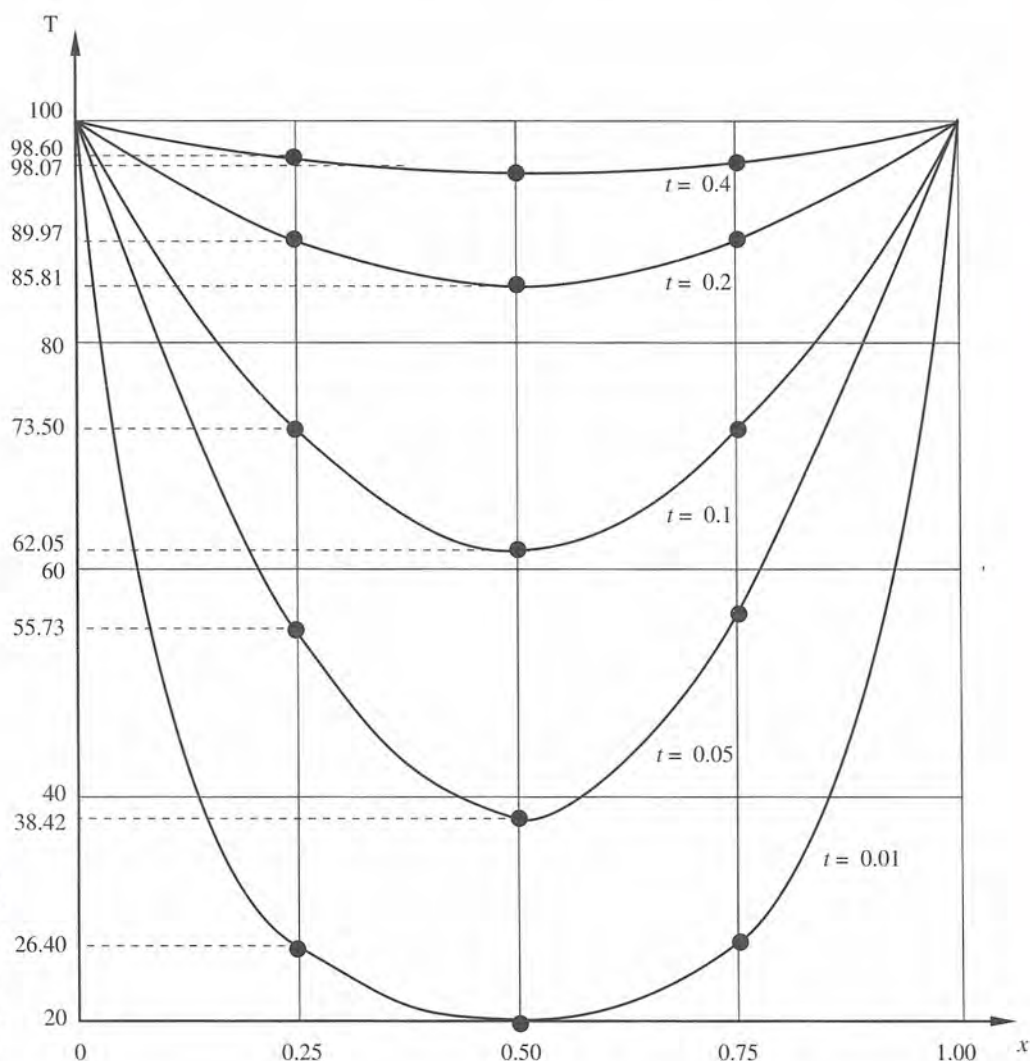
| Tiempo<br>(hrs) | $x$ (pies) |        |        |        |     |
|-----------------|------------|--------|--------|--------|-----|
|                 | 0.00       | 0.25   | 0.5    | 0.75   | 1.0 |
| 0.00            | 60         | 20.000 | 20.000 | 20.000 | 60  |
| 0.01            | 100        | 26.400 | 20.000 | 26.400 | 100 |
| 0.02            | 100        | 37.152 | 22.048 | 37.152 | 100 |
| 0.03            | 100        | 44.791 | 26.881 | 44.791 | 100 |
| 0.04            | 100        | 50.759 | 32.612 | 50.759 | 100 |
| 0.05            | 100        | 55.734 | 38.419 | 55.734 | 100 |
| 0.06            | 100        | 60.046 | 43.960 | 60.046 | 100 |
| 0.07            | 100        | 63.865 | 49.108 | 63.865 | 100 |
| 0.08            | 100        | 67.285 | 53.830 | 67.285 | 100 |
| 0.09            | 100        | 70.367 | 58.136 | 70.367 | 100 |
| 0.10            | 100        | 73.151 | 62.050 | 73.151 | 100 |
| 0.20            | 100        | 89.968 | 85.812 | 89.968 | 100 |
| 0.40            | 100        | 98.599 | 98.018 | 98.599 | 100 |
| 0.60            | 100        | 99.804 | 99.723 | 99.804 | 100 |
| 0.80            | 100        | 99.973 | 99.961 | 99.973 | 100 |
| 1.00            | 100        | 99.996 | 99.995 | 99.996 | 100 |

- La temperatura en el centro de la barra es un mínimo, de manera que se satisface

$$\left. \frac{dT}{dx} \right|_{x=\frac{1}{2}} = 0$$

(véase Fig. 8.8), ya que es el punto más alejado de los extremos, los cuales tienen las temperaturas que impulsan el flujo de calor hacia el centro de la barra (donde se da el gradiente máximo,  $\frac{dT}{dx}$  máxima).

- Hay que observar que cuando  $t = 0.01$  la temperatura en el punto central es igual a la inicial, o sea  $T(0.5, 0.01) = 20^\circ\text{F}$ . Esta situación no es congruente con el fenómeno que ocurre, ya que es de esperar que la temperatura cambie después del instante cero. El resultado se debe a que la estimación de la temperatura en un nodo depende de las temperaturas de los nodos en un tiempo previo.
- La temperatura en la barra tiende al régimen permanente a medida que transcurre el tiempo, es decir,  $T \rightarrow 100^\circ\text{F}$  cuando  $t \rightarrow \infty$ .
- Sólo se encontró la temperatura en tres puntos interiores de la barra. Si se desea información de mayor número de puntos interiores, debe construirse una malla más cerrada; es decir, subdividir la longitud  $L$  en más subintervalos.



**Figura 8.8**  
Distribución  
de la  
temperatura  
en la barra  
a diferentes  
tiempos.

**ALGORITMO 8.1** Método explícito

Para aproximar la solución al problema de valor en la frontera

$$\alpha \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t}$$

$$T(x, 0) = f(x), \quad 0 < x < x_F$$

$$T(0, t) = g_1(t)$$

$$T(x_F, t) = g_2(t), \quad t > 0$$

Proporcionar las funciones  $CI(X)$ ,  $CF1(T)$  y  $CF2(T)$  y los

**DATOS:** El número  $NX$  de puntos de la malla en el eje  $x$ , el número  $NT$  de puntos de la malla en el eje  $t$ , la longitud total  $XF$  del eje  $x$ , el tiempo máximo  $TF$  por considerar y el coeficiente  $ALFA$  de la derivada de segundo orden.

**RESULTADOS:** Los valores de la variable dependiente  $T$  a lo largo del eje  $x$  a distintos tiempos  $t$ :  $T$ .

PASO 1. Hacer  $DX = XF/(NX - 1)$ .

PASO 2. Hacer  $DT = TF/(NT - 1)$ .

PASO 3. Hacer  $LAMBDA = ALFA * DT / DX^2$ .

PASO 4. Hacer  $I = 2$ .

PASO 5. Mientras  $I \leq NX - 1$ , repetir los pasos 6 y 7.

PASO 6. Hacer  $T(I) = CI(DX * (I - 1)) / 2$ .

PASO 7. Hacer  $I = I + 1$ .

PASO 8. Hacer  $T(1) = (CI(0) + CF1(0)) / 2$ .

PASO 9. Hacer  $T(NX) = (CI(XF) + CF2(0)) / 2$ .

PASO 10. IMPRIMIR  $T$

PASO 11. Hacer  $J = 1$

PASO 12. Mientras  $J \leq NT$  repetir los pasos 13 a 24.

PASO 13. Hacer  $I = 2$

PASO 14. Mientras  $I \leq NX - 1$ , repetir los pasos 15 y 16.

PASO 15. Hacer  $T1(I) = LAMBDA * T(I - 1) + (1 - 2 * LAMBDA) * T(I) + LAMBDA * T(I + 1)$

PASO 16. Hacer  $I = I + 1$

PASO 17. Hacer  $I = 2$

PASO 18. Mientras  $I \leq NX - 1$ , repetir los pasos 19 y 20.

PASO 19. Hacer  $T(I) = T1(I)$

PASO 20. Hacer  $I = I + 1$

PASO 21. Hacer  $T(1) = CF1(DT * J)$

PASO 22. Hacer  $T(NX) = CF2(DT * J)$

PASO 23. IMPRIMIR  $T$ .

PASO 24. Hacer  $J = J + 1$

PASO 25. TERMINAR.

**Ejemplo 8.1**

Calcule la temperatura como una función de  $x$  y  $t$  en una barra aislada de longitud unitaria (en pies), sujeta a las siguientes condiciones inicial y de frontera

$$CI \quad T(x, 0) = 50 \text{ sen } \pi x \quad 0 < x < 1$$

$$CF1 \quad T(0, t) = 100^\circ F$$

$$t > 0$$

$$CF2 \quad T(1, t) = 50^\circ F$$

y con  $\alpha = 1 \text{ pie}^2 / h$ .



## Solución



El problema de condiciones en la frontera queda establecido como sigue

$$\text{PVF} \quad \begin{cases} \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \\ T(x, 0) = 50 \text{ sen } \pi x \\ T(0, t) = 100^\circ \text{F} \\ T(1, t) = 50^\circ \text{F} \end{cases}$$

Ahora se divide la barra en  $n = 8$  segmentos o subintervalos, de tal manera que se tiene un total de nueve nodos en cada fila, de los cuales siete son interiores; con esto,  $a = 0.125$  pies. El fenómeno se estudiará durante media hora y se dividirá este tiempo de interés en  $m = 100$ , que da  $b = 0.005$  horas. La malla queda como se muestra en la figura 8.9.

Para mayor facilidad del uso de la ecuación 8.26 se despeja el término  $T_{i,j+1}$ , ya que representa la temperatura desconocida, y se denominará  $\lambda$  el término  $\alpha b/a^2$ ; después de algunas manipulaciones algebraicas, dicha ecuación queda

$$T_{i,j+1} = \lambda T_{i-1,j} + (1 - 2\lambda) T_{i,j} + \lambda T_{i+1,j} \quad (8.27)$$

cuya aplicación en  $(i, j) = (1, 0)$  produce

$$T_{1,1} = \lambda T_{0,0} + (1 - 2\lambda) T_{1,0} + \lambda T_{2,0}$$

se calcula el valor de  $\lambda$

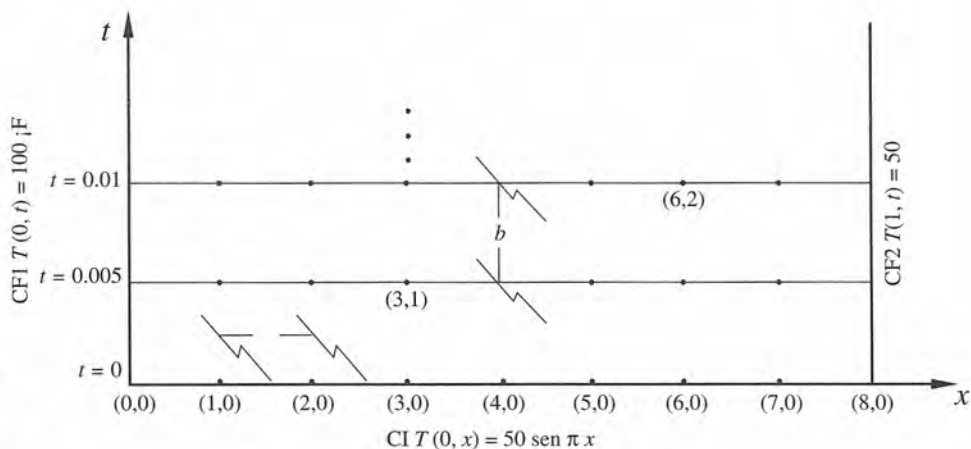
$$\lambda = 1(0.005)/(0.125)^2 = 0.32$$

y se substituyen valores

$$T_{1,1} = 0.32 (50) + (1 - 2(0.32)) (50) \text{ sen } (0.125\pi) + 0.32 (50) \text{ sen } (0.25\pi) = 34.2$$

Para  $(i, j) = (2, 0)$  se tiene:

$$\begin{aligned} T_{2,1} &= \lambda T_{1,0} + (1 - 2\lambda) T_{2,0} + \lambda T_{3,0} \\ T_{2,1} &= 0.32 (50) \text{ sen } (0.125\pi) + (1 - 2(0.32)) (50) \text{ sen } (0.250\pi) \\ &\quad + 0.32 (50) \text{ sen } (0.375\pi) = 33.63 \end{aligned}$$



**Figura 8.9**  
Malla del  
ejemplo 8.1.

Al continuar

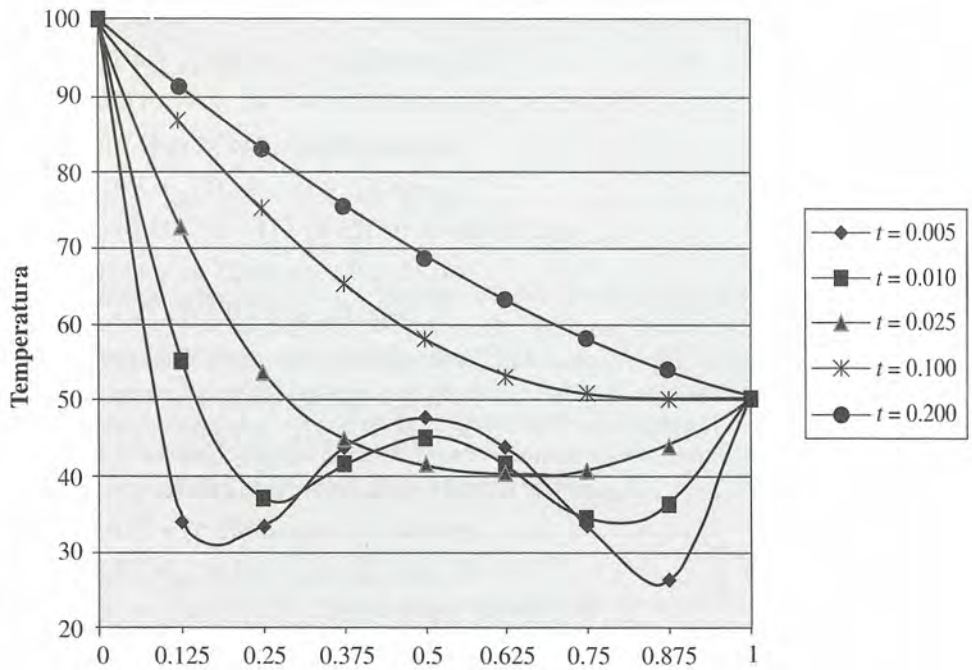
$$\begin{aligned}
 T_{3,1} &= \lambda T_{2,0} + (1 - 2\lambda)T_{3,0} + \lambda T_{4,0} \\
 T_{3,1} &= 0.32 (50) \sin (0.25 \pi) + 1 - 2 (0.32) (50) \sin (0.375 \pi) \\
 &\quad + 0.32 (50) \sin (0.5 \pi) = 43.94 \\
 T_{4,1} &= \lambda T_{3,0} + (1 - 2\lambda)T_{4,0} + \lambda T_{5,0} \\
 T_{4,1} &= 0.32 (50) \sin (0.375 \pi) + (1 - 2 (0.32)) (50) \sin (0.5 \pi) \\
 &\quad + 0.32 (50) \sin (0.625 \pi) = 47.56 \\
 T_{5,1} &= \lambda T_{4,0} + (1 - 2\lambda)T_{5,0} + \lambda T_{6,0} \\
 T_{5,1} &= 0.32 (50) \sin (0.5 \pi) + (1 - 2 (0.32)) (50) \sin (0.625 \pi) \\
 &\quad + 0.32 (50) \sin (0.75 \pi) = 43.94 \\
 T_{6,1} &= \lambda T_{5,0} + (1 - 2\lambda)T_{6,0} + \lambda T_{7,0} \\
 T_{6,1} &= 0.32 (50) \sin (0.625 \pi) + (1 - 2 (0.32)) (50) \sin (0.75 \pi) \\
 &\quad + 0.32 (50) \sin (0.875 \pi) = 33.63 \\
 T_{7,1} &= \lambda T_{6,0} + (1 - 2\lambda)T_{7,0} + \lambda T_{8,0} \\
 T_{7,1} &= 0.32 (50) \sin (0.75 \pi) + (1 - 2 (0.32)) (50) \sin (0.875 \pi) \\
 &\quad + 0.32 (25) = 26.2
 \end{aligned}$$

Estas temperaturas corresponden a puntos discretos sobre la barra a un tiempo igual a 0.005 horas.

Para obtener la temperatura en los mismos puntos de la barra dados arriba, pero ahora a un tiempo de 0.01 h (tercera fila de la malla de la Fig. 8.9), se aplica nuevamente la ecuación 8.27. De la misma manera se obtienen los valores de temperatura para los tiempos de 0.015, 0.02 h, etc.; o sea, la temperatura en los nodos interiores de las filas 4, 5, etc. Los resultados obtenidos con el **PROGRAMA 8.1** del CD son:

**Tabla 8.2** Temperaturas (°F) del ejemplo 8.1.

| $t$<br>(horas) | $x$ (pies) |       |       |       |       |       |       |       |     |
|----------------|------------|-------|-------|-------|-------|-------|-------|-------|-----|
|                | 0.0        | 0.125 | 0.25  | 0.375 | 0.5   | 0.625 | 0.75  | 0.875 | 1.0 |
| 0.000          | 50         | 19.13 | 35.36 | 46.19 | 50.00 | 46.19 | 35.36 | 19.13 | 25  |
| 0.005          | 100        | 34.20 | 33.63 | 43.94 | 47.56 | 43.94 | 33.63 | 26.20 | 50  |
| 0.010          | 100        | 55.08 | 37.11 | 41.80 | 45.25 | 41.80 | 34.55 | 36.20 | 50  |
| 0.015          | 100        | 63.70 | 44.36 | 41.40 | 43.04 | 40.59 | 37.40 | 40.09 | 50  |
| 0.020          | 100        | 69.13 | 49.61 | 42.88 | 41.73 | 40.35 | 39.28 | 42.40 | 50  |
| 0.025          | 100        | 72.76 | 53.70 | 44.66 | 41.66 | 40.45 | 40.62 | 43.83 | 50  |
| 0.030          | 100        | 75.38 | 56.91 | 46.59 | 42.23 | 40.89 | 41.59 | 44.78 | 50  |
| .              | .          | .     | .     | .     | .     | .     | .     | .     | .   |
| .              | .          | .     | .     | .     | .     | .     | .     | .     | .   |
| .              | .          | .     | .     | .     | .     | .     | .     | .     | .   |
| 0.050          | 100        | 81.19 | 65.19 | 53.68 | 46.92 | 44.17 | 44.41 | 46.68 | 50  |
| 0.100          | 100        | 86.98 | 75.08 | 65.18 | 57.81 | 53.06 | 50.61 | 49.86 | 50  |
| 0.150          | 100        | 89.73 | 80.09 | 71.59 | 64.57 | 59.14 | 55.16 | 52.28 | 50  |
| 0.200          | 100        | 91.32 | 83.02 | 75.40 | 68.67 | 62.90 | 58.03 | 53.83 | 50  |
| 0.300          | 100        | 92.86 | 85.85 | 79.10 | 72.67 | 66.60 | 60.85 | 55.36 | 50  |
| 0.400          | 100        | 93.42 | 86.89 | 80.46 | 74.14 | 67.96 | 61.89 | 55.92 | 50  |
| 0.500          | 100        | 93.63 | 87.28 | 80.96 | 74.68 | 68.46 | 62.28 | 56.13 | 50  |



**Figura 8.10**  
Gráfica de la  
distribución de  
temperatura a  
diferentes  
tiempos.

### DISCUSIÓN DE LOS RESULTADOS

- En este caso, la distribución de temperaturas con respecto al centro de la barra no es simétrica, a pesar de que la distribución inicial sí lo es,  $50 \sin(\pi x)$ . Esto se debe a que la temperatura en los extremos es diferente, lo cual genera un flujo de calor más intenso del extremo izquierdo hacia el centro de la barra.
- La temperatura en el centro de la barra y sus cercanías disminuye en el intervalo  $0 < t < 0.02$  (ver tabla 8.2). Esto se debe a que cuando  $t = 0$ , la temperatura en el centro de la barra es mayor que en sus vecindades, de tal modo que en un lapso hay flujo de calor del centro de la barra hacia los lados (más pronunciado hacia el extremo derecho), razón por la cual la temperatura en la zona central disminuye aproximadamente hasta 0.025 segundos (en la figura 8.10 gráficas a  $t = 0.05$ , 0.010 y 0.025 segundos) para después empezar a aumentar (gráfica a  $t = 0.1$  segundos).
- Cuando el lapso es amplio ( $t > 0.2$  segundos), la distribución de la temperatura es casi lineal a lo largo de la barra (en la figura 8.10 gráfica a  $t = 0.2$  segundos). Es de esperarse que sea lineal cuando  $t \rightarrow \infty$ , esto es, que se alcance el régimen permanente en un tiempo muy grande. Al no cambiar la temperatura en el tiempo, el término  $\frac{\partial T}{\partial t} = 0$  y la

EDP de nuestro problema se simplifica a  $\frac{\partial^2 T}{\partial x^2} = 0$  o  $\frac{d^2 T}{dx^2} = 0$ , una ecuación diferencial

ordinaria, cuya solución analítica es  $T = c_1 x + c_2$ .

Los cálculos pueden hacerse con Matlab:





```

Xf=1; Nx=9; tf=0.5; Nt=101; Alfa=1;
Dx=Xf/(Nx-1); Dt=tf/(Nt-1);
Lambda=Alfa*Dt/Dx^2
for i=2:Nx-1
    T(i)=CI8_1(Dx*(i-1));
end
T(1)=(CI8_1(0)+CF1_8_1(0))/2;
T(Nx)=(CI8_1(Xf)+CF2_8_1(0))/2;
fprintf('0.000')
for i=1:Nx
    fprintf(' %6.2f',T(i))
end
fprintf('\n')
for j=1:Nt-1
    for i=2:Nx-1
        Tl(i)=Lambda*T(i-1)+(1-2*Lambda)*T(i)+...
            Lambda*T(i+1);
    end
    for i=2:Nx-1
        T(i)=Tl(i);
    end
    T(1)=CF1_8_1(Dt*j);
    T(Nx)=CF2_8_1(Dt*j);
    fprintf('%5.3f',j*Dt)
    for i=1:Nx
        fprintf(' %6.2f',T(i))
    end
    fprintf('\n')
end

```

También se puede usar Microsoft Excel para realizar los cálculos. Para ello escriba en las celdas indicadas en la primera columna de la siguiente tabla la fórmula de la segunda columna. Después arrastre la fórmula de la celda C6 hasta la celda J6; arrastre ahora las fórmulas de las celdas C7 y C8 hasta las celdas I7 e I8, respectivamente. Por último, arrastre el rango de celdas A8:J8 hasta las celdas A107:J107.

| Celda | Fórmula                      |
|-------|------------------------------|
| A6    | Tiempo                       |
| A7    | 0                            |
| A8    | =A7+\$D\$2                   |
| B 1   | alfa                         |
| B2    | 1                            |
| B5    | Distancia                    |
| B6    | 0                            |
| B7    | =(50*SENO(PI()*B6)+\$C\$4)/2 |
| B8    | =\$C\$4                      |
| C 1   | a=Deltax                     |
| C2    | 0.125                        |
| C3    | CF1                          |

|     |                                      |
|-----|--------------------------------------|
| C4  | 100                                  |
| C6  | =B6+\$C\$2                           |
| C7  | =50*SENO(PI()*C6)                    |
| C8  | =\$E\$2*B7+(1-2*\$E\$2)*C7+\$E\$2*D7 |
| D 1 | b=Deltat                             |
| D2  | 0.005                                |
| D3  | CF2                                  |
| D4  | 50                                   |
| E 1 | Lambda                               |
| E2  | =B2*D2/C2^2                          |
| J7  | =(5 0* SEÑO(PI()*J6)+\$D\$4)/2       |
| J8  | =\$D\$4                              |

### MÉTODO IMPLÍCITO



Para ilustrar este método se resolverá nuevamente el ejemplo

$$\text{PVP} \left\{ \begin{array}{l} \text{EDP } \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \text{ y } \alpha = 1 \text{ pie}^2/\text{h} \\ \text{C.I. } T(x, 0) = 20^\circ\text{F, } L = 1 \text{ pie} \\ \text{CF1 } T(0, t) = 100^\circ\text{F, } t_{\text{máx}} = 1 \text{ h} \\ \text{CF2 } T(1, t) = 100^\circ\text{F} \end{array} \right.$$

ya utilizado para mostrar el método explícito.

Primero se obtendrá la ecuación básica del algoritmo.

