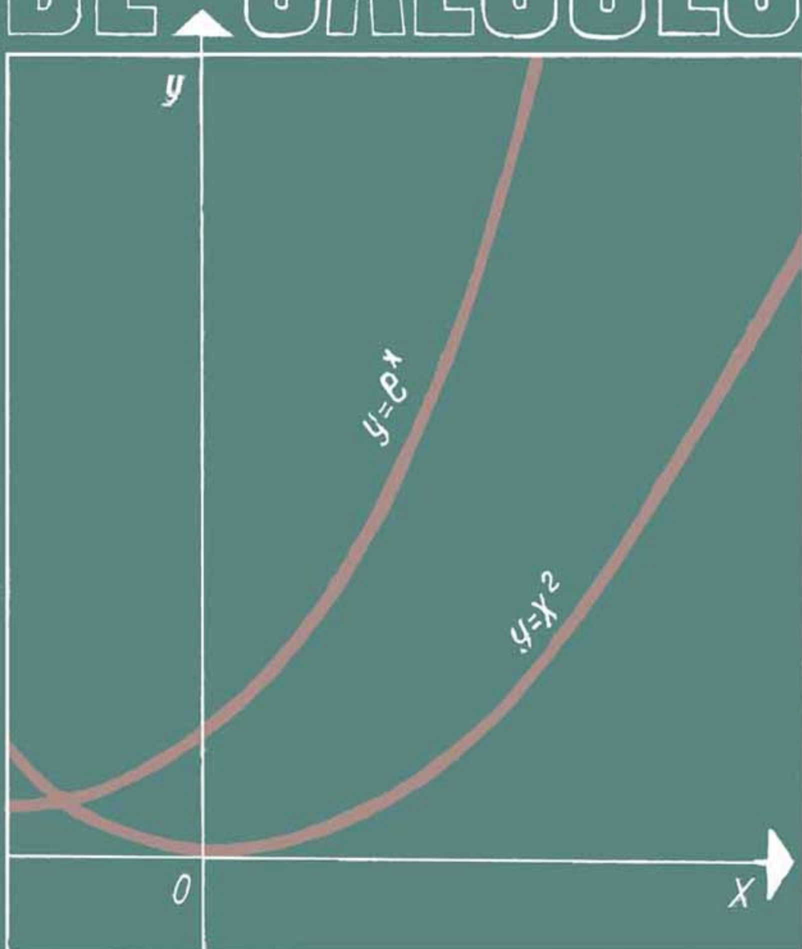
MATEMÁTICA DE CÁLCULO



Editorial Mir Moscú

Н. И. Данилина, Н. С. Дубровская,
О. П. Кваша, Г. Л. Смирнов

ВЫЧИСЛИТЕЛЬНАЯ МАТЕМАТИКА

Издательство «Высшая школа»

N.I. Danílina, N.S. Dubróvskaya. O.P. Kvashá,
G.L. Smirnov

MATEMÁTICA DE CÁLCULO

(ANÁLISIS NUMÉRICO)

Editorial Mir Moscú

Traducido del ruso por A. I. Samojválov

Impreso en la URSS

На испанском языке

ISBN 5-03-001813-1

Indice

Prefacio

El proceso actual de desarrollo impetuoso de la técnica de cálculo origina la ampliación constante de la esfera de aplicación de las partes modernas de matemáticas. Los métodos cuantitativos se introducen prácticamente en todos los campos de la actividad humana. Al mismo tiempo la utilización de la técnica de calculación en la economía nacional exige especialistas de alta calificación que posean los métodos de matemática de cálculo.

La matemática de cálculo es una de las asignaturas principales indispensables para preparar especialistas de alto nivel profesional en distintas ramas de la economía nacional. El objetivo de su estudio consiste en enculcar a los alumnos los fundamentos teóricos y hábitos prácticos para la resolución de diferentes problemas aplicados con el empleo de los modelos matemáticos y métodos numéricos realizables en el ordenador.

El material principal del Manual abarca todas las cuestiones contenidas en el programa de la asignatura «Matemática de cálculo» para la especialidad «Programación para los ordenadores de alta velocidad» y constituye un curso entero, acabado. Al mismo tiempo, comprendiendo que incluso actualmente el contenido de la materia no se ha formado definitivamente, así como teniendo en cuenta las perspectivas de desarrollo de los métodos numéricos, los autores consideraron necesario incluir algunas cuestiones que hoy día salen de los marcos del programa vigente. El material respectivo compuesto con caracteres menudos no es obligatorio para el estudio, pero puede ser recomendado al trabajar con los alumnos más preparados.

El contenido del manual comprende tres secciones grandes: «Métodos del álgebra» (caps. II—VI), «Métodos numéricos del análisis matemático» (caps. I, VII, VIII) y «Métodos numéricos de resolución de las ecuaciones diferenciales (caps. IX, X).

El material teórico está ilustrado por numerosos ejemplos. Al final de cada capítulo se dan ejercicios para el trabajo individual.

En el libro están adoptadas las siguientes designaciones: el comienzo y el final de la demostración de cualquier afirmación se señalan con signos □ y ■, respectivamente, el comienzo y el final de la resolución de un ejemplo, con signos △ y ▲.

Autores

Introducción

Los métodos de cálculo fueron desarrollados por tales eminentes científicos como K. Gauss, I. Newton, O. Cauchy, C. Hermite, B. G. Galerkin, A. N. Krylov, N. I. Lobachevski, P. L. Chébyshev, L. Euler.

En la formación y desarrollo de los métodos modernos de la matemática de cálculo desempeña gran papel la escuela soviética de matemáticas que está en vanguardia de la ciencia mundial.

Así, en el dominio de aproximación, de estabilidad y convergencia do los esquemas de diferencias los trabajos fundamentales de importancia primordial fueron cumplidos por S. K. Godunov, V. S. Riábenki, A. A. Samarski, A. V. Filíppov y otros.

Las obras de A. A. Dorodnitsin, M. V. Kéldysh, V. I. Krylov, O. A. Ladýzhenskaya, G. I. Marchuk, A. A. Samarski, S. L. Sóbolev dieron gran aportación al desarrollo de los métodos numéricos y a su aplicación práctica a los problemas de la física matemática.

