DO

J. M. Sanz-Serna

# DIEZ LECCIONES DE CÁLCULO NUMÉRICO

segunda edición revisada y ampliada

Universidad de Valladolid

# Índice general

**Prólogo** **11**

**1. Introducción al análisis numérico** **13**
1.1. Objetivos del análisis numérico . . . . . 13
1.2. Necesidad del análisis numérico . . . . . 14
1.3. Perspectiva histórica . . . . . 15
1.4. Errores . . . . . 15
1.4.1. Error absoluto y error relativo . . . . . 15
1.4.2. Cotas y estimaciones del error . . . . . 16
1.5. Representación de los números reales . . . . . 17
1.5.1. Punto flotante . . . . . 17
1.5.2. Mantisa y exponente . . . . . 18
1.5.3. La aritmética de punto flotante . . . . . 18
1.6. Costo operativo y eficiencia . . . . . 20

**2. Interpolación de Taylor** **23**
2.1. Polinomios de Taylor . . . . . 23
2.2. Notaciones de Landau . . . . . 27
2.3. Caracterización analítica . . . . . 29
2.4. Cotas del error . . . . . 32
2.5. Convergencia . . . . . 34
2.6. Cambio de escala . . . . . 36

**3. Interpolación de Lagrange** **39**
3.1. Polinomio de Lagrange . . . . . 39
3.1.1. Planteamiento . . . . . 39
3.1.2. La forma de Lagrange . . . . . 41
3.2. Forma de Newton . . . . . 45

3.2.1. Construcción por recurrencia . . . . . 45
3.2.2. Cambio de orden . . . . . 47
3.2.3. Diferencias divididas de una función . . . . . 47
3.2.4. Cálculo de las diferencias divididas . . . . . 48
3.2.5. La tabla de diferencias divididas . . . . . 49
3.2.6. Comparación entre formas . . . . . 50
3.3. Cotas de error . . . . . 53
3.3.1. El error de interpolación . . . . . 53
3.3.2. Diferencias divididas y derivadas . . . . . 55
3.4. Convergencia . . . . . 56

**4. Interpolación polinómica a trozos** **59**
4.1. Interpolación lineal a trozos . . . . . 59
4.1.1. Funciones lineales a trozos . . . . . 59
4.1.2. Funciones lineales a trozos como interpolantes . . . . . 61
4.1.3. Comparación con la interpolación de Lagrange polinómica 62
4.2. Polinomios a trozos . . . . . 63

**5. Cuadratura numérica** **67**
5.1. Introducción . . . . . 67
5.1.1. El problema de la cuadratura numérica . . . . . 67
5.1.2. Reglas de cuadratura . . . . . 68
5.1.3. Grado de exactitud . . . . . 69
5.2. Obtención de reglas de cuadratura . . . . . 70
5.2.1. El método interpolatorio . . . . . 70
5.2.2. El método de coeficientes indeterminados . . . . . 73
5.2.3. El método del desarrollo de Taylor . . . . . 74
5.3. Error de cuadratura . . . . . 77
5.4. Reglas compuestas . . . . . 80
5.4.1. Concepto de regla compuesta . . . . . 80
5.4.2. Errores en las reglas compuestas . . . . . 81
5.4.3. Relación con la interpolación polinómica a trozos . . . . . 85

**6. Derivación numérica** **87**
6.1. Introducción . . . . . 87
6.2. Obtención de fórmulas . . . . . 88
6.3. Error de la derivación numérica . . . . . 90
6.4. La derivación numérica problema mal puesto . . . . . 92

**7. Operadores de diferencias** **93**
7.1. El operador $\Delta$ . . . . . 93
7.1.1. El operador $\Delta$ . . . . . 93
7.1.2. Las potencias de $\Delta$ . . . . . 94
7.1.3. Diferencias a partir de valores de la función . . . . . 94
7.1.4. Valores de la función a partir de diferencias . . . . . 95
7.1.5. Diferencias y diferencias divididas . . . . . 96
7.2. La fórmula de Newton . . . . . 98
7.3. Otros operadores de diferencias . . . . . 100
7.4. El cálculo de diferencias finitas . . . . . 101

**8. Ecuaciones en diferencias** **105**
8.1. Preliminares . . . . . 105
8.1.1. Concepto de ecuación en diferencias . . . . . 105
8.1.2. Problemas de valores iniciales . . . . . 106
8.1.3. Escritura en términos de diferencias . . . . . 107
8.2. Ecuaciones en diferencias lineales . . . . . 108
8.2.1. Definiciones . . . . . 108
8.2.2. Operador asociado . . . . . 109
8.2.3. Ecuación homogénea . . . . . 109
8.2.4. La ecuación no homogénea . . . . . 111
8.3. Coeficientes constantes . . . . . 112
8.3.1. La ecuación característica . . . . . 112
8.3.2. Caso de raíces multiples . . . . . 115
8.3.3. Comportamiento de las soluciones cuando $j \to \infty$ . . . . 115

**9. Ecuaciones no lineales** **119**
9.1. Introducción . . . . . 119
9.2. Bisección . . . . . 119
9.3. Método de la secante . . . . . 122
9.4. Iteración de punto fijo . . . . . 124
9.4.1. Descripción del método . . . . . 124
9.4.2. Atracción local . . . . . 126
9.4.3. Caso de convergencia cuadrática . . . . . 129
9.4.4. Puntos fijos repulsores . . . . . 132
9.4.5. Comportamiento global de la iteración . . . . . 132
9.5. El método de Newton . . . . . 136

**10. Solución de sistemas lineales** **141**
10.1. Introducción . . . . . 141
10.2. La eliminación Gaussiana . . . . . 142
10.2.1. Un ejemplo . . . . . 142
10.2.2. Eliminación de las incógnitas . . . . . 143
10.2.3. Sustitución regresiva . . . . . 145
10.2.4. Aplicabilidad del método . . . . . 146
10.3. Factorización LU . . . . . 148
10.3.1. La eliminación de incógnitas, proceso matricial . . . . . 149
10.3.2. La factorización implícita en la eliminación . . . . . 150
10.3.3. Factorizaciones LU de una matriz . . . . . 152
10.4. Implementación . . . . . 154
10.4.1. Subrutina de factorización . . . . . 155
10.4.2. Subrutina de sustitución progresiva . . . . . 156
10.4.3. Subrutina de sustitución regresiva . . . . . 156
10.4.4. Solución de sistemas lineales . . . . . 157
10.4.5. Cálculo de la inversa . . . . . 157
10.4.6. Cálculo de determinantes . . . . . 157
10.5. Costo operativo . . . . . 158
10.6. Pivotaje . . . . . 160
10.6.1. Motivación . . . . . 160
10.6.2. Pivotaje parcial . . . . . 161
10.6.3. Interpretación matricial . . . . . 163
10.6.4. Implementación . . . . . 164

**Indice de materias** **168**

# Prólogo

*Del prólogo de la primera edición:*

Persigue esta obra ofrecer a los alumnos de facultades y escuelas técnicas un primer contacto con el cálculo numérico. Si se considera en su totalidad, cubre un curso cuatrimestral de cálculo numérico. También pueden extraerse de ella materiales para cubrir aspectos de cálculo numérico en asignaturas de álgebra lineal, cálculo infinitesimal, etc.

Decidí llamarla "Diez lecciones de cálculo numérico", y no "Cálculo numérico" u otra cosa más pretenciosa, para subrayar que se trata sólo de una iniciación al tema y no, en modo alguno, de un trabajo con aspiraciones de completitud. La obra deja fuera algunos temas básicos esenciales en el cálculo efectivo como mínimos cuadrados, splines, el algoritmo QR o la descomposición en valores singulares. La elección de los puntos incluidos se ha hecho siguiendo el criterio de abordar materias que pudieran tratarse adecuadamente con un equipaje matemático sucinto: sólo se presuponen al lector conocimientos básicos de álgebra lineal y cálculo infinitesimal de una variable. Por ello la obra puede emplearse en el primer curso de la enseñanza universitaria, como lo corrabora nuestra propia experiencia. Además, por razones pedagógicas, he buscado limitar el número de páginas, tratando de invertir el proceso según el cual cada generación de alumnos debe habérselas con libros de texto cuyo peso es doble del que acarrearon sus predecesores.

Si fui flexible en la elección de temas y en la amplitud de la obra, no lo he sido en otra serie de aspectos. Creo que es obligado exponer desde el principio al alumno a los rasgos peculiares del cálculo numérico, es decir enseñarle a convivir con errores y a apreciar la importancia del costo operativo y la eficiencia. También he tratado de usar parte del espacio disponible para intentar que se capte la idea de que el cálculo numérico, además de una disciplina académica que se estudia en la aulas, es parte de un esfuerzo humano global desarrollado a lo largo de la historia para resolver problemas científicos y técnicos.

El libro contiene un número no despreciable de ejercicios y problemas. Al-

gunos —no demasiados— piden aplicar con papel y lápiz los algoritmos descritos; los profesores que usen este texto o incluso los propios alumnos no tendrán mayor dificultad en crear ejercicios similares suplementarios con sólo cambiar los valores de los datos. Otros problemas son de índole más teórica y otros piden escribir programas de ordenador y ejecutarlos. Nunca se insistirá lo bastante en la necesidad de resolver problemas cuando se estudian matemáticas. Ante todo, las teorías se han ido construyendo en la historia para resolver problemas específicos; puede haber matemáticas sin teoremas mas no sin problemas. De otra parte, aprender matemáticas es un proceso donde se trata de aprender a *hacer* algo, no de adquirir conocimientos. Por eso el proceso de estudio de las matemáticas debe ser más similar al aprendizaje del piano, de la guitarra, del tenis, de la conducción de automóviles que al de la historia o la literatura. Para alcanzar el objetivo perseguido, usar racionalmente los métodos numéricos, es pues menester practicar y practicar hasta lograr soltura. Se necesitan problemas de todo tipo: de rutina que afiancen el cálculo y teóricos que obliguen a la reflexión. Y en el caso particular del cálculo numérico es *imprescindible* usar el ordenador, o al menos una calculadora de bolsillo, para experimentar tangiblemente con las realidades que se estudian.

*Notas a la segunda edición:*

En esta nueva edición, que dedico a la memoria de mi padre, he añadido contenidos y eliminado erratas, al tiempo que mejorado la redacción en muy numerosos puntos. Para facilitar el estudio del alumno, ejemplos y problemas aparecen ahora con letra de menor cuerpo.[1] Con ello el número de páginas es el mismo, a pesar de los nuevos materiales.

En el Indice de Materias, las letras C, S, E, P indican, respectivamente Capítulo, Sección, Ejemplo y Problema. Así, S1.1 remite a la Sección 1.1, mientras E1.1.1 que lo hace al Ejemplo 1.1.1 y E1.6.5–7 a los Ejemplos desde el 1.6.5 al 1.6.7.

[1]Con este mismo objetivo algunos comentarios se han trasladado a notas a pie de página.

# Capítulo 1

# Introducción al análisis numérico

## 1.1 Objetivos del análisis numérico

El análisis numérico (o cálculo numérico) es la rama de la matemática que tiene como objetivo el diseño y estudio de métodos que permitan obtener *efectivamente* soluciones numéricas de problemas matemáticos.

**Ejemplo 1.1.1** En muchas situaciones aparece en diversas ciencias la necesidad de calcular la probabilidad[1] de que una variable aleatoria con distribución normal, digamos p. ej. de media 0 y varianza 1, tome un valor comprendido entre ciertos límites, digamos, 0.5 y 0.8. Esta probabilidad viene dada por

$$\frac{1}{\sqrt{2\pi}}\int_{0.5}^{0.8} e^{-x^2/2}dx. \tag{1.1}$$

Hay que conocer pues el valor numérico de esta expresión, que es $\approx 0.096$. El análisis numérico propone métodos que permiten hallar el valor de integrales $\int_a^b f(x)dx$, siendo $f$ una función conocida y $a$, $b$ números reales conocidos. □

En la definición anterior hemos subrayado la palabra *efectivamente*. Queremos manifestar con ella que en análisis numérico no basta con métodos que *en principio* pudieran permitir el cálculo de la cantidad buscada.

**Ejemplo 1.1.2** El valor del determinante de una matriz cuadrada $A$ con elementos $(a_{ij})$ y dimensión $d \times d$ viene dado por

$$\det(A) = \sum \pm a_{1i_1} \cdots a_{di_d}, \tag{1.2}$$

[1] Entender este ejemplo no requiere entender los términos *probabilidad, variable aleatoria,* etc. Comentarios análogos podrían hacerse en muchos otros lugares de este libro.

donde la suma se extiende a todas las $d!$ permutaciones $(i_1, \ldots, i_d)$ de los números $1, 2, \ldots, d$. La fórmula (1.2) permite, en principio, hallar $\det(A)$ cuando $A$ es conocida. Sin embargo, a menos que $d$ sea muy pequeño, no permite el cálculo efectivo. Si $d = 25$ hay más de 15 cuatrillones de permutaciones posibles. Aun imaginando que pudieramos calcular un término de los 25! que componen (1.2) cada billónesima de segundo (cosa imposible con los actuales ordenadores[2]) para hallar todos necesitaríamos 15 billones de segundos, es decir, más de cuatrocientos mil años.

## 1.2 Necesidad del análisis numérico

¿Por qué queremos resolver problemas matemáticos? La ciencia y tecnología son parte esencial del mundo actual. En todas las tareas científicas, desde la creación de nuevos fármacos al cálculo de las órbitas de los satélites de comunicaciones, desde la predicción meteorológica al diseño de coches y aviones, se usan modelos matemáticos. En la simulación de tales modelos y en el análisis de los datos experimentales correspondientes es imprescindible resolver problemas matemáticos.

¿Por qué hacen falta las técnicas del análisis numérico para resolver los problemas matemáticos? Consideremos por ejemplo el cálculo de integrales $\int_a^b f(x)dx$, como en la sección precedente. En algunos casos sencillos, p. ej. $f(x) = x^n$, la integral puede hallarse *analíticamente* sin recurrir a los métodos numéricos, p. ej. hallando una primitiva $F$ del integrando (esto es una función que satisfaga $dF/dx = f$) y formando la diferencia $F(b) - F(a)$. Sin embargo para la 'mayoría' de los integrandos $f$ la primitiva $F$ es o difícil de encontrar o incluso (esto es lo que pasa para (1.1)) no expresable mediante funciones *elementales,* es decir mediante las funciones usuales: polinomios, trigonométricas, etc. En esas circunstancias es o conveniente o imprescindible acudir a los métodos numéricos. De hecho los métodos no numéricos tienen una aplicabilidad muy restringida.

**Ejemplo 1.2.1** La suma $S$ de la serie

$$1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots$$

puede determinarse *analíticamente*, pero con gran dificultad: Euler (1707–1783) se hizo famoso por probar en 1735 que $S = \pi^2/6$. *Numéricamente* podemos obtener el valor de $S$ con tanta precisión como deseemos (pero nunca exactamente) sumando los $n$ primeros términos

$$S_n = 1 + \frac{1}{4} + \frac{1}{9} + \frac{1}{16} + \cdots + \frac{1}{n^2}. \qquad \square$$

[2]Como refencia: en un ordenador personal, el muy sencillo bucle del Ejemplo 1.5.5 con diez millones de términos tarda en ejecutarse decenas de segundos. Así, cada término de ese ejemplo requiere millonésimas —no billónesimas— de segundo.

## 1.3 Perspectiva histórica

A pesar del interés enorme en la resolución numérica de problemas, el análisis numérico es, como disciplina autónoma, muy joven: ha nacido en la decada 1940–1950. ¿Por qué? Hasta entonces efectuar cálculos era muy costoso en tiempo —y por tanto en dinero. Ramas científicas completas, como la mecánica de fluidos, se desarrollaron lentamente por la imposibilidad de resolver matemáticamente las ecuaciones necesarias.

Justamente la necesidad de llevar a cabo cálculos intensivamente fue lo que impulsó a concebir y construir los primeros ordenadores en la decada 1940–1950. Abierta la posibilidad de calcular a precio razonable, los métodos numéricos han florecido desde entonces. Los avances en la tecnología de los ordenadores —que, claro es, hoy se aplican por doquier y no sólo en las tareas de cálculo numérico para las que se inventaron— permiten la introducción de nuevos métodos numéricos. A su vez, las necesidades de cálculo cada vez mayores llevan construir ordenadores cada día más potentes.

## 1.4 Errores

Por desgracia, al usar métodos numéricos es casi imposible obtener exactamente la respuesta del problema: por diversas razones —una de ellas la precisión finita de los ordenadores, que comentaremos más abajo— el resultado es sólo aproximado.

### 1.4.1 Error absoluto y error relativo

Por sencillez supongamos que el resultado buscado sea un sólo número real $V$—en ocasiones será una matriz, un vector, etc. La diferencia $E = N - V$ entre la solución numérica calculada y la verdadera se llama *error*, o mejor *error absoluto*. En algunas aplicaciones el signo de $E$ es importante. Si $E > 0$, es decir $N > V$, el error se llama *por exceso*. Si $E < 0$, ó $N < V$, el error es *por defecto*. Un error por exceso al calcular la dosis de un medicamento puede ser mortal, mientras que uno por defecto en las mismas circunstancias quizá sea poco transcendente, o

viceversa. No obstante, en la mayoría de las aplicaciones el signo de $E$ es de poca importancia y conviene considerar tan sólo el error sin su signo, es decir $|E|$.

Por razones fáciles de comprender, el mero tamaño de $|E|$ suele ser poco relevante y lo que importa es su relación con $V$. Cuando vaya a pagar ¿que pasará si me faltan 5 euros del precio? Si estoy comprando un coche me los perdonarán, no si estoy comprando un libro de bolsillo. Si $V \neq 0$, el cociente $|E|/|V|$ se llama *error relativo*; multiplicando por 100 obtenemos el error *porcentual*. En la vida diaria errores del 10 % son a menudo despreciables, por ejemplo no es difícil conseguir descuentos del 10 %. En muchas aplicaciones científicas hay que trabajar con mucha más precisión.

### 1.4.2 Cotas y estimaciones del error

Es imposible conocer en la práctica *exactamente* el valor de $E$ correspondiente a una aproximación $N$: conociendo $E$ conocemos la solución verdadera $V = N - E$ y entonces ¿para qué queremos $N$?[3] Deberemos contentarnos con acotar o estimar $E$, sin conocerlo.

Una *cota del error* es un número positivo $C$ para el que se pueda probar (rigurosamente en sentido matemático) que $|E| < C$. Lamentablemente, en numerosas situaciones las cotas de que se puede disponer son grandes sobreestimaciones del verdadero error (es decir $C$ es mucho mayor que $|E|$). En tal caso, la obtención de la cota puede ser un interesante ejercicio matemático ... sin ningún valor práctico.

Muchos métodos numéricos proporcionan junto con el valor de $N$ una *estimación* $\mathcal{E}$ de $E$; se espera que $\mathcal{E}$ coincida aproximadamente con $E$, pero tal cosa pudiera no ser cierta.

**Ejemplo 1.4.1** En el Ejemplo 1.2.1 la diferencia $S - S_n$ entre el verdadero valor $S$ de la serie y la aproximación $S_n$ es el llamado *resto* de la serie

$$R_n = \frac{1}{(n+1)^2} + \frac{1}{(n+2)^2} + \cdots$$

Como $S_n < S$ el error es por defecto. Además puede demostrarse (vea el Problema 1.4.1) que

$$\frac{1}{n+1} < S - S_n < \frac{1}{n},$$

[3]Esta observación se refiere al uso real de los métodos numéricos. En el aprendizaje de los mismos es útil aplicarlos a situaciones en que la solución exacta es conocida; esto permite hacerse una idea de los errores que el método produce.

de modo que $1/n$ es una cota (superior) del error (absoluto) y $1/(n+1)$ es una cota inferior de tal error. Sumar un millón de terminos permite conocer la suma de la serie con error inferior a una millónesima. □

**Problema 1.4.1** Para $0 \leq x < \infty$, considere la función $f(x)$ que toma el valor $1/(k+1)^2$ en el intervalo $k \leq x < (k+1)$, $k = 0, 1, \ldots$ Esboce su gráfica y observe que $\int_{1/n}^{\infty} f(x)\,dx$ es el resto $R_n$ del Ejemplo 1.4.1. En los mismos ejes de coordenadas, dibuje las gráficas de la funciones $m(x) = 1/(x+1)^2$ y $M(x) = 1/x^2$. Concluya que $m(x) \leq f(x) \leq M(x)$. Obtenga las cotas superior e inferior del Ejemplo 1.4.1. Obtenga cotas para los restos de las series $\sum(1/k^{(1+\alpha)})$ con $\alpha > 0$.

## 1.5 Errores resultantes de la representación en el ordenador de los números reales

### 1.5.1 Punto flotante

La forma más sencilla de manejar un número real es por su desarrollo decimal, p. ej. $3/8 = 0.375$, $-11/5 = -2.2$, o, si el desarrollo es infinito, por una truncación adecuada, p. ej. $2/3 = 0.66$ ó $2/3 = 0.666666$, $\pi = 3.141592$. (Usaremos siempre el punto y no la coma para separar las unidades de las décimas.) El formato decimal es sin embargo incómodo si el número es, en valor absoluto, muy grande o muy chico. En un mol hay una cantidad de moléculas igual al número de Avogadro (1776–1856), que aproximadamente es

$$602213670000000000000000, \tag{1.3}$$

la masa de un protón en gramos es aproximadamente

$$0.00000000000000000000000000016726231\ldots$$

Conviene entonces usar respectivamente las representaciones $6.0221367 \times 10^{23}$, $1.6726231 \times 10^{-27}$. Los ordenadores usan internamente siempre esta representación, que se llama de *punto flotante*,[4] en independencia de que al producir salidas por pantalla o impresora puedan transformar su representación interna a otros formatos.

[4]El corcho que flota en el puerto aparenta no moverse, pero sube y baja con la marea; el punto de la representación (normalizada) en punto flotante aparece tipográficamente siempre en el mismo lugar y sin embargo se mueve ...

### 1.5.2 Mantisa y exponente

Notemos que $60.221367 \times 10^{22}$ y $0.60221367 \times 10^{24}$ son también representaciones en punto flotante de (1.3). Para evitar tal multiplicidad de formas podemos *normalizar* para que, si el número no es cero, su primera cifra no nula aparezca inmediatamente a la izquierda del punto decimal. Con este convenio de normalización, debemos usar $6.0221367 \times 10^{23}$ y no $60.221367 \times 10^{22}$ ni $0.60221367 \times 10^{24}$. La parte 6.0221367 se llama *mantisa* del número y 23 es el *exponente*. El número cero es especial: para él la mantisa y el exponente no están definidos.

Para representar *exactamente* los números reales deberíamos usar mantisas de longitud infinita, cosa evidentemente imposible en la práctica. Al calcular con lápiz y papel la longitud (finita) de la mantisa puede variarse de unos números a otros, p. ej. podemos escribir $(2.0 \times 10^3)/(3.0 \times 10^1) = 6.666 \times 10^1$; datos con mantisa de dos cifras, cociente con mantisa de cuatro. En los ordenadores (salvo excepciones) la longitud de la mantisa está fijada y es la misma para todos los números.

**Ejemplo 1.5.1** Un sistema rudimentario que tuviese una mantisa de sólo dos cifras ¿cuántos números del intervalo $1 \leq x \leq 2$ puede representar? Respuesta: once, $1.0 \times 10^0$, $1.1 \times 10^0$, ..., $2.0 \times 10^0$; el continuo $1 \leq x \leq 2$ se reemplaza por un conjunto discreto. En ese sistema sólo podemos manejar $\sqrt{2}$ como $1.4 \times 10^0$, con un error aproximadamente del 1 %. Entre 1 y 10, la distancia entre números de punto flotante consecutivos es una décima; entre 10 y 100 los números van de unidad en unidad, ...

Los sistemas reales poseen mantisas más largas, pero el principio es el mismo: en el ordenador se substituyen intervalos continuos por conjuntos discretos. □

**Ejemplo 1.5.2** En el entorno de MATLAB[5] con cuya ayuda se ha escrito este libro, el número 'siguiente' al 1 es 1+eps, donde eps $= 2^{-52} \approx 2.2 \times 10^{-16}$. El hecho de que eps sea una potencia de dos es consecuencia de que los ordenadores usan aritmética binaria (de base dos) y no de base 10. Ejecutar el comando $(1 + 0.4 * \text{eps}) - 1$ da por resultado 0: el número de punto flotante más próximo al $1 + 0.4 * \text{eps}$ es el 1 ¿Cuál es el resultado de ejecutar el comando $(1 + 0.6 * \text{eps}) - 1$? □

### 1.5.3 La aritmética de punto flotante

Como consecuencia de la representación en punto flotante, la aritmética del ordenador es diferente de la 'verdadera'. Estudiemos algunos ejemplos.

[5]En MATLAB la mantisa es más larga que la que por defecto tienen otros sistemas (que por tanto son menos precisos).

**Ejemplo 1.5.3** *Substracción de números próximos.* Consideremos otra vez el sistema rudimentario de mantisa de dos cifras en base 10. En él la mejor representación posible del número real $15\pi/34 \approx 1.38$ es $1.4 \times 10^0$ (con error de aproximadamente 1 %). Si en ese ordenador calculamos $x^* - y^*$ donde $x^*$ es la versión del ordenador de $x = \sqrt{2}$ e $y^*$ es la versión del ordenador de $y = 15\pi/34$ obtendremos 0, cantidad errónea por completo como aproximación a $x - y \approx 0.0282$ (que, por otra parte, se puede representar en el ordenador por $2.8 \times 10^{-2}$ con error de aproximadamente el 1 %). *En general, al restar dos números de coma flotante $x^*$, $y^*$ aproximaciones a números reales $x$, $y$ muy próximos entre sí, se obtiene un número $x^* - y^*$ que aproxima mal a la verdadera diferencia $x - y$. Conviene pues formular siempre los algoritmos de modo que se evite substraer cantidades casi iguales.* □

**Ejemplo 1.5.4** *Falta de asociatividad de la aritmética de coma flotante.* Si las variables $x$, $y$, $z$ tienen valores $1 \times 10^{20}$, $-1 \times 10^{20}$, $z = 1$, en MATLAB los comandos $x + (y + z)$ y $(x + y) + z$ conducen a los valores 0, 1 respectivamente. □

**Ejemplo 1.5.5** Para hallar el valor de la suma $S_n$ del Ejemplo 1.2.1 hemos usado primero el bucle

```
suma = 0;
for k = 1:n
suma = suma + 1/(k*k);
end
```

y luego el bucle

```
sumainv = 0;
for k = n:-1,1
sumainv =1/(k*k)+ sumainv;
end
```

En el segundo los términos se incorporan a la suma de derecha a izquierda, esto es comenzando por los menores. Con números reales se tendría que **suma** coincidiría con **sumainv**. Con $n$ igual a cien millones, en el ordenador se obtiene **suma** $= 1.64493396684726 \neq$ **sumainv** $= 1.64493396684823$.[6]

Recordemos que para esta serie se conoce analíticamente el valor de la suma $\pi^2/6$, expresión que en MATLAB toma el valor 1.64493406684823. El resultado de **sumainv** difiere de él en exactamente una diezmillonésima, en armonía con la cota del Ejemplo 1.4.1. Sin embargo el valor de **suma** viola la cota; esto muestra que el resultado del primer bucle se ve afectado por errores que provienen de las operaciones aritméticas de coma flotante (además del que se debe a reemplazar $S$ por la suma parcial $S_n$, y del resultante de no ser $S$ exactamente representable en coma flotante). *En general ha de evitarse sumar reiteradamente sumandos de pequeña magnitud a una cantidad grande, siendo preferible sumar primero las cantidades pequeñas entre sí y luego agregarlas a la grande.* □

[6]Esto ilustra de nuevo la falta de asociatividad de la aritmética de coma flotante. Los bucles difieren sólo en la forma en que se agrupan los sumandos de $1 + 1/4 + \cdots + 1/n^2$.

A veces a los errores derivados de la mantisa finita que usa el ordenador se les llama, con denominación un tanto imprecisa, *errores de redondeo.*[7][8]

**Problema 1.5.1** *Resolución de la ecuación de segundo grado.* La ecuación de segundo grado $ax^2 + bx + c = 0$ se resuelve de ordinario por las fórmulas

$$x_1 = \frac{-b + \sqrt{b^2 - 4ac}}{2a}, \quad x_2 = \frac{-b - \sqrt{b^2 - 4ac}}{2a}. \tag{1.4}$$

Pruebe que si se define

$$y_1 = \frac{2c}{-b - \sqrt{b^2 - 4ac}}, \quad y_2 = \frac{2c}{-b + \sqrt{b^2 - 4ac}}, \tag{1.5}$$

entonces $x_1 = y_1$ y $x_2 = y_2$. Cuando $a = 1$, $b = 12345678987$, $c = 1$ evalúe en su ordenador $x_1$, $x_2$, $y_1$, $y_2$. Repita cambiando el signo de $b$. Observe cómo cada resultado que se ha obtenido mediante resta de cantidades próximas es claramente erróneo.

## 1.6 Costo operativo y eficiencia

Un mismo problema suele poder resolverse por diversos métodos numéricos. ¿Cómo decidirse por uno? El tamaño de los errores es ciertamente un criterio útil: podríamos elegir el método que diese menores errores. No obstante hay un criterio mejor: nos quedaremos con el método que, dando errores dentro de unos límites predeterminados, necesite el menor trabajo. Esta preocupación por estudiar la *eficiencia* de los diversos métodos es un rasgo distintivo del análisis numérico. La eficiencia depende tanto del tamaño de los errores como del *costo operativo* del método. Aquellas tareas para las que, en un momento histórico dado, no se dispone de un método aceptablemente eficiente no se pueden llevar a cabo.

Ilustraremos la idea de costo operativo con un algoritmo sumamente últil: el de Horner (1744–1834).[9]

---

[7]En rigor, el error de redondeo es el que se comete al 'redondear' un número no representable en el sistema de coma flotante utilizado para convertirlo en el número de coma flotante más próximo.

[8]No confunda la aritmética que usa el ordenador (sujeta a errores de redondeo) con la que emplearemos en el libro para enunciar y probar teoremas sobre los métodos numéricos. La segunda es la aritmética 'verdadera' (sin errores de redondeo).

[9]Horner propuso en 1819 este algoritmo, que en España se suele asociar al nombre de Ruffini (1765–1822) (quien lo usó en 1809). Hay ejemplos de su empleo en la Edad Media en China y Persia. Y en Europa era bien conocido de Newton (1643–1727).

**Ejemplo 1.6.1** *Método de Horner para evaluar polinomios.* Supongamos que deseamos hallar el valor de un polinomio cuártico $a_0 + a_1x + a_2x^2 + a_3x^3 + a_4x^4$ (los $a_i$ son números reales conocidos) para un valor dado $x_0$ de la variable $x$. El método más ingenuo halla sucesivamente los productos $x_0^2 = x_0 * x_0$, $x_0^3 = x_0^2 * x_0$, $x_0^4 = x_0^3 * x_0$ con un coste operativo de tres multiplicaciones, luego los productos $a_i * x_0^i$ (otras cuatro multiplicaciones), y finalmente hace las cuatro sumas/restas. (Que sean efectivamente sumas o restas depende del signo concreto que vayan teniendo los sumandos; para sumar 48 y $-27$ hay que restar.) En total cuatro sumas/restas y siete multiplicaciones. Sin embargo haciendo las operaciones en orden

$$a_0 + x_0 * \Big[a_1 + x_0 * \big[a_2 + x_0 * [a_3 + x_0 * a_4]\big]\Big] \tag{1.6}$$

sólo son precisas cuatro sumas/restas y cuatro multiplicaciones. Para un polinomio de grado $N$ el método ingenuo requiere $N$ sumas/restas y $2N - 1$ multiplicaciones, el método (1.6) llamado de *Horner* o de multiplicación encajada usa $N$ sumas/restas y $N$ multiplicaciones. El de Horner es pues más eficiente. Si sólo se va a evaluar una vez un polinomio y se usa un ordenador, no hay diferencia entre emplear uno u otro método. Si, como a veces ocurre, dentro de un programa complejo hay que evaluar millones de veces, usar uno u otro método puede significar tener que esperar el resultado dos horas o sólo una. Si calculamos con papel y lápiz en un examen, usar un mal método puede implicar no tener el resultado a tiempo. □

**Problema 1.6.1** *Definición formal del algoritmo de Horner.* El método o algoritmo de Horner para evaluar en $x_0$ el polinomio $P(x) = a_0 + a_1x + \cdots + a_Nx^N$ consiste formalmente en las siguientes operaciones

$$\begin{aligned} q_{N-1} &= a_N, \\ q_{N-i-1} &= q_{N-i}x_0 + a_{N-i}, \qquad i = 1, \ldots, N, \\ P(x_0) &= q_{-1}. \end{aligned} \tag{1.7}$$

Compruebe que cuando $N = 4$, este algoritmo lleva a cabo los cálculos indicados en (1.6)

**Problema 1.6.2** *Horner con lápiz y papel.* Para el cálculo con lápiz y papel el algoritmo (1.7) suele disponerse en la forma

| | | | | | | |
|---|---|---|---|---|---|---|
| | $a_N$ | $a_{N-1}$ | $a_{N-2}$ | $\cdots$ | $a_1$ | $a_0$ |
| $x_0$ | | $q_{N-1}x_0$ | $q_{N-2}x_0$ | $\cdots$ | $q_1x_0$ | $q_0x_0$ |
| | $q_{N-1}$ | $q_{N-2}$ | $q_{N-3}$ | $\cdots$ | $q_0$ | $q_{-1}$ |

Use este esquema para evaluar $Q(x) = -12 - x^2 - 4x^3 + x^4 + x^5$ en $x_0 = \pm1, \pm2, \pm3, \pm4, \pm6$ y $\pm12$. ¿Cuáles son las soluciones enteras de la ecuación $Q(x) = 0$?

**Problema 1.6.3** *Horner en la calculadora.* Algunas calculadoras de bolsillo baratas sólo pueden guardar en memoria un número ¿cómo llevaría a cabo en ellas el algoritmo de Horner?

**Problema 1.6.4** *Horner y división de polinomios.* Demuestre que en (1.6), $q_{-1}$ es el resto de la división de $P(x)$ entre $x - x_0$ y que los $q_i$, $i = 0, \ldots, N-1$ son los coeficientes de las potencias $x^i$ en el correspondiente (polinomio) cociente.

**Problema 1.6.5** *Cambio de origen en un polinomio.* Dado el polinomio $P(x) = a_0 + a_1 x + \cdots + a_N x^N$ (conocemos los $a_i$), deseamos expresarlo en potencias de $x - x_0$, es decir hallar los $b_i$ para que $P(x) = b_0 + b_1(x - x_0) + \cdots + b_N(x - x_0)^N$. Escribiendo

$$P(x) = b_0 + (x - x_0)\Big[b_1 + (x - x_0)\big[\cdots[b_{N-1} + b_N(x - x_0)]\cdots\big]\Big]$$

vemos que los $b_i$ son los restos que se obtienen al dividir sucesivamente entre $x - x_0$. Dé un esquema para calcular con lápiz y papel los $b_i$ añadiendo filas al esquema de Horner usual. Halle el costo operativo. Escriba un programa de ordenador que lleve a cabo la obtención de los $b_i$.

**Problema 1.6.6** Usando el esquema manual del problema precedente escriba $1+x+(1/2)x^2+(1/6)x^3$ en potencias de $x-1$. (La respuesta es $(8/3)+(5/2)(x-1)+(x-1)^2+(1/6)(x-1)^3$.)

**Problema 1.6.7** *Cambio de origen en un polinomio. Continuación.* ¿Cómo escribir en potencias de $x - x_0$ un polinomio cuyos coeficientes en potencias de $x - x_1$ se conocen? (Indicación: reduzca el problema al de expresar en potencias de $y - y_0$ un polinomio cuya expresión en potencias de $y$ es conocida.)

# Capítulo 2

# Interpolación de Taylor

## 2.1 Polinomios de Taylor, estudio algebraico

Consideremos la función[1]

$$f(x) = \int_0^x e^{-s^2}\, ds, \tag{2.1}$$

cuyos valores se necesitan al estudiar la distribución normal en cálculo de probabilidades (vea (1.1)). Hallar tales valores no es cosa inmediata. Sin embargo, sí es fácil encontrar los valores $f(0)$, $f'(0)$, $f''(0)$, ..., de la función y sus derivadas sucesivas evaluadas en $x = 0$. En efecto, por el teorema fundamental del cálculo,

$$f'(x) = e^{-x^2},$$

de donde

$$\begin{aligned} f''(x) &= -2xe^{-x^2}, \\ f'''(x) &= -2e^{-x^2} + 4x^2e^{-x^2}, \end{aligned}$$

y evaluando en $x = 0$

$$f(0) = 0, \quad f'(0) = 1, \quad f''(0) = 0, \quad f'''(0) = -2.$$

Evidentemente podríamos seguir calculando derivadas de orden superior, pero lo dejaremos en la tercera. Nos planteamos ahora encontrar un polinomio $P$ tal que

$$P(0) = 0, \quad P'(0) = 1, \quad P''(0) = 0, \quad P'''(0) = -2. \tag{2.2}$$

[1] Ya indicamos que esta función no es elemental: no se puede expresar combinando un número finito de funciones 'usuales': polinomios, trigonométricas, etc.

Presumiblemente, al $P$ *interpolar* a $f$ en el sentido de compartir con ella en $x = 0$ los valores de las derivadas de órdenes 0 a 3, no será una función muy diferente de la propia $f$ y quizá, en vez de evaluar $f$ podríamos evaluar $P$. [2] ¿Podremos hallar un polinomio que satisfaga las condiciones (2.2)?

Conviene plantearse la siguiente cuestión más general, cuya solución conduce a los llamados polinomios de Taylor (1685–1731) de una función, polinomios de uso constante en todas las matemáticas.

**El problema de interpolación de Taylor. Dados** un entero no negativo $N$, un punto $x_0$ de la recta y los valores $f(x_0), f'(x_0), \ldots, f^{(N)}(x_0)$ de una función y sus $N$ primeras derivadas en $x_0$, **encontrar** un polinomio de grado $\leq N$, tal que

$$P(x_0) = f(x_0), \quad P'(x_0) = f'(x_0), \quad \ldots, \quad P^{(N)}(x_0) = f^{(N)}(x_0). \tag{2.3}$$

(Naturalmente se requiere que la función $f$ tenga $N$ derivadas en el punto $x_0$.)

**Teorema 2.1.1** *El problema de interpolación de Taylor tiene solución única, que se llama polinomio de Taylor de grado $\leq N$ de la función $f$ en el punto $x_0$.*

*Primera demostración.* En la demostración más ingenua buscamos $P$ por *coeficientes indeterminados,* esto es escribimos

$$P(x) = a_0 + a_1 x + \cdots + a_N x^N \tag{2.4}$$

y tratamos de buscar los $N + 1$ coeficientes desconocidos $a_i$ para satisfacer las $N + 1$ condiciones en (2.3). Derivando obtenemos los polinomios $P'(x)$, $P''(x)$, $\ldots$, $P^{(N)}(x)$:

$$\begin{array}{ccccccccc}
a_0 & + & a_1 x & + & a_2 x^2 & + & \cdots & + & a_N x^N, \\
 & & a_1 & + & 2a_2 x & + & \cdots & + & N a_N x^{N-1}, \\
 & & & & 2a_2 & + & \cdots & + & N(N-1) a_N x^{N-2}, \\
 & & & & & & \ddots & & \vdots \\
 & & & & & & & & N(N-1)\cdots 1 a_N,
\end{array}$$

donde observamos que los términos en la diagonal son de la forma $i! a_i$. Evaluando las derivadas en $x_0$ e igualando a los datos del problema tenemos un sistema lineal *triangular* de $N + 1$ ecuaciones y otras tantas incógnitas:

$$\begin{array}{ccccccccc}
a_0 & + & a_1 x_0 & + & \cdots & + & a_N x_0^N & = & f(x_0), \\
 & & a_1 & + & \cdots & + & N a_N x_0^{N-1} & = & f'(x_0), \\
 & & & & \ddots & & \vdots & \vdots & \vdots \\
 & & & & & & N(N-1)\cdots 1 a_N & = & f^{(N)}(x_0).
\end{array}$$

[2] Los polinomios son muy sencillos de evaluar, vea la Sección 1.6.

La última ecuación determina $a_N = f^{(N)}(x_0)/N!$; llevando este valor a la penúltima ecuación hallamos $a_{N-1}$; llevando éste a la antepenúltima se tiene $a_{N-2}$ y así siguiendo este proceso, llamado de *sustitución regresiva,* determinamos todos los $a_i$ y con ello el polinomio $P$. □

*Segunda demostración.* El método anterior es laborioso. Puede encontrarse $P$ más fácilmente si en vez de buscarse en potencias de $x$ como en (2.4) se busca como una combinación lineal

$$P(x) = b_0 + b_1(x - x_0)^2 + b_2(x - x_0)/2! + \cdots + b_N(x - x_0)^N/N!$$

de los polinomios $(x - x_0)^i/i!$, $i = 0, 1, \ldots, N$. Derivando el nuevo formato de $P$, evaluando en $x_0$ e imponiendo (2.3) como en la demostración anterior, obtenemos un sistema de ecuaciones lineales para los $b_i$. La ventaja es que ahora el sistema es trivial

$$\begin{array}{lllll} b_0 & & & = & f(x_0), \\ & b_1 & & = & f'(x_0), \\ & & \ddots & \vdots & \vdots \\ & & b_N & = & f^{(N)}(x_0), \end{array}$$

con lo que

$$\begin{aligned} P(x) &= f(x_0) + f'(x_0)(x - x_0) \\ &\quad + f''(x_0)\frac{(x - x_0)^2}{2!} + \cdots + f^{(N)}(x_0)\frac{(x - x_0)^N}{N!}, \end{aligned} \tag{2.5}$$

*que es la forma usual de escribir el polinomio de Taylor.* □

Escrito en la forma (2.5), $P(x)$ tiene una *propiedad notable:* al suprimir el último sumando se obtiene el polinomio de Taylor de grado $\leq N - 1$, al suprimir los dos últimos el de grado $\leq N - 2$, etc.

**Ejemplo 2.1.1** Para (2.2) el polinomio de Taylor (de grado $\leq 3$) es $P(x) = x - (1/3)x^3$. La Figura 2.1 muestra la función $f$ (línea continua) y el polinomio $P$ (trazos); para $|x| \leq 0.5$ ambos son indistinguibles, para otros valores de $x$ se separan apreciablemente. ¿Cuál es el polinomio de Taylor de grado $\leq 1$ de nuestra función en $x_0 = 0$? ¿Qué significa geométricamente la recta cuya ecuación es este polinomio? □

**Problema 2.1.1** Para la función (2.1) escriba, usando la fórmula (2.5), los polinomios de Taylor en $x_0 = 0$ para $N = 0, 1, 2, 3, 4, 5, 6, 7$. ¿Son distintos dos a dos?

**Problema 2.1.2** *Polinomios de Taylor que conviene retener de memoria.* Use la fórmula (2.5) para hallar los polinomios de Taylor de grado $\leq N$ en $x_0 = 0$ de las funciones $e^x$, $\sin x$, $\cos x$, $1/(1 + x)$, $\log(1 + x)$.

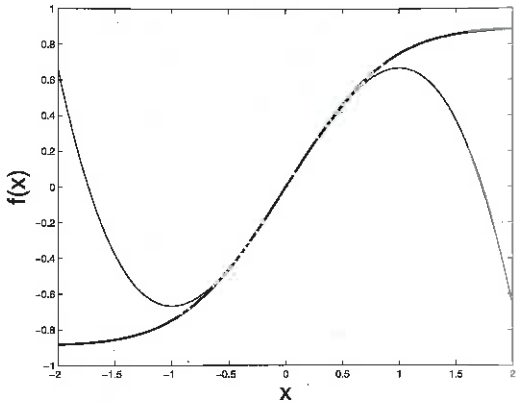


Figura 2.1: Función (línea gruesa) y polinomio de Taylor de grado $\leq 3$ para el Ejemplo 2.1.1

**Problema 2.1.3** *Obtención práctica de polinomios de Taylor.* Use (2.5) para hallar para $x_0 = 0$ y $N = 1, 2, \ldots, 6$ el polinomio de Taylor de $e^{-x^2}$. ¿Observa que esos polinomios son muy fáciles de escribir para quien sabe de memoria los polinomios de Taylor de $e^x$? Integre ahora en $[0, x]$ término a término los polinomios de Taylor de $e^{-x^2}$ y obtendrá los polinomios de Taylor del Problema 2.1.1. Compare el trabajo requerido en el presente problema con el que se necesitó en el 2.1.1. *En general el cálculo de las derivadas sucesivas de la función en el punto es el peor camino para encontrar un polinomio de Taylor. Es mejor ir combinando desarrollos ya conocidos partiendo de unos cuantos que se saben de memoria.*

**Problema 2.1.4** Escriba el polinomio de Taylor de grado dos de $e^x$ en $x_0 = 0$. Usando el Problema 1.6.6 pase ese polinomio a potencias de $x - 1$. Por otro lado halle el polinomio de Taylor de grado dos de $e^x$ en $x_0 = 1$. ¿Coinciden ambos resultados?

**Problema 2.1.5** **¡Importante!** Si en el problema de Taylor la función $f$ es ella misma un polinomio de grado $\leq N$, ¿cuál es su polinomio de Taylor de grado $\leq N$ en $x_0$? (Solución: el polinomio de partida. No se precisa hacer ningún cálculo si se utiliza la unicidad garantizada por el teorema de esta sección.)

Sin derivar ninguna función y sin usar el lápiz ¿cuánto valen las derivadas sucesivas de $f(x) = 1 - \pi x + \pi^2 x^2 - x^3$ evaluadas en el origen? (Respuesta: $f'(0) = -\pi$, $f''(0) = 2\pi^2$, $f'''(0) = -6$, y las restantes son nulas.)

**Problema 2.1.6** *Cálculo de las derivadas sucesivas de un polinomio en un punto.* Use el problema precedente para demostrar que si un polinomio se escribe en potencias de $x - x_0$, los coeficientes correspondientes son las derivadas sucesivas del polinomio evaluadas en $x_0$ divididas por los factoriales. Explique entonces cómo usar el algoritmo de Horner para evaluar en un punto $x_0$ un polinomio dado y sus derivadas hasta la $m$-ésima (use el Problema 1.6.5). Escriba un programa de ordenador que efectúe esa tarea. Cuente las operaciones requeridas.

**Problema 2.1.7** Si en el problema de Taylor el polinomio sujeto a las condiciones (2.3) se buscase de grado $\leq N+1$, la solución existiría pero ya no sería única. Encuentre todos los polinomios de grado $\leq N+1$ que satisfagan (2.3) (solución: $P(x)+c(x-x_0)^{N+1}$, con $c$ constante arbitraria y $P(x)$ el polinomio de Taylor de grado $N$). Encuentre todos los polinomios que satisfagan (2.3) (solución: $P(x)+(x-x_0)^{N+1}Q(x)$, con $Q(x)$ polinomio arbitrario).

**Problema 2.1.8** *Paridad del polinomio de Taylor.* Una función $f$ real de variable real es par (respectivamente impar) si está definida en un intervalo simétrico $(-a,a)$ y $f(x)\equiv f(-x)$ (resp. $f(x)\equiv -f(-x)$). Demuestre que un polinomio es una función par (resp. impar) si y sólo si al escribirlo en potencias de $x$ sólo contiene potencias pares (resp. impares). ¿Qué se puede decir de la derivada de una función par (resp. impar)? (Antes de contestar haga una gráfica.) Demuestre que el polinomio de Taylor en $x_0=0$ de grado $\leq N$ de una función par (resp. impar) es par (resp. impar). Estudie a continuación el caso $x_0\neq 0$.

## 2.2 Notaciones o y O de Landau

En la Figura 2.1 hemos observado cómo el error $f(x)-P(x)$ que se comete al tomar el valor $P(x)$ del polinomio de Taylor como aproximación al valor $f(x)$ tiende hacia 0 rápidamente cuando $x\to x_0=0$. Antes de medir la velocidad de esa convergencia, presentaremos las notaciones de Landau (1877–1938), que vamos a emplear. Comencemos reflexionando un poco.

**Ejemplo 2.2.1** Pensemos en un cubo de aristas de longitud $x>0$. Cuando $x$ decrece, el cubo tiene cada vez menos volumen y masa *relativamente* a su superficie: por eso el viento arrastra granos de arena y, salvo huracanes, no arrastra adoquines. Para $x>0$ chica, la función volumen $V(x)=x^3$ es pequeña en términos *absolutos* (los granos de arena tienen poco volumen) pero no es esa pequeñez la que es relevante aquí (los granos de arena también tienen poca superficie y el viento hace poca fuerza sobre cada uno). Aquí lo decisivo es que el volumen es pequeño *comparado* con la superficie $S(x)=6x^2$.

¿Cómo traducir esto a matemáticas? La pequeñez *absoluta* de la superficie y el volumen de un cubo de arista pequeña se corresponde con las afirmaciones "$S(x)\to 0$ para $x\to 0$" y "$V(x)\to 0$ para $x\to 0$." Para subrayar: si usted me da un valor de la superficie $\epsilon$ que considere pequeño, le encontraré un tamaño de arista $\delta$ por debajo del cual los cubos tienen superficie inferior a $\epsilon$, y análogamente para el volumen. La pequeñez *relativa* del volumen respecto a la superficie se corresponde con la afirmación "$V(x)/S(x)=x/6\to 0$ cuando $x\to 0$." Ahora lo que se afirma es que, elegido $\epsilon$ existe $\delta$ tal que si $0<x<\delta$ entonces $|V(x)|/|S(x)|<\epsilon$ ó $|V(x)|<\epsilon|S(x)|$. □

**Ejemplo 2.2.2** El economista Malthus (1766–1834) pensaba que en el futuro el progreso de la humanidad se detendría porque la cantidad de alimentos $A$ crece linealmente como función del tiempo $t$ y la población $P$ crece exponencialmente. Según él, aunque *absolutamente* cada vez habrá más alimentos, cada vez habrá más hambre porque el cociente $A/P$ (alimentos para cada habitante) tiende hacia cero. □

**Ejemplo 2.2.3** *La aproximación de Stirling (1692–1770).* Frecuentemente se aproxima el valor de $n!$ por el de $s_n = n^n \exp(-n)\sqrt{2\pi n}$. Para $n = 6$, es $6! = 720$ y la aproximación vale 710.0... con error menor que 10. Duplicando el valor de $n$, para $n = 12$, el error supera los tres millones. ¿Cómo puede ser la aproximación útil? En realidad, $12! = 479\,001\,600$ y la aproximación es 475 687 486.4..., con un error *relativo* inferior al 7 por mil; para $n = 6$ el error relativo es del 14 por mil. Puede demostrarse que el error absoluto $n! - s_n \to \infty$, pero el cociente (error relativo) $[n! - s_n]/n!$ tiende a 0, es decir el error crece más despacio que el factorial.[3] □

*Escribimos $f = o(g)$, $x \to x_0$ y leemos "$f$ es o pequeña de $g$ cuando $x$ tiende a $x_0$," si para $x$ próximo a $x_0$, $x \neq x_0$, $f$ y $g$ están definidas, $g$ no se anula y* $\lim_{x\to x_0} f(x)/g(x) = 0$. ("Para $x$ próximo a $x_0$" es una manera breve de decir "existe $r > 0$ tal que para cada $x$ con $|x - x_0| < r$".) La afirmación $f = o(g)$ traduce matemáticamente la idea de que al acercar $x$ a $x_0$, $f(x)$ se vuelve despreciable frente a $g(x)$. (No se afirma nada sobre el tamaño absoluto de $f(x)$.) La notación de Landau también se usan cuando $x_0$ es $\pm\infty$ y para sucesiones. Si $f_1 - f_2 = o(g)$ cuando $x \to x_0$, también se escribe $f_1 = f_2 + o(g)$, etc.

**Ejemplo 2.2.4** En los ejemplos anteriores, $V = o(S)$ cuando $x \to 0$, y $A = o(P)$ cuando $t \to \infty$. En el primer ejemplo $V$ es (en términos absolutos) pequeña, en el segundo $A$ es (en términos absolutos) grande. En la aproximación de Stirling el error $n! - s_n$ es $o(n!)$ cuando $n \to \infty$. □

Además de la notación o pequeña, aparece frecuentemente otra llamada "O grande". *Escribimos $f = O(g)$, $x \to x_0$ y leemos "$f$ es O grande de $g$ cuando $x$ tiende a $x_0$," si existe una constante $K > 0$ tal que para $x$ próximo a $x_0$, $f$ y $g$ están definidas y satisfacen $|f(x)| \leq K|g(x)|$.*

**Problema 2.2.1** De las siguientes afirmaciones ¿cuáles son ciertas cuando $x \to 0$? (i) $x^2 = o(x^3)$, (ii) $x^2 = O(x^3)$, (iii) $x^3 = o(x^2)$, (iv) $x^3 = O(x^2)$, (v) $\sin x = o(1)$, (vi) $e^x = o(1)$, (vii) $\sin x = O(1)$, (viii) $\sin x = x + o(x)$, (ix) $\sqrt{|x|} = o(x)$, (x) $\sin x = o(x\sqrt{|x|})$. Repita el ejercicio suponiendo ahora que $x \to \infty$.

**Problema 2.2.2** ¿Dónde falla el siguiente razonamiento? Cuando $x \to 0$, se tiene $x = o(1)$, $x^2 = o(1)$, luego $x = x^2$.

**Problema 2.2.3** Cuando $x \to 0$ ¿es $1000x^2$ o pequeña (despreciable) frente a $x$? ¿Cuánto valen las funciones $1000x^2$ y $x$ en $x = 0.1$? ¿Ve alguna contradicción en todo esto?

---

[3] En otras palabras $s_n/n! \to 1$ ó, como se suele decir, $n!$ y la aproximación de Stirling son (infinitos) *equivalentes:* $n! \sim s_n$. Para el error relativo se demuestra que $[n! - s_n]/n! \sim 1/(12n)$ (para $n = 12$, $1/(12n) \approx 0.007$).

## 2.3 Polinomios de Taylor, caracterización analítica

Las funciones

$$1, \quad (x-x_0), \quad \frac{(x-x_0)^2}{2!}, \quad \dots \quad , \quad \frac{(x-x_0)^N}{N!}, \tag{2.6}$$

que sirven de base para escribir el polinomio de Taylor (2.5) tienen la propiedad de que, cuando $x \to x_0$, cada una es o pequeña frente a cada una de las que la preceden. Por tanto, cuando $x$ se acerca a $x_0$ *cada sumando de (2.5) es una corrección pequeña de los sumandos anteriores.* Esta pequeñez es en el sentido del límite $x \to x_0$: para un valor concreto de $x$ muy bien puede ser cada sumando mayor (en módulo) que todos los precedentes (Problema 2.2.3).

Veremos seguidamente que el error de interpolación de Taylor $f(x) - P(x)$ resulta a su vez despreciable frente a la más despreciable función de la escala (2.6).

**Teorema 2.3.1** *Para $N \geq 1$, sea $f$ una función para la que el problema de Taylor se pueda proponer, esto es, una función $N$ veces derivable en $x_0$. El polinomio de Taylor $P$ verifica*

$$f - P = o((x-x_0)^N), \qquad x \to x_0. \tag{2.7}$$

*Además, $P$ es el único polinomio de grado $\leq N$ con esta propiedad, y por ello (2.7) caracteriza al polinomio de Taylor entre todos los de grado $\leq N$.*

*Demostración.* Tenemos que probar que

$$\lim_{x \to x_0} \frac{f(x) - P(x)}{(x-x_0)^N}$$

es nulo. Nos encontramos ante una indeterminación $0/0$. Si $N \geq 2$, aplicando la regla de l'Hôpital (1661–1704), hemos de estudiar el nuevo límite

$$\lim_{x \to x_0} \frac{f'(x) - P'(x)}{N(x-x_0)^{N-1}},$$

que es también indeterminado. Continuando las aplicaciones de la regla hasta un total de $N-1$ (cosa posible pues $f$ es $N-1$ veces derivable con continuidad en el entorno de $x_0$) se llega al límite

$$\lim_{x \to x_0} \frac{f^{(N-1)}(x) - P^{(N-1)}(x)}{N!(x-x_0)},$$

| $x$ | $f$ | $P_1$ | $f-P_1$ | $S_1$ | $f-S_1$ |
|---|---|---|---|---|---|
| 0.1 | 1.136 788 | 1.100 000 | 0.036 788 | 1.050 000 | 0.086 788 |
| 0.01 | 1.011 050 | 1.010 000 | 0.001 050 | 1.005 000 | 0.006 050 |
| 0.001 | 1.001 000 | 1.001 000 | 0.000 032 | 1.000 500 | 0.000 532 |
| 0.000 1 | 1.000 101 | 1.000 100 | 0.000 001 | 1.000 050 | 0.000 051 |

Tabla 2.1: Errores al aproximar $f(x) = \exp(x) + \sin(x\sqrt{|x|})$ por su recta de Taylor en $x_0 = 0$ y por la recta $y = S_1(x) = 1 + 0.5x$

o, cálculando la derivada $N-1$ del polinomio,

$$\lim_{x\to x_0} \frac{f^{(N-1)}(x) - [f^{(N-1)}(x_0) + f^{(N)}(x_0)(x-x_0)]}{N!(x-x_0)} = \frac{1}{N!}\left[\lim_{x\to x_0} \frac{f^{(N-1)}(x) - f^{(N-1)}(x_0)}{x-x_0} - f^{(N)}(x_0)\right].$$

Este es 0 por definición de $f^{(N)}(x_0)$.