Se toma el nodo  $(i, j)$  de la malla construida sobre el dominio de definición  $0 = t_0 < t < t_{\text{máx}} = 1, 0 < x < L = 1$  (Fig. 8.11) y se evalúa la EDP, entonces

$$\left. \frac{\partial T}{\partial t} \right|_{(x_i, t_j)} = \alpha \left. \frac{\partial^2 T}{\partial x^2} \right|_{(x_i, t_j)}$$

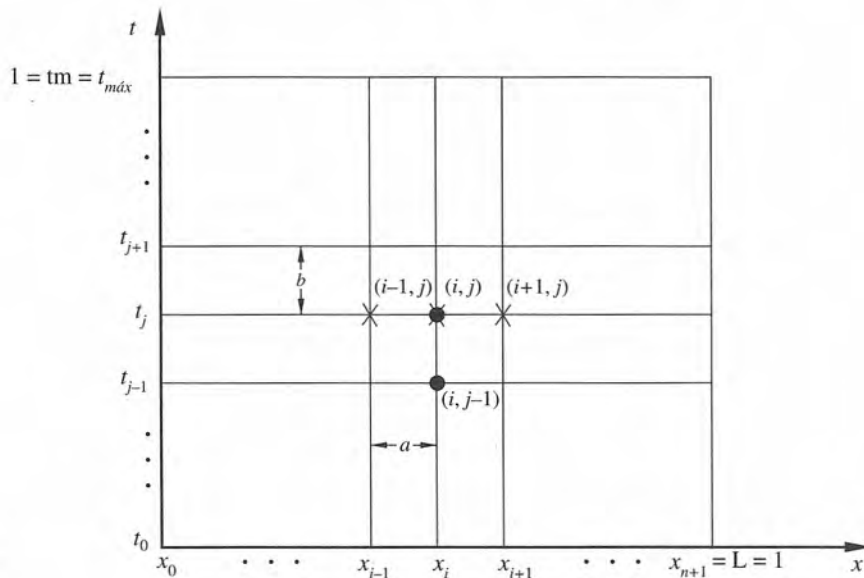


Figura 8.11

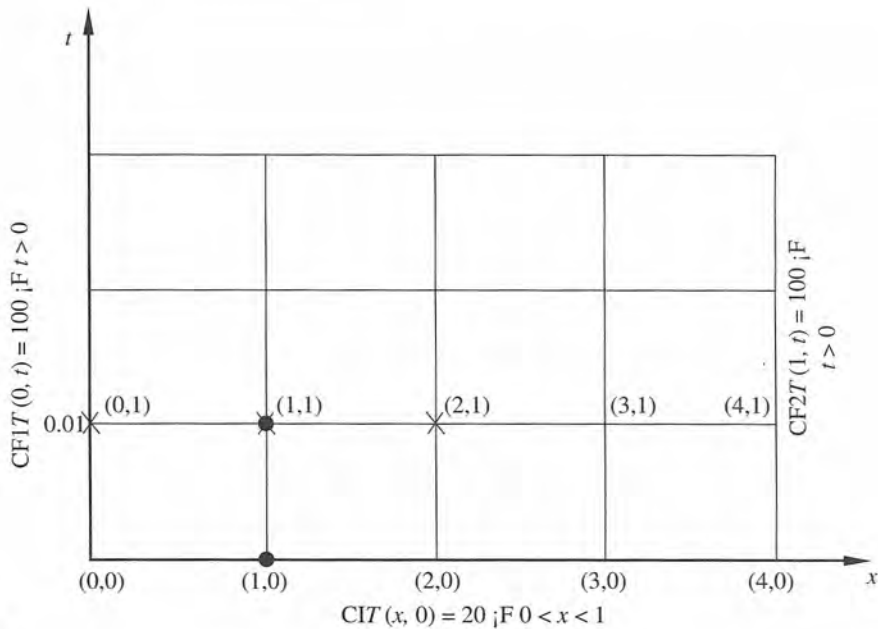


Figura 8.12

Ahora se sustituye  $\partial T / \partial t$  en  $(x_i, t_j)$  por diferencias hacia atrás, y  $\partial^2 T / \partial x^2$  en  $(x_i, t_j)$  por diferencias centrales, lo que da

$$\frac{T_{i,j} - T_{i,j-1}}{b} = \alpha \frac{T_{i-1,j} - 2T_{i,j} + T_{i+1,j}}{a^2} \quad (8.28)$$

De acuerdo con la notación de punto negro (•) para los nodos empleados a fin de aproximar a  $\partial T / \partial t$  y cruz (x) para aquellos que se usan en la aproximación de  $\partial^2 T / \partial x^2$ , se tiene el esquema de la figura 8.12.

**Solución** Se construye la malla con  $n = 4$  y  $m = 100$ , con lo que  $a = 0.25$  y  $b = 0.01$ .

Si  $(i, j) = (1, 1)$ , la ecuación 8.28 se aproxima así

$$\frac{T_{1,1} - T_{1,0}}{b} = \alpha \frac{T_{0,1} - 2T_{1,1} + T_{2,1}}{a^2}$$

La temperatura en los nodos (0,1) y (1,0) está dada por las condiciones frontera e inicial, respectivamente, pero se desconoce la temperatura en los nodos (1,1) y (2,1). Entonces se tiene una ecuación con dos incógnitas que rearmada queda

$$(1 + 2\lambda) T_{1,1} - \lambda T_{2,1} = T_{1,0} + \lambda T_{0,1} \quad (8.29)$$

donde, como se sabe,  $\lambda = \alpha b / a^2$  (que es un parámetro adimensional).

El procedimiento se repite en el nodo (2,1) y la ecuación diferencial parcial queda aproximada por:

$$\frac{T_{2,1} - T_{2,0}}{b} = \alpha \frac{T_{1,1} - 2T_{2,1} + T_{3,1}}{a^2}$$

En esta ecuación hay tres incógnitas  $T_{1,1}$ ,  $T_{2,1}$  y  $T_{3,1}$ ; así pues, al rearmarla queda:

$$-\lambda T_{1,1} + (1 + 2\lambda) T_{2,1} - \lambda T_{3,1} = T_{2,0} \quad (8.30)$$



Análogamente para el nodo (3,1), la ecuación diferencial parcial (EDP) queda aproximada por:

$$\frac{T_{3,1} - T_{3,0}}{b} = \alpha \frac{T_{2,1} - 2T_{3,1} + T_{4,1}}{a^2}$$

En esta ecuación sólo hay dos incógnitas, que son  $T_{2,1}$  y  $T_{3,1}$ ; así pues, al reorganizarla resulta:

$$-\lambda T_{2,1} + (1 + 2\lambda) T_{3,1} = T_{3,0} + \lambda T_{4,1} \quad (8.31)$$

Las ecuaciones 8.29 a 8.31 constituyen un sistema de ecuaciones algebraicas lineales en las incógnitas  $T_{1,1}$ ,  $T_{2,1}$  y  $T_{3,1}$ , que son precisamente las temperaturas que se desea conocer. Esto es:

$$\begin{aligned} (1 + 2\lambda) T_{1,1} - \lambda T_{2,1} &= \lambda T_{0,1} + T_{1,0} \\ -\lambda T_{1,1} + (1 + 2\lambda) T_{2,1} - \lambda T_{3,1} &= T_{2,0} \\ -\lambda T_{2,1} + (1 + 2\lambda) T_{3,1} &= T_{3,0} + \lambda T_{4,1} \end{aligned}$$

Con la sustitución de valores

$$\lambda = 0.16, \quad T_{1,0} = T_{2,0} = T_{3,0} = 20^\circ\text{F}, \quad T_{0,1} = T_{4,1} = 100^\circ\text{F}$$

y resolviendo por alguno de los métodos del capítulo 3, se obtiene:

$$T_{1,1} = 29.99, \quad T_{2,1} = 22.42, \quad T_{3,1} = 29.99,$$

Hay que observar que estas temperaturas obtenidas para  $t = 0.01$  h son diferentes a las obtenidas con el método explícito; además, la temperatura del punto central es distinta de la condición inicial. Esta situación es más congruente con la realidad del fenómeno que ocurre (recuérdese que con el método explícito la temperatura es  $20^\circ\text{F}$ ). Lo anterior se explica porque para el cálculo se han tomado en cuenta todos los nodos de la primera y segunda filas, excepto los de las esquinas  $T(0,0)$  y  $T(4,0)$ .

Mediante la ecuación 8.28 y los mismos razonamientos para la segunda y tercera filas se llega a

$$\begin{aligned} (1 + 2\lambda) T_{1,2} - \lambda T_{2,2} &= \lambda T_{0,2} + T_{1,1} \\ -\lambda T_{1,2} + (1 + 2\lambda) T_{2,2} - \lambda T_{3,2} &= T_{2,1} \\ -\lambda T_{2,2} + (1 + 2\lambda) T_{3,2} &= T_{3,1} + \lambda T_{4,2} \end{aligned}$$

Al sustituir valores conocidos

$$\lambda = 0.16, \quad T_{0,2} = T_{4,2} = 100, \quad T_{1,1} = T_{3,1} = 29.99, \quad T_{2,1} = 22.42,$$

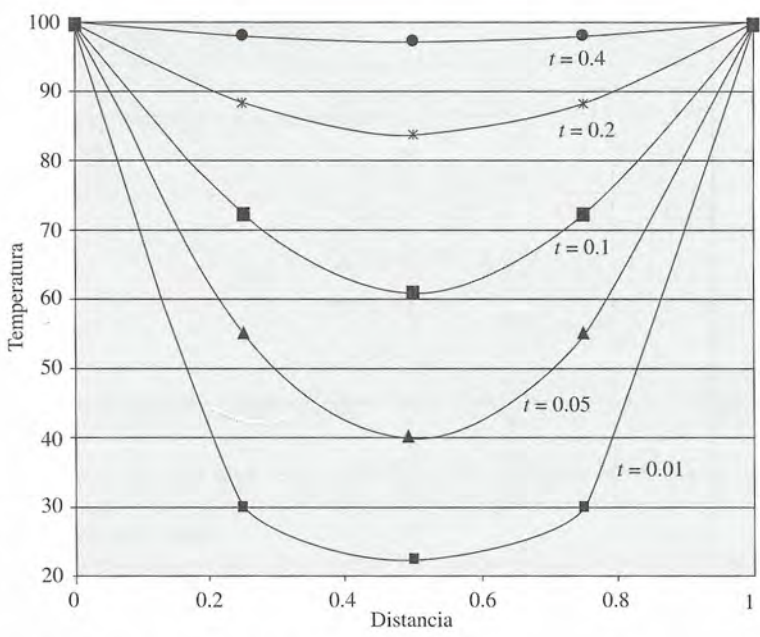
y resolver se obtiene:

$$T_{1,2} = 38.02, \quad T_{2,2} = 26.2, \quad T_{3,2} = 38.02,$$

que son las temperaturas correspondientes a  $t = 0.02$  h y a  $x = 0.25$ ,  $x = 0.5$ , y  $x = 0.75$  pies, respectivamente.

Al aproximar la EDP por diferencias divididas en la fila  $j + 1$  (véase Fig. 8.12) se obtiene el siguiente sistema

$$\begin{aligned} (1 + 2\lambda) T_{1,j+1} - \lambda T_{2,j+1} &= \lambda T_{0,j+1} + T_{1,j} \\ -\lambda T_{1,j+1} + (1 + 2\lambda) T_{2,j+1} - \lambda T_{3,j+1} &= T_{2,j} \\ -\lambda T_{2,j+1} + (1 + 2\lambda) T_{3,j+1} &= \lambda T_{4,j+1} + T_{3,j} \end{aligned}$$



**Figura 8.13**  
Distribución de temperatura en la barra a diferentes tiempos.

Hay que observar que en todos los casos el sistema por resolver tiene la misma matriz coeficiente, que es tridiagonal y simétrica.

Todo el sistema se soluciona estableciendo y resolviendo secuencialmente los sistemas de tres ecuaciones simultáneas para cada fila a partir de la segunda. Los resultados obtenidos con el **PROGRAMA 8.2** del CD se presentan en la tabla 8.3 y en la gráfica de la figura 8.13.

**Tabla 8.3** Resultados de la solución del PVF de conducción de calor en una barra metálica.

| <i>t</i> (horas) | <i>x</i> (pies) |       |       |       |      |
|------------------|-----------------|-------|-------|-------|------|
|                  | 0.00            | 0.25  | 0.50  | 0.75  | 1.00 |
| 0.00             | 60              | 20.00 | 20.00 | 20.00 | 60   |
| 0.01             | 100             | 29.99 | 22.43 | 29.99 | 100  |
| 0.02             | 100             | 38.02 | 26.20 | 38.02 | 100  |
| 0.03             | 100             | 44.64 | 30.67 | 44.64 | 100  |
| 0.04             | 100             | 50.23 | 35.41 | 50.23 | 100  |
| 0.05             | 100             | 55.04 | 40.17 | 55.04 | 100  |
| 0.06             | 100             | 59.25 | 44.80 | 59.25 | 100  |
| 0.07             | 100             | 62.97 | 49.20 | 62.97 | 100  |
| 0.08             | 100             | 66.29 | 53.35 | 66.29 | 100  |
| 0.09             | 100             | 69.28 | 57.21 | 69.28 | 100  |
| 0.10             | 100             | 71.97 | 60.79 | 71.97 | 100  |
| .                | .               | .     | .     | .     | .    |
| .                | .               | .     | .     | .     | .    |
| .                | .               | .     | .     | .     | .    |
| 0.20             | 100             | 88.62 | 83.91 | 88.62 | 100  |
| 0.40             | 100             | 98.10 | 97.32 | 98.10 | 100  |
| 0.60             | 100             | 99.68 | 99.55 | 99.68 | 100  |
| 0.80             | 100             | 99.95 | 99.93 | 99.95 | 100  |
| 1.00             | 100             | 99.99 | 99.99 | 99.99 | 100  |

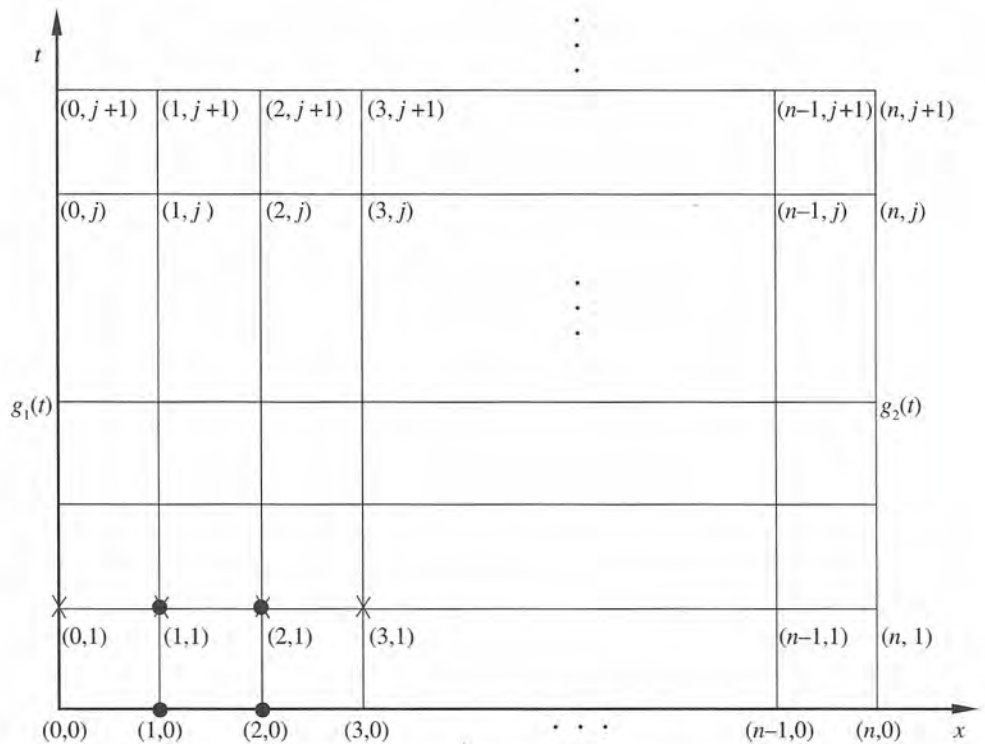


Figura 8.14

En general, si se divide la longitud de la barra en  $n$  subintervalos, o sea con  $n-1$  nodos interiores (véase Fig. 8.14), el sistema de  $n-1$  ecuaciones simultáneas con  $n-1$  incógnitas para la fila  $j+1$  queda

$$\begin{aligned}
 (1 + 2\lambda)T_{1,j+1} - \lambda T_{2,j+1} &= \lambda T_{0,j+1} + T_{1,j} \\
 -\lambda T_{1,j+1} + (1 + 2\lambda)T_{2,j+1} - \lambda T_{3,j+1} &= T_{2,j} \\
 -\lambda T_{2,j+1} + (1 + 2\lambda)T_{3,j+1} - \lambda T_{4,j+1} &= T_{3,j} \\
 &\vdots \\
 -\lambda T_{n-3,j+1} + (1 + 2\lambda)T_{n-2,j+1} - \lambda T_{n-1,j+1} &= T_{n-2,j} \\
 -\lambda T_{n-2,j+1} + (1 + 2\lambda)T_{n-1,j+1} &= T_{n-1,j} + \lambda T_{n,j+1}
 \end{aligned}$$

La solución de este sistema corresponde a las temperaturas en los puntos seleccionados de la barra a un tiempo  $(j+1)b$ .

Hay que observar que la simetría de la matriz coeficiente y su característica tridiagonal. Además, los elementos de esta matriz son constantes para cualquier fila (o tiempo) y son:

$$\begin{bmatrix}
 (1 + 2\lambda) & -\lambda & 0 & & 0 \\
 -\lambda & (1 + 2\lambda) & -\lambda & & \vdots \\
 0 & -\lambda & (1 + 2\lambda) & -\lambda & \vdots \\
 \vdots & & & & \vdots \\
 \vdots & & & & 0 \\
 0 & & \dots & -\lambda & (1 + 2\lambda) & -\lambda \\
 & & & 0 & -\lambda & (1 + 2\lambda)
 \end{bmatrix}$$

**ALGORITMO 8.2** Método implícito

Para aproximar la solución al problema de valor en la frontera

$$\text{PVF} \left\{ \begin{array}{l} \alpha \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t} \\ T(x, 0) = f(x), \quad 0 < x < x_F \\ T(0, t) = g_1(t) \\ T(x_F, t) = g_2(t) \end{array} \right. \quad t > 0$$

Proporcionar las funciones  $CI(X)$ ,  $CF1(T)$  y  $CF2(T)$  y los

**DATOS:** El número  $NX$  de puntos de la malla en el eje  $x$ , el número  $NT$  de puntos de la malla en el eje  $t$ , la longitud total  $XF$  del eje  $x$ , el tiempo máximo  $TF$  por considerar y el coeficiente  $ALFA$  de la derivada de segundo orden.

**RESULTADOS:** Los valores de la variable dependiente  $T$  a lo largo del eje  $x$  a distintos tiempos  $t$ :  $T$ .

- PASO 1. Realizar los paso 1 a 10 del algoritmo 8.1  
 PASO 2. Hacer  $I = 1$   
 PASO 3. Mientras  $I \leq NX - 2$ , repetir los pasos 4 a 7.  
     PASO 4. Hacer  $A(I) = -LAMBDA$   
     PASO 5. Hacer  $B(I) = 1 + 2 * LAMBDA$   
     PASO 6. Hacer  $C(I) = -LAMBDA$   
     PASO 7. Hacer  $I = I + 1$   
 PASO 8. Hacer  $J = 1$   
 PASO 9. Mientras  $J \leq NT$ , repetir los paso 10 a 24.  
     PASO 10. Hacer  $T(1) = CF1(DT * J)$   
     PASO 11. Hacer  $T(NX) = CF2(DT * J)$   
     PASO 12. Hacer  $I = 1$   
     PASO 13. Mientras  $I \leq NX - 2$ , repetir los paso 14 y 15.  
         PASO 14. Hacer  $D(I) = T(I + 1)$   
         PASO 15. Hacer  $I = I + 1$   
     PASO 16. Hacer  $D(1) = D(1) + LAMBDA * T(1)$   
     PASO 17. Hacer  $D(NX - 2) = D(NX - 2) + LAMBDA * T(NX)$   
     PASO 18. Realizar los pasos 1 a 12 del algoritmo 3.5 con  $N = NX - 2$   
     PASO 19. Hacer  $I = 1$   
     PASO 20. Mientras  $I \leq NX - 2$ , repetir los paso 21 y 22.  
         PASO 21. Hacer  $T(I + 1) = X(I)$   
         PASO 22. Hacer  $I = I + 1$   
     PASO 23. IMPRIMIR  $T$   
     PASO 24. Hacer  $J = J + 1$   
 PASO 25. TERMINAR.

**Ejemplo 8.2**

Resolver el ejemplo 8.1 con el método implícito.

**Solución**

A fin de comparar resultados con el método implícito, se divide la barra también en 8 segmentos,  $a = 0.125$  pies y se estudiará también el fenómeno de conducción de calor durante media hora con subintervalos en el tiempo de tamaño  $b = 0.005$  horas. De esto  $\lambda = 0.32$ .



El primer sistema lineal que permite calcular las temperaturas en la barra a  $t = 0.005$  horas es:

$$\begin{aligned}
 + 1.64 T_{1,1} - 0.32 T_{2,1} &= 51.1342 \\
 - 0.32 T_{1,1} + 1.64 T_{2,1} - 0.32 T_{3,1} &= 35.3553 \\
 - 0.32 T_{2,1} + 1.64 T_{3,1} - 0.32 T_{4,1} &= 46.1940 \\
 - 0.32 T_{3,1} + 1.64 T_{4,1} - 0.32 T_{5,1} &= 50.0000 \\
 - 0.32 T_{4,1} + 1.64 T_{5,1} - 0.32 T_{6,1} &= 46.1940 \\
 - 0.32 T_{5,1} + 1.64 T_{6,1} - 0.32 T_{7,1} &= 35.3553 \\
 - 0.32 T_{6,1} + 1.64 T_{7,1} &= 35.1342
 \end{aligned}$$

cuya solución por el algoritmo de Thomas es:

$$T_{1,1} = 38.56, T_{2,1} = 37.84, T_{3,1} = 44.90, T_{4,1} = 47.93, T_{5,1} = 44.50, T_{6,1} = 35.78, T_{7,1} = 28.41$$

El segundo sistema lineal que permite calcular las temperaturas en la barra a  $t = 0.005$  horas es:

$$\begin{aligned}
 + 1.64 T_{1,1} - 0.32 T_{2,1} &= 70.5636 \\
 - 0.32 T_{1,1} + 1.64 T_{2,1} - 0.32 T_{3,1} &= 37.8445 \\
 - 0.32 T_{2,1} + 1.64 T_{3,1} - 0.32 T_{4,1} &= 44.9041 \\
 - 0.32 T_{3,1} + 1.64 T_{4,1} - 0.32 T_{5,1} &= 47.9329 \\
 - 0.32 T_{4,1} + 1.64 T_{5,1} - 0.32 T_{6,1} &= 44.5021 \\
 - 0.32 T_{5,1} + 1.64 T_{6,1} - 0.32 T_{7,1} &= 35.7840 \\
 - 0.32 T_{6,1} + 1.64 T_{7,1} &= 44.4055
 \end{aligned}$$

cuya solución por el algoritmo de Thomas es:

$$T_{1,1} = 51.17, T_{2,1} = 41.76, T_{3,1} = 44.58, T_{4,1} = 46.40, T_{5,1} = 43.40, T_{6,1} = 36.98, T_{7,1} = 34.29$$

Usando el **PROGRAMA 8.2** del CD con las modificaciones correspondientes, se obtiene:

| $t$ (horas) | $x$ (pies) |       |       |       |       |       |       |       |       |
|-------------|------------|-------|-------|-------|-------|-------|-------|-------|-------|
|             | 0.000      | 0.125 | 0.250 | 0.375 | 0.500 | 0.625 | 0.750 | 0.875 | 1.000 |
| 0.000       | 50         | 19.13 | 35.36 | 46.19 | 50.00 | 46.19 | 35.56 | 19.13 | 25    |
| 0.005       | 100        | 38.56 | 37.84 | 44.90 | 47.93 | 44.50 | 35.78 | 28.41 | 50    |
| 0.010       | 100        | 51.17 | 41.76 | 44.58 | 46.40 | 43.40 | 36.98 | 34.29 | 50    |
| 0.015       | 100        | 59.67 | 45.89 | 45.00 | 45.42 | 42.81 | 38.35 | 38.15 | 50    |
| 0.020       | 100        | 65.60 | 49.74 | 45.92 | 44.97 | 42.62 | 39.65 | 40.75 | 50    |
| 0.025       | 100        | 69.89 | 53.17 | 47.15 | 44.96 | 42.72 | 40.82 | 42.57 | 50    |
| 0.030       | 100        | 73.08 | 56.15 | 48.54 | 45.28 | 43.05 | 41.85 | 43.88 | 50    |
| .           | .          | .     | .     | .     | .     | .     | .     | .     | .     |
| .           | .          | .     | .     | .     | .     | .     | .     | .     | .     |
| .           | .          | .     | .     | .     | .     | .     | .     | .     | .     |
| 0.050       | 100        | 80.31 | 64.65 | 54.36 | 48.48 | 45.60 | 45.06 | 46.73 | 50    |
| 0.100       | 100        | 86.83 | 74.90 | 65.13 | 57.94 | 53.31 | 50.87 | 50.02 | 50    |
| 0.150       | 100        | 89.62 | 79.90 | 71.37 | 64.38 | 59.00 | 55.07 | 52.24 | 50    |
| 0.200       | 100        | 91.21 | 82.82 | 75.14 | 68.40 | 62.66 | 57.85 | 53.73 | 50    |
| 0.300       | 100        | 92.77 | 85.70 | 78.89 | 72.45 | 66.39 | 60.70 | 55.27 | 50    |
| 0.400       | 100        | 93.37 | 86.80 | 80.34 | 74.02 | 67.84 | 61.80 | 55.87 | 50    |
| 0.500       | 100        | 93.60 | 87.23 | 80.90 | 74.62 | 68.40 | 62.23 | 56.10 | 50    |

Se deja al lector construir la gráfica de distribución de temperatura a lo largo de la barra para ciertos valores de  $t$  y hacer una discusión de resultados como se hizo en el ejemplo 8.1

## 8.4 Convergencia (método explícito), estabilidad y consistencia

### CONVERGENCIA

Hasta ahora no se ha analizado la importante pregunta de si los valores obtenidos aproximan “convenientemente” la solución del PVF en los nodos de la malla. En esta sección se contesta parcialmente ese punto.

El **error de discretización** se define en cada nodo como:

$$e = T - U,$$

donde  $U$  es la solución verdadera del PVF y  $T$  la aproximación obtenida con el esquema explícito.

Se dice que un esquema de diferencias es convergente si al hacer  $a = \Delta x \rightarrow 0$ ,  $b = \Delta t \rightarrow 0$  en la malla, el error de discretización  $e$  también tiende a cero. Con estas definiciones presentes se demuestra a continuación que una condición suficiente para convergencia del método explícito en la solución de:

$$\frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t} \quad (\text{adimensionalizando las variables } \lambda = 1)$$

es que  $0 < (\Delta t / \Delta x^2) < 0.5$ . Se aceptará que no se comenten errores de redondeo, lo cual es prácticamente imposible, pero aun así, este criterio de convergencia es de uso práctico.