Los métodos do resolución de los problemas incorrectos se desarrollan a base del concepto de regulación, introducido por el académico A. N. Tíjonov.

En el álgebra lineal los métodos de cálculo obtuvieron gran desarrollo, en primer lugar, merced a las investigaciones realizadas por V. V. Voevodin, L. V. Kantoróvich, D. K. Faddéev y N. V. Faddéeva.

En el campo de optimización de los métodos numéricos es necesario mencionar importantes trabajos de los matemáticos soviéticos N. S. Bajválov, A. N. Kolmogórov, S. L. Sóbolev y otros.

Antes de pasar a la exposición inmediata del material básico procuremos dar una característica breve de la matemática de cálculo. Para esto vamos a responder a las tres preguntas siguientes:

1) ¿Qué cosa es la matemática de cálculo?

2) ¿Cuáles distinciones de principio de la matemática de cálculo permitieron separarla de las matemáticas generales y convertirla en autónoma?

3) ¿Qué importancia tiene la matemática de cálculo para la economía nacional?

1. Actualmente por término «Matemática de cálculo» suele entenderse la parte de matemáticas que estudia un círculo de cuestiones relacionadas con la aplicación de los ordenadores.

En la matemática de cálculo pueden destacarse tres tendencias. La primera es la vinculada con el uso de los ordenadores en diferentes esferas de la actividad científica y práctica que incluye, en particular, la solución numérica de diferentes problemas matemáticos. La segunda es la tendencia ligada con el desarrollo de nuevos métodos numéricos y algoritmos y con el perfeccionamiento de los existentes. La tercera es la relacionada con las cuestiones de la interacción del hombre con el ordenador.

Este libro está dedicado a la primera del las tendencias mencionadas, es decir, a la aplicación de los métodos numéricos para la resolución de los problemas prácticos.

De fundamento sobre el cual se apoya la matemática de cálculo sirven los medios de cálculo y, en primer lugar, los ordenadores que están desarrollándose hoy día con ritmos sin precedentes y esto es un rasgo característico del período actual del progreso técnico. Así, durante 30 años, la velocidad de cómputo ha aumentado de una operación por segundo (en la regla de cálculo) hasta 3 000 000 operaciones por segundo, o sea, $3 \cdot 10^6$ veces. Cabe señalar que desde el tiempo de invención de la máquina de vapor la velocidad de viaje ha aumentado de 13 km/h (velocidad del caballo) a 40 000 km/h (velocidad de la nave cósmica), es decir, sólo $3 \cdot 10^3$ veces.

2. Vamos a ilustrar las particularidades características de la matemática de cálculo, las cuales la distinguen de las así llamadas matemáticas «puras», citando los ejemplos siguientes.

Desde el punto de vista del matemático «puro» resolver un problema significa demostrar la existencia de su solución y señalar el proceso que converge a la solución. En cambio, para el que calcula el tiempo de obtención de la solución, o sea, la velocidad de convergencia del proceso, resulta ser con frecuencia un factor de mayor importancia. Así, es sabido que la solución del sistema de n ecuaciones algebraicas puede obtenerse teóricamente para todo n fijado como resultado del número finito de operaciones, por ejemplo, con ayuda de los métodos de Cramer o de Gauss. Por eso desde el punto de vista del matemático «puro» tal problema se considera ya resuelto. Sin embargo, durante la realización práctica de los métodos indicados se encuentran dos dificultades de principio que no siempre, ni mucho menos, son superables. La primera dificultad consiste en lo siguiente. Con n bastante grande el número de operaciones, aunque sea finito, alcanza una magnitud tan enorme que la ejecución de todas estas operaciones llega a ser imposible incluso para los ordenadores más potentes. Así, para resolver el sistema de n ecuaciones con el método de Cramer se necesita cumplir $n \cdot n!$ operaciones lo que para $n = 20$ constituye $4{,}6 \cdot 10^{19}$ operaciones. Si el ordenador tiene una rapidez de $3 \cdot 10^6$ operaciones por segundo, debería estar trabajando ininterrumpidamente acerca de medio millón de años. El método de Gauss resulta mucho más económico: el número de operaciones, al resolver el mismo problema, tiene el orden de n^3.

No obstante, un número de operaciones tan grande engendra la segunda dificultad de principio: el error de cálculo acumulado en cada operación ejerce una influencia tan grande sobre el resultado final que a menudo está muy lejos de la solución verdadera.

Actualmente los métodos exactos suelen utilizarse para resolver los sistemas cuyo orden no supere 10^3. Por eso, desde el punto de vista del «calculador» el problema de resolver sistemas cuyo orden es mayor que 10^3 dista mucho de ser trivial. Para resolver tales sistemas se emplean los métodos iterativos los cuales, a pesar de ser aproximados, poseen una ventaja importante de no acumular el error de cálculo al pasar de una iteración a la otra. Ahora bien, se trata de un hecho que parece paradójico a primera vista, pero a pesar de esto es característico para los métodos numéricos: preferencia de un algoritmo aproximado al exacto.

3. La ampliación considerable de las esferas de aplicación de la matemática de cálculo, incluyendo su introducción en la economía nacional, se explica principalmente por el hecho de que los fenómenos de la naturaleza y de la vida social, diferentes por su esencia, a menudo tienen una estructura formal semejante y, por consiguiente, pueden ser descritos por los mismos modelos matemáticos. Por eso la solución de los problemas, descritos por estos modelos, se puede obtener con ayuda de los mismos métodos numéricos.

Hoy día la Unión Soviética tiene un rumbo decisivo hacia la formentación ulterior de la economía y el aumento de la efectividad de la producción a costa de su intensificación. En la realización de este rumbo el papel primordial pertenece al progreso científico-técnico con la utilización de los últimos alcances de las ciencias fundamentales y aplicadas. Uno de los factores fundamentales que contribuyen a la aceleración del progreso científico-técnico y al cumplimiento de los programas complejos y de destinación especial encaminados a resolver los problemas científico-técnicos más importantes y elevar ulteriormente la productividad del trabajo lo constituye la técnica de cálculo electrónico. El desarrollo de los ordenadores y de la matemática de cálculo permitirá, por ejemplo, pasar de la automatización del mando de los sistemas técnicos y de los procesos tecnológicos a la del mando de los procesos de producción.