Probemos ahora la unicidad, razonando por reducción al absurdo. Si hubiese dos polinomios distintos, digamos $Q$ y $R$, de grado $\leq N$ que satisficiesen (2.7), entonces, restando $f - R = o((x-x_0)^N)$ de $f - Q = o((x-x_0)^N)$, concluimos que $R - Q = o((x-x_0)^N)$. Pero, según el Problema 2.3.1, entonces $R - Q \equiv 0$ y tenemos una contradicción. □

**Ejemplo 2.3.1** Según el teorema, el polinomio de Taylor de grado $\leq 1$, $P_1(x) = f(x_0) + f'(x_0)(x-x_0)$ (cuya gráfica es la recta tangente en $(x_0, f(x_0))$ a la gráfica de $y = f(x)$) es el único de grado $\leq 1$ para el que $[f(x) - P_1(x)]/(x-x_0) \to 0$ cuando $x_0 \to 0$.[4] Para la función $f(x) = \exp(x) + \sin(x\sqrt{|x|})$ y $x_0 = 0$, es $P_1(x) = 1 + x$. En la Tabla 2.1 compare los tamaños de $f(x) - P_1(x)$ y de $x$ al irse acercando $x$ a 0.

Hemos considerado también el polinomio $S_1(x) = 1 + 0.5x$. El error $f(x) - S_1(x)$ tiende a 0 cuando $x \to 0$, pero no es $o(x)$ (de hecho, $[f(x) - S_1(x)]/x$ tiende a $1/2$ —demuéstrelo— y no a 0). Para $x = 0.1$ el error en $P_1(x)$ es aproximadamente la mitad del error en $S_1(x)$; para $x = 0.0001$ es sólo la quincuagésima parte. □

[4]Esta afirmación no es sino la definición de derivada, ligeramente enmascarada. Para que $f(x) - [A + B(x-x_0)]$ sea $o(x-x_0)$ es claro que hay que tomar $A = f(x_0)$ y entonces imponer que

$$\lim \frac{f(x) - [f(x_0) + B(x-x_0)]}{x-x_0} = 0$$

conduce a que $B$ ha de ser el límite de $[f(x) - f(x_0)]/(x-x_0)$.

**Ejemplo 2.3.2** Para el polinomio de Taylor de grado $\leq 3$ del Ejemplo 2.1.1, será $f(x) - P(x) = o(x^3)$ cuando $x \to 0$. Todavía más, $x - (1/3)x^3$ es también el polinomio de Taylor de grado $\leq 4$ y por ello el error será incluso $o(x^4)$.[5] Para $x = 0.1$ el error es $\approx 9.97 \times 10^{-7}$; para $x = 0.01$, es[6] $\approx 1.00 \times 10^{-11}$. □

**Ejemplo 2.3.3** En la función $f(x) = 1 - x + \pi x^2 + x^3\sqrt{1+x^2}$, es obvio que $x^3\sqrt{1+x^2} = o(x^2)$ cuando $x \to 0$. Del teorema se desprende, sin hacer cálculo alguno, que el polinomio de Taylor de grado $\leq 2$ relativo a $x_0 = 0$ es $1 - x + \pi x^2$. Sin necesidad de derivar, se concluye que $f'(0) = -1$ y $f''(0) = 2\pi$. □

El análisis matemático emplea el teorema anterior para determinar las propiedades de una función $f$ cerca de un punto $x_0$ a través del estudio de las propiedades de uno de sus polinomios de Taylor en $x_0$ de grado adecuado.

**Ejemplo 2.3.4** Consideremos una función con $f'(x_0) = 0$, $f''(x_0) > 0$. Entonces el polinomio de Taylor $P_2(x) = f(x_0) + f''(x_0)(x - x_0)^2/2!$ tiene un mínimo en $x = x_0$, pues para $x \neq x_0$ el segundo sumando es $> 0$, luego $P_2(x) > P_2(x_0)$. ¿Qué le ocurre a la función? Escribamos

$$f(x) - f(x_0) = f''(x_0)\frac{(x-x_0)^2}{2!} + [f(x) - P_2(x)];$$

en el segundo miembro, según el teorema, el corchete es despreciable frente al primer sumando cuando $x \to x_0$. Para $x$ suficientemente próximo a $x_0$ el corchete es, en valor absoluto, menor que el 10 % del primer sumando, de modo que, aunque el corchete sea negativo, el segundo miembro es positivo. En resumen $f(x) - f(x_0) > 0$ para $x$ próximo a $x_0$ y decimos que $f$ tiene un mínimo relativo en $x_0$. Observe que en el razonamiento no basta con que el corchete sea pequeño; debemos saber que es más pequeño que el primer sumando, y éste tiende a 0 cuando $x \to x_0$. □

**Problema 2.3.1** Pruebe que si un polinomio $P$ tiene grado $\leq N$ ($N$ es un entero no negativo) y es tal que $P = o((x - x_0)^N)$, $x \to x_0$, entonces $P$ es idénticamente nulo . (Indicación: Escriba $P$ en potencias de $x - x_0$ y vaya probando que todos los coeficientes son nulos, comenzando por el término independiente.)

**Problema 2.3.2** Para la función $f$ en (2.1) y cuando $x \to 0$ encuentre todos los polinomios $Q$ de grado $\leq 3$ tales $f - Q = o(1)$. Idem $f - Q = o(x)$. Idem $f - Q = o(x^2)$.

---

[5] Usando el teorema de la sección siguiente se tiene $f(x) - P(x) = (1/120)f^{(5)}(\xi_x)x^5$ con $|\xi_x| < |x|$. Cuando $x$ tiende hacia 0 lo hace $\xi_x$ y por ello $f(x) - P(x) = (1/120)f^{(5)}(0)x^5 + o(x^6)$, resultado más fuerte que $f(x) - P(x) = o(x^4)$.

[6] A la vista de la nota precedente, esto sugiere que $f^{(5)}(0) = 12$. Derive (2.1) cinco veces y evalúe en 0 para comprobar la validez de esta conjetura.

## 2.4 Cotas del error

Los resultados de la sección precedente dan información sobre $f(x) - P(x)$ en el límite $x \to x_0$, pero *no permiten medir la diferencia* $f(x) - P(x)$ *en un valor de* $x$ *concreto* que hayamos fijado. Acotemos ahora este error[7] de interpolación.

**Teorema 2.4.1** Error en el polinomio de Taylor; forma de Lagrange (1736–1813). *Sean* $x$, $x_0$ *dos números reales distintos y sea* $f$ *una función con* $N$ *derivadas continuas en el intervalo cerrado de extremos* $x$, $x_0$. *Supongamos, además, que la derivada* $f^{(N+1)}$ *existe en el intervalo abierto de esos mismos extremos. Entonces, hay un punto* $\xi$, *comprendido estrictamente entre* $x$ *y* $x_0$ *tal que*

$$f(x) - P(x) = f^{(N+1)}(\xi)\frac{(x - x_0)^{N+1}}{(N+1)!}. \tag{2.8}$$

*Demostración.* En toda la prueba es importante que piense que $x$ es un número fijado. Definamos $M = [f(x) - P(x)]/(x - x_0)^{N+1}$, con lo que $f(x) - P(x) = M(x - x_0)^{N+1}$ y hay que probar que existe $\xi$ en el intervalo anunciado de suerte que $f^{(N+1)}(\xi) = M(N+1)!$. Se forma la función auxiliar

$$F(t) = f(t) - P(t) - M(t - x_0)^{N+1},$$

en la que hemos llamado $t$ a la variable independiente porque el símbolo $x$ ya se está utilizando para otro fin. Como

$$F^{(N+1)}(t) = f^{(N+1)}(t) - M(N+1)!$$

basta probar que $F^{(N+1)}(t)$ se anula en algún punto del intervalo anunciado.

De la definición de $M$ se sigue que la función $F$ se anula cuando a su variable $t$ se le da el valor numérico $x$. Por otro lado se calcula inmediatamente que en $t = x_0$ se anulan $F, F', \ldots, F^{(N)}$. Por el teorema de Rolle (1652–1719), $F(x) = F(x_0)$ implica que $F'$ se anula en un punto $\eta_1$ entre $x$ y $x_0$. Del mismo modo, $F'(\eta_1) = F'(x_0)$ implica que $F''$ se anula en un punto $\eta_2$ entre $\eta_1$ y $x_0$. Iterando el argumento se llega al cero $\xi$ de $F^{(N+1)}$. □

Cuando la función $f$ es ella misma un polinomio de grado $\leq N$, el polinomio de Taylor $P(x)$ coincide con ella (Problema 2.1.5) y el error $f(x) - P(x)$ es idénticamente nulo; el segundo miembro de (2.8) es idénticamente nulo por serlo la función derivada $N+1$ de un polinomio de grado $\leq N$.

[7] A veces a la diferencia $f(x) - P(x)$ se le llama *resto* en vez de error y se habla de forma de Lagrange del resto, etc.

**Corolario 2.4.1** *Además de las hipótesis del teorema, supongamos que, para cada $t$ en el intervalo abierto de extremos $x$, $x_0$, se verifica que $|f^{(N+1)}(t)| \leq K_{N+1}$. Entonces*

$$|f(x) - P(x)| \leq \frac{|x - x_0|^{N+1} K_{N+1}}{(N+1)!}. \tag{2.9}$$

**Ejemplo 2.4.1** Para el problema de Taylor en (2.2), $N = 3$, tenemos $f^{(4)}(x) = (12 - 8x^2)xe^{-x^2}$. Supongamos que $-1/2 \leq x \leq 1/2$, entonces $|12 - 8x^2| \leq 12$ y podemos tomar $K_4 = 12 \times (1/2) \times 1 = 6$. De acuerdo con (2.9), $x - (1/3)x^3$ aproxima a la función con error menor que $(1/16) \times 6/24 = 1/64 \leq 0.02$, si $-1/2 \leq x \leq 1/2$. Vea también el Problema 2.4.2. □

Además de la forma de Lagrange, hay otras formas útiles de escribir la diferencia $f(x) - P(x)$; la siguiente es la más relevante. Note que se requieren más hipótesis sobre $f$ que en la forma de Lagrange y que, a su vez, ésta requiere más hipótesis que las que se necesitan para poder construir $P$.

**Teorema 2.4.2** Error en el polinomio de Taylor; forma integral. *Sean $x$, $x_0$ dos números reales distintos y sea $f$ una función con $N+1$ derivadas continuas en el intervalo cerrado de extremos $x$, $x_0$. Entonces,*

$$f(x) - P(x) = \frac{1}{N!} \int_{x_0}^{x} (x - s)^N f^{(N+1)}(s)\, ds. \tag{2.10}$$

*Demostración.* Vea el Problema 2.4.3. □

**Problema 2.4.1** Se va a diseñar una calculadora de bolsillo. Para que evalúe la función $\sin x$ se la reemplaza por un polinomio, y se desea que para cada $x$ real el error sea menor que $5 \times 10^{-7}$. Explique por qué se puede suponer que $0 \leq x \leq \pi/2$. Si se usa un polinomio de Taylor en $x_0 = \pi/4$ ¿qué valor de $N$ hay que tomar?

**Problema 2.4.2** En el Ejemplo 2.4.1 mejore la cota para $f(x) - [x - (1/3)x^3]$ teniendo en cuenta que $x - (1/3)x^3$ es también el polinomio de Taylor de grado $\leq 4$ en $x_0 = 0$, de modo que es posible tomar $N = 4$ en (2.9).

**Problema 2.4.3** Demuestre (2.10) integrando por partes reiteradamente.

**Problema 2.4.4** *Cálculo aproximado de raíces cuadradas.* Sean $a$, $b$ números reales con $a > 0$ y $a + b > 0$. La fórmula

$$\sqrt{a+b} \approx \sqrt{a} + \frac{b}{2\sqrt{a}}$$

ya se usó en Babilonia hace casi 4000 años. Por ejemplo $\sqrt{31} = \sqrt{25+6} \approx 5 + 6/10 = 5.6$ (el verdadero valor es 5.567764). Pruebe que esta fórmula equivale a usar el polinomio de Taylor de grado 1 de la función $\sqrt{x}$ en $x_0 = a$ y evaluarlo en $x = a + b$. Demuestre que la aproximación dada por esta fórmula es siempre por exceso. Dé una cota del error absoluto que se comete al usar la fórmula.

**Problema 2.4.5** *Raíces cuadradas. Continuación.* Hace 4000 años no se conocía el cálculo infinitesimal ¿cómo encontraron los babilonios la fórmula anterior? Si $b$ es pequeño $\sqrt{a+b} = \sqrt{a} + \delta$ donde $\delta$ es pequeño. Elevando al cuadrado $a + b = a + 2\sqrt{a}\delta + \delta^2$ y, despreciando $\delta^2$ frente a $\delta$, será $b \approx 2\sqrt{a}\delta$, y de aquí la fórmula. Utilice esta técnica para aproximar $\sqrt{a+b}$ por un polinomio de grado $\leq 2$ en $b$. *Históricamente los polinomios de Taylor de las funciones más usuales se hallaron por diversos artificios antes de que Taylor demostrara que los correspondientes coeficientes no son otra cosa que los valores de las derivadas sucesivas escalados con los factoriales.*

**Problema 2.4.6** *Raíces de ecuaciones.* Se considera la ecuación $1 - x + \lambda(x + x^2 - x^7) = 0$, donde $x$ es la incógnita buscada y $\lambda$ un parámetro conocido que toma valores pequeños. Si $\lambda = 0$, la única raíz es $x = 1$. Para $\lambda$ cercano a 0, se busca una solución $x = x(\lambda) = 1 + a\lambda + b\lambda^2 + \cdots$. Sustituya esta expresión en la ecuación y determine los valores de $a$ y $b$. De esta forma, ha hallado el polinomio de Taylor de grado $\leq 2$ de la función $x = x(\lambda)$, cuya expresión analítica desconocemos.[8] Evidentemente el procedimiento puede usarse para hallar polinomios de Taylor de grado tan alto como se desee.

## 2.5 Convergencia de la sucesión de los polinomios de Taylor

Hasta aquí manteníamos, para una función $f$ dada, fijos el entero $N$ y el punto $x_0$. Manteniendo $f$ y $x_0$, vamos ahora a ir considerando los polinomios de Taylor de grados $0, 1, \ldots, N, \ldots$, que denotaremos por $P_0(x), P_1(x), \ldots, P_N(x), \ldots$ Nos preguntamos si incrementando $N$ podemos acercar tanto como queramos $P_N(x)$ al valor de la función $f(x)$, es decir, si será cierto que

$$f(x) = \lim_{N\to\infty} P_N(x) = \lim_{N\to\infty} \sum_{n=0}^{N} f^{(n)}(x_0)\frac{(x-x_0)^n}{n!}.$$

En otras palabras nos preguntamos si $f(x)$ es la suma de la correspondiente *serie de Taylor*

$$f(x) = \sum_{n=0}^{\infty} f^{(n)}(x_0)\frac{(x-x_0)^n}{n!}.$$

Si la respuesta a esta cuestión fuese afirmativa, podríamos hallar los valores de $f(x)$ con tanta precisión como deseasemos sumando un número (finito) suficientemente grande de términos de la serie.

[8] La *existencia* de esta función $x = x(\lambda)$ y de sus derivadas de cualquier orden puede garantizarse por el teorema de las funciones implícitas.

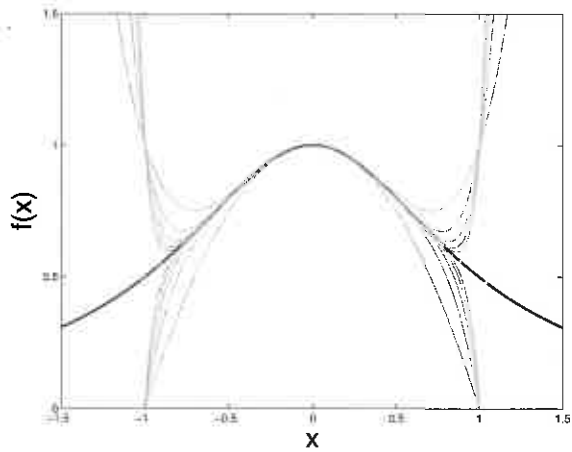


Figura 2.2: La función $f(x) = 1/(1+x^2)$ y sus polinomios de Taylor en $x_0 = 0$ hasta el grado 20, $P_2(x) = 1 - x^2$, $P_4(x) = 1 - x^2 + x^4, \ldots, P_{20}(x) = 1 - x^2 + \cdots + x^{20}$. En $-1 < x < 1$ los polinomios quedan alternativamente bajo y sobre la función, en cada $x$ fijo convergen hacia $f(x)$ —aunque no uniformemente: cuanto más próximo se halle $x$ de $\pm 1$ más hay que elevar el grado del polinomio para obtener una precisión dada. Para $|x| \geq 1$ la serie diverge. Aparentemente no hay nada en la gráfica de $f$ que permita explicar el cambio de comportamiento en $x = \pm 1$

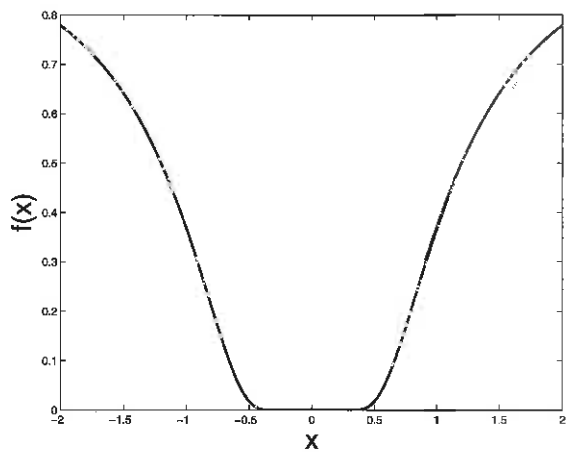


Figura 2.3: La función $\exp(-1/x^2)$. Para $x \neq 0$ es estrictamente positiva. En $x_0 = 0$ todas las derivadas son nulas y por ello el polinomio de Taylor de grado $\leq N$ es, para cada valor de $N$, idénticamente nulo. De acuerdo con la caracterización analítica de los polinomios de Taylor, la función, sea cual sea el valor de $N$, es $o(x^N)$ cuando $x \to 0$

- Notemos ante todo que para que la cuestión esté definida se necesita poder formar todos los $P_N$, lo que requiere que $f$ tenga derivadas de todos los órdenes en $x_0$.

- Aunque todas las derivadas existan, la serie quizá no converja para algunos (o todos) los valores de $x \neq x_0$ en que estemos interesados (Figura 2.2).[9]

- Aunque las derivadas existan y la serie converja, puede ser que la suma no coincida con $f(x)$. El ejemplo arquetípico lo da la función $\exp(-1/x^2)$ cuyas derivadas en $x_0 = 0$ son todas nulas (Figura 2.3). Por tanto la suma de la serie es la función nula y **no** coincide con $\exp(-1/x^2)$.

Al investigar la convergencia de la serie de Taylor, $x$ está fijo y $N \to \infty$. Como acabamos de ver, en este límite los polinomios de Taylor *pueden* no ser útiles. Al estudiar la caracterización analítica el proceso era otro: $x \to x_0$ con $N$ fijo. En ese límite los polinomios de Taylor *siempre* se comportan adecuadamente.

**Problema 2.5.1** Usando las cotas de error de la interpolación de Taylor demuestre que las funciones $\sin x$, $\cos x$, $e^x$ son, para cada $x$ real, la suma de sus correspondientes series de Taylor en $x_0 = 0$. En el intervalo $0 \leq x \leq a$ represente gráficamente, para cada una de las funciones, la propia función y los polinomios de Taylor en $x_0 = 0$ de grados uno a diez. Ensaye varios valores de $a$.

## 2.6 Cambio de escala

Partamos de una función $f(x)$ y de su polinomio de Taylor (de grado $\leq N$) en el origen ($x_0 = 0$)

$$p(x) = f(0) + f'(0)\frac{x}{1!} + \cdots + f^{(N)}(0)\frac{x^N}{N!}. \tag{2.11}$$

Si $\lambda \neq 0$ es una constante, hagamos el cambio de variable $x = \lambda X$ y definamos una nueva función $F$ por $F(X) = f(\lambda X)$. De este modo $F$ y $f$ toman el mismo valor cuando se evalúan en puntos $X$, $x$ que se corresponden por el cambio. Para las derivadas, los valores de $F'$ en $X$ y $f'$ en $x$ **no** coinciden, puesto que, en virtud de la regla de la cadena, $F'(X) = \lambda f'(\lambda X)$, $F''(X) = \lambda^2 f''(\lambda X)$, etc.

[9] Se puede demostrar que dada una sucesión cualquiera de números reales $(a_n)$ existe una función para la cual $f^{(n)}(x_0) = a_n$ para cada entero no negativo $n$. Eligiendo p. ej. $a_n = (n!)^2$ la serie no converge para ningún $x \neq x_0$.

**Ejemplo 2.6.1** Sea $f(x)$ la función que da la abscisa (medida en metros) de un móvil cuando lleva moviéndose $x$ segundos y tomemos $\lambda = 3600$. Entonces $X$ es el valor del tiempo transcurrido cuando la unidad de medida es la hora. Matemáticamente $F$ y $f$ son funciones distintas entre sí, porque evaluadas en *un mismo punto* (digamos en el punto 1) producen —en general— valores distintos ($F(1) \neq f(1)$). Sin embargo esas funciones son representaciones de un mismo movimiento: $F(1) = f(3600)$ porque la abscisa es la misma tras una hora o tras 3600 segundos. Las derivadas $F'$ y $f'$ describen *velocidades* medidas en metros por hora o en metros por segundo. Evaluar esas funciones derivadas en puntos que se corresponden, no conduce al mismo valor: la velocidad en metros por hora es 3600 mayor que la velocidad en metros por segundo. Para las derivadas segundas (aceleraciones) el factor de escala es $3600 \times 3600$. □

Usando (2.5), el polinomio de Taylor de $F$ en $X_0 = 0$ es

$$P(X) = F(0) + F'(0)\frac{X}{1!} + \cdots + F^{(N)}(0)\frac{X^N}{N!}.$$

Expresemos las derivadas de $F$ en términos de las de $f$, empleemos (2.11) y recordemos que $\lambda X = x$:

$$P(X) = f(0) + \lambda f'(0)\frac{X}{1!} + \cdots + \lambda^N f^{(N)}(0)\frac{X^N}{N!} = p(\lambda X) = p(x).$$

En palabras: *para el cambio lineal de variable $x = \lambda X$, se llega al mismo resultado (i) cambiando de variable en una función y luego hallando el polinomio de Taylor (en el origen $X_0 = 0$) o (ii) hallando primero el polinomio de Taylor (en el origen $x_0 = 0$) y luego cambiando de variable en él.*[10]

**Ejemplo 2.6.2** *Análisis dimensional.* Consideremos el término $f'(0)x/1!$ del polinomio de Taylor (2.11). En física, si $f$ y $x$ son dimensionales (p. ej. representan una longitud y un tiempo), entonces $f'$ es también dimensional y sus dimensiones son el cociente entre las dimensiones de $f$ y las de $x$ (esto es consecuencia de la definición de derivada como límite de cocientes de incrementos). Por tanto el producto $f'(0)x/1!$ tiene las dimensiones de $f$. Análogamente se ve que cada término $f^{(k)}(0)x^k/k!$ tiene las dimensiones de $f$. La expresión (2.11) es entonces *dimensionalmente correcta:* cada sumando del segundo miembro tiene las dimensiones de $f$.

Las dimensiones de $f^{(k)}(0)x^k/k!$ aparecen más claras usando la notación de Leibniz (1646–1716) para las derivadas: $f^{(k)} = d^k f/dx^k$.[11]

*El análisis dimensional es útil para recordar fórmulas y evitar equivocaciones.* Si en (2.11) se comete la equivación de escribir, p. ej. $f^{(k)}(0)x^{k+1}/k!$ la fórmula no queda dimensionalmente correcta — en términos matemáticos no se comporta adecuadamente frente a cambios de escala en las variables. □

[10]Esta conclusión podía esperarse: ¿Creeríamos que la calidad de una interpolación de Taylor puede incrementarse con sólo cambiar la unidad de medida?

[11]Observe que el exponente $k$ no afecta a $f$, mientras que $dx$ aparece elevado a la potencia $k$.

**Problema 2.6.1** Con la notación usada en la presente sección, dé una demostración alternativa de la relación $P(X) = p(x)$. Use la siguiente idea: $p(\lambda X)$ es un polinomio de grado $\leq N$ en $X$ que satisface las condiciones de interpolación que definen $P(X)$.[12]

**Problema 2.6.2** *Cambio lineal de variable independiente.* Repita la discusión de esta sección para el caso, más general, del cambio de variable $x = \lambda X + \mu$, $\lambda \neq 0$. Si parte de (2.11) ¿es $p(\lambda X + \mu)$ el polinomio de Taylor de $F(X) = f(\lambda X + \mu)$ en $X_0 = 0$?

**Problema 2.6.3** *Cambio lineal de variable dependiente.* En $y = f(x)$ efectuamos el cambio de variable dependiente $Y = Ay + B$ ($A \neq 0$). Estudie cómo obtener, a partir de los de $f(x)$, polinomios de Taylor de la función $Y = F(x) = Af(x) + B$ resultante del cambio .

**Problema 2.6.4** ¿Son dimensionalmente correctas las expresiones de Lagrange (2.8) e integral (2.10) para el error en el polinomio de Taylor?

[12]Para cambios generales de variables, $x = \phi(X)$, $\phi(0) = 0$, la función $p(\phi(X))$ no necesariamente es un polinomio en $X$ y por ende $p(\phi(X))$ no puede ser el polinomio de Taylor de $F(X) = f(\phi(X))$ en $X_0 = 0$.

# Capítulo 3

# Interpolación polinómica de Lagrange

## 3.1 Polinomio de Lagrange

### 3.1.1 Planteamiento

Ahora conocer el valor de funciones como $e^x$, $\sin x$, $\cos x$, etc. es cosa de apretar una tecla. Antes de los años 70 del pasado siglo no había calculadoras de bolsillo ni ordenadores personales, y la mejor forma de acceder a los valores de las funciones era mediante *tablas*. He aquí un fragmento de una tabla de la función seno, con el argumento medido en grados,

| $x$ | $\sin x$ |
|---|---|
| 60 | 0.866025 |
| 65 | 0.906308 |
| 70 | 0.939693 |
| 75 | 0.965926 |
| 80 | 0.984808 |

¿Como hallar el seno de 72? La idea básica es sustituir la función seno por un polinomio que tome los mismos valores que ella en los puntos de la columna izquierda y luego tomar el valor de $P(72)$ como aproximación al correspondiente seno. ¿Existirá tal polinomio?

Conviene plantearse la siguiente cuestión más general, que aparece constantemente en análisis numérico.

**El problema de interpolación de Lagrange. Dados** un entero no negativo $N$, $N+1$ puntos $x_0, \ldots, x_N$ de la recta, dos a dos distintos, y los correspondientes valores $f(x_0), \ldots, f(x_N)$ de una función, **encontrar** un polinomio de grado $\leq N$, tal que

$$P(x_0) = f(x_0), \quad P(x_1) = f(x_1), \quad \ldots \quad , \quad P(x_N) = f(x_N). \tag{3.1}$$

Los $x_i$ se llaman *abscisas* o *nodos* de la interpolación. No se supone que estén ordenadas, por ejemplo puede ser $x_0 = 1000$, $x_1 = -\pi$, $x_2 = 4e$.

**Teorema 3.1.1** *El problema de interpolación de Lagrange tiene solución única, que se llama polinomio interpolador de Lagrange de grado $\leq N$ de la función $f$ en los puntos $x_0, \ldots, x_N$.*

*Demostración.* En la demostración más ingenua buscamos $P$ por *coeficientes indeterminados,* es decir escribimos $P(x) = a_0 + a_1 x + \cdots + a_N x^N$ y tratamos de buscar los $N+1$ coeficientes desconocidos $a_i$ para satisfacer las $N+1$ condiciones en (3.1). Evaluando $P$ en los nodos e igualando a los datos del problema tenemos un sistema lineal de $N+1$ ecuaciones y otras tantas incógnitas

$$\begin{array}{ccccccccc}
a_0 & + & a_1 x_0 & + & \cdots & + & a_N x_0^N & = & f(x_0), \\
a_0 & + & a_1 x_1 & + & \cdots & + & a_N x_1^N & = & f(x_1), \\
\cdots & & \cdots & & \cdots & & \cdots & = & \cdots \\
a_0 & + & a_1 x_N & + & \cdots & + & a_N x_N^N & = & f(x_N).
\end{array}$$

La matriz de este sistema es de tipo Vandermonde (1735–1796) con nodos dos a dos distintos y por tanto no singular (Problema 3.1.6). Entonces el sistema tiene una única solución. □

**Ejemplo 3.1.1** Hallemos un polinomio cúbico que interpole los valores $f(1) = 1$, $f(2) = 1$, $f(3) = 2$, $f(4) = 6$ (éstos han sido tomados de la función llamada Gamma de Euler y denotada $\Gamma(x)$, Problema 3.1.17). Como en la demostración, tenemos

$$\begin{array}{ccccccccc}
a_0 & + & a_1 1 & + & a_2 1 & + & a_3 1 & = & 1, \\
a_0 & + & a_1 2 & + & a_2 4 & + & a_3 8 & = & 1, \\
a_0 & + & a_1 3 & + & a_2 9 & + & a_3 27 & = & 2, \\
a_0 & + & a_1 4 & + & a_2 16 & + & a_3 64 & = & 6.
\end{array}$$

Para resolver este sistema restamos a cada ecuación la que le precede

$$\begin{array}{ccccccc}
a_1 & + & 3a_2 & + & 7a_3 & = & 0, \\
a_1 & + & 5a_2 & + & 19a_3 & = & 1, \\
a_1 & + & 7a_2 & + & 37a_3 & = & 4.
\end{array}$$

Reiterando el proceso

$$\begin{array}{rcrcl} 2a_2 & + & 12a_3 & = & 1, \\ 2a_2 & + & 18a_3 & = & 3, \end{array}$$

y una última resta da $6a_3 = 2$, ó $a_3 = 1/3$. Seguidamente $2a_2 + 12a_3 = 1$, da $a_2 = -3/2$, y $a_1 + 3a_2 + 7a_3 = 0$ da $a_1 = 13/6$. Finalmente $a_0 + a_1 + a_2 + a_3 = 1$ lleva a $a_0 = 0$, con lo que

$$P(x) = 0 + \frac{13}{6}x - \frac{3}{2}x^2 + \frac{1}{3}x^3. \tag{3.2}$$

(Este polinomio es una aproximación muy mala a la función $\Gamma(x)$ porque las abscisas están muy distantes entre sí; lo hemos elegido porque, para $x$ entero positivo, $\Gamma(x)$ es entero positivo y los cálculos son sencillos.)□

### 3.1.2 La forma de Lagrange

El método anterior es laborioso. Se puede encontrar $P$ más fácilmente si en vez de buscarse en potencias de $x$, $P(x) = a_0 + a_1x + \cdots + a_Nx^N$, se busca como una combinación lineal $P(x) = b_0\ell_0(x) + b_1\ell_1(x) + \cdots + b_N\ell_N(x)$ de $N+1$ polinomios $\ell_i(x)$ de grado $N$, introducidos por Lagrange y que no hay que confundir con el propio $P$ (polinomio interpolador de Lagrange de $f$). Por definición

$$\ell_i(x) = \frac{(x-x_0)\ldots(x-x_{i-1})(x-x_{i+1})\cdots(x-x_N)}{(x_i-x_0)\ldots(x_i-x_{i-1})(x_i-x_{i+1})\cdots(x_i-x_N)}; \tag{3.3}$$

el factor $x - x_i$ está ausente del numerador así que el grado es $N$ como necesitamos. Si evaluamos $\ell_i(x)$ en un nodo que no sea el $x_i$ vamos a obtener cero; para $x = x_i$ resultará $\ell_i(x_i) = 1$ al coincidir numerador y denominador. Si ahora imponemos las condiciones de interpolación (3.1) caemos en el sistema trivial

$$\begin{array}{ccccccc} b_0 & & & & = & f(x_0), \\ & b_1 & & & = & f(x_1), \\ & & \ddots & & \vdots & \vdots \\ & & & b_N & = & f(x_N), \end{array}$$

con lo que

$$P(x) = f(x_0)\ell_0(x) + f(x_1)\ell_1(x) + \cdots + f(x_N)\ell_N(x). \tag{3.4}$$

Esta construcción, debida a Lagrange, no precisa de la resolución de sistemas lineales. ¿Observa el paralelismo de esta sección con otra anterior?

**Ejemplo 3.1.2** Volvamos al caso del Ejemplo 3.1.1 que ahora resolveremos en la forma de Lagrange para apreciar el ahorro de trabajo que ésta supone en relación con el método de coeficientes indeterminados. Los polinomios de Lagrange son

$$\begin{aligned}
\ell_0(x) &= \frac{(x-2)(x-3)(x-4)}{(1-2)(1-3)(1-4)}, \\
\ell_1(x) &= \frac{(x-1)(x-3)(x-4)}{(2-1)(2-3)(2-4)}, \\
\ell_2(x) &= \frac{(x-1)(x-2)(x-4)}{(3-1)(3-2)(3-4)}, \\
\ell_3(x) &= \frac{(x-1)(x-2)(x-3)}{(4-1)(4-2)(4-3)},
\end{aligned}$$

y el polinomio de interpolación es

$$\begin{aligned}
P(x) &= 1\ell_0(x) + 1\ell_1(x) + 2\ell_2(x) + 6\ell_3(x) \\
&= -\frac{1}{6}(x-2)(x-3)(x-4) + \frac{1}{2}(x-1)(x-3)(x-4) \\
&\quad - (x-1)(x-2)(x-4) + (x-1)(x-2)(x-3). \qquad (3.5)
\end{aligned}$$

Por la unicidad, este polinomio debe ser el mismo que encontramos en (3.2). Si redujésemos (3.5) a potencias de $x$ deberíamos obtener los coeficientes de $x^i$ en (3.2). En general, *no* debe reducirse la forma de Lagrange a potencias de $x$ porque ello puede implicar problemas de pérdida de precisión al usar mantisas de longitud finita. □

**Problema 3.1.1** Usando la tabla del comienzo de la sección y la forma de Lagrange, calcule aproximadamente el seno de 72 grados, mediante interpolación (por un polinomio) lineal en las abscisas 70 y 75. (Resultado: 0.950186, el valor verdadero es 0.951056.) Luego use interpolación cuadrática basada en las abscisas 65, 70 y 75. (Resultado: 0.951044.)

**Problema 3.1.2** El peso específico $p$ del agua a diversas temperaturas centígradas $t$ es como sigue

| $t$ | $p$ |
|---|---|
| 0 | 0.999871 |
| 1 | 0.999928 |
| 2 | 0.999969 |
| 3 | 0.999991 |

Aproxime el valor en $t = 4$ usando la forma de Lagrange e interpolación lineal en las abscisas 2, 3; cuadrática en las abscisas 1, 2, 3; cúbica en las abscisas 0, 1, 2, 3. (El verdadero valor es 1.000000, las aproximaciones son 1.000013, 0.999994, 0.999991; la cúbica es peor que la cuadrática.)

**Problema 3.1.3** La tensión de vapor (en kilopascales) del helio líquido a varias temperaturas (en grados Kelvin) es

| $T$ | $p$ |
|---|---|
| 3.0 | 24.0 |
| 3.1 | 27.8 |
| 3.2 | 32.0 |
| 3.3 | 36.5 |

Aproxime la tensión para 3.4 K usando interpolación lineal en 3.3, 3.2; cuadrática en 3.3, 3.2, 3.1 y cúbica. El verdadero valor es 41.5.

**Problema 3.1.4** Use la siguiente tabla de valores de la función $\Gamma$ para hallar $\Gamma(1.23)$ mediante una interpolación cúbica (verdadero 0.91075)

| $x$ | $\Gamma(x)$ |
|---|---|
| 1.0 | 1.00000 |
| 1.1 | 0.95135 |
| 1.2 | 0.91817 |
| 1.3 | 0.89747 |

**Problema 3.1.5** *Interpolación inversa.* En el contexto del Problema 3.1.3, podemos necesitar saber para qué temperatura $T$ la tensión de vapor vale 30 kilopascales. Podríamos pensar en interpolar la función $p = f(T)$ por un polinomio cúbico, $p = P_3(T)$, y luego resolver la ecuación $P_3(T) = 30$. Hay un camino mejor, llamado interpolación inversa: cambie los papeles de las variables $p$ y $T$, construya un interpolante cúbico $T = Q_3(p)$ y evalúe en $p = 30$.

**Problema 3.1.6** *Matrices de Vandermonde.* Dados $N+1$ números $x_i$, $i = 0, 1, \ldots, N$, no necesariamente distintos dos a dos, la matriz cuadrada cuyo elemento $(i,j)$, $i, j = 0, 1, \ldots, N$, es $x_i^j$, se llama de Vandermonde. Calcule el correspondiente determinante operando con las columnas de la matriz. (Solución: el determinante es el producto de los $N(N+1)/2$ factores de la forma $x_i - x_j$ donde $i > j$.) Una matriz de Vandermonde es no singular si y solo si los $x_i$ son dos a dos distintos.

**Problema 3.1.7** *Vandermonde otra vez.* Calcule el determinante de la matriz de Vandermonde sin necesidad de hacer combinaciones lineales en la matriz. Para ello fije los valores de $x_1, \ldots, x_N$ y considere $x_0$ como una variable. Entonces el determinante es un polinomio en $x_0$ de grado $\leq N$. Escriba este polinomio como producto de su coeficiente director (coeficiente de $x_0^N$) y de $N$ factores $(x_0 - \xi_i)$ donde los $\xi_i$ son los correspondientes ceros (cuyo valor es fácil de identificar).

**Problema 3.1.8** *Polinomios de Lagrange.* Dado un problema de interpolación de Lagrange, pruebe que los correspondientes $\ell_i$ forman una base del espacio vectorial de los polinomios de grado $\leq N$.

**Problema 3.1.9** *Identidades satisfechas por los polinomios de Lagrange.* Sin efectuar ningún cálculo pruebe que los $\ell_i(x)$ suman $\equiv 1$. (Indicación: la suma de los $\ell_i(x)$ es un polinomio de grado ..., que en los nodos ...; aplique la unicidad.) Pruebe a continuación que $x_0\ell_0(x) + x_1\ell_1(x) + \cdots + x_N\ell_N(x) \equiv x$. Generalice.

**Problema 3.1.10** *Una expresión para los $\ell_i(x)$.* Si $W(x) = (x - x_0)\cdots(x - x_N)$, pruebe que $\ell_i(x) = W(x)/[(x - x_i)W'(x_i)]$.

**Problema 3.1.11** Si en el problema de Lagrange, la propia función $f$ es un polinomio de grado $\leq N$ ¿qué polinomio de interpolación se obtiene? Dados $N + 1$ nodos dos a dos distintos, se interpola en ellos una función y luego se interpola el polinomio interpolante ¿cuál es el resultado?

**Problema 3.1.12** Si en el problema de interpolación de Lagrange el polinomio se buscase de grado $\leq N + 1$ la solución ya no sería única. Escriba todos los polinomios de grado $\leq N + 1$ que satisfacen (3.1). Escriba todos los polinomios que satisfacen (3.1).

**Problema 3.1.13** *Cambio de variable en el polinomio de interpolación.* Sea $y = f(x)$ una función y sea $p(x)$ el polinomio interpolador de Lagrange (grado $\leq N$) en las $N + 1$ abscisas dos a dos distintas $x_0, \ldots, x_N$. Si $\lambda \neq 0$ y $\mu$ son dos constantes, se introducen la nueva variable $X$ tal que $x = \lambda X + \mu$ y la nueva función $F(X) = f(\lambda X + \mu)$. Sin hacer ningún cálculo, pruebe que $p(\lambda X + \mu)$ es el polinomio interpolador $P(X)$ de $F(X)$ en las abscisas $X_0, \ldots, X_N$, tales que $x_0 = \lambda X_0 + \mu, \ldots, x_N = \lambda X_N + \mu$. ¿Qué ocurre si el cambio de variable es más general, $x = \phi(X)$?

**Problema 3.1.14** *Continuación.* Un científico norteamericano posee una tabla que expresa la densidad del agua en gramos por centímetro cúbico a diferentes temperaturas Fahrenheit; un ciéntifico alemán posee una tabla que difiere de la anterior en que las temperaturas vienen dadas en grados centígrados. Ignorando los errores que resulten de utilizar mantisas de longitud finita ¿qué tabla es más conveniente para obtener valores interpolados?

**Problema 3.1.15** *Cambio de variable dependiente en el polinomio de interpolación.* Sea $y = f(x)$ una función y sea $p(x)$ el polinomio interpolador de Lagrange (grado $\leq N$) en las $N + 1$ abscisas dos a dos distintas $x_0, \ldots, x_N$. Se considera el cambio $Y = Ay + B$ con $A \neq 0$ y $B$ constantes. Explique cómo construir, a partir de $p(x)$, el polinomio interpolador $P(x)$ de la función $Af(x) + B$ en las abscisas $x_0, \ldots, x_N$.

**Problema 3.1.16** *Cambios de variable en la forma de Lagrange.* Resuelva los problemas 3.1.13 y 3.1.15 escribiendo $p(x)$ en forma de Lagrange y llevando a cabo de modo explícito los cambios de variable.

**Problema 3.1.17** *La función Gamma.* Para $x > 0$ se define

$$\Gamma(x) = \int_0^\infty e^{-s} s^{x-1}\, ds.$$

Pruebe que, para cada $x > 0$, $\Gamma(x + 1) = x\Gamma(x)$ y que $\Gamma(1) = 1$. Concluya que, si $x > 0$ es un entero, $\Gamma(x) = (x - 1)!$.

## 3.2 La forma de Newton; diferencias divididas

### 3.2.1 Construcción por recurrencia del polinomio interpolador de Lagrange

Hemos visto ya dos formas de obtener la solución de un problema de Lagrange: el método de coeficientes indeterminados y la forma de Lagrange. En esta sección veremos una tercera, debida al científico más importante de la historia, Isaac Newton. La idea básica es construir la solución en pasos sucesivos. Primero construiremos el polinomio $P_0$ de grado $\leq 0$ (es decir constante) que coincide con la función en el punto $x_0$. Luego, construiremos el polinomio $P_1$ de grado $\leq 1$ que coincide con la función en $x_0$, $x_1$. A continuación, el polinomio $P_2$ de grado $\leq 2$ que coincide con la función en $x_0$, $x_1$, $x_2$ y así hasta llegar al polinomio que buscamos (con grado $\leq N$ que satisface (3.1)). Observemos que en cada paso el polinomio $P_i$ existe y es único en virtud de la existencia y unicidad de la solución del problema de Lagrange.

¿Cómo es $P_0$? Evidentemente $P_0$ es el polinomio constante igual a $f(x_0)$. Para hallar $P_1$ escribamos $P_1 = P_0 + Q_1$, donde $Q_1$ es un polinomio que hay que determinar. El grado de $Q_1$ ha de ser $\leq 1$, porque queremos que $P_1$ tenga grado $\leq 1$ y $P_0$ es una constante. Por otro lado, $Q_1$ debe anularse en $x = x_0$, ya que $P_1(x_0) = P_0(x_0) = f(x_0)$. Así, $Q_1(x)$ debe ser de la forma $c_1(x - x_0)$. Para determinar $c_1$ imponemos que $P_1(x_1) = f(x_1)$, es decir $P_0 + c_1(x_1 - x_0) = f(x_1)$ y de aquí despejamos $c_1$. La construcción de $P_2(x)$ conocido $P_1(x)$ es similar. Escribimos $P_2 = P_1 + Q_2$ y hay que determinar $Q_2$. Este tiene grado $\leq 2$ (¿por qué?) y debe anularse en $x_0$, $x_1$ (porque esos puntos $P_1$ coincide con $P_2$ por coincidir ambos con $f$). Por ello, $Q_2(x) = c_2(x-x_1)(x-x_2)$, donde $c_2$ es una constante que se determina imponiendo que $\cancel{Q_2}(x_2) = f(x_2)$. Siguiendo esta marcha se obtiene el polinomio buscado $P_2(x_2) = f(x_2)$

$$
\begin{aligned}
P(x) &= P_{N-1}(x) + Q_N(x) \\
&= P_{N-1}(x) + c_N(x - x_0)(x - x_1)\cdots(x - x_{N-1}) \\
&= P_{N-2}(x) + Q_{N-1}(x) + c_N(x - x_0)(x - x_1)\cdots(x - x_{N-1}) \\
&= P_{N-2}(x) + c_{N-1}(x - x_0)(x - x_1)\cdots(x - x_{N-2}) \\
&\quad + c_N(x - x_0)(x - x_1)\cdots(x - x_{N-1}) \\
&= \cdots
\end{aligned}
$$

y, en definitiva,

$$P(x) = c_0$$

$$
\begin{aligned}
&+ c_1(x - x_0) \\
&+ c_2(x - x_0)(x - x_1) \\
&\dots \\
&+ c_N(x - x_0)(x - x_1)\cdots(x - x_{N-1}).
\end{aligned} \tag{3.6}
$$

Por consiguiente, $P$ queda ahora escrito como combinación lineal de

$$1, \quad (x - x_0), \quad (x - x_0)(x - x_1), \quad \dots, \quad (x - x_0)\cdots(x - x_{N-1}); \tag{3.7}$$

esto ha de compararse con las construcciones anteriores de $P$ donde las bases eran las de los monomios $x^i$ o la de los $\ell_i$. *¿Se fija en que $x_N$ no aparece en (3.7)?*

En (3.6), la primera fila da la constante que interpola en $x_0$, las dos primeras filas la recta que interpola en $x_0$ y $x_1$, etc. Esta propiedad recuerda otra análoga del polinomio de Taylor (2.5).

**Ejemplo 3.2.1** Resolvamos en esta forma recurrente el caso del Ejemplo 3.1.1. Sucesivamente,

$$
\begin{aligned}
P_0 &= 1, \\
P_1(x) &= 1 + c_1(x - 1), \\
1 &= 1 + c_1(2 - 1), \\
c_1 &= 0, \\
P_1(x) &= 1, \\
P_2(x) &= 1 + c_2(x - 1)(x - 2), \\
2 &= 1 + c_2(3 - 1)(3 - 2), \\
c_2 &= \frac{1}{2}, \\
P_2(x) &= 1 + \frac{1}{2}(x - 1)(x - 2), \\
P_3(x) &= 1 + \frac{1}{2}(x - 1)(x - 2) + c_3(x - 1)(x - 2)(x - 3), \\
6 &= 1 + \frac{1}{2}(4 - 1)(4 - 2) + c_3(4 - 1)(4 - 2)(4 - 3), \\
c_3 &= \frac{1}{3},
\end{aligned}
$$

y el polinomio buscado es (para comparar con (3.6) se incluye el término con $c_1$, que es nulo)

$$P(x) = 1 + 0(x - 1) + \frac{1}{2}(x - 1)(x - 2) + \frac{1}{3}(x - 1)(x - 2)(x - 3). \tag{3.8}$$

Este $P$ debe ser, enmascarado, el mismo que ya encontramos en (3.2) y (3.5). Omitiendo en (3.8) el último sumando se obtiene la parábola cuadrática que interpola en $x = 1, 2, 3$; omitiendo los dos últimos sumandos la recta que interpola en $x = 1, 2$; omitiendo los tres últimos la constante que interpola en $x = 1$. □

### 3.2.2 Cambio de orden en la construcción

En la construcción anterior se van imponiendo las condiciones de interpolación (3.1) en el orden $x_0, x_1, \ldots, x_N$. En realidad, éstas se pueden ir añadiendo en uno de los $(N+1)!$ órdenes posibles de las abscisas $x_i$. Con todos estos órdenes se ha de llegar al mismo $P$ final, que bien sabemos es único. Sin embargo las étapas intermedias de la construcción serán distintas.

**Ejemplo 3.2.2** Construyamos una vez más el polinomio (3.8) por recurrencia, esta vez considerando los nodos en el orden 4, 3, 2, 1. Las cantidades intermedias ahora se denotarán con estrellas para no confundirlas con las que aparecieron en el Ejemplo 3.2.1.

$$
\begin{aligned}
P_0^* &= 6, \\
P_1^*(x) &= 6 + c_1^*(x-4), \\
2 &= 6 + c_1^*(3-4), \\
c_1^* &= 4, \\
P_1^*(x) &= 6 + 4(x-4), \\
P_2^*(x) &= 6 + 4(x-4) + c_2^*(x-4)(x-3), \\
1 &= 6 + 4(2-4) + c_2^*(2-4)(2-3), \\
c_2^* &= \frac{3}{2}, \\
P_2^*(x) &= 6 + 4(x-4) + \frac{3}{2}(x-4)(x-3),, \\
P_3^*(x) &= 6 + 4(x-4) + \frac{3}{2}(x-4)(x-3) + c_3^*(x-4)(x-3)(x-2), \\
1 &= 6 + 4(1-4) + \frac{3}{2}(1-4)(1-3) + c_3^*(1-4)(1-3)(1-2), \\
c_3^* &= \frac{1}{3},
\end{aligned}
$$

de donde

$$P(x) = 6 + 4(x-4) + \frac{3}{2}(x-4)(x-3) + \frac{1}{3}(x-4)(x-3)(x-2), \tag{3.9}$$

polinomio que no puede ser otro sino (3.2), alias (3.5), alias (3.8). Ahora (compare con (3.8)), al suprimir el último sumando se halla la parábola cuadrática que interpola en 4, 3, 2; al suprimir los dos últimos la recta que interpola en 4, 3; al suprimir los tres últimos la constante que interpola en 4. □

### 3.2.3 Diferencias divididas de una función

Vamos a dar un medio de calcular las constantes $c_i$ que aparecen en la forma de Newton (3.6). Introduciremos una definición. Dados un entero no negativo $N$,

$N+1$ puntos $x_0, \ldots, x_N$ dos a dos distintos y una función $f$ definida en ellos, se llama *diferencia dividida* de la función en los puntos al coeficiente de $x^N$ en el desarrollo en potencias de $x$ del correspondiente polinomio interpolador de Lagrange. Se representa esta diferencia dividida por $f[x_0, x_1, \ldots, x_N]$. El entero $N$ se llama *orden* de la diferencia dividida. De la propia definición, se sigue que el valor de una diferencia dividida es *independiente del orden* en que se escriban sus argumentos; p. ej. $f[1, \pi] = f[\pi, 1]$, ambos son el coeficiente de $x$ en la recta que interpola a la función en $x = 1$ y $x = \pi$, que es lo mismo que interpolarla en $x = \pi$ y $x = 1$.

**Ejemplo 3.2.3** Para la función $\Gamma$ que en el Ejemplo 3.1.1 interpolamos en los puntos 1, 2, 3 y 4 , vemos inmediatamente en (3.2) que $\Gamma[1,2,3,4] = 1/3$. Esta es una diferencia dividida *tercera* de la función Gamma. El mismo valor $1/3$ puede obtenerse en las expresiones alternativas (3.5) (sumando $-1/6 + 1/2 - 1 + 1$) y (3.8) ó (3.9). De (3.8) podemos recuperar más valores de diferencias divididas. En efecto, puesto que al suprimir el último sumando obtenemos el polinomio cuadrático que interpola en 1, 2, 3, tendremos que $\Gamma[1,2,3] = 1/2$; ésta es una diferencia *segunda*. Análogamente $\Gamma[1,2] = 0$, porque al suprimir los dos últimos términos se obtiene la recta que interpola en 1, 2, y $\Gamma[1] = 1$, porque suprimiendo los tres últimos términos se recae en la constante que interpola en 1. A su vez, de (3.9) obtenemos por consideraciones análogas $\Gamma[2,3,4] = 3/2$, $\Gamma[3,4] = 4$ y $\Gamma[4] = 6$. $\square$

Por el argumento dado en el ejemplo, los coeficientes $c_i$ que aparecen en la forma de Newton (3.6) son diferencias divididas de la función; concretamente

$$\begin{aligned} P(x) = \quad & f[x_0] \\ & + f[x_0, x_1](x - x_0) \\ & + f[x_0, x_1, x_2](x - x_0)(x - x_1) \\ & \cdots \\ & + f[x_0, x_1, \ldots, x_N](x - x_0)(x - x_1) \cdots (x - x_{N-1}). \end{aligned} \tag{3.10}$$

### 3.2.4 Cálculo de las diferencias divididas

Es claro que las diferencias divididas de orden cero de una función son los propios valores de la función, $f[x_i] = f(x_i)$. Las diferencias divididas de orden $N \geq 1$ pueden calcularse fácilmente en términos de diferencias divididas de orden $N-1$, tal y como afirma el siguiente teorema, que también explica el nombre 'diferencias divididas'.

**Teorema 3.2.1** *Si $x_0, \ldots, x_N$ son $N+1$ puntos, $N \geq 1$, dos a dos distintos en*

*los que una función $f$ está definida,*

$$f[x_0, x_1, \ldots, x_N] = \frac{f[x_1, \ldots, x_N] - f[x_0, \ldots, x_{N-1}]}{x_N - x_0}. \tag{3.11}$$

*Demostración.* Construyamos el polinomio interpolador de Lagrange en los puntos $x_0, \ldots, x_N$ por recurrencia de dos maneras: primero añadiendo los nodos en el orden $x_0, \ldots, x_N$ y luego en el inverso $x_N, \ldots, x_0$. Con el primer orden llegaremos a (3.10). Con el segundo orden llegaremos a la expresión que se obtiene al reemplazar en (3.10) $x_0$ por $x_N$, $x_1$ por $x_{N-1}$ etc., es decir a

$$\begin{aligned} P(x) = \quad & f[x_N] \\ & + f[x_N, x_{N-1}](x - x_N) \\ & + f[x_N, x_{N-1}, x_{N-2}](x - x_N)(x - x_{N-1}) \\ & \cdots \\ & + f[x_N, x_{N-1}, \ldots, x_0](x - x_N)(x - x_{N-1}) \cdots (x - x_1). \end{aligned} \tag{3.12}$$

En (3.10) el coeficiente de $x^{N-1}$ del desarrollo en potencias del segundo miembro es

$$f[x_0, \ldots, x_{N-1}] + f[x_0, \ldots, x_N](-x_0 - \cdots - x_{N-1}),$$

mientras que en (3.13) el coeficiente de $x^{N-1}$ del desarrollo en potencias del segundo miembro es

$$f[x_1, \ldots, x_N] + f[x_0, \ldots, x_N](-x_1 - \cdots - x_N)$$

(hemos tenido en cuenta que las diferencias divididas no dependen del orden en que se escriban sus argumentos). Igualando estas expresiones y despejando la diferencia dividida de orden $N$ se llega a (3.11). □

### 3.2.5 La tabla de diferencias divididas

Usando reiteradamente el teorema anterior podemos construir la siguiente tabla. Los números de la columna de la izquierda son las abscisas e inmediatamente a su derecha aparecen los correspondientes valores de la función. Cada uno de los restantes elementos de la tabla se obtiene a partir de los que están a su oeste

y a su noroeste

$$\begin{array}{lllll}
x_0 & f[x_0] & & & \\
x_1 & f[x_1] & f[x_0,x_1] & & \\
x_2 & f[x_2] & f[x_1,x_2] & f[x_0,x_1,x_2] & \\
\vdots & \vdots & \vdots & \vdots & \ddots \\
x_N & f[x_N] & f[x_{N-1},x_N] & f[x_{N-2},x_{N-1},x_N] & \cdots \quad f[x_0,\ldots,x_N]
\end{array}.$$

Hallada la tabla, los elementos diagonales nos dan los coeficientes por los que hay que multiplicar los polinomios de la base (3.7) para formar el polinomio solución del problema de Lagrange en su forma de Newton (3.10). *Esta es la forma práctica usual de construir la forma de Newton y no el proceso recurrente que presentamos más arriba.*

**Ejemplo 3.2.4** Para la construcción del Ejemplo 3.2.1, la tabla de diferencias divididas es

$$\begin{array}{llll}
1 & 1 & & \\
2 & 1 & \frac{1-1}{2-1}=0 & \\
3 & 2 & \frac{2-1}{3-2}=1 & \frac{1-0}{3-1}=\frac{1}{2} \\
4 & 6 & \frac{6-2}{4-3}=4 & \frac{4-1}{4-2}=\frac{3}{2} \quad \frac{3/2-1/2}{4-1}=\frac{1}{3}
\end{array}$$

Hemos empleado fracciones pues los datos son enteros. Para datos no enteros o en un ordenador se van haciendo las divisiones, e.g. $1/3 = 0.333$. *Preste particular atención a los denominadores, donde son frecuentes las equivocaciones.* Formada esta tabla podemos leer de ella los cuatro coeficientes que hacen falta para escribir (3.8). □

### 3.2.6 Comparación entre las formas de coeficientes indeterminados, Lagrange y Newton

Hemos visto ya tres maneras de construir la solución de un problema de interpolación de Lagrange dado: por coeficientes indeterminados, en la forma de Lagrange y en la forma de Newton. Comparemos.