Al expandir en serie de Taylor alrededor del nodo  $(i, j)$  la solución verdadera  $U$  (es la variable  $t$  solamente), se obtiene\*

$$U_{i,j+1} = U_{i,j} + \Delta t U_t + \frac{(\Delta t)^2}{2!} U_{tt} + O[(\Delta t)^3] \quad (8.32)$$

Al expandir  $U$  en la variable  $x$  adelante y atrás del nodo  $(i, j)$  se obtienen, respectivamente,

$$U_{i+1,j} = U_{i,j} + \Delta x U_x + \frac{\Delta x^2}{2!} U_{xx} + \frac{\Delta x^3}{3!} U_{xxx} + \frac{\Delta x^4}{4!} U_{xxxx} + \frac{\Delta x^5}{5!} U_{xxxxx} + O[(\Delta x)^6] \quad (8.33)$$

$$U_{i-1,j} = U_{i,j} - \Delta x U_x + \frac{\Delta x^2}{2!} U_{xx} - \frac{\Delta x^3}{3!} U_{xxx} + \frac{\Delta x^4}{4!} U_{xxxx} - \frac{\Delta x^5}{5!} U_{xxxxx} + O[(\Delta x)^6] \quad (8.34)$$

Hay que observar que en las ecuaciones 8.32 a 8.34 las derivadas se evalúan en el nodo  $(i, j)$ , cuyas coordenadas son  $x = i\Delta x$  y  $t = j\Delta t$ .

Con la suma de las ecuaciones 8.33 y 8.34 se obtiene:

$$U_{i+1,j} + U_{i-1,j} = 2U_{i,j} + \Delta x^2 U_{xx} + \frac{\Delta x^4}{12} U_{xxxx} + O[(\Delta x)^6] \quad (8.35)$$

\*  $U_t$  representa la primera derivada de  $U$  con respecto a  $t$ ,  $U_{tt}$  la segunda derivada de  $U$  con respecto a  $t$ , etcétera.

Al multiplicar por  $\lambda$

$$\lambda [U_{i+1,j} + U_{i-1,j}] = 2\lambda U_{i,j} + \Delta x^2 \lambda U_{xx} + \frac{\Delta x^4}{12} \lambda U_{xxxx} + O[(\Delta x)^6] \quad (8.36)$$

Se despeja  $U_{xx}$  y se sustituye  $\lambda$  con  $\Delta t/\Delta x^2$  en algunos términos y resulta:

$$U_{xx} = \frac{\lambda}{\Delta t} [U_{i+1,j} + U_{i-1,j}] - \frac{2\lambda}{\Delta t} U_{i,j} - \frac{\Delta x^2}{12} U_{xxxx} + \frac{\lambda}{\Delta t} O[(\Delta x)^6] \quad (8.37)$$

$U_t$  se despeja de la ecuación 8.32

$$U_t = \frac{1}{\Delta t} [U_{i,j+1} - U_{i,j} - \frac{\Delta t^2}{2!} U_{tt}] - \frac{O[(\Delta t)^3]}{\Delta t} \quad (8.38)$$

Al sustituir las ecuaciones 8.37 y 8.38 en la ecuación diferencial parcial,  $U_t = U_{xx}$

$$\begin{aligned} U_{i,j+1} - U_{i,j} - \frac{\Delta t^2}{2!} U_{tt} - O[(\Delta t)^3] = \\ \lambda U_{i+1,j} + \lambda U_{i-1,j} - 2\lambda U_{i,j} - \frac{\Delta x^2 \Delta t}{12} U_{xxxx} + \lambda O[(\Delta x)^6] \end{aligned} \quad (8.39)$$

Se despeja  $U_{i,j+1}$

$$\begin{aligned} U_{i,j+1} = \lambda U_{i-1,j} + (1 - 2\lambda) U_{i,j} + \lambda U_{i+1,j} - \frac{\Delta x^2 \Delta t}{12} U_{xxxx} \\ + \frac{\Delta t^2}{2!} U_{tt} + O[(\Delta t)^3] + \lambda O[(\Delta t)^6] \end{aligned} \quad (8.40)$$

Si se hace

$$\frac{Z_{i,j}}{\Delta t} = \frac{\Delta t}{2} U_{tt} - \frac{\Delta x^2}{12} U_{xxxx} + O[(\Delta t)^2] + O[(\Delta t)^4] \quad (8.41)$$

y se sustituye la ecuación 8.41 en la 8.40

$$U_{i,j+1} = \lambda U_{i-1,j} + (1 - 2\lambda) U_{i,j} + \lambda U_{i+1,j} + Z_{i,j} \quad (8.42)$$

Se resta del esquema explícito  $T_{i,j+1} = \lambda T_{i-1,j} + (1 - 2\lambda) T_{i,j} + \lambda T_{i+1,j}$  miembro a miembro la ecuación 8.42

$$\begin{aligned} T_{i,j+1} - U_{i,j+1} = \lambda(T_{i-1,j} - U_{i-1,j}) + (1 - 2\lambda)(T_{i,j} - U_{i,j}) \\ + \lambda(T_{i+1,j} - U_{i+1,j}) - Z_{i,j} \end{aligned} \quad (8.43)$$

Este desarrollo algebraico expresa el error de discretización  $e_{i,j+1} = (T_{i,j+1} - U_{i,j+1})$  en función de los errores en los nodos vecinos  $e_{i-1,j}$ ,  $e_{i,j}$  y  $e_{i+1,j}$  que se usan en el esquema explícito, o sea,

$$e_{i,j+1} = \lambda e_{i-1,j} + (1 - 2\lambda) e_{i,j} + \lambda e_{i+1,j} - Z_{i,j} \quad (8.44)$$

Supóngase ahora que  $0 < \lambda \leq 0.5$ , con lo que los coeficientes  $\lambda$  y  $(1 - 2\lambda)$  son **no negativos**. Si, por otro lado, se saca el valor absoluto en ambos miembros de la ecuación 8.44 y se aplica la desigualdad del triángulo, se obtiene:

$$|e_{i,j+1}| \leq \lambda |e_{i-1,j}| + (1 - 2\lambda) |e_{i,j}| + \lambda |e_{i+1,j}| + |-Z_{i,j}| \quad (8.45)$$



Si se llama  $e_{\text{máx}}(k)$  con  $0 \leq k \leq m$  la cota superior de  $|e_{i,k}|$  con  $1 \leq i \leq n-1$ , se denota por  $Z_{\text{máx}}(k)$  con  $k$  —igual que antes— la cota superior de  $|Z_{i,k}|$  con  $1 \leq i \leq n-1$ , y se sustituye  $e_{i-1,j}$ ,  $e_{i,j}$ ,  $e_{i+1,j}$ ,  $e_{i,j+1}$  y  $Z_{i,j}$  con sus respectivas cotas superiores  $e_{\text{máx}}(j)$ ,  $e_{\text{máx}}(j+1)$  y  $Z_{\text{máx}}(j)$ , simplificando se llega a

$$e_{\text{máx}}(j+1) \leq e_{\text{máx}}(j) + Z_{\text{máx}}(j) \quad (8.46)$$

Si se analiza esta desigualdad en un periodo  $0 \leq t \leq t_{\text{máx}} = t_m$  se tiene

$$\begin{aligned} e_{\text{máx}}(1) &\leq e_{\text{máx}}(0) + Z_{\text{máx}}(0) \\ e_{\text{máx}}(2) &\leq e_{\text{máx}}(1) + Z_{\text{máx}}(1) \\ e_{\text{máx}}(3) &\leq e_{\text{máx}}(2) + Z_{\text{máx}}(2) \\ &\vdots \\ e_{\text{máx}}(m) &\leq e_{\text{máx}}(m-1) + Z_{\text{máx}}(m-1) \end{aligned}$$

Al sustituir el término  $e_{\text{máx}}(1)$  de la segunda desigualdad con el lado derecho de la primera, aquélla permanece e incluso se refuerza, con lo cual queda

$$e_{\text{máx}}(2) \leq Z_{\text{máx}}(0) + Z_{\text{máx}}(1) + e_{\text{máx}}(0)$$

Válgase la consideración de que los valores iniciales son exactos,  $e_{\text{máx}}(0) = 0$ .

Este resultado se sustituye por el término  $e_{\text{máx}}(2)$  de la tercera desigualdad, con lo que

$$e_{\text{máx}}(3) \leq Z_{\text{máx}}(0) + Z_{\text{máx}}(1) + Z_{\text{máx}}(2)$$

Este procedimiento se repite hasta  $e_{\text{máx}}(m)$ ; por tanto

$$e_{\text{máx}}(m) \leq m Z_{\text{máx}}(m-1)$$

Se tiene

$$e_{\text{máx}}(m) \leq t_{\text{máx}} \left[ \frac{\Delta t}{2} U_{tt} - \frac{\Delta x^2}{12} U_{xxxx} + O[(\Delta t)^2] + O[(\Delta x)^4] \right]$$

recordando que  $t_{\text{máx}} = m \Delta t$  y la ecuación 8.41. De esta desigualdad se deduce que  $e_{\text{máx}}(m)$  tiende a cero si  $\Delta x$  y  $\Delta t$  tienden a cero; y ya que esta deducción se desarrolló para  $0 < \lambda \leq 0.5$ , la conclusión sólo será válida para estos valores de  $\lambda$ ; por ello, se constituye como una condición suficiente para convergencia —pero no necesaria— ya que ésta puede ocurrir por otras razones.

## ESTABILIDAD

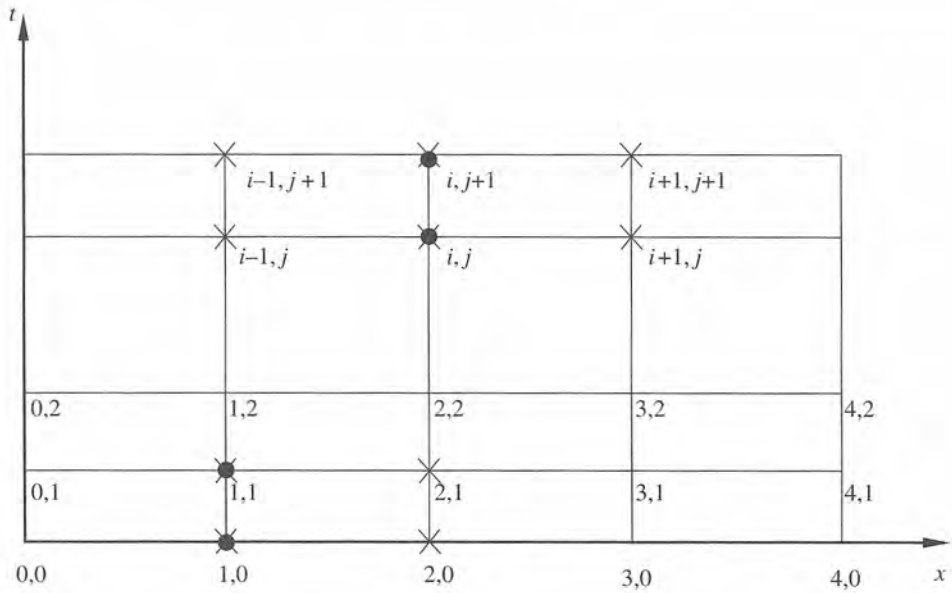
El concepto de estabilidad se refiere a la propiedad de una ecuación de diferencias particular (base de un algoritmo), y significa que cuando  $\Delta t \rightarrow 0$ , el error introducido por cualquier motivo (condiciones iniciales, frontera, redondeo, etc.) está acotado. Lo anterior no significa que la desviación entre la solución verdadera de cierta ecuación diferencial parcial y su aproximación con una ecuación de diferencias sea pequeña, ya que esto está determinado por el concepto de consistencia.

## CONSISTENCIA

Se dice que una ecuación de diferencias tiene consistencia cuando **solamente aproxima** la ecuación diferencial parcial que representa. Aunque esta propiedad parece cumplirse en todos los casos, no es así para algunos esquemas iterativos; por ejemplo, el algoritmo explícito de Dufort-Frankel no es consistente en ciertas circunstancias.\*

\* Dufort, E. C. y Frankel, S.P. "Stability Conditions in the Numerical Treatment of Parabolic Differential Equations", *Math Tables Aids Comput.*, 7, (1953) pp. 135-152.





**Figura 8.15**  
Nodos usados  
en el método  
de Crank-  
Nicholson.

## 8.5 Método de Crank-Nicholson

Además de los métodos vistos para resolver los PVF de las secciones 8.2 y 8.3, existen otros métodos de solución con diferencias. Entre éstos uno de los más importantes por su estabilidad incondicional y alto orden de convergencia\* es el algoritmo de Crank-Nicholson.

Este método consiste en combinar las aproximaciones de  $\partial T / \partial t$  con diferencias hacia delante apoyándose en la fila  $j$  y la aproximación con diferencias hacia atrás apoyándose en la fila  $j+1$ , con lo que se obtiene un algoritmo implícito. Por ejemplo, al aproximar  $\partial T / \partial t$  en el nodo  $(i, j)$  con diferencias hacia delante y de  $\partial^2 T / \partial x^2$  con diferencias centrales (véase Fig. 8.15) se obtiene:

$$\frac{T_{i,j+1} - T_{i,j}}{\Delta t} = \alpha \frac{T_{i-1,j} - 2T_{i,j} + T_{i+1,j}}{\Delta x^2} \quad (8.47)$$

Al aproximar  $\partial T / \partial t$  en el nodo  $(i, j+1)$  con diferencias hacia atrás y a  $\partial^2 T / \partial x^2$  con diferencias centrales (véase Fig. 8.15) se llega a

$$\frac{T_{i,j+1} - T_{i,j}}{\Delta t} = \alpha \frac{T_{i-1,j+1} - 2T_{i,j+1} + T_{i+1,j+1}}{\Delta x^2} \quad (8.48)$$

Luego de sumar las ecuaciones 8.47 y 8.48 y reorganizar resulta

$$T_{i,j+1} - T_{i,j} = \frac{\lambda}{2} [T_{i-1,j} - 2T_{i,j} + T_{i+1,j} + T_{i-1,j+1} - 2T_{i,j+1} + T_{i+1,j+1}] \quad (8.49)$$

que es el algoritmo de Crank-Nicholson.

Si este algoritmo (ecuación 8.49) se aplica a los nodos  $(1,0)$  y  $(1,1)$ , o sea  $i = 1, j = 0$  (véase Figura 8.15), se tiene

\* Isaacson, E. y Keller, H.B. *Analysis of Numerical Methods*, John Wiley and Sons, Nueva York, 1966 pp. 390, 392, 1082, 1088.

$$T_{1,1} - T_{1,0} = \frac{\lambda}{2} [T_{0,0} - 2T_{1,0} + T_{2,0} + T_{0,1} - 2T_{1,1} + T_{2,1}] \quad (8.50)$$

donde los nodos (0,0), (1,0), (2,0) y (0,1) son conocidos a partir de las condiciones inicial y de frontera; en cambio, los nodos (1,1) y (2,1) son incógnitas. Al reorganizar la ecuación 8.50 queda

$$(1 + \lambda) T_{1,1} - \frac{\lambda}{2} T_{2,1} = (1 - \lambda) T_{1,0} + \frac{\lambda}{2} [T_{0,0} + T_{0,1} + T_{2,0}] \quad (8.51)$$

Al aplicar el mismo algoritmo (Ec. 8.49) a los nodos (2,0) y (2,1); es decir,  $i = 2, j = 0$ , se tiene

$$T_{2,1} - T_{2,0} = \frac{\lambda}{2} [T_{1,0} - 2T_{2,0} + T_{3,0} + T_{1,1} - 2T_{2,1} + T_{3,1}] \quad (8.52)$$

donde las incógnitas son  $T_{1,1}$ ,  $T_{2,1}$  y  $T_{3,1}$ , ya que los demás nodos están dados por la condición inicial. Al reorganizar resulta

$$-\frac{\lambda}{2} T_{1,1} + (1 + \lambda) T_{2,1} - \frac{\lambda}{2} T_{3,1} = \frac{\lambda}{2} T_{1,0} + (1 - \lambda) T_{2,0} + \frac{\lambda}{2} T_{3,0} \quad (8.53)$$

Análogamente, al aplicar la ecuación 8.49 a los nodos (3,0) y (3,1), es decir,  $i = 3, j = 0$ , queda

$$T_{3,1} - T_{3,0} = \frac{\lambda}{2} [T_{2,0} - 2T_{3,0} + T_{4,0} + T_{2,1} - 2T_{3,1} + T_{4,1}] \quad (8.54)$$

donde los nodos desconocidos son solamente (2,1) y (3,1), ya que los otros son conocidos por la condición inicial.

La ecuación 8.54 se reorganiza y queda

$$-\frac{\lambda}{2} T_{2,1} + (1 + \lambda) T_{3,1} = \frac{\lambda}{2} [T_{2,0} + T_{4,0} + T_{4,1}] + (1 - \lambda) T_{3,0} \quad (8.55)$$

Las ecuaciones 8.51, 8.53 y 8.55 forman un sistema cuya solución es la temperatura  $T$  en los nodos (1,1), (2,1) y (3,1); o sea,

$$\begin{aligned} (1 + \lambda) T_{1,1} - \frac{\lambda}{2} T_{2,1} &= (1 - \lambda) T_{1,0} + \frac{\lambda}{2} [T_{0,0} + T_{0,1} + T_{2,0}] \\ -\frac{\lambda}{2} T_{1,1} + (1 + \lambda) T_{2,1} - \frac{\lambda}{2} T_{3,1} &= \frac{\lambda}{2} [T_{1,0} + T_{3,0}] + (1 - \lambda) T_{2,0} \\ -\frac{\lambda}{2} T_{2,1} + (1 + \lambda) T_{3,1} &= \frac{\lambda}{2} [T_{2,0} + T_{4,0} + T_{4,1}] + (1 - \lambda) T_{3,0} \end{aligned} \quad (8.56)$$

Una vez resuelto el sistema de ecuaciones 8.56 se puede seguir el mismo procedimiento, pero aplicado ahora en los nodos (1,1) (1,2); (2,1), (2,2) y (3,1), (3,2), con lo cual resulta

$$\begin{aligned} (1 + \lambda) T_{1,2} - \frac{\lambda}{2} T_{2,2} &= (1 - \lambda) T_{1,1} + \frac{\lambda}{2} [T_{0,1} + T_{0,2} + T_{2,1}] \\ -\frac{\lambda}{2} T_{1,2} + (1 + \lambda) T_{2,2} - \frac{\lambda}{2} T_{3,2} &= \frac{\lambda}{2} [T_{1,1} + T_{3,1}] + (1 - \lambda) T_{2,1} \\ -\frac{\lambda}{2} T_{2,2} + (1 + \lambda) T_{3,2} &= \frac{\lambda}{2} [T_{2,1} + T_{4,1} + T_{4,2}] + (1 - \lambda) T_{3,1} \end{aligned}$$

cuya solución proporciona las temperaturas de los nodos interiores de la segunda fila; o sea,  $t = 2\Delta t$ . Este procedimiento se repite un número  $m$  de veces, hasta obtener las temperaturas en ciertos puntos de la barra a lo largo del tiempo, hasta un  $t_{\text{máx}} = m\Delta t$ .

Si en lugar de dividir la barra en cuatro subintervalos se dividiera en  $n$  subintervalos, se tendrían  $n-1$  nodos interiores, a los que al aplicarse la ecuación 8.49 como en el caso anterior (cuatro subintervalos) se generaría un sistema de  $n-1$  ecuaciones con  $n-1$  incógnitas:  $T_{1,1}, T_{2,1}, T_{3,1}, \dots, T_{n-1,1}$  (para la primera fila); o sea

$$\begin{aligned}
 (1 + \lambda) T_{1,1} - \frac{\lambda}{2} T_{2,1} &= (1 - \lambda) T_{1,0} + \frac{\lambda}{2} [T_{0,0} + T_{0,1} + T_{2,0}] \\
 -\frac{\lambda}{2} T_{1,1} + (1 + \lambda) T_{2,1} - \frac{\lambda}{2} T_{3,1} &= \frac{\lambda}{2} [T_{1,0} + T_{3,0}] + (1 - \lambda) T_{2,0} \\
 &\vdots \\
 -\frac{\lambda}{2} T_{n-3,1} + (1 + \lambda) T_{n-2,1} - \frac{\lambda}{2} T_{n-1,1} &= \frac{\lambda}{2} [T_{n-3,0} + T_{n-1,0}] + (1 - \lambda) T_{n-2,0} \\
 -\frac{\lambda}{2} T_{n-2,1} + (1 + \lambda) T_{n-1,1} &= \frac{\lambda}{2} [T_{n-2,0} + T_{n,0} + T_{n,1}] + (1 - \lambda) T_{n-1,0}
 \end{aligned} \tag{8.57}$$

Este procedimiento se aplica en las filas  $j$  y  $j+1$  para tener

$$\begin{aligned}
 (1 + \lambda) T_{1,j+1} - \frac{\lambda}{2} T_{2,j+1} &= (1 - \lambda) T_{1,j} + \frac{\lambda}{2} [T_{0,j} + T_{0,j+1} + T_{2,j}] \\
 -\frac{\lambda}{2} T_{1,j+1} + (1 + \lambda) T_{2,j+1} - \frac{\lambda}{2} T_{3,j+1} &= \frac{\lambda}{2} [T_{1,j} + T_{3,j}] + (1 - \lambda) T_{2,j} \\
 &\vdots \\
 -\frac{\lambda}{2} T_{n-3,j+1} + (1 + \lambda) T_{n-2,j+1} - \frac{\lambda}{2} T_{n-1,j+1} &= \frac{\lambda}{2} [T_{n-3,j} + T_{n-1,j}] + (1 - \lambda) T_{n-2,j} \\
 -\frac{\lambda}{2} T_{n-2,j+1} + (1 + \lambda) T_{n-1,j+1} &= \frac{\lambda}{2} [T_{n-2,j} + T_{n,j} + T_{n,j+1}] + (1 - \lambda) T_{n-1,j}
 \end{aligned}$$

que en notación matricial queda

$$A \mathbf{t}^{(j+1)} = B \mathbf{t}^{(j)} + \mathbf{c}$$

donde:

$$A = \begin{bmatrix}
 (1 + \lambda) & -\frac{\lambda}{2} & 0 & \dots & 0 \\
 -\frac{\lambda}{2} & (1 + \lambda) & -\frac{\lambda}{2} & & \vdots \\
 0 & & & & 0 \\
 \vdots & & & & \vdots \\
 0 & \dots & 0 & -\frac{\lambda}{2} & (1 + \lambda) \\
 & & & -\frac{\lambda}{2} & (1 + \lambda)
 \end{bmatrix}$$

$$\mathbf{t}^{(j+1)} = [T_{1,j+1} \quad T_{2,j+1} \quad T_{3,j+1} \quad \dots \quad T_{n-1,j+1}]^T$$



$$B = \begin{bmatrix} (1-\lambda) & \frac{\lambda}{2} & 0 & \cdots & 0 \\ \frac{\lambda}{2} & (1-\lambda) & \frac{\lambda}{2} & & \vdots \\ 0 & & & & 0 \\ \vdots & & & & \vdots \\ \vdots & & \frac{\lambda}{2} & (1-\lambda) & \frac{\lambda}{2} \\ 0 & \cdots & 0 & \frac{\lambda}{2} & (1-\lambda) \end{bmatrix}$$

$$\mathbf{t}^{(j)} = [T_{1,j} \ T_{2,j} \ T_{3,j} \ \cdots \ T_{n-1,j}]^T$$

y

$$\mathbf{c} = \left[ \frac{\lambda}{2} (T_{0,j} + T_{0,j+1}) \ 0 \ \cdots \ 0 \ \frac{\lambda}{2} (T_{n,j} + T_{n,j+1}) \right]^T$$

**Ejemplo 8.3**

Resuelva el siguiente problema por el método de Crank-Nicholson

$$\text{PVF} \begin{cases} \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \\ T(x, 0) = 20^\circ\text{F} \\ T(0, t) = 100^\circ\text{F} \\ T(L, t) = 100^\circ\text{F} \end{cases}$$

$$\alpha = 1 \text{ pie}^2 / \text{hr}$$

$$L = 1 \text{ pie}$$

$$t_{\text{máx}} = 1 \text{ hr}$$

**Solución**

Al dividir la longitud de la barra en cuatro subintervalos (véase Fig. 8.15), el primer sistema de ecuaciones por resolver, ya sustituidos los datos, es



$$\begin{bmatrix} 1.16 & -0.08 & 0.00 \\ -0.08 & 1.16 & -0.08 \\ 0.00 & -0.08 & 1.16 \end{bmatrix} \begin{bmatrix} T_{1,1} \\ T_{2,1} \\ T_{3,1} \end{bmatrix} = \begin{bmatrix} 31.2 \\ 20 \\ 31.2 \end{bmatrix}$$

Este sistema se resuelve por alguno de los métodos del capítulo 3 y se obtiene

$$T_{1,1} = 28.36 \quad T_{2,1} = 21.15 \quad T_{3,1} = 28.36$$

temperaturas que corresponden a un tiempo  $t = 0.01$  horas.

Para calcular las temperaturas de la segunda línea se conserva la matriz coeficiente y sólo se varía el vector de términos independientes; o sea

$$\begin{bmatrix} 1.16 & -0.08 & 0.00 \\ -0.08 & 1.16 & -0.08 \\ 0.00 & -0.08 & 1.16 \end{bmatrix} \begin{bmatrix} T_{1,2} \\ T_{2,2} \\ T_{3,2} \end{bmatrix} = \begin{bmatrix} 41.5144 \\ 22.3036 \\ 41.5144 \end{bmatrix}$$

El sistema se resuelve para obtener

$T_{1,2} = 37.47 \qquad T_{2,2} = 24.40 \qquad T_{3,2} = 37.47$

temperaturas que corresponden a un tiempo  $t = 0.02$  horas. Al continuar este procedimiento con el PROGRAMA 8.3 del CD se obtienen los resultados de la tabla 8.4.

**Tabla 8.4** Resultados de la solución del ejemplo 8.2. Se usó  $t = 0.01$  constante y sólo se muestran algunos de los resultados.

| $t$ (horas) | $x$ pies |       |       |       |     |
|-------------|----------|-------|-------|-------|-----|
|             | 0.0      | 0.25  | 0.5   | 0.75  | 1.0 |
| 0.00        | 60       | 20.00 | 20.00 | 20.00 | 60  |
| 0.01        | 100      | 28.36 | 21.15 | 28.36 | 100 |
| 0.02        | 100      | 37.47 | 24.40 | 37.47 | 100 |
| 0.03        | 100      | 44.61 | 28.99 | 44.61 | 100 |
| 0.04        | 100      | 50.45 | 34.10 | 50.45 | 100 |
| 0.05        | 100      | 55.38 | 39.29 | 55.38 | 100 |
| 0.06        | 100      | 59.67 | 44.32 | 59.67 | 100 |
| 0.07        | 100      | 63.44 | 49.07 | 63.44 | 100 |
| 0.08        | 100      | 66.81 | 53.50 | 66.81 | 100 |
| 0.09        | 100      | 69.83 | 57.59 | 69.83 | 100 |
| 0.10        | 100      | 72.56 | 61.44 | 72.56 | 100 |
| 0.20        | 100      | 89.28 | 84.84 | 89.28 | 100 |
| 0.40        | 100      | 98.36 | 97.68 | 98.36 | 100 |
| 0.60        | 100      | 99.75 | 99.64 | 99.75 | 100 |
| 0.80        | 100      | 99.96 | 99.95 | 99.96 | 100 |
| 1.00        | 100      | 99.99 | 99.99 | 99.99 | 100 |

Los resultados obtenidos con el método de Crank-Nicholson son en general un promedio de los resultados de los métodos explícito e implícito; esto puede explicarse con base en que el método de Crank-Nicholson combina ambos.