La amplia implantación de los modelos matemáticos y de los ordenadores en la práctica de planificación y la utilización de los mismos en las cuestiones de optimización contribuirá a resolver felizmente los problemas de desarrollo económico y social.

CAPÍTULO I

Teoría elemental de los errores

§ 1.1. Números exactos y aproximados. Fuentes de errores y su clasificación

En el proceso de resolución de un problema nos vemos obligados a tratar diferentes números que pueden ser exactos o aproximados. Los números exactos presentan el valor verdadero del número y los aproximados, un valor próximo al verdadero con la particularidad de que el grado de proximidad se determina por el error de cálculo.

Por ejemplo, en las afirmaciones «el cubo tiene 6 caras», «en la mano hay 5 dedos», «en la clase hay 32 alumnos», «en el libro hay 582 páginas» los números 6, 5, 32 y 582 son exactos.

En las afirmaciones «la anchura de la casa es de 14,25 m», «el radio de la Tierra es igual a 6000 km», «la masa de una caja de cerillas es de 10 g» los números 14,25; 6000; 10 son aproximados. Esto está relacionado, ante todo, con la imperfección de los instrumentos (medios) de medida que utilizamos. No existen los aparatos de medida absolutamente exactos, cada uno tiene su precisión, o sea, admite cierto error en las mediciones. En el segundo ejemplo, además, el carácter aproximado del número está encerrado en el mismo concepto de radio de la Tierra. La cosa consiste en que, hablando en rigor, la Tierra no es un globo y hablar sobre su radio se puede sólo aproximadamente. En el tercer ejemplo el carácter aproximado del número se determina, además, por el hecho de que diferentes cajas pueden tener diferente masa y el número 10 determina la masa de cierta caja media.

En otros casos un mismo número puede ser tanto exacto como aproximado. Así, por ejemplo, el número 3 es exacto si se trata del número de lados de un triángulo y es aproximado si se utiliza en vez del número π al calcular el área del círculo con ayuda de la fórmula $S = \pi R^2$.

En la práctica de cálculos por *número aproximado a* se entiende un número que se distingue insignificantemente del número exacto A y lo sustituye en los cálculos.

La resolución de la mayoría de problemas prácticos puede ser representada, con cierto grado de convencionalismo, en forma de dos etapas sucesivas: 1) descripción matemática del problema dado; 2) resolución del problema matemático formulado.

En la primera etapa hay dos fuentes características de errores. En primer lugar, los procesos en desarrollo real no siempre pueden ser descritos matemáticamente y las simplificaciones introducidas ofrecen sólo la posibilidad de obtener modelos más o menos idealizados. En segundo lugar, la fijación de los parámetros iniciales que se obtienen, por lo general, de un experimento que ofrece no más que un resultado aproximado, es insuficientemente exacta.

En consonancia con lo dicho el error sumario del modelo matemático y de los datos de partida da el *error de información inicial*. Teniendo presente que este error es independiente de la segunda etapa de resolución del problema, se llama con frecuencia *error inevitable*.

La solución exacta de un problema matemático (la segunda etapa), por vía analítica o en un ordenador, es, como regla, irrealizable. Así, por ejemplo, sólo para una clase muy restringida de ecuaciones diferenciales, se puede lograr la solución exacta. Por eso en los cálculos prácticos se utilizan los métodos de obtención de las soluciones aproximadas y, en primer lugar, soluciones numéricas.

Es precisamente tal sustitución forzada de la solución exacta por una aproximada que engendra el *error del método* o, como suele llamarse, *error de aproximación*.

Por último, en el proceso de resolución del problema se lleva a cabo el redondeo de los datos iniciales y de los resultados intermedios y finales. Estos errores, así como los que aparecen durante la ejecución de las operaciones aritméticas sobre los números aproximados, se transportan en una u otra medida, a los resultados de cálculos y forman el así llamado *error de cálculo* (o *error de redondeos*).

En relación con lo expuesto, al plantear un problema, se señala la precisión requerida de la solución, es decir, se asigna el error máximamente admisible en el proceso de todos los cálculos o se limita por el requisito de computar el error total del resultado. Por eso al trabajar con los números aproximados es necesario saber resolver los problemas siguientes:

1) dar las características matemáticas de la exactitud de los números aproximados;

2) conociendo el grado de precisión de los datos iniciales, estimar el grado de precisión del resultado;

3) elegir los datos iniciales con el grado de precisión que asegure la precisión prefijada del resultado;

4) construir del modo óptimo el proceso de cálculo para no efectuar cómputos que no influyan sobre las cifras exactas del resultado.

§ 1.2. Notación decimal y redondeo de los números

Todo número decimal positivo a puede ser representado en forma de una fracción decimal finita o infinita:

$$a = \alpha_1 \cdot 10^m + \alpha_2 \cdot 10^{m-1} + \ldots$$
$$\ldots + \alpha_n \cdot 10^{m-n+1} + \ldots, \qquad (1)$$

donde α_i son las cifras del número $(i = 1, 2, \ldots, n, \ldots)$ con la particularidad de que $\alpha_1 \neq 0$; m, el orden (posición) decimal superior del número a.

Ejemplo 1. Representemos el número 1905,0778 en la forma (1):

$$1905{,}0778 = 1\cdot 10^3 + 9\cdot 10^2 + 0\cdot 10^1 + 5\cdot 10^0 + \\ + 0\cdot 10^{-1} + 7\cdot 10^{-2} + 7\cdot 10^{-3} + 8\cdot 10^{-4}.$$

Cada unidad del i-ésimo orden respectivo, si se considera de la izquierda a la derecha, tiene el propio valor 10^{m-i+1} llamado *valor posicional*. Así, el primer (a la izquierda) valor posicional es 10^m y el segundo valor posicional, 10^{m-1}, etc.