- Costo de escribir el polinomio interpolador Es *máximo* con coeficientes indeterminados donde se requiere resolver un sistema lineal $(N+1)\times(N+1)$, *nulo* para la forma de Lagrange e *intermedio* para la de Newton, donde hay que construirse la tabla de diferencias divididas.

- Costo de, una vez formado el polinomio, evaluarlo en un punto. Es *mínimo* en la forma de coeficientes indeterminados, donde se aplica el algoritmo de Horner. Para la forma de Newton el costo es también *bajo* pues existe un algoritmo tipo Horner (Problema 3.2.5). La forma de Lagrange es la más enojosa (Problema 3.2.6).

- Supongamos que se ha hallado en forma de Newton, mediante la correspondiente tabla de diferencias divididas, el polinomio interpolador (de grado $\leq N$) de una función $f$ en $N+1$ abscisas $x_0, \ldots, x_N$ y que se desea construir el de grado $\leq N+1$ que interpola en un nodo más $x_{N+1}$. Basta con añadir un término al polinomio ya obtenido y el coeficiente necesario se obtiene añadiendo una *fila* a la tabla de diferencias divididas ya calculada. En las formas de coeficientes indeterminados y Lagrange no es inmediatamente posible utilizar los cálculos que se hicieron para hallar el polinomio de grado $\leq N$.

- En la forma de Newton, cada sumando de los que componen el polinomio tiene un significado: son las cantidades que hay que añadir a un polinomio interpolador de un cierto grado para transformarlo en el de un grado más que interpola en un punto más. Por eso, al ir sumando sucesivamente, es posible observar los efectos de pasar de la interpolación lineal a la cuadrática, de ésta a la cúbica, etc. En las otras dos formas los sumandos individuales carecen de significado aprovechable.

En términos generales la forma de Newton es la más aconsejable; pero la de Lagrange también se emplea, sobre todo en cuestiones teóricas.

**Problema 3.2.1** Resuelva los Problemas 3.1.1, 3.1.2, 3.1.3, 3.1.4 usando la forma de Newton.

**Problema 3.2.2** *Diferencias divididas de un polinomio.* Pruebe, usando la definición, que si $f$ es un polinomio de grado $\leq N$ entonces cualquier diferencia dividida de orden $\geq N+1$ de $f$ es nula.

**Problema 3.2.3** ¿Cuántas operaciones se hacen al formar una tabla de diferencias divididas?

**Problema 3.2.4** ¿Están en la tabla del Ejemplo 3.2.4 los coeficientes necesarios para escribir (3.9)? Dados $N+1$ nodos, hay $(N+1)!$ maneras de formar el polinomio interpolador por el método de Newton, correspondientes a reordenadaciones de los nodos; hay también $(N+1)!$ tablas de diferencias divididas posibles. En una tabla dada ¿para cuántas formas de Newton podemos leer los coeficientes? (Respuesta: para $2^N$.)

**Problema 3.2.5** *Evaluación de la forma de Newton.* Formule un algoritmo tipo Horner para evaluar polinomios interpoladores escritos en la forma de Newton. Deben usarse $N$ multiplicaciones y $2N$ sumas/restas.

**Problema 3.2.6** *Evaluación de la forma de Lagrange.* Pruebe que el polinomio interpolador en forma de Lagrange puede reescribirse en la forma

$$P(x) = W(x)\left[\frac{F_0}{x-x_0} + \cdots + \frac{F_N}{x-x_N}\right],$$

donde $W(x) = (x-x_0)\cdots(x-x_N)$ y las $F_i$ son constantes. *Reescrito* de esta forma el polinomio se evalúa con $2N+1$ sumas/restas, $N+1$ multiplicaciones y $N+1$ divisiones.

**Problema 3.2.7** *Una fórmula para las diferencias divididas.* Pruebe que

$$f[x_0, x_1, \ldots, x_N] = \sum_{i=0}^{N} \frac{f(x_i)}{(x_i - x_0)\cdots(x_i - x_{i-1})(x_i - x_{i+1})\cdots(x_i - x_N)}.$$

**Problema 3.2.8** *Cambio de variable en la forma de Newton.* Las notaciones son las del Problema 3.1.13. ¿Cuál es la relación entre las diferencias divididas de $f$ en los nodos $x_0, \ldots, x_N$ y las de $F$ en los nodos $X_0, \ldots, X_N$? Demuestre que $P(X) = p(x)$ llevando a cabo explícitamente el cambio de variable en $p(x)$ escrito en forma de Newton. Analice la forma de Newton desde el punto de vista del análisis dimensional (vea antes el Ejemplo 2.6.2).

**Problema 3.2.9** *Diferencias divididas de un producto.* Pruebe que si la función $f$ es el producto de otras dos $g$ y $h$ entonces

$$f[x_0, \ldots, x_N] = \\ g[x_0]h[x_0, \ldots, x_N] + g[x_0, x_1]h[x_1, \ldots, x_N] + \cdots + g[x_0, \ldots, x_N]h[x_N].$$

(Indicación. Escriba en forma de Newton el polinomio interpolador $G$ de $g$ tomando los nodos en el orden $x_0, \ldots, x_N$; escriba después el polinomio interpolador $H$ de $h$ pero ahora tomando los nodos en orden inverso. El producto $GH$ coincide con la función en los nodos pero no es la solución del problema de Lagrange porque su grado es $\leq 2N$. Suprima términos del producto $GH$ hasta encontrar la solución del problema de Lagrange y luego use la definición de diferencia dividida como coeficiente de $x^N$.)

**Problema 3.2.10** *Lema de Aitken (1895–1967).* En el problema de Lagrange, denotemos por $W$ al conjunto de los nodos y, para cada subconjunto $S$ no vacío de $W$, llamemos $P_S$ al polinomio interpolador de Lagrange de $f$ en $S$. (Esto es, si $S$ tiene $n+1$ puntos $x_i$, entonces $P_S$ tiene grado $\leq n$ y $P_S(x_i) = f(x_i)$ para $x_i \in S$.) Pruebe que si $S$ y $T$ son dos subconjuntos no vacíos de $W$ que tienen todos sus puntos en común salvo el $x_i$, que está en $S$ y no en $T$, y el $x_j$ que está en $T$ y no en $S$, entonces

$$P_{S\cup T} = \frac{(x_i - x)P_T(x) - (x_j - x)P_S(x)}{x_i - x_j}.$$

¿A qué se reduce la fórmula si $S$ y $T$ sólo tuviesen un punto cada uno?

**Problema 3.2.11** *El algoritmo de Neville.* Basándonos en el Lema de Aitken podemos construir $P = P_W$ a partir de dos polinomios que interpolen cada uno en $N$ puntos. Cada uno de estos se puede construir a su vez a partir de dos que interpolen en $N-1$ puntos y así hasta llegar a polinomios constantes. Hay varias maneras de implementar esta idea para hallar el polinomio de interpolación. En una debida a Neville se define para $k = 0, \ldots, N$, $P_{k0} = f(x_k)$ y para $g = 0, 1, \ldots, N-1$, $k = g+1, \ldots, N$,

$$P_{k,g+1}(x) = \frac{(x_k - x)P_{k-1,g}(x) - (x_{k-g-1} - x)P_{k,g}(x)}{x_k - x_{k-g-1}}.$$

Pruebe, usando el Lema de Aitken, que $P_{k,g}$ es el polinomio de grado $\leq g$ que interpola en los $g+1$ puntos $x_{k-g}, \ldots, x_k$. En particular $P_{N,N}$ es la solución $P$ del problema de Lagrange. Este proceso de Neville tiene poca utilidad para hallar el *polinomio* de interpolación $P$, pero a veces se usa para encontrar el valor $P(x)$ en un valor númerico dado. La ventaja es que cada una de las cantidades intermedias $P_{k,g}(x)$ es una aproximación al valor $f(x)$ que buscamos.

**Problema 3.2.12** *El algoritmo de Neville en el cálculo con papel y lápiz.* Construya una tabla triangular (semejante a la de diferencias divididas) que tenga los valores $P_{k,0}$ en la primera columna, los $P_{k,1}(x)$ en la segunda, etc. Al completar la tabla ha hallado $P(x)$ y también muchas otras aproximaciones al valor de la función $f(x)$; comparándolas entre sí se puede conjeturar el tamaño del error de interpolación. Lleve a cabo los cálculos en el caso del Problema 3.1.1.

# 3.3 Cotas de error

## 3.3.1 El error de interpolación

Formado, por el procedimiento que sea, el polinomio $P$ solución del problema de Lagrange ¿Qué diferencia hay entre su valor $P(x)$ en un punto $x$ dado y el verdadero valor $f(x)$ de la función?

**Teorema 3.3.1** *Supongamos que $f$ es una función con $N \geq 1$ derivadas continuas en un intervalo $[a,b]$ y tal que $f^{(N+1)}$ existe en $(a,b)$. Sean $x_0, \ldots, x_N$, $N+1$ nodos en $[a,b]$, dos a dos distintos, y sea $P$ el correspondiente polinomio interpolador. Entonces a cada punto $x$ en $[a,b]$ le corresponde un punto $\xi$ con*

$$\text{mín}(x_0, \ldots, x_N, x) < \xi < \text{máx}(x_0, \ldots, x_N, x)$$

*para el cual*

$$f(x) - P(x) = \frac{(x-x_0)\cdots(x-x_N)}{(N+1)!} f^{(N+1)}(\xi). \tag{3.13}$$

*Demostración.* Elijamos un número $x$ en $[a,b]$. Si $x$ es uno de los $x_i$, la igualdad (3.13) es válida elijamos como elijamos $\xi$. Si $x$ no es uno de los nodos, definamos

$$M = \frac{f(x) - P(x)}{(x-x_0)\cdots(x-x_N)},$$

con lo que hay que probar que existe $\xi$ en el intervalo anunciado de suerte que $f^{(N+1)}(\xi) = M(N+1)!$. Se forma la función auxiliar

$$F(t) = f(t) - P(t) - M(t-x_0)\cdots(t-x_N),$$

en la que hemos llamado $t$ a la variable porque la letra $x$ ya la estábamos ultilizando para otra cosa. Se tiene que $F^{(N+1)}(t) = f^{(N+1)}(t) - M(N+1)!$; tras esto basta probar que $F^{(N+1)}(t)$ se anula en algún punto del intervalo anunciado.

De la definición de $M$ se sigue que la función $F$ se anula cuando a su variable $t$ se le da el valor numérico $x$. Por otro lado $F$ se anula en los $N+1$ nodos $x_i$. Apliquemos el terorema de Rolle: puesto que $F$ tiene $N+2$ ceros distintos, su derivada tiene $N+1$ (uno entre cada dos de $F$), la derivada segunda $N, \ldots,$ y la derivada $N+1$ uno. □

Los siguientes comentarios ayudan a entender y a memorizar la expresión (3.13):

- La presencia de los factores $(x - x_i)$, $i = 0, \ldots, N$ en el segundo miembro garantiza que el primero es nulo en los nodos, de acuerdo con la definición de $P$.

- Cuando la función $f$ que se interpola es ella misma un polinomio de grado $\leq N$, el primer miembro es nulo para cada $x$ porque $P$ coincide con $f$ y el segundo lo es porque la derivada $N+1$ de un polinomio de grado $\leq N$ es idénticamente nula.

- Los $N+1$ factores $(x - x_i)$ combinados con la derivada de orden $N+1$ hacen la fórmula dimensionalmente correcta (vea el Ejemplo 2.6.2).

**Corolario 3.3.1** *Supongamos, además de las hipótesis del teorema, que, para cada $t$ en $(a, b)$, se tiene $|f^{(N+1)}(t)| \leq K_{N+1}$. Entonces*[1]

$$|f(x) - P(x)| \leq \frac{|x - x_0| \cdots |x - x_N| K_{N+1}}{(N+1)!}. \tag{3.14}$$

**Ejemplo 3.3.1** Para la interpolación lineal basada en los nodos $x_0, x_1, x_1 > x_0$, la cota anterior es $(1/2)|(x-x_0)(x-x_1)|K_2$, con $K_2$ cota de la derivada segunda de $f$. Supongamos que el punto $x$ donde evaluamos el interpolante está entre $x_0$ y $x_1$, es decir se evita lo que a veces se llama *extrapolación* (evaluar fuera del menor intervalo que contenga a los nodos). Para estos $x$ el valor mayor posible de $|(x-x_0)(x-x_1)|$ se alcanza en el punto medio $(x_0+x_1)/2$ y es $(x_1-x_0)^2/4$ (para esos $x$, se tiene $|(x-x_0)(x-x_1)| = (x-x_0)(x_1-x)$, parábola cuadrática nula en $x_0$ y $x_1$; por simetría su vértice está en el punto medio, haga una gráfica). En resumen para el interpolante lineal $P_1$ podemos escribir

$$|f(x) - P_1(x)| \leq \frac{1}{8}(x_1 - x_0)^2 K_2, \quad x \in [x_0, x_1]. \square \tag{3.15}$$

[1] La cota se refiere al *verdadero valor* $P(x)$. En la práctica no se conocerá $P(x)$ exactamente debido a la precisión finita de los coeficientes de $P$ y de las operaciones aritméticas de evaluarlo en $x$.

**Ejemplo 3.3.2** Se utiliza una tabla de la función seno, tabulada de cinco en cinco grados desde 0 a 90 (como la del comienzo de la lección) para hallar senos, interpolando linealmente entre el nodo inmediatamente mayor e inmediatamente menor al ángulo deseado, ¿en qué errores se incurre? La función que se está considerando no es la que en matemáticas se denota por $\sin x$, ya que los ángulos no vienen en radianes; estamos tratando con $f(x) = \sin(2\pi x/360)$. Su derivada segunda es $-(2\pi/360)^2 \sin(2\pi x/360)$ y $K_2 = (2\pi/360)^2 \approx 0{,}0003$. Entonces la cota (3.15) es $(0.0003 \times 25)/8 \approx 0.001$. En el Problema 3.1.1 para 72 grados (con seno $= 0.951056$) se obtuvo un valor del interpolante 0.950186; luego la cota es bastante fina. □

## 3.3.2 Diferencias divididas y derivadas

Para concluir la sección probaremos en (3.17), que salvo un factor, las diferencias divididas de las funciones suficientemente derivables son valores de derivadas de la función y que los polinomios de Taylor son límite de polinomios de Lagrange. Comencemos probando una identidad notable.

**Teorema 3.3.2** *Si $P$ es el polinomio de grado $\leq N$ que interpola a una función $f$ en $N+1$ nodos dos a dos distintos $x_0, \ldots, x_N$, entonces para cada $x$ que no coincida con ningún $x_i$ y para el que $f$ esté definida*

$$f(x) - P(x) = f[x_0, \ldots, x_N, x](x - x_0) \cdots (x - x_N). \tag{3.16}$$

*Demostración.* Elijamos un valor de $x$ para el cual deseemos probar la igualdad. Consideremos el polinomio $Q$ de grado $\leq N+1$ que interpola a $f$ en los $N+1$ nodos $x_i$ y en $x$. En este polinomio no podemos llamar $x$ a la variable, porque este símbolo ya lo estamos usando para otra cosa. Usemos la letra $t$ para denotar la variable. Entonces por la construcción de la sección precedente

$$Q(t) - P(t) = f[x_0, \ldots, x_N, x](t - x_0) \cdots (t - x_N);$$

basta ahora evaluar en $t = x$. □

**Corolario 3.3.2** *Si $N$ es un entero no negativo, sean $x_0, \ldots, x_{N+1}$, $N+2$ puntos, dos a dos distintos, del intervalo $[a, b]$. Si $f$ es una función con $N$ derivadas continuas en $[a, b]$ y $f^{(N+1)}$ existe en $(a, b)$ entonces hay un punto $\xi$ con*

$$\text{mín}(x_0, \ldots, x_{N+1}) < \xi < \text{máx}(x_0, \ldots, x_{N+1})$$

*para el que*

$$f[x_0, \ldots, x_{N+1}] = \frac{f^{(N+1)}(\xi)}{(N+1)!}. \tag{3.17}$$

*Demostración.* Basta combinar la identidad (3.16) (usada con $x_{N+1}$ en lugar de $x$) con la fórmula (3.13). □

**Corolario 3.3.3** *Sea $f$ una función con $N$ derivadas continuas en un intervalo y $P$ el polinomio interpolador de Lagrange en $N+1$ nodos $x_0, \ldots, x_N$, dos a dos distintos, del intervalo. Por otro lado sea $Q(x)$ el polinomio de Taylor de grado $\leq N$ de $f$ en $x_0$. Entonces, si se fijan $x_0$ y $x$, mientras que $x_1 \to x_0, \ldots, x_N \to x_0$, se verifica que $P(x) \to Q(x)$.*

*Demostración.* Basta escribir $P(x)$ en la forma de Newton (3.10) y aplicar el corolario precedente: Los coeficientes $f[x_0, \ldots, x_i]$ de la forma de Newton tienden a los $f^{(i)}(x_0)/i!$ del polinomio de Taylor, mientras que $(x - x_0) \cdots (x - x_{i-1}) \to (x - x_0)^i$. □

**Ejemplo 3.3.3** *El polinomio de Taylor como límite de polinomios de Lagrange.* La recta tangente en $(x_0, f(x_0))$ a la gráfica de $y = f(x)$ corresponde al polinomio de Taylor de grado $\leq 1$. Es límite, cuando $x_1 \to x_0$ de rectas secantes que interpolan lagrangianamente en $x_0$ y $x_1$. Por eso se decía clásicamente que la recta tangente y la gráfica de la función tienen en común *dos* puntos confundidos (en uno sólo). Se sigue diciendo que las ecuaciones $y = f(x)$, $y = f(x_0) + f'(x_0)(x - x_0)$ poseen una raíz *doble* en el punto $x = x_0$.

Análogamente la gráfica del polinomio de Taylor de grado $\leq 2$ en $x_0$ de la función $y = f(x)$ es una parábola, límite, cuando $x_1 \to x_0$ y $x_2 \to x_0$, de parábolas que interpolan lagrangianamente a la función en tres puntos $x_0, x_1, x_2$. La parábola de Taylor y la gráfica de la función tienen pues en común tres puntos confundidos en $x_0$. □

**Problema 3.3.1** Se va a tabular la función seno (con el argumento medido en grados) en abscisas equiespaciadas para luego efectuar interpolación lineal en la misma entre los nodos inmediatamente inferior y superior al ángulo deseado. Si se desean errores menores que $5 \times 10^{-6}$ ¿bastará tabular de grado en grado? ¿Y de medio grado en medio grado?

**Problema 3.3.2** Para $N = 0$ y $N = 1$ explicite el primer y segundo miembros de (3.16) y compruebe que coinciden.

**Problema 3.3.3** Para $N = 0$ el Corolario 3.3.2 reproduce un conocido teorema ¿cuál?

## 3.4 Convergencia de los polinomios de interpolación de Lagrange

Como hicimos para el problema de Taylor, nos planteamos ahora si $f(x) - P(x)$, error de interpolación, puede hacerse arbitrariamente pequeño incrementando el grado del polinomio. Más precisamente, dada una función $f$ definida en

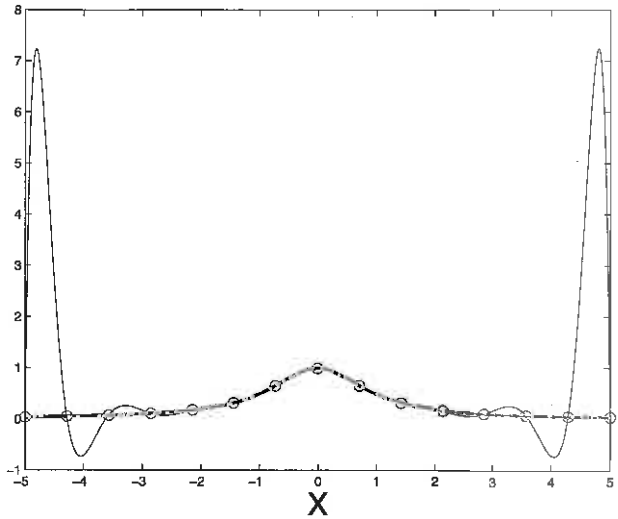


Figura 3.1: La función $1/(1+x^2)$ y el polinomio que la interpola en 15 puntos equiespaciados en $-5 \leq x \leq 5$

un intervalo $[a, b]$, elijamos un punto $x_0^{(0)}$ e interpolemos en él por un polinomio constante $P_0$, elijamos dos puntos distintos entre sí $x_0^{(1)}$, $x_1^{(1)}$ e interpolemos en ellos por una recta $P_1(x)$, elijamos tres puntos dos a dos distintos $x_0^{(2)}$, $x_1^{(2)}$, $x_2^{(2)}$ e interpolemos por una parábola cuadrática, etc. ¿Será cierto que $\lim_{N\to\infty} P_N(x) = f(x)$? La respuesta es, en general negativa, con la conclusión de que *incrementar el grado de los polinomios de interpolación no siempre es aconsejable.*

**Ejemplo 3.4.1** En 1900, Runge (1856–1927) demostró que si se interpola la función $1/(1+x^2)$ (que posee derivadas continuas de todos los órdenes) en $N+1$ abscisas equiespaciadas en el intervalo $[-5, 5]$ (esto es en $x_i^{(N)} = -5 + 10i/N$, $i = 0, \ldots, N$, $N = 1, 2, \ldots$) y se denota por $P_N$ el polinomio interpolador entonces, cuando $N \to \infty$, $P_N(x)$ no converge al valor $f(x)$ si $|x| > 3.6$. En la Fig. 3.1 apreciamos los grandes errores a que da origen $P_{14}$.[2] (Naturalmente la gráfica no prueba la divergencia de la sucesión de los $P_N$.) □

No siempre el resultado es tan pesimista: todo depende, según (3.14), de cómo crezcan las derivadas de la función al incrementarse el orden de la derivada.

[2]Este ejemplo parece más notable si se tiene en cuenta que K. Weierstrass (1815–1897) (director en 1880 de la tesis de Runge) había probado en 1885 el siguiente importante teorema de aproximación: Dado un intervalo acotado $a \leq x \leq b$, una función $f$ continua en él y un margen de error $\epsilon$ (arbitrariamente pequeño), existe un polinomio $Q$ que aproxima a $f$ con error menor que $\epsilon$, es decir $|f(x) - Q(x)| < \epsilon$ para cada $x$ en el intervalo. La gráfica de una función continua puede siempre aproximarse por la de un polinomio de grado suficientemente alto. El problema es que ese polinomio no se puede hallar interpolando $f$.

**Ejemplo 3.4.2** Consideremos la función $\cos x$ en un intervalo acotado $a \leq x \leq b$, interpolada en $N+1$ puntos dos a dos distintos cualesquiera del intervalo. La cota (3.14) muestra que

$$|f(x) - P_N(x)| \leq \frac{(b-a)^{N+1}}{(N+1)!},$$

donde el segundo miembro tiende a 0 cuando $N \to \infty$. Por ello los $P_N$ convergen a la función (uniformemente en $a \leq x \leq b$).

**Problema 3.4.1** Se interpola la función $e^x$ en $N+1$ puntos dos a dos distintos cualesquiera de un intervalo $[a, b]$. Pruebe que, al hacer $N \to \infty$, la sucesión de los polinomios de interpolación converge a $e^x$ para cada $x \in [a, b]$ y que la convergencia es uniforme.

**Problema 3.4.2** Escriba un programa que realice gráficas como la Fig. 3.1 para un valor cualquiera $N$ del grado del polinomio. Use la forma de Newton de los polinomios interpolantes y evalúe por un algoritmo tipo Horner.

# Capítulo 4

# Interpolación polinómica a trozos

## 4.1 Interpolación lineal a trozos

### 4.1.1 Funciones lineales a trozos

Antes de considerar un nuevo tipo de interpolación debemos estudiar una familia de funciones que usaremos. Note que hasta nuevo aviso no estamos tratando de interpolar ninguna función.

**Ejemplo 4.1.1** Al escribir este libro, la cuantía de la cuota que cada contribuyente ha de pagar en el Impuesto sobre la Renta de las Personas Físicas (IRPF) se fija según la siguiente tablilla

| Base | Cuota íntegra | % |
|---|---|---|
| | | 24 |
| 17707.20 | 4249.73 | 28 |
| 33007.20 | 8533.73 | 37 |
| 53407.20 | 16081.73 | 43 |

Se usa como sigue. A una persona cuyas rentas en el año fuesen de p. ej. 24000 euros le corresponden (a) 4249.73 euros de impuestos por sus primeros 17707.20 euros de renta y (b) el 28 % de los restantes 6292.80, esto es 1761.98; en total una cuota de 6011.71 euros. Así en el intervalo de rentas $[17707.20, 33007.20]$ la cuota como función de la renta es una recta de pendiente 0.28. Análogamente, en el intervalo $[33007.20, 53407.20]$ tenemos otra recta, ahora de pendiente 0.37, etc. La gráfica de la cuota de impuesto en función de la renta puede verse en la Fig. 4.1. Está hecha de segmentos de recta soldados; la derivada tiene saltos en 17707.20, 33007.20 y 53407.20; pero estos saltos no son demasiado aparentes. □

En general, consideremos un intervalo $[a, b]$ y una partición $\Delta$ del mismo

$$\Delta : a = x_0 < x_1 < x_2 < \ldots < x_N = b. \tag{4.1}$$

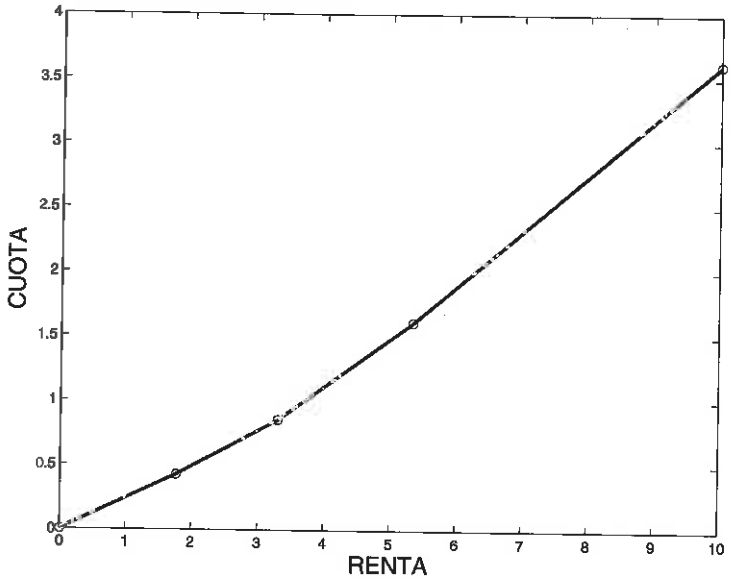


Figura 4.1: Cuota íntegra del IRPF en función de la renta del contribuyente (rentas hasta cien mil euros). Cada unidad en los ejes representa diez mil euros

Denotamos por $M_0^1(\Delta)$ al conjunto de todas las funciones $s$ continuas en $[a, b]$ que restringidas a cada intervalo $[x_{i-1}, x_i]$, $i = 1, \ldots, N$ de la partición coinciden con un polinomio de grado $\leq 1$ (polinomio que, en general, cambia con $i$ para una $s$ dada). Si $s \in M_0^1(\Delta)$ decimos que $s$ es una *función lineal a trozos* (en la partición $\Delta$). En los $x_i \in \Delta$ las funciones $s \in M_0^1(\Delta)$ presentan saltos en la derivada.

La manera más adecuada de determinar una función lineal a trozos $s$ en la partición (4.1) consiste en almacenar, para cada $i = 1, \ldots, N$, los dos coeficientes reales $\alpha_i$, $\beta_i$ tales que

$$\alpha_i + \beta_i(x - x_{i-1}), \qquad x \in [x_{i-1}, x_i] \tag{4.2}$$

sea la ecuación del $i$-ésimo segmento de los que componen $s$; note que $\alpha_i = s(x_{i-1})$ y que $\beta_i$ es la pendiente. Esta técnica de especificar $s$ a través de las $\alpha_i$ y $\beta_i$ es por cierto la que usa la Hacienda Pública al dar la tabla del IRPF.

Para evaluar $s$ en un valor dado de $x \notin \Delta$ primero hemos de encontrar el valor de $i$ para el que $x \in [x_{i-1}, x_i]$; después hemos de hallar el valor de (4.2) sumando a $\alpha_i = s(x_{i-1})$ la cantidad $\beta_i(x - x_{i-1})$; esto es precisamente lo que hicimos para hallar el impuesto correspondiente a una renta dada.

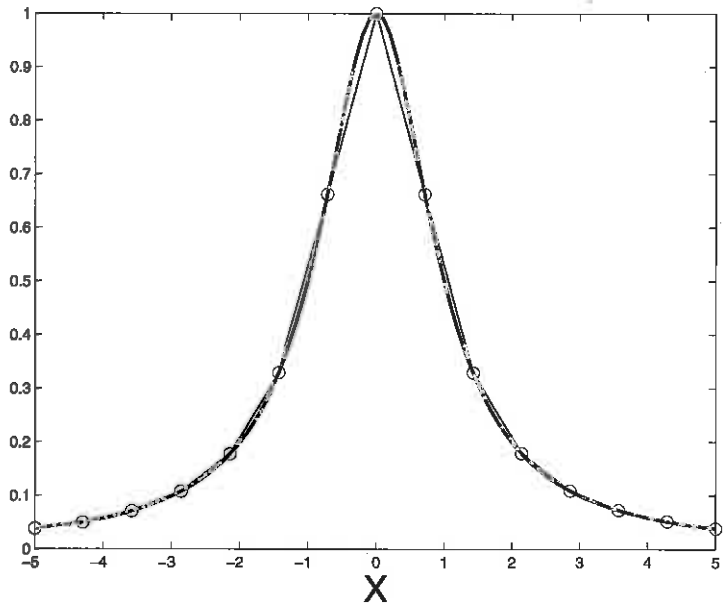


Figura 4.2: La función $1/(1+x^2)$ y su interpolante lineal a trozos en 15 puntos equidistantes en $-5 \leq x \leq 5$

### 4.1.2 Funciones lineales a trozos como interpolantes

Es obvio que, dada la partición (4.1) y dados los valores de una función $f(x_0)$, $\ldots$, $f(x_N)$ en los nodos de la misma, existe una única función lineal a trozos $s \in M_0^1(\Delta)$ tal que

$$s(x_0) = f(x_0), s(x_1) = f(x_1), \quad \ldots \quad , s(x_N) = f(x_N).$$

Esta $s$ se llama el interpolante lineal a trozos de $f$. La Fig. 4.2 muestra un ejemplo; los errores de interpolación son acusados cerca de $x = 0$ (justamente donde la derivada segunda de la función es mayor) pero aceptables en otras partes de la gráfica.

Para hallar el interpolante $s$ de una función $f$ dada hemos de encontrar los $\alpha_i$ y $\beta_i$ que aparecen en (4.2). Esto es inmediato, ya que $\alpha_i = f(x_{i-1})$, $\beta_i = [f(x_i) - f(x_{i-1})]/[x_i - x_{i-1}]$ (es decir $\beta_i = f[x_i, x_{i-1}]$).

Para el error de interpolación notemos que para $x \in [x_{i-1}, x_i]$, la función $s$ coincide con la recta de interpolación de Lagrange de $f$ en $x_{i-1}$, $x_i$. Por eso, usando (3.15), si $f$ tiene derivada segunda en $[a, b]$ acotada en valor absoluto por

$K_2$,

$$|f(x) - s(x)| \leq \frac{1}{8}(x_i - x_{i-1})^2 K_2, \quad x \in [x_{i-1}, x_i].$$

Si denotamos por $h$ a la mayor de las cantidades $x_i - x_{i-1}$, $i = 1, \ldots, N$, llamada *diámetro* de la partición, entonces

$$|f(x) - s(x)| \leq \frac{1}{8}h^2 K_2, \quad x \in [a, b]. \tag{4.3}$$

La cota (4.3) nos permite estudiar la cuestión de la *convergencia*. Consideremos en $[a, b]$ una sucesión $\Delta_0$, $\Delta_1$, $\Delta_2$, ..., de particiones, con correspondientes diámetros $h_0$, $h_1$, $h_2$, ... que tiendan a 0. Dada una función $f$ podemos considerar la correspondiente sucesión de funciones interpolantes a trozos $s_0$, $s_1$, $s_2$, ... De acuerdo con (4.3), tendremos, si $f$ tiene derivada segunda acotada, que, para $N \to \infty$, $s_N(x) \to f(x)$ uniformemente en $[a, b]$. Además, como la cota involucra el *cuadrado* de $h$, la convergencia se llama cuadrática. Al refinar una partición reduciendo su diámetro a la mitad, la cota del error de interpolación se divide entre cuatro.

### 4.1.3 Comparación con la interpolación de Lagrange polinómica

Dados la partición $\Delta$ en (4.1) y los correspondientes $N + 1$ valores de una función $f$, podemos interpolar ó por un polinomio $P$ de grado $\leq N$ ó por una función lineal $s$ a trozos. Vea las Figuras 3.1 y 4.2. Comparemos.

- Si $N$ es grande el costo de evaluar $P$ es grande. Sin embargo el costo de evaluar $s$ no crece con $N$.

- Como vimos en el Capítulo 3, no está en absoluto garantizado que al incrementar $N$ los polinomios $P$ se acerquen a $f$, ni aun suponiendo que $f$ es indefinidamente derivable. Sin embargo al refinar $\Delta$ sabemos que los correspondientes interpolantes lineales a trozos convergen a la función y lo hacen cuadráticamente, con sólo suponer que $f$ tiene derivada segunda acotada.

- El polinomio $P$ es indefinidamente derivable, mientras que $s$ no tiene ni siquiera derivada primera en los nodos. Esta falta de regularidad hace que los interpolantes lineales a trozos sean inaplicables en ciertas situaciones.

**Problema 4.1.1** Escriba las ecuaciones de los segmentos de recta que componen la gráfica de la Fig. 4.1.

**Problema 4.1.2** Para dar una función lineal a trozos en la partición (4.1) se van a prescribir los $N$ pares $(\alpha_i, \beta_i)$ ¿a cada elección de los $N$ pares le corresponde un $s$? ¿Qué $N-1$ ligaduras (ecuaciones) lineales (independientes) deben satisfacer los $2N$ coeficientes $\alpha_i$, $\beta_i$? Compruebe que esas ligaduras se satisfacen en el ejemplo del IRPF.

**Problema 4.1.3** *Una base para el espacio* $M_0^1(\Delta)$. Pruebe que $M_0^1(\Delta)$ es un espacio vectorial para las operaciones usuales. Pruebe que es de dimensión $N+1$. (Indicación: la transformación $s \mapsto (s(x_0), \ldots, s(x_N))$ es lineal, aplica $M_0^1(\Delta)$ en $\mathcal{R}^{N+1}$ y es biyectiva, luego $M_0^1(\Delta)$ y $\mathcal{R}^{N+1}$ tienen la misma dimensión.) Para $i = 0, \ldots, N$, denote por $L_i(x)$ la función lineal a trozos $\in M_0^1(\Delta)$ que toma el valor 1 en $x = x_i$ y se anula en los restantes nodos de la partición. Dibuje las gráficas de las $L_i$ y explique por qué algunos las llaman funciones 'sombrero'. Demuestre que las $N+1$ funciones $L_i$ constituyen una base de $M_0^1(\Delta)$. Demuestre que, dada una función $f$, su interpolante lineal a trozos es

$$f(x_0)L_0(x) + \cdots + f(x_N)L_N(x).$$

¿Qué le recuerda esta expresión?

**Problema 4.1.4** Pruebe que para la función en la Fig. 4.2 la derivada segunda es máxima en valor absoluto en el origen donde vale –2. Considerando particiones uniformes en $[-5, 5]$ ¿cómo hay que tomar $N$ para que al interpolar esta función linealmente a trozos los errores sean menores que 0.005? ¿Y menores que 0.000005?

**Problema 4.1.5** Escriba un programa que realice figuras como la 4.2 para particiones uniformes arbitrarias de $[-5, 5]$. Experimente con varios valores de diámetro.

## 4.2 Interpolación por polinomios a trozos de grado $> 1$

Además de las funciones 'hechas' de segmentos de recta podemos considerar funciones hechas de segmentos de parábolas cuadráticas. Denotamos por $M_0^2(\Delta)$ al conjunto de todas las funciones continuas en $[a, b]$ que restringidas a cada intervalo $[x_{i-1}, x_i]$, $i = 1, \ldots, N$ de la partición coinciden con un polinomio de grado $\leq 2$ (polinomio que en general cambia con $i$). Si $q \in M_0^2(\Delta)$ decimos que $s$ es una *función cuadrática a trozos* (en la partición $\Delta$). En los $x_i \in \Delta$ las funciones $s \in M_0^2(\Delta)$ presentan saltos en las derivadas primera y segunda.

Para determinar una función cuadrática a trozos en una partición dada hemos de dar para cada subintervalo $[x_{i-1}, x_i]$ los tres coeficientes de la ecuación de la

parábola con la que $q$ coincide en él. Para evaluar $q$ en un valor de $x$ hay que encontrar un subintervalo que lo contenga y luego evaluar la correspondiente parábola cuadrática.

¿Cómo usar las funciones cuadráticas a trozos para interpolar? Con sólo pedir

$$q(x_0) = f(x_0), q(x_1) = f(x_1), \ldots, q(x_N) = f(x_N) \tag{4.4}$$

no determinamos $q$ de modo único. En cada subintervalo $[x_{i-1}, x_i]$, (4.4) impone dos condiciones $q(x_{i-1}) = f(x_{i-1})$, $q(x_i) = f(x_i)$ a $q$, que en ese subintervalo es una parábola cuadrática que sabemos se determina por sus valores en *tres* puntos. Entonces hay que prescribir en cada subintervalo $[x_{i-1}, x_i]$ que $q$ coincida con $f$ en un punto adicional $x_i^* \in (x_{i-1}, x_i)$. Por razones de simetría se elige siempre $x_i^*$ igual al punto medio $(x_{i-1} + x_i)/2$ que denotaremos $x_{i+\frac{1}{2}}$. Así además de (4.4) imponemos

$$q(x_{\frac{1}{2}}) = f(x_{\frac{1}{2}}), q(x_{\frac{3}{2}}) = f(x_{\frac{3}{2}}), \ldots, q(x_{N-\frac{1}{2}}) = f(x_{N-\frac{1}{2}}). \tag{4.5}$$

Las condiciones (4.4)–(4.5) determinan únicamente un elemento de $M_0^2(\Delta)$ que se llama el *interpolante cuadrático a trozos de $f$*.

**Ejemplo 4.2.1** La Fig. 4.3 muestra la función $1/(1+x^2)$ en $[-5, 5]$ y un interpolante cuadrático a trozos. La partición $\Delta$ es uniforme y tiene siete subintervalos. Los puntos $x_i \in \Delta$ son (las abscisas de) los círculos; entre dos círculos consecutivos la gráfica del interpolante es un fragmento de parábola. Los puntos medios de los subintervalos son (las abscisas de) las estrellas; inmediatamente a la derecha de una estrella se tiene *la misma* parábola cuadrática que inmediatamente a la izquierda de la misma. El interpolante coincide con la función en las estrellas y los círculos, 15 puntos en total. Compare con las Fig. 3.1 y 4.2; en cada una de ellas el interpolante coincide con la función en 15 puntos y no obstante la aproximación es muy inferior. □

Si $f$ tiene derivada tercera acotada en $[a, b]$ puede demostrarse (Problema 4.2.4) que el error de interpolación tiene la siguiente cota

$$|f(x) - q(x)| \leq \frac{\sqrt{3}}{216} h^3 K_3, \quad x \in [a, b], \tag{4.6}$$

siendo $h$ el diámetro de la partición y $K_3$ una cota del valor absoluto de la derivada tercera. Al igual que en el caso lineal a trozos, se produce la *convergencia* de los interpolantes a la función al refinar la partición. La ventaja es que aquí la convergencia es *cúbica* y al dividir por dos el diámetro la cota se divide entre ocho; en (4.3) el exponente de $h$ era sólo dos.

Por otro lado la interpolación cuadrática a trozos *no subsana* uno de los inconvenientes de la interpolación lineal a trozos: los interpolantes cuadráticos a trozos son continuos en $[a, b]$ pero no son derivables en los puntos $x_i$ de $\Delta$.

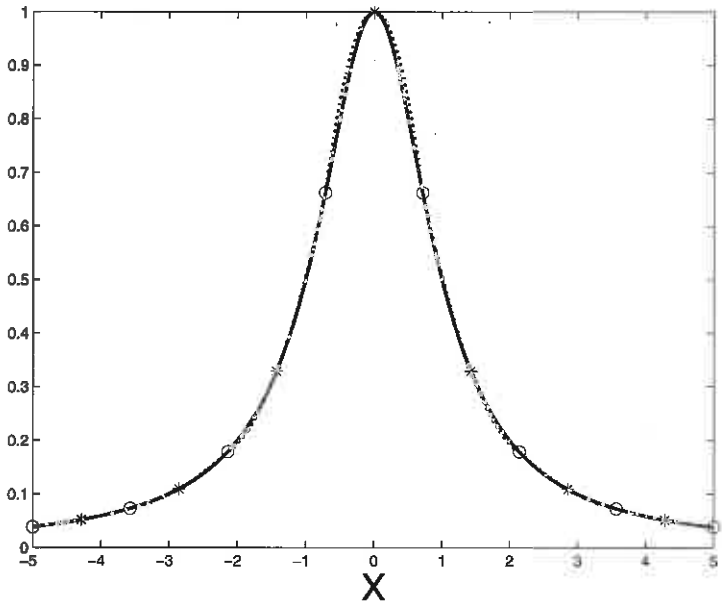


Figura 4.3: La función $1/(1+x^2)$ y su interpolante cuadrático a trozos (línea de puntos ) en una partición de $-5 \leq x \leq 5$ en siete subintervalos de la misma longitud

Es claro que se pueden considerar interpolantes cúbicos a trozos, cuárticos a trozos, quínticos a trozos, etc. para los que la convergencia va siendo como $h^4$, $h^5$, $h^6$, etc. El interpolante de grado $\nu$ a trozos interpola en los puntos de la partición y en los puntos que dividen los subintervalos en $\nu$ partes iguales. En la práctica pocas veces se pasa del caso cúbico a trozos y ninguna del quíntico a trozos, para evitar costos de evaluación y otras dificultades. Justamente *la enseñanza de este capítulo es que para interpolar en muchos puntos, en general, es mejor usar polinomios de grado bajo en particiones finas que un sólo polinomio de grado alto.*

**Problema 4.2.1** En la partición $x_0 = 0 < x_1 = \pi/2 < x_2 = \pi$, y para la función $\sin x$, halle los interpolantes lineal a trozos, cuadrático a trozos y cúbico a trozos.

**Problema 4.2.2** Dada una partición $\Delta$ con $N$ subintervalos, se eligen $N$ ecuaciones $y = a_i + b_i x + c_i x^2$ ¿existe necesariamente un elemento de $M_0^2(\Delta)$ cuya restricción al $i$-ésimo subintervalo sea, para cada $i$, $y = a_i + b_i x + c_i x^2$?

**Problema 4.2.3** *Una base para el espacio* $M_0^2(\Delta)$. Pruebe que $M_0^2(\Delta)$ es un espacio vectorial. Pruebe que es de dimensión $2N+1$. (Indicación: $q \mapsto (q(x_0), q(x_{1/2}), \ldots, q(x_N))$ aplica $M_0^2(\Delta)$ en $\mathcal{R}^{2N+1}$ es lineal y biyectiva, luego $M_0^2(\Delta)$ y $\mathcal{R}^{2N+1}$ tienen la misma dimensión.) Para

$i = 0, \ldots, N$ denote por $Q_i(x)$ a la función cuadrática a trozos $\in M_0^2(\Delta)$ que toma el valor 1 en $x = x_i$ y se anula en los restantes nodos de $\Delta$, así como en todos los puntos medios de los subintervalos de $\Delta$. Dibuje las gráficas de las $Q_i$ y explique por qué algunos las llaman funciones 'pagoda'. Para $i = 1, \ldots, N$ denote por $Q_{i-1/2}$ a la función cuadrática a trozos $\in M_0^2(\Delta)$ que toma el valor 1 en $x_{i-1/2}$ y se anula en cada punto medio $x_{j-1/2}$, $j \neq i$, así como en cada punto de $\Delta$. Dibuje las gráficas de las $Q_{i-1/2}$. Demuestre que las $2N+1$ funciones $Q$ que ha construido constituyen una base de $M_0^2(\Delta)$. Demuestre que, dada una función $f$, su interpolante cuadrático a trozos es

$$f(x_0)Q_0(x) + f(x_{1/2})Q_{1/2}(x) + f(x_1)Q_1(x) + \cdots + f(x_N)Q_N(x)$$

¿qué le recuerda esta expresión?

**Problema 4.2.4** Demuestre la cota (4.6). (Indicación: todo se reduce a hallar la expresión del error para la interpolación en $x_0 \leq x \leq x_1$ por un polinomio cuadrático con nodos $x_0$, $x_{1/2}$, $x_1$ —vea el Ejemplo 3.15— y luego aplicar esa expresión en cada subintervalo de la partición. Los cálculos le resultarán más sencillos si primero supone que $x_0 = -1$, $x_1 = 1$ —donde debe probar que el error está acotado por $\sqrt{3}K_3/27$— y luego aborda el caso general cambiando de variable independiente/usando análisis dimensional —vea la Sección 2.6 y los Problemas 3.1.13, 3.1.15.)

**Problema 4.2.5** Escriba un programa que realice figuras como la 4.3 para particiones uniformes arbitrarias de $[-5, 5]$. Halle en forma de Newton cada una de las parábolas cuadráticas que intervienen y evalúe por el algoritmo de Horner. Experimente con varios valores del diámetro.

# Capítulo 5

# Cuadratura numérica

## 5.1 Introducción

### 5.1.1 El problema de la cuadratura numérica

Estudiaremos en este capítulo métodos para el cálculo aproximado de integrales

$$\mathcal{I}(f) = \int_a^b f(x)\,dx, \tag{5.1}$$

siendo $f$ una función integrable en el intervalo acotado $[a, b]$. Por razones históricas el cálculo de integrales, o lo que es lo mismo el de áreas, se conoce en matemáticas con el nombre de *cuadratura,* reservándose el término *integración* para la resolución de ecuaciones diferenciales.[1]

¿Por qué usar cuadratura numérica en vez de calcular (5.1) por la regla de Barrow (1630–1677) (teorema fundamental del cálculo)?

- Tal vez no se disponga de la expresión analítica del integrando $f$, sino sólo de una tabla de valores de $f$ o de una subrutina que permita evaluar $f$ en cualquier punto $x$.

[1]Por ejemplo el problema de la cuadratura del círculo pedía construir con regla y compás un cuadrado cuya área coincida con la de un círculo dado (esto es muy diferente de la opinión vulgar que supone que la cuadratura del círculo es hallar razonamientos sofísticos que prueben que los círculos son cuadrados). De modo similar, clásicamente se concebía el problema de evaluar (5.1) como el de hallar un cuadrado cuya área coincidiese con la del cuadrilátero curvilíneo delimitado por el eje de abscisas, las rectas $x = a$, $x = b$ y la gráfica de $f$.

- Aunque se conozca la expresión analítica de $f$ y ésta sea una función elemental, la primitiva de $f$ quizá no sea elemental. Ya vimos que esto ocurre p. ej. si $f(x) = \exp(-x^2)$.

- Aunque la primitiva de $f$ sea elemental, encontrarla puede requerir alguna feliz idea (p. ej. un cambio de variables) con la que no seamos capaces de atinar.

- Aunque la primitiva sea elemental y obtenible sin recurso a ideas felices, encontrarla puede ser tedioso o precisar técnicas numéricas más complejas que las requeridas para hallar (5.1). Por ejemplo si $f$ es una función racional (cociente de dos polinomios) la primitiva se puede encontrar aplicando técnicas sistemáticas; por desgracia éstas requieren hallar todas las raíces del denominador, tarea que en general hay que realizar por aproximación numérica y es más difícil que aproximar (5.1) directamente.

### 5.1.2 Reglas de cuadratura

La mayor parte de los métodos numéricos aproximan (5.1) por una combinación lineal de valores de $f$

$$\mathcal{I}_{N+1}(f) = \alpha_0 f(x_0) + \cdots + \alpha_N f(x_N). \tag{5.2}$$

Elegidos y fijados $N \geq 0$, los $\alpha_i$ y los $x_i$, la expresión (5.2) asocia a cada función $f$ un número real y constituye lo que se llama una *regla de cuadratura*. Los $x_i$ se llaman *abscisas* o *nodos* de la regla; suele convenir que se hallen en $[a, b]$, pero no es imprescindible.[2] Supondremos, naturalmente, que los $x_i$ son dos a dos distintos. Los $\alpha_i$ se denominan *pesos* y también *coeficientes* de la regla.

Conviene retener de memoria las cuatro reglas siguientes.

*La regla de los rectángulos*

$$\mathcal{I}^R(f) = (b-a)f(a). \tag{5.3}$$

*La regla del punto medio*

$$\mathcal{I}^{PM}(f) = (b-a)f(c), \qquad c = \frac{a+b}{2}. \tag{5.4}$$

[2]Si algún nodo no pertenece a $[a, b]$ entonces la regla sólo es aplicable a funciones que además de en $[a, b]$ estén definidas en los nodos.

*La regla de los trapecios*

$$\mathcal{I}^T(f) = \frac{b-a}{2}f(a) + \frac{b-a}{2}f(b). \tag{5.5}$$

*La regla de Simpson (1710–1761)*

$$\mathcal{I}^S(f) = \frac{b-a}{6}f(a) + \frac{4(b-a)}{6}f(c) + \frac{b-a}{6}f(b), \qquad c = \frac{a+b}{2}. \tag{5.6}$$

### 5.1.3 Grado de exactitud

Decimos que una regla de cuadratura (5.2) tiene grado de exactitud $M \geq 0$ si halla exactamente la integral de cada polinomio de grado $\leq M$ (es decir, $\mathcal{I}(f) = \mathcal{I}_{N+1}(f)$ para cada $f$ polinomio de grado $\leq M$), pero no halla exactamente la integral de al menos un polinomio de grado $M+1$. Parece razonable suponer que es conveniente considerar reglas de grado de exactitud alto.

**Ejemplo 5.1.1** *Regla del rectángulo.* Para una constante $A$ la integral vale $A(b-a)$ y coincide con la aproximación (5.3); así el grado de exactitud es a lo menos 0. Para la recta $A + Bx$, el verdadero valor es $A(b-a)+B(b^2-a^2)/2$, la aproximación $(b-a)(A+Ba)$; como no coinciden para $B \neq 0$, el grado de exactitud es 0. □

**Ejemplo 5.1.2** *Regla del punto medio.* Para una constante $A$ la integral vale $A(b-a)$ y coincide con la aproximación dada por (5.4); así el grado de exactitud es a lo menos 0. Para la recta $A+Bx$, el verdadero valor es $A(b-a) + B(b^2-a^2)/2$, que también coincide con la aproximación $(b-a)\big(A+B(b+a)/2\big)$; el grado de exactitud es pues $\geq 1$. Para la parábola cuadrática $A+Bx+Cx^2$, la integral $A(b-a)+B(b^2-a^2)/2+C(b^3-a^3)/3$ y la aproximación $(b-a)\big(A+B(b+a)/2+C(b+a)^2/4\big)$ son distintas si $C \neq 0$. Por ello el grado de exactitud es 1. □

**Ejemplo 5.1.3** *Reglas de los trapecios y Simpson.* Es fácil probar (vea los problemas que siguen) que para las reglas de los trapecios y Simpson los grados de exactitud son 1 y 3 respectivamente. □

Con cada regla, además de memorizar los coeficientes y nodos, conviene acordarse del grado de exactitud.

**Problema 5.1.1** Aproxime $\int_{0.5}^{1.5} \cos x\, dx$ mediante las reglas del rectángulo, punto medio, trapecios y Simpson. (Solución: 0.87758, 0.54030, 0.47416, 0.51825; el verdadero valor es 0.51807, así que los errores son respectivamente –0.36, –0.022, 0.044, –0.00019.)

**Problema 5.1.2** Aproxime $\int_{0.95}^{1.05} \cos x\, dx$ por las mismas reglas. (Ahora resulta 0.058168, 0.054030, 0.053962, 0.0540077227; el verdadero valor es 0.0540077208, así que los errores son respectivamente −0.004 1, −0.000 022, 0.000 046, 0.000 000 001 9.)

**Problema 5.1.3** Consideremos reglas de cuadratura (5.2) que integren exactamente las funciones constantes. Demuestre que el grado de exactitud es $M$ si y solamente si la regla no da error al cuadrar las funciones $1, x, x^2, \ldots, x^M$ y da error no nulo al cuadrar $x^{M+1}$. Utilice este resultado para demostrar que el grado de exactitud de la regla de los trapecios es 1 y el de la regla de Simpson 3.

**Problema 5.1.4** Demuestre que en el problema anterior las funciones $x^k$ pueden sustituirse por las funciones $(x-\xi)^k$, donde $\xi$ es un número real cualquiera. La elección $\xi = (a+b)/2$ es particularmente útil, porque entonces la integral de $(x-\xi)^k$ es nula si $k$ es impar (entienda esto con una gráfica). Use este resultado con $\xi = (a+b)/2$ para comprobar de nuevo que el grado de exactitud de la regla de Simpson es 3.

## 5.2 Obtención de reglas de cuadratura

Supondremos ahora que en (5.2) hemos elegido $N$ y los nodos $x_i$ y tratemos de determinar de manera razonable los pesos $\alpha_i$. Para tal obtención vamos a presentar tres técnicas distintas, que se usan también en otras muchas áreas del análisis numérico. Conviene observar que *la calidad de un método no depende de la solidez del razonamiento que nos llevó a construirlo, sino de sus prestaciones prácticas y de los resultados matemáticos que sobre él podamos probar una vez construido.*

### 5.2.1 El método interpolatorio

Sustituyamos el integrando $f$ por el polinomio $P$ de grado $\leq N$ que coincide con él en los $N+1$ nodos $x_i$. Podremos tomar $\int_a^b P(x)\,dx$ como aproximación al verdadero valor (5.1), es decir *definir* una regla por

$$\mathcal{I}_{N+1}(f) = \mathcal{I}(P) = \int_a^b P(x)\,dx. \tag{5.7}$$

Si expresamos $P$ en la forma de Lagrange (3.4), tendremos

$$\mathcal{I}_{N+1}(f) = f(x_0)\int_a^b \ell_0(x)\,dx + \cdots + f(x_N)\int_a^b \ell_N(x)\,dx,$$

con lo que los pesos buscados son

$$\alpha_i = \int_a^b \ell_i(x)\,dx, \qquad i = 0, \ldots, N \tag{5.8}$$

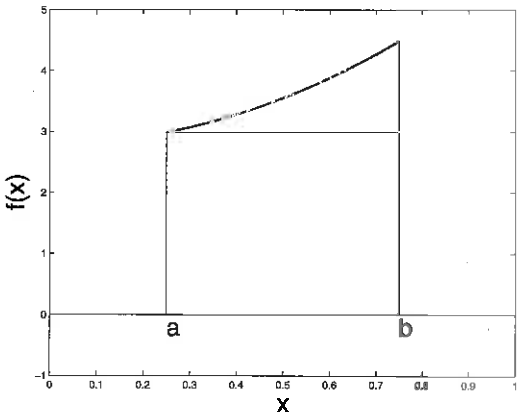


Figura 5.1: Regla del rectángulo

(note que el segundo miembro de (5.8) depende sólo de las abscisas y es independiente del integrando $f$). La regla (5.2) con los pesos (5.8) se llama *regla interpolatoria* correspondiente a las abscisas $x_i$. El siguiente teorema nos garantiza una cota inferior para el grado de las reglas interpolatorias.