Enseguida se presenta un algoritmo para este método

**ALGORITMO 8.3** Método de Crank-Nicholson

Para aproximar la solución al

PVF  $\left\{ \begin{array}{l} \text{EDP } \alpha \frac{\partial^2 T}{\partial x^2} = \frac{\partial T}{\partial t} \\ \text{CI } T(x, 0) = f(x), \quad 0 \leq x \leq x_F \\ \text{CF1 } T(0, t) = g_1(t) \\ \text{CF2 } T(x_F, t) = g_2(t) \end{array} \right. \quad t > 0$

proporcionar las funciones CI(X), CF1(T) y CF2(T) y los

**DATOS:** El número  $NX$  de puntos de la malla en el eje  $x$ , el número  $NT$  de puntos de la malla en el eje  $t$ , la longitud total  $XF$  a considerar del eje  $x$ , el tiempo máximo  $TF$  por considerar y el coeficiente  $ALFA$  de la derivada de segundo orden.

**RESULTADOS:** Los valores de la variable dependiente  $T$  a lo largo del eje  $x$  a distintos tiempos  $t$ :  $T$ .

PASO 1. Realizar los pasos 1 a 10 del algoritmo 8.1

PASO 2. Hacer  $I = 1$

PASO 3. Mientras  $I \leq NX - 2$ , repetir los pasos 4 a 7.

PASO 4. Hacer  $A(I) = LAMBDA$

PASO 5. Hacer  $B(I) = -2 - 2 * LAMBDA$

PASO 6. Hacer  $C(I) = LAMBDA$

PASO 7. Hacer  $I = I + 1$

PASO 8. Realizar los pasos 8 a 24 del algoritmo 8.2 con los siguientes cambios:

En el paso 15 hacer  $D(I) = -LAMBDA * T(I) - (2 - 2 * LAMBDA) * T(I+1) - LAMBDA * T(I+2)$

En el paso 17 hacer  $D(1) = D(1) - LAMBDA * T(1)$

En el paso 18 hacer  $D(NX-2) = D(NX-2) - LAMBDA * T(NX)$

PASO 9. TERMINAR.

## 8.6 Otros métodos para resolver el problema de conducción de calor unidimensional

### MÉTODO DE RICHARDSON

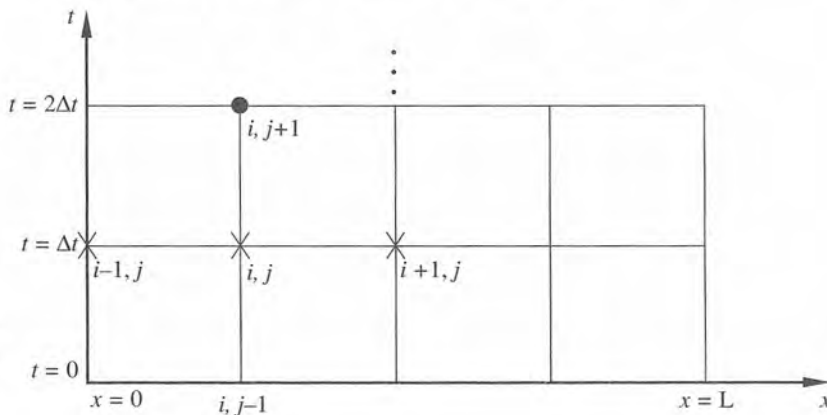
Este método usa diferencias divididas centrales para aproximar  $\partial T / \partial t$  en la ecuación de conducción. De acuerdo con la malla de la figura 8.16 se tiene

$$\frac{T_{i,j+1} - T_{i,j-1}}{2\Delta t} = \alpha \frac{T_{i-1,j} - 2T_{i,j} + T_{i+1,j}}{\Delta x^2} \quad (8.58)$$

Hay que observar que si se conocen las dos primera filas (la primera podría ser la condición inicial y la segunda se calcula por alguno de los métodos de las secciones anteriores), el método resulta explícito en el nodo  $(i, j + 1)$ ; o sea,

$$T_{i,j+1} = 2\lambda [T_{i-1,j} - 2T_{i,j} + T_{i+1,j}] + T_{i,j-1} \quad (8.59)$$

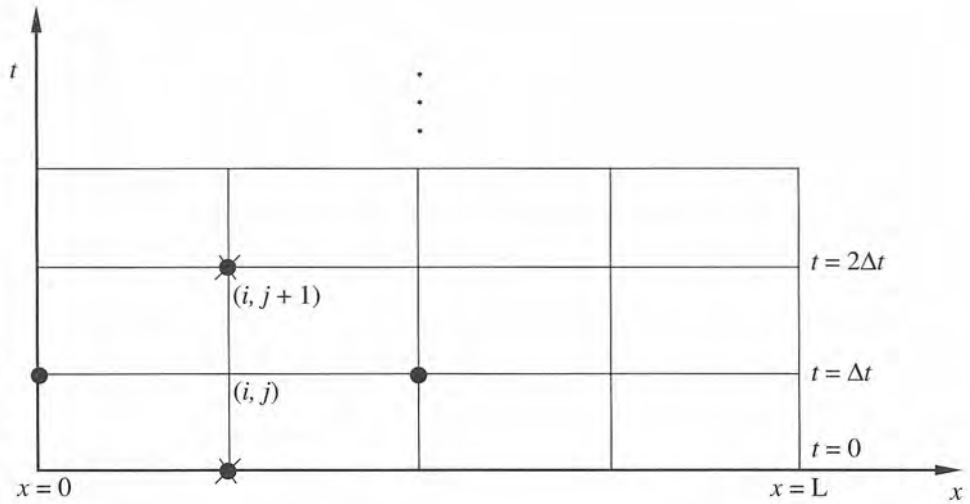
con lo que pueden calcularse la tercera, cuarta, etc., filas.



**Figura 8.16**  
Nodos usados  
en el método  
de Richardson.



**Figura 8.17**  
Nodos usados  
en el método  
de Dufort-  
Frankel.



### MÉTODO DE DUFORT-FRANKEL

Young y Gregory\* demuestran que el método de Richardson es poco satisfactorio, ya que presenta considerables problemas de estabilidad; si embargo, sustituyendo  $T_{i,j}$  con  $(T_{i,j+1} + T_{i,j-1})/2$  en la ecuación 8.58 se obtiene el método de Dufort-Frankel con mejores propiedades de estabilidad.

$$\frac{T_{i,j+1} - T_{i,j-1}}{2\Delta t} = \alpha \frac{T_{i-1,j} - T_{i,j-1} - T_{i,j+1} + T_{i+1,j}}{\Delta x^2}$$

Como el método de Richardson, si se conocen dos filas el algoritmo resulta explícito para el cálculo de las temperaturas de la siguiente fila; es decir,

$$T_{i,j+1} = \frac{2\lambda}{1+2\lambda} [T_{i-1,j} + T_{i+1,j}] + \left(\frac{1-2\lambda}{1+2\lambda}\right) T_{i,j-1} \quad (8.60)$$

#### Ejemplo 8.4

Mediante el método de Dufort-Frankel, resuelva el ejemplo 8.3 con los mismos valores para  $\Delta x$ ,  $\Delta t$  y  $\alpha$ .

#### Solución

La primera fila está dada por las condiciones iniciales y para la segunda fila ( $t = 0.01$ ) se tomarán los resultados obtenidos con el método implícito (véase la tabla 8.3).

Se aplica la ecuación 8.60 para conocer  $T_{i,j+1} = T_{1,2}$  y se obtiene:

$$T_{1,2} = \frac{2\lambda}{1+2\lambda} [T_{0,1} + T_{2,1}] + \left(\frac{1-2\lambda}{1+2\lambda}\right) T_{1,0} \quad (8.60)$$

Al sustituir los valores  $\lambda = 0.16$ ,  $T_{0,1} = 100$ ,  $T_{2,1} = 22.43$  y  $T_{1,0} = 20$  se tiene

$$T_{1,2} = \frac{2(0.16)}{1+2(0.16)} [100 + 22.43] + \frac{1-2(0.16)}{1+2(0.16)} [20] = 39.98$$

\* Young, D.M. y Gregory, R.T. *A Survey of Numerical Mathematics*. Vol. II Addison Wesley (1973); pp. 1084-1086.

Con el cálculo del siguiente punto  $T_{i,j+1} = T_{2,2}$  queda

$$T_{2,2} = \frac{2\lambda}{1+2\lambda} [T_{1,1} + T_{3,1}] + \left(\frac{1-2\lambda}{1+2\lambda}\right) T_{2,0}$$

y al sustituir valores se obtiene  $T_{2,2} = 24.84$

El algoritmo se aplica en la misma forma para las filas siguientes. Los resultados se presentan en la tabla 8.5.

**Tabla 8.5** Resultados del ejemplo 8.3.

| $t$ (h) | $x$ pies |       |       |       |      |
|---------|----------|-------|-------|-------|------|
|         | 0.0      | 0.25  | 0.5   | 0.75  | 1.00 |
| 0.00    | 60       | 20.00 | 20.00 | 20.00 | 60   |
| 0.01    | 100      | 29.99 | 22.43 | 29.99 | 100  |
| 0.02    | 100      | 39.98 | 24.84 | 39.98 | 100  |
| 0.04    | 100      | 52.34 | 34.96 | 52.34 | 100  |
| 0.06    | 100      | 61.22 | 45.29 | 61.22 | 100  |
| 0.08    | 100      | 68.14 | 54.46 | 68.14 | 100  |
| 0.10    | 100      | 73.72 | 62.25 | 73.72 | 100  |
| 0.20    | 100      | 89.85 | 85.38 | 89.85 | 100  |
| 0.40    | 100      | 98.48 | 97.81 | 98.48 | 100  |
| 0.60    | 100      | 99.77 | 99.67 | 99.97 | 100  |
| 0.80    | 100      | 99.97 | 99.95 | 99.97 | 100  |
| 1.00    | 100      | 99.99 | 99.99 | 99.99 | 100  |

## 8.7 Solución de la ecuación de onda unidimensional

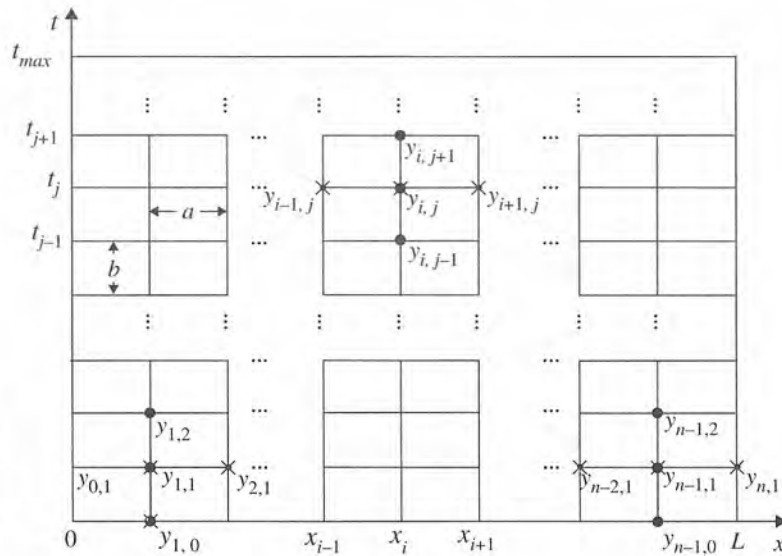
Se puede aproximar la ecuación de onda unidimensional obtenida en la sección 8.1 con ecuaciones de diferencias. Sea entonces el problema de valores a la frontera a resolver

$$\left\{ \begin{array}{ll} \text{EDP: } \frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2} \\ \text{CI1: } y(x, 0) = f(x), & 0 < x < L \\ \text{CI2: } \left. \frac{\partial y}{\partial t} \right|_{(x, 0)} = g(x), & 0 < x < L \\ \text{CF1: } y(0, t) = 0, & t > 0 \\ \text{CF2: } y(L, t) = 0, & t > 0 \end{array} \right.$$

donde la posición original de la cuerda está dada por CI 1 (condición inicial 1), y la velocidad inicial que se le imprime a la cuerda por CI2 ( $\partial y / \partial t = 0$  en el caso de que la cuerda simplemente se suelte).

Para obtener un método de diferencias finitas que resuelva el PVF anterior, empezaremos construyendo una malla en el dominio de interés  $0 < x < L$ ,  $0 < t < t_{\max}^*$  en  $n$  y  $m$

\*  $t_{\max}$  es una cota superior para el tiempo que se usará para detener los cálculos.



**Figura 8.18**  
Representación  
de algunos  
nodos de la  
malla para la  
cuerda vibrante.

subintervalos de tamaño  $a = \Delta x = L/n$  y  $b = \Delta t = t_{\max}/m$ , respectivamente, como se muestra en la figura 8.18.

Dado que la EDP se cumple en todo el dominio, también es cierta para el nodo  $(x_i, t_j)$ , esto es:

$$\left. \frac{\partial^2 y}{\partial t^2} \right|_{(x_i, t_j)} = c^2 \left. \frac{\partial^2 y}{\partial x^2} \right|_{(x_i, t_j)}$$

Usando diferencias centrales para las derivadas parciales, se tiene:

$$\frac{y_{i,j-1} - 2y_{i,j} + y_{i,j+1}}{b^2} = c^2 \frac{y_{i-1,j} - 2y_{i,j} + y_{i+1,j}}{a^2}$$

Se ha remplazado  $y(x_i, t_j)$  con  $y_{i,j}$  para simplificar la notación.

Los nodos marcados con punto negro (•) en la figura 8.18 son los nodos usados para aproximar  $\partial^2 y / \partial t^2$ , y los marcados con cruz (×) se emplean a fin de aproximar  $\partial^2 y / \partial x^2$ .

Si se hace ahora  $\lambda = bc/a$ , se puede escribir la última ecuación como

$$y_{i,j-1} - 2y_{i,j} + y_{i,j+1} = \lambda^2 y_{i-1,j} - 2\lambda^2 y_{i,j} + \lambda^2 y_{i+1,j}$$

Despejando  $y_{i,j+1}$  (el punto en el tiempo más avanzado)

$$y_{i,j+1} = 2(1 - \lambda^2)y_{i,j} + \lambda^2(y_{i+1,j} + y_{i-1,j}) - y_{i,j-1} \quad (8.61)$$

En la frontera izquierda y derecha (los extremos de la cuerda), a cualquier tiempo se tiene:  $y_{0,j} = y_{n,j} = 0$ , para cada  $j = 1, 2, \dots, m$  dadas por las condiciones de frontera CF1 y CF2. Por otro lado, la condición inicial CI1 permite tener  $y_{i,0} = f(x_i)$ , para cada  $i = 1, 2, 3, \dots, n-1$ .

Para concretar, consideremos el nodo  $(i, j) = (1, 1)$  donde la ecuación 8.61 queda

$$y_{1,2} = 2(1 - \lambda^2)y_{1,1} + \lambda^2(y_{2,1} + y_{0,1}) - y_{1,0}$$

donde, como se dijo antes  $y_{0,1} = 0$ ,  $y_{1,0} = f(x_1)$  y se desconoce  $y_{1,1}$ ,  $y_{2,1}$ , y obviamente  $y_{1,2}$ . A fin de reunir más ecuaciones apliquemos 8.61 al nodo  $(i, j) = (2, 1)$ , de donde

$$y_{2,2} = 2(1 - \lambda^2)y_{2,1} + \lambda^2(y_{3,1} + y_{1,1}) - y_{2,0}$$



que adiciona dos incógnitas más. Continuando de esta manera se tendría una última ecuación para el nodo  $(i, j) = (n-1, 1)$

$$y_{n-1,2} = 2(1-\lambda^2)y_{n-1,1} + \lambda^2(y_{n,1} + y_{n-2,1}) - y_{n-1,0}$$

donde  $y_{n-1,0} = f(x_{n-1})$  y  $y_{n,1} = 0$

Agrupando las ecuaciones anteriores y recurriendo a la notación matricial se tiene

$$\begin{bmatrix} y_{1,2} \\ y_{2,2} \\ \vdots \\ y_{n-2,2} \\ y_{n-1,2} \end{bmatrix} = \begin{bmatrix} 2(1-\lambda^2) & \lambda^2 & 0 & \dots & 0 \\ \lambda^2 & 2(1-\lambda^2) & \lambda^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \lambda^2 & 2(1-\lambda^2) & \lambda^2 \\ 0 & \dots & 0 & \lambda^2 & 2(1-\lambda^2) \end{bmatrix} \begin{bmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{n-2,1} \\ y_{n-1,1} \end{bmatrix} - \begin{bmatrix} y_{1,0} \\ y_{2,0} \\ \vdots \\ y_{n-2,0} \\ y_{n-1,0} \end{bmatrix} \quad (8.62)$$

El vector  $[y_{1,0}, y_{2,0}, \dots, y_{n-2,0}, y_{n-1,0}]^T$  está dado por la condición inicial CI1, de modo que para obtener el lado izquierdo de 8.62 se requiere una aproximación del vector  $[y_{1,1}, y_{2,1}, \dots, y_{n-2,1}, y_{n-1,1}]^T$  que se puede obtener de la condición inicial CI2:  $\left. \frac{\partial y}{\partial t} \right|_{(x,0)} = g(x)$ . Un método\*

consiste en sustituir a  $\partial y / \partial t$  por una aproximación en diferencias hacia delante en el nodo  $(i, 0)$ ,  $\left. \frac{\partial y}{\partial t} \right|_{(i,0)} \approx \frac{y_{i,1} - y_{i,0}}{b}$  para  $i = 1, 2, \dots, n-1$

Despejando  $y_{i,1}$  y sustituyendo  $\left. \frac{\partial y}{\partial t} \right|_{(i,0)}$  por  $g(x_i)$  se llega a la aproximación buscada

$$y_{i,1} \approx y_{i,0} + bg(x_i) \quad \text{para } i = 1, 2, \dots, n-1 \quad (8.63)$$

Una vez obtenida esta aproximación se puede obtener el lado izquierdo de la ecuación 8.62 operando matricialmente en el lado derecho\*\*; los resultados proporcionan el desplazamiento en ciertos puntos de la cuerda al tiempo  $t_2$ . Para obtener el desplazamiento de los puntos mencionados al tiempo  $t_3$  simplemente tendría que operarse matricialmente en el lado derecho de la ecuación

$$\begin{bmatrix} y_{1,3} \\ y_{2,3} \\ \vdots \\ y_{n-2,3} \\ y_{n-1,3} \end{bmatrix} = \begin{bmatrix} 2(1-\lambda^2) & \lambda^2 & 0 & \dots & 0 \\ \lambda^2 & 2(1-\lambda^2) & \lambda^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \lambda^2 & 2(1-\lambda^2) & \lambda^2 \\ 0 & \dots & 0 & \lambda^2 & 2(1-\lambda^2) \end{bmatrix} \begin{bmatrix} y_{1,2} \\ y_{2,2} \\ \vdots \\ y_{n-2,2} \\ y_{n-1,2} \end{bmatrix} - \begin{bmatrix} y_{1,1} \\ y_{2,1} \\ \vdots \\ y_{n-2,1} \\ y_{n-1,1} \end{bmatrix}$$

y así sucesivamente hasta obtener los desplazamientos en tiempo  $t_{\max}$ . Al igual que como se vio en el método explícito (sección 8.4) hay condiciones de convergencia que se pueden establecer mediante el parámetro  $\lambda$ . Para el caso que nos ocupa conviene observar que se cumpla  $0 < \lambda < 1$ .

\* En el problema 8.19 se da otro método de aproximación, que utiliza diferencias centrales.

\*\* Nótese que no se trata de resolver un sistema de ecuaciones lineales.



Ejemplo 8.5

Resolver el

EDP:  $\frac{\partial^2 y}{\partial t^2} = c^2 \frac{\partial^2 y}{\partial x^2}$

CI1:  $y(x, 0) = \text{sen}(\pi x), \quad 0 < x < 1$

PVF { CI2:  $\frac{\partial y}{\partial t} \Big|_{(x, 0)} = 2\pi \text{sen}(2\pi x) \quad 0 < x < 1$

CF1:  $y(0, t) = 0, \quad t > 0$

CF2:  $y(1, t) = 0, \quad t > 0$

Nótese que para simplificar se ha adimensionado con lo que  $c^2 = 1$  y  $L = 1$ . Considere  $t_{\text{max}} = 1$ .

Solución



Se divide la longitud de la cuerda en 10 subintervalos, esto es  $a = 0.1$  y el tiempo máximo en 100 subintervalos, con lo que  $b = 0.01$  y  $\lambda = bc/a = 0.01(1)/0.1=0.1$

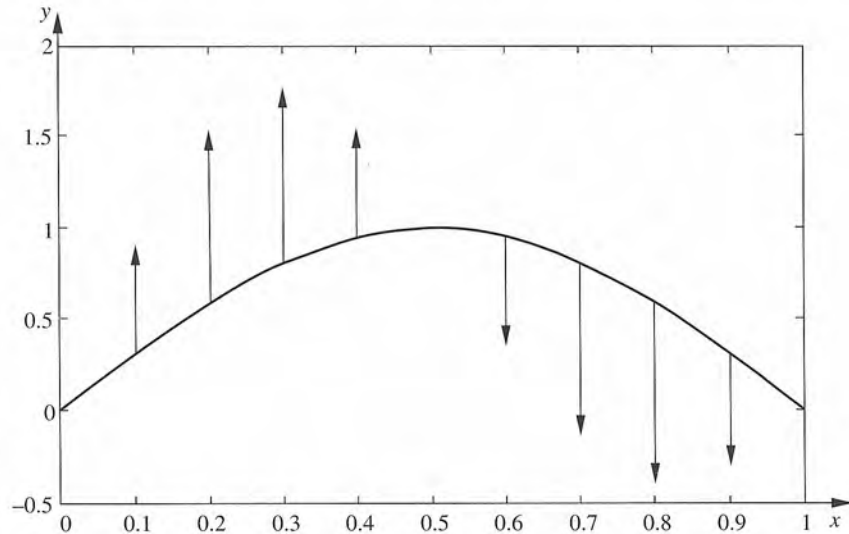
La posición inicial de la cuerda  $y_{i,0}$ ,  $i = 1, 2, \dots, 9$  dada por CI1 es:

| $x$                 | 0.1      | 0.2      | 0.3      | 0.4      | 0.5      | 0.6      | 0.7      | 0.8      | 0.9      |
|---------------------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| $\text{sen}(\pi x)$ | 0.309017 | 0.587785 | 0.809017 | 0.951057 | 1.000000 | 0.951057 | 0.809017 | 0.587785 | 0.309017 |

y la velocidad inicial  $\partial y/\partial t = g(x_i)$ ,  $i = 1, 2, \dots, 9$  que se le imprime a la cuerda dada por CI2 en tales puntos es:

| $x$                      | 0.1      | 0.2      | 0.3      | 0.4      | 0.5   | 0.6       | 0.7       | 0.8       | 0.9       |
|--------------------------|----------|----------|----------|----------|-------|-----------|-----------|-----------|-----------|
| $2\pi \text{sen}(\pi x)$ | 3.693164 | 5.975664 | 5.975664 | 3.693164 | 0.000 | -3.693164 | -5.975664 | -5.975664 | -3.693164 |

En la gráfica de la figura 8.19 se representa la CI1 por la curva y la CI2 por las flechas, cuya longitud es proporcional a la magnitud de la velocidad.



**Figura 8.19**  
Representación gráfica de las condiciones iniciales del problema.

A1 aproximar el vector  $[y_{1,1}, y_{2,1} \dots y_{n-2,1} y_{n-1,1}]^T$  con la ecuación 8.63

$$\begin{bmatrix} y_{1,1} \\ y_{2,1} \\ y_{3,1} \\ y_{4,1} \\ y_{5,1} \\ y_{6,1} \\ y_{7,1} \\ y_{8,1} \\ y_{9,1} \end{bmatrix} = \begin{bmatrix} y_{1,0} \\ y_{2,0} \\ y_{3,0} \\ y_{4,0} \\ y_{5,0} \\ y_{6,0} \\ y_{7,0} \\ y_{8,0} \\ y_{9,0} \end{bmatrix} + b \begin{bmatrix} g(x_1) \\ g(x_2) \\ g(x_3) \\ g(x_4) \\ g(x_5) \\ g(x_6) \\ g(x_7) \\ g(x_8) \\ g(x_9) \end{bmatrix} = \begin{bmatrix} 0.309017 \\ 0.587785 \\ 0.809017 \\ 0.951057 \\ 1.000000 \\ 0.951057 \\ 0.809017 \\ 0.587785 \\ 0.309017 \end{bmatrix} + 0.01 \begin{bmatrix} 3.693164 \\ 5.975664 \\ 5.975664 \\ 3.693164 \\ 0.000000 \\ -3.693164 \\ -5.975664 \\ -5.975664 \\ -3.693164 \end{bmatrix} = \begin{bmatrix} 0.345949 \\ 0.647542 \\ 0.868774 \\ 0.987988 \\ 1.000000 \\ 0.914125 \\ 0.749260 \\ 0.528029 \\ 0.272085 \end{bmatrix}$$

Con esto se tiene

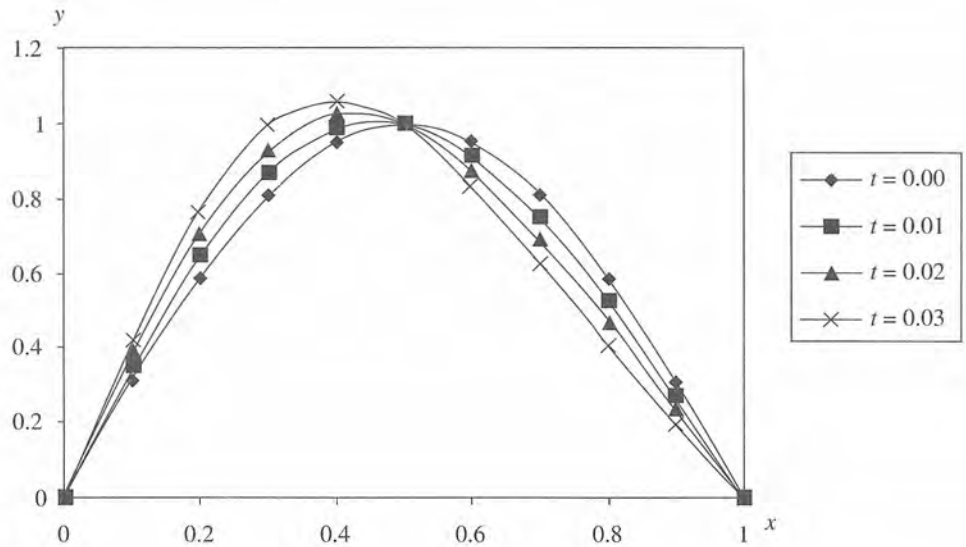
$$\begin{bmatrix} y_{1,2} \\ y_{2,2} \\ y_{3,2} \\ y_{4,2} \\ y_{5,2} \\ y_{6,2} \\ y_{7,2} \\ y_{8,2} \\ y_{9,2} \end{bmatrix} = \begin{bmatrix} 1.98 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.01 & 1.98 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.01 & 1.98 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 1.98 & 0.01 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.01 & 1.98 & 0.01 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.01 & 1.98 & 0.01 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.01 & 1.98 & 0.01 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 1.98 & 0.01 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 1.98 & 0.01 \end{bmatrix} \begin{bmatrix} 0.345949 \\ 0.647542 \\ 0.868774 \\ 0.987988 \\ 1.000000 \\ 0.914125 \\ 0.749260 \\ 0.528029 \\ 0.272085 \end{bmatrix} - \begin{bmatrix} 0.309017 \\ 0.587785 \\ 0.809017 \\ 0.951057 \\ 1.000000 \\ 0.951057 \\ 0.809017 \\ 0.587785 \\ 0.309017 \end{bmatrix}$$

de donde

$$\begin{bmatrix} y_{1,2} \\ y_{2,2} \\ y_{3,2} \\ y_{4,2} \\ y_{5,2} \\ y_{6,2} \\ y_{7,2} \\ y_{8,2} \\ y_{9,2} \end{bmatrix} = \begin{bmatrix} 0.382437 \\ 0.706495 \\ 0.927511 \\ 1.023847 \\ 0.999021 \\ 0.876403 \\ 0.688939 \\ 0.467926 \\ 0.234992 \end{bmatrix}$$

cuya gráfica se muestra en la figura 8.20.