En el ejemplo considerado el valor posicional correspondiente a la cifra 9 es $10^{3-2+1} = 100$; a la cifra 5 es $10^{3-4+1} = 1$; a la cifra 8 es $10^{3-8+1} = 0{,}0001$.

En la práctica de cálculos surge a menudo la necesidad de *redondear* un número, o sea, reemplazarlo con otro que tiene una menor cantidad de cifras. En este caso se conservan una o más cifras, contando de la izquierda a la derecha, y se omiten todas las sucesivas.

Las más usadas son las siguientes **reglas de redondeo**:

1°. *Si las cifras a omitir constituyen un número que sea mayor de la mitad de la unidad del último orden conservado, la última cifra quedada se toma con exceso (aumenta en unidad).*

En cambio, *si las cifras a suprimir constituyen un número que sea menor que la mitad de la unidad del último orden conservado, las cifras que se quedan no varían.*

2°. *Si las cifras a omitir constituyen un número que sea igual a la mitad de la unidad del último orden conservado, entonces la última cifra quedada se toma con exceso si es impar y no se somete a la variación si es par.*

Esta regla suele llamarse **regla de la cifra par.**

Ejemplo 2. Redondear los números siguientes: $A_1 = 12{,}7852$; $A_2 = 394{,}261$; $A_3 = 6{,}265001$; $A_4 = 147{,}5$; $A_5 = 148{,}5$ hasta tres cifras.

△ Siguiendo la 1ª regla de redondeo, obtenemos $a_1 = 12{,}8$; $a_2 = 394$; $a_3 = 6{,}27$, puesto que $0{,}0852 > 0{,}5\cdot 10^{-1}$; $0{,}261 < 0{,}5\cdot 10^0$; $0{,}005001 > 0{,}5\cdot 10^{-2}$.

Siguiendo la 2ª regla de redondeo, obtenemos $a_4 = 148$; $a_5 = 148$, puesto que la cifra 7 es impar y 8 es par. ▲

En algunos casos, que actualmente tienen cada vez mayor aplicación, se utiliza una regla de redondeo más simple. Esta regla consiste en omitir todas las cifras comenzando con cierto orden (posición). Valiéndonos de esta regla, obtendríamos los siguientes valores al redondear los números del ejemplo considerado: $a_1 = 12{,}7$; $a_2 = 394$; $a_3 = 6{,}26$; $a_4 = 147$; $a_5 = 148$.

§ 1.3. Errores absoluto y relativo

Sea A un número exacto y a, su valor aproximado. Si $a < A$, se dice que el número a es un *valor aproximado del número A por defecto*; si $a > A$, *un valor aproximado por exceso*.

La diferencia entre el número exacto A y su valor aproximado a se llama *error*.

Por lo general, la magnitud del error $A - a$ e incluso su signo no se puede determinar, puesto que se desconoce el número exacto A. Por eso en vez del mismo error se utiliza la cota superior de su magnitud absoluta.

Se denomina *error absoluto* de un número aproximado a a la magnitud Δ_a que satisface la desigualdad

$$\Delta_a \geqslant |A - a|. \tag{1}$$

El error absoluto es la cota superior de desviación del número exacto A respecto al aproximado:

$$a - \Delta_a \leqslant A \leqslant a + \Delta_a. \tag{2}$$

La desigualdad (2) se escribe frecuentemente en la forma siguiente:

$$A = a \pm \Delta_a. \tag{3}$$

En calidad de error absoluto se toma, en la medida de lo posible, el número mínimo. Por ejemplo, midiendo la longitud de un segmento, nos hemos cerciorado de que el error admitido no supera 0,5 cm; con mayor razón éste no excede 1, 2 y 3 cm. Cada uno de estos números puede considerarse valor absoluto. Sin embargo, por error absoluto conviene tomar el menor de estos números, ya que cuanto menor sea el error absoluto tanto más estrecho será el intervalo dentro del cual se asigna el número exacto.

En la práctica se emplean a menudo las expresiones siguientes: «con precisión hasta 0,01», «con precisión hasta 1 cm», etc. Esto quiere decir que el error absoluto es igual, respectivamente, a 0,01; 1 cm, etc.

Ejemplo 1. Hemos medido la longitud de un segmento L con precisión hasta 0,05 cm y obtenido $l = 18{,}4$ cm. Aquí el error absoluto $\Delta_l = 0{,}05$ cm. De acuerdo con la fórmula (3) es necesario escribir $l = 18{,}4 \pm 0{,}05$ cm. La magnitud exacta de la longitud del segmento está encerrada, según la fórmula (2), entre las cotas siguientes: $18{,}35 \leqslant L \leqslant 18{,}45$.

Ejemplo 2. Se ha medido la longitud de un segmento L con ayuda de una regla en la cual el valor de una división es de 0,1 cm. Resultó que el valor exacto de L está entre 4,6 y 4,7 cm. En este caso como valor aproximado debe tomarse $l = 4{,}65$, o, sea, el centro del intervalo dentro del cual se encuentra el número exacto L. Es evidente que en este caso de error absoluto sirve la mitad del valor de una división de la regla: $\Delta_l = 0{,}05$. Ahora bien, $L = 4{,}65 \pm 0{,}05$ cm.

El error absoluto representa sólo el aspecto cuantitativo del error sin reflejar el aspecto cualitativo, es decir, sin mostrar cómo hemos realizado la medición o el cálculo: bien o mal. Efectivamente, supongamos que midiendo con la misma regla, en la cual el valor de una división es de 1 cm, la longitud del tablero de una mesa y el espesor del mismo, hemos obtenido los resultados siguientes (en cm): el espesor $L_1 = 2 \pm 0{,}05$ y la longitud $L_2 = 100 \pm 0{,}5$. Tanto en la primera medición como en la segunda el error absoluto es igual y constituye 0,5 cm. No obstante, se ve que la segunda medición está cumplida más cualitativamente que la primera. Con el fin de estimar la calidad de los cálculos cumplidos o las mediciones respectivas se introduce el concepto de error relativo.