**Teorema 5.2.1** *Dados $N \geq 0$ y $N+1$ abscisas $x_i$ dos a dos distintas, la correspondiente regla de cuadratura interpolatoria (5.2), (5.8) tiene grado de exactitud $\geq N$.*

*Demostración.* Si el integrando $f$ es un polinomio de grado $\leq N$ entonces su polinomio interpolador $P$ es $\equiv f$ y la igualdad (5.7) revela que $\mathcal{I}_{N+1}(f) = \mathcal{I}(f)$. □

**Ejemplo 5.2.1** *Regla del rectángulo.* Encontremos la regla interpolatoria con $N = 0$ basada en la abscisa $x_0 = a$. El interpolante es la constante $f(a)$, cuya integral vale

$$\int_a^b f(a)\,dx = (b-a)f(a),$$

con lo que la regla resulta ser la del rectángulo (5.3). La regla sustituye la verdadera integral, que es el área de un cuadrilátero curvilíneo, por el área del *rectángulo* delimitado por las rectas $x = a$, $x = b$, el eje de abscisas y la recta $y = f(a)$ (Figura 5.1). Establecimos anteriormente —y es evidente en la gráfica— que el grado de exactitud es 0. Así el grado coincide con la cota inferior $N$ suministrada por el Teorema 5.2.1. □

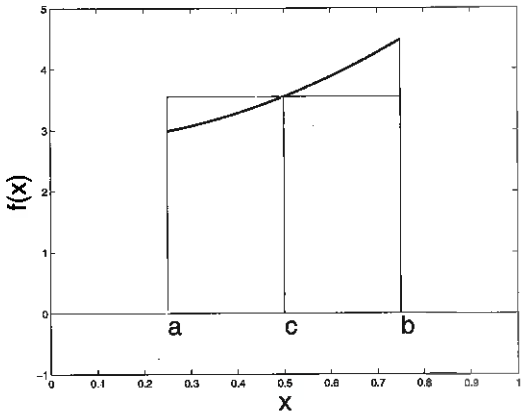


Figura 5.2: Regla del punto medio

**Ejemplo 5.2.2** *Regla del punto medio.* Encontremos la regla interpolatoria con $N = 0$ basada en la abscisa $x_0 = c = (a+b)/2$. El interpolante es la constante $f(c)$ cuya integral es

$$\int_a^b f(c)\,dx = (b-a)f(c),$$

y la regla resulta ser la del punto medio (5.4). La regla sustituye la verdadera integral por el área del *rectángulo* delimitado por las rectas $x = a$, $x = b$, el eje de abscisas y la recta $y = f(c)$ (Figura 5.2). Si el integrando es constante coincide con la recta $y = f(c)$ y la regla no da error; por tanto su grado de exactitud es $\geq 0$, cota dada por el teorema anterior. Si cuadramos una recta $y = f(x) = A + Bx$, $B \neq 0$, hay diferencia entre el interpolante $y = A + Bc$ y el verdadero integrando; sin embargo *las integrales* de ambos coinciden según se comprueba en la gráfica. Por ello, el grado de exactitud es $\geq 1$, a pesar de que la regla se construye interpolando por un polinomio constante. Demostramos en la sección precedente que, de hecho, el grado es exactamente 1; esto también se sigue inmediatamente de la intepretación gráfica. □

**Ejemplo 5.2.3** *Regla de los trapecios.* Supongamos ahora $N = 1$, $x_0 = a$, $x_1 = b$. Escriba la recta de interpolación de Lagrange e intégrela y concluya que la regla resultante no es otra que la regla del trapecio (5.5). Ahora el verdadero valor de la integral se sustituye por el área de un trapecio (Figura 5.3). La interpretación geométrica muestra que el grado de exactitud es 1. □

**Ejemplo 5.2.4** *Regla de Simpson.* Finalmente tomemos $N = 2$, $x_0 = a$, $x_1 = c = (a+b)/2$, $x_2 = b$. Escriba la parábola de interpolación de Lagrange e intégrela y concluya que la regla resultante no es otra que la regla de Simpson (5.6). En el Problema 5.1.3 vimos que el grado de exactitud es 3, una unidad más que la cota dada por el teorema: la regla se basa en interpolar por parábolas cuadráticas pero integra sin error incluso las parábola cúbicas. □

**Ejemplo 5.2.5** *Cambio de variable.* Consideremos el cambio de variable $x = \lambda X + \mu$, donde $\lambda > 0$ y $\mu$ son constantes. Sea $A \leq X \leq B$ el intervalo que se corresponde con $a \leq x \leq b$, esto

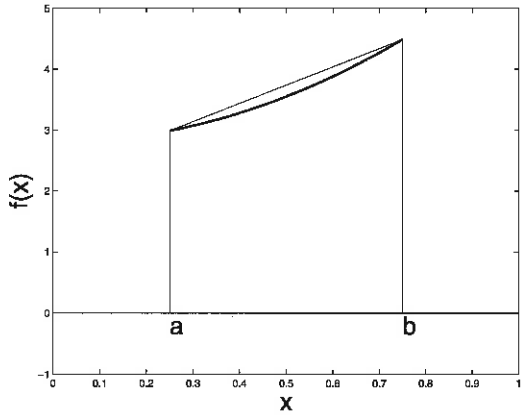


Figura 5.3: Regla del trapecio

es, $a = \lambda A + \mu$, $b = \lambda B + \mu$. A cada función $f(x)$ en $[a, b]$ le corresponde una nueva función $F(X) = f(\lambda X + \mu)$. Las integrales $\int_a^b f(x)dx$, $\int_A^B F(X)dX$ *no son iguales*, sino que se hallan en la proporción de $b - a$ a $B - A$.[3] De acuerdo con el Problema 3.1.13, el polinomio interpolador de $F(X)$ en los nodos $X_i$ es $p(\lambda X + \mu)$, donde $p(x)$ es el polinomio que interpola a $f(x)$ en los $x_i$. La regla interpolatoria en el intervalo $[a, b]$ con nodos $x_i$ y la regla interpolatoria en $[A, B]$ con los nodos $X_i$ correspondientes ($x_i = \lambda X_i + \mu$) tienen sus pesos en la proporción de $b - a$ a $B - A$, porque los valores de estas reglas son las integrales de $p(x)$ y $p(\lambda X + \mu)$ respectivamente.

Se puede usar esta observación para hallar los pesos, encontrándolos primero en un intervalo sencillo como $[0, 1]$ ó $[-1, 1]$. Así para los nodos $-1, 0, 1$ en el intervalo $-1 \leq X \leq 1$ se halla sin dificultad que los pesos de la regla interpolatoria son respectivamente $1/3$, $4/3$, $1/3$. Entonces para el intervalo $[a, b]$ la regla interpolatoria basada en los nodos $a$, $c$, $b$ (regla de Simpson) tendrá pesos $(b - a)/2 \times 1/3$, $(b - a)/2 \times 4/3$, $(b - a)/2 \times 1/3$. □

## 5.2.2 El método de coeficientes indeterminados

Se basa éste en notar que la regla (5.2) tiene grado de exactitud $\geq k$ si y solo si $\mathcal{I}_{N+1}(f) = \mathcal{I}(f)$ cuando $f(x) = 1, x, \ldots, x^k$, es decir

$$\begin{array}{ccccccccc} \alpha_0 & + & \alpha_1 & + & \cdots & + & \alpha_N & = & b - a \\ \alpha_0 x_0 & + & \alpha_1 x_1 & + & \cdots & + & \alpha_N x_N & = & (b^2 - a^2)/2 \\ \vdots & + & \vdots & + & \cdots & + & \vdots & = & \vdots \\ \alpha_0 x_0^k & + & \alpha_1 x_1^k & + & \cdots & + & \alpha_N x_N^k & = & (b^{k+1} - a^{k+1})/(k+1) \end{array} \tag{5.9}$$

[3]Lleve a cabo el cambio de variable en la integral, o mejor observe que las dos integrales son áreas de sendas figuras en que comparten las mismas alturas pero cuyas bases están en la proporción mencionada de $a - b$ a $A - B$, es decir de $\lambda$ a 1.

Esto es un sistema de $k+1$ ecuaciones para los $N+1$ coeficientes $\alpha_i$. Si pretendemos alcanzar grado $\geq N$ estaremos imponiendo tantas condiciones como coeficientes tenemos: además, como la matriz del sistema es de Vandermonde con abscisas dos a dos distintas, el sistema tendrá solución única (Problema 3.1.6). De este modo hemos probado el siguiente resultado:

**Teorema 5.2.2** *Dados $N \geq 0$ y $N+1$ abscisas $x_i$ dos a dos distintas, hay una única elección de pesos $\alpha_i$ para los cuales la regla (5.2) tenga grado de exactitud $\geq N$. Los pesos pueden hallarse resolviendo el sistema (5.9) con $k = N$.*

Combinando este aserto con el Teorema 5.2.1 tendremos:

**Teorema 5.2.3** *Dados $N \geq 0$ y $N+1$ abscisas $x_i$ dos a dos distintas, la correspondiente regla interpolatoria es la única con esas abscisas que tiene grado de exactitud $\geq N$. Los pesos pueden hallarse resolviendo el sistema (5.9) con $k = N$ o por la fórmula (5.8).*

**Ejemplo 5.2.6** Consideremos reglas con un sólo nodo $N = 0$. Con un coeficiente $\alpha_0$ por determinar pongamos $k = 0$ en (5.9), es decir, exijamos que la fórmula integre sin error las constantes. Obtenemos $\alpha_0 = b - a$. Así se hallan, en particular, la regla del rectángulo (5.3) y la del punto medio (5.4) sin pasar por los correspondientes interpolantes. Observe que, cuando $x_0 = (a+b)/2$ (regla del punto medio) también se satisface la segunda ecuación de (5.9) porque $\alpha_0 x_0 = (b-a)(a+b)/2 = (b^2 - a^2)/2$; eso ocurre porque la fórmula cuadra exactamente las rectas. □

### 5.2.3 El método del desarrollo de Taylor

**Ejemplo 5.2.7** Ilustremos éste obteniendo una regla con $N = 1$, $x_0 = a$, $x_1 = b$. Suponiendo que el integrando $f$ es suficientemente diferenciable, tendremos $f(b) = f(a) + f'(a)(b-a) + f''(a)(b-a)^2/2 + \cdots$, donde los puntos suspensivos quieren indicar que procederemos con el desarrollo de Taylor hasta un orden todavía sin determinar y añadiremos el término de error correspondiente. Sustituyendo en la regla $\mathcal{I}_2(f) = \alpha_0 f(a) + \alpha_1 f(b)$, tenemos

$$\mathcal{I}_2(f) = (\alpha_0 + \alpha_1) f(a) + \alpha_1 (b-a) f'(a) + \alpha_1 \frac{(b-a)^2}{2} f''(a) + \cdots. \tag{5.10}$$

Por otro lado, integrando término a término

$$\begin{aligned} \mathcal{I}(f) &= \int_a^b \left[ f(a) + f'(a)(x-a) + \frac{1}{2} f''(a)(x-a)^2 + \cdots \right] dx \\ &= (b-a) f(a) + \frac{(b-a)^2}{2} f'(a) + \frac{(b-a)^3}{6} f''(a) + \cdots. \end{aligned} \tag{5.11}$$

Ahora impongamos que los términos de los segundos miembros de (5.10) y (5.11) vayan coincidiendo. Obtenemos la jerarquía de ecuaciones

$$\begin{aligned} \alpha_0 + \alpha_1 &= b - a, \\ (b-a)\alpha_1 &= \frac{(b-a)^2}{2}, \\ \frac{(b-a)^2}{2}\alpha_1 &= \frac{(b-a)^3}{6}, \\ \cdots &= \cdots, \end{aligned}$$

donde tratamos de satisfacer el mayor número de ecuaciones *consecutivas* que nos sea posible empezando desde la *primera* (de ahí el nombre *jerarquía*). Claramente satisfaremos las dos primeras ecuaciones si tomamos $\alpha_1 = \alpha_0 = (b-a)/2$; estos son justamente los pesos de la regla de los trapecios (5.5). No se puede satisfacer también la tercera pues da un valor de $\alpha_1$ incompatible con el ya obtenido de la segunda.

¿Qué significan todas las manipulaciones que hemos hecho y por qué hemos llegado a los coeficientes de la regla de los trapecios? Supongamos que el integrando sea un polinomio de grado $\leq k$. Entonces efectuando los desarrollos de Taylor de $f$ hasta ese orden, el polinomio de Taylor coincide con el propio integrando y los segundos miembros de (5.10)–(5.11) tienen justamente $k+1$ sumandos, sin que haya términos de error. Por ello, las primeras $k+1$ ecuaciones de la jerarquía son necesarias y suficientes para garantizar que la regla y la integral coinciden para polinomios de grado $\leq k$, o, en otras palabras, que el grado de exactitud es $\geq k$. Al considerar dos ecuaciones de la jerarquía recaemos en la regla de los trapecios (única regla con esos nodos con grado $\geq 1$, Teorema 5.2.3). Y no se pueden satisfacer las tres primeras ecuaciones porque la regla de los trapecios no es exacta para polinomios cuadráticos. □

En general el método del desarrollo de Taylor pide:

- Desarrollar en Taylor las ordenadas $f(x_i)$. El punto en torno al cual se efectúa el desarrollo de Taylor puede ser uno de los nodos ($a$ en el ejemplo precedente) o cualquier otro punto que conduzca a simplificaciones en los cálculos; el punto medio $c = (a+b)/2$ del intervalo suele ser particularmente útil.

- Desarrollar en Taylor el integrando $f(x)$ en torno al mismo punto e integrar término a término el desarrollo.

- Imponer que los términos de ambos desarrollos vayan coincidiendo hasta donde sea posible.

**Problema 5.2.1** Obtenga los coeficientes de la regla de los trapecios y los de la regla de Simpson por el método de coeficientes indeterminados.

**Problema 5.2.2** En el método de coeficientes indeterminados se puede exigir grado de exactitud $\geq k$ imponiendo que la regla cuadre sin error las funciones $1, x-\xi, (x-\xi)^2, \ldots, (x-\xi)^k$, siendo $\xi$ cualquier número real. Escriba el sistema que reemplaza al (5.9). Obtenga los coeficientes de las reglas de los trapecios y Simpson usando esta idea con $\xi = a$. Luego repita con $\xi = c = (a+b)/2$. Compare la dificultad de los cálculos con la del problema anterior.

**Problema 5.2.3** Halle los coeficientes de las reglas del punto medio y Simpson mediante desarrollos de Taylor en torno al punto $c = (a+b)/2$.

**Problema 5.2.4** *Regla de los tres octavos.* Halle por el método interpolatorio la única regla de grado de exactitud $\geq 3$ que usa las abscisas equiespaciadas $a$, $(2a+b)/3$, $(a+2b)/3$, $b$. Vuelva a obtener esos coeficientes usando coeficientes indeterminados, imponiendo exactitud en funciones de la forma $(x-c)^k$, $c = (a+b)/2$. Obtenga una vez más los coeficientes, ahora por desarrollo de Taylor en torno a $c$. (La solución es $(b-a)/8$, $3(b-a)/8$, $3(b-a)/8$, $(b-a)/8$, y por eso la regla se llama de los tres octavos.) Halle el grado de exactitud (solución: 3).

**Problema 5.2.5** *La regla de los trapecios modificada.* Además de reglas de cuadratura de la forma (5.2) se pueden considerar otras que, por ejemplo, usen valores de la derivada de $f$. Buscamos una regla de la forma $\alpha_0 f(a) + \alpha_1 f(b) + \beta_0 f'(a) + \beta_1 f'(b)$. Determine los cuatro pesos para lograr que el grado de exactitud sea $\geq 3$; para ello use coeficientes indeterminados basados en las funciones $(x-c)^k$, $c = (a+b)/2$. Vuelva a determinar esos pesos ahora desarrollando en Taylor en torno a $c$. (La solución es $\alpha_0 = \alpha_1 = (b-a)/2$, $\beta_0 = -\beta_1 = (b-a)^2/12$.[4]) Demuestre que el grado de exactitud es justamente 3.

**Problema 5.2.6** Demuestre que la única regla que usa una sola abscisa y tiene grado de exactitud $\geq 1$ es la del punto medio. Concluya que con una sola abscisa no es posible alcanzar grado de exactitud $\geq 2$.

**Problema 5.2.7** *Regla Gaussiana con dos nodos.* En el intervalo $[-1, 1]$ halle por coeficientes indeterminados la regla de grado de exactitud máximo posible basada en las dos abscisas $\pm\sqrt{3}/3$. ¿Cuál es el grado de exactitud? (Solución: los pesos son ambos 1 y el grado es exactamente 3.) La regla obtenida se basa en un interpolante lineal y con todo integra exactamente las cúbicas.[5]

---

[4]Aquí los coeficientes $\beta$ no son proporcionales a la longitud del intervalo $b-a$ sino a su cuadrado. Esto no contradice lo establecido en el Ejemplo 5.2.5, porque ahora no trabajamos con una regla de la forma (5.2).

[5]Más generalmente, Gauss (1777–1855) demostró que para cada entero $N$ hay una única elección de $N+1$ abscisas $x_i$ para las cuales la correspondiente regla de cuadratura interpolatoria posee grado de exactitud $2N+1$. Esta regla se llama gaussiana y la determinación de las abscisas pasa por la teoría de los *polinomios ortogonales*. No hay ninguna regla con $N+1$ abscisas que alcance el grado de exactitud $2N+2$.

## 5.3 Error de cuadratura

¿Qué podremos decir del error $\mathcal{E}_{N+1}(f) = \mathcal{I}(f) - \mathcal{I}_{N+1}(f)$ que se comete al cambiar el verdadero valor de una integral por la aproximación obtenida por (5.2)?

Un estudio detallado de la obtención de cotas de error para las reglas de cuadratura se sale del nivel de este libro (con todo, los problemas de esta sección dan algunas ideas útiles). Para las reglas que hemos venido estudiando puede demostrarse:

*La regla de los rectángulos.* Si $f$ tiene derivada continua en $[a, b]$, entonces existe un punto $\eta \in [a, b]$ para el que

$$\mathcal{E}^R(f) = \frac{(b-a)^2}{2} f'(\eta). \tag{5.12}$$

*La regla del punto medio.* Si $f$ tiene *dos* derivadas continuas en $[a, b]$, entonces existe un punto $\eta \in [a, b]$ para el que

$$\mathcal{E}^{PM}(f) = \frac{(b-a)^3}{24} f''(\eta). \tag{5.13}$$

*La regla de los trapecios.* Si $f$ tiene *dos* derivadas continuas en $[a, b]$, entonces existe un punto $\eta \in [a, b]$ para el que

$$\mathcal{E}^T(f) = -\frac{(b-a)^3}{12} f''(\eta). \tag{5.14}$$

*La regla de Simpson.* Si $f$ tiene *cuatro* derivadas continuas en $[a, b]$, entonces existe un punto $\eta \in [a, b]$ para el que

$$\mathcal{E}^S(f) = -\frac{(b-a)^5}{2880} f^{(iv)}(\eta). \tag{5.15}$$

El punto $\eta$ naturalmente cambia al cambiar $f$ o al cambiar de regla.

De modo similar a nuestra discusión del error de interpolación, observe:

- El orden de la derivada $M + 1$ que aparece es una unidad superior al grado de exactitud $M$. Si el integrando es un polinomio de grado $\leq M$ entonces el error ha de ser cero, lo cual se logra usando la derivada $M+1$ del integrando.

- Con la derivada $M + 1$ aparece la longitud del intervalo elevada a $M + 2$. Esto se justifica por *análisis dimensional* como en el Ejemplo 2.6.2: las dimensiones de la integral son el producto de las de $f$ por las de $x$.

**Ejemplo 5.3.1** Para la integral $\int_{0.5}^{1.5} \cos x\, dx$ considerada en el Problema 5.1.1 los errores para las cuatro reglas son $-0.36$, $-0.022$, $0.044$, $-0.00019$. ¿Observa que el signo de los errores se podía haber previsto de las fórmulas (5.12)–(5.15)? Acotando el módulo de las derivadas de la función $\cos x$ por 1, las fórmulas dan las siguientes cotas para el valor absoluto del error $1/2 = 0.50$, $1/24 \approx 0.042$, $1/12 \approx 0.083$, $1/2880 \approx 0.00035$. □

**Ejemplo 5.3.2** Para la integral $\int_{0.95}^{1.05} \cos x\, dx$ considerada en el Problema 5.1.2, el error en la regla del rectángulo es $-0.0041$. La misma regla en la integral precedente daba error $-0.36$, unas 100 veces mayor —como predice la cota (5.12) ya que antes $b - a$ era 10 veces mayor que ahora. Para las reglas del punto medio y trapecios los errores ahora son $-0.000\,022$, $0.000\,046$, apreciándose una división entre 1000 respecto al ejemplo anterior. Para la regla de Simpson ahora el error es $0.000\,000\,001\,9$ y la división es entre $100\,000$. *Todas las reglas se comportan mejor en este ejemplo que en el precedente, pero las de grado de exactitud alto sacan más partido de la pequeñez del intervalo.* □

**Problema 5.3.1** *Obtención de la regla de Simpson a partir de la del punto medio y la de los trapecios.* Compare (5.13) y (5.14) para concluir que cuando $f$ es un polinomio de segundo grado, el valor absoluto del error en la regla de los trapecios es exactamente el doble del correspondiente a la regla del punto medio. Definiendo una nueva regla por $\mathcal{I}(f) = (2/3)\mathcal{I}^{\mathcal{PM}}(f) + (1/3)\mathcal{I}^{\mathcal{T}}(f)$, tendremos $\mathcal{E}(f) = (2/3)\mathcal{E}^{\mathcal{PM}}(f) + (1/3)\mathcal{E}^{\mathcal{T}}(f)$ y por tanto $\mathcal{E}(f) = 0$ para $f$ polinomio de grado $\leq 2$. Explicite los pesos y las abscisas de la nueva regla y comente.[6]

**Problema 5.3.2** *Un teorema de valor medio para integrales.* Sean $\alpha$ y $g$ funciones reales definidas en $[a, b]$ tales que las integrales $A = \int_a^b \alpha(x)\, dx \neq 0$, $B = \int_a^b g(x)\alpha(x)\, dx$ existan, $\alpha$ tome valores $\geq 0$ y $g$ sea continua. Pruebe que existe un punto $\eta \in [a, b]$ para el que

$$\int_a^b g(x)\alpha(x)\, dx = g(\eta) \int_a^b \alpha(x)\, dx.$$

(Indicación: pruebe que la cantidad $B/A$ queda comprendida entre los valores máximo y mínimo tomados por $g$.)

**Problema 5.3.3** *Error en la regla del rectángulo.* Vamos a demostrar en varias etapas (5.12), suponiendo que $f$ tiene derivada continua en $[a, b]$.

(i) Demuestre que

$$\mathcal{E}^R(f) = \int_a^b (f(x) - f(a))\, dx$$

(el error de cuadratura es la cuadratura del error de interpolación).

(ii) Demuestre que la función de $x$, $a < x \leq b$, dada por $g(x) = (f(x) - f(a))/(x - a)$ se puede definir también para $x = a$ de suerte que después de esa prolongación $g$ sea continua en $[a, b]$.

[6] Este problema ilustra cómo combinar linealmente los resultados de *dos* procesos numéricos para aproximar una misma cantidad a efectos de obtener una aproximación mejor que las que se combinan.

(iii) Pruebe que para cada $\eta \in [a, b]$ existe otro $\eta^* \in [a, b]$ tal que $g(\eta) = f'(\eta^*)$.

(iv) Use el apartado (i) y el Problema 5.3.2 para concluir que

$$\mathcal{E}^R(f) = g(\eta) \int_a^b (x - a)\, dx = \frac{(b-a)^2}{2} g(\eta).$$

(v) Use los apartados (iii) y (iv) para obtener el resultado (salvo un cambio de notación).

**Problema 5.3.4** *Error en la regla del punto medio.* Vamos a demostrar en varias etapas (5.13), suponiendo que $f$ tiene dos derivadas continuas en $[a, b]$.

(i) Demuestre que

$$\mathcal{E}^{PM}(f) = \int_a^b (f(x) - f(c))\, dx$$

(el error de cuadratura es la cuadratura del error de interpolación). Demuestre que también

$$\mathcal{E}^{PM}(f) = \int_a^b (f(x) - f(c) - f'(c)(x - c))\, dx.$$

(ii) Demuestre que la función de $x$, $x \in [a, b]$, $x \neq c$, dada por $g(x) = (f(x) - f(c) - f'(c)(x - c))/(x - c)^2$ se puede definir también para $x = c$ de suerte que después de esa prolongación $g$ sea continua en $[a, b]$.

(iii) Pruebe que para cada $\eta \in [a, b]$ existe otro $\eta^* \in [a, b]$ tal que $g(\eta) = f''(\eta^*)/2$ (recuerde la forma de Lagrange del error en la interpolación de Taylor y tenga especial cuidado en el caso $\eta = c$).

(iv) Use el apartado (i) y el Problema 5.3.2 para concluir que

$$\mathcal{E}^{PM}(f) = g(\eta) \int_a^b (x - c)^2\, dx = \frac{(b-a)^3}{12} g(\eta).$$

(v) Use los apartados (iii) y (iv) para obtener el resultado (salvo un cambio de notación).

**Problema 5.3.5** *Error en la regla del trapecio.* Vamos a demostrar en varias etapas (5.14), suponiendo que $f$ tiene dos derivadas continuas en $[a, b]$.

(i) Demuestre que si $P(x)$ es la recta que interpola a $f$ en $a$ y $b$, entonces

$$\mathcal{E}^T(f) = \int_a^b (f(x) - P(x))\, dx$$

(el error de cuadratura es la cuadratura del error de interpolación).

(ii) Demuestre que la función de $x$, $x \in (a, b)$, dada por $g(x) = (f(x) - P(x))/((x-a)(x-b))$ se puede definir también para $x = a$ y $x = b$ de suerte que tras esa prolongación $g$ sea continua en $[a, b]$.

(iii) Pruebe que para cada $\eta \in [a, b]$ existe otro $\eta^* \in [a, b]$ tal que $g(\eta) = f''(\eta^*)/2$ (si $\eta$ no es ni $a$ ni $b$ use la forma de Lagrange del error de interpolación, para $\eta = a$ ó $\eta = b$ puede razonar por continuidad).

(iv) Use el apartado (i) y el Problema 5.3.2 para concluir que

$$\mathcal{E}^T(f) = g(\eta) \int_a^b (x - a)(x - b)\, dx = -\frac{(b-a)^3}{6} g(\eta).$$

(v) Use los apartados (iii) y (iv) para obtener el resultado (salvo un cambio de notación).

## 5.4 Reglas compuestas

### 5.4.1 Concepto de regla compuesta

Las reglas con las que hemos venido trabajando hasta ahora (rectángulos, punto medio, trapecios, Simpson) son de utilidad limitada. Se observa de (5.12)–(5.15) que para que el error al usarlas sea aceptable es menester que la longitud $b - a$ del intervalo de integración sea pequeña relativamente al tamaño de las derivadas del integrando. ¿Qué haremos si, habiendo aproximado una integral por —digamos— la regla de Simpson, nos encontramos con un error que es demasiado grande para nuestras necesidades?

Una primera idea —que, por razones que luego expondremos, es poco útil— consiste en construir por uno de los métodos que vimos más arriba una nueva regla cuyo grado de exactitud sea más alto que el de la de Simpson.

Otra idea, mucho más fructífera, consiste en introducir una partición

$$\Delta : a = x_0 < x_1 < x_2 < \cdots < x_N = b,$$

escribir

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_1} f(x)\,dx + \cdots + \int_{x_{N-1}}^{x_N} f(x)\,dx \tag{5.16}$$

y aproximar cada una de las integrales en los subintervalos por la regla de Simpson. Así obtenemos una aproximación

$$\begin{aligned} \mathcal{I}^{SC}(f) = & \qquad (5.17) \\ & \frac{x_1 - x_0}{6} f(x_0) + \frac{4(x_1 - x_0)}{6} f(x_{1/2}) + \frac{x_1 - x_0}{6} f(x_1) \\ & + \cdots \\ & + \frac{x_N - x_{N-1}}{6} f(x_{N-1}) + \frac{4(x_N - x_{N-1})}{6} f(x_{N-\frac{1}{2}}) + \frac{x_N - x_{N-1}}{6} f(x_N) \end{aligned}$$

(con $x_{i-1/2} = (x_{i-1} + x_i)/2$, $i = 1, \ldots, N$) que se llama *regla de Simpson compuesta* relativa a la partición $\Delta$. Observemos que esta es una regla de la forma (5.2) cuyas abscisas son los puntos de la partición y los puntos medios de los subintervalos de la partición.[7]

Si en vez de aproximar las integrales en el segundo miembro de (5.16) por la regla de Simpson, lo hacemos por las reglas[8] del rectángulo, punto medio o

[7] ¿Cuáles es el peso correspondiente a $x_1$? ¿Y a $f(x_{3/2})$?

[8] También sería posible utilizar reglas distintas en subintervalos distintos, p. ej. la regla de los trapecios en ciertos subintervalos y la del punto medio en los restantes.

trapecios, obtenemos las reglas del *rectángulo compuesta, punto medio compuesta o trapecios compuesta* relativas a la partición $\Delta$

$$\begin{aligned} \mathcal{I}^{RC}(f) &= (x_1 - x_0)f(x_0) \\ &+ \cdots \\ &+ (x_N - x_{N-1})f(x_{N-1}), \end{aligned} \tag{5.18}$$

$$\begin{aligned} \mathcal{I}^{PMC}(f) &= (x_1 - x_0)f(x_{1/2}) \\ &+ \cdots \\ &+ (x_N - x_{N-1})f(x_{N-1/2}), \end{aligned} \tag{5.19}$$

$$\begin{aligned} \mathcal{I}^{TC}(f) &= \frac{x_1 - x_0}{2}f(x_0) + \frac{x_1 - x_0}{2}f(x_1) \\ &+ \cdots \\ &+ \frac{x_N - x_{N-1}}{2}f(x_{N-1}) + \frac{x_N - x_{N-1}}{2}f(x_N). \end{aligned} \tag{5.20}$$

### 5.4.2 Errores en las reglas compuestas

Cuando se usa una regla compuesta el error cometido al cuadrar la integral en $[a, b]$ es la suma de los errores que se cometen en cada una de las $N$ cuadraturas correspondientes a los subintervalos $[x_{i-1}, x_i]$. Por ejemplo para la regla del rectángulo compuesta, restando (5.18) de (5.16) tenemos

$$\begin{aligned} \mathcal{E}^{RC}(f) &= \mathcal{I}(f) - \mathcal{I}^{RC}(f) \\ &= \int_{x_0}^{x_1} f(x)\,dx - (x_1 - x_0)f(x_0) \\ &+ \cdots \\ &+ \int_{x_{N-1}}^{x_N} f(x)\,dx - (x_N - x_{N-1})f(x_{N-1}); \end{aligned}$$

cada línea es un error para la regla del rectángulo simple (es decir no compuesta). Usando la fórmula (5.12) podremos escribir para integrandos con derivada continua en $[a, b]$

$$\begin{aligned} \mathcal{E}^{RC}(f) &= \frac{(x_1 - x_0)^2}{2}f'(\eta_1) \\ &+ \cdots \\ &+ \frac{(x_N - x_{N-1})^2}{2}f'(\eta_N), \end{aligned}$$

siendo los $\eta_i$ puntos de $[x_{i-1}, x_i]$, $i = 1, \ldots, N$. Es fácil probar (Problema 5.4.3) que existirá $\eta \in [a, b]$ tal que

$$\mathcal{E}^{RC}(f) = \frac{1}{2}\left[(x_1 - x_0)^2 + \cdots + (x_N - x_{N-1})^2\right] f'(\eta).$$

Obtengamos una *cota* a partir de esta expresión. Para ello tomemos valores absolutos y mayoremos cada $(x_i - x_{i-1})^2$ por $h(x_i - x_{i-1})$, con $h = \text{máx}_i(x_i - x_{i-1})$, diámetro de la partición. Como

$$(x_1 - x_0) + \cdots + (x_N - x_{N-1}) = x_N - x_0 = b - a,$$

el resultado es

$$\left|\mathcal{E}^{RC}(f)\right| \leq \frac{1}{2}(b-a)hK_1, \tag{5.21}$$

con $K_1$ una cota del valor absoluto la derivada de $f$ en $[a, b]$. De aquí se concluye la *convergencia:* al hacer la partición más y más fina el error de cuadratura tiende a 0. Y esta convergencia es de *primer orden:* dividir por dos el diámetro de la partición divide por dos la cota del error.

Siguiendo los mismos pasos se tiene, para funciones con dos derivadas continuas en $[a, b]$,

$$|\mathcal{E}^{PMC}(f)| \leq \frac{1}{24}(b-a)h^2K_2, \tag{5.22}$$

$$|\mathcal{E}^{TC}(f)| \leq \frac{1}{12}(b-a)h^2K_2, \tag{5.23}$$

y, para funciones con cuatro derivadas continuas en $[a, b]$,

$$|\mathcal{E}^{SC}(f)| \leq \frac{1}{2880}(b-a)h^4K_4, \tag{5.24}$$

con $K_i$ una cota de la derivada $i$-ésima en $[a, b]$. Estas fórmulas muestran la convergencia de las respectivas reglas; convergencia que es cuadrática para la regla del punto medio y la de los trapecios y cuártica para la regla de Simpson. *¿Observa que en general la convergencia es de orden $M + 1$ para una regla compuesta basada en una simple de grado de exactitud $M$?*

**Ejemplo 5.4.1** Hemos calculado aproximadamente la integral

$$\int_0^1 \frac{\exp(-x^2/2)}{\sqrt{2\pi}}\, dx$$

cuyo integrando carece de primitiva elemental. El resultado verdadero, tomado de una tabla de la distribución normal, es 0.341 344 736 0. Hemos usado la regla de los trapecios en una partición

Tabla 5.1: Regla de los trapecios compuesta, integrando indefinidamente derivable

| $h$ | Numérico | Error | Error$/h^2$ |
|---|---|---|---|
| 1. | 0.320456 | $2.1 \times 10^{-2}$ | 0.02089 |
| 0.5 | 0.336261 | $5.1 \times 10^{-3}$ | 0.02034 |
| 0.25 | 0.340082 | $1.3 \times 10^{-3}$ | 0.02021 |
| 0.125 | 0.341030 | $3.2 \times 10^{-4}$ | 0.02017 |
| 0.0625 | 0.341266 | $7.9 \times 10^{-5}$ | 0.02016 |

Tabla 5.2: Regla de Simpson compuesta, integrando indefinidamente derivable

| $h$ | Numérico | Error | Error$/h^4$ |
|---|---|---|---|
| 1. | 0.341529051 | $1.8 \times 10^{-4}$ | $1.84 \times 10^{-4}$ |
| 0.5 | 0.341355487 | $1.0 \times 10^{-5}$ | $1.71 \times 10^{-4}$ |
| 0.25 | 0.341345406 | $6.6 \times 10^{-7}$ | $1.68 \times 10^{-4}$ |
| 0.125 | 0.341344787 | $4.1 \times 10^{-8}$ | $1.68 \times 10^{-4}$ |
| 0.0625 | 0.341344748 | $2.5 \times 10^{-9}$ | $1.68 \times 10^{-4}$ |

uniforme de diámetro $h$ y los resultados están dados en la Tabla 5.1. Observe cómo decrece el error al refinar la partición: al dividir entre dos el diámetro el error se divide aproximadamente entre cuatro (vea (5.23)). Esto se aprecia mejor en la última columna, que da el error dividido por el diámetro al cuadrado. □

**Ejemplo 5.4.2** Este ejemplo es como el anterior, con la diferencia de que ahora la regla es la de Simpson. Los resultados están en la Tabla 5.2. No sólo los errores, para cada $h$ dado, son menores, sino que la reducción del error con $h$ es ahora mucho más rápida que en la Tabla 5.1. Al dividir $h$ entre dos, el error lo hace aproximadamente por 16. ¿Era esto de esperar? □

**Ejemplo 5.4.3** *Es muy importante* notar que, en el ejemplo precedente, es *injusto* comparar los errores de los trapecios y Simpson correspondientes a un valor de $h$ dado. La regla de Simpson con $h = 0.0625$ precisa evaluar el integrando en 33 puntos; la de los trapecios con el mismo valor de $h$ sólo evalúa el integrando en 17 puntos, aproximadamente la mitad. *Los métodos númericos deben compararse por trabajo para un error deseado o por error para un trabajo dado.* Una manera ilustrativa de cotejar los resultados de las Tablas 5.1 y 5.2 es construir una gráfica de eficiencia como la Fig. 5.4. Esta es una representación doblemente logarítmica: en cada eje, las sucesivas potencias de diez aparecen a intervalos de longitud constante. En ordenadas (de $10^{-1}$ a $10^{-9}$) tenemos el error, en abscisas (de $10^0$ a $10^2$) el trabajo medido en número de evaluaciones del integrando. La línea superior corresponde a la regla de los trapecios compuesta, la inferior a

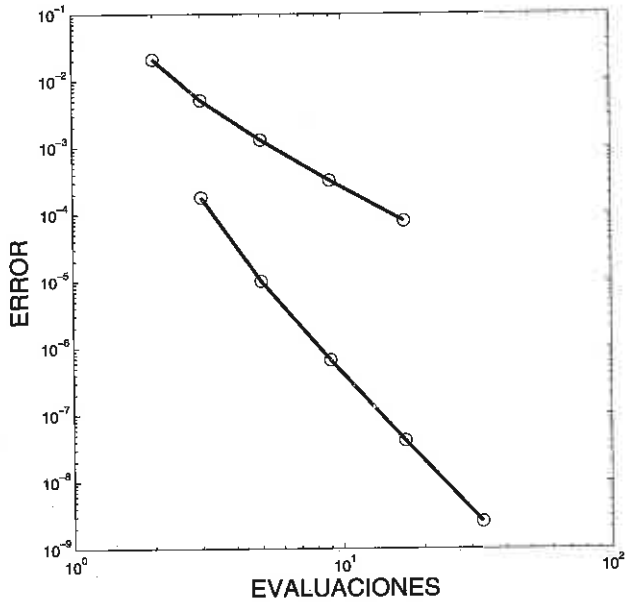


Figura 5.4: Comparación entre la eficiencia de las reglas de los trapecios y Simpson compuestas

la de Simpson compuesta. Las líneas son aproximadamente rectas, con pendientes una doble de otra ¿por qué? Para un trabajo dado, la regla de Simpson da error menor que la de los trapecios. Prolongando la línea de la regla de los trapecios, fórmese una idea del trabajo que se necesitaría con ella si se precisaran errores menores que $10^{-8}$. □

**Ejemplo 5.4.4** Este ejemplo es como el 5.4.1 (regla de los trapecios compuesta) con la única salvedad que ahora el integrando es $\sqrt{x}$ y el verdadero valor de la integral es 2/3. Los resultados se hallan en la Tabla 5.3. Los errores son mayores que en el Ejemplo 5.4.1 y no se dividen entre cuatro al dividir el diámetro entre dos. ¿Qué está fallando? Respuesta: el integrando no tiene dos derivadas continuas, y las hipótesis necesarias para escribir (5.23) no se verifican. □

Cuando se pueden estimar las derivadas del integrando las fórmulas (5.21)–(5.24) permiten decidir qué tamaño del diámetro hace falta para obtener errores menores que una tolerencia prescrita. Cuando las derivadas no se poseen o no son fáciles de estimar (el caso más común en la práctica), podemos cuadrar en varias particiones y examinar los resultados numéricos.

**Ejemplo 5.4.5** En la Tabla 5.1 observamos en la segunda columna que la aproximación numérica conserva la cifra de las milésimas al refinar de $h = 0.125$ a $h = 0.0625$; esto permite *suponer* que la integral es 0.341.

Tabla 5.3: Regla de los trapecios compuesta, integrando con derivada segunda singular

| $h$ | Numérico | Error | Error$/h^2$ |
|---|---|---|---|
| 1. | 0.500000 | $1.6 \times 10^{-1}$ | 0.16 |
| 0.5 | 0.603553 | $6.3 \times 10^{-2}$ | 0.25 |
| 0.25 | 0.643283 | $2.3 \times 10^{-2}$ | 0.37 |
| 0.125 | 0.663581 | $8.5 \times 10^{-3}$ | 0.54 |
| 0.0625 | 0.665559 | $3.1 \times 10^{-3}$ | 0.79 |

### 5.4.3 Relación con la interpolación polinómica a trozos

Al usar la regla de los trapecios compuesta sustituimos en cada subintervalo $[x_{i-1}, x_i]$ el verdadero integrando $f$ por el segmento de recta que coincide con él cuando $x = x_{i-1}$ y $x = x_i$. Por tanto estamos sustituyendo $f$ por su interpolante lineal a trozos $s \in M_0^1(\Delta)$ y tomando $\mathcal{I}(s)$ como aproximación a $\mathcal{I}(f)$. Análogamente la regla de Simpson compuesta toma como aproximación a $\mathcal{I}(f)$ la integral $\mathcal{I}(q)$ del interpolante cuadrático a trozos $q \in M_0^2(\Delta)$ de $f$. Como observamos en el capítulo precedente, al refinar la partición los interpolantes lineal a trozos y cuadrático a trozos tienen gráficas arbitrariamente próximas a la de $f$, y esto explica la convergencia de las reglas de los trapecios compuesta y Simpson compuesta. Vea las Figs. 4.2 y 4.3.

Las reglas de los rectángulos compuesta y punto medio compuesta también corresponden a tomar como aproximación a la integral la integral de un interpolante polinómico a trozos. Para ambas reglas el interpolante es una función constante a trozos discontinua en los puntos de la partición $\Delta$. Este tipo de interpolante no se estudió en el capítulo precedente.

En conclusión vemos que todas las reglas (5.2) que hemos estudiado se basan en tomar como aproximación a $\mathcal{I}(f)$ la integral exacta de una función que interpola a $f$. Para las reglas simples con $N + 1$ abscisas el interpolante es el polinomio interpolador de Lagrange de grado $\leq N$. Para las reglas compuestas el interpolante es una función polinómica a trozos.

¿Qué ocurre si en vez de acudir a la idea de regla compuesta usamos reglas (5.2) con $N + 1$ abscisas y grado de exactitud $\geq N$, con $N$ grande? Por el Teorema 5.2.3 las reglas en cuestión serían interpolatorias basadas en aproximar $f$ por un polinomio de grado alto. La Fig. 3.1 nos muestra que este modo de proceder no es muy aconsejable.

**Problema 5.4.1** Se va a calcular la integral de $\exp(-x^2)$ entre 0 y 1 por la regla del punto medio compuesta usando una partición uniforme. Se desean errores menores que $5 \times 10^{-7}$. ¿Cuántos subintervalos hay que usar? (Si toma 2 como cota superior de la derivada segunda la respuesta es 409.)

**Problema 5.4.2** Escriba un programa que le permita realizar tablas como las de esta sección para cada una de las reglas compuestas que hemos estudiado. Realice gráficas de eficiencia cuando se integran en $[0, 1]$ las funciones $\sin x$, $x^{1/2}$, $x^{3/2}$, $x^{5/2}$, $x^{7/2}$. Comente sus hallazgos.

**Problema 5.4.3** *Un teorema de valor medio para sumas.* Sea $g$ una función continua en el intervalo $[a, b]$ y consideremos números no negativos $\alpha_i$, no todos nulos, y puntos $\eta_i \in [a, b]$, $i = 1, \ldots, N$. Pruebe que existe $\eta \in [a, b]$ para el cual

$$\alpha_1 g(\eta_1) + \cdots + \alpha_N g(\eta_N) = g(\eta)\left[\alpha_1 + \cdots + \alpha_N\right].$$

(Indicación: vea el Problema 5.3.2, del cual el presente es una variante discreta.)

**Problema 5.4.4** *Particiones adaptadas.* ¿Por qué es útil considerar particiones no uniformes? Si se va a hallar el área bajo la curva de Fig. 4.2 mediante la regla de los trapecios compuestos usando un número fijado de puntos en la partición ¿es mejor una partición uniforme u otra que tenga intervalos grandes cerca del extremos del intervalo y pequeños hacia el centro? Las particiones que son finas donde la función que se estudia varía mucho y groseras donde varía poco se llaman *adaptadas*. La adaptación se puede hacer ó por el usuario analizando la función ó automáticamente por el programa que realice la cuadratura. En la segunda opción, el programa muestrea los valores de $f$ para identificar las zonas de variación rápida.

**Problema 5.4.5** *La extrapolación al límite de Richardson (1881–1953).*[9] En la Tabla 5.1 ya observamos que el resultado $\mathcal{I}_h$ de usar la regla de los trapecios compuesta en una partición uniforme de diámetro $h$ tiene un comportamiento de la forma $\mathcal{I}_h = \mathcal{I} + ch^2 + o(h^2)$, donde $\mathcal{I}$ es el verdadero valor de la integral y $c$ una constante (que depende del intervalo en que se trabaja y de la función que se cuadra). Por tanto la *nueva aproximación* a $\mathcal{I}$ definida por $\mathcal{I}_h^{\mathcal{R}} = (4/3)\mathcal{I}_{h/2} - (1/3)\mathcal{I}_h$ satisface $\mathcal{I}_h^{\mathcal{R}} = \mathcal{I} + (4/3)c(h/2)^2 - (1/3)ch^2 + o(h^2) = \mathcal{I} + o(h^2)$, valor más cercano a $\mathcal{I}$ de los $\mathcal{I}_h$, $\mathcal{I}_{h/2}$ que se combinan. (Por ejemplo en la tabla mencionada es $\mathcal{I}_1^{\mathcal{R}} = (4/3) \times 0.336262 - (1/3) \times 0.320456 = 0.341529$ con error $1.8 \times 10^{-4}$.) Se dice que $\mathcal{I}_h^{\mathcal{R}}$ es la aproximación obtenida por *extrapolación al límite*. Hállela para $h = 1/2$, $h = 1/4$ en la Tabla 5.1. Compare las aproximaciones obtenidas con la aproximación 0.341266, la mejor de las que figuran en la tabla.

**Problema 5.4.6** *La extrapolación al límite de Richardson (continuación).* En la Tabla 5.2 los errores se comportan como $ch^4 + o(h^4)$ y no como $ch^2 + o(h^2)$. ¿Con qué coeficientes hemos de combinar $\mathcal{I}_h$, $\mathcal{I}_{h/2}$ para obtener una aproximación con error $o(h^4)$? Obtenidos esos coeficientes úselos para extrapolar las aproximaciones $\mathcal{I}_1$, $\mathcal{I}_{1/2}$, dadas en la tabla.

[9] Richardson fue la primera persona en propugnar (ya en 1922) el uso, hoy tan familiar, del cálculo numérico para predecir el tiempo meteorológico.

# Capítulo 6

# Derivación numérica

## 6.1 Introducción

Nos ocuparemos ahora de aproximar el valor de la derivada

$$\mathcal{D}(f) = f'(x^*) \tag{6.1}$$

de una función en un punto fijado $x^*$ o también el valor de la derivada segunda

$$\mathcal{D}^2(f) = f''(x^*) \tag{6.2}$$

o de las derivadas de orden superior.

Las razones que llevan a la introducción de fórmulas de derivación numérica son análogas a las que conducen a la cuadratura numérica: quizá no se conozca la expresión analítica de $f$ o ella sea muy complicada.[1]

Para aproximar (6.1) consideraremos fórmulas del tipo

$$\mathcal{D}_{N+1}(f) = \alpha_0 f(x_0) + \cdots + \alpha_N f(x_N). \tag{6.3}$$

Fijados $x^*$, $N$, los *nodos* o *abscisas* $x_i$ (dos a dos distintos), y los *coeficientes o pesos* $\alpha_i$, la fórmula (6.3) asocia a cada $f$ un número real. Para (6.2) consideraremos análogamente

$$\mathcal{D}^2_{N+1}(f) = \beta_0 f(x_0) + \cdots + \beta_N f(x_N). \tag{6.4}$$

Diremos que una regla es de *grado de exactitud* $M$ si no da error cuando $f$ es un polinomio de grado $\leq M$ y da error en algún polinomio de grado $M+1$.

[1]Por otro lado, las fórmulas de derivación numérica son esenciales para construir métodos numéricos para ecuaciones diferenciales; éstos se emplean cada día en multitud de aplicaciones como predicción meteorológica, diseño de vehículos, prospección petrolífera, etc.

**Problema 6.1.1** Consideremos fórmulas (6.3) que den 0 al aplicarse a las funciones constantes. Demuestre que el grado de exactitud es $M$ si y solamente si la fórmula no da error para las funciones $1, x, x^2, \ldots, x^M$ y da error no nulo para $x^{M+1}$.

**Problema 6.1.2** Demuestre que en el problema anterior las funciones $x^k$ pueden sustituirse por las funciones $(x-\xi)^k$, donde $\xi$ es un número real cualquiera.

## 6.2 Obtención de fórmulas

Supondremos ahora que en (6.3) ó (6.4) hemos elegido $N$ y los nodos $x_i$ y tratemos de determinar de manera razonable los coeficientes $\alpha_i$, ó $\beta_i$. Esto puede efectuarse por los métodos que presentamos al hablar de la cuadratura numérica: interpolatorio, coeficientes indeterminados y desarrollo de Taylor.

**Ejemplo 6.2.1** *Aproximación a la derivada primera usando un nodo.* Supongamos que en (6.3) $N = 0$. Si usamos el método interpolatorio, el interpolante es constante y la aproximación a la derivada es nula para cada $f$. La fórmula carece de interés. □

**Ejemplo 6.2.2** *Aproximación a la derivada primera usando dos nodos.* Usemos el método interpolatorio escribiendo el interpolante en forma de Newton $P(x) = f(x_0) + f[x_0, x_1](x - x_0)$, con lo que la fórmula buscada es

$$\mathcal{D}_2(f) = \mathcal{D}_2(P) = P'(x^*) = f[x_0, x_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}. \tag{6.5}$$

Note que los coeficientes de la fórmula $\pm 1/(x_1 - x_0)$ han resultado ser independientes del punto $x^*$ en el que se busca la derivada. Por la forma en que ha sido encontrada, la fórmula no da error cuando $f$ es una recta y por ello su grado de exactitud es al menos 1. Para la función $x^2$ el verdadero valor es $2x^*$ y la aproximación $(x_1^2 - x_0^2)/(x_1 - x_0) = x_1 + x_0$. Por consiguiente el grado de exactitud es 1 siempre que no acontezca de que el punto $x^*$ donde deseamos hallar la derivada sea el punto medio de $x_0$ y $x_1$. Cuando $x^* = (x_1 + x_0)/2$ el grado de exactitud es 2, según se comprueba de inmediato. □

**Ejemplo 6.2.3** *La diferencia progresiva.* En la situación del ejemplo anterior supongamos que $x^*$ coincide con $x_0$ y que $x_0 < x_1$. La fórmula (6.5) es entonces

$$f'(x^*) \approx \frac{f(x^* + h) - f(x^*)}{h}, \tag{6.6}$$

con $h = x_1 - x_0 > 0$ y se llama *progresiva* ¿por qué? Su grado de exactitud es 1, según vimos. □

**Ejemplo 6.2.4** *La diferencia central.* En la situación del Ejemplo 6.2.2 supongamos que $x^*$ coincide con el punto medio $(x_1 + x_0)/2$. La fórmula (6.5) es entonces

$$f'(x^*) \approx \frac{f(x^* + h) - f(x^* - h)}{2h}, \qquad h = \frac{x_1 - x_0}{2}. \tag{6.7}$$

Su grado de exactitud es 2, según discutimos arriba. □

**Ejemplo 6.2.5** *Aproximación a la derivada segunda usando tres nodos.* Con sólo dos nodos hay que aproximar la derivada segunda por 0, sea cual sea $f$. Supongamos entonces que disponemos de tres nodos. Usemos el método interpolatorio escribiendo el interpolante en forma de Newton $P(x) = f(x_0) + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1)$, con lo que la fórmula buscada es

$$\mathcal{D}_3^2(f) = \mathcal{D}^2(P) = P''(x^*) = 2f[x_0, x_1, x_2] = \frac{\frac{f(x_2)-f(x_1)}{x_2-x_1} - \frac{f(x_1)-f(x_0)}{x_1-x_0}}{\frac{x_2-x_0}{2}}. \tag{6.8}$$

Los coeficientes de la fórmula han resultado ser independientes del punto $x^*$ en el que se busca la derivada segunda. El hecho de que el doble de una diferencia dividida segunda aproxime a la derivada segunda ya nos es conocido, recuerde el Corolario 3.3.2.

Por el modo en que ha sido encontrada, la fórmula no da error cuando $f$ es una parábola cuadrática y por ello su grado de exactitud es al menos 2. Para la función $x^3$ el verdadero valor es $6x^*$ y la aproximación (efectúe los cálculos) $2(x_0 + x_1 + x_2)$. Por consiguiente el grado de exactitud es 2 siempre que no suceda que el punto $x^*$ donde deseamos hallar la derivada sea la media aritmética de $x_0$, $x_1$, $x_2$. Cuando $x^* = (x_0 + x_1 + x_2)/3$ el grado de exactitud es 3, según se comprueba de inmediato. □

**Ejemplo 6.2.6** *La fórmula central para la derivada segunda.* Se desea aproximar $f''(x^*)$ mediante una combinación lineal de los valores $f(x^* + h)$, $f(x^*)$, $f(x^* - h)$, $h \neq 0$. La fórmula (6.8) se escribe ahora

$$f''(x^*) \approx \frac{f(x^* + h) - 2f(x^*) + f(x^* - h)}{h^2} \tag{6.9}$$

y se llama fórmula de diferencias centrales. Según dijimos arriba (¿dónde?) su grado es tres. □

**Ejemplo 6.2.7** *Cambio de variable.* Con las notaciones del Ejemplo 5.2.5, sean $x^*$ y $X^*$ puntos que se corresponden en el cambio. Se verifica $F'(X^*) = \lambda f'(x^*)$. Razonando como en aquel ejemplo, se prueba que los pesos de la regla interpolatoria con nodos $x_i$ para aproximar $f'(x^*)$ se obtienen dividiendo entre $\lambda$ los pesos de la regla interpolatoria para aproximar $F'(X^*)$ con nodos $X_i$. □

**Problema 6.2.1** *La fórmula central de cinco puntos para la derivada primera.* Se va a aproximar $f'(x^*)$ mediante una combinación lineal de $f(x^* + 2h)$, $f(x^* + h)$, $f(x^*)$, $f(x^* - h)$, $f(x^* - 2h)$. Utilice las técnicas interpolatoria, de coeficientes indeterminados y Taylor para encontrar los coeficientes. ¿Cuál es el grado de precisión? (Solución: $-1/(12h)$, $8/(12h)$, 0, $-8/(12h)$, $1/(12h)$; cuatro.)

**Problema 6.2.2** *La fórmula de diferenciación regresiva de tres puntos.* Se desea aproximar $f'(x^*)$ por una combinación lineal de $f(x^* - 2h)$, $f(x^* - h)$, $f(x^*)$. Use el método del desarrollo de Taylor (desarrollando en torno a $x^*$), para hallar los coeficientes $1/(2h)$, $-2/h$, $3/(2h)$. ¿Por qué se llama regresiva esta fórmula? Pruebe que su grado de exactitud es 2. Vuelva a obtener los coeficientes, esta vez por el método interpolatorio.

**Problema 6.2.3** *La fórmula central de cinco puntos para la derivada tercera.* Se desea aproximar $f'''(x^*)$ mediante una combinación lineal de $f(x^* + 2h)$, $f(x^* + h)$, $f(x^*)$, $f(x^* - h)$, $f(x^* - 2h)$. Utilice las técnicas interpolatoria, de coeficientes indeterminados y Taylor para encontrar los coeficientes. ¿Cuál es el grado de exactitud? (Solución: $1/(2h^3)$, $-2/(2h^3)$, 0, $2/(2h^3)$, $-1/(2h^3)$, cuatro.)

**Problema 6.2.4** Sea $h > 0$; se desea aproximar el valor $f''(x^*)$ por una combinación lineal de los cuatro valores $f(x^* \pm h)$, $f'(x^* \pm h)$. Utilice el método del desarrollo de Taylor para encontrar los coeficientes 0, 0, $\pm 1/(2h)$.

**Problema 6.2.5** Demuestre que cuando se obtiene la fórmula (6.3) por el método interpolatorio los coeficientes vienen dados por $\alpha_i = \ell_i'(x^*)$, siendo los $\ell_i(x)$ los polinomios de base de Lagrange. Pruebe que una fórmula interpolatoria con $N+1$ abscisas tiene grado de exactitud $\geq N$. Extienda estos resultados a las fórmulas (6.4) para aproximar la derivada segunda.

**Problema 6.2.6** Usando el método de coeficientes indeterminados pruebe que hay una única elección de coeficientes en la fórmula (6.3) que haga el grado de exactitud $\geq N$. Concluya del problema anterior que esos coeficientes son justamente los de la fórmula interpolatoria. Extienda estos resultados a las fórmulas (6.4) para aproximar la derivada segunda.

## 6.3 Error de la derivación numérica

Presentaremos algunos ejemplos.

**Ejemplo 6.3.1** *Error en la diferenciación progresiva.* El error para (6.6) será

$$\mathcal{E}^{PROG}(f) = f'(x^*) - \frac{f(x^* + h) - f(x^*)}{h}.$$

Desarrollando en Taylor en torno a $x^*$,

$$f(x^* + h) = f(x^*) + hf'(x^*) + \frac{h^2}{2} f''(\xi)$$

(hay que suponer que la derivada primera es continua en $[x^*, x^* + h]$ y la segunda existe en $(x^*, x^* + h)$) y de ahí

$$\mathcal{E}^{PROG}(f) = -\frac{h}{2} f''(\xi).$$

Si $K_2$ es una cota del valor absoluto de la derivada segunda,

$$|\mathcal{E}^{PROG}(f)| \leq \frac{h}{2} K_2,$$

y tenemos *convergencia* de primer orden: tomando $h$ suficientemente pequeño, la fórmula de derivación numérica da errores tan pequeños como se desee. □

Tabla 6.1: Derivación numérica por diferencias centrales

| $h$ | $\exp(h)$ | $\exp(-h)$ | Aproximación |
|---|---|---|---|
| 1 | 2.718282 | 0.3678795 | 1.175201 |
| $10^{-1}$ | 1.105171 | 0.9048374 | 1.001668 |
| $10^{-2}$ | 1.010050 | 0.9900498 | 1.000017 |
| $10^{-3}$ | 1.001001 | 0.9990005 | 1.000017 |
| $10^{-4}$ | 1.000100 | 0.9999000 | 1.000166 |
| $10^{-5}$ | 1.000010 | 0.9999900 | 1.001358 |

**Ejemplo 6.3.2** *Error en la diferenciación central.* El error en (6.7) será

$$\mathcal{E}^{CENT}(f) = f'(x^*) - \frac{f(x^*+h) - f(x^*-h)}{2h}.$$

Desarrollando en Taylor en torno a $x^*$,

$$f(x^* \pm h) = f(x^*) \pm hf'(x^*) + \frac{h^2}{2}f''(x^*) \pm \frac{h^3}{6}f'''(\xi_\pm),$$

donde $\xi_+$ y $\xi_-$ son puntos intermedios (hay que suponer que la derivada segunda es continua en $[x^* - h, x^* + h]$ y la tercera existe en $(x^* - h, x^* + h)$) y de ahí

$$\mathcal{E}^{CENT}(f) = -\frac{h^2}{12}\left[f'''(\xi_+) + f'''(\xi_-)\right]. \tag{6.10}$$

Si $K_3$ es una cota del valor absoluto de la derivada tercera, tenemos *convergencia* de segundo orden

$$|\mathcal{E}^{CENT}(f)| \leq \frac{h^2}{6}K_3. \quad \square$$

¿Observa en los ejemplos anteriores que el orden de la convergencia resulta ser igual al grado de exactitud? Con grado $M$ esperamos una derivada $M+1$ de $f$ en la expresión del error (¿por qué?) y tal derivada debe ir acompañada de un factor $h^M$ si hemos de tener dimensiones correctas. (Para la aproximación de derivadas segundas el orden de convergencia será el grado de exactitud menos uno.)