**Figura 8.20**  
Posición  
de la cuerda  
a  $t = 0.00$ ,  
 $t = 0.01$ ,  $t = 0.02$   
y  $t = 0.03$ .



El cálculo para el tiempo  $t = 0.03$  se obtiene de realizar las operaciones que se indican enseguida

$$\begin{bmatrix} y_{1,3} \\ y_{2,3} \\ y_{3,3} \\ y_{4,3} \\ y_{5,3} \\ y_{6,3} \\ y_{7,3} \\ y_{8,3} \\ y_{9,3} \end{bmatrix} = \begin{bmatrix} 1.98 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.01 & 1.98 & 0.01 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.01 & 1.98 & 0.01 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.01 & 1.98 & 0.01 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.01 & 1.98 & 0.01 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.01 & 1.98 & 0.01 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.01 & 1.98 & 0.01 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 1.98 & 0.01 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.01 & 1.98 \end{bmatrix} \begin{bmatrix} 0.382437 \\ 0.706495 \\ 0.927511 \\ 1.023847 \\ 0.999021 \\ 0.876403 \\ 0.688939 \\ 0.467926 \\ 0.234992 \end{bmatrix} - \begin{bmatrix} 0.345949 \\ 0.647542 \\ 0.868774 \\ 0.987988 \\ 1.000000 \\ 0.914125 \\ 0.749260 \\ 0.528029 \\ 0.272085 \end{bmatrix}$$

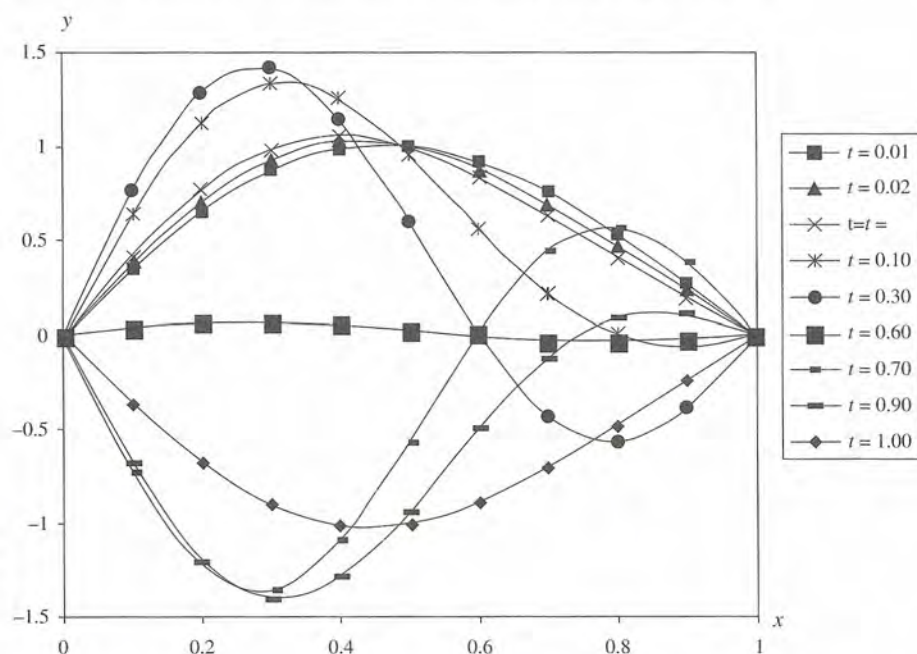
cuyo resultado es  $[0.418342 \ 0.764418 \ 0.985001 \ 1.058494 \ 0.997064 \ 0.838033 \ 0.628283 \ 0.407704 \ 0.197878]^T$ , cuya gráfica se muestra en la figura 8.20

Este procedimiento se repite para obtener la posición de la cuerda a tiempo  $t = 0.03, 0.04, \dots, t_{\max}$ . De las posiciones obtenidas para los 100 tiempos, en la tabla 8.6 se muestran sólo algunos para facilitar su presentación. En el CD se encuentra el **PROGRAMA 8.4** en Visual Basic que permite observar el movimiento de la cuerda en modo rápido y en modo cámara lenta a fin de que el lector pueda analizar este fenómeno que ha desempeñado un papel importantísimo en el desarrollo de la ciencia y la tecnología.

**Tabla 8.6** Algunos resultados del ejemplo 8.5.

| $t$  | $x$   |        |        |        |        |        |        |        |        |        |       |
|------|-------|--------|--------|--------|--------|--------|--------|--------|--------|--------|-------|
|      | 0.0   | 0.1    | 0.2    | 0.3    | 0.4    | 0.5    | 0.6    | 0.7    | 0.8    | 0.9    | 1.0   |
| 0.00 | 0.000 | 0.309  | 0.588  | 0.809  | 0.951  | 1.000  | 0.951  | 0.809  | 0.588  | 0.309  | 0.000 |
| 0.01 | 0.000 | 0.346  | 0.648  | 0.869  | 0.988  | 1.000  | 0.914  | 0.749  | 0.528  | 0.272  | 0.000 |
| 0.02 | 0.000 | 0.382  | 0.706  | 0.928  | 1.024  | 0.999  | 0.876  | 0.689  | 0.468  | 0.235  | 0.000 |
| 0.03 | 0.000 | 0.418  | 0.764  | 0.985  | 1.058  | 0.997  | 0.838  | 0.628  | 0.408  | 0.198  | 0.000 |
| 0.10 | 0.000 | 0.642  | 1.123  | 1.334  | 1.256  | 0.956  | 0.563  | 0.213  | 0.001  | -0.051 | 0.000 |
| 0.30 | 0.000 | 0.760  | 1.283  | 1.417  | 1.148  | 0.604  | -0.000 | -0.440 | -0.574 | -0.387 | 0.000 |
| 0.50 | 0.000 | 0.037  | 0.062  | 0.067  | 0.051  | 0.022  | -0.009 | -0.031 | -0.036 | -0.024 | 0.000 |
| 0.70 | 0.000 | -0.729 | -1.230 | -1.356 | -1.094 | -0.568 | 0.014  | 0.437  | 0.563  | 0.379  | 0.000 |
| 0.90 | 0.000 | -0.685 | -1.192 | -1.400 | -1.291 | -0.942 | -0.502 | -0.125 | 0.084  | 0.103  | 0.000 |
| 1.00 | 0.000 | -0.370 | -0.686 | -0.907 | -1.012 | -1.000 | -0.890 | -0.710 | -0.489 | -0.248 | 0.000 |

Las gráficas correspondientes a estos tiempos se dan en la figura 8.21



**Figura 8.21**  
Gráfica de las  
posiciones de  
la cuerda a  
diferentes  
tiempos.

## 8.8 Tipos de condiciones frontera en procesos físicos y tratamientos de condiciones frontera irregulares

Dependiendo de las características del proceso físico modelado y de las circunstancias que rodean al proceso de estudio, se tendrán en general tres tipos de condiciones frontera en un PVF.



### 1. CONDICIONES DE DIRICHLET

Se dan estas condiciones cuando la variable dependiente es conocida en todos los puntos frontera. Los ejemplos de las secciones anteriores tienen este tipo de condiciones frontera.

### 2. CONDICIONES DE NEUMANN

Cuando se conoce la derivada de la variable dependiente en los puntos frontera, se dice que se tienen las condiciones de Neumann. Por ejemplo, el problema de conducción de calor de la barra con condiciones de este tipo, quedaría formulado así.

$$\text{PVF} \left\{ \begin{array}{l} \text{EDP: } \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \\ \text{CI: } T(x, 0) = f(x), \quad 0 \leq x \leq L \\ \text{CF1: } \left. \frac{dT}{dx} \right|_{x=0} = g_1(t) \\ \quad \quad \quad t > 0 \\ \text{CF2: } \left. \frac{dT}{dx} \right|_{x=L} = g_2(t) \end{array} \right.$$

Estas condiciones pueden obtenerse físicamente, por ejemplo aislando térmicamente una frontera, ya que en este caso

$$\left. \frac{dT}{dx} \right|_{x=L} = 0,$$

es decir, no habría cambio de temperatura en la frontera. O bien si se tiene una frontera en contacto con un fluido (que puede ser aire), la ley de enfriamiento de Newton proporcionaría esta condición

$$\left. \frac{dT}{dx} \right|_{x=L} = h(T - T_0),$$

donde  $h$  es el coeficiente de transmisión de calor y  $T_0$  la temperatura del fluido.

### 3. CONDICIONES COMBINADAS

Esta condición aparece cuando se tiene una combinación de las dos anteriores. De nuevo, el problema de conducción de calor en la barra quedaría formulado con este tipo de condiciones así

$$\text{PVF} \left\{ \begin{array}{l} \text{EDP: } \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \\ \text{CI: } T(x, 0) = f(x), \quad 0 \leq x \leq L \\ \text{CF1: } \left. \frac{dT}{dx} \right|_{x=0} = g_1(t) \\ \quad \quad \quad t > 0 \\ \text{CF2: } T(L, t) = g_2(t) \end{array} \right.$$

En los ejercicios al final de capítulo, se resuelven problemas con condiciones de Neumann y combinadas.

### FRONTERAS IRREGULARES

Según la geometría del sistema, se pueden tener fronteras irregulares; esto es, casos como el de la figura 8.22.

Si se tienen, por ejemplo, las condiciones frontera de Dirichlet, los valores de la variable dependiente en C y D son conocidos; por tanto, la aproximación de la variable dependiente en el punto P puede hacerse con una interpolación. El caso más simple es una interpolación lineal entre los puntos A y C o entre B y D.

Para la interpolación entre los puntos A y C sería

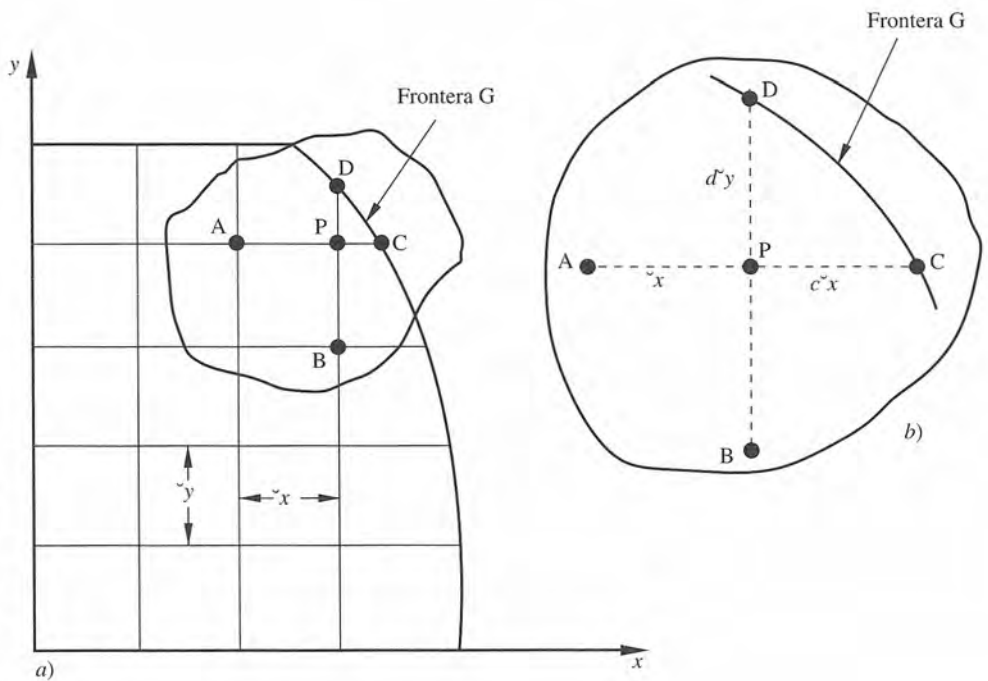
$$\frac{T_C - T_A}{\Delta x + c\Delta y} = \frac{T_P - T_A}{\Delta x}, \text{ de donde } T_P = \frac{T_C - T_A}{1 + c} + T_A \quad (8.64)$$

donde  $0 < c < 1$ .

Para la interpolación entre los puntos B y D.

$$\frac{T_D - T_B}{\Delta y + d\Delta x} = \frac{T_P - T_B}{\Delta y}, \text{ de donde } T_P = \frac{T_D - T_B}{1 + d} + T_B \quad (8.65)$$

Si se quisiera una aproximación mayor de  $T_P$ , cabe promediar los valores obtenidos por medio de las ecuaciones 8.64 y 8.65, o se cierra la malla (con lo que se aumentan los cálculos) y se usa alguna de las ecuaciones 8.64 u 8.65 o bien se toman TC o TD como aproximación de TP, según la que esté más cerca.



**Figura 8.22**

a) Malla sobre un dominio con frontera irregular.  
b) Ampliación de la región con puntos frontera D y C.

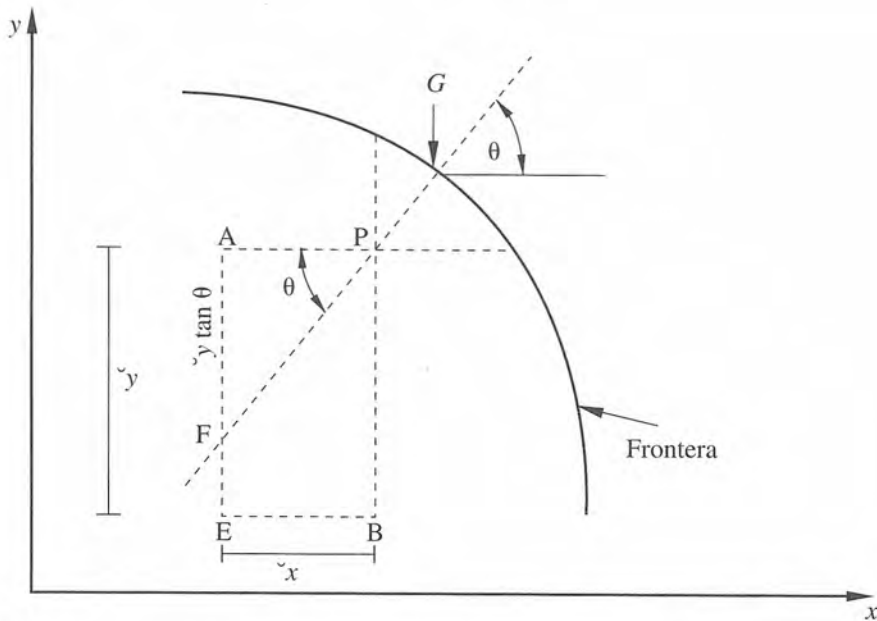


Figura 8.23

Cuando en los puntos de la frontera irregular G (véase Fig. 8.23) se conoce

$$\left. \frac{\partial T}{\partial N} \right|_G$$

en vez de  $T$ , donde  $N$  es el vector normal a la frontera (condiciones de Neumann), el problema de estimar el valor de los puntos cercanos a la frontera se torna un poco más difícil. Supóngase que se tiene una rejilla como en la figura 8.23. Ya que se conoce

$$\left. \frac{\partial T}{\partial N} \right|_G$$

este valor se puede igualar según

$$\left. \frac{\partial T}{\partial N} \right|_G = \frac{T_P - T_F}{\overline{FP}}, \text{ de donde } T_P = \left. \frac{\partial T}{\partial N} \right|_G \overline{FP} + T_F \quad (8.66)$$

Por construcción de la malla

$$\overline{FP} = \Delta x / \cos \theta \quad (8.67)$$

y también

$$\frac{T_E - T_A}{\Delta y} = \frac{T_F - T_A}{\Delta y \operatorname{tg} \theta}, \text{ de donde } T_F = (T_E - T_A) \operatorname{tg} \theta + T_A \quad (8.68)$$

Se sustituyen las ecuaciones 8.67 y 8.68 en la ecuación 8.66

$$T_P = \frac{\Delta x}{\cos \theta} \left. \frac{\partial T}{\partial N} \right|_G + (T_E - T_A) \operatorname{tg} \theta + T_A$$

En los problemas por resolver (al final del capítulo) se pide determinar  $T_P$  cuando el punto  $F$  cae entre los puntos  $E$  y  $B$  (véase Fig. 8.24).

Por último, si se tienen condiciones frontera combinadas, se aplica alguno de los tratamientos anteriores a cada punto frontera, según corresponda.

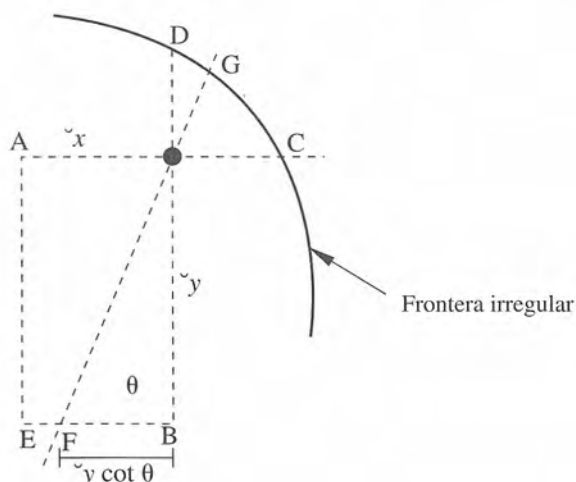


Figura 8.24

## Ejercicios

8.1 Confirme que las siguientes ecuaciones diferenciales parciales

a)  $\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} = 0$  ecuación de Laplace

b)  $\frac{\partial^2 U}{\partial x^2} = \frac{\partial^2 U}{\partial t^2}$  ecuación de onda

c)  $\frac{\partial^2 U}{\partial x^2} = \frac{\partial U}{\partial t}$  ecuación de difusión

son elíptica, hiperbólica y parabólica, respectivamente, en cualquier punto donde  $T$  y  $U$  estén definidas.

**Solución** a) La función solución  $T$  es —en este caso— función de  $x$  y  $y$  solamente; esto es,

$$T = T(x, y)$$

Al identificar los coeficientes  $A$ ,  $B$  y  $C$  de la ecuación de Laplace con los del modelo general (Ec. 8.1), se tiene

$$A = 1, \quad B = 0, \quad C = 1$$

Hay que observar que los tres coeficientes son constantes y, por tanto, independientes del punto  $(x, y)$  donde se desee establecer su clasificación. Así pues, en un punto cualquiera  $(x, y)$  donde  $T$  esté definida, se ve que

$$0^2 - 4(1)(1) < 0$$

por lo que la **ecuación es elíptica** en todo el dominio de definición de  $T$ .

b) De la misma manera que en (a), aunque ahora con  $U$  como función de  $x$  y  $t$ , se tiene

$$A = 1, \quad B = 0, \quad C = -1$$

y para un punto cualquiera  $(x, t)$  donde  $U$  esté definida

$$0^2 - 4(1)(-1) = 4 > 0$$

por lo que la ecuación de onda es **hiperbólica** en el punto  $(x, t)$  dado.



c) En este caso para un punto  $(x, t)$  donde  $U$  esté definida

$$A = 1, \quad B = 0, \quad C = 0$$

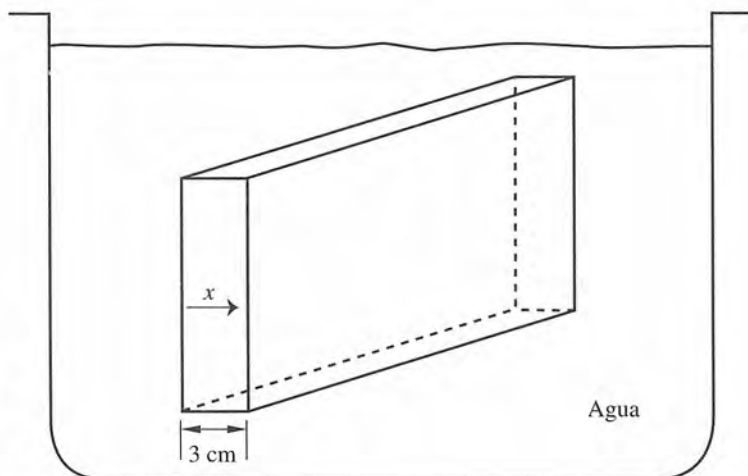
y

$$0^2 - 4(1)(0) = 0$$

por lo que la ecuación de difusión es **parabólica** en dicho punto  $(x, t)$ .



- 8.2** Una loza de gel de agar contiene una concentración uniforme de urea de  $2 \times 10^{-4}$  gmol/cm<sup>3</sup>; la loza tiene 3 cm de espesor (véase Fig. 8.25). Determine la concentración de urea en la parte central de la loza después de 2, 4, 6 y 8 horas de inmersión en agua (la urea es soluble en agua).



**Figura 8.25**  
Loza de agar sumergida en agua.

**Solución** El modelo matemático que permite establecer la concentración está dado por

$$\frac{\partial C}{\partial t} = \mathcal{D} \frac{\partial^2 C}{\partial x^2}$$

donde:

$C$  es la concentración de urea en la loza

$t$  el tiempo

$x$  la distancia

$\mathcal{D}$  la constante de difusividad (equivalente a  $\alpha$  en el fenómeno de conducción de calor).

Por el problema se sabe que  $C = 2 \times 10^{-4}$  gmol/cm<sup>3</sup>, que es la condición inicial (concentración inicial de la urea en la loza).

Por otro lado, se puede establecer que

$$\begin{aligned} C(0, t) &= 0 \\ t &> 0 \\ C(1, t) &= 0 \end{aligned}$$

lo cual físicamente significa que al sumergirse la loza en el agua, la urea de la superficie se disuelve de inmediato y la concentración de las caras (fronteras de la loza) es cero cualquier tiempo después.

El problema de valores en la frontera queda formulado

$$\text{PVP} \left\{ \begin{array}{l} \text{EDP: } \frac{\partial C}{\partial t} = \mathcal{D} \frac{\partial^2 C}{\partial x^2} \\ \text{CI: } C(x, 0) = 2 \times 10^{-4}, 0 \leq x \leq L \\ \text{CF1: } C(0, t) = 0 \\ \text{CF2: } C(1, t) = 0 \end{array} \right. \quad t > 0$$

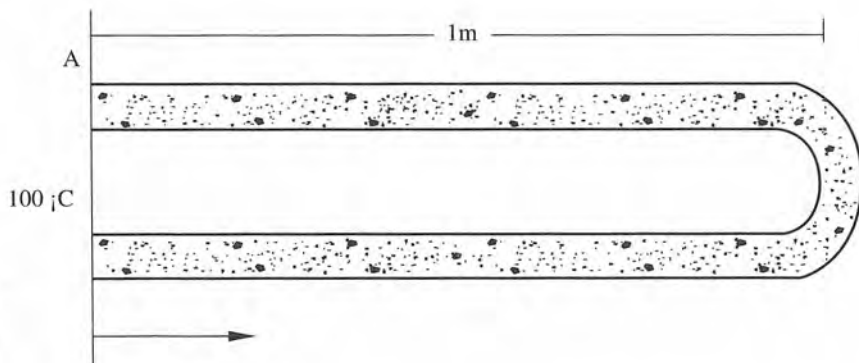
Si se toma  $\mathcal{D} = 1.7 \times 10^{-2} \text{ cm}^2/\text{h}$  y se aplica el **PROGRAMA 8.3** del CD, se obtienen los resultados siguientes para  $x = 1.5 \text{ cm}$  (el centro de la loza) transcurridas 2, 4, 6 y 8 horas, con  $\Delta x = 0.3$  y  $\Delta t = 0.01$ .

| $t \text{ (h)}$ | $x \text{ (cm)}$ |          |          |          |          |          |
|-----------------|------------------|----------|----------|----------|----------|----------|
|                 | 0.00             | 0.30     | 0.60     | 0.90     | 1.20     | 1.50     |
| 0.0             | 0.000100         | 0.000200 | 0.000200 | 0.000200 | 0.000200 | 0.000200 |
| 2.0             | 0.000000         | 0.000146 | 0.000191 | 0.000199 | 0.000200 | 0.000200 |
| 4.0             | 0.000000         | 0.000117 | 0.000176 | 0.000195 | 0.000199 | 0.000200 |
| 6.0             | 0.000000         | 0.000099 | 0.000162 | 0.000188 | 0.000197 | 0.000199 |
| 8.0             | 0.000000         | 0.000088 | 0.000149 | 0.000181 | 0.000194 | 0.000197 |



**8.3** Calcule la distribución de temperatura  $T(x, t)$  en una barra cilíndrica de vidrio y aislada térmicamente excepto en el plano A (véase Fig. 8.26). Inicialmente la barra está a  $20^\circ\text{C}$  y en el instante cero se ajusta con el plano A una placa cuya temperatura es de  $100^\circ\text{C}$  y permanece constante durante el tiempo de estudio (tres horas). La barra es lo suficientemente delgada como para despreciar la distribución de temperatura radial y se sabe que para el material vidrio  $\alpha = 1.23 \times 10^{-3} \text{ m}^2/\text{h}$ .