Se llama *error relativo* de un número aproximado a la magnitud δ_a que satisface la desigualdad:

$$\delta_a \geqslant \left| \frac{A-a}{a} \right|, \quad a \neq 0. \tag{4}$$

En particular, por error relativo se puede tomar

$$\delta_a = \frac{\Delta_a}{|a|}, \quad a \neq 0 \tag{5}$$

y la relación (3) puede ser representada en la forma

$$A = a\,(1 \pm \delta_a). \tag{6}$$

Nótese que el error relativo es un número abstracto y se expresa, a veces, en tanto por ciento.

Retornando a la medición del espesor del tablero de la mesa y de la longitud del mismo, determinemos sus errores relativos:

$$\delta_{L_1} = 0{,}5/2 = 0{,}25 \text{ o bien } 25\%;$$
$$\delta_{L_2} = 0{,}5/100 = 0{,}005 \text{ o bien } 0{,}5\%.$$

En tales casos se dice que la longitud del tablero de la mesa está medida relativamente mejor (en 50 veces) que su espesor.

Ejemplo 3. Un valor exacto A se encuentra en el intervalo [23,07; 23,10]. Determinar su valor aproximado, así como los errores absoluto y relativo.

△ Tomamos por valor aproximado el centro del intervalo dado: $a = 23{,}085$. El error absoluto es la mitad de su longitud: $\Delta_a = 0{,}015$. Tomemos por error relativo $\delta_a = \Delta_a/a = 0{,}000604...$ La magnitud del error suele redondearse hasta una o dos cifras distintas del cero. Por eso se puede poner $\delta_a = 0{,}07\%$. Nótese que en los problemas de semejante género el redondeo del error se ejecuta en la dirección de aumento del mismo para asegurar el cumplimiento de las desigualdades (2) y (4). ▲

Ejemplo 4. Determinar (en tanto por ciento) el error relativo del número aproximado $a = 35{,}148$, si $A = 35{,}148 \pm 0{,}00074$.

△ Hagamos uso de la fórmula (5). Tenemos

$$\delta_a = \Delta_a/|a| = 0{,}00074/35{,}148 = 0{,}000022 \approx 0{,}003\%. \quad ▲$$

Ejemplo 5. Determinar el valor absoluto del número aproximado $a = 4{,}123$, si $\delta_a = 0{,}01\%$.

△ Escribamos los porcentajes en forma de la fracción decimal y para determinar el error absoluto utilicemos la fórmula (5). Entonces

$$\Delta_a = |a| \cdot \delta_a = 4{,}123 \cdot 0{,}0001 \approx 0{,}0005;$$
$$A = 4{,}123 \pm 0{,}0005. \ \blacktriangle$$

Ejemplo 6. Determinar en qué caso la calidad de cálculos es superior:

$$A_1 = 13/19 \approx 0{,}684, \text{ o bien } A_2 = \sqrt{52} \approx 7{,}21.$$

$\triangle$ Para hallar los errores absolutos tomemos los números a_1 y a_2 con mayor número de signos decimales: $13/19 \approx 0{,}68421$; $\sqrt{52} \approx$ $\approx 7{,}2111$ ... Determinamos los valores absolutos rodendeándolos con exceso:

$$\Delta_{a_1} = |0{,}68421 \ldots - 0{,}684| \approx 0{,}00022;$$
$$\Delta_{a_2} = |7{,}2111 \ldots - 7{,}21| \approx 0{,}0012.$$

Encontramos los errores relativos:

$$\delta_{a_1} = \Delta_{a_1}/|a_1| = 0{,}00022/0{,}684 \approx 0{,}00033 \approx 0{,}04\%;$$
$$\delta_{a_2} = \Delta_{a_2}/|a_2| = 0{,}0012/7{,}21 \approx 0{,}00017 \approx 0{,}02\%.$$

En el segundo caso la calidad de cálculos resulta superior, ya que $\delta_{a_2} < \delta_{a_1}$. $\blacktriangle$

§ 1.4. Cifras significativas justas

Al resolver los problemas se impone con frecuencia la condición: calcular el resultado con precisión hasta 0,1; 0,01, etc. Puede crearse la impresión de que la precisión de cálculos se determina por el número de cifras decimales puestas después de la coma. No obstante, no es cierto. La exactitud de cálculo se determina por el número de cifras del resultado que son fiables.

Se llaman *cifras significativas* de un número todas sus cifras, a excepción de los ceros, puestas a la izquierda de la primera cifra distinta del cero.

Los ceros puestos al fin de un número son siempre cifras significativas (en el caso contrario no se escriben).

Ejemplo 1. Los números 0,001604 y 30,500 tienen, respectivamente, 4 y 5 cifras significativas.

Al escribir los números enteros pueden haber algunos pormenores. Si, por ejemplo, queremos mostrar que en el número 400 000 los últimos tres ceros no son significativos, el número dado ha de escribirse en la forma de dos factores: $400 \cdot 10^3$ ó $40{,}0 \cdot 10^4$, o bien $0{,}400 \cdot 10^6$. La última forma de la notación se llama *normalizada* y es preferible. En este caso se dice que 400 es *mantisa* del número y 6, su *orden*.

Recuérdese que todo número decimal positivo, exacto y aproximado, puede ser representado en la forma

$$a = \alpha_1 \cdot 10^m + \alpha_2 \cdot 10^{m-1} + \ldots + \alpha_n \cdot 10^{m-n+1} \quad + \ldots$$

La cifra α_n del número aproximado a se llama *cifra significativa justa* (o simplemente *justa*) siempre que se cumpla la desigualdad

$$|A-a| \leqslant 0{,}5 \cdot 10^{m-n+1}, \tag{1}$$

o sea siempre que el valor absoluto de la diferencia entre el número exacto y su valor aproximado no sobrepase la mitad de la unidad del orden decimal en que está α_n.

Puesto que de ordinario en vez de $|A-a|$ se considera el error absoluto Δ_a, la desigualdad (1) se reemplaza a menudo por la siguiente:

$$\Delta_a \leqslant 0{,}5 \cdot 10^{m-n+1}, \tag{2}$$

puesto que al cumplirse esta desigualdad se cumple la desigualdad inicial (1).