**Ejemplo 6.3.3** Hemos usado en una calculadora la fórmula del ejemplo precedente con $x^* = 0$ y varios valores de $h$ y $f(x) = \exp(x)$. Naturalmente el valor verdadero es 1. Los resultados están en la Tabla 6.1. En la última columna apreciamos cómo el error se divide entre 100 al dividir $h$ de 1 a 0.1, y de 0.1 a 0.01. Después las cosas parecen complicarse. ¿Qué está pasando? $\square$

**Problema 6.3.1** Use el Problema 5.4.3 para probar que en (6.10) se puede representar el error como $-(h^2/6)f'''(\xi)$ si la derivada tercera es continua.

**Problema 6.3.2** *Error en la aproximación central a la diferencia segunda.* Pruebe que el error de (6.9) es $-(h^2/12)f^{(iv)}(\xi)$, si $f$ es suficientemente regular.

## 6.4 La derivación numérica problema mal puesto

Cuando $h$ es pequeño en la Tabla 6.1 no se aprecia la convergencia cuadrática que cabría esperar. Esto se debe a la longitud finita de mantisa, que no nos permite disponer de los valores exactos de $\exp(\pm h)$. Estudiaremos en esta sección el efecto que sobre *las fórmulas de diferenciación numérica* tiene el hecho de no conocer exactamente los valores $f'(x_i)$.

Antes conviene esclarecer el efecto que perturbar los valores de $f$ tiene sobre *el verdadero valor de la derivada* $f(x^*)$, no sobre las aproximaciones numéricas al mismo.

**Ejemplo 6.4.1** Consideremos una función $f$ que tome valores cuyo valor absoluto sea del orden de 1. La función $\widetilde{f}(x) = f(x) + 10^{-12}\sin(10^{18}x)$ toma valores sumamente parecidos a los de $f$. Sin embargo $\widetilde{f}'(0) = f'(0) + 10^6$. □

Cambios arbitrariamente pequeños en los valores de $f$ pueden ocasionar cambios arbitrariamente grandes en el valor de la derivada en un punto. Se dice entonces que evaluar la derivada en un punto es un problema *mal puesto.*

El mal comportamiento de la operación de derivar pasa a las fórmulas de derivación numérica. Como ejemplo, analicemos la fórmula de diferenciación central (6.7). Vimos que para que el error fuese pequeño hay que tomar $h$ pequeño. Pero entonces un error de $\epsilon$ en el valor de $f(x^* \pm h)$ (como el que se perpetra al representar $f(x^* \pm h)$ en el ordenador mediante un número de precisión finita) cambia la aproximación numérica a la derivada en una gran cantidad $\pm\epsilon/h$. Además, para $h$ pequeño $f(x^* + h)$ y $f(x^* - h)$ son parecidos y la fórmula implica la substracción de cantidades vecinas, operación peligrosa para la artimética de precisión finita según estudiamos en el Capítulo 1.

**Ejemplo 6.4.2** Vea la Tabla 6.1: en la última fila el error relativo de 1.000010 como aproximación a $\exp(0.00001)$ es pequeño, y lo propio ocurre en el error de 0.9999900 en relación con $\exp(-0.00001)$. Pero hay que restar para obtener 0.000020 y esta diferencia, como aproximación a $\exp(0.00001) - \exp(-0.00001)$, contiene un gran error relativo pues en ella sólo disponemos de dos dígitos significativos, 2 y 0. *Nosotros* obtendríamos $0.000020/(2 * 10^{-5}) = 1$ como valor de la derivada, ¿de dónde sacó la calculadora 1.001358? Respuesta: ella trabaja internamente en base 2 y aunque nos muestre números como 1.000010, 0.9999900 sus aproximaciones binarias internas son ligeramente distintas. Ese *ligeramente* se amplifica al dividir entre $2 * 10^{-5}$ y ... □

# Capítulo 7

# Operadores de diferencias

## 7.1 El operador $\Delta$ y sus potencias. Tablas de diferencias

### 7.1.1 El operador $\Delta$

Es muy frecuente en los problemas de interpolación, cuadratura numérica y derivación numérica que las abscisas $x_i$ estén equiespaciadas, esto es tengamos

$$x_i = x_0 + ih, \qquad i = 0, 1, \ldots, N, \tag{7.1}$$

siendo $h$ un número positivo. Cuando se da esta circunstancia las fórmulas se pueden simplificar notablemente usando los llamados operadores de diferencias finitas, que estudiaremos en este capítulo.

El operador de diferencias más sencillo es el llamado de *diferencias progresivas* $\Delta$. Fijado un valor del parámetro $h$, el operador $\Delta$ asocia a cada función real $f$ una nueva función real $\Delta f$. La función $\Delta f$ toma en el punto $x$ el valor $\Delta f(x)$ definido como $f(x+h) - f(x)$. Por consiguiente, el dominio de definición de $\Delta f$ consta de los puntos $x$ tales que tanto $x$ como $x+h$ estén en el dominio de definición de $f$. Es importante observar que $\Delta$ cambia al variar el valor de $h$; sin embargo esta dependencia se sobreentiende y no se hace explícita en la notación.

**Ejemplo 7.1.1** Para $f(x) = x$, $h = 1$, la función $\Delta f$ vale constantemente 1. Para $f(x) = x^2$, $h = 1$, $\Delta f(x) = 2x + 1$. $\square$

A $\Delta$ se le llama operador de 'diferencias' por razones obvias. El adjetivo 'progresivo' se aplica porque el punto $x + h$ está *por delante* del punto $x$. A veces se

dice que $\Delta$ es un operador en diferencias *finitas*.[1]

### 7.1.2 Las potencias de $\Delta$

El operador $\Delta$ puede iterarse, es decir se puede aplicar a la función $\Delta f$ para obtener una nueva función $\Delta(\Delta f)$ que se escribe $\Delta^2 f$. De este modo se define un nuevo operador $\Delta^2$ que se llama de *diferencias (progresivas) segundas.*

Análogamente escribimos $\Delta^3 f$ para referirnos a la función $\Delta(\Delta^2 f)$ y, en general, para $n = 0, 1, \ldots$, $\Delta^{n+1} f$ para referirnos a la función $\Delta(\Delta^n f)$. Así definimos los operadores de *diferencias progresivas terceras, cuartas, etc.*

### 7.1.3 Diferencias a partir de valores de la función

¿Cuánto vale $\Delta^2 f(x) = [\Delta(\Delta f)](x)$? Por definición, tendremos $g(x+h) - g(x)$ donde $g$ es la función $\Delta f$. Entonces, como

$$g(x) = f(x+h) - f(x), \qquad g(x+h) = f(x+h+h) - f(x+h),$$

se verificará

$$\Delta^2 f(x) = f(x+2h) - 2f(x+h) + f(x).$$

De manera similar, puede demostrarse

$$\Delta^3 f(x) = f(x+3h) - 3f(x+2h) + 3f(x+h) - f(x)$$

y en general, para $n = 0, 1, \ldots$, (Problema 7.1.2)

$$\begin{aligned} \Delta^n f(x) &= f(x+nh) - \binom{n}{1} f(x+(n-1)h) \\ &\quad + \binom{n}{2} f(x+(n-2)h) + \cdots + (-1)^n \binom{n}{n} f(x), \end{aligned} \tag{7.2}$$

donde el *número combinatorio*

$$\binom{n}{k} = \frac{n(n-1)\cdots(n-k+1)}{k!} \tag{7.3}$$

[1]Esta terminología es clásica y se remonta a los siglos en que, siguiendo a Leibniz, se concebía la diferencial $df$ de una función $f$ como diferencia 'infinitamente pequeña' $f(x+dx) - f(x)$ entre los valores de $f$ en dos puntos 'infinitamente' próximos $x$ y $x+dx$ (concepción que sigue vigente entre muchos físicos e ingenieros). Entonces se contraponía a la diferencial infinitesimal la diferencia *finita* $f(x+h) - f(x)$.

cuenta las maneras de escoger $k$ elementos de un conjunto de $n$.

Aunque la fórmula (7.2) tiene cierto interés (por ejemplo muestra que el valor $\Delta^n f(x)$ sólo depende de los valores que $f$ tome en los $n+1$ puntos $x, \ldots, x+nh$) *no debe nunca emplearse para hallar el valor de* $\Delta^n f(x)$. Por el contrario debe usarse la definición, es decir hallar $\Delta^n f(x)$ como $\Delta^{n-1} f(x+h) - \Delta^{n-1} f(x)$, lo cual exige conocer los valores de dos diferencias de orden $n-1$, lo que a su vez se vuelve a hacer por la definición restando diferencias de orden $n-2$, etc. En el cálculo a mano las operaciones se ordenan de manera sencilla en la llamada *tabla de diferencias,* que no debemos confundir con la tabla de diferencias *divididas* que estudiamos en relación con la forma de Newton del polinomio interpolador.

**Ejemplo 7.1.2** Deseamos hallar $\Delta^3 \sin 0$ cuando $h = \pi/6$. Como vimos, el número buscado depende de los valores $\sin 0 = 0$, $\sin \pi/6 = 0.500000$, $\sin \pi/3 = 0.866025$, $\sin \pi/2 = 1$, y con ellos formamos la siguiente tabla

| | | | |
|---|---|---|---|
| 0.000000 | 0.500000 | −0.133975 | −0.098075 |
| 0.500000 | 0.366025 | −0.232050 | |
| 0.866025 | 0.133975 | | |
| 1.000000 | | | |

La primera columna tiene los valores de la función. Restando en esa columna cada número del siguiente se obtiene la segunda, que por tanto contiene diferencias primeras: $0.500000 = \Delta \sin 0$, $0.366025 = \Delta \sin \pi/6$, $0.133975 = \Delta \sin \pi/3$. Restando en la segunda columna obtenemos la tercera, que contiene $\Delta^2 \sin 0$ y $\Delta^2 \sin \pi/6$. En la última columna encontramos la cantidad buscada. La primera fila lleva valores en 0, la segunda en $\pi/6$, etc. □

Observe que, en contraposición a (7.2), formar una tabla de diferencias no requiere ninguna multiplicación. Además es más conveniente desde otro punto de vista: en el ejemplo anterior no ha habido que efectuar ningún redondeo en las cantidades intermedias.

### 7.1.4 Valores de la función a partir de diferencias

La fórmula (7.2) que expresa las potencias de $\Delta$ en términos de valores de $f$ puede invertirse. Por ejemplo, de la definición de $\Delta$ es

$$f(x+h) = \Delta f(x) + f(x),$$

y de la expresión para $\Delta^2$

$$f(x+2h) = \Delta^2 f(x) + 2f(x+h) - f(x),$$

de donde

$$f(x+2h) = \Delta^2 f(x) + 2[\Delta f(x) + f(x)] - f(x) = \Delta^2 f(x) + 2\Delta f(x) + f(x),$$

y, en general (vea el Problema 7.1.3),

$$f(x+nh) = f(x) + \binom{n}{1}\Delta f(x) + \binom{n}{2}\Delta^2 f(x) + \cdots + \binom{n}{n}\Delta^n f(x). \quad (7.4)$$

Esta fórmula tiene aplicaciones sorprendentes.

**Ejemplo 7.1.3** *Suma de los cubos.* ¿Cuánto vale la suma $1 + 8 + 27 + \cdots + n^3$ de los $n$ primeros cubos? Esta suma $f(n)$ es una función de $n$, definida sólo para $n = 1, 2, 3, \ldots$, puntos equiespaciados con $h = 1$. Podemos poner $f(0) = 0$ (la suma de ningún cubo no es nada). Usemos (7.4) con $h = 1$, $x = 0$ y para ello empecemos calculando las diferencias

| | $\Delta$ | $\Delta^2$ | $\Delta^3$ | $\Delta^4$ | $\Delta^5$ | $\Delta^6$ |
|---|---|---|---|---|---|---|
| 0 | 1 | 7 | 12 | 6 | 0 | 0 |
| 1 | 8 | 19 | 18 | 6 | 0 | |
| 9 | 27 | 37 | 24 | 6 | | |
| 36 | 64 | 61 | 30 | | | |
| 100 | 125 | 91 | | | | |
| 225 | 216 | | | | | |
| 441 | | | | | | |

¿Hay que seguir calculando eternamente? No: los siguientes elementos que aparecerán en la primera fila son todos nulos. En efecto en la segunda columna nos encontramos valores de un polinomio cúbico y según el Problema 7.1.1 sus diferencias cuartas o mayores son nulas, así que las columnas para $\Delta^5, \Delta^6, \ldots$ sólo tienen ceros. Entonces la suma buscada es, para $n \geq 4$,

$$n + 7\frac{n(n-1)}{2} + 12\frac{n(n-1)(n-2)}{6} + 6\frac{n(n-1)(n-2)(n-3)}{24}.$$

Para $n = 3$ habría en principio que suprimir el último término correspondiente a la derivada cuarta, pero como ese término contiene el factor $n - 3$ tal cosa no es necesaria. Análogas consideraciones valen para $n = 2, 1, 0$ y por ello la fórmula es cierta para cada $n = 0, 1, 2, \ldots$. Simplificándola se obtiene

$$\frac{1}{4}n^2(n+1)^2.$$

Una persona astuta y perezosa no hubiese formado las sumas que encontramos en la primera columna de la tabla: la segunda columna es obvia y a partir de ellas podemos sacar las restantes. □

### 7.1.5 Diferencias y diferencias divididas

Como indicábamos al comienzo de la sección los operadores de diferencias finitas simplifican muchas cuestiones. Como un primer ejemplo veamos su aplicación al cálculo de las diferencias divididas. Si las abscisas están dadas por (7.1)

es claro que, para $i = 0, \ldots, N-1$,

$$f[x_i, x_{i+1}] = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i} = \frac{\Delta f(x_i)}{h}$$

y, por tanto, para $i = 0, \ldots, N-2$,

$$\begin{aligned} f[x_i, x_{i+1}, x_{i+2}] &= \frac{f[x_{i+1}, x_{i+2}] - f[x_i, x_{i+1}]}{x_{i+2} - x_i} \\ &= \frac{\Delta f(x_{i+1}) - \Delta f(x_i)}{2h^2} \\ &= \frac{\Delta^2 f(x_i)}{2h^2} \end{aligned}$$

y en general, para $n = 1, 2, \ldots$, $i = 0, 1, \ldots, N-n$, (Problema 7.1.5)

$$f[x_i, x_{i+1}, \ldots, x_{i+n}] = \frac{\Delta^n f(x_i)}{n! h^n}. \tag{7.5}$$

Comparando con el Corolario 3.3.2, sacamos la conclusión de que, para $f$ suficientemente regular, $\Delta^n f(x_i)/h^n$ es uno de los valores que toma la derivada $f^{(n)}$ en el intervalo $[x_i, x_{i+n}]$.

**Problema 7.1.1** *Diferencias de un polinomio.* Demuestre que si $f$ es un polinomio de grado exactamente $N$, $N = 1, 2, \ldots$, entonces $\Delta f$ es también un polinomio, mas de grado exactamente $N-1$. Pruebe que si $f$ es una constante, $\Delta f$ es la función idénticamente nula. Concluya que, si $f$ es un polinomio de grado exactamente $N$, $N = 0, 1, 2 \ldots$, entonces $\Delta^n f$ es idénticamente nula si $n = N+1, N+2, \ldots$ y un polinomio de grado exactamente $N-n$ si $n = 1, 2, \ldots, N$.

**Problema 7.1.2** Pruebe por inducción en $n$ la fórmula (7.2).

**Problema 7.1.3** Pruebe por inducción en $n$ la fórmula (7.4).

**Problema 7.1.4** Halle las sumas $1 + 16 + \cdots + n^4$, y $1 + 32 + \cdots + n^5$.

**Problema 7.1.5** Pruebe por inducción en $n$ la fórmula (7.5).

**Problema 7.1.6** Se toma $\Delta^n f(x_i)/h^n$, $n = 1, 2, \ldots$ como aproximación a $f^{(n)}(x^*)$. Demuestre que el grado de exactitud es $\geq n$. Demuestre que el grado de exactitud es justamente $n$ salvo que $x^*$ sea el punto medio entre $x_i$ y $x_{i+n}$ en cuyo caso es $n+1$. Note que, si $n$ es par, ese punto medio es una de las abscisas $x_i, \ldots, x_{i+n}$.

## 7.2 La fórmula de Newton para interpolar en abscisas equiespaciadas

Supongamos que hemos de escribir la forma de Newton (3.10) del polinomio interpolador en el caso particular en que las abscisas (7.1) están equiespaciadas. Ante todo, podemos evitar la formación de la tabla de diferencias divididas formando en su lugar la tabla de diferencias (que ahorra las divisiones) y luego recurriendo a (7.5). Pero aún son posibles más simplificaciones. Introduzcamos una variable auxiliar $s$ tal que

$$x = x(s) = x_0 + hs.$$

Los nodos $x_i$ de la variable $x$ corresponden, en términos de la variable $s$, a los enteros $i$, $i = 0, 1, \ldots, N$. El producto $(x - x_0)(x - x_1) \cdots (x - x_{n-1})$ que en la forma de Newton acompaña a la diferencia dividida $n$-ésima, $n = 1, \ldots, N$, vale

$$\begin{aligned}
&(x_0 + hs - x_0)(x_0 + hs - x_1) \cdots (x_0 + hs - x_{n-1}) = \\
&(x_0 + hs - x_0)(x_0 + hs - x_0 - h) \cdots (x_0 + hs - x_0 - (n-1)h) = \\
&h^n s(s-1) \cdots (s - n + 1).
\end{aligned}$$

Usando (7.5)

$$\begin{aligned}
&(x - x_0)(x - x_1) \cdots (x - x_{n-1}) f[x_0, x_1, \ldots, x_n] \\
&\qquad = s(s-1) \cdots (s - n + 1) \frac{\Delta^n f(x_0)}{n!},
\end{aligned}$$

y en definitiva la forma de Newton del polinomio interpolador es

$$\begin{aligned}
P(x_0 + hs) = \quad & f(x_0) \\
& + s\Delta f(x_0) \\
& + s(s-1)\frac{\Delta^2 f(x_0)}{2!}, \\
& \cdots \\
& + s(s-1) \cdots (s - N + 1)\frac{\Delta^N f(x_0)}{N!}.
\end{aligned} \tag{7.6}$$

**Ejemplo 7.2.1** Deseamos aproximar $\sin 1$ interpolando por un polinomio cúbico en las abscisas $0, \pi/6, \pi/3, \pi/2$. La correspondiente tabla de diferencias la hallamos en el Ejemplo 7.1.2 y el polinomio será

$$0.000000 + 0.500000 \times s - \frac{0.133975}{2} \times s(s-1) - \frac{0.098075}{6} \times s(s-1)(s-2).$$

Evaluando en $s = 6/\pi = 1.909859$ que se corresponde con $x = 1$, encontramos que los sumandos del polinomio valen 0, 0.954930, −0.116404, 0.002560. Por tanto aproximando $\sin 1 = 0.841471$ por la recta que interpola en 0, $\pi/6$ obtenemos 0.954930; por la parábola que interpola en 0, $\pi/6$, $\pi/3$ obtenemos $0.954930 - 0.116404 = 0.838526$, y finalmente la interpolación cúbica da $0.838526 + 0.002560 = 0.841086$. *Note que usando la forma (7.6) estamos obligados a ir añadiendo los nodos de menor* 0 *a mayor* $\pi/2$. La forma (3.10) es más versátil pues en ella no hay que suponer que los $x_i$ aparecen ordenados; por tanto al usarla podemos ir añadiendo los nodos en el orden que creamos más conveniente, digamos primero el $\pi/6$, luego el $\pi/3$, luego el 0 y luego el $\pi/2$. □

Hay una forma más sencilla de escribir (7.6). Para ello extenderemos el símbolo (7.3) que está definido para $n = 1, 2, \ldots$, $k = 0, 1, \ldots, n$, definiendo para $s$ *real* cualquiera y $k = 0, 1, \ldots$

$$\binom{s}{k} = \frac{s(s-1)\cdots(s-k+1)}{k!}; \tag{7.7}$$

visto como función de $s$ nos encontramos ante un polinomio de grado exactamente $k$. Note que, a menos que $s$ sea un entero $\geq k$, este símbolo carece de interpretación combinatoria como maneras de elegir $k$ de entre $s$. Observe asímismo que

$$\binom{s}{k} = 0, \qquad k = 1, 2, \ldots, \qquad s = 0, 1, \ldots, k-1. \tag{7.8}$$

Con la notación (7.7), la forma de Newton (7.6) se escribe

$$P(x_0 + sh) = f(x_0) + \binom{s}{1}\Delta f(x_0) + \binom{s}{2}\Delta^2 f(x_0) + \cdots + \binom{s}{N}\Delta^N f(x_0). \tag{7.9}$$

Podemos emplear (7.4) para comprobar que el segundo miembro coincide con $f$ en los nodos $x_0, \ldots, x_N$, esto es, cuando $s = 0, 1, \ldots, N$. Para ello en (7.4) hay que dar a $x$ el valor $x_0$, y a $n$ sucesivamente los valores $0, 1, \cdots, N$ y usar (7.8).

**Problema 7.2.1** Resuelva los Problemas 3.1.1, 3.1.2, 3.1.3, 3.1.4 usando la forma de Newton para abscisas equiespaciadas.

**Problema 7.2.2** *El binomio de Newton.* Un uso importante de (7.7) aparece en el desarrollo de Taylor de $(1+x)^s$ para $s$ real. Pruebe que tal desarrollo es

$$1 + \binom{s}{1}x + \binom{s}{2}x^2 + \cdots$$

Esta es la célebre fórmula del binomio de Newton, grabada en la tumba del matemático en la abadía de Westminster. Cuando $s$ es un entero positivo $n$, la suma es finita y tiene $n+1$ términos porque, según (7.8), $n$ sobre $k$ es nulo si $n = k+1, k+2, \ldots$ El caso en que $s$ no es un entero positivo es el interesante descubierto por Newton: entonces la serie no acaba. Este es uno de los desarrollos de Taylor que hay que saberse de memoria. Converge si $-1 < x < 1$.

**Problema 7.2.3** *Las fórmulas de diferenciación progresiva.* Derivando la forma de Newton del polinomio interpolador, demuestre que

$$\frac{1}{h}\left[\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \cdots \pm \frac{1}{N}\Delta^N\right] f(x_0)$$

es la fórmula interpolatoria para aproximar $f'(x_0)$ combinando linealmente $f(x_0), \ldots, f(x_N)$. ¿Qué error produce al aplicarse a $x^{N+1}$? (Indicación: añadiendo un término a la suma la fórmula es exacta para $x^{N+1}$.)

## 7.3 Otros operadores de diferencias

Además del operador $\Delta$ hay otros operadores de diferencias. Los más frecuentes son el de *diferencias regresivas* $\nabla$, con $\nabla f(x) = f(x) - f(x-h)$, y los dos de *diferencias centrales* $\delta$ y $\Delta_0$, con $\delta f(x) = f(x+h/2) - f(x-h/2)$, $\Delta_0 f(x) = f(x+h) - f(x-h)$. El símbolo $\nabla$ se lee *nabla*, por tener la figura de un arpa antigua del mismo nombre.

Ninguno de estos tres nuevos operadores es imprescindible; $\nabla f(x)$ no es otra cosa que $\Delta f(x-h)$, $\delta f(x)$ coincide con $\Delta f(x-h/2)$, y $\Delta_0$ es $\delta$ con un valor doble del parámetro $h$. Sin embargo muchas fórmulas son notacionalmente más sencillas si se usan los operadores adecuados en vez de limitarse a $\Delta$. Veamos algunos ejemplos de uso de estos operadores.

**Ejemplo 7.3.1** *La forma de Newton regresiva del polinomio interpolador.* Deseamos interpolar en los puntos (7.1) por un polinomio $P$ de grado $\leq N$. Ya notamos en el Ejemplo 7.2.1 que al usar (7.6) se añaden los nodos de menor a mayor. Esto puede no ser aconsejable: quizá se desea $P$ para evaluarlo en puntos próximos a $x_N$ y entonces es preferible añadir los nodos en el orden $x_N, x_{N-1}, \ldots, x_0$ para introducir lo antes posibles los nodos cercanos a los puntos donde vamos a evaluar. La forma regresiva (Problema 7.3.1)

$$\begin{aligned} P(x_N + ht) = \quad & f(x_N) \\ & + t\nabla f(x_N) \\ & + t(t+1)\frac{\nabla^2 f(x_N)}{2!}, \\ & \cdots \\ & + t(t+1)\cdots(t+N-1)\frac{\nabla^N f(x_N)}{N!} \end{aligned} \tag{7.10}$$

tiene la propiedad de que el primer término es la constante que interpola en $f(x_N)$, los dos primeros la recta que además interpola en $x_{N-1}$, etc. □

**Ejemplo 7.3.2** Particularicemos (7.10) al polinomio cúbico que interpola a la función seno en $0, \pi/6, \pi/3, \pi/2$. Las diferencias regresivas que necesitamos pueden leerse en diagonal en la tabla del Ejemplo 7.1.2 (¿por qué?) y los sumandos son 1.000000, 0.133975, −0.232050, −0.098975. □

**Ejemplo 7.3.3** *Aproximación central a la derivada.* La expresión $\Delta_0 f(x)/(2h)$ da la aproximación central (6.7) a $f'(x)$ como combinación lineal de $f(x \pm h)$. □

**Ejemplo 7.3.4** *Aproximación central a la derivada segunda.* Puesto que

$$\delta\delta f(x) = \delta f(x + \frac{h}{2}) - \delta f(x - \frac{h}{2}) = f(x+h) - f(x) - [f(x) - f(x-h)],$$

la expresión $\delta^2 f(x)/h^2$ suministra la aproximación central usual (6.9) a $f''(x)$ como combinación lineal de $f(x+h)$, $f(x)$, $f(x-h)$. □

**Problema 7.3.1** Demuestre (7.10).

**Problema 7.3.2** *Las fórmulas de diferenciación regresiva.* (Compare con el Problema 7.2.3.) Demuestre, derivando la forma regresiva del polinomio interpolador, que

$$\frac{1}{h}\left[\nabla + \frac{1}{2}\nabla^2 + \frac{1}{3}\nabla^3 + \cdots + \frac{1}{N}\nabla^N\right] f(x_N) \tag{7.11}$$

es la fórmula interpolatoria para aproximar $f'(x_N)$ combinando linealmente $f(x_0), \ldots, f(x_N)$. ¿Qué error produce al aplicarse a $x^{N+1}$?

## 7.4 El cálculo de diferencias finitas

Además de los operadores de diferencias podemos introducir otros. El más sencillo es la *identidad* $Id$ que 'transforma' cada función $f$ en sí misma. También tenemos el operador de *desplazamiento* $E$; por definición $Ef$ es la función que en $x$ toma el valor $f(x+h)$.

Los operadores pueden de modo obvio (i) sumarse, (ii) multiplicarse por números reales y (iii) componerse entre sí.

**Ejemplo 7.4.1** $3Id$ es el operador que asocia a $f$ la función que en $x$ vale $3f(x)$; la composición de $\Delta$ consigo mismo es el operador $\Delta^2$ que encontramos más arriba, etc. ¿Qué es el operador $Id + \Delta$? Respuesta: $Id + \Delta$ asocia a $f$ la función que en $x$ vale $Idf(x) + \Delta f(x)$; pero esto es $f(x) + f(x+h) - f(x)$ y entonces

$$Id + \Delta = E. \quad \square$$

Las igualdades entre operadores como las que acabamos de obtener conducen a muchas fórmulas de interés.

**Ejemplo 7.4.2** Elevemos la igualdad anterior a la potencia $n$ (es decir compongamos cada miembro consigo mismo $n$ veces)

$$E^n = (Id + \Delta)^n.$$

El segundo miembro puede calcularse por la fórmula del binomio, porque los operadores $Id$ y $\Delta$ conmutan,[2] es decir $Id\Delta = \Delta Id = \Delta$. Obtenemos

$$E^n = Id + \binom{n}{1}\Delta + \cdots + \binom{n}{n}\Delta^n,$$

y aplicando estos operadores a una función $f$ y evaluando en $x$ obtenemos (7.4). □

Análogamente a lo que acabamos de ilustrar en el ejemplo, todas las fórmulas de este capítulo son obtenibles combinando operadores. Las manipulaciones necesarias a menudo requieren manejar series de operadores y pueden no ser rigurosas. Cuando tales series se aplican a una función polinómica las series suelen reducirse a sumas finitas y las manipulaciones son fáciles de justificar. El cálculo de operadores da un medio sencillo de inventar fórmulas, o, cuando menos, de acordarse de ellas. Pueden verse en los problemas numerosos casos particulares de interés, como la fórmula sumatoria de Euler y Maclaurin (1698–1746). Esta permite encontrar fórmulas para la suma de las potencias de los números naturales, hallar aproximadamente la suma de series, etc.

**Problema 7.4.1** ¿Cómo obtener la gráfica de $Ef$ a partir de la de $f$? La respuesta *no* es desplazándola $h$ unidades hacia la *derecha*.

**Problema 7.4.2** Utilice la identidad $\Delta^n = (E - Id)^n$ para probar (7.2).

**Problema 7.4.3** *El polinomio interpolador otra vez.* Demuestre que el cuadrado del operador que asocia a la función $f$ la función que en $x$ vale $f(x + h/2)$ es el operador $E$. Por ello ese operador puede denotarse como $E^{1/2}$ de suerte que la ley de multiplicación de exponentes $(E^{1/2})^2 = E$ sea válida. Pruebe que el inverso de $E$ es el operador que asocia a $f$ la función que en $x$ vale $f(x - h)$. Por consideraciones similares es fácil comprobar que, si se define para cada racional $p/q$ la potencia $E^{p/q}$ por $E^{p/q}f(x) = f(x + (p/q)h)$, entonces son válidas las leyes usuales para exponentes. Extendemos la definición a $s$ irracional poniendo también $E^s f(x) = f(x + sh)$. Desarrolle $E^s = (Id + \Delta)^s$ por la fórmula del binomio de Newton (Problema 7.2.2) para concluir que

$$f(x_0 + sh) = f(x_0) + \binom{s}{1}\Delta f(x_0) + \binom{s}{2}\Delta^2 f(x_0) + \cdots$$

Este es un desarrollo similar al de Taylor pero con diferencias en vez de derivadas. Truncando se van obteniendo las formas de Newton de los polinomios interpoladores de Lagrange de grados más y más altos.

[2] Si $A$ y $B$ no conmutan, $(A+B)(A+B) = A^2 + AB + BA + B^2$ es diferente del desarrollo binomial $A^2 + 2AB + B^2$.

**Problema 7.4.4** *Derivación y diferencias finitas.* Definamos el operador $D$ que asocia a cada función derivable la función $Df = f'$. Demuestre que el operador $\exp(hD) = Id + hD + (h^2/2)D^2 + \dots$ al aplicarse a un polinomio $P$ (cualquiera que sea el grado) le asocia la función $EP$. Note que, aunque $\exp(hD)$ tiene en principio infinitos términos, la suma $P + hDP + (h^2/2)D^2P + \cdots$ posee sólo un número finito de términos no nulos. Entonces, $\exp(hD) = E$ y $D = h^{-1}\log E$.

**Problema 7.4.5** *Las fórmulas de diferenciación progresiva otra vez.* Combinando la identidad $E = Id + \Delta$ con el resultado $D = h^{-1}\log E$ obtenido en el problema anterior, se tiene $D = h^{-1}\log(Id + \Delta)$, y desarrollando el logaritmo

$$D = \frac{1}{h}\left[\Delta - \frac{1}{2}\Delta^2 + \frac{1}{3}\Delta^3 - \cdots\right].$$

El sentido de esta igualdad es que ambos miembros coinciden al aplicarse a un polinomio, sea cual sea su grado. Obtenga a partir de aquí la fórmula del Problema 7.2.3.

**Problema 7.4.6** *Las fórmulas de diferenciación regresiva otra vez.* Demuestre la identidad $E^{-1} = Id - \nabla$. Demuestre que $D = -h^{-1}\log(Id - \nabla)$ y

$$D = \frac{1}{h}\left[\nabla + \frac{1}{2}\nabla^2 + \frac{1}{3}\nabla^3 + \cdots\right].$$

Obtenga de aquí la fórmula de diferenciación regresiva (7.11).

**Problema 7.4.7** *Las funciones hiperbólicas.* Las funciones *seno hiperbólico* y *coseno hiperbólico* se definen para cada valor real de $x$ por las relaciones $\sinh x = (\exp(x) - \exp(-x))/2$, $\cosh x = (\exp(x) + \exp(-x))/2$. Pruebe que $\cosh^2 x - \sinh^2 x = 1$. Pruebe que $(d/dx)\sinh x = \cosh x$, $(d/dx)\cosh x = \sinh x$. Concluya que $\sinh x$ es una función impar que crece monótonamente de $-\infty$ a $+\infty$ cuando $x$ crece de $-\infty$ a $\infty$. Esboce su gráfica.

**Problema 7.4.8** *La función inversa de* $\sinh x$. Por su monotonía, la función $y = \sinh x$ tiene una inversa. Esta se llama *argumento seno hiperbólico* y se denota por $x = \operatorname{arg\,sinh} y$. Crece monótonamente de $-\infty$ a $+\infty$ cuando $y$ lo hace de $-\infty$ a $\infty$. Demuestre que $(d/dy)\operatorname{arg\,sinh} y = (1 + y^2)^{-1/2}$. Demuestre que la serie de Taylor de $\operatorname{arg\,sinh} y$ es

$$y - \frac{1}{2}\frac{y^3}{3} + \frac{1\cdot 3}{2\cdot 4}\frac{y^5}{5} - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\frac{y^7}{7} + \cdots$$

(Indicación: mediante el binomio de Newton, halle la serie de Taylor de la derivada y después integre término a término.)

**Problema 7.4.9** *Fórmulas centrales para aproximar la derivada.* Demuestre la relación $\delta = 2\sinh(hD/2)$ y que por tanto $D = (2/h)\operatorname{arg\,sinh}(\delta/2)$. Use el problema precendente para encontar la fórmula

$$D = \frac{1}{h}\left[\delta - \frac{1}{2}\frac{\delta^3}{3\cdot 2^2} + \frac{1\cdot 3}{2\cdot 4}\frac{\delta^5}{5\cdot 2^4} - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6}\frac{\delta^7}{7\cdot 2^6} + \cdots\right].$$

Truncando este desarrollo se obtienen aproximaciones a $D$ dadas por $h^{-1}\delta$, $h^{-1}[\delta - (1/24)\delta^3]$, ... exactas para polinomios de grado 2, 4, ...

**Problema 7.4.10** *Fórmulas centrales para aproximar la derivada segunda.* Eleve al cuadrado el desarrollo del problema precedente para encontrar

$$D^2 = \frac{1}{h^2}\left[\delta^2 - \frac{1}{12}\delta^4 + \frac{1}{90}\delta^6 - \frac{1}{560}\delta^8 + \frac{1}{3150}\delta^{10} - \cdots\right]$$

Truncando, se obtienen fórmulas $h^{-2}\delta^2$, $h^{-2}[\delta^2 - (1/12)\delta^4]$, ... exactas para polinomios de grado $3, 5, \ldots$

**Problema 7.4.11** *Lós números de (Jakob) Bernoulli (1654–1708).* Estos números, $B_0 = 1$, $B_1 = -1/2$, $B_2 = 1/6$, $B_3 = 0, \ldots$ aparecen en muchas fórmulas. Se definen por la relación

$$\frac{x}{\exp(x) - 1} = \sum_{n=0}^{\infty} \frac{1}{n!} B_n x^n.$$

Halle los once primeros. (Indicación: multiplique ambos miembros por $\exp(x) - 1$ e identifique coeficientes.)

**Problema 7.4.12** *La integral definida y la suma en términos de operadores.* La operación inversa de la derivación $DF(x) = F'(x) = f(x)$ es la primitivación (integral indefinida). Por eso $D^{-1}$ no está definido de manera unívoca sino salvo constante aditiva (si $DF(x) = f(x)$, entonces todas las funciones que difieren de $F$ en una constante también se aplican por $D$ en $f$). Sin embargo la combinación $E^n D^{-1} - D^{-1}$ sí posee sentido único pues la constante de integración desaparece al restar. Recordando que $x_k = x_0 + kh$, pruebe la igualdad

$$(E^n - Id)D^{-1}f(x_0) = \int_{x_0}^{x_n} f(x)\, dx.$$

Demuestre que, de modo análogo,

$$(E^n - Id)\Delta^{-1}f(x_0) = \sum_{k=0}^{n-1} f(x_k).$$

**Problema 7.4.13** *Fórmula sumatoria de Euler–Maclaurin.* Demuestre:

$$\sum_{k=0}^{n-1} f(x_k) = h^{-1}\int_{x_0}^{x_n} f(x)\, dx + \frac{B_1}{1!}[f(x_n) - f(x_0)] + \frac{B_2}{2!}h[f'(x_n) - f'(x_0)] + \frac{B_3}{3!}h^2[f''(x_n) - f''(x_0)] + \cdots,$$

donde los $B_k$ son los números de Bernoulli. (Indicación: use el problema precedente. Escriba $(E^n - Id)\Delta^{-1} = (E^n - Id)D^{-1}[D\Delta^{-1}] = (E^n - Id)D^{-1}[D(\exp(hD) - Id)^{-1}]$. Use la definición de los números de Bernoulli.)

**Problema 7.4.14** *Sumas de potencias.* Emplee la fórmula de Euler–Maclaurin para hallar la suma de los $n$ primeros números naturales, la suma de los correspondientes cuadrados, la suma de los correspondientes cubos. Dé una expresión general, usando los números de Bernoulli, para $\sum_{k=1}^{n} n^\ell$, donde $\ell = 0, 1, 2, \ldots$

# Capítulo 8

# Ecuaciones en diferencias

## 8.1 Preliminares

En este capítulo trataremos con sucesiones $u_0$, $u_1$, $u_2$, ... Para referirnos a la globalidad de la sucesión (y no a uno de sus términos) usaremos la notación $\{u_j\}_{j=0}^{\infty}$, que a veces abreviaremos a $\{u_j\}$. Por razones que se aclararán posteriormente, es conveniente considerar sucesiones cuyos términos son números *complejos*; esto hace las cosas más *fáciles* que si considerásemos sólo números reales.

Denotaremos por $\mathcal{S}$ al conjunto de todas las sucesiones de números complejos. Las sucesiones se suman entre sí y se multiplican por un escalar de manera obvia,

$$\{u_j\} + \{v_j\} = \{u_j + v_j\}, \qquad \alpha\{u_j\} = \{\alpha u_j\},$$

y, con estas operaciones, $\mathcal{S}$ es un espacio vectorial de dimensión infinita (Problema 8.1.1).

### 8.1.1 Concepto de ecuación en diferencias

Si $F$ es una función compleja de dos variables complejas, la expresión

$$u_{j+1} = F(u_j, j) \tag{8.1}$$

se llama *ecuación en diferencias* (de primer orden). Una sucesión $\{u_j\}$ se llama *solución* de (8.1) si, para cada valor de $j = 0, 1, \ldots$, se verifica la igualdad (8.1).

**Ejemplo 8.1.1** Sea $F$ la función $F(\alpha, \beta) = (\beta + 1)\alpha$. La correspondiente ecuación en diferencias es $u_{j+1} = (j+1)u_j$. La sucesión $\{j!\}$ es una solución porque para cada $j$ es cierto que $u_{j+1} = (j+1)! = (j+1)j! = (j+1)u_j$. $\square$

Más generalmente se pueden considerar ecuaciones en diferencias de segundo orden

$$u_{j+2} = G(u_{j+1}, u_j, j) \tag{8.2}$$

(ahora $G$ es una función compleja de tres variables complejas), de tercer orden

$$u_{j+3} = H(u_{j+2}, u_{j+1}, u_j, j),$$

etc.

**Ejemplo 8.1.2** En un corral se deja una pareja de conejos recién nacidos, macho y hembra. Pasado un mes, los conejos son adultos y se aparean. Acabado el segundo mes la coneja pare una nueva pareja, macho y hembra, y acto seguido se produce un nuevo apareamiento. El proceso se va reiterando y cada pareja se aparea por primera vez al mes de nacer y luego lo hace cada mes, originando, cada mes, una nueva pareja descendiente. Se trata de saber el número de parejas al cabo de un año o más generalmente al cabo de $j$ meses.[1] Observemos que, si $u_j$ es el número de parejas transcurridos $j$ meses,

$$u_{j+2} = u_{j+1} + u_j, \tag{8.3}$$

porque las parejas al cabo de $j+2$ meses se obtienen sumando a las $u_{j+1}$ que había el mes anterior los partos que se han producido, y estos son tantos como parejas había al cabo de $j$ meses. □

### 8.1.2 Problemas de valores iniciales

Cada ecuación en diferencias tiene infinitas soluciones..

**Ejemplo 8.1.3** Compruebe que las sucesiones $3j!$, $\pi j!$ e $ij!$ ($i$ es la unidad imaginaria) son soluciones de la ecuación del Ejemplo 8.1.1. Encuentre infinitas más. □

Al existir infinitas soluciones para cada ecuación en diferencias, la información de que una sucesión que deseemos conocer satisface una ecuación en diferencias dada no es suficiente para determinar dicha sucesión.

**Ejemplo 8.1.4** En el Ejemplo 8.1.2 haber planteado (8.3) no es suficiente para determinar el número de conejos. Por ejemplo, la sucesión formada toda por ceros satisface (8.3) y no es la buscada. □

Para una ecuación de primer orden (8.1) un *problema de valores iniciales* pide tratar de encontrar una solución de la ecuación tal que $u_0$ sea un número complejo $\alpha$ dado, llamado la *condición inicial.*

[1] Este problema aparece en el *Liber abaci* (1202) del matemático de Pisa Leonardo Fibonacci ($\sim$ 1170 – $\sim$ 1250), introductor en Europa del sistema de numeración indoarábigo que hoy usamos y sustituyó al romano.

**Ejemplo 8.1.5** Tomemos el problema de valores iniciales dado por la ecuación del Ejemplo 8.1.1 y la condición inicial $u_0 = 7$. De la ecuación, con $j = 0$, obtenemos $u_1 = 1 \times 7 = 7$; con $j = 1$, $u_2 = 2 \times 7 = 14$; con $j = 3$, $u_3 = 3 \times 14 = 42$; ... . *Note que este proceso, por su carácter iterativo, es muy fácil de implementar en un ordenador.* Por otra parte vemos que cada término de la sucesión está determinado unívocamente y por ello el problema de valores iniciales tiene una única solución. Es fácil escribirla en forma cerrada: $u_j = 7j!$. □

Para una ecuación de segundo orden (8.2) un *problema de valores iniciales* trata de encontrar una solución de la ecuación tal que $u_0$ y $u_1$ sean respectivamente números complejos $\alpha$, $\beta$ dados, llamados las *condiciones iniciales.*

**Ejemplo 8.1.6** Para el Ejemplo 8.1.2 conocemos las condiciones iniciales $u_0 = 1$, $u_1 = 1$. Usando la ecuación con $j = 0$, $u_2 = 1+1 = 2$, y sucesivamente $u_3 = 2+1 = 3$, $u_4 = 3+2 = 5$, $u_5 = 5 + 3 = 8$, $u_6 = 8 + 5 = 13$, $u_7 = 13 + 8 = 21$, etc. (De nuevo notemos que el proceso iterativo es muy fácilmente implementable en el ordenador.) Los términos de la sucesión quedan determinados y la solución del problema de valores iniciales es única. La sucesión hallada se llama de Fibonacci y, al menos según algunos, sus términos aparecen frecuentemente en la naturaleza, p.ej. en el número de semillas en diversos frutos. □

Para una ecuación de orden $N$ el problema de valores iniciales pide hallar la solución en que $u_0, \ldots, u_{N-1}$ son números complejos dados. La ecuación tiene infinitas soluciones, el problema de valores iniciales una sola.

### 8.1.3 Escritura en términos de diferencias

¿Dónde están las diferencias en las ecuaciones que hemos venido considerando?

**Ejemplo 8.1.7** Para la ecuación del Ejemplo 8.1.1, si escribimos $u_{j+1} = \Delta u_j + u_j$, tendremos la forma alternativa $\Delta u_j = j u_j$; antes se ligaba $u_{j+1}$ a $u_j$, ahora se liga $\Delta u_j$ a $u_j$. □

**Ejemplo 8.1.8** Para (8.3) tendremos, usando (7.4), $u_{j+2} = \Delta^2 u_j + 2\Delta u_j + u_j$, $u_{j+1} = \Delta u_j + u_j$, y llevando esto a la ecuación $\Delta^2 u_j = -\Delta u_j + u_j$. □

Una ecuación de orden $N$ puede escribirse siempre ó ligando $N + 1$ términos consecutivos $u_{j+N}, \ldots, u_j$ ó ligando $u_j$ y las diferencias $\Delta u_j, \ldots, \Delta^N u_j$.

**Problema 8.1.1** *Espacio vectorial de sucesiones.* Pruebe que el conjunto $S$ de todas las sucesiones de números complejos, con las operaciones usuales de suma de sucesiones y producto por un escalar, es un espacio vectorial. Pruebe que las sucesiones

$$\begin{matrix} 1 & 0 & 0 & 0 & \ldots \\ 0 & 1 & 0 & 0 & \ldots \\ 0 & 0 & 1 & 0 & \ldots \\ 0 & 0 & 0 & 1 & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots \end{matrix} \tag{8.4}$$

son linealemente independientes y por tanto la dimensión de $\mathcal{S}$ es infinita. ¿Son esas sucesiones una base del espacio? (Respuesta: no; la sucesión cuyos elementos son todos iguales a 1 no es combinación lineal de ellas.)

**Problema 8.1.2** Encuentre una ecuación en diferencias cuyas soluciones sean todas y solas las progresiones geométricas de razón 2. Idem para las progresiones aritméticas de razón 7. Idem para todas las progresiones aritméticas, sea cual sea su razón. En cada caso escriba la ecuación resultante tanto con términos consecutivos como usando diferencias.

**Problema 8.1.3** Se escribe una ecuación en diferencias de orden $N$ en términos de las diferencias $u_j, \Delta u_j, \ldots, \Delta^N u_j$. Para imponer los valores iniciales vamos a prescribir valores de diferencias $\Delta^n u_0$ ¿cuáles necesitamos además de $u_0 = \Delta^0 u_0$?

## 8.2 Ecuaciones en diferencias lineales

Cualquier objeto matemático *lineal* es más fácil de analizar que si fuese no lineal. Por ello los modelos lineales son siempre los primeros en estudiarse.

### 8.2.1 Definiciones

La ecuación de primer orden (8.1) se llama *lineal* si es de la forma

$$u_{j+1} + a_j u_j = f_j, \tag{8.5}$$

siendo $\{a_j\}$ y $\{f_j\}$ sucesiones conocidas. La primera se llama sucesión de *coeficientes* y la segunda consituye el *término independiente, término no homogéneo o término fuente.*

**Ejemplo 8.2.1** Las ecuaciones $u_{j+1} = \sin(ju_j)$ y $u_{j+1} = (u_j)^2 + j$ no son lineales. La del Ejemplo 8.1.1 lo es; en ella, para cada $j$, $a_j = -(j+1)$, $f_j = 0$. □

La ecuación de segundo orden (8.2) se llama *lineal* si es de la forma

$$u_{j+2} + a_j u_{j+1} + b_j u_j = f_j. \tag{8.6}$$

Aquí los coeficientes $\{a_j\}$, $\{b_j\}$ y el término independiente, no homogéneo o fuente $\{f_j\}$ son sucesiones conocidas.

**Ejemplo 8.2.2** La ecuación de Fibonacci (8.3) es lineal. Para cada $j$, es $a_j = b_j = -1$, $f_j = 0$. La ecuación $u_{j+2} = u_{j+1}u_j$ no es lineal. □

En general una ecuación en diferencias de orden $N$ es lineal si es de la forma

$$u_{j+N} + a_j^{[N-1]} u_{j+N-1} + \cdots + a_j^{[0]} u_j = f_j. \tag{8.7}$$

Ahora hay $N$ sucesiones conocidas de coeficientes $\{a_j^{[N-1]}\}, \ldots, \{a_j^{[0]}\}$ y $\{f_j\}$ se denomina, una vez más, término independiente, no homogéneo o fuente.

Una ecuación en diferencias, sea cual sea su orden, se llama *homogénea* si su término fuente es la sucesión idénticamente nula. En otro caso se llama *no homogénea*.

### 8.2.2 Operador asociado

Al primer miembro de la ecuación (8.7) se le asocia la aplicación u operador $\mathcal{L}$ que transforma cada sucesión $\{v_j\} \in \mathcal{S}$ en la nueva sucesión $\mathcal{L}\{v_j\}$ cuyo término $j$-ésimo, $j = 0, 1, \ldots$, es

$$v_{j+N} + a_j^{[N-1]} v_{j+N-1} + \cdots + a_j^{[0]} v_j.$$

Esta aplicación es *lineal* (Problema 8.2.2).

**Ejemplo 8.2.3** Para la ecuación $u_{j+1} - (j+1)u_j = 0$ que hemos venido considerando, $\mathcal{L}$ transforma la sucesión $\{1\}$ en la sucesión $\{-j\}$, la sucesión $\{j\}$ en la sucesión $\{-j^2+1\}$. □

### 8.2.3 Ecuación homogénea

Consideremos la ecuación homogénea asociada a la ecuación (8.7), es decir,

$$u_{j+N} + a_j^{[N-1]} u_{j+N-1} + \cdots + a_j^{[0]} u_j = 0. \tag{8.8}$$

Habiendo introducido el operador $\mathcal{L}$ podemos decir que una sucesión $\{u_j\}$ es solución de (8.8) si y solamente si $\mathcal{L}\{u_j\} = \{0\}$. En otras palabras, las soluciones de la ecuación homogénea son las contraimágenes por $\mathcal{L}$ del 'cero' de $\mathcal{S}$, o, con la nomenclatura usual del álgebra lineal, las soluciones son justamente los elementos del núcleo de $\mathcal{L}$. Es bien sabido (Problema 8.2.3) que ese núcleo es un subespacio vectorial. Por ello *las soluciones de una ecuación homogénea forman un subespacio vectorial de* $\mathcal{S}$ o, equivalentemente, *cualquier combinación lineal de soluciones de la ecuación homogénea es una nueva solución.*

**Ejemplo 8.2.4** En el Ejemplo 8.1.1 se descubre inmediatamente la solución $\{j!\}$. Por consiguiente, para cualquier número complejo $\alpha$, la sucesión $\{\alpha u_j\}$ es también una solución. □

**Teorema 8.2.1** *Sean*

$$\begin{bmatrix} u_0^{[1]} \\ u_1^{[1]} \\ \vdots \\ u_{N-1}^{[1]} \end{bmatrix}, \begin{bmatrix} u_0^{[2]} \\ u_1^{[2]} \\ \vdots \\ u_{N-1}^{[2]} \end{bmatrix}, \ldots, \begin{bmatrix} u_0^{[N]} \\ u_1^{[N]} \\ \vdots \\ u_{N-1}^{[N]} \end{bmatrix} \tag{8.9}$$

*$N$ vectores $N$-dimensionales linealmente independientes. Entonces las sucesiones $\{u_j^{[1]}\}, \ldots, \{u_j^{[N]}\}$ soluciones de (8.8) que los tienen por condiciones iniciales son una base del conjunto de soluciones de (8.8). Este conjunto es un espacio vectorial de dimensión $N$.*

*Demostración.* Las $N$ sucesiones $\{u_j^{[i]}\}$ son linealmente independientes: una combinación lineal nula de ellas da origen, tras descartar los subíndices $j = N, N+1, \ldots$, a una combinación lineal nula de los vectores libres (8.9) y por tanto ha de tener coeficientes todos nulos.

Consideremos para acabar una solución cualquiera $\{u_j\}$ y veamos que se puede escribir como combinación lineal de las $N$ dadas. Desde luego el vector $[u_0, u_1, \ldots, u_{N-1}]^T$ de condiciones iniciales de $\{u_j\}$ será combinación lineal de los vectores (8.9). Existen entonces escalares $\alpha^{[1]}, \ldots, \alpha^{[N]}$, tales que

$$u_j = \alpha^{[1]} u_j^{[1]} + \cdots + \alpha^{[N]} u_j^{[N]}$$

cuando $j = 0, 1, \ldots, N-1$. Nuestro objetivo es ver que esta igualdad vale también para $j = N, N+1, \ldots$ Consideremos las sucesiones

$$\{u_j\}, \quad \{\alpha^{[1]} u_j^{[1]} + \cdots + \alpha^{[N]} u_j^{[N]}\};$$

ambas tienen las mismas condiciones iniciales (por nuestra elección de las alfas) y ambas son soluciones de la misma ecuación en diferencias de orden $N$ (la primera por hipótesis; la otra es combinación lineal de soluciones). Al satisfacer ambas el mismo problema de valores iniciales han de ser la misma. Vea el Problema 8.2.4. □

Ahora que sabemos que las *infinitas* soluciones de una ecuación homogénea forman un espacio de dimensión $N$, podemos encontrar *todas* hallando $N$ que sean independientes y combinándo éstas linealmente. Y para cerciorarnos de que $N$ soluciones dadas sean independientes nos basta con comprobar que sus $N$ vectores de condiciones iniciales lo son.

### 8.2.4 La ecuación no homogénea

Pasemos ahora a (8.7). En términos del operador $\mathcal{L}$ estamos buscando sucesiones $\{u_j\}$ que por $\mathcal{L}$ se transformen en $\{f_j\}$. La combinación lineal de soluciones *NO* es una nueva solución, p.ej. si $\{u_j\}$ satisface $\mathcal{L}\{u_j\} = \{f_j\}$ entonces $\{2u_j\}$ se aplica por $\mathcal{L}$ en $\{2f_j\}$, que no es lo que nos interesa.

Supongamos que $\{u_j\}$ es una solución de (8.7) y que $\{v_j\}$ satisface la ecuación *homogénea* asociada (8.8). Entonces $\{u_j + v_j\}$ es una nueva solución de (8.7); en efecto

$$\mathcal{L}\{u_j + v_j\} = \mathcal{L}\{u_j\} + \mathcal{L}\{v_j\} = \{f_j\} + \{0\} = \{f_j\}.$$

Además, si *fijamos* $\{u_j\}$ y hacemos que $\{v_j\}$ recorra todas las posibles soluciones de la ecuación homogénea, obtenemos todas las soluciones posibles de la no homogénea (8.7). Pues usando una vez más la linealidad, si $\{w_j\}$ es una nueva solución del no homogéneo, entonces $\mathcal{L}\{w_j - u_j\} = \{0\}$, de modo que al poner $\{w_j\} = \{u_j\} + \{w_j - u_j\}$ estamos escribiendo $\{w_j\}$ como suma de la solución fijada $\{u_j\}$ y una solución $\{w_j - u_j\}$ de la homogénea. Esto se soliía tradicionalmente sintetizar con la frase *"la solución general de una ecuación no homogénea es suma de una particular y la general de la homogénea."*[2]

**Problema 8.2.1** Suponga que al corral donde están los conejos de Fibonacci echa el dueño, transcurridos dos meses desde que llevó la primera pareja, una nueva pareja recién nacida. Transcurridos tres, cuatro, ..., meses sigue trayendo nuevas parejas recién nacidas, una al mes. Cada pareja traída se multiplica como la primera. Escriba la ecuación en diferencias que gobierna el proceso. Observe que las aportaciones al corral del exterior aparecen como término fuente.[3]

**Problema 8.2.2** Demuestre que el operador $\mathcal{L}$ es lineal, esto es, para cada par de escalares $\alpha$ y $\beta$ y cada par de sucesiones $\{v_j\}$, $\{w_j\}$, se verifica $\mathcal{L}(\alpha\{v_j\} + \beta\{w_j\}) = \alpha\mathcal{L}\{v_j\} + \beta\mathcal{L}\{w_j\}$.

**Problema 8.2.3** *Núcleo de una aplicación lineal.* Demuestre que si $X$ e $Y$ son espacios vectoriales y $T$ es una aplicación lineal de $X$ en $Y$, el núcleo de $T$, conjunto de los $x \in X$ con $Tx = 0 \in Y$ es un subespacio vectorial de $X$, es decir contiene todas las combinaciones lineales de sus elementos.