**Solución** Este problema es semejante al del ejemplo resuelto al inicio de la sección 8.3 con la diferencia de que un extremo está aislado, lo que modifica la condición frontera correspondiente. Un aislamiento térmico “ideal” significa que no hay flujo de calor en dirección alguna y matemáticamente se expresa



**Figura 8.26**  
Barra cilíndrica  
de vidrio  
aislada.

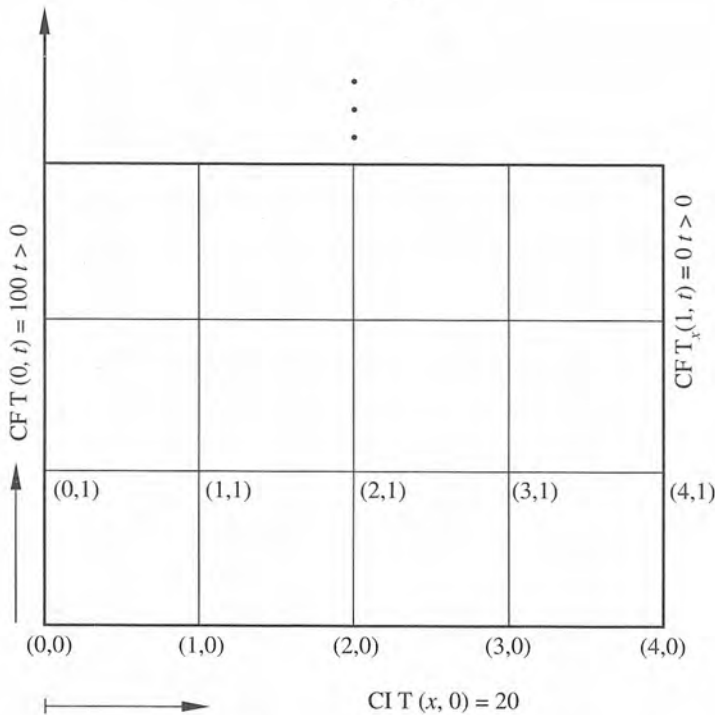
$$\left. \frac{\partial T}{\partial x} \right|_{x=1} = T_x = 0$$

Por lo anterior, el problema de valor en la frontera con condiciones frontera combinadas queda formulado por

$$\text{PVF} \left\{ \begin{array}{l} \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} \\ T(x, 0) = 20^\circ \text{C}, \quad 0 \leq x < L \\ T(0, t) = 100^\circ \text{C} \\ T_x(1, t) = 0 \end{array} \quad t > 0 \right.$$

Con el empleo del método explícito y la selección de  $\Delta x = 0.25$  y  $\Delta t = 0.1$ ,

$$\lambda = \frac{\alpha \Delta t}{\Delta x^2} = 1.968 \times 10^{-3}$$



**Figura 8.27**

la malla queda como se ilustra en la figura 8.27, y se tiene

Para  $i = 1, j = 0$  en la ecuación 8.27

$$T_{1,1} = 0.001968(60) + (1 - 2(0.001968))20 + 0.001968(20) = 20.08$$

donde  $T_{0,0}$  se aproxima con la media aritmética de los valores límites de  $T(0, t)$  cuando  $t \rightarrow 0$  y  $T(x, 0)$  cuando  $x \rightarrow 0$ , que en este caso es la media de 100 y  $20^\circ \text{C}$ .

Al aplicar el mismo algoritmo al nodo (2,1) se tiene

$$T_{2,1} = 0.001968(20) + (1 - 2(0.001968))20 + 0.001968(20) = 20$$

de igual manera para el nodo (3,1) resulta

$$T_{3,1} = 0.001968(20) + (1 - 2(0.001968))20 + 0.001968(20) = 20$$

Hay que observar que la temperatura del nodo (4,0) es 20 °C, ya que la condición inicial lo establece y esa frontera está aislada.

Para el cálculo del nodo (4,1) se usa la condición frontera  $T_x = 0$  y su aproximación con diferencias hacia atrás como sigue

$$T_x(1,t) = 0 \approx \frac{T_{4,1} - T_{3,1}}{\Delta x}$$

por lo que  $T_{4,1} \approx T_{3,1} \approx 20$  °C

Con este procedimiento se calculan las temperaturas de los nodos de las filas superiores; aquí debe notarse que por la condición frontera  $T_x = 0$ , la temperatura en el extremo aislado de la barra será aproximadamente igual a la temperatura de la barra en un nodo anterior ( $x = 0.75$ ).

Los resultados de la tabla 8.5, obtenidos con el **PROGRAMA 8.1** muestran lo anterior.

**Tabla 8.5**

| $t$<br>(en horas) | $x$ (m) |        |        |        |        |
|-------------------|---------|--------|--------|--------|--------|
|                   | 0.00    | 0.250  | 0.500  | 0.75   | 1.00   |
| 0.0               | 60.000  | 20.00  | 20.000 | 20.000 | 20.000 |
| 0.1               | 100.000 | 20.079 | 20.000 | 20.000 | 20.000 |
| 0.2               | 100.000 | 20.236 | 20.000 | 20.000 | 20.000 |
| 0.4               | 100.000 | 20.548 | 20.001 | 20.000 | 20.000 |
| 0.6               | 100.000 | 20.858 | 20.004 | 20.000 | 20.000 |
| 0.8               | 100.000 | 21.166 | 20.007 | 20.000 | 20.000 |
| 1.0               | 100.000 | 21.471 | 20.012 | 20.000 | 20.000 |
| 1.5               | 100.000 | 22.223 | 20.029 | 20.000 | 20.000 |
| 2.0               | 100.000 | 22.961 | 20.053 | 20.001 | 20.001 |
| 2.5               | 100.000 | 23.685 | 20.084 | 20.001 | 20.001 |
| 3.0               | 100.000 | 24.395 | 20.121 | 20.002 | 20.002 |

- 8.4** Encuentre la distribución de temperatura  $T(x,t)$  en una aleta delgada de cobre (véase Fig. 8.28), unida por la cara sombreada a un radiador cuya temperatura constante es 200 °F. La función de la aleta es disipar calor por convención a la atmósfera, cuya temperatura es de 68 °F. Considere que la aleta está inicialmente a 68 °F y que el coeficiente de transmisión de calor  $h$  es 30 BTU/(h pie<sup>2</sup> °F).

**Solución** Al efectuar un balance de calor en un elemento diferencial de la aleta, de dimensiones  $\Delta x$ ,  $l = 1$  pie y  $a = 0.5$  pie se tiene, de acuerdo con la ley de continuidad (Ec. 8.3)

$$(A \Delta x \rho C_p) \frac{\partial T}{\partial t} = -k A \left. \frac{\partial T}{\partial x} \right|_x - (-k A \left. \frac{\partial T}{\partial x} \right|_{x + \Delta x}) - 2 \Delta x (ah) (T - 68)$$



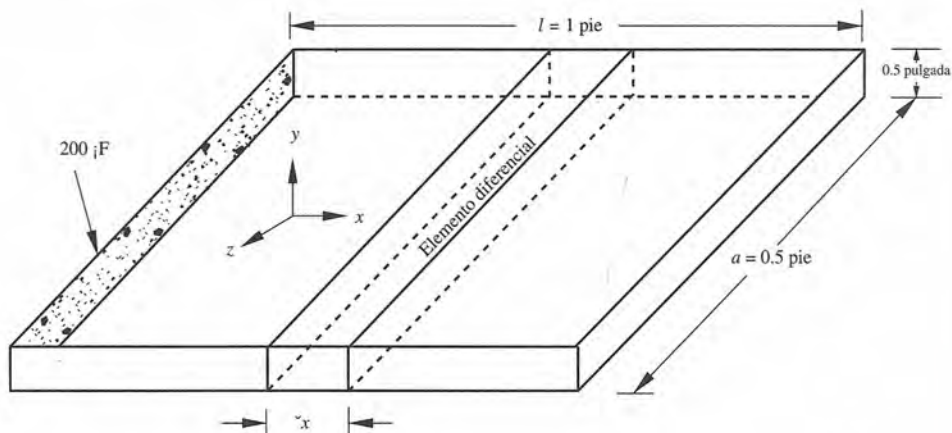


Figura 8.28

donde el primer y segundo términos del lado derecho se refieren al calor que entra y que sale, respectivamente, del elemento diferencial por las caras perpendiculares al eje  $x$  y de área  $A = 0.5(0.5)/12 = 0.020833$  pies<sup>2</sup>. En cambio, el tercer término se refiere al calor que sale del elemento diferencial hacia la atmósfera; con el factor 2 de éste se incluyen las dos caras perpendiculares al eje  $y$ . Nótese que se ha depreciado el calor que sale por las caras perpendiculares al eje  $z$ , ya que la placa es muy delgada y  $Q = 0$ .

El lado izquierdo de la ecuación representa la acumulación de calor en el elemento diferencial considerado.

Toda la ecuación se divide entre  $A\Delta x \rho C_p$  y después se hace que  $\Delta x \rightarrow 0$ , con lo cual

$$\frac{\partial T}{\partial t} = \frac{k}{\rho C_p} \frac{\partial^2 T}{\partial x^2} - \frac{2(ah)}{C_p A \rho} (T - 68)$$

de tal manera que se obtiene el modelo matemático que rige el fenómeno descrito.

Si a este modelo se unen las condiciones

$$T(x, 0) = 68^\circ\text{F}$$

que describen la temperatura en las fronteras de la aleta, se tiene un problema de valores en la frontera.

Las propiedades físicas del cobre requeridas para resolver la ecuación se listan enseguida.

$$k = 223 \text{ BTU}/(\text{h ft}^2 ^\circ\text{F}/\text{ft})$$

$$C_p = 0.09 \text{ BTU}/\text{lb}^\circ\text{F}$$

$$\rho = 560 \text{ lb}/\text{ft}^3$$

Para resolver este PVF se ha utilizado el método de Crank-Nicholson, para lo cual se ha modificado el programa 8.3 a fin de incluir el término

$$\frac{2ah}{A\rho C_p} (T - 68)$$

El programa resultante utiliza  $\Delta t = 0.001 \text{ h}$  y la longitud de la aleta (1 pie) se dividió en intervalos de 0.05 cada uno.

En la tabla siguiente se presentan algunos de los resultados obtenidos

| $t$<br>(en hora) | (pies) |        |        |       |       |       |
|------------------|--------|--------|--------|-------|-------|-------|
|                  | 0.00   | 0.20   | 0.40   | 0.60  | 0.80  | 1.00  |
| 0.000            | 134    | 68.00  | 68.00  | 68.00 | 68.00 | 68.00 |
| 0.001            | 200    | 71.28  | 68.05  | 68.00 | 68.00 | 68.00 |
| 0.002            | 200    | 81.30  | 68.42  | 68.01 | 68.00 | 68.00 |
| 0.004            | 200    | 102.00 | 71.56  | 68.20 | 68.01 | 68.00 |
| 0.006            | 200    | 113.86 | 77.05  | 68.99 | 68.07 | 68.00 |
| 0.008            | 200    | 121.43 | 82.53  | 70.52 | 68.28 | 68.00 |
| 0.010            | 200    | 126.66 | 87.31  | 72.48 | 68.71 | 68.00 |
| 0.015            | 200    | 134.51 | 96.16  | 77.62 | 70.51 | 68.00 |
| 0.020            | 200    | 138.76 | 101.85 | 81.95 | 72.57 | 68.00 |
| 0.040            | 200    | 144.85 | 111.07 | 90.42 | 77.43 | 68.00 |
| 0.060            | 200    | 146.16 | 113.17 | 92.50 | 78.71 | 68.00 |
| 0.080            | 200    | 146.46 | 113.66 | 93.00 | 79.02 | 68.00 |
| 0.100            | 200    | 146.53 | 113.78 | 93.11 | 79.09 | 68.00 |

## Problemas

8.1 Clasifique las siguientes ecuaciones diferenciales parciales (consúltese el ejercicio 8.1).

a)  $\operatorname{sen} x \frac{\partial^2 u}{\partial x^2} + y^2 \frac{\partial^2 u}{\partial y^2} = 0$

a1) en  $0 < x < \pi$ ,  $-\infty < y < \infty$

a2) en  $x = 0$ ,  $-\infty < y < \infty$

a3) en  $\pi < x < 2\pi$ ;  $-\infty < y < \infty$

b)  $y \frac{\partial^2 u}{\partial x^2} - x \frac{\partial^2 u}{\partial y^2} + \frac{\partial u}{\partial x} + y \frac{\partial u}{\partial y} = 0$

c)  $y \frac{\partial^2 u}{\partial x^2} + 2e^{x+y} \frac{\partial^2 u}{\partial x \partial y} + e^{2y} \frac{\partial^2 u}{\partial y^2} = 0$

d)  $\frac{\partial^2 u}{\partial x^2} + (1 + y^2) \frac{\partial^2 u}{\partial y^2} = 0$

e)  $\operatorname{sen}^2 y \frac{\partial^2 u}{\partial x^2} - e^{2x} \frac{\partial^2 u}{\partial y^2} + 3 \frac{\partial u}{\partial x} - 5u = 0$

8.2 ¿En qué regiones la ecuación

$$\frac{\partial^2 u}{\partial x^2} + y \frac{\partial^2 u}{\partial y^2} = 0$$

es hiperbólica, elíptica y parabólica? (consúltese ejercicio 8.1).

- 8.3 Obtenga las ecuaciones (8.18) a (8.22) a partir de la expansión en serie de Taylor de  $T(x, t)$ , alrededor del punto  $(x_p, t_p)$ , aplicando los mismos razonamientos que condujeron a las ecuaciones (8.12) y (8.14) a (8.16).
- 8.4 Expresé las siguientes ecuaciones diferenciales en términos de diferencias finitas

a)  $\frac{d^2y}{dx^2} - y \frac{dy}{dx} + 2y = 0$  con diferencias centrales

b)  $\sin x \frac{\partial^2 u}{\partial x^2} + y^2 \frac{\partial^2 u}{\partial y^2} = 0$  con diferencias hacia delante

c)  $y \frac{\partial^2 u}{\partial x^2} - x \frac{\partial^2 u}{\partial y^2} + \frac{\partial u}{\partial x} + y \frac{\partial u}{\partial y} = 0$  con diferencias centrales

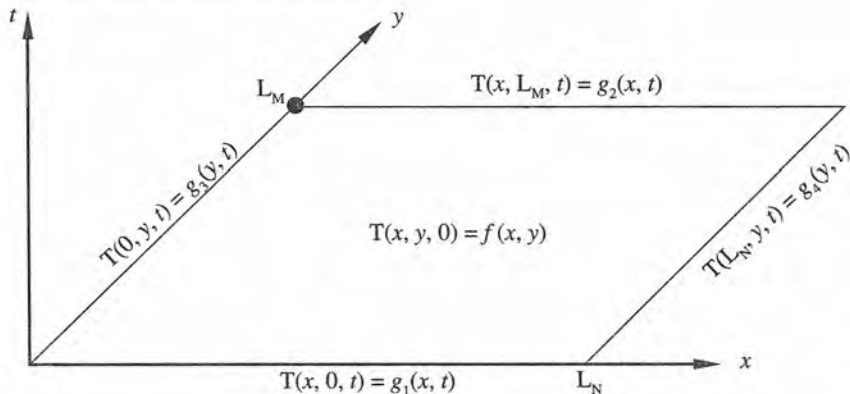
d)  $\frac{\partial^2 u}{\partial x^2} + (1 + y^2) \frac{\partial^2 u}{\partial y^2} = 0$  con diferencias hacia atrás

- 8.5 La ecuación

$$\frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2} + \alpha \frac{\partial^2 T}{\partial y^2}$$

describe la conducción de calor en régimen transitorio en dos dimensiones. Exprésela en términos de diferencias finitas. El término de la izquierda en diferencias hacia adelante y los términos de la derecha en diferencias centrales.

- 8.6 La ecuación del problema 8.5 describe la conducción de calor en una lámina delgada (espesor despreciable) y aislada y permite calcular la temperatura en cualquier punto de la lámina a cualquier tiempo en régimen transitorio. Si las condiciones inicial y frontera son en general como se muestra en la figura 8.29



**Figura 8.29**  
Conducción de calor en una lámina rectangular.

Establezca el problema de valor en la frontera, encuentre el algoritmo correspondiente al método explícito y resuelva con  $\alpha = 0.01$  y las siguientes condiciones inicial y de frontera

CI:  $T(x, y, 0) = 20^\circ\text{C}; 0 \leq x \leq 0.1 \text{ m}; 0 \leq y \leq 0.2 \text{ m}$

CF1:  $T(x, 0, t) = 100^\circ\text{C}; 0 \leq x \leq 0.1 \text{ m}; 0 \leq t \leq 1 \text{ hora}$

CF2:  $T(x, 0.2, t) = 50^\circ\text{C}; 0 \leq x \leq 0.1 \text{ m}; 0 \leq t \leq 1 \text{ hora}$

CF3:  $T(0, y, t) = 100^\circ\text{C}; 0 \leq y \leq 0.2 \text{ m}; 0 \leq t \leq 1 \text{ hora}$

CF4:  $T(0.1, y, t) = 50^\circ\text{C}; 0 \leq y \leq 0.2 \text{ m}; 0 \leq t \leq 1 \text{ hora}$

Nota: Elabore una malla tal que  $0 < \lambda \leq 0.5$ .



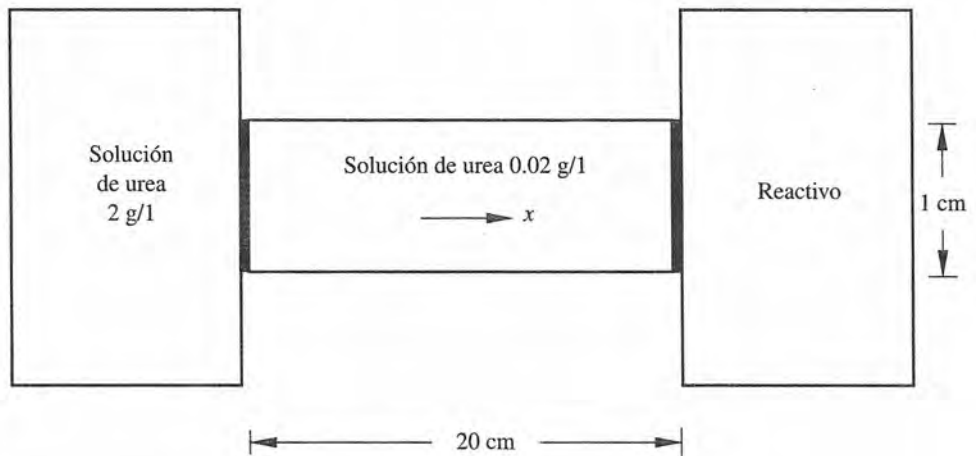
- 8.7 Resuelva el problema de valor en la frontera del problema 8.6 con el método implícito correspondiente.
- 8.8 Resuelva el PVF del problema 8.6 con el método de Crank-Nicholson correspondiente.
- 8.9 Se tiene una solución de urea contenida en un tubo de 1 cm de diámetro interior (véase Fig. 8.30), con un concentración inicial de 0.02 g/litro. Una membrana semipermeable conecta el tubo con un frasco que contiene una solución de urea con 2 g/litro. Otra membrana la conecta con un reactivo con el cual la urea reacciona para desaparecer instantáneamente. Si se considera que la difusión de la urea ocurre únicamente en el eje  $x$ , calcule la concentración de ésta a lo largo del tubo en los primeros 10 minutos. La difusividad de la urea es  $\mathcal{D} = 0.017 \text{ cm}^2/\text{h}$  (véase Ejer. 8.2).
- 8.10 Resuelva el problema 8.9 considerando que en el extremo derecho del tubo se tiene un frasco que contiene una solución con 1 g/L de urea en lugar del reactivo. Todas las demás condiciones permanecen.
- 8.11 Resuelva el siguiente PVF por los métodos explícito e implícito

$$\text{EDP: } \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2}$$

$$\text{CI: } T(x, 0) = 20 \text{ sen } x \quad \alpha = 1 \text{ pie}^2/\text{h}$$

$$\text{CF1: } T(0, t) = 100^\circ\text{C} \quad L = 1 \text{ pie}$$

$$\text{CF2: } T(L, t) = 50^\circ\text{C} \quad t_{\text{máx}} = 1 \text{ hora}$$



**Figura 8.30**  
Difusión de urea en una solución.

- 8.12 Resuelva el siguiente PVF por el método de Crank-Nicholson

$$\text{EDP: } \frac{\partial T}{\partial t} = \alpha \frac{\partial^2 T}{\partial x^2}, \quad \alpha = 1 \text{ pie}^2/\text{h}$$

$$\text{CI: } T(x, 0) = 20^\circ\text{C} \quad L = 1 \text{ pies}$$

$$\text{CF1: } T(0, t) = 100^\circ\text{C} \quad 0 < t \leq 12 \text{ minutos}$$

$$T(0, t) = 20^\circ\text{C} \quad 12 < t \leq 60 \text{ minutos}$$

$$\text{CF2: } T(L, t) = 100^\circ\text{C} \quad 0 < t \leq 12 \text{ minutos}$$

$$T(L, t) = 20^\circ\text{C} \quad 12 < t \leq 60 \text{ minutos}$$

- 8.13 Resuelva el ejercicio 8.3 por el método de Crank-Nicholson. Compare resultados.



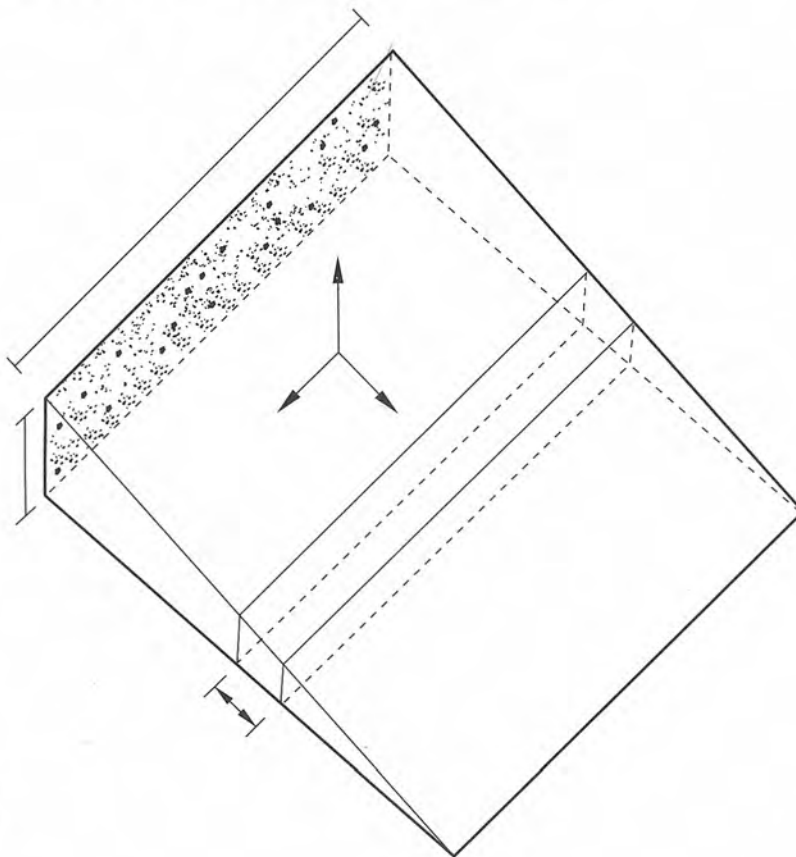
**8.14** Resuelva el EDP del ejercicio 8.4 con las siguientes condiciones

$$\text{CI: } T(x, 0) = (80 - 10x)^\circ\text{F}$$

$$\text{CF1: } T(0, t) = 200^\circ\text{F}$$

$$\text{CF2: } T(1, t) = 68^\circ\text{C}$$

**8.15** Si en el ejercicio 8.4 se modifica la geometría de la aleta para tenerla como se muestra en la figura 8.27, plantee y resuelva el PVF resultante.



**Figura 8.31**  
Conducción de  
calor en una  
aleta triangular.

**8.16** Utilice el método visto en la sección 8.7 para resolver el PVF siguiente

$$\text{PVF} \left\{ \begin{array}{l} \text{EDP: } \frac{\partial^2 y}{\partial t^2} = \frac{\partial^2 y}{\partial x^2} \\ \text{CI: } y(x, 0) = \sin(2\pi x), \quad 0 < x < 1 \\ \text{CI2: } \left. \frac{\partial y}{\partial t} \right|_{(x, 0)} = 2\pi \sin(2\pi x) \quad 0 < x < 1 \\ \text{CF1: } y(0, t) = 0, \quad t > 0 \\ \text{CF2: } y(1, t) = 0, \quad t > 0 \end{array} \right.$$

Utilice  $a = 0.1$  y  $b = 0.01$ . Observe que sólo se modificó la posición inicial de la cuerda.

- 8.17** Un problema interesante se obtiene cuando la condición inicial CII es una función dada en dos partes, como sigue:

$$\text{CII: } y(x, 0) = \begin{cases} 2x, & 0 < x \leq 0.5 \\ 2(1-x), & 0.5 \leq x < 1 \end{cases}$$

y la cuerda simplemente se suelta, esto es la condición inicial CI2 es  $\left. \frac{\partial y}{\partial t} \right|_{(x, 0)} = 0$ .

Empiece graficando la condición inicial CII y después haga los cálculos con el método visto en la sección 8.7. Puede modificar el **PROGRAMA 8.4** del CD.

- 8.18** El programa 8.4 del disco permite observar el movimiento de la cuerda en modo rápido y en modo cámara lenta. Nótese que la cuerda no parece tender en el tiempo a una posición de reposo que sería una línea horizontal.

- ¿Cómo explica usted este fenómeno que va en contra de lo observado en la realidad?
- Si se considera que está actuando sobre la cuerda un campo eléctrico, magnético o gravitacional, diga usted cómo se modificaría la EDP.

- 8.19** Se puede obtener una mejor aproximación que la que proporciona la ecuación 8.63 si se cuenta con la segunda derivada de  $f(x)$  en  $x_i$ , de la siguiente manera

- $y_{i,1} \approx y_{i,0} + bg(x_i) + c^2 b^2 / 2 f''(x_i)$  para  $i = 1, 2, \dots, n-1$
- En caso de no contar con  $f''(x_i)$  puede aproximarse por diferencias centrales como sigue

$$f''(x_i) = \frac{f(x_{i-1}) - 2f(x_i) + f(x_{i+1}))}{a^2}$$

De donde la ecuación dada en el inciso a) quedaría

$$y_{i,1} = (1 - \lambda^2) f(x_i) + \frac{\lambda^2}{2} [f(x_{i+1}) + f(x_{i-1}))] + bg(x_i)$$

Modifique el **PROGRAMA 8.4** del CD o elabore otro a fin de emplear alguna de estas aproximaciones para resolver el problema de la cuerda vibrante.