Por otro lado, si se da un número n de cifras justas del número aproximado a, por valor absoluto se puede tomar

$$\Delta_a = 0{,}5 \cdot 10^{m-n+1}. \tag{3}$$

Si la desigualdad (2) no se cumple, la cifra α_n se llama *dudosa*. Es evidente que si la cifra α_n es justa, todas las precedentes (a la izquierda de ella) también son justas.

Ejemplo 2. El número $a = 23{,}10$ se ha obtenido por redondeo de cierto número exacto. ¿Cuántas cifras justas contiene el número a?

△ Al redondear los números según la regla de la cifra par el error absoluto no puede superar la mitad de la unidad del último orden conservado. Esto quiere decir que en el número obtenido por redondeo todas las cifras quedadas son justas. Es evidente que en el caso en cuestión todas las cuatro cifras son justas y el error $\Delta_a = 0{,}005$. ▲

Ejemplo 3. El número $a = 23{,}071937$ contiene cinco cifras justas. Determinar el error absoluto del mismo.

△ Hagamos uso de la fórmula (3). Aquí $m = 1$, $n = 5$, por lo que por error absoluto se puede tomar $\Delta_a = 0{,}5 \cdot 10^{1-5+1} = 0{,}0005$. ▲

Ejemplo 4. El error absoluto del número $a = 705{,}1978$ es igual a $\Delta_a = 0{,}3$. Determinar qué cifras en el número a son justas y redondear el número a, conservando sólo las cifras justas.

△ Utilicemos la fórmula (2). Aquí $m = 2$, $\Delta_a = 0{,}3$ y n ha de determinarse de la desigualidad $0{,}3 \leqslant 0{,}5 \cdot 10^{3-n}$. Por verificación directa nos convencemos de que el n máximo que satisfaga esta desigualdad es igual a 3 y la cifra 5 es justa: $0{,}3 < 0{,}5 \cdot 10^{2-3+1}$, y la cifra 1 es dudosa: $0{,}3 > 0{,}5 \cdot 10^{2-4+1}$.

Por consiguiente, el número $a = 705{,}1978$ tiene tres cifras justas. Vamos a redondearlo hasta tres cifras: $a_1 = 705$. En este caso el error total es igual a la suma del error inicial y del error de redondeo: $\Delta_a = 0{,}3 + 0{,}2 = 0{,}5$, así que se puede escribir: $A = 705 \pm 0{,}5$. ▲

En las tablas matemáticas todas las cifras significativas colocadas son, como regla, justas. Así, en las conocidas Tablas de V. M. Bradis los valores de seno se dan con error absoluto que no excede de $0{,}5 \cdot 10^{-4}$.

El último tiempo está utilizándose con más frecuencia el concepto de cifras significativas justas en sentido lato. Este concepto está vinculado con la regla elemental de redondeo la cual hemos mencionado al fin del § 1.2.

La cifra α_n del número aproximado

$$a = \alpha_1 \cdot 10^m + \alpha_2 \cdot 10^{m-1} + \ldots + \alpha_n \cdot 10^{m-n+1} + \ldots$$

se llama *cifra significativa justa en sentido lato*, si se cumple la desigualdad

$$\Delta_a \leqslant 1 \cdot 10^{m-n+1}, \tag{4}$$

o sea, si el error absoluto del número a no supera la unidad del orden decimal que incluye α_n.

§ 1.5. Relación entre el número de cifras justas y el error del número

El número de cifras justas de un número aproximado se determina por la desigualdad

$$|A - a| \leqslant 0{,}5 \cdot 10^{m-n+1}, \tag{1}$$

lo que se desprende de la definición de la cifra significativa justa.

Dividiendo ambos miembros de la desigualdad (1) por $|a|$, obtenemos

$$\left|\frac{A-a}{a}\right| \leqslant \frac{0{,}5 \cdot 10^{m-n+1}}{|\alpha_1 \cdot 10^m + \alpha_2 \cdot 10^{m-1} + \ldots + \alpha_n \cdot 10^{m-n+1} + \ldots|} \leqslant$$
$$\leqslant \frac{0{,}5 \cdot 10^{m-n+1}}{\alpha_1 \cdot 10^m} = \frac{0{,}5}{\alpha_1 \cdot 10^{n-1}}. \tag{2}$$

Ahora bien, si la cifra α_n del número aproximado a es justa, por error relativo se puede tomar

$$\delta_a = \frac{0{,}5}{\alpha_1 \cdot 10^{n-1}}. \tag{3}$$

Por otro lado, para que la cifra α_n del número aproximado a sea justa es necesario que se cumpla la desigualdad

$$\delta_a \leqslant \frac{0{,}5}{(\alpha_1 + 1) \cdot 10^{n-1}}, \tag{4}$$

puesto que en este caso se cumplen las desigualdades (2) y (1).

En el caso en que se trata de cifras significativas justas en sentido lato, se puede obtener una fórmula análoga:

$$\delta_a = \frac{1}{\alpha_1 \cdot 10^{n-1}}. \tag{5}$$

Ejemplo 1. ¿Cuál es el error relativo del número aproximado $a = 4{,}176$ si todas sus cifras son justas?

△ Puesto que en el número 4,176 todas las cuatro cifras son justas, con ayuda de la fórmula (3) encontramos el error relativo

$$\delta_a = \frac{0,5}{\alpha_1 \cdot 10^{n-1}} = \frac{1}{2 \cdot 4 \cdot 10^3} \approx 0,00013 = 0,013\%.$$

Nótese que el error relativo del número a puede ser hallado valiéndose de la fórmula $\delta_a = \Delta_a/|a|$. Ya que en el número dado a todas las cifras son justas, $\Delta_a = 0,0005$. Así pues,

$$\delta_a = 0,0005/4,176 \approx 0,00012 = 0,012\%.$$

Como vemos, la diferencia no es grande, pero la aplicación de la fórmula (3) simplifica algo el cálculo. ▲

Ejemplo 2. ¿Cuál es el error relativo del número $a = 14,278$ si todas sus cifras son justas en sentido lato?