---

[2]Estas consideraciones no son sino un resultado básico de álgebra lineal: si $T$ es una aplicación lineal e $y$ es un elemento dado, entonces las soluciones $x$ de la ecuación $Tx = y$, contraimagen de $y$ por $T$, son de la forma $x_0 + n$ donde $x_0$ es una solución particular y $n$ recorre el correspondiente núcleo.

[3]En general los procesos que cursan sin intervención del exterior se rigen por ecuaciones homogéneas; los que interactúan con el medio por ecuaciones no homogéneas. El término fuente mide las aportaciones del medio.

**Problema 8.2.4** Este ejercicio clarifica la demostración del Teorema 8.2.1. Escriba los términos $u_j$, $j = 0, \ldots, 10$, de la solución $s_1$ de la ecuación de Fibonacci con $u_0 = 0$, $u_1 = 1$. Escriba los términos $u_j$, $j = 0, \ldots, 10$, de la solución $s_2$ de la ecuación de Fibonacci con $u_0 = 1$, $u_1 = 1$. Escriba los términos $u_j$, $j = 0, \ldots, 10$, de la solución $s_3$ de la ecuación de Fibonacci con $u_0 = 2$, $u_1 = 3$. Halle los coeficientes $\alpha$ y $\beta$ por los que hay que multiplicar los vectores planos $[0, 1]$ y $[1, 1]$ para obtener $[2, 3]$. Compruebe que, para $j = 0, \ldots, 10$, el término $j$-ésimo de $s_3$ coincide con el término $j$-ésimo de $\alpha s_1 + \beta s_2$.

## 8.3 Caso de coeficientes constantes

La ecuación (8.7) se llama de coeficientes constantes si las sucesiones de coeficientes $\{a_j^{[N-1]}\}, \ldots, \{a_j^{[0]}\}$ son constantes, es decir si es de la forma

$$u_{j+N} + a^{[N-1]}u_{j+N-1} + \cdots + a^{[0]}u_j = f_j. \tag{8.10}$$

Si además la ecuación es homogénea tendremos

$$u_{j+N} + a^{[N-1]}u_{j+N-1} + \cdots + a^{[0]}u_j = 0. \tag{8.11}$$

No hay pérdida de la generalidad si se supone de aquí en adelante que $a^{[0]} \neq 0$: si $a^{[0]}$ fuese nulo la ecuación se podrá reducir a una de menor orden y lo mismo ocurre para el problema de valores iniciales.

**Ejemplo 8.3.1** La ecuación de segundo orden $u_{j+2} = 5u_{j+1}$, $j = 0, 1, \ldots$ se reescribe como $v_{j+1} = 5v_j$ con sólo poner $v_j = u_{j+1}$. En el problema de valores iniciales para la ecuación de segundo orden con $u_0 = \alpha$, $u_1 = \beta$, los valores de $u_j$ con $j = 2, 3, \ldots$ son independientes de $\alpha$ y se determinan resolviendo el problema de valores iniciales $v_{j+1} = 5v_j$, $v_0 = \beta$. □

### 8.3.1 La ecuación característica

Para las ecuaciones *homogéneas de coeficientes constantes* es posible hallar *explícitamente* todas las soluciones. Para ello supongamos que la solución es de la forma $\{r^j\}$ con $r$ un número complejo *no nulo*, es decir que es una progresión geométrica, cuya razón $r$ está por determinar. Sustituyendo en (8.11) necesitamos que

$$r^{j+N} + a^{[N-1]}r^{j+N-1} + \cdots + a^{[0]}r^j = 0,$$

para cada $j = 0, 1, \ldots$; esto son infinitas condiciones. Afortunadamente dividiendo entre $r^j \neq 0$ todas se reducen a una sola,

$$r^N + a^{[N-1]}r^{N-1} + \cdots + a^{[0]} = 0$$

que debe satisfacerse para que $\{r^j\}$ sea una solución. Recíprocamente si $r$ satisface esta condición debe ser $r \neq 0$ (puesto que suponemos que $a^{[0]} \neq 0$) y $\{r^j\}$ es una solución no trivial (es decir no coincidente con la idénticamente nula) de la ecuación en diferencias.

Se suele introducir el polinomio

$$\rho(z) = z^N + a^{[N-1]}z^{N-1} + \cdots + a^{[0]} \tag{8.12}$$

que se llama *característico* de la ecuación (8.11). Con esta terminología, hemos probado que si $r$ es una raíz de la ecuación característica $\rho(z) = 0$ (es decir un cero del polinomio característico $\rho$) entonces $\{r^j\}$ es solución no trivial de (8.11).

Supongamos que la ecuación característica (8.12) tiene todas sus raíces $r_n$, $n = 1, \ldots, N$ distintas. Entonces tenemos $N$ soluciones particulares de (8.11) dadas por $\{r_n^j\}_{j=0}^{\infty}$, $n = 1, \ldots, N$. Los vectores de condiciones iniciales,

$$\begin{bmatrix} r_1^0 \\ r_1^1 \\ \vdots \\ r_1^{N-1} \end{bmatrix}, \begin{bmatrix} r_2^0 \\ r_2^1 \\ \vdots \\ r_2^{N-1} \end{bmatrix}, \ldots, \begin{bmatrix} r_N^0 \\ r_1^1 \\ \vdots \\ r_N^{N-1} \end{bmatrix},$$

son linealmente independientes al constituir las columnas de una matriz no singular (de Vandermonde, Problema 3.1.6). Aplicando el Teorema 8.2.1 nuestras $N$ soluciones particulares son una base del espacio de soluciones y la solución más general es

$$\{A_1 r_1^j + \cdots + A_N r_N^j\}$$

con las $A_n$ constantes arbitrarias.

**Ejemplo 8.3.2** Retomemos el ejemplo de la sucesión de Fibonacci. Para (8.3) la ecuación característica es $z^2 = z + 1$ con raíces

$$\Omega = (1 + \sqrt{5})/2 \approx 1.618, \qquad \omega = (1 - \sqrt{5})/2 \approx -0.618;$$

la positiva $\Omega$ se llama *razón áurea* y es un número notable. La solución más general es pues $\{A\Omega^j + B\omega^j\}$, con $A$ y $B$ constantes cualesquiera.

Para encontrar la sucesión de Fibonacci imponemos que $u_0 = u_1 = 1$ ó

$$\begin{aligned} A + B &= 1, \\ A\frac{1+\sqrt{5}}{2} + B\frac{1-\sqrt{5}}{2} &= 1, \end{aligned}$$

condiciones que nos sirven para determinar las constantes $A$ y $B$. La solución del sistema da $A = (5 + \sqrt{5})/10$ y $B = (5 - \sqrt{5})/10$ y por ello la sucesión de Fibonacci es

$$u_j = \frac{5+\sqrt{5}}{10}\left(\frac{1+\sqrt{5}}{2}\right)^j + \frac{5-\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{2}\right)^j.$$

El segundo miembro, a pesar de su apariencia, es entero para cada $j = 0, 1, \ldots$ Como $\Omega > 1$ y $|\omega| < 1$, el primer sumando del segundo miembro crece exponencialmente con $j$ mientras que el segundo sumando tiende exponencialmente a 0. Por ello, para $j$ no muy pequeño, $u_j$ es muy aproximadamente $A\Omega^j$, término de una progresión geométrica de razón $\Omega$. Cada mes los conejos aumentan aproximadamente un 61.8 %, resultado que distaba de ser obvio antes de resolver el problema. □

**Ejemplo 8.3.3** Consideremos ahora el problema de valores iniciales $u_0 = u_1 = 1$, $u_{j+2} = u_{j+1} + u_j + 1$ que describe una población de conejos con aportes mensuales, Problema 8.2.1. La ecuación es no homogénea y vamos a buscar una solución particular. Como el segundo miembro es constante buscamos también una solución con valor constante, digamos $c$. Sustituyendo en la ecuación $c = c + c + 1$, es decir $c = -1$. Ahora hemos de encontrar todas las soluciones de la homogénea, pero esto lo tenemos hecho en el ejemplo precedente donde hallamos la expresión $\{A\Omega^j + B\omega^j\}$. La solución más general de la ecuación es pues $\{-1 + A\Omega^j + B\omega^j\}$.

Imponiendo las condiciones iniciales, escribimos

$$\begin{aligned} A + B &= 2, \\ A\frac{1+\sqrt{5}}{2} + B\frac{1-\sqrt{5}}{2} &= 2. \end{aligned}$$

El segundo miembro es doble del que teníamos en el ejemplo precedente y, por linealidad, $A$ y $B$ son aquí el doble de lo que allí eran. Ahora el número de parejas al cabo de $j$ meses vale

$$-1 + 2\frac{5+\sqrt{5}}{10}\left(\frac{1+\sqrt{5}}{2}\right)^j + 2\frac{5-\sqrt{5}}{10}\left(\frac{1-\sqrt{5}}{2}\right)^j;$$

uno menos del doble que el valor encontrado en el ejemplo precedente. □

Las soluciones $\{r^j\}$ son en general complejas, aunque la ecuación en diferencias (8.11) y su polinomio característico tengan coeficientes reales. Por esto nos hemos venido situando de entrada en el caso complejo. Cuando interesan soluciones reales podemos proceder como en el ejemplo que sigue.

**Ejemplo 8.3.4** Para la ecuación $u_{j+2} - 2u_{j+1} + 2u_j = 0$ las raíces de la ecuación característica son $1 \pm i$ y la solución general

$$\{A(1+i)^j + B(1-i)^j\} \tag{8.13}$$

con $A$ y $B$ constantes *complejas* arbitrarias. Si estamos interesados en sucesiones *reales* elegiremos para $B$ el complejo conjugado $\bar{A}$, y entonces $B(1-i)^j$ será conjugado de $A(1+i)^j$ dando lugar a una suma real $A(1+i)^j + \bar{A}(1-i)^j$. Sea entonces $A = a + ib$ con $a$ y $b$ reales, y tomemos $B = a - ib$. Para manejar potencias la forma módulo/argumento es preferible, así que escribamos $1 \pm i = \sqrt{2}\exp(\pm i\pi/4)$, con lo que

$$(1 \pm i)^j = \sqrt{2^j}\exp(\pm\frac{i\pi j}{4}) = \sqrt{2^j}\cos\frac{\pi j}{4} \pm i\sqrt{2^j}\sin\frac{\pi j}{4}.$$

Sustituyendo en (8.13) y operando obtenemos

$$\{2a\sqrt{2^j}\cos\frac{\pi j}{4} + 2b\sqrt{2^j}\sin\frac{\pi j}{4}\},$$

que es la solución *real* más general. □

### 8.3.2 Caso de raíces multiples

Eliminemos la hipótesis de que las raíces de la ecuación característica son todas distintas. Sean entonces $r_1, r_2, \ldots, r_M$ las raíces dos a dos distintas, que sabemos son no nulas. Para $n = 1, \ldots, M$, supongamos que $r_n$ tiene multiplicidad $\mu_n$, esto es que el factor $z - r_n$ figura exactamente $\mu_n$ veces en la factorización de $\rho(z)$ en $N$ factores lineales. Entonces a cada raíz $r_n$ le corresponden las siguientes $\mu_n$ soluciones de (8.11)

$$\{r_n^j\}, \{jr_n^j\}, \{j^2 r_n^j\}, \ldots, \{j^{\mu_n - 1} r_n^j\}; \tag{8.14}$$

las $\mu_1$ soluciones de este tipo que corresponden a $r_1$, unidas a las $\mu_2$ correspondientes a $r_2, \ldots$ y a las $\mu_M$ correspondientes a $r_M$ constituyen una base del espacio de soluciones. La demostración de estos asertos puede verse en los problemas de esta sección.

**Ejemplo 8.3.5** Para la ecuación en diferencias lineal, homogénea y de coeficientes constantes de orden 6 con raíces de la ecuación característica 1 triple, 2 simple y 3 doble, la solución más general es

$$\{A + Bj + Cj^2 + D2^j + E3^j + Fj3^j\}. \square$$

### 8.3.3 Comportamiento de las soluciones cuando $j \to \infty$

A menudo, dada una ecuación como (8.11), interesa saber si cada una de sus soluciones $\{u_j\}$ tiende a 0 cuando $j \to \infty$ ó si es una sucesión acotada (es decir, si existe una constante $M$, independiente de $j$ pero que puede variar de una solución a otra, tal que $|u_j| \leq M$ para cada $j$). A este respecto tenemos el siguiente resultado.

**Teorema 8.3.1** *Es condición necesaria y suficiente para que cada sucesión solución de (8.11) tienda a cero que todas las raíces de la ecuación característica tengan módulo menor que la unidad.*

*Es condición necesaria y suficiente para que cada sucesión solución de (8.11) sea acotada que todas las raíces de la ecuación característica tengan módulo menor o igual que la unidad y las de módulo uno sean simples.*

*Demostración.* Haremos tan solo el caso de convergencia a 0, el de acotación es análogo. Necesidad: si hay una raíz $r_n$ de módulo $\geq 1$ entonces $\{r_n^j\}$ es una solución de la ecuación en diferencias que no tiende a 0. Suficiencia: si las raíces $r_n$ son todas de módulo menor que la unidad entonces todas las soluciones particulares (8.14) tienden hacia 0 (las exponenciales vencen a los polinomios); cualquier otra solución es combinación lineal de ellas y también tenderá hacia 0. $\square$

**Problema 8.3.1** *Razón áurea.* Un rectángulo de papel tiene su altura $h$ y su base $b$ en razón áurea, $h/b = \Omega$. Para hacer una pajarita, se segrega un cuadrado de lado $b$. Pruebe que en el rectángulo restante la base y la altura vuelven a estar en razón áurea. Los rectángulos en que la base y altura están en razón áurea son particularmente agradables a la vista y esta proporción aparece en numerosas obras de arte.

**Problema 8.3.2** Halle la solución más general de $u_{j+2} - (3/2)u_{j+1} + u_j = 1$ (busque una solución particular constante). Resuelva el problema de valores iniciales para la ecuación anterior con $u_0 = \alpha$, $u_1 = \beta$ y determine para qué valores de $\alpha$ y $\beta$ las soluciones están acotadas.

**Problema 8.3.3** Halle la solución más general de $u_{j+2} = -u_j + 1$, buscando una solución particular constante. Luego escriba la forma de todas las soluciones reales.

**Problema 8.3.4** Halle la solución más general de $u_{j+2} = -u_j + j$, buscando una solución particular que sea un polinomio de primer grado en $j$. Luego escriba la forma de todas las soluciones reales.

**Problema 8.3.5** Halle la solución más general de $u_{j+3} = u_j + 4^j$ buscando una solución particular de la forma $\alpha 4^j$. Luego escriba todas las soluciones reales.

**Problema 8.3.6** *Ceros múltiples de un polinomio.* Pruebe que $r$ es un cero de multiplicidad $\mu$ de un polinomio $\rho(z)$ (esto es $\rho(z)$ tiene $\mu$ veces el factor $(z - r)$) si y solamente si $r$ anula a $\rho$ y a sus $\mu - 1$ primeras derivadas.

**Problema 8.3.7** *Caso de raíces múltiples en la ecuación característica.* En este problema y los siguientes vamos a justificar la construcción de la base de soluciones de (8.11) en el caso de raíces múltiples. Consideraremos sólo un caso particular que tiene todos los ingredientes requeridos para abordar el general. Supongamos que las raíces son 1 simple y 2 triple.

Demuestre que la ecuación en diferencias es $u_{j+4} - 7u_{j+3} + 18u_{j+2} - 20u_{j+1} + 8u_j = 0$.

**Problema 8.3.8** *Continuación.* Las sucesiones $s_1 = \{1\}$ y $s_2 = \{2^j\}$ son, de acuerdo con la teoría de esta sección, soluciones. Sustituya $s_3 = \{j2^{j-1}\}$ en el primer miembro de la ecuación y demuestre que el resultado coincide con el valor en $z = 2$ de la derivada de la función $F_j(z) = z^j\rho(z)$ con respecto a $z$. Demuestre que, al ser $F_j'(z) = (z^j)'\rho(z) + z^j\rho'(z)$ y al ser 2 cero de $\rho$ y $\rho'$ (Problema 8.3.6), se verificará que $F_j'(z) = 0$. Concluya que $s_3$ es solución.

Considere ahora $s_4 = \{j(j-1)2^{j-2}\}$. Sustituya en el primer miembro de la ecuación en diferencias y demuestre que el resultado coincide con el valor en $z = 2$ de la derivada segunda de la función $F_j(z)$. Use que $F_j''(z) = (z^j)''\rho(z) + 2(z^j)'\rho'(z) + z^j\rho''(z)$ junto con el hecho que una raíz triple de $\rho = 0$ anula a $\rho$, $\rho'$ y $\rho''$ para concluir que $F_j''(2) = 0$ y por tanto que $s_4$ es solución.

**Problema 8.3.9** *Continuación.* Demostremos que $s_1$, $s_2$, $s_3$ y $s_4$ son linealmente independientes. Formemos con tal fin los vectores de condiciones iniciales

$$\begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \begin{bmatrix}1\\2\\4\\8\end{bmatrix}, \begin{bmatrix}0\\1\\4\\12\end{bmatrix}, \begin{bmatrix}0\\0\\2\\12\end{bmatrix},$$

y supongamos que no son independientes con lo que la matriz $A$ que los tiene por *columnas* es singular. Existen escalares $a_0$, $a_1$, $a_2$ y $a_3$, no todos nulos, que al multiplicarse por las *filas* de $A$ y sumar dan el vector fila nulo. Pruebe que entonces el polinomio cúbico, no idénticamente nulo, $a_0 + a_1 z + a_2 z^2 + a_3 z^3$ tiene cuatro ceros (el 1 simple y el 2 triple), lo cual es absurdo. En síntesis, $s_1$, $s_2$, $s_3$ y $s_4$ son una base del espacio de soluciones.

**Problema 8.3.10** *Conclusión.* Pruebe que $S_1 = \{1\}$, $S_2 = \{2^j\}$, $S_3 = \{j2^j\}$, $S_4 = \{j^2 2^j\}$ son otra base del espacio de soluciones, comprobando que las $s_n$ son combinaciones lineales de las $S_n$, con lo que toda solución puede expresarse combinando linealmente las $S_n$.

**Problema 8.3.11** ¿Para qué valores del parámetro real $\lambda$ son acotadas (respectivamente convergentes a 0) todas las soluciones de $u_{j+2} + \lambda u_{j+1} + u_j = 0$?

# Capítulo 9

# Ecuaciones no lineales

## 9.1 Introducción

Nos ocuparemos ahora de encontrar numéricamente las soluciones (raíces) de ecuaciones

$$f(x) = 0, \tag{9.1}$$

donde $f$ es una función dada, real de variable real. Estas soluciones se llaman también *ceros* de $f$. El caso más sencillo es aquel en que $f$ es un polinomio $P$. Si $P$ es lineal, $P(x) = ax + b$, la solución de (9.1) sólo requiere una división; el caso cuadrático, $P(x) = ax^2 + bx + c$, es igualmente simple y necesita la extracción de raíces cuadradas. La fórmula para resolver la ecuación cúbica se debe a Tartaglia (1499–1557) y dal Ferro (1465–1526) y la correspondiente a la cuártica a Ferrari (1522–1565). Para la quíntica la fórmula se buscó durante siglos; en 1826 Abel (1802–1829) probó que no existe. Por esto las ecuaciones polinómicas de grado $\geq 5$ han de resolverse por técnicas numéricas. *A fortiori* (es decir, con más razón) hay que usar técnicas numéricas para las ecuaciones en que $f$ no es un polinomio.

## 9.2 Bisección

El método de bisección es el más sencillo posible. Hay que suponer que la función $f$ en (9.1) está definida y es continua en un intervalo cerrado y acotado $I_0 = [a, b]$ y que toma valores de signo contrario en $a$ y $b$ (esto es, que ó bien $f(a) > 0$, $f(b) < 0$, ó bien $f(a) < 0$, $f(b) > 0$). En estas condiciones, el teorema de Bolzano (1781–1848) garantiza que $f$ se anula en algún punto del interior de $I_0$.

El método comienza calculando $f(c)$, siendo $c = (a+b)/2$ el punto medio del intervalo. Si tenemos la fortuna de ser $f(c) = 0$, hemos hallado una solución a nuestro problema y paramos. Caso contrario, $f(c)$ tiene ó bien el signo de $f(b)$ ó bien el signo de $f(a)$, y uno de los dos intervalos $[a, c]$, $[c, b]$ tiene la propiedad de que en sus extremos $f$ toma valores de signo contrario.

Llamemos $I_1$ a ese intervalo en cuyos extremos la función toma valores de signo contrario. Ahora estamos justamente como en la situación de partida pero con $I_1$ reemplazando a $I_0$; la ventaja es que la longitud de $I_1$ es la mitad que la de $I_0$. Repetimos el procedimiento: evaluamos $f$ en el punto medio de $I_1$ y ó bien damos con una solución del problema ó bien podemos quedarnos con un intervalo $I_2$, de longitud la mitad de la longitud de $I_1$, en cuyos extremos la función tiene signos distintos.

Iterando el proceso o hallamos una solución del problema o, tras $n$ pasos, dispondremos de un intervalo $I_n$ de longitud $(b-a)/2^n$ que contiene una raíz.

**Ejemplo 9.2.1** Es bien sabido que los planetas al girar en torno al Sol y los satélites artificiales al hacerlo en torno a la Tierra describen órbitas elípticas. Para hallar en qué punto de la elipse se encuentra el móvil en un tiempo dado $t$ hay que resolver la ecuación llamada de Kepler (1571–1630)

$$x - e\sin x = z, \tag{9.2}$$

donde $e$ es la excentricidad conocida de la elipse ($e = 0$ para una órbita circular y $e$ se acerca a 1 a medida que la elipse se va haciendo más oblonga) y $z$ es un número conocido que se ha calculado a partir de $t$. Aquí $f(x) = x - e\sin x - z$.

Resolvamos en el caso $e = 0.5$, $z = 0.7$. Claramente $f(0) = -0.7 < 0$, mientras que $f(2) = 1.3 - 0.5\sin 2 > 1.3 - 0.5 = 0.8 > 0$, así que el intervalo $I_0 = [0, 2]$ nos vale para empezar. Los resultados están en la Tabla 9.1. Al ser $f(1) < 0$ el punto medio 1 de $I_0$ sustituye al extremo $x = 0$ de $I_0$ para pasar a $I_1 = [1, 2]$ (Fig. 9.1). Tras cinco bisecciones, hemos evaluado la función $f$ cinco veces y sabemos que la raíz está entre 1.125 y 1.1875. Tomando como aproximación a la solución el promedio $1.15625 = (1.125 + 1.1875)/2$ estamos seguros de que el valor absoluto del error no supera 0.03125. □

El método de bisección:

- Funciona siempre.
- Da cotas superior e inferior para la raíz buscada.
- Es más lento que otros, que, a cambio, son más inseguros.

**Problema 9.2.1** Escriba un programa que implemente el método de bisección. El cálculo debe interrumpirse cuando se produzca el primero de estos dos eventos: (i) el número de evaluaciones de $f$ excede un número $M$ prescrito por el usuario, (ii) la longitud del intervalo es menor que una tolerancia $\epsilon$ prescrita por el usuario.

Tabla 9.1: Método de bisección para $f(x) = x - 0.5\sin x - 0.7 = 0$

| $n$ | Intervalo $I_n$ | Punto medio de $I_n$ | Valor de $f$ en punto medio |
|---|---|---|---|
| | $f(0) < 0,\ f(2) > 0$ | | |
| 0 | [0,2] | 1 | −0.121 |
| 1 | [1,2] | 1.5 | +0.301 |
| 2 | [1,1.5] | 1.25 | +0.075 |
| 3 | [1,1.25] | 1.125 | −0.026 |
| 4 | [1.125,1.25] | 1.1875 | +0.023 |
| 5 | [1.125,1.1875] | | |

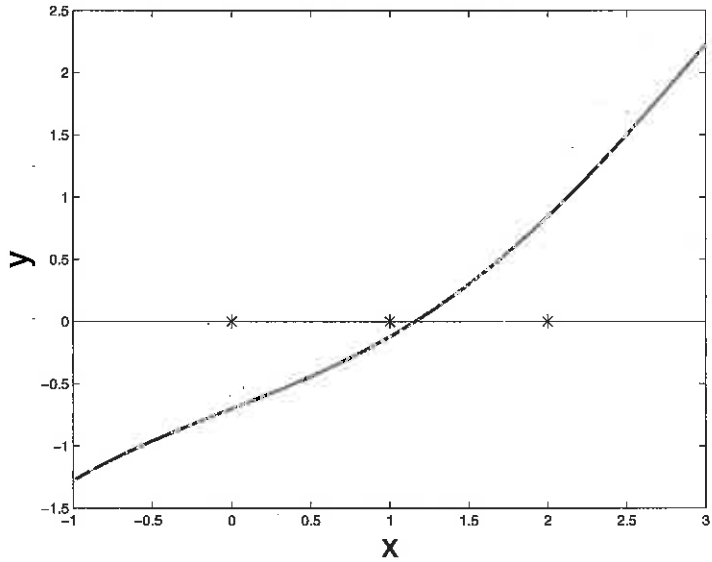


Figura 9.1: Método de bisección para $f(x) = x - 0.5\sin x - 0.7 = 0$

**Problema 9.2.2** En el ejemplo de la Tabla 9.1 ¿hasta qué $n$ hay que llegar para conocer la raíz con error menor que $5 \times 10^{-6}$?

## 9.3 Método de la secante

Otro método clásico es el de la secante, que se basa en la interpolación lineal.

Partimos de dos aproximaciones (más o menos buenas) $x_0$ y $x_1$ a la raíz que buscamos $\alpha$. Construimos la recta *secante* $P_1(x)$ que interpola a $f$ en $x_0$ y $x_1$, y, en vez de resolver (9.1), resolvemos $P_1(x) = 0$, llamando $x_2$ a su raíz. Si la interpolación fuese exacta y $P_1$ coincidiese con $f$, $x_2$ coincidiría con el valor buscado $\alpha$; si el error de interpolación es pequeño esperamos que $x_2$ sea una aproximación a $\alpha$ mejor que las de partida $x_0$, $x_1$ (Fig. 9.2).

Ahora el proceso sigue con el par $x_1$, $x_2$ en el papel que antes tenían $x_0$ y $x_1$. Es decir, interpolamos $f$ en $x_1$, $x_2$ y llamamos $x_3$ al cero del interpolante. Similarmente calcularemos $x_4$, $x_5$, etc.

Como $x_{n+2}$ es el cero de la recta que interpola en $x_n$, $x_{n+1}$, tendremos

$$f(x_{n+1}) + f[x_{n+1}, x_n](x_{n+2} - x_{n+1}) = 0,$$

de modo que los llamados *iterantes* $x_n$ del método se calculan por

$$x_{n+2} = x_{n+1} - \frac{f(x_{n+1})}{f[x_{n+1}, x_n]}, \quad n = 0, 1, \ldots \tag{9.3}$$

¿Observa que estamos usando una *ecuación en diferencias* no lineal de segundo orden con condiciones iniciales $x_0$, $x_1$ dadas?

La aplicación sucesiva de (9.3) no es siempre posible, pudiera ser que $x_{n+1}$ no estuviese en el dominio en que $f$ esté definida, o pudiera ser que $f[x_{n+1}, x_n] = 0$ en cuyo caso la fórmula no se puede aplicar (en tal caso la recta secante es horizontal y, salvo que $x_{n+1}$ fuese ya un cero de $f$, no corta al eje de abscisas).

Abandonamos el cálculo de los iterantes $x_n$ cuando se han hallado valores para los cuales $f$ es en módulo suficientemente pequeño, o cuando dos iterantes consecutivos distan poco entre sí, o cuando se ha excedido un número de iteraciones que se juzga razonable.

**Ejemplo 9.3.1** Resolvamos el caso de la Tabla 9.1 por el método de la secante con $x_0 = 0$, $x_1 = 2$. Los resultados están en la Tabla 9.2. La tercera columna da los errores; para calcularla hemos tomado como verdadero valor de $\alpha$ la aproximación que da el propio método de la secante cuando se calculan muchos iterantes. Lo más interesante de la tabla es que el método funciona

Tabla 9.2: Método de la secante para $f(x) = x - 0.5\sin x - 0{,}7 = 0$

| $n$ | $x_n$ | $x_n - \alpha$ | $f(x_n)$ |
|---|---|---|---|
| 0 | 0 | $-1.2 \times 10^{0}$ | $-7.00 \times 10^{-1}$ |
| 1 | 2 | $+8.4 \times 10^{-1}$ | $+8.45 \times 10^{-1}$ |
| 2 | 0.90594288 | $-2.5 \times 10^{-1}$ | $-1.88 \times 10^{-1}$ |
| 3 | 1.10460660 | $-5.3 \times 10^{-2}$ | $-4.20 \times 10^{-2}$ |
| 4 | 1.16199419 | $+4.0 \times 10^{-3}$ | $+3.20 \times 10^{-3}$ |
| 5 | 1.15794015 | $-6.1 \times 10^{-5}$ | $-4.91 \times 10^{-5}$ |
| 6 | 1.15800155 | $-7.0 \times 10^{-8}$ | $-5.60 \times 10^{-8}$ |

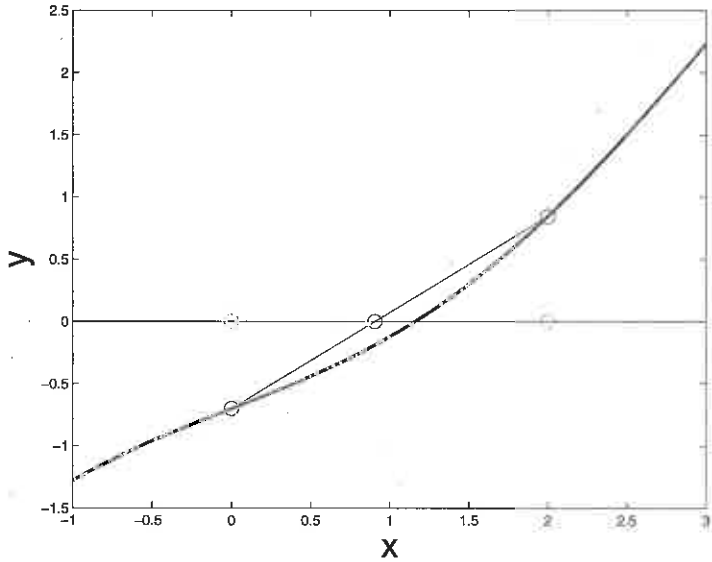


Figura 9.2: Método de la secante para $f(x) = x - 0.5\sin x - 0.7 = 0$

cada vez mejor al ir incrementando $n$: pasar de $n = 2$ a $n = 3$ divide el error entre 4, mientras que pasar de $n = 5$ a $n = 6$ divide el error entre 1000. Como consecuencia es posible obtener muy buenas aproximaciones a la raíz con un número moderado de iteraciones. □

El método de la secante

- No siempre genera aproximaciones convergentes a la raíz buscada.
- No da cotas superior e inferior para la raíz buscada.
- Cuando da aproximaciones que convergen a la raíz es mucho más rápido que el de bisección.

**Problema 9.3.1** Escriba un programa que implemente el método de la secante. El cálculo debe interrumpirse cuando se produzca el primero de estos dos eventos: (i) el número de evaluaciones de $f$ excede un número $M$ prescrito por el usuario, (ii) la distancia entre dos iterantes consecutivos es menor que una tolerancia prescrita por el usuario.

# 9.4 Iteración de punto fijo

La iteración de punto fijo *(o método de las aproximaciones sucesivas)* es una técnica general de gran importancia en la resolución de todo tipo de ecuaciones, incluyendo las ecuaciones diferenciales, integrales y en derivadas parciales que surgen al modelar matemáticamente los fenómenos del mundo real. Por otra parte está íntimamente relacionada con la idea de sistema dinámico discreto (vea los problemas sobre dinámica de poblaciones al final de esta sección). Aquí estudiaremos la aplicación de la iteración de punto fijo a la resolución numérica de (9.1).

## 9.4.1 Descripción del método

Comencemos con un ejemplo.

**Ejemplo 9.4.1** La excentricidad de la órbita de Venus es $e = 0.007$. Tratemos de resolver la correspondiente ecuación de Kepler (9.2) cuando $z = 0.7$, es decir resolvamos

$$f(x) = x - 0.007 \sin x - 0.7 = 0. \tag{9.4}$$

- El término $0.007 \sin x$ es despreciable frente a 0.7 de modo que, ignorándolo en la ecuación, concluimos que la solución será aproximadamente 0.7.
- Ahora que sabemos que $x \approx 0.7$, resulta que $0.007 \sin x$ (que supusimos 0) será $\approx 0.007 \times \sin 0.7 = 0.004510$, y, llevando este valor a la ecuación, $x \approx 0.7 + 0.004510 = 0.704510$.

- Puesto que ahora creemos que $0.007 \sin x \approx 0.007 \sin 0.704510 = 0.004534$, usando la ecuación, $x \approx 0.704534$.
- Finalmente $0.007 \sin x \approx 0.007 \sin 0.704534 = 0.004534$, que llevado a la ecuación da $x \approx 0.704534$. Es decir $x = 0.704534$ satisface la ecuación, dentro de la precisión de las millonésimas con la que hemos trabajado.

Notemos que cada aproximación a la raíz se obtiene de la precedente $p$ por la fórmula $0.7 + 0.007 \sin p$. □

En la iteración de punto fijo:

- Se generan aproximaciones sucesivas $x_0$, $x_1$, $x_2$, ..., a la raíz buscada $\alpha$ con la esperanza de podernos acercar más y más a ella. Matemáticamente deseamos que $x_n \to \alpha$ para $n \to \infty$, lo cual garantiza que, fijada una tolerancia, existe un valor finito de $n$ para el que los iterantes $x_n$, $x_{n+1}$, ... distan de $\alpha$ menos que la tolerancia.
- Cada aproximación $x_{n+1}$, $n = 0, 1, \ldots$, se obtiene como función de la precedente $x_n$, esto es $x_{n+1} = g(x_n)$, siendo $g$ la llamada *función de iteración*, que no hay que *confundir con la función* $f$ cuyos ceros queremos hallar. En otras palabras, estamos usando una *ecuación en diferencias* de primer orden.

**Ejemplo 9.4.2** En el Ejemplo 9.2, $x_{n+1} = 0.7 + 0.007 \sin x_n$, con lo que

$$g(x) = 0.7 + 0.007 \sin x. \qquad \square \tag{9.5}$$

Dada una ecuación (9.1) ¿cómo hallaremos una función de iteración $g$ que nos permita obtener las aproximaciones sucesivas? Ante todo notemos que, suponiendo la continuidad de $g$ y que $x_n \to \alpha$, se verifica que

$$\alpha = \lim x_{n+1} = \lim g(x_n) = g(\lim x_n) = g(\alpha).$$

Esto muestra que una condición *necesaria* para que la iteración tenga éxito es que $g(\alpha) = \alpha$, o, como también se dice, $\alpha$ sea un *punto fijo* de $g$ (que no se mueve al aplicarle $g$). Esta condición es muy fácil de entender: significa que, si por casualidad llegamos a un iterante $x_n$ que coincide con la solución buscada $\alpha$, entonces el siguiente iterante $g(x_n)$ también coincide con $\alpha$.

**Ejemplo 9.4.3** En el Ejemplo 9.2, si $\alpha$ es cero de $f$ en (9.4), entonces es evidentemente punto fijo de (9.5). □

En síntesis, para aplicar el método hay que reescribir el problema $f(x) = 0$ en la forma equivalente $x = g(x)$. Esto siempre se puede hacer de muchas formas; pero no todas son útiles.

**Ejemplo 9.4.4** La ecuación (9.4) se puede reescribir como

$$x = g(x) = \arcsin \frac{x - 0.7}{0.007}. \tag{9.6}$$

Con la aproximación inicial 0.7 que antes usamos, tenemos $x_1 = 0$ (que es una aproximación muy mala a la raíz), $x_2 = \arcsin(-100)$ (que ni siquiera está definido). ¿Por qué la función de iteración (9.6) falla y (9.5) tiene éxito? En (9.5) el valor de $x$ influye poco en el valor de $g$ debido al factor 0.007. En (9.6) el valor de $g$ es muy sensible a cambios en $x$. □

En general interesa que $g(x)$ varíe poco al variar $x$: el caso ideal sería tener $g$ constante; entonces, para $x_0$ arbitrario, $x_1$ nos daría exactamente el valor $\alpha$ buscado.

## 9.4.2 Atracción local

Supongamos que hayamos construido una función $g$ que admita por punto fijo el cero $\alpha$ de $f$ que busquemos. Daremos una condición suficiente para que la sucesión de iterantes converja hacia $\alpha$.

**Teorema 9.4.1** *Supongamos que $g$ es una función definida y con derivada continua en (un conjunto que contenga) un intervalo de la forma $(\alpha - r_0, \alpha + r_0)$, $r_0 > 0$, con $g(\alpha) = \alpha$. Si*

$$|g'(\alpha)| < 1, \tag{9.7}$$

*existe $r$ positivo, $r < r_0$, tal que, para $x_0 \in [\alpha - r, \alpha + r]$, se puede construir la sucesión $x_{n+1} = g(x_n)$, $n = 0, 1, \ldots$, con elementos en $[\alpha - r, \alpha + r]$ y convergente hacia $\alpha$. Además, si los errores $e_n = x_n - \alpha$ son no nulos, los cocientes $e_{n+1}/e_n$ tienden hacia $g'(\alpha)$ cuando $n \to \infty$.*

Antes de demostrar el teorema son pertinentes algunos comentarios.

- Si un error $e_n$ es cero, entonces $x_n = \alpha$ y todos los iterantes ulteriores también coinciden con $\alpha$; naturalmente este es el caso que nunca tendremos la fortuna de encontrar en la práctica.

- La posibilidad de la construcción de la sucesión de iterantes viene garantizada por el teorema. Esta posibilidad no es obvia: en el Ejemplo 9.4.4 vimos una situación en que obtuvimos un $x_1$ fuera del dominio de definición de $f$ y no se pudo seguir con la iteración.

- El teorema es *local*: garantiza que para $x_0$ *suficientemente* próximo a $\alpha$ la iteración de punto fijo tiene éxito. En la aplicación práctica no conocemos de antemano quien es $\alpha$ y frecuentemente no podremos comprobar (9.7). Además tampoco sabemos el valor de $r$ y no podemos comprobar la hipótesis $x_0 \in [\alpha - r, \alpha + r]$.

*Demostración.* Por la condición (9.7) podemos tomar un número $M$ menor que 1 y mayor que $|g'(\alpha)|$. Dada la continuidad de $g'$, existe $r < r_0$ tal que, para $x \in [\alpha - r, \alpha + r]$, $|g'(x)| \leq M$.

Tomemos $x_0 \in [\alpha - r, \alpha + r]$ y supongamos que hemos calculado ya $x_1, \ldots,$ $x_n$ y que están en $[\alpha - r, \alpha + r]$. Entonces podemos hallar $x_{n+1} = g(x_n)$. Puesto que $g(\alpha) = \alpha$,

$$x_{n+1} - \alpha = g(x_n) - g(\alpha);$$

*esta es la igualdad crucial en el estudio de la iteración de punto fijo.* Usando el teorema de los incrementos finitos

$$x_{n+1} - \alpha = g'(\xi_n)(x_n - \alpha),$$

con $\xi_n$ un punto intermedio entre $x_n$ y $\alpha$, que por tanto está en $[\alpha - r, \alpha + r]$. Tomando valores absolutos

$$|x_{n+1} - \alpha| = |g'(\xi_n)|\,|x_n - \alpha| \leq M|x_n - \alpha| < |x_n - \alpha| \leq r,$$

de modo que también $x_{n+1} \in [\alpha - r, \alpha + r]$. Por inducción concluimos que se puede construir la sucesión de iterantes en $[\alpha - r, \alpha + r]$.

Por otro lado, acabamos de ver que $|e_{n+1}| \leq M|e_n|$ para cada $n$, luego

$$|e_n| \leq M|e_{n-1}| \leq M(M|e_{n-2}|) \leq \ldots \leq M^n|e_0|.$$

Cuando $n \to \infty$, tendremos $e_n \to 0$, es decir $x_n \to \alpha$.

Finalmente, si los $e_n$ son no nulos,

$$\frac{e_{n+1}}{e_n} = \frac{x_{n+1} - \alpha}{x_n - \alpha} = \frac{g(x_n) - g(\alpha)}{x_n - \alpha} \to g'(\alpha),$$

por definición de derivada. □

**Ejemplo 9.4.5** Consideremos la ecuación que resolvimos en las Tablas 9.1 y 9.2, usando ahora la iteración de punto fijo con $g(x) = 0.7 + 0.5 \sin x$. Partiendo de $x_0 = 0$, los resultados están en la Tabla 9.3. El cociente entre errores consecutivos se acerca a $0.5 \cos \alpha \approx 0.200$, es decir cada

Tabla 9.3: Iteración de punto fijo con $g(x) = 0.7+0.5\sin x$ para la ecuación $f(x) = x-0.5\sin x-0.7 = 0$

| $n$ | $x_n$ | $e_n = x_n - \alpha$ | $e_n/e_{n-1}$ |
|---|---|---|---|
| 0 | 0 | $-1.1580$ | |
| 1 | 0.700000 | $-0.4580$ | 0.395 |
| 2 | 1.022109 | $-0.1359$ | 0.297 |
| 3 | 1.126605 | $-0.0314$ | 0.231 |
| 4 | 1.151479 | $-0.0065$ | 0.208 |
| 5 | 1.156684 | $-0.0013$ | 0.203 |

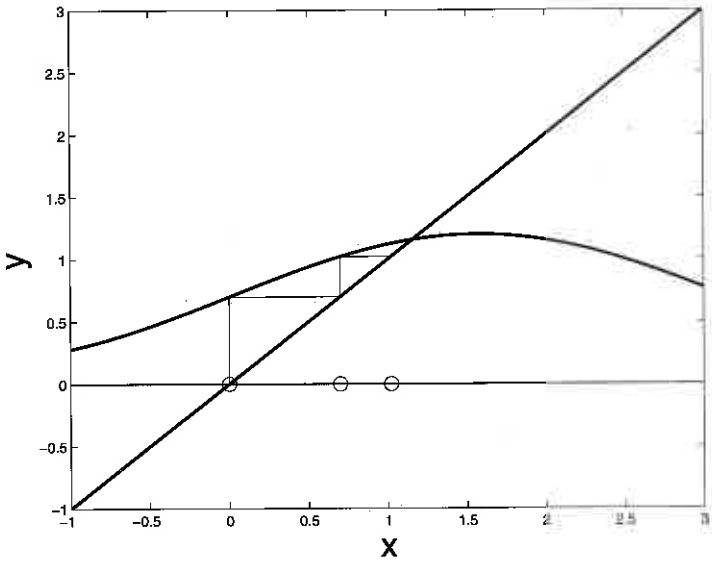


Figura 9.3: Iteración de punto fijo con $g(x) = 0.7 + 0.5\sin x$ para la ecuación $f(x) = x - 0.5\sin x - 0.7 = 0$

error es, asintóticamente, la quinta parte del precedente. Para este ejemplo, la iteración de punto fijo es más rápida que la bisección, pero más lenta que el método de la secante.

Es importante saber obtener los iterantes *gráficamente*. En la Figura 9.3 tenemos la gráfica de la función $y = g(x)$ y la bisectriz $y = x$ del primer y tercer cuadrantes. La abscisa (u ordenada) de la intersección es el punto fijo buscado. Desde el punto $x_0$ del eje de abscisas levantamos una *vertical* hasta intersecar la *gráfica* de $g$. La ordenada de la intersección es $x_1$. Por la intersección obtenida trazamos una *horizontal* hasta encontrar la *bisectriz* $y = x$, la abscisa del punto de encuentro es $x_1$. Estamos en condiciones de repetir el proceso "vertical a la gráfica de $g$, horizontal a la bisectriz." En la figura vemos que, con nuestra $g$, obtenemos una escalera que sube a la intersección de gráfica de $g$ y bisectriz. Con un $x_0$ mayor que el punto fijo hubiésemos obtenido una escalera *descendente*. □

**Ejemplo 9.4.6** Resolvamos una vez más la ecuación de Kepler (9.2), ahora con $e = 0.5$, $z = 2$. Partiendo de $x_0 = 2.5$, los resultados están en la Tabla 9.4. El cociente entre errores consecutivos se acerca a $0.5 \cos \alpha \approx -0.353$, es decir cada error es en magnitud, asintóticamente, el 35 % del precedente. El signo $-$ en el cociente de errores indica que a un error por defecto sigue uno por exceso y viceversa.

La Figura 9.4 da la marcha gráfica de la iteración. En vez de una escalera se obtiene algo parecido a una espiral. □

### 9.4.3 Caso de convergencia cuadrática

Hemos visto en la subsección precedente que la magnitud de $g'(\alpha)$ gobierna la rapidez de la convergencia de la iteración de punto fijo. El caso más favorable será aquel en que $g'(\alpha) = 0$; según el Teorema 9.4.1 entonces $e_{n+1}/e_n \to 0$, o, con la notación de Landau, $e_{n+1} = o(e_n)$, cuando $n \to \infty$. Aún más: $e_{n+1}$ es, asintóticamente, proporcional al *cuadrado* de $e_n$, tal y como vamos a establecer.

**Teorema 9.4.2** *Supongamos satisfechas las hipótesis del Teorema 9.4.1 y que además*

$$g'(\alpha) = 0, \tag{9.8}$$

*y exista $g''(\alpha)$. Entonces existe $r$ positivo, $r < r_0$ tal que si $x_0 \in [\alpha - r, \alpha + r]$ se puede construir la sucesión $x_{n+1} = g(x_n)$, $n = 0, 1, \ldots$, cuyos elementos están en $[\alpha - r, \alpha + r]$ y que converge hacia $\alpha$. Además, si los errores $e_n = x_n - \alpha$ son no nulos, los cocientes $e_{n+1}/e_n^2$ tienden hacia $(1/2)g''(\alpha)$ cuando $n \to \infty$.*

*Demostración.* La posibilidad de la construcción de los iterantes y la convergencia son consecuencias del Teorema 9.4.1, que se puede aplicar por tener menos hipótesis que el presente. Sólo hay que probar el aserto sobre el cociente de los errores.

Tabla 9.4: Iteración de punto fijo con $g(x) = 2 + 0.5 \sin x$ para la ecuación $f(x) = x - 0.5 \sin x - 2 = 0$

| $n$ | $x_n$ | $e_n = x_n - \alpha$ | $e_n/e_{n-1}$ |
|---|---|---|---|
| 0 | 2.5000 | +0.1458 | |
| 1 | 2.2992 | −0.0550 | −0.377 |
| 2 | 2.3731 | +0.0189 | −0.343 |
| 3 | 2.3475 | −0.0067 | −0.356 |

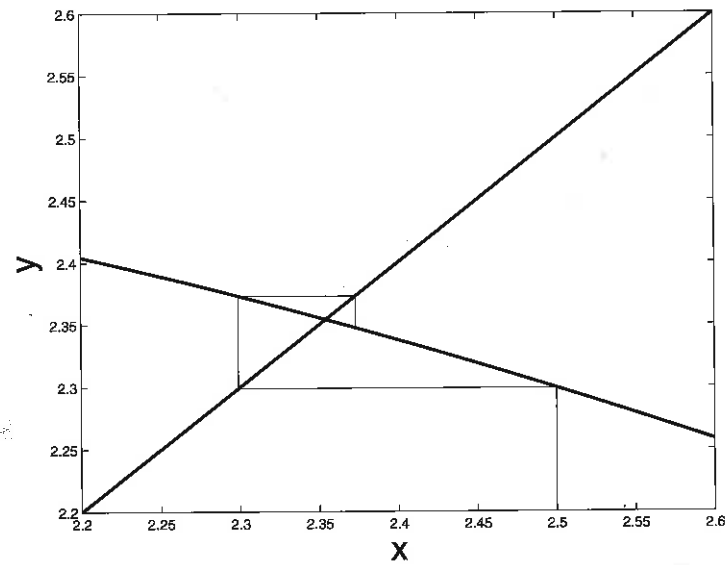


Figura 9.4: Iteración de punto fijo con $g(x) = 2 + 0.5 \sin x$ para la ecuación $f(x) = x - 0.5 \sin x - 2 = 0$

Tabla 9.5: Iteración de Newton para la ecuación $f(x) = x - 0.5\sin x + 2 = 0$

| $n$ | $x_n$ | $e_n = x_n - \alpha$ | $e_n/e_{n-1}^2$ |
|---|---|---|---|
| 0 | 0 | $-1.15 \times 10^{0}$ | |
| 1 | 1.400000000 | $+2.41 \times 10^{-1}$ | 0.18 |
| 2 | 1.173473865 | $+1.55 \times 10^{-2}$ | 0.21 |
| 3 | 1.158069898 | $+6.83 \times 10^{-5}$ | 0.29 |

Usando el polinomio de Taylor de segundo grado de $g$ en torno a $\alpha$ para aproximar $g(x_n)$ ,

$$x_{n+1} - \alpha = g(x_n) - g(\alpha) = g'(\alpha)(x_n - \alpha) + \frac{1}{2}g''(\alpha)(x_n - \alpha)^2 + o((x_n - \alpha)^2).$$

Aquí hemos empleado la caracterización analítica del polinomio de Taylor para escribir como $o(x_n - \alpha)$ la diferencia entre $g(x_n)$ y su aproximación. Invocando la hipótesis (9.8) y dividiendo entre $e_n^2 = (x_n - \alpha)^2$ (supuesto no nulo)

$$\frac{e_{n+1}}{e_n^2} = \frac{1}{2}g''(\alpha) + \frac{o(e_n^2)}{e_n^2}.$$

Por definición del símbolo de Landau el último sumando tiende a 0 y la prueba está completa. □

La convergencia que aquí tenemos se llama cuadrática (error asintóticamente proporcional al cuadrado del precedente); es mucho más rápida que la convergencia *lineal* suministrada por el Teorema 9.4.1 (error asintóticamente proporcional al precedente). [1]

**Ejemplo 9.4.7** Consideraremos por cuarta vez la ecuación resuelta en las Tablas 9.1, 9.2, 9.3. Usaremos de nuevo iteración de punto fijo con $x_0 = 0$, pero ahora la función de iteración es la debida a Newton,

$$g(x) = x - \frac{x - 0.5\sin x - 0.7}{1 - 0.5\cos x}.$$

Es obvio que el cero buscado $\alpha$ es punto fijo de $g$. Además un pequeño cálculo le mostrará, sin necesidad de conocer explícitamente $\alpha$, que tenemos la condición (9.8) que garantiza la convergencia cuadrática. Los resultados están en la Tabla 9.5. Las ventajas de este método sobre los otros tres probados en la misma ecuación son manifiestas. □

[1] Cuando $e_{n+1}/e_n \to 0$ hablamos de convergencia superlineal. Se demuestra que el método de la secante converge superlinealmente, pero sin alcanzar la convergencia cuadrática.

### 9.4.4 Puntos fijos repulsores

La condición (9.7) que garantiza la atracción local al punto fijo es (casi) necesaria.

**Teorema 9.4.3** *Supongamos que $g$ es una función definida y con derivada continua en (un conjunto que contenga) un intervalo de la forma $(\alpha - r_0, \alpha + r_0)$, $r_0 > 0$, con $g(\alpha) = \alpha$ y*

$$|g'(\alpha)| > 1.$$

*Las únicas sucesiones de la forma $x_{n+1} = g(x_n)$, $n = 0, 1, \ldots$, que convergen a $\alpha$ son aquellas cuyos términos son, de uno en adelante, iguales a $\alpha$.*

*Demostración.* Elijamos $M$ mayor que 1 y menor que $|g'(\alpha)|$. Por la continuidad de $g'$, existe un número $r$ positivo $r$, $r < r_0$, tal que, para $x \in [\alpha - r, \alpha + r]$, $|g'(x)| \geq M$. Supongamos que tenemos una sucesión de términos $\neq \alpha$ y que converge a $\alpha$. Para $n$ a partir de un cierto $n_0$, los $x_n$ pertenecen todos a $[\alpha - r, \alpha + r]$. Entonces, para $n \geq n_0$,

$$|x_{n+1} - \alpha| = |g(x_n) - g(\alpha)| = |g'(\xi_n)(x_n - \alpha)| \geq M|x_n - \alpha|,$$

de donde

$$|x_n - \alpha| \geq M|x_{n-1} - \alpha| \geq M^2|x_{n-2} - \alpha| \geq \ldots \geq M^{n-n_0}|x_{n_0} - \alpha|.$$

Haciendo tender $n$ a $\infty$, vemos que las distancias $|x_n - \alpha|$ se acercan a $\infty$, lo cual contradice la convergencia a $\alpha$ de la sucesión. □

Con el razonamiento que acabamos de usar en la demostración, se prueba que, en la situación del teorema, para $x_n$ próximo al punto fijo, $x_{n+1}$ dista de $\alpha$ más que $x_n$. El punto fijo se llama entonces *repulsor*.

### 9.4.5 Comportamiento global de la iteración

En las subsecciones precedentes hemos estudiado el comportamiento *local* de los iterantes. Local quiere decir cuando están cerca (en un entorno adecuado) del punto fijo. Veamos algo sobre el comportamiento *global*.

**Ejemplo 9.4.8** Consideremos la ecuación $x = -3 + 4\,\mathrm{sech}(x - 1)$, resuelta con iteración de punto fijo con $g(x) = -3 + 4\,\mathrm{sech}(x - 1)$. Hay (Fig. 9.5) tres raíces de la ecuación, o puntos fijos de $g$. Uno, llamémoslo $\alpha$, negativo, que es atractivo, aproximadamente $-2.8257$. Otro, llamémosleo $\beta$, en el intervalo $(0, 1)$, que es repulsivo (aproximadamente 0.4327) y, finalmente, el punto $x = 1$ que es atractivo (con convergencia cuadrática). En la figura estos puntos fijos están indicados con círculos.

¿Para qué condiciones iniciales $x_0$ hay convergencia de los iterantes al punto fijo repulsor $\beta$?

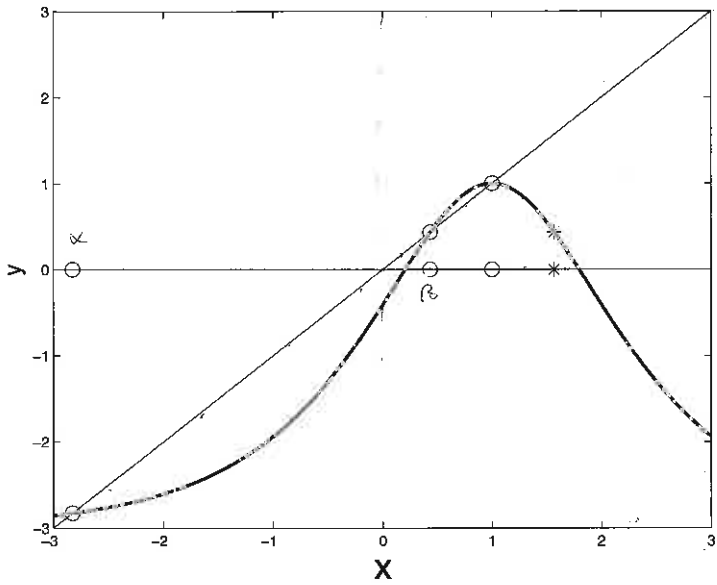


Figura 9.5: Iteración de punto fijo con $g(x) = -3 + 4\ \mathrm{sech}(x-1)$ para la ecuación $f(x) = -x - 3 + 4\ \mathrm{sech}(x-1) = 0$

- Con la condición inicial $x_0 = \beta$ son todos los iterantes constantemente iguales a $\beta$ y hay trivialmente convergencia.
- ¿Hay otras sucesiones de iterantes que converjan a $\beta$? Sabemos que los iterantes habrían de coincidir con $\beta$ de uno en adelante. Si $x_n$ es el último iterante $\neq \beta$, $g(x_n) = \beta$ y forzosamente $x_n = 2 - \beta$ ($2 - \beta$ es el punto simétrico de $\beta$ respecto de $x = 1$ marcado con una estrella en la figura). Ahora que sabemos que $x_n = 2 - \beta$, ¿quién será $x_{n-1}$? Respuesta: un punto que por $g$ se aplique en $2 - \beta > 1$; pero tales puntos no existen pues siempre $g \leq 1$. Solución al enigma: $x_n$ no tiene antecesor porque es $x_0$. Entonces $x_0 = 2 - \beta$ y se alcanza el punto fijo tras una sola iteración.

¿Para qué condiciones iniciales $x_0$ hay convergencia a $\alpha$?

- De acuerdo con la construcción gráfica de los iterantes, por lo pronto, para todas las del intervalo $I_1 = (-\infty, \beta)$. Para $x_0 \in (-\infty, \alpha)$ los iterantes crecen monótonamente a $\alpha$; para $x_0 \in (\alpha, \beta)$ decrecen monótonamente a $\alpha$.
- Pero también si $x_0 \in I_2 = (2 - \beta, \infty)$ la iteración converge a $\alpha$ *a pesar que esos* $x_0$ *disponen de un punto fijo atractivo* $x = 1$ *más cercano a ellos*. En efecto: partiendo de $x_0$ en $I_2$ tenemos $x_1 \in I_1$. Entonces y según acabamos de ver, la sucesión de iterantes que parte de $x_1$, es decir,

$$x_1, g(x_1) = x_2, g(x_2) = x_3, \ldots,$$

converge hacia $\alpha$. Evidentemente también la sucesión que empieza en $x_0$,

$$x_0, x_1, x_2, \ldots$$

convergerá a $\alpha$.