# RESPUESTAS A PROBLEMAS

## SELECCIONADOS

### CAPÍTULO 1

1.2

$$\begin{array}{lcl} a) & 536_{10} = 1030_8 = & 1000011000_2 \\ b) & 923_{10} = 1633_8 = & 1110011011_2 \\ c) & 1536_8 = 3000_8 = & 11000000000_2 \\ d) & 8_{10} = 10_8 = & 1000_2 \\ e) & 2_{10} = 2_8 = & 10_2 \\ f) & 10_{10} = 12_8 = & 1010_2 \\ g) & 0_{10} = 0_8 = & 0_2 \end{array}$$

1.3

$$\begin{array}{lcl} a) & 7778 = & 111111111_2 \\ b) & 5738 = & 101111011_2 \\ c) & 78 = & 111_2 \\ d) & 28 = & 10_2 \\ e) & 108 = & 1000_2 \\ f) & 08 = & 0_2 \end{array}$$

1.4

$$\begin{array}{lcl} a) & 1000_2 = & 8_{10} \\ b) & 10101_2 = & 21_{10} \\ c) & 111111_2 = & 63_{10} \end{array}$$

1.9

$$\begin{array}{lcl} a) & 985.34_{10} \approx 1731.256_8 \approx & 1111011001.010101110_2 \\ b) & 10.1_{10} \approx 12.063_8 \approx & 1010.000110011_2 \\ c) & 888.222_{10} \approx 1570.1615_8 \approx & 1101111000.001100011010_2 \\ d) & 3.57_{10} \approx 3.4436_8 \approx & 11.100100011110_2 \\ e) & 977.93_{10} \approx 1721.7341_8 \approx & 1111010001.111011100001_2 \\ f) & 0.357_{10} \approx 0.2666_8 \approx & 0.010110110110_2 \\ g) & 0.9389_{10} \approx 0.740557_8 \approx & 0.111100000101101111_2 \\ h) & -0.9389_{10} \approx -0.740557_8 \approx & -0.111100000101101111_2 \end{array}$$

1.13

$$\begin{array}{lcl} a) & 0.19921875 \\ b) & -160 \\ c) & 9306112 \end{array}$$

1.14

$$\begin{array}{lcl} a) & 0.1011010011100011101100101 \times 2^{1010} \\ b) & -0.1111010100101111 \times 2^{100} \\ c) & 0.11100100011001001 \times 2^{-100} \\ d) & 0.1111101 \times 2^{1101} \end{array}$$

1.23 La mantisa normalizada más pequeña en binario es  $0.10000000$  ( $=1/2$  en decimal), no  $0.00000001$  ( $2^{-8}$ ) y la mayor es  $0.11111111$  ( $\approx 1$ ).  
Por esto, los números de máquina positivos deben quedar en el intervalo cerrado  $[s, L]$ , donde

$$s = \text{número de máquina positivo más pequeño} = (0.10000000) \times 2^{-64} \\ = 2^{-65} \approx 0.2710505 \times 10^{-19}$$

y

$$L = \text{número de máquina positivo mayor} = + (0.11111111) \times 2^{63} \\ \approx 0.91873437 \times 10^{17}$$

El intervalo  $[s, L]$  puede dividirse en 128 subintervalos

$$\begin{array}{cccccc} [s, 2s), & [2s, 2^2s), & [2^2s, 2^3s), & \dots, & [2^{126}s, 2^{127}s), & [2^{127}s, L] \\ E = -64 & E = -63 & E = -62 & \dots & E = +62 & E = +63 \end{array}$$

donde  $E$  es la característica.

Nótese que cada subintervalo es dos veces más grande que su predecesor. Para cada  $E$  hay  $2^8$  posibles mantisas normalizadas. Por tanto, una computadora con una palabra de 16 bits puede almacenar un total de

$$128 \times 2^8 = 32768 \text{ números positivos de máquina en el} \\ \text{intervalo } [s, L]$$

$$1.26x = -278; y = 248.67$$

## CAPÍTULO 2

$$2.1 \quad a) \quad g'(x) = -\frac{2}{(x+1)^3}; \quad \begin{array}{l} x > 0.26 \\ x < -2.26 \end{array}$$

$$b) \quad g'(x) = -\frac{2(x-1)^{-2/3}}{3(x+1)^{4/3}}; \quad x > 1.125$$

$$c) \quad g'(x) = \cos x; \quad x = n\pi; \quad n = 0, 1, 2, \dots$$

$$d) \quad g'(x) = x + \frac{1}{2} \frac{n x}{\tan x} - \frac{1}{2}$$

$$g'(x) = 1 + \frac{(1/x) \tan x - \ln x \sec^2 x}{2 \tan^2 x}; \quad x = 3.8; \quad x = 4.2$$

$$e) \quad g'(x) = \frac{1 + 3x^2}{4\sqrt{6-x-x^3}}; \quad x = 0.8; \quad x = 1.2$$

$$f) \quad g'(x) = \frac{\sec x}{2}; \quad g'(x) = \frac{\sec x \tan x}{2}; \quad x = 0; \quad x = 0.5$$

2.3 (del problema 2.1)

$$a) \quad \bar{x} \approx 0.46557 \quad b) \quad \bar{x} \approx 4.87035 \quad c) \quad \bar{x} \approx 0.$$

$$d) \quad \bar{x} \approx 4.09546 \quad e) \quad \bar{x} \approx 1 \quad f) \quad \bar{x} \approx 0.61003$$

2.6 a)  $n$  multiplicaciones y  $n$  sumas

b)  $2n$  multiplicaciones y  $n$  sumas

$$2.8 \quad a) \quad \bar{x} \approx 3.14619 \quad b) \quad \bar{x} \approx 0.85261$$

$$c) \quad \bar{x} \approx 1.02987 \quad d) \quad \bar{x} \approx 0.20164$$



- 2.9 a)  $\bar{x} \approx -0.88616$   $\bar{y} \approx -1.39177$   
 b)  $\bar{x} \approx 0.74798$   $\bar{y} \approx 1.11894$   
 c)  $\bar{x} \approx 1.31555$   $\bar{y} \approx 0.32104$   $\bar{z} \approx 1.1362$   
 d)  $\bar{x} \approx 4.20447$   $\bar{y} \approx 0.76345$
- 2.17 a)  $\bar{x} \approx 10$  b)  $\bar{x} \approx 0.66624$   
 c)  $\bar{x} \approx 1.82938$  d)  $\bar{x} \approx 1.2032$
- 2.20 Si  $X_I = 2$  y  $X_D = 4$ ,  $n \approx 11$
- 2.26 a)  $\bar{x} \approx 0.25753$  b)  $\bar{x} \approx 3.83910$  c)  $\bar{x} \approx -0.94157$
- 2.28  $\bar{x}_{1,2} = + 2i$
- 2.30  $\bar{x}_{1,2} \approx -1.6844 \pm 3.43133 i$
- 2.33 a)  $\bar{x}_1 \approx 1.1$ ;  $\bar{x}_2 \approx 1.1$ ;  $\bar{x}_{3,4} \approx 3 + 4 i$   
 b)  $\bar{x}_1 \approx 1.24144$ ;  $\bar{x}_2 \approx 10.01798$   
 $\bar{x}_3 \approx 2.96396$ ;  $\bar{x}_4 \approx 0.97661$   
 c)  $\bar{x}_1 \approx 1$ ;  $\bar{x}_2 \approx 2$ ;  $\bar{x}_3 \approx 3$ ;  $\bar{x}_{4,5} \approx 2 \pm i$   
 d)  $\bar{x}_1 \approx 1.7$ ;  $\bar{x}_{2,3} \approx 1 \pm i$   $\bar{x}_{4,5} \approx + 2 i$
- 2.36  $\bar{V}_{\text{He}} \approx 0.62542 \text{ L}$ ;  $\bar{V}_{\text{H}_2} \approx 0.63983 \text{ L}$ ;  $\bar{V}_{\text{O}_2} \approx 0.6106 \text{ L}$ ;
- 2.41  $T \approx 105.33 \text{ }^\circ\text{C}$
- 2.43  $T \approx -102.3 \text{ }^\circ\text{C}$
- 2.47  $t \approx 3.041 \text{ hrs.}$
- 2.48  $f \approx 0.04878$
- 2.50  $\lambda \approx 0.08747$

### CAPÍTULO 3

- 3.15 a)  $e_1 = [1, -2, 5, 7, 8, 0.3]^T$   
 $e_2 = [-3.0343, 3.0686, 1.8284, -4.2402, 3.7255, -0.3103]^T$   
 $e_3 = [-1.0029, 5.8915, 0.4998, 3.4940, -1.8232, 1.3820]^T$   
 $e_4 = [5.7399, -3.2717, 0.1600, -6.7681, 2.8280, 38.8973]^T$   
 $e_5 = [4.8912, 2.1153, -0.6869, -1.0594, 1.3045, -0.8202]^T$
- b)  $e_1 = [4, 2, 1]^T$   
 $e_2 = [-0.42857, 0.28571, 1.14286]^T$   
 $e_3 = [-1.21212, 3.0303, -1.21212]^T$

$$\begin{aligned} c) \quad e_1 &= [10, \quad -20, \quad 5]^T \\ e_2 &= [1.66667, 1.66667, 3.33333]^T \\ e_3 &= [-1.07143, -0.35714, 0.71429]^T \end{aligned}$$

$$\begin{aligned} d) \quad e_1 &= [-1, \quad 1, \quad 0, \quad 2]^T \\ e_2 &= [4.33333, \quad 7.66667, \quad 1., \quad 1.66667]^T \\ e_3 &= [1.5, \quad -0.5, \quad -1, \quad 1]^T \\ e_4 &= [0.27322, -0.20036, 0.74681, 0.23679]^T \end{aligned}$$

**3.16** 3, 3, 3 y 4, respectivamente

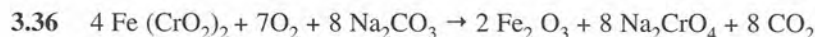
**3.17** Número de reacciones independientes = 8

**3.20**  $w = 2$  y  $w = 3$  para solución única  
 $w = 1$  para número infinito de soluciones

$$\begin{aligned} 3.26 \quad a) \quad x &= [-0.14114, \quad 1.56229, \quad -1.09371, \quad 0.30210]^T \\ b) \quad x &= [4, \quad 3, \quad 1]^T \\ c) \quad x &= [0.925, \quad -4.7, \quad 15.7, \quad 10.625, \quad -2.975]^T \end{aligned}$$

$$3.33 \quad C_{A1} = 0.4507; C_{A2} = 0.33803; \quad C_{A3} = 0.25352$$

$$3.35 \quad x = [0.04, 9.6 \times 10^{-4}, 2.304 \times 10^{-5}, 5.5264 \times 10^{-7}, 1.29525 \times 10^{-8}]^T$$



$$3.49 \quad p = [0.69118, \quad -9.9278, \quad 6.41471, \quad -5.32941, \quad 5., \quad -1.35379]^T$$

$$3.50 \quad x = [0.89052, \quad 0.99421, \quad 1.07371]^T$$

$$\begin{aligned} 3.54 \quad a) \quad x &= [2, \quad 5.33333, \quad 1.66666]^T \quad b) \quad x = [1, 1, 1]^T \\ d) \quad x &= [1.99464, 6.27889, -0.30459, 8.43005 \times 10^{-3}, -9.54861 \times 10^{-5}]^T \\ e) \quad x_1 &= -0.76592; \quad x_2 = 1.06005 \\ x_3 &= 1.19592; \quad x_4 = -0.04580 \\ x_5 &= 0.91064; \quad x_6 = 0.16144 \\ x_7 &= -0.81218; \quad x_8 = 0.15305 \\ x_9 &= -0.25689; \quad x_6 = 0.03420 \\ x_{11} &= 0.02692; \end{aligned}$$

$$3.60 \quad \lambda_1 = 3.28636; \quad \lambda_2 = 8.17744$$

$$\lambda_{3,4} = 4.26809 \pm 1.46356 i$$

$$3.61 \quad [1, \quad 2.96710, \quad 1.24679, \quad -0.83630]^T$$

$$3.62 \quad \lambda_{\text{dominante}} = 3; \quad e = [1, \quad 1, \quad 0]^T$$

**CAPÍTULO 4**

- 4.1  $\mathbf{x} = [3, 2, 1, 2, 4, 6]^T$
- 4.2  $\mathbf{x} = [2, 4, 1, 1]^T$
- 4.3  $x^0 = 0.8$  ;  $y^0 = 0.5$ ;  $\bar{x} \approx 0.7718$ ;  $\bar{y} \approx 0.4197$
- 4.4  $g_1(x, y) = 37 - y$ ,  $g_2(x, y) = x - 5$   
 $x > 5$        $y$        $y < 37$
- 4.5 a)  $\bar{x} = 6$ ;       $\bar{y} = 1$   
 b)  $\bar{x} \approx 6.17107$        $\bar{y} \approx -1.08216$
- 4.6 a)  $[0, 0]^T$ ;     $[8000, 4000]^T$   
 b)  $\mathbf{x} \approx [0.529164, 0.399996, 0.100006]^T$   
 c)  $\mathbf{x} = [1, 1, 1]^T$   
 d)  $\mathbf{x} \approx [0.14996, 0.000599, 0.410824]^T$
- 4.9 a)  $\mathbf{x} = [0, 0.1, 1]^T$   
 b)  $\mathbf{x} \approx [-0.110949, 0.4110824]^T$
- 4.11  $\mathbf{x} \approx [6.95, 2.5, -0.15]^T$
- 4.12  $\bar{C}_{A1} \approx 0.53292$ ;  $\bar{C}_{A2} \approx 0.42435$ ;  $\bar{C}_{A3} \approx 0.32879$
- 4.15  $\mathbf{x} \approx [0.61089, 0.37899, 0.24919, 0.51622, 0.07728]^T$
- 4.17  $\mathbf{x} \approx [1.4695, 0., -0.22777]^T$
- 4.18  $T = 57.85488935 e^{(-0.10941272 t)} + 33.4992038$
- 4.24  $t_{\text{opt}} = -1.15$
- 4.26 a)  $\bar{x} \approx -2.13147$ ;       $\bar{y} \approx 0.97941$ ;       $\bar{z} \approx -1.36122$   
 b) No tiene solución  
 c)  $\mathbf{x} \approx [4.37534, 1.66657, -3.46619]^T$
- 4.31 a)  $z_{\min} = -3$  en  $x = -7.85396$ ,  $y = -1.33128$  E-6  
 b)  $z_{\min} = 0$  en  $x_1 = 0$ ,       $x_2 = 0$ ,       $x_3 = 0$

**CAPÍTULO 5**

- 5.1 a) 1.1303      b) 1.2597      c) 1.2034      d) 16
- 5.2 203.35
- 5.3  $x(2) = 5.8$

5.8  $J_0(0.8) = 0.8463$

5.14  $p = 2.59$

5.15  $v = 67.8$

5.18  $a_0 = 99600 \quad a_1 = -1209.166667$   
 $a_2 = 5.375 \quad a_3 = -0.00833333$

5.19  $C_B(0.82) = 1.12$

5.23  $R_2(10) \approx 98$

5.28  $f(1;3,0.13) = 0.925$ , con un polinomio de segundo grado

5.29  $r = 10.12223 + 0.027975 T$

5.30  $P = 481.03743 v^{-1.06533}$

5.31  $a = 0.24033$

5.32  $z = 7.993487 \times 10^{10}; \quad E = 19999.73634$

5.35  $n = 3$

5.36  $\tau = 0.92893$

5.37  $a = 1.78752 \quad b = 0.0006533 \quad c = 1.84624 \times 10^{-5}$

5.39  $a_0 = 161.33646 \quad a_1 = 32.96875 \quad a_2 = -0.0855$

## CAPÍTULO 6

6.5 a) 20.9 kg/min. b) 30097 kg. c) 0.38153 d) 114.863 kg

6.7 1.64711

6.8 81792338.66 con Simpson 1/3 y  $N = 100$

6.12 a) 1.71125 b) 0.56343 c) 0.40546 d) 1.29584

6.13 a)  $I_1^{(3)} = 1.26613$  b)  $I_0^{(2)} = 0.07921$   
 c)  $K = 1: I_1^{(2)} = 1.04417; K = 2: I_1^{(2)} = 0.87105$

6.15 a)  $I_0^{(3)} = 0.006303$  b)  $I_0^{(3)} = 0.946083$

6.17 0.84338 o bien 84.388 %

6.18 3.70387 con Simpson 1/3

6.19 analíticamente  $= \sin^{-1}(1) = 1.570796327$



- 6.20 a) 0 b) 0.25 c) 1/3  
 6.23 50403593.58 con  $N = 2$ , 50021079.17 con  $N = 3$   
 6.24 1.21484  
 6.26 2.61198 con  $N = 2$ , 2.61945 con  $N = 3$   
 6.27 a) -0.57722 c) 0.22532 d) 1/3  
 6.28 a) 0.02 b) 0.22532 c) 0.08428  
 6.30 con  $N = 10$  y  $M = 10$  a) 1.47627 b) 1.47623 c) 0.35593  
 con  $N = 2$  y  $M = 2$  d) 0.25  
 6.32 a) 6.93463 con  $N = 10$  y  $M = 10$  b) 0.33424 con  $N = 20$  y  $M = 20$   
 c) 0.83333 con  $N = 2$  y  $M = 2$  d) 4.38911 con  $N = 10$  y  $M = 10$   
 6.35  $\bar{x} = 0.53802$   $\bar{y} = 0.52466$   
 6.36 a) 14951.02 con  $N = 20$  y  $M = 20$  b) 0.11267 con  $N = 10$  y  $M = 10$   
 6.39 -0.014121 con  $P = 2, 8$  y 15  
 6.40  $f''(3.7) = 0.02503$  con  $N = 2$   
 6.45  $v = 96.62$  m/s  
 6.46  $E(0.95) = 118.466$   $E(0.96) = 79.809$   $E(0.97) = 46.916$   
 $E(0.98) = 24.653$   $E(0.99) = 11.096$   $E(1.00) = 1.404$   
 6.47 0.72 para  $t = 10$  0.56 para  $t = 35$  0.36 para  $t = 60$   
 6.48 a) -0.0006725 b) 0.01218 c) -0.00038 d) 0.008829

## CAPÍTULO 7

- 7.1 tiempo  $\approx 432$  s con Euler y  $h = 0.5$   
 7.2 gasto  $\approx 0.0555$  m<sup>3</sup>/s  
 7.3 tiempo  $\approx 8000$  s con  $h = 25$  s y Euler simple  
 7.4 gasto = 7.73 l/s tiempo  $\approx 32$  horas  
 7.6
 

|           |     |     |     |      |     |
|-----------|-----|-----|-----|------|-----|
| $C_{A0}$  | 0.5 | 1.0 | 1.5 | 1.0  | 2.0 |
| $C_{B0}$  | 1.0 | 1.5 | 2.0 | 1.0  | 0.5 |
| $t$ (min) | 7   | 6   | 5   | 18.5 | 3   |

- 7.7  $h$  0.1 0.5 1.0 0.05  
 $I_1(3)$  2.047 1.992 5.280 2.052  
 $I_2(3)$  0.658 0.590 2.736 0.665
- 7.11 a) 8.8 s b) 220 m c) 95.56 m
- 7.12 tiempo  $\approx 30$  min
- 7.13  $C_1 = 49.00$   $C_2 = 44.57$   $C_3 = 41.84$
- 7.14  $C_1 = 46.31$   $C_2 = 44.58$   $C_3 = 42.38$  con  $h = 1$  min RK-4
- 7.15  $(C_{A1}, C_{A2}) =$  a) (0.67, 0.62) b) (0.68, 0.86) c) (0.73, 0.61)  
d) (0.813, 0.62) e) (0.655, 0.63)
- 7.16  $t$  200 400 600 800 1000  
 $C_A$  5.0000 4.2347 3.0407 1.7833 1.6097  
 $T$  300.0 310.6 322.3 333.2 333.8
- 7.17  $C_A = 4.33$ ,  $T = 307.31$  a  $t = 1200$  para  $T_j = 310$   
 $C_A = 0.488$ ,  $T = 348.38$  a  $t = 1200$  para  $T_j = 350$   
 $C_A = 0.531$ ,  $T = 344.28$  a  $t = 1200$  para  $T_j = 340$
- Cuando la temperatura  $T$  alcanza un valor mayor que  $T_j$ , se lleva a cabo violentamente la reacción, por lo que se acostumbra enfriar el reactor cuanto  $T$  está cerca de  $T_j$  o la rebasa.
- 7.18 a) 

|     |       |       |       |       |       |       |       |       |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $x$ | 0.5   | 1.0   | 1.5   | 2.0   | 2.5   | 3.0   | 3.5   | 4.0   | 4.5   | 5.0   |
| $y$ | 0.004 | 0.014 | 0.029 | 0.048 | 0.069 | 0.093 | 0.118 | 0.143 | 0.169 | 0.195 |

  
b) 

|     |       |       |       |       |       |       |       |       |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $x$ | 0.5   | 1.0   | 1.5   | 2.0   | 2.5   | 3.0   | 3.5   | 4.0   | 4.5   | 5.0   |
| $y$ | 0.004 | 0.016 | 0.034 | 0.057 | 0.083 | 0.111 | 0.141 | 0.172 | 0.203 | 0.234 |
- 7.19 

|     |       |       |       |       |       |       |       |       |       |       |
|-----|-------|-------|-------|-------|-------|-------|-------|-------|-------|-------|
| $x$ | 0.5   | 1.0   | 1.5   | 2.0   | 2.5   | 3.0   | 3.5   | 4.0   | 4.5   | 5.0   |
| $y$ | 0.006 | 0.023 | 0.050 | 0.084 | 0.123 | 0.168 | 0.216 | 0.266 | 0.317 | 0.369 |
- 7.21  $T(5 \text{ días}) = 66.82$  con  $h = 12$  horas y RK-4
- 7.22 57 g, usando RK-4 como inicializador y  $h = 1$  día
- 7.23 ciclo = 26 unidades de tiempo
- 7.24  $T_2 = 106.7$  con  $h = 0.25$
- 7.25 a)  $y(1) = 1$  b)  $y(2) = 3.38629$   
c)  $y(2) = 10.04277$  d)  $y(.5) = 2$
- 7.28 a)  $y(2) = 0.13534$  b)  $y(2.5) = 5.25193$  c)  $y(1) = 0.87628$   
d)  $y(-1) = 1.35914$  e)  $y(1.5) = 2.272$
- 7.31  $y(1) = 0.36788$   $z(1) = -0.36788$

7.32  $y(1) = 0.19876$

7.36 con RK-4

|                   |                   |                |
|-------------------|-------------------|----------------|
| a) $y(1) = 1$     | $z(1) = -0.36788$ |                |
| b) $y(2) = 1.97$  | $z(2) = 1.62$     |                |
| c) $y(3) = 34.04$ | $u(3) = 37.37$    | $v(3) = 40.74$ |

7.37  $N_A = 0.02383$        $N_B = 0.12319$        $N_C = 0.85298$   
 con  $h = 10$  min y RK-4

## CAPÍTULO 8

8.1 a1) Elíptica      a2) parabólica      a3) hiperbólica

b) Elíptica en  $x < 0$  o en  $y < 0$

Parabólica en  $x = 0$  o en  $y = 0$

Hiperbólica en  $x > 0, y > 0$  o en  $x < 0, y < 0$

c) Parabólica en el plano  $x$ - $y$

d) Elíptica en el plano  $x$ - $y$

e) Hiperbólica en el plano  $x$ - $y$

8.2 Elíptica en  $y > 0$ ;       $-\infty < x < \infty$

Parabólica en  $y = 0$ ;       $-\infty < x < \infty$

Hiperbólica en  $y < 0$ ;       $-\infty < x < \infty$

8.5 
$$\frac{T_{i,j,k+1} - T_{i,j,k}}{\Delta t} = \alpha \frac{T_{i+1,j,k} - 2T_{i,j,k} + T_{i-1,j,k}}{\Delta x^2}$$

$$+ \alpha \frac{T_{i,j+1,k} - 2T_{i,j,k} + T_{i,j-1,k}}{\Delta y^2}$$

8.6 El algoritmo matemático para resolver este PVF por el método explícito es

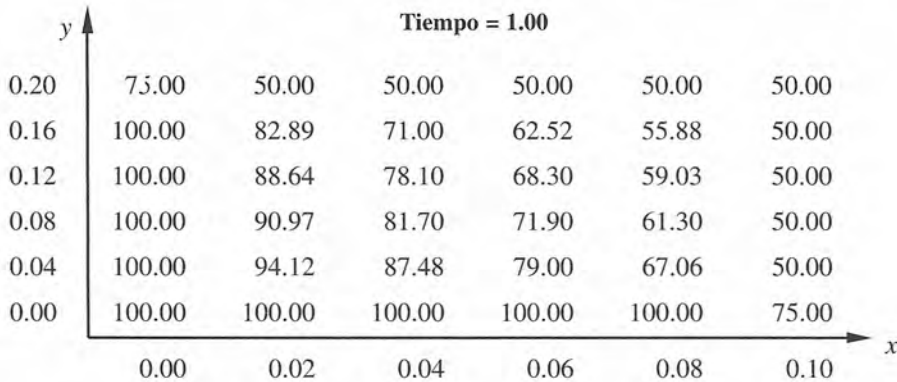
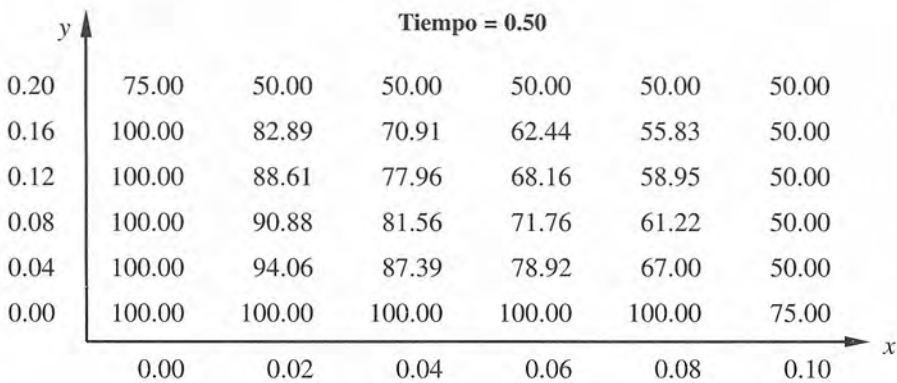
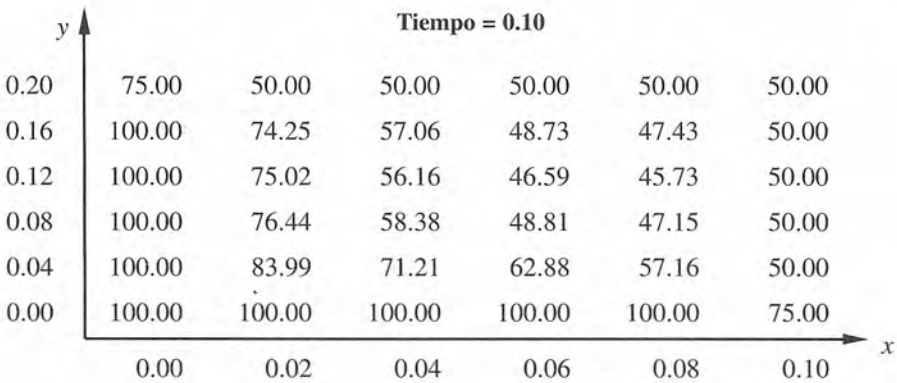
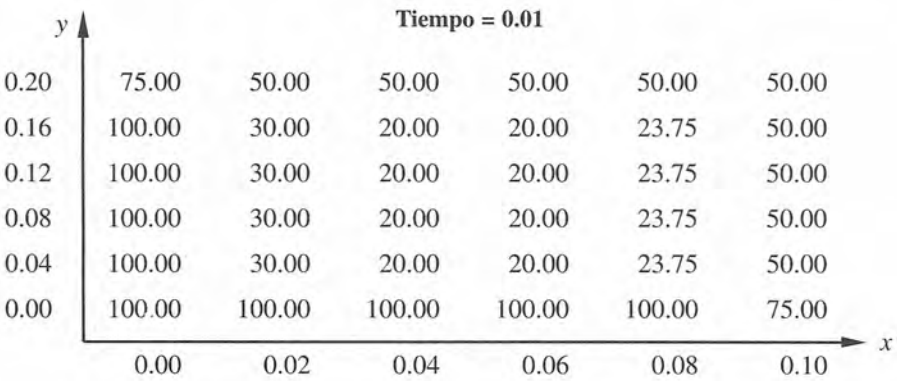
$$T_{i,j,k+1} = T_{i,j,k} + \alpha \frac{\Delta t}{\Delta x^2} [T_{i+1,j,k} - 2T_{i,j,k} + T_{i-1,j,k}]$$

$$+ \alpha \frac{\Delta t}{\Delta y^2} [T_{i,j+1,k} - 2T_{i,j,k} + T_{i,j-1,k}]$$