△ Puesto que todas las cinco cifras son justas en sentido lato, entonces, utilizando la fórmula (5), obtenemos

$$\delta_a = \frac{1}{\alpha_1 \cdot 10^{n-1}} = \frac{1}{1 \cdot 10^4} = 0,0001 = 0,01\%. \ ▲$$

Ejemplo 3. ¿Con cuántas cifras decimales justas hace falta tomar $\sqrt{18}$ para que el error no exceda del 0,1%?

△ Aquí $A = \sqrt{18} \approx 4, \ldots$; $\delta_a \leqslant 0,1\%$, o sea, $\delta_a \leqslant 0,001$. Tenemos $\delta_a = \frac{1}{2 \cdot 4 \cdot 10^{n-1}} \leqslant 0,001$, de donde $125 \leqslant 10^{n-1}$; $1,25 \cdot 10^2 \leqslant \leqslant 10^{n-1}$; $\log 1,25 + 2 \leqslant n - 1$; $n \geqslant 3 + \log 1,25$, es decir $n \geqslant 4$. ▲

§ 1.6. Errores de la suma y de la diferencia

Consideremos los números exactos $A_1, A_2, \ldots, A_n$ y sus valores aproximados $a_1, a_2, \ldots, a_n$. Sea $A = \sum_{i=1}^{n} A_i$ la suma de todos los números exactos y $a = \sum_{i=1}^{n} a_i$, suma de sus valores aproximados. Planteemos el problema: conociendo los errores absolutos Δ_{a_1}, $\Delta_{a_2}, \ldots, \Delta_{a_n}$ de todos los números aproximados, estimar el error absoluto de su suma a. Planteemos la diferencia

$$A - a = (A_1 - a_1) + (A_2 - a_2) + \ldots + (A_n - a_n).$$

Pasando a los valores absolutos de los miembros segundo y primero de esta relación y utilizando la propiedad de los valores absolutos, obtenemos

$$|A - a| \leqslant |A_1 - a_1| + |A_2 - a_2| + \ldots + |A_n - a_n|.$$

Por lo tanto,

$$|A - a| \leqslant \Delta_{a_1} + \Delta_{a_2} + \ldots + \Delta_{a_n} \qquad (1)$$

y por error absoluto del número aproximado a, o sea, de la suma de los números aproximados $a_1, a_2, \ldots, a_n$ se puede tomar la suma de los errores absolutos de los sumandos:

$$\Delta_a = \Delta_{a_1} + \Delta_{a_2} + \ldots + \Delta_{a_n}. \qquad (2)$$

De la última fórmula se deduce que, hablando en general, el error absoluto de la suma algebraica no debe ser menor que el error absoluto del menos exacto entre los sumandos. Por eso, con el fin de no realizar cálculos superfluos, no se debe conservar las cifras superfluas también en sumandos más exactos.

Al adicionar los números de exactitud absoluta distinta se suele proceder del modo siguiente:

1) se separa el número (o los números) de la exactitud mínima (es decir, el número que tiene el error absoluto máximo);

2) se redondean números más exactos de modo que en ellos se conserve una cifra más que en el número separado (es decir, se deja una cifra de reserva),

3) se efectúa la adición, teniendo en cuenta todas las cifras conservadas;

4) el resultado obtenido se redondea suprimiendo una cifra.

Observación. Si hay una gran cantidad de sumandos ($n > 10$), la estimación del error absoluto de la suma realizada con ayuda de la fórmula (2) resulta fuertemente aumentada, ya que de ordinario ocurre una compensación parcial de errores de signos opuestos. Si todos los sumandos están redondeados hasta el m-ésimo orden decimal, o sea, sus errores se evalúan por la magnitud $0{,}5 \cdot 10^{-m}$, la estimación estadística del error absoluto de la suma se calcula por la fórmula siguiente:

$$\Delta_a = \sqrt{n} \cdot 0{,}5 \cdot 10^{-m}. \qquad (3)$$

Ejemplo 1. Adicionar los números aproximados:

$a = 0{,}1732 + 17{,}45 + 0{,}000333 + 204{,}4 + 7{,}25 + 144{,}2 + 0{,}0112 + 0{,}634 + 0{,}0771$ cada uno de los cuales tiene justas todas las cifras escritas.

△ Elegimos los números de la mínima exactitud (de máximo error absoluto). Tales números son: 204,4 y 144,2. El error de cada uno de ellos constituye 0,05. Redondeamos los demás números, dejando una cifra (de reserva) más y sumamos todos los números:

$$\begin{array}{r} 0{,}17 \\ 17{,}45 \\ 0{,}00 \\ 204{,}4 \\ +\quad 7{,}25 \\ 144{,}2 \\ 0{,}01 \\ 0{,}63 \\ 0{,}08 \\ \hline 374{,}19 \end{array}$$

Redondeamos la suma obtenida, suprimiendo una cifra: 374,2.

Evaluemos la exactitud del resultado. El error absoluto de la suma se compone de dos sumandos:

1) del error inicial, o sea, de la suma de los errores de los números menos exactos y de los errores de redondeo de los demás números: $0{,}05\cdot 2 + 0{,}005\cdot 7 \approx 0{,}14$;

2) del error de redondeo del resultado: 0,01.

Ahora bien, el error absoluto de la suma es 0,15 y el resultado ha de escribirse en la forma $A = 374{,}2 \pm 0{,}15$. Es posible también tal forma de la notación: $A = 374{,}2 \pm 0{,}2$. ▲

De un modo análogo se procede también en el caso en que uno o varios números aproximados son negativos.

Ejemplo 2. Hallar la diferencia de los números aproximados $a = a_1 - a_2$ y estimar el error absoluto y relativo del resultado si $A_1 = 17{,}5 \pm 0{,}02$, $A_2 = 45{,}6 \pm 0{,}03$.