Finalmente para $x_0 \in (\beta, 2-\beta)$ hay convergencia hacia 1.

Los conjuntos abiertos $(-\infty, \beta) \cup (2-\beta, \infty)$ y $(\beta, 2-\beta)$ son los *dominios de atracción* de $\alpha$ y 1 respectivamente. Observe que uno de ellos no es conexo. En realidad un dominio de atracción puede tener una estructura topológica complicadísima. Los puntos repulsivos no tienen dominio de atracción. □

Del ejemplo debe aprender una cosa: una vez identificada una condición inicial $x_0$ que lleva a convergencia a un punto fijo (atractivo o no) $\alpha$, cada preimagen de $x_0$ por $g$, tomada como condición inicial, también conduce a convergencia a $\alpha$. Y lo mismo ocurre con las preimágenes de las preimágenes de $x_0$, las preimágenes de las preimágenes de las preimágenes de $x_0$, etc.

**Problema 9.4.1** Explique, usando alguno de los teoremas dados, el fracaso de la iteración de punto fijo de (9.6).

**Problema 9.4.2** Demuestre que una función continua en $[a, b]$ que toma valores en $[a, b]$ tiene, cuando menos, un punto fijo. (Indicación: si $g(a) = a$ ó $g(b) = b$ no hay nada que probar. En otro caso la función $x - g(x) \ldots$ )

**Problema 9.4.3** Sea $g$ una función definida en un intervalo y monótona creciente en sentido estricto, esto es $g(x) < g(y)$ para $x < y$. Supongamos que los iterantes $x_n$ están definidos desde $n = 0$ hasta $n = N > 1$ y que ninguno es un punto fijo de $g$. Pruebe que entonces la sucesión $x_0, x_1, x_2, \ldots, x_N$ es ó monótona creciente en sentido estricto ó monótona decreciente en sentido estricto.

**Problema 9.4.4** Sea $g$ una función definida en un entorno de un punto fijo de ella $\alpha$ y monótona decreciente en sentido estricto. Pruebe que $\alpha$ es el único punto fijo que $g$ tiene. Pruebe que si los iterantes $x_n$ están definidos desde $n = 0$ hasta $n = N > 1$ y no coinciden con $\alpha$ entonces dos iterantes consecutivos están a distinto lado de $\alpha$.

**Problema 9.4.5** *Puntos fijos con* $|g'(\alpha)| = 1$. Las funciones $g(x) = \sin x$ y $g(x) = \tan x$ tienen ambas el punto fijo $\alpha = 0$ y para ambas $g'(0) = 1$. Pruebe que, para $|x_0|$ suficientemente pequeño, con el seno la iteración de punto fijo converge mientras que con la tangente diverge. Por tanto el caso $|g'(\alpha)| = 1$ es dudoso, la convergencia o divergencia depende de los valores de las derivadas superiores de $g$.

Pruebe que si con $|g'(\alpha)| = 1$ hay convergencia al punto fijo entonces cada error es asintóticamente de la misma magnitud del anterior, por lo que la convergencia es lentísima y la sucesión generada carece de utilidad para hallar el punto fijo.

**Problema 9.4.6** *Puntos fijos de la aplicación logística.* Demuestre que para $\lambda \in [0, 4]$ la función $g(x) = \lambda x(1-x)$ aplica el intervalo $[0, 1]$ en sí mismo. (Indicación: represente gráficamente $g$.) Consideremos en lo sucesivo sólo valores $x \in [0, 1]$, $\lambda \in [0, 4]$. ¿Cuántos puntos fijos tiene $g$? Demuestre que el punto fijo $\alpha = 0$ es atractor si $\lambda < 1$ y repulsor para $\lambda > 1$. Demuestre que para $\lambda = 1$ cualquier $x_0 \in [0, 1]$ conduce a una sucesión de iterantes que converge a $\alpha = 0$ (haga una gráfica) y que cada error $e_n$ es asintóticamente igual al precedente. Estudie para qué valores de $\lambda$ el punto fijo $\neq 0$ es atractivo. (Solución: para $1 < \lambda < 3$.)

**Problema 9.4.7** *Convergencia cúbica.* Demuestre que si en el punto fijo $\alpha$ de $g$ es $g'(\alpha) = g''(\alpha) = 0$ y entonces los errores (supuestos no nulos) de la iteración de punto fijo verifican que $e_{n+1}/e_n^3$ se acerca a un límite. ¿Qué hipótesis sobre la derivabilidad de $g$ se necesitan? Esta convergencia se llama cúbica.

**Problema 9.4.8** ¿Cuántos puntos fijos tiene $g(x) = (1/2)x - x^3$ (haga una gráfica)? Halle un punto $\xi > 0$ con la propiedad $g(\xi) = -\xi$. ¿Qué le ocurre a la iteración de punto fijo para $x_0 \in (0, \xi)$? ¿Y para $x_0 = \xi$? ¿Y para $x_0 > \xi$? Los casos en que $x_0$ es negativo no necesitan ser discutidos: cambiar el signo de $x_0$ cambia el signo de todos los iterantes.

**Problema 9.4.9** Encuentre los puntos fijos de $(\pi/2)\sin x$. Para cada $x_0$ real la sucesión de iterantes converge a un punto fijo. Determine, en función de $x_0$, cuál es ese límite.

**Problema 9.4.10** *Dominio de atracción.* Sea $g(x)$ una función real que, por sencillez, supondremos definida para todo $x$ real. Sea $\alpha$ uno de los puntos fijos de $g$ que tenga la propiedad de que existe $r > 0$ tal que si $x_0 \in I = (\alpha - r, \alpha + r)$ entonces los iterantes de la iteración de punto fijo convergen a $\alpha$ (esto ocurre en particular si $g$ es derivable con continuidad y $|g'(\alpha)| < 1$). Pruebe que el conjunto de los $x_0$ para los que la iteración de punto fijo converge hacia $\alpha$ (el dominio de atracción de $\alpha$) es la reunión de las imágenes inversas sucesivas $g^{-1}(I)$, $g^{-1}(g^{-1}(I))$, ... (Indicación: pruebe que $x_0$ está en el dominio de atracción si existe un $n$ para el que $x_n \in I$.)

Pruebe que el dominio de atracción es un subconjunto abierto de la recta. Concluya que a cada punto $x_0$ para el que la iteración converge a $\alpha$ se le puede asociar un valor $\rho$ de suerte que todas las condiciones iniciales en $(x_0 - \rho, x_0 + \rho)$ conduzcan a convergencia a $\alpha$.

**Problema 9.4.11** *Dinámica de poblaciones. Modelo lineal.*[2] Llamemos $x_n$, $n = 0, 1, 2, \ldots$ al número de individuos de una población de una especie de insectos al cabo de $n$ años de empezar a observarla. Se supone que los insectos mueren al final del verano y nacen en primavera y que el número $x_{n+1}$ en una temporada es una función conocida $g(x_n)$ del número en la temporada anterior. Supondremos siempre que $g(0) = 0$ ¿por qué?

El modelo más sencillo es, como siempre, lineal, es decir $g(x) = \mu x$ con $\mu$ una constante positiva.

Pruebe que, si $\mu > 1$, el número de individuos crece exponencialmente con $n$ para cada $x_0 > 0$. ¿Cada cuántos años dobla la población su número?

Pruebe que, si $\mu < 1$, el número de individuos decrece exponencialmente hacia cero para cada $x_0 > 0$. ¿Cuántos años necesita la población para dividir entre dos su número?

Pruebe que $100(\mu - 1)$ representa el tanto por ciento en que la población aumenta de una temporada a la siguiente.

**Problema 9.4.12** *Dinámica de poblaciones. Modelo logístico.* En el modelo lineal del problema anterior, para $\mu > 1$ y $n$ suficientemente grande, $x_n$ supera cualquier número prescrito, lo cual evidentemente no se puede dar en la realidad. Para disponer de un modelo mejor, suponemos que el parámetro $\mu$ que gobierna el crecimiento no permanece constante sino que es de la forma

[2]En este problema y los siguientes se estudian sucesiones con $x_{n+1} = g(x_n)$ pero la finalidad del proceso *no* es encontrar el valor de los puntos fijos.

$\mu = \mu_0 - cx_n$ con $\mu_0$ y $c$ constante positivas; cuando la población es muy pequeña hay un factor de crecimiento $\mu \approx \mu_0$, y al crecer la población el factor de crecimiento va disminuyendo (por ejemplo porque hay individuos que no llegan a reproducirse por falta de alimentos). Tenemos entonces el llamado modelo logístico $x_{n+1} = g(x_n)$, $g(x) = (\mu_0 - cx)x$. El modelo sólo tiene sentido si $0 \leq x_n \leq \mu_0/c$, pues evidentemente $x_n \geq 0$ y si $x_n > \mu_0/c$ el factor de crecimiento sería negativo. Por tanto $\mu_0/c = x_{\max}$ es un tamaño máximo de la población. Pruebe que ha de ser $\mu_0 \leq 4$ para conseguir que $g(x) \in [0, x_{\max}]$ para $x \in [0, x_{\max}]$.

**Problema 9.4.13** *Continuación.* Pruebe que si $\mu_0 \leq 1$ entonces para cualquier $x_0 \in (0, x_{\max}]$ se tiene $\lim x_n = 0$, es decir los insectos tienden a desaparecer.

Pruebe que si $\mu_0 \in (1, 4]$ entonces hay un tamaño de la población $x_{\text{eq}} > 0$ que es de equilibrio, esto es, si un año $x_n = x_{\text{eq}}$ entonces en los años siguientes $x_{n+1} = x_{n+2} = \ldots = x_{\text{eq}}$.

Pruebe que para $\mu_0 \in (1, 3)$ y $x_0 \in (0, x_{\max}]$ el tamaño de la población, cuando $n \to \infty$, tiende a ser $x_{\text{eq}}$.

Pruebe que para $\mu_0 \in (3, 4]$ el tamaño de la población no tiende en general a ser el de equilibrio.

**Problema 9.4.14** *Continuación.* Experimente numéricamente hallando $x_n$, $n = 1, 2, \ldots, 50$, con varios valores de $x_0$ y $\mu_0$, tomando siempre $c = 10^{-6}\mu_0$ (lo cual fija el tamaño máximo de la población en un millón).

## 9.5 El método de Newton

El método de Newton (también llamado de Newton-Raphson) tiene analogía con el de la secante. Partimos de *una* aproximación $x_0$ (en la secante son dos) a la raíz buscada $\alpha$. Y en vez de construir una recta de interpolación de Lagrange para $f$ construimos una recta de Taylor, es decir la recta tangente $y = f(x_0) + f'(x_0)(x - x_0)$. El cero $x_1$ de esta recta se toma como nueva aproximación a la raíz (Fig. 9.6), se traza la tangente a $f$ en $x_1$, etc. *La idea de linealizar, esto es, sustituir funciones por sus tangentes para reemplazar problemas no lineales por lineales, es una de las más importantes en todas las matemáticas.*

Para esta iteración

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

de modo que los iterantes se calculan por la fórmula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}, \qquad n = 0, 1, \ldots \tag{9.9}$$

La aplicación de (9.9) no es siempre posible, pudiera ser que $x_n$ no estuviese en el dominio de definición de $f$, que $f$ no fuese derivable en $x_n$ o que $f'(x_n) = 0$.

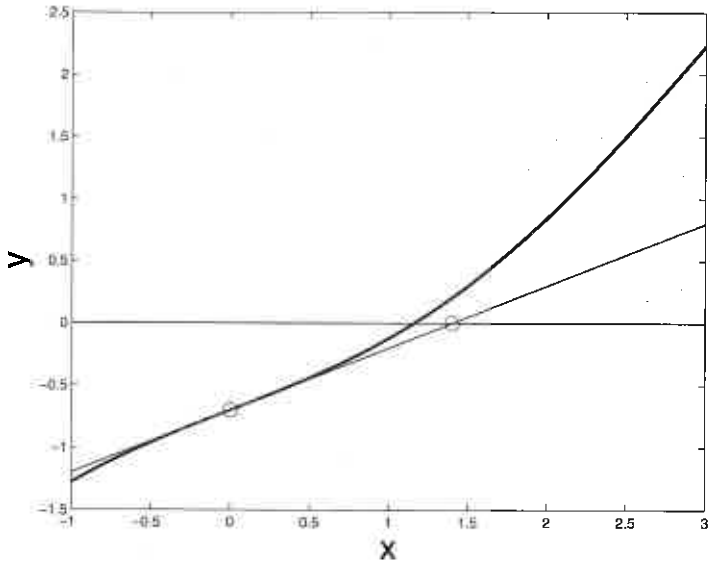


Figura 9.6: Método de Newton para $f(x) = x - 0.5 \sin x - 0.7 = 0$

En este último caso la tangente es horizontal y, salvo que $x_n$ sea ya un cero de $f$, no corta al eje de abscisas. Los criterios de parada son los mismos que en los métodos anteriores.

La fórmula (9.9) revela que la iteración de Newton, en independencia de su significado geométrico, no es nueva para nosotros. No es otra cosa que la iteración de punto fijo con función de iteración

$$g(x) = x - \frac{f(x)}{f'(x)}. \tag{9.10}$$

Notemos que efectivamente $f(\alpha) = 0$ conduce a $g(\alpha) = \alpha$. ¿Cuánto vale $|g'(\alpha)|$, cantidad que gobierna la convergencia de una iteración de punto fijo? Será

$$g'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2};$$

luego en un punto $\alpha$ con $f(\alpha) = 0$, $f'(\alpha) \neq 0$, $g$ tiene derivada nula, lo cual conduce a *convergencia cuadrática.*

**Teorema 9.5.1** *Supongamos que $f$ es una función definida y con dos derivadas continuas en (un conjunto que contenga) un intervalo de la forma $(\alpha - r_0, \alpha + r_0)$,*

*$r_0 > 0$, con $f(\alpha) = 0$, $f'(\alpha) \neq 0$. Supongamos además que existe $f'''(\alpha)$. Existe $r$ positivo, $r < r_0$, tal que para $x_0 \in [\alpha - r, \alpha + r]$ se puede construir la sucesión de iterantes de Newton (9.9) y converge hacia $\alpha$. Además si los errores $e_n = x_n - \alpha$ son no nulos, los cocientes $e_{n+1}/e_n^2$ tienen límite finito.*

*Demostración.* Basta con aplicar el Teorema 9.4.2 a la función (9.10). Para $x$ cerca de $\alpha$, $g$ está bien definida pues el denominador $f'(x)$ no se anula (no se anula en $x = \alpha$ por hipótesis y es una función continua). Además $g$ tiene derivada continua, pues $f$ tiene dos derivadas continuas y $g''(\alpha)$ existe por existir $f'''(\alpha)$. Finalmente $g'(\alpha) = 0$ según acabamos de comprobar más arriba. □

Un punto $\alpha$ con $f(\alpha) = 0$, $f'(\alpha) \neq 0$ se llama un cero *simple* de $f$. En el caso particular de que $f$ fuese un polinomio, el factor $x - \alpha$ figura una única vez en la descomposición de $f$ en producto de factores lineales. Así pues el método de Newton da convergencia cuadrática para ceros simples. Algunos se desorientan porque la hipótesis $f'(\alpha) \neq 0$ sobre la función cuyos ceros queremos hallar es la que conduce a $g'(\alpha) = 0$ para la función de iteración de la que nos valemos.

**Ejemplo 9.5.1** La iteración que usamos en el Ejemplo 9.4.7 es justamente la dada por el método de Newton. Allí ya notamos la convergencia cuadrática. □

El método de Newton para (9.1):

- No siempre converge a un cero de $f$ y acaso lo haga pero no al cero que buscábamos. Su éxito está sólo garantizado partiendo de una aproximación inicial cercana al cero buscado.

- Cuando converge lo hace mucho más rápidamente que los métodos de bisección y secante.

- Requiere la derivabilidad de $f$, lo cual no es un inconveniente serio pues la mayor parte de las funciones en las aplicaciones son derivables.

- Requiere valores de la derivada de $f$. Esto supone ó escribir una subrutina para $f'$ invocándola una vez por paso, ó acudir a una de las fórmulas de derivación numérica ya estudiadas. Por ejemplo podemos aproximar $f'(x_n)$ por $f[x_n, x_n + h]$ con $h$ pequeño y eso obliga a evaluar $f$ dos veces por cada paso, una en $x = x_n$, otra en $x = x_n + h$. Con cualquier implementación que elijamos, el trabajo a realizar por paso es superior al de los métodos de bisección y de la secante.

**Problema 9.5.1** *El método de Newton para raíces múltiples.* Sea $f$ una función definida en un intervalo abierto que contenga a un punto $\alpha$ con $f(\alpha) = 0$. Por sencillez supongamos que $f$ tiene derivadas de cualquier orden en todo punto y que $\alpha$ es el único cero de $f$. Demuestre que, si $p$ es un entero $\geq 1$, las dos condiciones siguientes son equivalentes. (i) $f$ y sus derivadas hasta la $p-1$ inclusive se anulan en $\alpha$, pero $f^{(p)}(\alpha) \neq 0$. (ii) $f$ es de la forma $f(x) = (x-\alpha)^p F(x)$ donde $F$ es una función continua con $F(\alpha) \neq 0$. Cuando se verifican estas condiciones se dice que $\alpha$ es un cero de multiplicidad $p$ de $f$.

En las condiciones anteriores, use (ii) para probar que la función de iteración del método de Newton tiene una derivada que en $\alpha$ vale $1 - 1/p$. Por ello, a menos que la raíz sea simple, el método de Newton converge pero la convergencia no es cuadrática.

Aplique el método de Newton a resolver $x^2 = 0$ a partir de $x_0$ dado. Halle la expresión de los $x_n$ y compruebe la convergencia lineal.

**Problema 9.5.2** Para la ecuación $f(x) = x^3 - x = 0$ vamos a determinar para que valores de $x_0$ la iteración de Newton converge a cada una de las tres raíces, $+1$, $0$, $-1$ (es decir vamos a contrar los *dominios de atracción* de las raíces).

Realice una figura con la gráfica de $f$, hallando el máximo y el mínimo de $f$.

Compruebe, usando la interpretación gráfica del método de Newton, que para $x_0 \in I_1 = (b_1, \infty)$, $b_1 = \sqrt{3}/3$, hay convergencia hacia $+1$. Por simetría, para $x_0 \in J_1 = (-\infty, -b_1)$ hay convergencia hacia $-1$.

Dibuje en la figura $I_2 = (-b_1, -b_2)$, donde $-b_2 < 0$ es el punto que el método de Newton lleva en $b_1$. Pruebe que $I_2$ está en el dominio de atracción de $+1$. Por simetría $J_2 = (b_2, b_1)$ está en el dominio de atracción de $-1$.

Dibuje ahora $I_3 = (b_3, b_2)$ donde $b_3$ es llevado por el método de Newton en $-b_2$ y pruebe que está en el dominio de $+1$, etc.

El dominio de atracción de $+1$ es la reunión de los intervalos abiertos disjuntos $I_1$, $I_2$, $I_3$, ... Los extremos de estos intervalos se acumulan a los puntos $\pm\gamma$ tales que el método de Newton lleva $\gamma$ en $-\gamma$ y viceversa. El dominio de $-1$ es simétrico del de $+1$. El dominio de 0 es $(-\gamma, \gamma)$. Empezado en los puntos $\pm b_n$ sólo se puede efectuar un número finito de iteraciones. En cualquier pequeño intervalo que contenga a $\gamma$ hay (i) puntos del dominio de $+1$, (ii) puntos del dominio de $-1$, (iii) puntos del dominio de 0, (iiii) puntos en los que la iteración acaba tras un número finito de pasos.

**Problema 9.5.3** *El método de Euler.* En el método de Newton $x_{n+1}$ es el cero del polinomio de Taylor de primer grado en $x_n$. En el método de Euler aproximamos $f$ por el polinomio de Taylor de segundo grado en $x_n$ y tomamos para $x_{n+1}$ *uno* de los ceros de esa parábola cuadrática.

Demuestre que esos ceros son

$$x_n + \frac{-f'(x_n) \pm \sqrt{(f'(x_n))^2 - 2f(x_n)f''(x_n)}}{f''(x_n)}.$$

¿Se quedaría usted con la raíz con $+$ o con la raíz con $-$? (Indicación: lo menos que podemos pedir al método es que $x_n = \alpha$ conduzca a $x_{n+1} = \alpha$.)

La fórmula que hemos encontrado tiene el inconveniente de que, para $x_n$ cerca de $\alpha$, exige substraer cantidades muy parecidas ¿cuáles? (Vea el Problema 1.5.1) Pruebe que $x_{n+1}$ puede

reescribirse, para evitar errores de redondeo, como

$$x_n - \frac{2}{1+\sqrt{1-2\frac{f(x_n)f''(x_n)}{(f'(x_n))^2}}}\frac{f(x_n)}{f'(x_n)};$$

esta es la expresión que tomamos como definición del método de Euler.

Pruebe que, para raíces simples y funciones $f$ suficientemente regulares, el método de Euler converge cúbicamente, es decir que $g'(\alpha) = g''(\alpha) = 0$. (Indicación: basta probar que el polinomio de Taylor de segundo grado de $g$ en torno a $\alpha$ es la constante $g(\alpha) = \alpha$. Este polinomio de Taylor de $g$ se construye a partir de los polinomios de Taylor de las partes de la fórmula. Por ejemplo, el polinomio de Taylor de grado dos de $x$ es $\alpha + (x - \alpha)$; el de $f(x)/f'(x)$ se obtiene dividiendo, etc.)

**Problema 9.5.4** *Métodos de Chebyshev y Halley* . Pruebe que las iteraciones siguientes, llamadas respectivamente de Chebyshev (1821–1894) y Halley (1656–1742), son cúbicas para ceros simples y funciones regulares

$$\begin{aligned} g(x) &= x - \left(1 + \frac{f(x)f''(x)}{2(f'(x))^2}\right)\frac{f(x)}{f'(x)}, \\ g(x) &= x - \frac{1}{1-\frac{f(x)f''(x)}{2(f'(x))^2}}\frac{f(x)}{f'(x)}. \end{aligned}$$

# Capítulo 10

# Solución de sistemas lineales

## 10.1 Introducción

Nos ocuparemos ahora de resolver sistemas lineales de ecuaciones de la forma

$$A\mathbf{x} = \mathbf{b}, \tag{10.1}$$

donde $A$ es una matriz real $d \times d$ conocida, $\mathbf{b}$ un vector (columna) conocido con $d$ componentes $b_i$ y $\mathbf{x}$ el vector (columna) que buscamos, con $d$ componentes $x_j$. En términos de cantidades escalares (10.1) se reescribe

$$\begin{array}{ccccccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1d}x_d & = & b_1, \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2d}x_d & = & b_2, \\ \vdots & & \vdots & & \cdots & & \vdots & \vdots & \vdots \\ a_{d1}x_1 & + & a_{d2}x_2 & + & \cdots & + & a_{dd}x_d & = & b_d; \end{array}$$

la notación vectorial es claramente preferible.

Si la matriz $A$ es invertible, entonces (10.1) posee una única solución $A^{-1}\mathbf{b}$ y, si no hace falta calcularla efectivamente, la discusión de (10.1) está concluida y no hay más que decir. Sin embargo, en muchísimas aplicaciones, es preciso resolver sistemas de la forma (10.1) *efectivamente*. De hecho, la resolución de ecuaciones lineales es la tarea que más frecuentemente realizan los ordenadores, toda vez que son numerosos los procesos de todo tipo que requieren como paso intermedio la resolución de sistemas lineales. Como consecuencia, la creación de métodos para resolver (10.1) es todavía hoy un campo de investigación muy activo. Nosotros nos limitaremos al método más importante, llamado de eliminación Gaussiana por haber sido Gauss el primero en describirlo sistemáticamente (aunque no el primero en utilizarlo, pues es antiquísimo).

## 10.2 La eliminación Gaussiana

### 10.2.1 Un ejemplo

Ilustraremos primero el método mediante un ejemplo.

**Ejemplo 10.2.1** Sea

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 5, \\ 6x_1 & + & 15x_2 & + & 19x_3 & + & 23x_4 & = & 30, \\ 8x_1 & + & 42x_2 & + & 60x_3 & + & 70x_4 & = & 98, \\ 12x_1 & + & 60x_2 & + & 1x_3 & + & 17x_4 & = & 144, \end{array} \tag{10.2}$$

el sistema por resolver. Eliminemos $x_1$ de las ecuaciones segunda, tercera y cuarta. A tal efecto restamos a la segunda ecuación 3 veces la primera, a la tercera 4 veces la primera y a la cuarta 6 veces las primera. Obtenemos

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 5, \\ & & 6x_2 & + & 7x_3 & + & 8x_4 & = & 15, \\ & & 30x_2 & + & 44x_3 & + & 50x_4 & = & 78, \\ & & 42x_2 & - & 23x_3 & - & 13x_4 & = & 114. \end{array} \tag{10.3}$$

Ahora eliminamos $x_2$ de las ecuaciones tercera y cuarta y para ello restamos a la tercera ecuación 5 veces la segunda y a la cuarta 7 veces la segunda. El resultado es

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 5, \\ & & 6x_2 & + & 7x_3 & + & 8x_4 & = & 15, \\ & & & & 9x_3 & + & 10x_4 & = & 3, \\ & & & - & 72x_3 & - & 69x_4 & = & 9. \end{array} \tag{10.4}$$

Finalmente eliminamos $x_3$ de la cuarta ecuación, restando a la cuarta ecuación $-8$ veces la tercera y obteniendo

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 5, \\ & & 6x_2 & + & 7x_3 & + & 8x_4 & = & 15, \\ & & & & 9x_3 & + & 10x_4 & = & 3, \\ & & & & & & 11x_4 & = & 33. \end{array} \tag{10.5}$$

Hemos concluido la *eliminación* de las incógnitas y llegado a un sistema equivalente a (10.2) pero *triangular superior*. Este sistema se resuelve inmediatamente por *sustitución regresiva*, es decir de la última ecuación se obtiene la última incógnita, llevando el valor de esta a la penúltima ecuación se obtiene la penúltima incógnita, etc. Así concluimos

$$\begin{aligned} x_4 &= \frac{33}{11}, \\ x_4 &= 3, \\ x_3 &= \frac{3 - 10 \times 3}{9}, \\ x_3 &= -3, \end{aligned}$$

$$
\begin{aligned}
x_2 &= \frac{15 - 7 \times (-3) - 8 \times 3}{6}, \\
x_2 &= 2, \\
x_1 &= \frac{5 - 3 \times 2 - 4 \times (-3) - 5 \times 3}{2}, \\
x_1 &= -2,
\end{aligned}
$$

y el sistema está resuelto. □

Observemos que hay dos etapas bien definidas.

- La *eliminación* de las incógnitas, pasando, para un sistema de dimensión $d$, por $d$ sistemas equivalentes; el primero es el de partida que deseamos resolver, el último es triangular superior.
- La solución del sistema triangular superior por *sustitución regresiva.*

A continuación describiremos cada una de las dos etapas para el sistema general (10.1).

### 10.2.2 Eliminación de las incógnitas

Para el sistema $d \times d$ (10.1) denotemos por

$$A^{(k)}\mathbf{x} = \mathbf{b}^{(k)}, \qquad k = 1, 2, \ldots, d,$$

los $d$ sistemas equivalentes. Comenzamos con $k = 1$, donde tenemos el sistema de partida,

$$A^{(1)} = A, \qquad \mathbf{b}^{(1)} = \mathbf{b}.$$

Pasamos del $k$-ésimo al $(k+1)$-ésimo, $k = 1, \ldots, d-1$, eliminando la incógnita $x_k$. *Note que $x_d$ no se elimina.*

El sistema $k$-ésimo tiene la siguiente estructura

$$
\begin{array}{rcl}
a_{11}^{(k)}x_1 + a_{12}^{(k)}x_2 + \cdots + a_{1k}^{(k)}x_k + \cdots + a_{1j}^{(k)}x_j + \cdots + a_{1d}^{(k)}x_d & = & b_1^{(k)}, \\
a_{22}^{(k)}x_2 + \cdots + a_{2k}^{(k)}x_k + \cdots + a_{2j}^{(k)}x_j + \cdots + a_{2d}^{(k)}x_d & = & b_2^{(k)}, \\
\ddots & \vdots & \vdots \\
a_{kk}^{(k)}x_k + \cdots + a_{kj}^{(k)}x_j + \cdots + a_{kd}^{(k)}x_d & = & b_k^{(k)}, \\
 & \vdots & \vdots \\
a_{ik}^{(k)}x_k + \cdots + a_{ij}^{(k)}x_j + \cdots + a_{id}^{(k)}x_d & = & b_i^{(k)}, \\
 & \vdots & \vdots \\
a_{dk}^{(k)}x_k + \cdots + a_{dj}^{(k)}x_j + \cdots + a_{dd}^{(k)}x_d & = & b_d^{(k)}.
\end{array}
$$

## 10.2 La eliminación Gaussiana

### 10.2.1 Un ejemplo

Ilustraremos primero el método mediante un ejemplo.

**Ejemplo 10.2.1** Sea

$$\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 5, \\ 6x_1 & + & 15x_2 & + & 19x_3 & + & 23x_4 & = & 30, \\ 8x_1 & + & 42x_2 & + & 60x_3 & + & 70x_4 & = & 98, \\ 12x_1 & + & 60x_2 & + & 1x_3 & + & 17x_4 & = & 144, \end{array} \tag{10.2}$$

el sistema por resolver. Eliminemos $x_1$ de las ecuaciones segunda, tercera y cuarta. A tal efecto restamos a la segunda ecuación 3 veces la primera, a la tercera 4 veces la primera y a la cuarta 6 veces las primera. Obtenemos

$$\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 5, \\ & & 6x_2 & + & 7x_3 & + & 8x_4 & = & 15, \\ & & 30x_2 & + & 44x_3 & + & 50x_4 & = & 78, \\ & & 42x_2 & - & 23x_3 & - & 13x_4 & = & 114. \end{array} \tag{10.3}$$

Ahora eliminamos $x_2$ de las ecuaciones tercera y cuarta y para ell restamos a la tercera ecuación 5 veces la segunda y a la cuarta 7 veces la segunda. El resultado es

$$\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 5, \\ & & 6x_2 & + & 7x_3 & + & 8x_4 & = & 15, \\ & & & & 9x_3 & + & 10x_4 & = & 3, \\ & & & - & 72x_3 & - & 69x_4 & = & 9. \end{array} \tag{10.4}$$

Finalmente eliminamos $x_3$ de la cuarta ecuación, restando a la cuarta ecuación $-8$ veces la tercera y obteniendo

$$\begin{array}{rcrcrcrcl} 2x_1 & + & 3x_2 & + & 4x_3 & + & 5x_4 & = & 5, \\ & & 6x_2 & + & 7x_3 & + & 8x_4 & = & 15, \\ & & & & 9x_3 & + & 10x_4 & = & 3, \\ & & & & & & 11x_4 & = & 33. \end{array} \tag{10.5}$$

Hemos concluido la *eliminación* de las incógnitas y llegado a un sistema equivalente a (10.2) pero *triangular superior*. Este sistema se resuelve inmediatamente por *sustitución regresiva*, es decir de la última ecuación se obtiene la última incógnita, llevando el valor de esta a la penúltima ecuación se obtiene la penúltima incógnita, etc. Así concluimos

$$\begin{aligned} x_4 &= \frac{33}{11}, \\ x_4 &= 3, \\ x_3 &= \frac{3 - 10 \times 3}{9}, \\ x_3 &= -3, \end{aligned}$$

$$
\begin{aligned}
x_2 &= \frac{15 - 7 \times (-3) - 8 \times 3}{6}, \\
x_2 &= 2, \\
x_1 &= \frac{5 - 3 \times 2 - 4 \times (-3) - 5 \times 3}{2}, \\
x_1 &= -2,
\end{aligned}
$$

y el sistema está resuelto. □

Observemos que hay dos etapas bien definidas.

- La *eliminación* de las incógnitas, pasando, para un sistema de dimensión $d$, por $d$ sistemas equivalentes; el primero es el de partida que deseamos resolver, el último es triangular superior.
- La solución del sistema triangular superior por *sustitución regresiva.*

A continuación describiremos cada una de las dos etapas para el sistema general (10.1).

### 10.2.2 Eliminación de las incógnitas

Para el sistema $d \times d$ (10.1) denotemos por

$$A^{(k)}\mathbf{x} = \mathbf{b}^{(k)}, \qquad k = 1, 2, \ldots, d,$$

los $d$ sistemas equivalentes. Comenzamos con $k = 1$, donde tenemos el sistema de partida,

$$A^{(1)} = A, \qquad \mathbf{b}^{(1)} = \mathbf{b}.$$

Pasamos del $k$-ésimo al $(k+1)$-ésimo, $k = 1, \ldots, d-1$, eliminando la incógnita $x_k$. *Note que $x_d$ no se elimina.*

El sistema $k$-ésimo tiene la siguiente estructura

$$
\begin{array}{rcl}
a_{11}^{(k)}x_1 + a_{12}^{(k)}x_2 + \cdots + a_{1k}^{(k)}x_k + \cdots + a_{1j}^{(k)}x_j + \cdots + a_{1d}^{(k)}x_d & = & b_1^{(k)}, \\
a_{22}^{(k)}x_2 + \cdots + a_{2k}^{(k)}x_k + \cdots + a_{2j}^{(k)}x_j + \cdots + a_{2d}^{(k)}x_d & = & b_2^{(k)}, \\
\ddots & \vdots & \vdots \\
a_{kk}^{(k)}x_k + \cdots + a_{kj}^{(k)}x_j + \cdots + a_{kd}^{(k)}x_d & = & b_k^{(k)}, \\
 & \vdots & \vdots \\
a_{ik}^{(k)}x_k + \cdots + a_{ij}^{(k)}x_j + \cdots + a_{id}^{(k)}x_d & = & b_i^{(k)}, \\
 & \vdots & \vdots \\
a_{dk}^{(k)}x_k + \cdots + a_{dj}^{(k)}x_j + \cdots + a_{dd}^{(k)}x_d & = & b_d^{(k)}.
\end{array}
$$

Aquí la incógnita $x_1$ ya sólo figura en la primera ecuación, la $x_2$ en las dos primeras, ..., la $x_{k-1}$ en las $k-1$ primeras, siendo $x_k$ la primera de las incógnitas que todavía está presente en todas las ecuaciones. Hemos de eliminar $x_k$ de las ecuaciones por debajo de la $k$-ésima, esto es de las ecuaciones con índices $i = k+1, \ldots, d$. Para ello, para $i = k+1, \ldots, d$, restaremos a la ecuación $i$-ésima $\ell_{ik}$ veces la $k$-ésima, siendo

$$\ell_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}};$$

con este valor de $\ell_{ik}$ se consigue que, en el nuevo sistema con índice $k+1$, la incógnita $x_k$ figure con coeficiente nulo en la ecuación $i$-ésima. La ecuación $k$-ésima, que multiplicada por números adecuados se sustrae de las siguientes, se llama *pivotal*. Los números $\ell_{ik}$ por los que se multiplica la fila pivotal son los *multiplicadores*. El número $a_{kk}^{(k)}$ se denomina *pivote* de la eliminación de $x_k$.

**Ejemplo 10.2.2** En el sistema (10.2) para eliminar $x_1$ la ecuación pivotal es la primera y el pivote vale 2. Hay tres multiplicadores $\ell_{21} = 6/2 = 3$, $\ell_{31} = 8/2 = 4$, $\ell_{41} = 12/2 = 6$. La ecuación pivotal se multiplica sucesivamente por 3, 4 y 6 y se sustrae a las ecuaciones segunda, tercera y cuarta de (10.2). Esto origina ceros en los coeficientes de $x_1$ en las filas segunda, tercera y cuarta de (10.3).

En el mismo ejemplo cuando eliminamos $x_2$ la ecuación pivotal es la segunda de (10.3) y el pivote vale 6. Hay dos multiplicadores $\ell_{23} = 30/6 = 5$, $\ell_{24} = 42/6 = 7$, que sirven para multiplicar a la ecuación pivotal antes de sustraersela a las ecuaciones tercera y cuarta de (10.3).

Para eliminar $x_3$ hay un solo multiplicador $\ell_{34} = -72/9 = -8$, porque $x_3$ sólo se elimina de la cuarta ecuación de (10.4). □

En definitiva las fórmulas para ir hallando los sucesivos sistemas son como sigue.

Para $k = 1, \ldots, d-1$,
  Para $i = 1, \ldots, k$, %Estas ecuaciones no cambian
    Para $j = 1, \ldots, d$,
      $a_{ij}^{(k+1)} = a_{ij}^{(k)}$,
    $b_i^{(k+1)} = b_i^{(k)}$,
  Para $i = k+1, \ldots, d$, %Estas ecuaciones cambian
    $\ell_{ik} = \dfrac{a_{ik}^{(k)}}{a_{kk}^{(k)}}$,
    Para $j = 1, \ldots, k$, %Ceros en $x_j$ ya eliminadas

$$
\begin{aligned}
& a_{ij}^{(k+1)} = 0, \\
\text{Para } j = k+1, \ldots, d, & \\
& a_{ij}^{(k+1)} = a_{ij}^{(k)} - \ell_{ik} a_{kj}^{(k)}, \\
& b_i^{(k+1)} = b_i^{(k)} - \ell_{ik} b_k^{(k)}.
\end{aligned} \tag{10.6}
$$

La línea (10.6) es la más importante; note cómo en ella hay tres índices:

- El $k$, que cuenta qué incógnita estamos eliminando, $k = 1, \ldots, d-1$.
- El $i$, que cuenta de qué ecuación la estamos eliminando, $i = k+1, \ldots, d$.
- El $j$, que va contando qué coeficiente de la ecuación $i$-ésima estamos transformando, $j = k+1, \ldots, d$.

Acabaremos esta subsección con una nota. En el proceso aquí descrito para eliminar la incógnita $t$ de

$$2t + 3u - 4v + 5w + 6x - 7y + 8z = 123, \tag{10.7}$$

usando como ecuación pivotal

$$3t + 6u - 9v + 12w + 6x - 7y + 8z = 789, \tag{10.8}$$

restamos a (10.7) la ecuación (10.8) multiplicada por 0.666666. En muchas descripciones elementales, se multiplicaría (10.7) por 3 y (10.8) por 2 y se restaría. El interés de este proceso alternativo elemental radica en que, cuando los coeficientes son enteros, evita introducir números no enteros. Para el cálculo automático, donde el ordenador no 've' si los coeficientes son enteros, el proceso alternativo carece de toda utilidad: usándolo la eliminación elemental de $t$ requiere catorce multiplicaciones y siete adiciones/substracciones, el proceso descrito aquí necesita una división (para el multiplicador), siete multiplicaciones y siete adiciones/substracciones (el que sean adiciones o substracciones depende del signo de los operandos: sumar dos números de signo distinto es una resta).

### 10.2.3 Sustitución regresiva

Es usual llamar $U$ a la matriz triangular superior del sistema triangular $A^{(d)}$ que se obtiene tras acabar de eliminar las incógnitas (U es la primera letra de 'upper,' superior en inglés). Llamando $\mathbf{c}$ al segundo miembro correspondiente $\mathbf{b}^{(d)}$,

hemos de resolver $U\mathbf{x} = \mathbf{c}$. Las fórmulas de la sustitución regresiva son

$$x_d = \frac{c_d}{u_{dd}},$$

$$\text{Para } i = d-1, \ldots, 1,$$

$$\tilde{c}_i = c_i - \sum_{j=i+1}^{d} u_{ij} x_j,$$

$$x_i = \frac{\tilde{c}_i}{u_{ii}}.$$

Observe cómo ahora sólo hay dos índices: $i$, $j$.

### 10.2.4 Aplicabilidad del método

El método que hemos descrito no siempre puede completarse. En efecto, si encontramos un pivote $a_{kk}^{(k)}$ nulo, $k = 1, 2, \ldots, d-1$, la fórmula para los multiplicadores que sirven para eliminar $x_k$ no tiene sentido. Además las fórmulas de la sustitución regresiva requieren que esos mismos pivotes $a_{kk}^{(k)} = u_{kk}$, $k = 1, 2, \ldots, d-1$ sean no nulos y además que $a_{dd}^{(d)} = u_{dd}$ sea no nulo.

**Ejemplo 10.2.3** Supongamos que en el sistema de partida $a_{11} = 0$ (lo cual, claro, puede ocurrir aunque $A$ sea invertible). Entonces nuestro proceso de restar múltiplos de la primera ecuación para eliminar $x_1$ no puede llevarse a cabo. □

**Ejemplo 10.2.4** Supongamos que el sistema de partida tiene la forma

$$\begin{array}{ccccc} 3x_1 & - & 12x_2 & + & \cdots \\ 9x_1 & - & 36x_2 & + & \cdots \\ \vdots & & \vdots & & \end{array}$$

Podemos eliminar $x_1$, tras lo cual el sistema queda (hay que restar a la segunda fila el triple de la primera)

$$\begin{array}{ccccc} 3x_1 & - & 12x_2 & + & \cdots \\ 0x_1 & - & 0x_2 & + & \cdots \\ \vdots & & \vdots & & \end{array}$$

Ahora no podemos eliminar $x_2$. Note que el elemento $(2, 2)$ no era nulo en la matriz de partida, sino que se ha anulado en el curso de la eliminación. □

El siguiente resultado nos dice cuándo podremos completar el método de eliminación Gaussiana. Dada una matriz $A$ de tamaño $d \times d$, su *submatriz principal* $k$-ésima, $k = 1, 2, \ldots, d$, es, por definición, la matriz $k \times k$ formada por las primeras $k$ filas y columnas de $A$.

**Teorema 10.2.1** *Sea $A$ una matriz $d \times d$ cuyas submatrices menores principales de orden $k$, $k = 1, 2, \ldots, d$, sean invertibles (en particular $A$ es invertible). Entonces el método de eliminación Gaussiana para (10.1) anteriormente descrito puede completarse sin encontrar ningún divisor nulo.*

*Recíprocamente si, dado el sistema (10.1), el método de eliminación Gaussiana que hemos descrito puede completarse sin encontrar divisores nulos, entonces todas las submatrices menores principales de $A$ (incluyendo la propia $A$) son invertibles.*

*Demostración.* Demostraremos el primer resultado; la prueba del recíproco es, en lo esencial, la misma. Podemos eliminar $x_1$ toda vez que $a_{11}^{(1)} = a_{11}$ no puede anularse pues en otro caso la submatriz principal de orden $1 \times 1$ de $A$ no sería invertible. Razonemos por inducción y supongamos que ya hemos podido eliminar $x_1, x_2, \ldots, x_k$ ($k \leq d-1$) al haber sido los correspondientes pivotes no nulos. Consideremos la matriz $A^{(k+1)}$ del sistema que se obtiene tras eliminar $x_k$ y fijémonos en su submatriz principal $(k+1) \times (k+1)$ que llamaremos $B_{k+1}$. De un lado, $B_{k+1}$ es una matriz triangular superior cuyo determinante será igual al producto de los elementos diagonales

$$\det(B_{k+1}) = a_{11}^{(k+1)} a_{22}^{(k+1)} \cdots a_{k+1,k+1}^{(k+1)} = a_{11}^{(1)} a_{22}^{(2)} \cdots a_{k+1,k+1}^{(k+1)}. \tag{10.9}$$

De otra parte, $\det(B_{k+1})$ debe coincidir con el determinante *no nulo por hipótesis* de la submatriz $A_{k+1}$ menor principal de orden $k+1$ de la matriz de partida $A$: en efecto $B_{k+1}$ se ha obtenido a partir de $A_{k+1}$ restando a filas múltiplos de otras, proceso que deja invariable el determinante. Combinando (10.9) con $\det(B_{k+1}) \neq 0$ tenemos $a_{k+1,k+1}^{(k+1)} \neq 0$. Si $k+1 < d$ podemos pasar a eliminar $x_{k+1}$, usando el pivote $a_{k+1,k+1}^{(k+1)}$; si $k+1 = d$ hemos probado que $u_{dd} \neq 0$. □

La conclusión es que las hipótesis bajo las cuales la eliminación de Gauss que hemos descrito da resultado son *más restrictivas* que la hipótesis de invertibilidad de $A$ bajo la cual el sistema (10.1) posee una única solución. Esta deficiencia notable de la eliminación Gaussiana se subsana mediante el *pivotaje*, proceso que estudiaremos más adelante.

**Problema 10.2.1** Resuelva por eliminación Gaussiana, indicando los valores de los multiplicadores y de los pivotes,

$$\begin{array}{rcrcrcrcl} 2x_1 & + & 2x_2 & + & 2x_3 & + & 2x_4 & = & 12, \\ 4x_1 & + & 6x_2 & + & 6x_3 & + & 6x_4 & = & 34, \\ 6x_1 & + & 14x_2 & + & 16x_3 & + & 16x_4 & = & 82, \\ 2x_1 & + & 2x_2 & + & 12x_3 & + & 14x_4 & = & 42. \end{array}$$

**Problema 10.2.2** Resuelva por eliminación Gaussiana los sistemas que siguen. En cada caso, indique los valores de los multiplicadores y de los pivotes. Si se fija, puede usar en el segundo sistema la mayor parte de los cálculos hechos para el primero.

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 2x_2 & + & 2x_3 & + & 2x_4 & = & 20, \\ 4x_1 & + & 6x_2 & + & 6x_3 & + & 6x_4 & = & 58, \\ 6x_1 & + & 14x_2 & + & 16x_3 & + & 16x_4 & = & 146, \\ & & & & 10x_3 & + & 12x_4 & = & 78, \end{array}$$

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 2x_2 & + & 2x_3 & + & 2x_4 & = & 0, \\ 4x_1 & + & 6x_2 & + & 6x_3 & + & 6x_4 & = & 2, \\ 6x_1 & + & 14x_2 & + & 16x_3 & + & 16x_4 & = & 10, \\ & & & & 10x_3 & + & 12x_4 & = & 12. \end{array}$$

**Problema 10.2.3** Considere el sistema

$$\begin{array}{rcrcrcrcr} 2x_1 & + & 2x_2 & + & 2x_3 & + & 2x_4 & = & 20, \\ 4x_1 & + & 6x_2 & + & 6x_3 & + & 6x_4 & = & 58, \\ 6x_1 & + & 14x_2 & + & 16x_3 & + & 16x_4 & = & 146, \\ 2x_1 & + & 2x_2 & + & 12x_3 & + & 12x_4 & = & 90, \end{array}$$

e intente resolverlo por eliminación Gaussiana. Concluya que la matriz tiene rango tres, que sus tres primeras filas son independientes pero la cuarta es combinación lineal de las tres primeras y que las tres primeras columnas son independientes pero la cuarta es combinación lineal de las tres primeras. Concluya también que el sistema es compatible y que se puede fijar arbitrariamente el valor de $x_4$. Finalmente halle la solución que tiene $x_4 = 4$.

## 10.3 Factorización LU

El material presentado hasta ahora podría bastar para resolver en la práctica sistemas lineales (de hecho Vd. ha debido ser capaz de resolver los sistemas propuestos en los problemas anteriores). Sin embargo para entender completamente la eliminación Gaussiana y sus variantes, tal como se implementan actualmente, es preciso profundizar más en lo anterior, dando una nueva visión de los procesos ya presentados.

### 10.3.1 La eliminación de incógnitas, proceso matricial

Comencemos denotando, para $k = 1, 2, \ldots, d-1$,

$$M^{(k)} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & -\ell_{k+1,k} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & -\ell_{k+2,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & -\ell_{d,k} & \cdots & 1 \end{bmatrix}; \tag{10.10}$$

esta matriz solo difiere de la matriz unidad en la $k$-ésima columna.

**Ejemplo 10.3.1** Para el Ejemplo 10.2.2,

$$M^{(1)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -4 & 0 & 1 & 0 \\ -6 & 0 & 0 & 1 \end{bmatrix}, \qquad M^{(2)} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & -5 & 1 & 0 \\ 0 & -7 & 0 & 1 \end{bmatrix}. \tag{10.11}$$

¿Cómo es $M^{(3)}$? □

Al transformar un vector $d$-dimensional $\mathbf{v}$ por $M^{(k)}$ y formar $\mathbf{w} = M^{(k)}\mathbf{v}$ las componentes $v_i$ con $i = 1, 2, \ldots, k$ quedan inalteradas, mientras que, para $i = k+1, k+2, \ldots, d$, $w_i = v_i - \ell_{ik}v_k$. Por la misma razón, si $C$ es una matriz con $d$ filas y un número cualquiera de columnas, las filas primera a $k$-ésima de $M^{(k)}C$ coinciden con las de $C$, mientras que, para $i = k+1, k+2, \ldots, d$, la fila $i$-ésima de $M^{(k)}C$ se obtiene restando a la $i$-ésima fila de $C$ el producto de $\ell_{ik}$ por la fila $k$-ésima de $C$. Por consiguiente

$$A^{(k+1)} = M^{(k)}A^{(k)}, \qquad \mathbf{b}^{(k+1)} = M^{(k)}\mathbf{b}^{(k)},$$

o, en palabras, *cada sistema* $A^{(k+1)}\mathbf{x} = \mathbf{b}^{(k+1)}$ *puede obtenerse del precedente* $A^{(k)}\mathbf{x} = \mathbf{b}^{(k)}$ *multiplicando éste por* $M^{(k)}$.

**Ejemplo 10.3.2** La matriz y segundo miembro de (10.3) se obtienen de los de (10.2) multiplicando por la matriz (10.11).

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -4 & 0 & 1 & 0 \\ -6 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 4 & 5 \\ 6 & 15 & 19 & 23 \\ 8 & 42 & 60 & 70 \\ 12 & 60 & 1 & 17 \end{bmatrix} = \begin{bmatrix} 2 & 3 & 4 & 5 \\ 0 & 6 & 7 & 8 \\ 0 & 30 & 44 & 50 \\ 0 & 42 & -23 & -13 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ -3 & 1 & 0 & 0 \\ -4 & 0 & 1 & 0 \\ -6 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 \\ 30 \\ 98 \\ 144 \end{bmatrix} = \begin{bmatrix} 5 \\ 15 \\ 78 \\ 114 \end{bmatrix} \square$$

### 10.3.2 La factorización implícita en la eliminación

El siguiente resultado es importante.

**Teorema 10.3.1** *Dado el sistema (10.1), supongamos que se puede llevar a cabo la eliminación Gaussiana descrita en la sección anterior. Si $U\mathbf{x} = \mathbf{c}$ es el sistema triangular superior resultante de eliminar las incógnitas y $L$ es la matriz*

$$L = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ \ell_{21} & 1 & 0 & \cdots & 0 & 0 \\ \ell_{31} & \ell_{32} & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \ell_{d1} & \ell_{d,2} & \ell_{d,3} & \cdots & \ell_{d,d-1} & 1 \end{bmatrix}, \tag{10.12}$$

*entonces*

$$A = LU, \qquad \mathbf{b} = L\mathbf{c}. \tag{10.13}$$

*Demostración.* Según el resultado de la subsección precedente

$$\begin{aligned} U &= A^{(d)} \\ &= M^{(d-1)}A^{(d-1)} \\ &= M^{(d-1)}M^{(d-2)}A^{(d-2)} \\ &\cdots \\ &= M^{(d-1)}M^{(d-2)}\cdots M^{(1)}A^{(1)} \\ &= M^{(d-1)}M^{(d-2)}\cdots M^{(1)}A, \end{aligned}$$

y

$$\begin{aligned} \mathbf{c} &= \mathbf{b}^{(d)} \\ &= M^{(d-1)}\mathbf{b}^{(d-1)} \\ &= M^{(d-1)}M^{(d-2)}\mathbf{b}^{(d-2)} \\ &\cdots \\ &= M^{(d-1)}M^{(d-2)}\cdots M^{(1)}\mathbf{b}^{(1)} \\ &= M^{(d-1)}M^{(d-2)}\cdots M^{(1)}\mathbf{b}, \end{aligned}$$

de manera que basta probar que

$$L = \left[M^{(d-1)} M^{(d-2)} \cdots M^{(1)}\right]^{-1},$$

es decir

$$L = \left[M^{(1)}\right]^{-1} \left[M^{(2)}\right]^{-1} \cdots \left[M^{(d-1)}\right]^{-1}. \tag{10.14}$$

Comencemos calculando las inversas en el segundo miembro de (10.14). Para $k = 1, 2, \ldots, d-1$,

$$\left[M^{(k)}\right]^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \ell_{k+1,k} & \cdots & 0 \\ 0 & 0 & 0 & \cdots & \ell_{k+2,k} & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & \cdots & \ell_{d,k} & \cdots & 1 \end{bmatrix}; \tag{10.15}$$

esto se demuestra como sigue. Observamos arriba que para formar $\mathbf{w} = M^{(k)}\mathbf{v}$ dejamos las componentes hasta la $k$ inalteradas y, para las demás, $w_i = v_i - \ell_{ik} v_k$, es decir $v_i = w_i + \ell_{ik} v_k = w_i + \ell_{ik} w_k$. Entonces para invertir la acción de $M^{(k)}$ basta cambiar el signo a los $\ell_{ik}$.

Ahora que conocemos la expresión de las inversas que figuran en (10.14), demostrar que su producto es la matriz $L$ en (10.12) es un sencillo ejercicio de multiplicación. □

La matriz triangular inferior $L$ en (10.12) se llama así por ser L la primera letra de *lower*, inferior en inglés.

Estudiemos el significado de este teorema. Pongámonos en la situación de tener que resolver el sistema (10.1). E imaginemos que alguien nos da dos matrices $L$ (triangular inferior con unos en la diagonal) y $U$ (triangular superior), tales que $A = LU$. Esta información nos sería muy útil: resolver $A\mathbf{x} = \mathbf{b}$ sería resolver $LU\mathbf{x} = \mathbf{b}$ y, llamando $\mathbf{c} = U\mathbf{x}$, tendríamos que resolver primero $L\mathbf{c} = \mathbf{b}$ para hallar $\mathbf{c}$ y, hallado este, resolver $U\mathbf{x} = \mathbf{c}$ para obtener la solución $\mathbf{x}$. Gracias a la información $A = LU$, hemos pasado de tener que resolver un sistema a tener que resolver dos *pero con la gran ventaja de que ambos tienen matriz triangular*. El sistema con matriz triangular superior se resuelve por sustitución regresiva como ya sabemos; el sistema con matriz triangular inferior se soluciona por *sustitución*

*progresiva:* la primera incógnita se saca de la primera ecuación y se lleva a la segunda ecuación para obtener la segunda incógnita, etc.

¿Cómo se relaciona esto con la eliminación de Gauss?

- De acuerdo con el Teorema 10.3.1 (vea (10.13)), al eliminar las incógnitas nos estamos procurando nosotros mismos la matriz triangular inferior $L$ y la matriz triangular superior $U$ con $A = LU$. La matriz $U$ la encontrábamos en la sección anterior al concluir el proceso de eliminación de incógnitas. La matriz $L$ no la encontramos abiertamente en la sección anterior, pero sí de modo encubierto toda vez que para conocerla basta conocer los valores de sus elementos no diagonales y éstos no son otros que los multiplicadores que usábamos en la eliminación.

- Por otro lado, (10.13) también muestra que el segundo miembro $\mathbf{c}$ que obtuvimos en la eliminación Gaussiana tras eliminar las incógnitas satisface $L\mathbf{c} = \mathbf{b}$. En otra palabras, al ir operando en el segundo miembro estábamos resolviendo, sin ser conscientes de ello, el sistema triangular $L\mathbf{c} = \mathbf{b}$.

En resumen y, aunque no nos diéramos cuenta en su momento, la eliminación Gaussiana involucra tres procesos:

- *Factorización.* Escribir la matriz $A$ como producto $LU$ de una $L$ triangular inferior con unos en la diagonal y otra $U$ triagular superior.

- *Sustitución progresiva* para resolver un sistema triangular con matriz $L$.

- *Sustitución regresiva* para resolver un sistema con matriz $U$.

### 10.3.3 Factorizaciones LU de una matriz

Dada una matriz $A$, una factorización LU de ella es, por definición, un par de matrices $L$, triangular inferior con unos en la diagonal, y $U$ triangular superior con $A = LU$. Acabamos de estudiar cómo el éxito de la eliminación Gaussiana se debe que encuentra una factorización LU de la matriz, lo que reduce el problema a resolver dos más fáciles. ¿Además de la factorización LU que encontramos por eliminación Gaussiana, cuando ésta se puede llevar a cabo, existirán otras, quizá más ventajosas? Cuando no se puede llevar a cabo la eliminación Gaussiana por ser singular una de las submatrices menores principales de A ¿existirá todavía alguna factorización LU de la que valernos para resolver (10.1)? La respuesta a ambas preguntas es negativa como muestra el siguiente resultado.

**Teorema 10.3.2** *Sea $A$ una matriz cuadrada, real e invertible. Entonces $A$ tiene factorización LU si y solamente si todas las submatrices menores principales de $A$ son invertibles. Si $A$ tiene factorización LU la factorización es única y coincide con la que se obtiene por eliminación Gaussiana al resolver sistemas de la forma $A\mathbf{x} = \mathbf{b}$.*

*Demostración.* Si las matrices menores principales son todas invertibles, entonces el Teorema 10.2.1 nos garantiza que podemos llevar a cabo la eliminación Gaussiana en los sistemas de la forma $A\mathbf{x} = \mathbf{b}$ y el Teorema 10.3.1 muestra que la eliminación nos dará una factorización LU.