Con  $\Delta x = 0.02$ ,  $\Delta y = 0.02$ ,  $\Delta t = 0.01$ , se anotan algunos resultados

**Tiempo = 0.00**

| y \ x | 0.00  | 0.02  | 0.04  | 0.06  | 0.08  | 0.10  |
|-------|-------|-------|-------|-------|-------|-------|
| 0.20  | 56.67 | 35.00 | 35.00 | 35.00 | 35.00 | 40.00 |
| 0.16  | 60.00 | 20.00 | 20.00 | 20.00 | 20.00 | 35.00 |
| 0.12  | 60.00 | 20.00 | 20.00 | 20.00 | 20.00 | 35.00 |
| 0.08  | 60.00 | 20.00 | 20.00 | 20.00 | 20.00 | 35.00 |
| 0.04  | 60.00 | 20.00 | 20.00 | 20.00 | 20.00 | 35.00 |
| 0.00  | 73.33 | 60.00 | 60.00 | 60.00 | 60.00 | 56.67 |





- 8.9** Con  $\Delta x = 2$  cm,  $\Delta t = 0.5$  cm,  $\lambda = 0.1275$  y el método de Crank Nicholson, se anotan algunos resultados

| $t$ ( min ) | $x$ ( cm ) |        |        |        |        |      |
|-------------|------------|--------|--------|--------|--------|------|
|             | 0.0        | 4.0    | 8.0    | 12.0   | 16.0   | 20.0 |
| 0.0         | 1.0100     | 0.0200 | 0.0200 | 0.0200 | 0.0200 | 0.01 |
| 1.0         | 2.0000     | 0.0655 | 0.0203 | 0.0200 | 0.0195 | 0.00 |
| 5.0         | 2.0000     | 0.4481 | 0.0574 | 0.0213 | 0.0157 | 0.00 |
| 10.0        | 2.0000     | 0.7631 | 0.1815 | 0.0397 | 0.0143 | 0.00 |

- 8.10** Con  $\Delta x = 2$  cm,  $\Delta t = 0.5$  cm,  $\lambda = 0.1275$  y el método de Crank Nicholson, se anotan algunos resultados

| $t$ ( min ) | $x$ ( cm ) |        |        |        |        |      |
|-------------|------------|--------|--------|--------|--------|------|
|             | 0.0        | 4.0    | 8.0    | 12.0   | 16.0   | 20.0 |
| 0.0         | 1.0100     | 0.0200 | 0.0200 | 0.0200 | 0.0200 | 0.51 |
| 1.0         | 2.0000     | 0.0655 | 0.0203 | 0.0202 | 0.0425 | 1.00 |
| 5.0         | 2.0000     | 0.4481 | 0.0582 | 0.0402 | 0.2319 | 1.00 |
| 10.0        | 2.0000     | 0.7640 | 0.1923 | 0.1214 | 0.3896 | 1.00 |

- 8.11** Con el uso del método implícito con  $\Delta x = 0.25$ ,  $\Delta t = 0.01$  y  $\lambda = 0.01$ , se anotan algunos resultados

| $t$ ( hrs ) | $x$ ( pies ) |       |       |       |       |
|-------------|--------------|-------|-------|-------|-------|
|             | 0.0          | 0.25  | 0.50  | 0.75  | 1.00  |
| 0.0         | 100.00       | 4.95  | 9.59  | 13.63 | 50.00 |
| 1.0         | 100.00       | 63.47 | 42.86 | 40.67 | 50.00 |
| 5.0         | 100.00       | 86.87 | 74.10 | 61.87 | 50.00 |
| 10.0        | 100.00       | 87.49 | 75.00 | 62.49 | 50.00 |

- 8.12** Con  $\Delta x = 0.1$ ,  $\Delta t = 0.24$ ,  $\lambda = 0.4166666$ , se anotan algunos resultados para  $x < 0.5$ , ya que la distribución de temperaturas es simétrica.

| $t$ ( hrs ) | $x$ ( pies ) |       |       |       |       |       |
|-------------|--------------|-------|-------|-------|-------|-------|
|             | 0.0          | 0.1   | 0.2   | 0.3   | 0.4   | 0.5   |
| 0           | 60.00        | 20.00 | 20.00 | 20.00 | 20.00 | 20.00 |
| 2           | 100.00       | 75.07 | 54.02 | 39.09 | 30.62 | 27.93 |
| 6           | 100.00       | 88.02 | 77.22 | 68.65 | 63.16 | 61.27 |
| 8           | 100.00       | 91.36 | 83.57 | 77.38 | 73.41 | 72.05 |
| 12          | 100.00       | 95.50 | 91.44 | 88.23 | 86.16 | 85.45 |
| 14          | 20.00        | 44.44 | 64.04 | 76.62 | 82.93 | 84.74 |
| 20          | 20.00        | 27.96 | 35.15 | 40.85 | 44.51 | 45.77 |
| 40          | 20.00        | 20.30 | 20.58 | 20.80 | 20.94 | 20.99 |
| 60          | 20.00        | 20.01 | 20.02 | 20.03 | 20.04 | 20.04 |

8.14 Con  $\Delta x = 0.05$ ,  $\Delta t = 0.001$ ,  $\lambda = 1.76984127$ ,  $\beta = 28.57188572$ , se anotan algunos resultados

| $t$ (hr) | 0.0    | 0.2    | 0.4    | 0.6   | 0.8   | 1.0   |
|----------|--------|--------|--------|-------|-------|-------|
| 0.000    | 140.00 | 78.00  | 76.00  | 74.00 | 72.00 | 69.00 |
| 0.001    | 200.00 | 80.71  | 75.82  | 73.83 | 71.84 | 68.00 |
| 0.006    | 200.00 | 118.48 | 83.00  | 73.83 | 70.80 | 68.00 |
| 0.020    | 200.00 | 140.07 | 103.90 | 83.94 | 73.76 | 68.00 |
| 0.100    | 200.00 | 146.54 | 113.79 | 93.12 | 79.09 | 68.00 |

8.15

$$\frac{\partial T}{\partial t} = \frac{k}{\rho C_p} \frac{\partial^2 T}{\partial x^2} - \frac{k}{\rho C_p (1-x)} \frac{\partial T}{\partial t} - \frac{2 \cdot 1.0625 \cdot a \cdot h}{0.25 (1-x) \rho C_p} (T - 68)$$

C.I  $T(x, 0) = 68^\circ \text{F}$

C.F.1  $T(0, t) = 200^\circ \text{F}$

C.F.2  $T(1, t) = 68^\circ \text{F}$

# ÍNDICE ANALÍTICO

- ajuste exacto, 318
  - de mínimos cuadrados, 318
- algoritmo de Aitken, 63, 116
  - de Crank-Nicholson, 565
  - de Crout, 251
  - de la posición falsa, 115
  - de Thomas, 358
  - del método trapezoidal, 397
- algoritmos de Taylor, 475
  - de Runge- Kutta, 482
- ángulo entre vectores, 142, 143
- aproximación cúbica de trazador, 357
  - cúbica segmentaria de Bessel, 355, 356
  - cúbica segmentaria de Hermite, 354, 355
  - multilineal con mínimos cuadrados, 359, 367
  - polinomial, 348, 393
  - polinomial de Lagrange, 323
  - polinomial de Newton, 333, 335
  - polinomial por mínimos cuadrados, 318
  - polinomial segmentaria, 352
  - polinomial simple, 319, 322, 382, 455
- asíntotas, 68
- asociatividad de la multiplicación de matrices, 132
  - de la suma de matrices, 129
  - del producto de matrices, 141
- bit, 3
- byte, 9
- cálculo de inversas, 172, 173
  - del determinante, 165, 169
- característica, 10
- cifras significativas, 14
- combinación lineal de vectores, 145
- condición inicial, 539, 542, 557
  - suficiente, 40
- condiciones combinadas, 575
  - frontera, 539, 543, 574
  - frontera combinadas, 578, 581
  - frontera de Dirichlet, 574, 576
- conjuntos ortogonales de vectores, 148
- conmutatividad, 141
  - de la suma de matrices, 127
- convergencia, 40, 53, 60, 113, 214, 215, 219, 261, 275, 316, 561
  - aceleración de, 62, 63, 218, 222, 281
  - velocidad de, 115, 275
  - monotónica, 44
  - oscilatoria, 44
- conversión de números enteros, 4
  - de números fraccionarios, 7
- correctores de Adams-Moulton, 491
- criterio de ajuste exacto, 391, 434
  - de convergencia, 38, 48, 57, 62, 222
  - de convergencia, 458
  - de exactitud, 54, 56
  - de mínimos cuadrados, 312
  - de ortogonalidad, 150, 157
- cuadratura de Gauss-Legendre, 416, 417, 418, 420, 421, 422, 423, 425, 430, 447, 448, 449, 460, 462
  - de Gauss-Laguerre, 450, 461
- cuenta de operaciones, 87, 174
- curva de nivel, 291
- determinante de una matriz, 165, 206
  - normalizado, 199, 201
- diagonal principal, 133, 165, 239
- diferenciación numérica, 395, 434
- diferencias centrales, 540, 541, 544, 554, 565, 586
  - centrales de orden par, 387
  - divididas, 329, 370, 437
  - divididas centrales, 572
  - divididas de orden cero, 330
  - finitas, 542, 586
  - finitas hacia delante, 395, 339
  - hacia atrás, 488, 489, 540, 554, 556, 582
  - hacia delante, 437 540, 565
- dígitos binarios, 3
  - de exactitud, 12
  - de seguridad, 30
  - significativos, 15
- dirección de descenso más brusco, 291
  - de exploración, 281
- distancia entre dos vectores, 143
- distributividad, 141
  - de la suma de matrices, 129
  - del producto de matrices, 131
- divergencia, 42, 214
  - monotónica, 44
  - oscilatoria, 44
- división sintética, 235
- doble precisión, 12, 24
- dominio de concavidad, 68
  - de convexidad, 68
  - de definición, 68
- ecuación de Beattie- Bridgeman, 67, 120
  - de conducción de calor en régimen transitorio, 536
  - de estado, 67
  - de estado de Redlich-Kwong, 120
  - de estado de Van der Waals, 94, 120
  - de Fourier, 446
  - de una onda en dimensión, 536
  - de Poisselle, 254
  - general de la conducción de calor, 534
- ecuaciones polinomiales con coeficientes reales, 71



- eliminación de Gauss, 162, 165, 181, 185, 241, 246, 251, 268
  - de Gauss con pivoteo, 172, 187
  - de Jordan, 170, 173, 242
- error absoluto, 12, 16, 476
  - de discretización, 18, 561, 563
  - de redondeo, 12, 59, 222, 473, 477
  - de truncamiento, 411, 438, 457, 473
  - en por ciento, 12
  - porcentual, 476
  - relativo, 12, 16, 18
- errores de redondeo, 18, 58, 207
  - de salida, 19
- estabilidad, 22, 91
- estimación de errores en la aproximación, 437
- extrapolación de Richardson, 412, 413
- factor de fricción, 124
  - de tamaño de etapa, 282, 283
- factores cuadráticos, 92, 93
- factorización de matrices, 181, 183, 185
  - de matrices con pivoteo, 188
- fila pivote, 167
- fórmula de Chebyshev, 120
  - de Francis, 97, 113
  - de Halley
  - de inversión matricial, 277
  - de Newton en diferencias finitas hacia delante, 398
  - fundamental de Newton, 349, 435
  - hacia delante de Gauss, 388
  - modificada de Lin, 91
- fórmulas de cuadratura gaussiana, 395
  - de Newton-Cotes, 395, 401
- fronteras irregulares, 576
- función de transferencia, 99
  - escalar, 290
  - suma de residuos, 285, 288
- gradiente, 290
- independencia de conjuntos, 145, 146, 157
  - lineal, 152
- integración de Romberg, 412, 413, 457
  - numérica, 471
  - trapezoidal, 485
- integrales impropias, 450
  - múltiples, 425
- interpolación, 318, 319, 320, 460
  - inversa, 382, 383
  - lineal inversa, 494, 501
- interpretación geométrica de la independencia lineal, 147
- intervalo de búsqueda, 285, 288
- ley de acción de masas, 300
  - de Beer, 226
  - de Dalton, 106
  - de Henry, 223, 307
  - de Kirchhoff, 253
  - de Raoult, 106
  - de paralelogramo, 147
- longitud de un vector, 141
- lugar geométrico de las raíces, 113
- mantisa, 10
- matrices conformes, 130
  - elementales, 202
  - especiales, 133
  - sumables, 129
- matriz, 125
  - atómica, 231, 232, 233
  - aumentada, 160, 163, 243
  - bandeada, 177, 201, 251
  - casi singular, 129
  - cero, 127
  - coeficiente, 160, 163, 177, 181, 185, 201, 201, 207, 241, 253
  - coeficiente densa, 207
  - coeficiente diagonalmente dominante, 215, 227
  - coeficiente positivamente definida, 219
  - coeficiente simétrica, 222
  - columna, 137
  - de nodos, 225
  - de orden  $n$ , 126
  - diagonal, 133, 177, 239
  - diagonal dominante, 201
  - dispersa, 177
  - identidad, 133, 136, 202
  - inversa, 135
  - jacobiana, 270, 271, 283, 300, 301, 311, 313, 315
  - mal condicionada, 159
  - no singular, 135
  - pentadiagonal, 177, 246
  - permutadora, 135, 136, 239
  - positiva definida, 193, 201, 239, 249
  - simétrica, 133, 177, 201, 239, 249
  - singular, 135, 159, 241
  - transpuesta, 133, 138
  - triangular inferior, 194, 239
  - triangular superior, 133, 165, 182, 239
  - tridiagonal, 177, 219, 247
  - tridiagonal por bloques, 242, 243
  - unitaria, 133
- método de Aitken, 64
  - de bisección, 53, 56, 57, 66, 109, 115, 116
  - de Broyden, 272, 276, 313, 316
  - de Cholesky, 193, 251
  - de Crank-Nicholson, 564, 570, 584, 587, 588
  - de Crout, 182, 183, 251
  - de desplazamientos simultáneos, 209, 216, 219, 262, 273
  - de desplazamientos sucesivos, 209, 216, 219, 262, 269, 273
  - de Doolittle, 182, 183, 251
  - de Doolittle con pivoteo, 188, 192
  - de Dufort-Frankel, 572, 573
  - de Euler, 470, 475, 478, 480, 507
  - de Euler modificado, 477, 479, 482, 484, 484, 509
  - de Gauss-Seidel, 207, 209, 212, 214, 214, 218, 222, 234, 252, 259
  - de Gram-Schmidt, 162, 240
  - de Horner, 80, 84, 85, 87, 114



- de Jacobi, 207, 209, 211, 214, 218, 252, 259, 262
- de la secante, 49, 59, 60, 74, 115, 118
- de la secante, error, 61
- de la secante, interpretación geométrica, 54
- de Lagrange, 382
- de Laguerre, 118
- de Lin, 89, 90
- de mínimos cuadrados, 313, 391
- de Müller, 73, 79, 99, 117, 118
- de Newton-Raphson, 46, 49, 56, 71, 72, 98, 101, 105, 107, 114, 115, 117, 121, 234, 298
- de Newton-Raphson, error, 61
- de Newton-Raphson, fallas, 49
- de Newton-Raphson con optimización de  $t$ , 316
- de Newton-Raphson modificado, 272, 312
- de Newton-Raphson multivariable, 265, 300, 305, 310, 311, 313
- de Newton-Raphson-Horner, 88, 90
- de posición falsa, 27, 53, 54, 55, 56, 66, 67, 95, 115, 116, 383
- de punto fijo, 34, 46, 56, 62, 113, 207, 269, 311
- de punto fijo multivariable, 259, 311, 312, 313
- de Richardson, 572, 573
- de Richmond, 115, 121
- de Romberg, 416
- de segundo orden de convergencia, 46
- de Simpson, 398, 406
- de Simpson compuesto, 404, 410, 456
- de Simpson 1/3, 461, 485
- de Simpson 3/8, 456
- de Simpson 3/8 compuesto, 457
- de Steffensen, 64, 96, 116
- de Thomas, 178, 179
- de Wttaker, 114
- del descenso de máxima pendiente, 290, 316
- del eigenvalor dominante, 315, 316
- explícito, 545, 557, 581, 587
- Illinois, 66
- implícito, 554, 587
- Regula-Falsi, 53
- trapezoidal, 395, 404, 406, 417
- trapezoidal compuesto, 402, 408
- método de Bairstow, 299-332
  - factor cuadrático, 299-304
- método de disparo, 493-496
- método de Newton-Raphson, 266
  - interpretación geométrica, 268
  - suma de residuos al cuadrado, 286
- métodos cuasi-Newton, 313
  - de Adams-Bashford, 492, 495, 529
  - de Adams-Moulton, 492, 495, 529
  - de Bailey, 119
    - compuestos de integración, 402
  - de dos puntos, 59, 61, 66
  - de Lambert, 119
  - de mínimos cuadrados, 252, 452
  - de múltiples pasos, 484
  - de Newton-Cotes, 395
  - de predicción-corrección, 492, 495, 501
  - de primer orden, 61
  - de relajación, 218
  - de Runge-Kutta, 480, 482, 501, 510, 513, 514, 516, 520, 526, 529
  - de Taylor, 474, 475, 479, 528
  - de un solo paso, 484
  - SOR, 219, 253, 287
- modelo de Ostwald-De Waele, 124
- multiplicación de matrices, 130
  - de vectores, 139
- norma euclídeana, 236, 239
- número de máquina, 30
  - de Reynolds, 124
  - en una computadora, 9
  - reales (punto flotante), 10
  - enteros, 9
  - normalizados, 15
  - reales, 125
- operaciones elementales con matrices, 126
- operador de diferencias hacia atrás, 339
  - de diferencias hacia delante, 339
  - en diferencias centrales, 386
- orden de convergencia, 44, 59, 60, 61, 118, 564
  - de precedencia, 299
  - de una ecuación diferencial, 469
- otorgonalización, 150, 157, 233
  - de Gram-Schmidt, 150, 157, 232
- overflow, 16
- palabra de memoria, 9
- partición de ecuaciones, 257, 300
- pivote, 167
- pivoteo parcial, 167, 192, 202
  - total, 202, 235
- polinomio característico, 233, 234
  - de grado  $n$  en diferencias divididas, 349
  - de interpolación, 488, 493
  - de Lagrange, 441
  - de Newton, 337
  - de Newton en diferencias divididas, 384, 443, 463
  - de Newton en diferencias finitas, 338
  - de Newton en diferencias finitas hacia atrás, 340
  - de Newton en diferencias finitas hacia delante, 340
- polinomios complejos, 71
  - de Lagrange, 323, 325, 370, 542
- positividad, 141
- precisión sencilla, 18, 27
- predictor, 478
- primera diferencia central, 387
  - dividida, 330
  - hacia atrás, 339
- problema de valores en la frontera, 111-113, 492, 539, 543, 587
- problemas de valor inicial, 99, 469, 470
- producto de matrices por un escalar, 128
  - punto de vectores, 140

- propagación de errores, 19
- puntos de inflexión, 68
  - singulares de una función, 68
- raíces complejas, 71, 72, 73, 115, 117
  - reales, 73, 117
  - reales no repetidas, 46
  - repetidas, 73
- rango, 143, 158, 240
  - de la matriz coeficiente, 230, 241
  - de una matriz, 158, 232, 233
- reducción de ecuaciones, 256, 295
- regla de Cramer, 362
  - de Horner, 114
  - de las mallas de Kirchhoff, 228
  - de los nodos de Kirchhoff, 228
  - de Simpson, 404, 406, 407, 421, 485
  - del trapecioide, 445
  - trapezoidal, 395, 407, 447
- reordenamiento de ecuaciones, 259
- residuo de una función, 2, 4
- segunda diferencia dividida, 334
  - hacia atrás, 339
  - hacia delante, 339
- serie de Fibonacci, 285
  - de Taylor, 60, 411, 474, 475, 476, 477, 480, 482, 539, 561, 586
- sistema binario, 2, 3
  - consistente, 161
  - de control lineal, 99
  - decimal, 3
  - diagonal dominante, 216
  - homogéneo, 160, 230, 237
  - inconsistente, 161
  - no homogéneo, 160
  - octal, 4
  - simétrico, 191
  - tridiagonal, por bloques, 244
- sistemas de ecuaciones diferenciales, 501
  - de ecuaciones lineales, 160
  - de ecuaciones mal condicionados, 197, 198, 201, 219, 251
  - dispersos, 225
  - especiales, 177
  - lineales simétricos, 250
- solución única, 161
- suma de matrices, 126
- sustitución regresiva, 84, 163, 164, 169, 178, 179, 182, 243
- tanteo de ecuaciones, 257, 310
- teorema binominal, 61
  - de Bolzano, 55
- teoría de las vibraciones, 240
- tiempo de máquina, 473
- transformada inversa de Laplace, 100, 111
- transformada de Laplace, 99, 111
- triangulación, 163, 164, 165, 169, 178, 184, 197, 241
- underflow*, 16
- valor característico dominante, 253, 254
  - inicial, 234
- valores característicos, 234, 253
  - complejos, 117
  - iniciales, 67, 95, 104, 107, 109, 113, 257
  - iniciales, búsqueda, 66
- vector característico, 253, 254
  - cero, 143, 147
  - de exploración, 283
  - de términos independientes, 160
    - gradiente, 291
  - dominante, 253, 254
  - incógnita, 160
  - inicial, 209, 210, 214, 215
  - linealmente dependiente, 146, 233, 241
  - linealmente independiente, 146, 240
  - residuo, 219
  - solución, 207, 213, 219
- vectores, 137
  - propios, 235, 237
- vibración en estado estacionario, 510
- visualización de raíces, 25

Esta obra se terminó de imprimir en junio del 2006  
 en los talleres de Fusión Servicios Gráficos  
 Miguel Negrete No.34 - 401  
 Col. Niños Héroes C.P. 03440 México D.F.





**E**n el análisis numérico se concretan una serie de métodos o algoritmos, cuya característica principal es la posibilidad de obtener resultados numéricos de problemas matemáticos de cualquier tipo a partir de números y operaciones aritméticas. La finalidad de este libro es facilitar el estudio y uso racional de dichos algoritmos en diferentes áreas de la ingeniería y las ciencias.

En esta segunda edición de *Métodos Numéricos Aplicados a la Ingeniería* se intenta no caer en un curso más de recetas matemáticas desvinculadas y sin sentido, por lo que los autores han desarrollado el material del libro alrededor de tres ideas fundamentales: el punto fijo, la eliminación de Gauss y la aproximación de funciones. Para lograr esto, se emplean como recursos didácticos para cada método o problema diferentes sistemas de representación: el gráfico, el tabular y el algebraico.

Con el fin de que el lector vea claramente la relación entre los métodos que se estudian en el libro y su aplicación en el contexto real, al final de cada capítulo se resuelven problemas de aplicación de diferentes áreas. De igual manera, se hace hincapié en el uso de herramientas como la calculadora y la computadora así como en la importancia de la visualización de los problemas.

#### CAMBIOS ESENCIALES EN ESTA EDICIÓN

- Nuevos ejemplos, ejercicios y problemas de aplicación a diferentes ingenierías.
- Cambios importantes en las secciones de valores y vectores propios, en problemas de valores a la frontera y en el método de Bairstow, así como en la sección solución y análisis de la ecuación de onda en una dimensión.
- Problemas resueltos con guiones en Matlab y la calculadora TI-92 a lo largo de todo el texto.
- CD diseñado especialmente para esta edición con: sugerencias para resolver un gran número de ejemplos y ejercicios con Matlab, Mathcad y Mathematica; programas fuente en Visual Basic y sus respectivos ejecutables para la solución de más ejercicios; ligas a sitios de interés y otros apoyos adicionales para el lector.
- Sitio Web con: actualizaciones del material del libro (nuevas soluciones, problemas y programas, entre otros).



COMPAÑÍA EDITORIAL CONTINENTAL

ISBN 970-24-0258-1