Demostremos que, recíprocamente, la invertibilidad de las matrices menores es *necesaria* para la existencia de una factorización. Supongamos que la factorización existe, es decir $A = LU$, con lo que $\det(A) = \det(L)\det(U)$. El determinante de $A$ es, por hipótesis, $\neq 0$, el de $U$ es $u_{11}u_{22}\cdots u_{dd}$ y el de $L$ es 1. Por eso, los $u_{ii}$ son no nulos. Ahora para $1 \leq k \leq d$ partamos cada una de las matrices $A$, $L$ y $U$ en cuatro bloques, separando en cada una las $k$ primeras filas/columnas:

$$\begin{bmatrix} A_k & * \\ * & * \end{bmatrix} = \begin{bmatrix} L_k & 0 \\ * & * \end{bmatrix} \begin{bmatrix} U_k & * \\ 0 & * \end{bmatrix}$$

($A_k$, $L_k$, $U_k$ son entonces las correspondientes submatrices principales de orden $k$, las estrellas representan diversas submatrices de dimensiones adecuadas y los ceros son submatrices nulas). Efectuando el producto por bloques es $A_k = L_kU_k$ y, tomando determinantes,

$$\det(A_k) = \det(L_k)\det(U_k) = 1 \times u_{11}\cdots u_{kk} \neq 0.$$

Así $A$ tiene submatrices principales invertibles.

Veamos ahora que la factorización, de existir, es única. Supongamos que

$$A = L_1U_1 = L_2U_2.$$

Tanto $L_1$ como $U_2$ son invertibles (la primera tiene determinante 1, y la segunda —según vimos— determinante igual al de $A$). Entonces

$$U_1U_2^{-1} = L_1^{-1}L_2;$$

aquí el primer miembro es triangular superior (vea el Problema 10.3.3) y el segundo triangular inferior con diagonal de unos, como ambos coinciden ambos son triangulares inferiores y superiores y con diagonal de unos, es decir, ambos son la identidad. Se sigue que $U_1 = U_2$ y $L_1 = L_2$ y la factorización es única.

Por último, hemos visto que la condición para que la eliminación Gaussiana se pueda llevar a cabo (invertibilidad de las submatrices menores principales) es la misma que la condición para que la factorización exista. Por ello, cuando la factorización existe se puede hallar por eliminación Gaussiana. □

**Problema 10.3.1** Tras haber resuelto los Problemas 10.2.1 y 10.2.2 por eliminación Gaussiana, escriba para cada sistema las matrices $L$ y $U$ y el vector $\mathbf{c}$ y compruebe que $A = LU$, $\mathbf{b} = L\mathbf{c}$.

**Problema 10.3.2** Escriba al azar una matriz $U$ triangular superior invertible de dimensión $4 \times 4$ y una matriz $L$ triangular inferior $4 \times 4$ con diagonal de unos. Multiplíquelas y llame $A$ a su producto. Efectúe la eliminación Gaussiana en $A$ y compruebe que vuelve a los factores de partida.

**Problema 10.3.3** *Matrices triangulares.* Demuestre que el producto de dos matrices cuadradas triangulares superiores es también triangular superior. Demuestre que el producto de dos matrices triangulares superiores con diagonales de unos es una matriz triangular superior con diagonal de unos. Demuestre que la inversa de una matriz triangular superior invertible es triangular superior. Demuestre que la inversa de una matriz triangular superior con diagonal de unos es triangular superior y tiene diagonal de unos. Repita después cambiando superior por inferior.

**Problema 10.3.4** *La matriz $M^{(k)}$ escrita en términos de productos exteriores.* Consideremos la matriz $M^{(k)}$ en (10.10) y denotemos por $\mathbf{e}_k$ al $k$-ésimo vector coordenado (todas las componentes nulas excepto la $k$ que es uno) y por $\mathbf{l}_k$ al vector $[0, \ldots, 0, \ell_{k+1,k}, \ell_{k+2,k}, \ldots, \ell_{d,k}]^T$ (la transposición hace falta pues los vectores son siempre columna y habíamos escrito una fila). Pruebe que $M^{(k)} = I - \mathbf{l}_k\mathbf{e}_k^T$. Fíjese en la construcción que asocia a cada par de vectores $d$-dimensionales $\mathbf{a}$, $\mathbf{b}$ la matriz $\mathbf{a}\mathbf{b}^T$ (de dimensión $d \times d$) llamada producto exterior de ambos; aparece frecuentemente en álgebra lineal. No hay que confundir el producto exterior, que es una matriz, con el número real $\mathbf{a}^T\mathbf{b}$, producto interno de $\mathbf{a}$ y $\mathbf{b}$. Identifique la imagen y el núcleo de la transformación lineal de matriz $\mathbf{a}\mathbf{b}^T$.

**Problema 10.3.5** *Inversa de $M^{(k)}$.* Escriba $M^{(k)} = I - \mathbf{l}_k\mathbf{e}_k^T$ como en el problema anterior y multiplique por $I + \mathbf{l}_k\mathbf{e}_k^T$ para concluir que la inversa de $M^{(k)}$ es la matriz $I + \mathbf{l}_k\mathbf{e}_k^T$, cosa que ya demostramos en (10.15). Use la representación que acaba de obtener de la inversa de $M^{(k)}$ para probar (10.14).

## 10.4 Implementación

En esta sección estudiaremos cómo implementar prácticamente en un ordenador la eliminación Gaussiana que describimos en la Sección 10.2. Para ello conviene escribir tres subrutinas diferentes, que estudiaremos en cada una de las subsecciones que siguen.

### 10.4.1 Subrutina de factorización

Esta subrutina toma como dato la matriz $A$ y devuelve al salir las matrices $L$ y $U$. El único almacenamiento que se necesita es una matriz $d \times d$ que al entrar a la subrutina contiene los elementos de $A$ y al salir los de $L$ y los de $U$.

Tal como vimos en la Sección 10.2.2, se empieza calculando el multiplicador $\ell_{21} = a_{21}/a_{11}$ que sirve para eliminar $x_1$ de la segunda ecuación. Como el valor de $a_{21}$ no se vuelve a necesitar es posible almacenar $\ell_{21}$ en la posición en memoria antes ocupada por $a_{21}$. Esto es, sobreescribimos $a_{21} \leftarrow a_{21}/a_{11}$.

A continuación hallamos $a_{22}^{(2)}, a_{23}^{(2)}, \ldots, a_{2d}^{(2)}$ mediante la fórmula (10.6); estos valores se sobreescriben en las posiciones de memoria antes ocupadas respectivamente por $a_{22}, a_{23}, \ldots, a_{2d}$.

Después hallamos $\ell_{31} = a_{31}/a_{11}$, que sobreescribimos en $a_{31}$, y $a_{32}^{(2)}, a_{33}^{(2)}, \ldots, a_{3d}^{(2)}$, que sobreescribimos en $a_{32}, a_{33}, \ldots, a_{3d}$. Así continuamos el proceso. Al acabar la matriz $d \times d$, en que estuvo $A$, tiene los siguientes elementos

$$\begin{bmatrix} u_{11} & u_{12} & u_{13} & \cdots & u_{1,d-1} & u_{1d} \\ \ell_{21} & u_{22} & u_{23} & \cdots & u_{2,d-1} & u_{2d} \\ \ell_{31} & \ell_{32} & u_{33} & \cdots & u_{3,d-1} & u_{3d} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \ell_{d1} & \ell_{d,2} & \ell_{d,3} & \cdots & \ell_{d,d-1} & u_{dd} \end{bmatrix}.$$

**Ejemplo 10.4.1** Para el sistema (10.2), completadas las operaciones correspondientes a eliminar $x_1$, tiene la siguiente información

$$\begin{bmatrix} 2 & 3 & 4 & 5 \\ 3 & 6 & 7 & 8 \\ 4 & 30 & 44 & 50 \\ 6 & 42 & -23 & -13 \end{bmatrix};$$

comparando con (10.3) vemos que tenemos almacenados los coeficientes de la matriz $A^{(2)}$ del sistema resultante de eliminar $x_1$, excepto en las posiciones $(2,1)$, $(3,1)$, $(4,1)$ donde están los multiplicadores usados.

Tras completar las operaciones para eliminar $x_2$, la matriz es

$$\begin{bmatrix} 2 & 3 & 4 & 5 \\ 3 & 6 & 7 & 8 \\ 4 & 5 & 9 & 10 \\ 6 & 7 & -72 & -69 \end{bmatrix};$$

y a la salida del algoritmo tendremos

$$\begin{bmatrix} 2 & 3 & 4 & 5 \\ 3 & 6 & 7 & 8 \\ 4 & 5 & 9 & 10 \\ 6 & 7 & -8 & 11 \end{bmatrix}. \square$$

El algoritmo es como sigue:

$$
\begin{aligned}
&\text{Para } k = 1, \ldots, d-1, \\
&\qquad \text{Para } i = k+1, \ldots, d, \\
&\qquad\qquad a_{ik} \leftarrow \frac{a_{ik}}{a_{kk}}, \\
&\qquad\qquad \text{Para } j = k+1, \ldots, d, \\
&\qquad\qquad\qquad a_{ij} \leftarrow a_{ij} - a_{ik}a_{kj}.
\end{aligned}
$$

### 10.4.2 Subrutina de sustitución progresiva

Esta subrutina toma una matriz $L$ triangular inferior con diagonal de unos y un vector $\mathbf{b}$ y retorna la solución $\mathbf{c}$ del sistema $L\mathbf{c} = \mathbf{b}$. A la salida $\mathbf{c}$ se halla en las posiciones que a la entrada ocupaba $\mathbf{b}$. El algoritmo es

$$
\begin{aligned}
&\text{Para } j = 1, \ldots, d-1, \\
&\qquad \text{Para } i = j+1, \ldots, d, \\
&\qquad\qquad b_i \leftarrow b_i - \ell_{ij}b_j.
\end{aligned}
$$

Como vemos, estamos reproduciendo aquí las fórmulas de transformación del segundo miembro que ya dimos en la Subsección 10.2.2, con la salvedad que estamos ahora usando $j$ donde antes allí poníamos $k$.

Observe también que sólo se invocan los elementos subdiagonales de $L$ y poco importa qué tengamos almacenado en la diagonal por encima de la diagonal. Por eso esta subrutina puede tomar como matriz la resultante de la subrutina de factorización de la subsección precedente.

### 10.4.3 Subrutina de sustitución regresiva

Esta subrutina toma una matriz $U$ triangular superior y un vector $\mathbf{c}$ y retorna la solución $\mathbf{x}$ del sistema $U\mathbf{x} = \mathbf{c}$. A la salida $\mathbf{x}$ se halla en las posiciones que a la entrada ocupaba $\mathbf{c}$. El algoritmo es

$$
\begin{aligned}
&c_d \leftarrow \frac{c_d}{u_{dd}} \\
&\qquad \text{Para } i = d-1, \ldots, 1, \\
&\qquad\qquad \text{Para } j = i+1, \ldots, d,
\end{aligned}
$$

$$c_i \leftarrow c_i - u_{ij}c_j,$$

$$c_i \leftarrow \frac{c_i}{u_{ii}}$$

Como vemos estamos implementando aquí las fórmulas que ya vimos en la Subsección 10.2.3. Sólo se invocan los elementos diagonales y superdiagonales de $U$ y poco importa qué tengamos almacenado en la diagonal por debajo de la diagonal. Por eso esta subrutina puede tomar como matriz la resultante de la subrutina de factorización.

### 10.4.4 Solución de sistemas lineales

Para resolver un sistema $A\mathbf{x} = \mathbf{b}$ se usan sucesivamente las tres subrutinas que acabamos de describir. ¿Por qué hemos dispuesto tres subrutinas y no combinado todo en una sola? Supongamos que deseamos resolver un segundo sistema $A\mathbf{x}' = \mathbf{b}'$ *con la misma matriz del que resolvimos anteriormente,* circunstancia que se da muy a menudo en la práctica. Los factores $L$ y $U$ ya fueron hallados al resolver el primer sistema y no hemos de volver a usar la rutina de factorización. Basta con hacer las sustituciones progresiva y regresiva con el nuevo segundo miembro $\mathbf{b}'$.

### 10.4.5 Cálculo de la inversa

Para hallar la inversa $A^{-1}$ de $A$, basta tomar en cuenta que la igualdad $AA^{-1} = I$ nos dice que la $j$-ésima columna $\mathbf{f}_j$ de $A^{-1}$ es la solución del sistema $A\mathbf{f}_j = \mathbf{e}_j$, con $\mathbf{e}_j$ igual a la $j$-ésima columna de la matriz unidad, es decir, el $j$-ésimo vector coordenado. Debemos entonces resolver $d$ sistemas con la misma matriz.

### 10.4.6 Cálculo de determinantes

Llamar a la rutina de factorización y usar que

$$\det(A) = \det(U) = u_{11} \cdots u_{dd}$$

es la manera usual de computar el determinante de una matriz.

**Problema 10.4.1** Armado de papel, lápiz y eliminación Gaussiana halle el determinante y la inversa de las matrices en los Problemas 10.2.1, 10.2.2.

**Problema 10.4.2** Escriba programas de ordenador para las subrutinas descritas en esta sección.

**Problema 10.4.3** *El orden de los bucles.* Examine los algoritmos que hemos presentado en esta sección para las sustituciones progresiva y regresiva. Son análogos pero hay una diferencia. En el caso progresivo el bucle de índice $i$ que describe *filas* de $L$ es el más interno. Esto hace que tras utilizar un $\ell_{ij}$ el siguiente que se requiere es $\ell_{i+1,j}$, que está en la misma *columna*. En el caso regresivo el bucle más interno es el de índice $j$ y tras utilizar $u_{ij}$ se pasa a $u_{i,j+1}$, que está en la misma fila. El acceder los elementos sucesivamente a lo largo de una fila o a lo largo de una columna no es indiferente. Por ejemplo, en algunos lenguajes de programación, dos elementos de una matriz que aparecen seguidos en una columna ocupan posiciones adyacentes de memoria y es preferible que el bucle de filas sea el más interno. Reescriba el algoritmo de sustitución regresiva dejando como índice más interno el de filas.

**Problema 10.4.4** *Variantes de la eliminación Gaussiana.* En la rutina de factorización hay tres bucles, de fuera hacia dentro, $k$, $i$, $j$. Estos bucles se pueden disponer de $3! = 6$ maneras distintas y cada una tiene sus ventajas y desventajas dependiendo de la máquina y el lenguaje de programación; cada manera puede considerarse como una variante de la eliminación Gaussiana y alguna de ellas es clásicamente conocida como algoritmo de Crout ó Doolittle. Escriba una rutina en que el bucle más interno sea el $i$ y el más externo el $k$.

## 10.5 Costo operativo

Los resultados de esta sección son particularmente importantes para el análisis numérico. Supongamos que resolvemos por eliminación Gaussiana $N$ sistemas $d \times d$ con la misma matriz. Deseamos contar el número de operaciones necesarias. Sólo contaremos multiplicaciones y divisiones; las sumas suelen aparecer en número muy parecido al de multiplicaciones/divisiones y además, aunque esto depende de la máquina usada, suelen ser más baratas.

En la subrutina de factorización hemos de hacer una división por multiplicador. Hay un multiplicador para eliminar $x_{d-1}$, dos para $x_{d-2}, \ldots, d-1$ para $x_1$. Esto hace

$$1 + 2 + \cdots + d - 1 = \frac{d^2 - d}{2}$$

divisiones. Por otro lado, al eliminar $x_{d-1}$ hace falta una multiplicación para cambiar el elemento $(d, d)$, al eliminar $x_{d-2}$ hacen falta cuatro multiplicaciones, ..., y al eliminar $x_1$ se requieren $(d-1)^2$ multiplicaciones. Esto da

$$1^2 + 2^2 + \cdots + (d-1)^2 = \frac{d(d-1)(2d-1)}{6}$$

multiplicaciones y en total hay

$$\frac{d^2 - d}{2} + \frac{d(d-1)(2d-1)}{6} = \frac{d^3 - d}{3}$$

operaciones en la factorización.

Es fácil contar (¡hágalo!) que en una sustitución progresiva hay

$$\frac{d^2 - d}{2}$$

operaciones y en la regresiva

$$\frac{d^2 + d}{2}.$$

Para resolver $N$ sistemas se factoriza una vez y se llama $N$ veces a cada rutina de sustitución. Esto da un total de

$$\frac{d^3}{3} + Nd^2 - \frac{d}{3} \tag{10.16}$$

operaciones.

Observemos lo siguiente:

- El costo es un *polinomio* en $N$ y $d$, números que miden el tamaño del problema. Esto es mucho más ventajoso que el método 'Cramer + determinantes por fórmula' presentado en algunos libros y que *carece de interés:* en un determinante $d \times d$ calculado por la fórmula de definición hay $d!$ términos y el crecimiento de $d!$ con $d$ es tan grande que el método no se puede implementar en la práctica (Ejemplo 1.1.2).

- Fijado $N$ (por ejemplo $N = 1$) el crecimiento de (10.16) con $d$ es cúbico $d^3/3 + O(d^2)$. Asintóticamente, doblar la dimensión multiplica por 8 el costo. Por ejemplo, para $d = 20$, $d^3/3 + d^2 - d/3 = 3060$ y para $d = 40$, $d^3/3 + d^2 - d/3 = 22920$; el cociente 22920/3060 es $\approx 7.5$.

- Fijado $N$, la parte principal $d^3/3$ del costo al crecer $d$ proviene de la factorización. El costo de las sustituciones progresiva y regresiva se vuelve despreciable frente al costo de la factorización al crecer $d$. Para $d = 40$, de las 22920 operaciones para resolver un sistema, 21320 son de la factorización.

- El costo de hallar por eliminación Gaussiana $A^{-1}$ es (tomando $N = d$ en (10.16)) $(4d^3 - d)/3$. Esto es muy inferior al costo requerido cuando se usa para la inversa la fórmula con determinantes que figura en muchos libros.

- Si se ha de efectuar una operación $A^{-1}\mathbf{y}$ y se halla $A^{-1}$ por eliminación Gaussiana y luego se multiplica la inversa por $\mathbf{y}$, se requieren $(4d^3 - d)/3 +$

$d^2$ operaciones (matriz $\times$ vector cuesta $d^2$). Por otro lado si pensamos que $A^{-1}\mathbf{y}$ es la solución del sistema de matriz $A$ y segundo miembro $\mathbf{y}$ y usamos eliminación Gaussiana en este sistema sólo necesitamos $(d^3-d)/3+d^2$ operaciones. No interesa pues hallar la inversa explícitamente. La misma conclusión es cierta si se requieren $N$ productos $A^{-1}\mathbf{y}_j$. Por ello *hay que evitar en lo posible invertir matrices y procurar en lo posible resolver sistemas.*

**Problema 10.5.1** Se va a calcular $\mathbf{y} = A^{-1}(B\mathbf{z}+\mathbf{u})+\mathbf{x}$, donde $A$ y $B$ son matrices $d\times d$ conocidas y $\mathbf{u}$, $\mathbf{x}$, $\mathbf{z}$ vectores $d$-dimensionales conocidos. Explique cómo disponer los cálculos para no tener que invertir $A$. ¿Cuáles son los costos del proceso sugerido por Vd. y del obvio $B\mathbf{z}$, $B\mathbf{z}+\mathbf{u}$, $A^{-1}(B\mathbf{z}+\mathbf{u})$, $\mathbf{y}$?

**Problema 10.5.2** Si $A$ es $d\times d$ conocida e invertible y $B$ es $d\times\nu$ ¿cómo se halla $A^{-1}B$ sin encontrar la inversa?

# 10.6 Pivotaje

## 10.6.1 Motivación

En la Subsección 10.2.4 ya notamos que el método de eliminación descrito en la Sección 10.2 adolece del defecto de fallar para ciertas matrices invertibles $A$ por originar algún pivote nulo. Además de esta deficiencia importante, la eliminación Gaussiana, tal como la describimos, tiene la desventaja de que al usarse con aritmética inexacta como la de los ordenadores puede dar resultados inaceptablemente equivocados. La causa de tales resultados equivocados radica en la presencia de pivotes que, sin ser nulos, son pequeños en relación con las demás cantidades involucradas y que por tanto conducen a multiplicadores grandes. Los multiplicadores grandes pueden amplificar los errores introducidos por la inexactitud de la aritmética del ordenador.

**Ejemplo 10.6.1** Imaginemos un ordenador que trabajase en base 10 con mantisa de tres cifras y resolvamos en él

$$\begin{aligned} 1.00\times 10^{-4}x + 1.00\times 10^{0}y &= 1.00\times 10^{0},\\ 1.00\times 10^{0}x + 1.00\times 10^{0}y &= 2.00\times 10^{0}, \end{aligned}$$

sistema con solución exacta $x = 10000/9999 = 1.000100\ldots$, $y = 9998/9999 = 0.999899\ldots$ Los valores obtenidos son

$$\ell_{21} = \frac{1.00\times 10^{0}}{1.00\times 10^{-4}} = 1.00\times 10^{4},$$

$$
\begin{aligned}
a_{22}^{(2)} &= a_{22} - \ell_{21}a_{12} = 1.00 \times 10^0 - (1.00 \times 10^4)(1.00 \times 10^0) \\
&=^* -1.00 \times 10^4 \\
c_2 &= b_2 - \ell_{21}b_1 = 2.00 \times 10^0 - (1.00 \times 10^4)(1.00 \times 10^0) \\
&=^* -1.00 \times 10^4 \\
y &= \frac{c_2}{a_{22}^{(2)}} = 1.00 \times 10^0 \\
x &= \frac{b_1 - a_{12}y}{a_{11}} = \frac{1.00 \times 10^0 - (1.00 \times 10^0)(1.00 \times 10^0)}{1.00 \times 10^{-4}} = 0.
\end{aligned}
$$

¡El valor 0 obtenido para $x \approx 1$ es completamente erróneo! Y eso que en todo el proceso sólo hay dos errores en la aritmética —en las igualdades señaladas con una estrella— y ambos son pequeños relativamente. □

## 10.6.2 Pivotaje parcial

Para subsanar los inconvenientes que acabamos de señalar se reemplaza el método de eliminación Gaussiana descrito en la Sección 10.2 y a veces llamado *eliminación Gaussiana sin pivotaje* por otro llamado de *eliminación Gaussiana con pivotaje parcial.*

Para describir el método con pivotaje parcial, situémonos de nuevo en las circunstancias de la Subsección 10.2.2 y supongamos que hemos llegado ya al sistema $A^{(k)}\mathbf{x} = \mathbf{b}^{(k)}$ y vamos a eliminar $x_k$. Entonces, en vez de tomar directamente la ecuación $k$-ésima como pivotal, nos fijamos en las ecuaciones $k, k+1, \ldots, d$ que no han sido aún pivotales y de entre ellas elegimos aquella (llamémosla $p(k)$-ésima) en que la $x_k$ figure con coeficiente de valor absoluto lo mayor posible. Entonces permutamos (pivotamos) las ecuaciones $k$ y $p(k)$ y procedemos a eliminar $x_k$. Este proceso se ilustra en el siguiente ejemplo.

**Ejemplo 10.6.2** Sea

$$
\begin{array}{rcrcrcl}
-2x_1 & - & 3x_2 & + & 14x_3 & = & 4, \\
2x_1 & + & 2x_2 & - & 3x_3 & = & 5, \\
4x_1 & + & 2x_2 & - & 2x_3 & = & 10,
\end{array}
$$

el sistema por resolver. De los tres elementos de la primera columna el de mayor valor absoluto aparece en la ecuación $p(1) = 3$. Permutamos las ecuaciones 1 y $p(1)$ y tenemos

$$
\begin{array}{rcrcrcl}
4x_1 & + & 2x_2 & - & 2x_3 & = & 10, \\
2x_1 & + & 2x_2 & - & 3x_3 & = & 5, \\
-2x_1 & - & 3x_2 & + & 14x_3 & = & 4.
\end{array}
$$

Los multiplicadores son 0.5, −0.5, ambos de valor absoluto $\leq 1$ (¿es casualidad?), y el resultado

de la eliminación de $x_1$ es

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & - & 2x_3 & = & 10, \\ & & x_2 & - & 2x_3 & = & 0, \\ & - & 2x_2 & + & 13x_3 & = & 9. \end{array}$$

Para eliminar $x_2$ tenemos dos candidatos a pivote 1 y $-2$, el de mayor valor absoluto aparece en la ecuación que ahora ocupa la posición $p(2) = 3$. Permutando las ecuaciones 2 y $p(2)$

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & - & 2x_3 & = & 10, \\ & - & 2x_2 & + & 13x_3 & = & 9, \\ & & x_2 & - & 2x_3 & = & 0, \end{array}$$

y, eliminando con el multiplicador $-0.5$, tenemos

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & - & 2x_3 & = & 10, \\ & - & 2x_2 & + & 13x_3 & = & 9, \\ & & & & 4.5x_3 & = & 4.5. \end{array}$$

La sustitución regresiva da $x_3 = 1$, $x_2 = 2$, $x_1 = 2$. □

El método con pivotaje parcial subsana las deficiencias del método sin pivotaje.

- Ante todo, si se aplica con aritmética exacta a un sistema con matriz invertible $A$, puede completarse sin dar origen a ninguna división por cero. Para probar esto, supongamos que hemos podido llevar a cabo el método hasta formar el sistema intermedio $A^{(k)}\mathbf{x} = \mathbf{b}^{(k)}$. Se verifica, vistos los ceros en $A^{(k)}$,

$$\det(A^{(k)}) = a_{11}^{(k)} a_{22}^{(k)} \cdots a_{k-1,k-1}^{(k)} \det(B_k),$$

  donde $B_k$ es la matriz formada por las últimas $n-(k-1)$ filas y columnas de $A^{(k)}$. Si todos los elementos $a_{kk}^{(k)}, a_{k+1,k}^{(k)}, \ldots, a_{dk}^{(k)}$ fuesen nulos, $B_k$ tendría nula su primera columna, lo que conduce a $\det(B_k) = 0$ y a la contradicción $\det(A) = \pm\det(A^{(k)}) = 0$ ($A^{(k)}$ se obtiene de $A$ por operaciones que o mantienen el determinante o cambian su signo). Por consiguiente no todos los $a_{kk}^{(k)}, a_{k+1,k}^{(k)}, \ldots, a_{dk}^{(k)}$ son nulos y, si $k < d$, al quedarnos con el de mayor valor absoluto tendremos un pivote no nulo para eliminar $x_k$ (si $k = d$ hemos probado que $u_{dd} \neq 0$).

- Se puede demostrar que al llevar a cabo con aritmética de precisión finita la eliminación con pivotaje parcial el comportamiento de los errores es estable.

**Ejemplo 10.6.3** Resolvamos el sistema del Ejemplo 10.6.1, ahora pivotando. Hay que intercambiar las ecuaciones, y tras ello el multiplicador es $1.00 \times 10^{-4}$. La ecuación para determinar $y$ es $1.00 \times 10^0 y = 1.00 \times 10^0$, y por tanto $y = 1.00 \times 10^0$. La $x$ se obtiene de $1.00 \times 10^0 x + 1.00 \times 10^0 y = 2.00 \times 10^0$, con lo que $x = 1.00 \times 10^0$. Ni $x$ ni $y$ son exactos (los verdaderos valores no son representables en nuestro ordenador imaginario), pero al menos son los números representables en el ordenador más próximos a la verdadera solución. □

## 10.6.3 Interpretación matricial

Al igual que hicimos con el método sin pivotaje, conviene interpretar el método de eliminación con pivotaje en términos de factorizaciones de matrices. La observación crucial es: *eliminar con pivotaje parcial en un sistema dado es lo mismo que eliminar sin pivotaje en un sistema obtenido permutando las ecuaciones del dado.*

**Ejemplo 10.6.4** Consideremos el sistema del Ejemplo 10.6.2, reordenemos las ecuaciones como sigue

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & - & 2x_3 & = & 10, \\ -2x_1 & - & 3x_2 & + & 14x_3 & = & 4, \\ 2x_1 & + & 2x_2 & - & 3x_3 & = & 5, \end{array}$$

y resolvamos sin pivotaje. Obtenemos, con los multiplicadores $\ell_{21} = -0.5$ y $\ell_{31} = 0.5$,

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & - & 2x_3 & = & 10, \\ & - & 2x_2 & + & 13x_3 & = & 9. \\ & & x_2 & - & 2x_3 & = & 0, \end{array}$$

y, con el multiplicador $\ell_{32} = -0.5$,

$$\begin{array}{rcrcrcl} 4x_1 & + & 2x_2 & - & 2x_3 & = & 10, \\ & - & 2x_2 & + & 13x_3 & = & 9, \\ & & & & 4.5x_3 & = & 4.5. \end{array}$$

Hemos hecho justamente las mismas operaciones que en el Ejemplo 10.6.2, porque las ecuaciones *ya estaban en el orden adecuado para que fuesen saliendo los mayores pivotes posibles.* En un sistema dado no es obvio antes de comenzar la eliminación cual es el orden adecuado; ese orden *se va descubriendo a lo largo del proceso de eliminación.* □

Para representar matricialmente el proceso de permutación de ecuaciones necesitamos el concepto de *matriz de permutación.* Una matriz $d \times d$ es de permutación si todos sus elementos son nulos excepto $d$ que valen la unidad y no hay dos unos ni en una misma fila ni en una misma columna.

**Ejemplo 10.6.5** Con $d = 3$ tenemos estas matrices de permutación

$$\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \tag{10.17}$$

y otras tres ¿cuáles? □

Si $P$ es una matriz de permutación con el uno de la $i$-ésima fila situado en la columna $\pi(i)$, entonces el vector $P\mathbf{x}$ tiene por $i$-ésima componente $x_{\pi(i)}$. De la misma manera, si $B$ es una matriz $d \times d$ entonces $PB$ tiene por fila $i$-ésima la fila $\pi(i)$ de $B$.

**Ejemplo 10.6.6** Si $P$ es la primera matriz de (10.17), la primera componente de $P\mathbf{x}$ es la segunda de $\mathbf{x}$; la segunda de $P\mathbf{x}$, la primera de $\mathbf{x}$; las terceras coinciden. □

Equipados con las matrices de permutación, podemos decir que la eliminación con pivotaje parcial en $A\mathbf{x} = \mathbf{b}$ es equivalente a eliminar sin pivotaje en un sistema $PA\mathbf{x} = P\mathbf{b}$, siendo $P$ una matriz de permutación adecuada. Por tanto en el algoritmo con pivotaje parcial se genera una factorización $LU = PA$. Resolver $A\mathbf{x} = \mathbf{b}$ es entonces equivalente a resolver $LU\mathbf{x} = P\mathbf{b}$, y esto requiere una sustitución progresiva con matriz $L$ sobre el vector permutado $P\mathbf{b}$ y una sustitución regresiva. En el método que expusimos más arriba, al eliminar las incógnitas se encuentran, de modo encubierto, $L$ y $P$ y explícitamente $U$.

### 10.6.4 Implementación

Como en el caso sin pivotaje conviene distribuir la tarea en tres subrutinas.

La subrutina de factorización necesita almacenar una matriz $d \times d$ y un vector de $d-1$ componentes enteras $p(k)$, $k = 1, 2, \ldots, d-1$, donde $p(k)$ es el índice de la fila que pasa a ser pivotal para eliminar $x_k$.

**Ejemplo 10.6.7** En el sistema del Ejemplo 10.6.2 la matriz antes de comenzar es

$$\begin{bmatrix} -2 & -3 & 14 \\ 2 & 2 & -3 \\ 4 & 2 & -2 \end{bmatrix};$$

al intercambiar las filas primera y tercera hacemos $p(1) = 3$ y

$$\begin{bmatrix} 4 & 2 & -2 \\ 2 & 2 & -3 \\ -2 & -3 & 14 \end{bmatrix}.$$

Tras eliminar $x_1$ y ubicando los multiplicadores en los sitios donde aparecerían los ceros, tendremos

$$\begin{bmatrix} 4 & 2 & -2 \\ 0.5 & 1 & -2 \\ -0.5 & -2 & 13 \end{bmatrix}.$$

Intercambiamos filas (incluyendo los multiplicadores) y ponemos $p(2) = 3$,

$$\begin{bmatrix} 4 & 2 & -2 \\ -0.5 & -2 & 13 \\ 0.5 & 1 & -2 \end{bmatrix}.$$

Eliminando $x_2$

$$\begin{bmatrix} 4 & 2 & -2 \\ -0.5 & -2 & 13 \\ 0.5 & -0.5 & 4.5 \end{bmatrix},$$

matriz, que junto al vector $[3, 3]^T$ es la salida de la subrutina. Notemos que

$$LU = \begin{bmatrix} 1 & 0 & 0 \\ -0.5 & 1 & 0 \\ 0.5 & -0.5 & 1 \end{bmatrix} \begin{bmatrix} 4 & 2 & -2 \\ 0 & -2 & 13 \\ 0 & 0 & 4.5 \end{bmatrix} = \begin{bmatrix} 4 & 2 & -2 \\ -2 & -3 & 14 \\ 2 & 2 & -3 \end{bmatrix} = PA,$$

la matriz del tercer miembro se obtiene de $A$ permutando primero las filas primera y tercera y luego segunda y tercera. □

La subrutina de sustitución regresiva toma el vector de las $p(i)$, la matriz resultado de la factorización y el segundo miembro $\mathbf{b}$ y procede a hallar $P\mathbf{b}$ y resolver $L\mathbf{c} = P\mathbf{b}$.

**Ejemplo 10.6.8** En el caso del ejemplo precedente, tomamos el segundo miembro $[4, 5, 10]^T$ y con el vector de punteros $[3, 3]^T$ intercambiamos la primera componente con la tercera y luego la segunda con la tercera:

$$\begin{bmatrix} 4 \\ 5 \\ 10 \end{bmatrix} \to \begin{bmatrix} 10 \\ 5 \\ 4 \end{bmatrix} \to \begin{bmatrix} 10 \\ 4 \\ 5 \end{bmatrix}.$$

En este vector hacemos sustitución progresiva con los multiplicadores contenidos en $L$,

$$\begin{bmatrix} 10 \\ 4 \\ 5 \end{bmatrix} \to \begin{bmatrix} 10 \\ 9 \\ 0 \end{bmatrix} \to \begin{bmatrix} 10 \\ 9 \\ 4.5 \end{bmatrix}. \square$$

La sustitución regresiva no ofrece novedad.

Estas subrutinas pueden usarse para hallar la inversa y el determinante. Para el determinante hay que tener en cuenta que $\det(P)\det(A) = \det(L)\det(U)$, con lo que $\det(A) = \pm\det(U) = \pm u_{11} \cdots u_{dd}$; el signo es más (menos) si se ha permutado filas un número par (non) de veces.

**Problema 10.6.1** Resuelva por eliminación Gaussiana con pivotaje parcial el sistema $A\mathbf{x} = \mathbf{b}$ dado por (10.2) (efectúe los cálculos con papel y lápiz, usando fracciones). Escriba la matriz de permutación $P$ tal que la eliminación con pivotaje parcial en (10.2) es equivalente a la eliminación sin pivotaje en $PA\mathbf{x} = P\mathbf{b}$. Escriba los factores $L$ y $U$ de $PA$. Repita este problema usando los sistemas de los Problemas 10.2.1 y 10.2.2.

**Problema 10.6.2** Escriba subrutinas que lleven a cabo la eliminación Gaussiana con pivotaje parcial.

**Problema 10.6.3** *Matrices de permutación.* Sea $P$ una matriz de permutación $d \times d$, y sea $\rho(j)$ la posición que ocupa el 1 en la $j$-ésima *columna*; $\rho$ es una permutación, transformación biunívoca del conjunto $\{1, 2, \ldots, d\}$ en sí mismo. Demuestre que, si $C$ es una matriz con $d$ columnas y un número cualquiera de filas, $CP$ tiene por $j$-ésima columna la columna $\rho(j)$ de $C$.

**Problema 10.6.4** *Continuación.* Pruebe que el efecto geométrico de la transformación lineal de matriz $P$ es transformar el $j$-ésimo vector coordenado en el $\rho(j)$-ésimo, $P\mathbf{e}_j = \mathbf{e}_{\rho(j)}$, $j = 1, 2, \ldots, d$. Concluya de esto que (i) la matriz $P$ es ortogonal por llevar una base ortonormal en otra y por tanto $P^{-1} = P^T$, (ii) si $P$ y $Q$ son matrices de permutación con permutaciones asociadas $\rho$ y $\sigma$, entonces $QP$ es una nueva matriz de permutación asociada a la función $\sigma \circ \rho$ resultado de componer $\rho$ seguida de $\sigma$.

**Problema 10.6.5** *La eliminación Gaussiana con pivotaje total.* Además del pivotaje parcial que hemos estudiado, existe el pivotaje total, raramente usado en la práctica aunque tiene la ventaja de llevar a pivotes más grandes y multiplicadores más chicos que el pivotaje parcial. Al comenzar la eliminación se examinan los módulos los $d^2$ elementos de $A$ y se determina el mayor. Si éste figura en la ecuación $i$ y columna $j$, se permutan las ecuaciones primera e $i$-ésima, y también las columnas primera y $j$-ésima, con lo que el vector de incógnitas ahora tiene $x_j$ en primera posición. Se elimina la primera incógnita (que ahora es $x_j$) usando como pivotal la primera ecuación (que es la $i$-ésima antigua). A continuación, se examinan los módulos de los $(d-1)^2$ elementos de las últimas $d-1$ ecuaciones/incógnitas y se elige el mayor para hacer de nuevo pivote, etc. Resuelva, utilizando aritmética de números racionales y pivotaje total, el sistema del Ejemplo 10.6.2.

**Problema 10.6.6** *Continuación.* Dé una interpretación matricial del pivotaje total.

**Problema 10.6.7** *Sistemas escalonados.* Consideraremos sistemas lineales $A\mathbf{x} = \mathbf{b}$ donde hay $m$ ecuaciones y $n$ incógnitas. No se hace ninguna hipótesis sobre $m$ y $n$ de modo que son posibles todas las combinaciones: $m > n$ (más ecuaciones que incógnitas), $m = n$ (tantas ecuaciones como incógnitas), $m < n$ (más incógnitas que ecuaciones). En el caso $m = n$ no se supone $A$ invertible.

El tipo más sencillo de estos sistemas generales es el que tiene la matriz *escalonada*. Esto quiere decir que las últimas $m - r$ filas de $A$ , $(0 \leq r \leq m)$ son idénticamente nulas y para cada $i = 1, \ldots, r$ existe un índice $\ell(i)$ de manera que $a_{i,1} = 0$, $a_{i,2} = 0 \ldots$, $a_{i,\ell(i)-1} = 0$, $a_{i,\ell(i)} \neq 0$ y además $\ell(i)$ es función estrictamente creciente de $i$ (esto es, $\ell(i_1) < \ell(i_2)$ si $i_1 < i_2$). Los sistemas escalonados juegan, para sistemas generales, el papel que en los sistemas $n \times n$ de matriz invertible juegan los sistemas con matriz triangular superior.

Pruebe que, si $A$ es escalonada, sus $r$ primeros vectores fila son linealmente independientes y que, por tanto, $r$ es el rango de $A$.

**Problema 10.6.8** *Continuación.* Observe que, si en un sistema escalonado existe $b_i \neq 0$ con $i > r$, entonces el sistema carece de solución pues la $i$-ésima ecuación no la tiene. Suponga entonces que no existe $b_i \neq 0$ con $i > r$, descarte las últimas $m - r$ ecuaciones, que son triviales, y explique cómo resolver el sistema por un proceso de sustitución regresiva en el que se quizá se pueda dar a alguna de las incógnitas valores arbitrarios. ¿A cuáles? (Respuesta: a las $x_j$ con $j < \ell(1)$, $j > \ell(r)$ o con $j$ tal que $\ell(i) < j < \ell(i+1)$, $i = 1, \ldots, r-1$.)

**Problema 10.6.9** *Eliminación Gaussiana en sistemas generales.* Indique cómo llevar a cabo eliminación Gaussiana con pivotaje parcial en un sistema general $m \times n$ para reducirlo a uno con matriz escalonada, que se puede resolver acto seguido como en el problema precedente.

**Problema 10.6.10** *Eliminación Gaussiana y rango.* Explique por qué efectuando la reducción a forma escalonada mencionada en el problema precedente se pone de manifiesto el rango de la matriz de partida. (Por desgracia esta forma de hallar el rango es muy sensible a errores de redondeo.)

# Índice alfabético


Abel S9.1
Abscisas
De cuadratura S5.1.2
De derivación numérica S6.1
De interpolación S3.1.1
Equiespaciadas C7
Aitken, Lema de P3.2.10–12
Algoritmo
De Crout P10.4.4
De Doolittle P10.4.4
De Horner E1.6.1, P1.6.1–7, P3.2.5, P3.4.2, P4.2.5
De Neville P3.2.11–12
Análisis dimensional E2.6.2, P2.6.4, P3.2.8, S3.3.1, S5.3, S6.3
Análisis numérico
Historia S1.3
Necesidad S1.2
Objetivos S1.1
Aproximaciones sucesivas *ver* Iteración de punto fijo
Atracción de un punto fijo
De puntos alejados S9.4.5
Local S9.4.2
*ver también* Dominio de atracción
Avogadro *ver* Número

Barrow *ver* Regla
Base
Del espacio de soluciones de una ecuación en diferencias lineal homogénea S8.2.3
Del espacio de soluciones de una ecuación en diferencias lineal homogénea con coeficientes constantes S8.3.1–2
Del espacio $M_0^1(\Delta)$ P4.1.3
Del espacio $M_0^2(\Delta)$ P4.2.3
Bernoulli *ver* Número
Binomio de Newton P7.2.2, P7.4.3
Bisección *ver* Método de bisección
Bolzano *ver* Teorema

Cálculo de diferencias finitas S7.4
Cálculo numérico *ver* Análisis numérico
Cambio de escala o variable
En el polinomio de Taylor S2.6
En el polinomio interpolador de Lagrange P3.1.13–16
En la forma de Newton del polinomio interpolador P3.2.8
En las reglas de cuadratura E5.2.5
En las reglas de derivación E6.2.7
Cambio de origen en un polinomio P1.6.5–7, P2.1.6
Cero de una función C9
Múltiple P8.3.6, P9.5.1
Simple S9.5
*ver también* Raíz

Chebyshev *ver* Método de Chebyshev
Coeficientes
De una ecuación en diferencias lineal S8.2.1
Indeterminados S2.1, S3.1.1, S3.2.6, S5.2.2, S6.2
*ver también* Pesos
Condiciones iniciales *ver* Problema de valores iniciales
Contraejemplo de Runge E3.4.1, S4.1.3, E4.2.1
Convergencia
Cuadrática en la iteración de punto fijo S9.4.3, S9.5
Cúbica en la iteración de punto fijo P9.4.7, P9.5.3–4
De la cuadratura numérica S5.4.2
De la derivación numérica S6.3
De la interpolación cuadrática a trozos S4.2
De la interpolación lineal a trozos S4.1
De la iteración de punto fijo S9.4.2–3
De la serie de Taylor S2.5
De los polinomios de interpolación de Lagrange S3.4
Del método de la secante S9.3, S9.4.3
Superlineal S9.4.3
Costo operativo E1.1.2, S1.6
De evaluar el polinomio interpolador de Lagrange S3.2.6
De evaluar un polinomio S1.6
De formar el polinomio interpolador de Lagrange S3.2.6
De la eliminación Gaussiana S10.5
En la cuadratura numérica E5.4.3
Cota *ver* Error
Crout *ver* Algoritmo
Cuadratura del círculo S5.1.1
Cuadratura Gaussiana P5.2.7
Cuadratura numérica E1.1.1, S1.2, C5, P7.4.12

Dal Ferro S9.1
Derivación numérica C6, P7.2.3, P7.3.2, P7.4.4–6, P7.4.9–10
*ver también* Fórmula
Derivadas sucesivas de un polinomio P2.1.5–6
Determinante
Cálculo efectivo S10.4.6, S10.6.4
Cálculo por fórmula E1.1.2
De Vandermonde *ver* Matriz de Vandermonde
Diámetro de una partición S4.1.2
Diferencias (finitas)
De un polinomio P7.1.1
*ver también* Ecuaciones en diferencias, Operadores de diferencias
Diferencias divididas S3.2, S6.2
De un polinomio P3.2.2
De un producto P3.2.9
Relación con derivadas S3.3.2, S6.2
Relación con diferencias S7.1.5
Dimensiones *ver* Análisis dimensional
Dinámica de poblaciones *ver* Sistema dinámico discreto
Distribución normal E1.1.1, S2.1, E5.4.1–2
Dominio de atracción de un punto fijo S9.4.5, P9.4.10, P9.5.2
Doolittle *ver* Algoritmo

Ecuación

De Fibonacci E8.1.2, E8.1.4, E8.1.6, E8.2.2, P8.2.1, P8.2.4, E8.3.2–3
De Kepler E9.2.1, E9.3.1, E9.4.1–7, E9.5.1
Pivotal S10.2.2
Ecuación característica (ecuación en diferencias lineal de coeficientes constantes) S8.3.1
Ecuaciones
De segundo grado, error de redondeo P1.5.1, P9.5.3
Lineales *ver* Sistema lineal
No lineales C9
Polinómicas P2.4.6, S9.1
*ver también* Interpolación inversa
Ecuaciones en diferencias C9, S9.3, S9.4.1
Lineales S8.2–3
Lineales de coeficientes constantes S8.3
Lineales homogéneas S8.2.3
Lineales no homogéneas S8.2.4
Eficiencia S1.6, E5.4.3
Eliminación de incógnitas S10.2.2
Eliminación Gaussiana C10
Costo operativo S10.5
Implementación con pivotaje S10.6.4
Implementación sin pivotaje S10.4
Para sistemas rectangulares generales P10.6.9
Pivotaje S10.6
Variantes P10.4.4
Y rango P10.6.10
Error S1.4
Absoluto S1.4.1, S2.2
Cota S1.4.2
De cuadraura S5.3, S5.4.2
De derivación numérica S6.3
De redondeo S1.5, E3.1.2, S6.4, S7.1.3, P9.5.3, E10.6.1, E10.6.3
En el polinomio de Taylor S2.4
En la interpolación cuadrática a trozos S4.2
En la interpolación de Lagrange S3.3
En la interpolación lineal a trozos S4.1
Estimación S1.4.2
Por defecto S1.4.1
Por exceso S1.4.1
Porcentual S1.4.1
Relativo S1.4.1, S2.2, E6.4.2, E10.6.1
Espacio
$M_0^1(\Delta)$ S4.1
$M_0^2(\Delta)$ S4.2
Vectorial de sucesiones S8.1
*ver también* Base
Estimativo de error S1.4.2
Euler
Suma de los inversos de los cuadrados E1.2.1
*Ver* Fórmula sumatoria de Euler-Maclaurin, Función Gamma de Euler, Método de Euler
Evaluación
De derivadas de polinomios P2.1.6
De funciones cuadráticas a trozos S4.2
De funciones lineales a trozos S4.1.1
De polinomios S1.6
De polinomios de interpolación

S3.2.6
Excentricidad
De una órbita E9.2.1
De Venus E9.4.1
Exponente S1.5.2
Extrapolación al límite de Richardson P5.4.5–6

Factorización de una matriz S10.3
Ferrari S9.1
Fibonacci *ver* Ecuación de Fibonacci
Fila pivotal S10.2.2
Fórmula de derivación
Central (de cinco puntos) para la derivada primera P6.2.1
Central (de tres puntos) para la derivada primera E6.2.4, E6.3.2–3, E6.4.2, E7.3.3
Central (de tres puntos) para la derivada segunda E6.2.6, P6.3.2, E7.3.4
Central para la derivada primera P7.4.9
Central para la derivada segunda P7.4.10
Central para la derivada tercera P6.2.3
De dos nodos para la derivada primera E6.2.2
De tres nodos para la derivada segunda E6.2.5
Progresiva (de dos puntos) para la derivada primera E6.2.3, E6.3.1
Progresiva para la derivada primera P7.2.3, P7.4.5
Regresiva (de tres puntos) para la derivada primera P6.2.2
Regresiva para la derivada primera P7.3.2, P7.4.6
Fórmula de Stirling E2.2.3
Fórmula sumatoria de Euler-Macalurin P7.4.13–14
Función
Cuadrática a trozos S4.2
De iteración S9.4.1
Elemental S1.2, S5.1.1
Gamma de Euler E3.1.1, P3.1.17
Hiperbólica P7.4.7–8
Impar P2.1.8
Lineal a trozos S4.1
Pagoda P4.2.3
Par P2.1.8
Polinómica a trozos C4, S5.4.3
Racional S5.1.1
Sombrero P4.1.3

Gauss *ver* Cuadratura Gaussiana, Eliminación Gaussiana
Gamma *ver* Función
Grado de exactitud S5.1.3, S6.1

Halley *ver* Método de Halley
Historia
De las ecuaciones algebraicas S9.1
Del análisis numérico S1.3
Del polinomio de Taylor P2.4.5
Leibniz e infinitésimos S7.1.1
*Liber abaci* E8.1.2
Horner *ver* Algoritmo

Implementación (eliminación Gaussiana)
Con pivotaje parcial S10.6.4
Sin pivotaje S10.4
Integración
Numérica *ver* Cuadratura

Interpolación
De Lagrange *ver* Polinomio interpolador de Lagrange
De Taylor *ver* Polinomio de Taylor
Inversa P3.1.5
Lineal a trozos S4.1
Polinómica a trozos C4, S5.4.3
*ver también* Métodos de hallar fórmulas
Iteración de punto fijo S9.4–5
Iterantes S9.3–5

Kepler *ver* Ecuación de Kepler

Lagrange
Forma de Lagrange para el error del polinomio de Taylor S2.4
*ver también* Polinomio de base de Lagrange, Polinomio interpolador de Lagrange
Landau *ver* Notaciones
Leibniz *ver* Historia, Notación de Leibniz para la derivada
Leonardo de Pisa *ver* Fibonacci
L'Hôpital *ver* Regla
*Liber abaci ver* Historia

Maclaurin *ver* Fórmula sumatoria de Euler-Maclaurin
Malthus E2.2.2
Mantisa S1.5.2
Matriz
De permutación S10.6.3, P10.6.3–4
De Vandermonde S3.1.1, P3.1.6–7, S5.2.2, S8.3.1
Inversa, su cáculo S10.4.5, S10.6.4
Triangular S2.1, S10.3.3
*ver también* Determinante, Rango
Método
De bisección S9.2
De Chebyshev P9.5.4
De Euler P9.5.3
De Gauss *ver* Eliminación Gaussiana
De Halley P9.5.4
De la secante S9.3, S9.4.3
De Newton E9.4.7, S9.5
Métodos de hallar fórmulas
Del desarrollo de Taylor S5.2.3, S6.2
Interpolatorio S5.2.1, S6.2
*ver también* Coeficientes indeterminados
Mínimo de una función E2.3.4
Multiplicador S10.2.2

Nabla ($\nabla$) S7.3
Neville *ver* Algoritmo
Newton S1.6, *ver* Binomio de Newton, Método de Newton, Polinomio interpolador de Lagrange (forma de Newton)
Nodos *ver* Abscisas
Notación de Leibniz para la derivada E2.6.2
Notaciones de Landau S2.2
Núcleo de un operador lineal S8.2.3–4
Número
Combinatorio S7.1.3, S7.2
De Avogadro S1.5.1
De Bernoulli P7.4.11, P7.4.13
De Fibonacci *ver* Ecuación de Fibonacci
*ver también* Razón áurea

Operador
Asociado a una ecuación en diferencias S8.2.2

De desplazamiento $E$ S7.4
De diferencia central $\delta$, $\Delta_0$ S7.3
De diferencia progresiva $\Delta$ S7.1
De diferencia regresiva $\nabla$ S7.3
Identidad $Id$ S7.4
Operadores de diferencias C7, S8.1.3

Partición S4.1.1
Adaptada P5.4.4
Pesos
De cuadratura S5.1.2
De derivación numérica S6.1
Pivotaje
Parcial S10.6.2–4
Total P10.6.5–6
Pivote S10.2.2
Polinomio
Cambio de origen P1.6.5–7, P2.1.6
Ceros múltiples P8.3.6
División P1.6.4
Evaluación P1.6.1–3
Evaluación de las derivadas sucesivas P2.1.5–6
Sus diferencias P7.1.1
Sus diferencias divididas P3.2.2
Polinomio característico *ver* Ecuación característica
Polinomio de base de Lagrange S3.1.2, S5.2.1, P6.2.5
Polinomio de Taylor
Cambio de escala S2.6
Caracterización analítica S2.3
Convergencia S2.5
De $(1+x)^s$ P7.2.2
Error S2.4
Estudio algebraico S2.1
Historia P2.4.5
Obtención práctica P2.1.3, P9.5.3
Paridad P2.1.8
Polinomio interpolador de Lagrange
Comparación con interpolación lineal a trozos S4.1.3
Convergencia S3.4
Definición S3.1
Error S3.3
Evaluación S3.2.6
Forma de Lagrange S3.1.2, S3.2.6, S5.2.1
Forma de Newton S3.2, P3.4.2, P4.2.5, E6.2.2, E6.2.5, S7.2, E7.3.1, P7.4.3
Relación con el polinomio de Taylor E3.3.3
*ver también* Algoritmo de Neville, Lema de Aitken
Problema de valores iniciales S8.1.2, E8.3.2, S9.3–4
Problema mal puesto S6.4
Producto exterior de vectores P10.3.4
Programa de ordenador P1.6.5, P2.1.6, P3.4.2, P4.1.5, P4.2.5, P5.4.2, P9.2.1, P9.3.1, P10.4.2, P10.6.2
Punto fijo de una función *ver* Iteración de punto fijo
Punto flotante S1.5

Raíz
Cuadrada, cálculo aproximado P2.4.4–5
Doble E3.3.3
Ecuación algebraica P2.4.6
*ver también* Cero de una función, Ecuaciones
Rango de una matriz P10.6.10
Razón áurea E8.3.2, P8.3.1

Redondeo *ver* Error de redondeo
Regla
De Barrow S5.1.1
De cuadratura C5
De cuadratura compuesta S5.4
De L'Hôpital S2.3
De los rectángulos S5.2–3
De los rectángulos compuesta S5.4
De los tres octavos P5.2.4
De los trapecios S5.2–5.3
De los trapecios compuesta S5.4
De los trapecios modificada P5.2.5
De Simpson S5.2–3
De Simpson compuesta S5.4
Del punto medio S5.2–3
Del punto medio compuesta S5.4
Gaussiana P5.2.7
Representación de números reales en el ordenador S1.5
Repulsión local de un punto fijo S9.4.4
Richardson *ver* Extrapolación
Rolle *ver* Teorema de Rolle
Ruffini S1.6
Runge *ver* Contraejemplo

Secante *ver* Método de la secante
Serie de Taylor S2.5
Simpson *ver* Regla
Sistema
De numeración E8.1.2
Dinámico discreto P9.4.11–14
Escalonado P10.6.7–9
Lineal S2.1, S3.1.1, S5.2.2, C10
Triangular S2.1, S10.2.1, S10.2.3, S10.3
Stirling E2.2.3
Sumación de series E1.2.1, E1.4.1, P1.4.1, E1.5.5
Suma
De cubos E7.1.3
De potencias P7.1.4, P7.4.14
Sustitución
Progresiva S10.3.2
Regresiva S2.1, S10.2.3, S10.3.2

Tabla
De diferencias S7.1.3
De diferencias divididas S3.2.5, P3.2.4
De una función S3.1.1
Tartaglia S9.1
Taylor *ver* Método del desarrollo de Taylor, Polinomio de Taylor, Serie de Taylor
Teorema
De aproximación de Weierstrass S3.4
De Bolzano S9.2
De Rolle S2.4
De valor medio para integrales P5.3.2
De valor medio para sumas P5.4.3
Término independiente, no homogéneo o fuente de una ecuación en diferencias S8.2.1

Valor medio *ver* Teorema
Vandermonde *ver* Matriz

Weierstrass, *ver* Teorema

UNIVERSIDAD AUTÓNOMA DE MADRID
5410029932

REC

ISBN 978-84-8448-552-X
9 788484 485520

En las últimas décadas el cálculo científico ha experimentado una revolución al aparecer primero y más tarde popularizarse y desarrollarse los ordenadores digitales. En todas las ciencias e ingenierías son hoy posibles cálculos inconcebibles hace treinta años. Esta obra busca presentar al lector las bases del cálculo numérico, subrayando aspectos esenciales como la búsqueda de la eficiencia y el estudio de los errores resultantes de calcular aproximadamente. Los ejemplos y problemas constituyen una parte substancial del volumen.

Se ha perseguido reducir al mínimo los requisitos matemáticos previos y por ello el libro resulta adecuado para los primeros cursos de Facultades Universitarias y Escuelas Técnicas. Si se considera en su totalidad cubre holgadamente un cuatrimestre de cálculo numérico. También pueden extraerse materiales para tratar aspectos de cálculo numérico en asignaturas de álgebra lineal, cálculo infinitesimal, ampliación de matemáticas, etc.

UniversidaddeValladolid
Secretariado de Publicaciones
e Intercambio Editorial