△ Encontramos $a = a_1 - a_2 = 17{,}5 - 45{,}6 = -28{,}1$; $\Delta_a = \Delta_{a_1} + \Delta_{a_2} = 0{,}02 + 0{,}03 = 0{,}05$. Ahora bien, $A = -28{,}1 \pm 0{,}05$. Determinemos el error relativo: $\delta_a = 0{,}05/|-28{,}1| \approx 0{,}002 \approx 0{,}2\%$. ▲

Se puede mostrar que si el error absoluto de la suma de números aproximados se determina por la fórmula (2) y el error relativo de la suma $\delta_a = \Delta_a\,|\,a\,|$, entonces

$$\delta_{\text{mín}} \leqslant \delta_a \leqslant \delta_{\text{máx}}.$$

Ejemplo 3. Estimar el error relativo de la suma de números del ejemplo 1 y compararlo con los errores relativos de los sumandos.

△ Determinemos el error relativo de la suma:

$$\delta_a = 0{,}2/374{,}2 = 0{,}0006 = 0{,}06\%.$$

Los errores relativos de los sumandos constituyen:

$$\begin{gathered}
\delta_{a_1} = 0{,}005/0{,}17 = 3\%; \quad \delta_{a_2} = 0{,}005/17{,}45 = 0{,}03\%; \\
\delta_{a_4} = 0{,}05/204{,}4 = 0{,}03\%; \quad \delta_{a_5} = 0{,}005/7{,}25 = 0{,}07\%; \\
\delta_{a_6} = 0{,}05/144{,}2 = 0{,}04\%; \quad \delta_{a_7} = 0{,}005/0{,}01 = 50\%; \\
\delta_{a_8} = 0{,}005/0{,}63 = 0{,}8\%; \quad \delta_{a_9} = 0{,}005/0{,}08 = 7\%.
\end{gathered}$$

Así pues, $\delta_{\text{mín}} = 0{,}03\%$, $\delta_{\text{máx}} = 50\%$, $\delta_a = 0{,}06\%$, o sea, el error relativo de la suma está encerrado entre los errores relativos mínimo y máximo de los sumandos. ▲

Nótese que al sustraer los números próximos aparece con frecuencia la situación llamada *pérdida de exactitud*. Sea $x > 0$, $y > 0$ y $a = x - y$; entonces

$$\delta_a = \frac{\Delta_a}{|a|} = \frac{\Delta_x + \Delta_y}{|x - y|}.$$

Ahora bien, si los números x e y poco se distinguen uno de otro, entonces, incluso al ser pequeños los errores Δ_x y Δ_y, la magnitud del error relativo de la diferencia puede resultar considerable.

Ejemplo 4. Sea $x = 5{,}125$, $y = 5{,}135$; aquí $\Delta_x = 0{,}0005$, $\Delta_y = 0{,}0005$, $\delta_x \approx \delta_y \approx 0{,}01\%$. El error relativo de la diferencia $a = x - y$ constituye

$$\delta_a = \frac{0{,}0005 + 0{,}0005}{0{,}01} \cdot 100 = 10\%.$$

Es evidente que como resultado de la sustracción de dos números próximos puede tener lugar gran pérdida de exactitud. Para evitar esto, es necesario procurar que el esquema de cálculo se transforme de un modo tal que las pequeñas diferencias de las magnitudes se calculen inmediatamente.

Ejemplo 5. Hallar la diferencia $A = \sqrt{6{,}27} - \sqrt{6{,}26}$, y estimar el error relativo del resultado.

△ Sea $A_1 = \sqrt{6{,}27} \approx 2{,}504$; $\Delta_{a_1} = 0{,}0005$; $A_2 = \sqrt{6{,}26} \approx 2{,}502$; $\Delta_{a_2} = 0{,}0005$. Entonces $a = 2{,}504 - 2{,}502 = 0{,}2 \cdot 10^{-2}$; $\Delta_a = 0{,}0005 + 0{,}0005 = 0{,}001$, de donde

$$\delta_a = \frac{0{,}1 \cdot 10^{-2}}{0{,}2 \cdot 10^{-2}} = 0{,}5 = 50\%.$$

No obstante, cambiando el esquema de cálculo se puede obtener una estimación mucho mejor del error relativo:

$$A = \sqrt{6{,}27} - \sqrt{6{,}26} = \frac{(\sqrt{6{,}27} - \sqrt{6{,}26})(\sqrt{6{,}27} + \sqrt{6{,}26})}{\sqrt{6{,}27} + \sqrt{6{,}26}} =$$

$$= \frac{6{,}27 - 6{,}26}{\sqrt{6{,}27} + \sqrt{6{,}26}} = \frac{0{,}01}{\sqrt{6{,}27} + \sqrt{6{,}26}} \approx 0{,}2 \cdot 10^{-2} = a;$$

$$\delta_a = \frac{\Delta_{a_1} + \Delta_{a_2}}{a_1 + a_2} = \frac{0{,}001}{5{,}006} = 0{,}2 \cdot 10^{-3} = 0{,}02\%.$$

Ahora bien, al calcular con las mismas cuatro cifras justas de a_1 y a_2 hemos obtenido un resultado mucho mejor desde el punto de vista del error relativo. ▲

Ejemplo 6. Calcular el valor de la función $y = 1 - \cos x$ para los siguientes valores del argumento: 1) $x_1 = 80°$; 2) $x_2 = 1°$. Calcular los errores absoluto y relativo del resultado.

△ 1) Con ayuda de las «Tablas matemáticas de cuatro cifras» de Bradis encontramos $\cos 80° \approx 0{,}1736$ y puesto que todas las cifras de este número son justas, tenemos $\Delta_{0,1736} = 0{,}00005$. Entonces $y_1 = 1 - 0{,}1736 = 0{,}8264$ y $\Delta_{y_1} = 0{,}00005$ (del número exacto, igual a la unidad, se sustraye el número aproximado con error absoluto que no excede de 0,00005).

Por consiguiente,

$$\delta_{y_1} = 0{,}00005/0{,}8264 = 0{,}00006 = 0{,}006\%.$$

2) Tenemos $\cos 1° \approx 0{,}9998$; $\Delta_{0,9998} = 0{,}00005$; $y_2 = 1 - 0{,}9998 = 0{,}0002$; $\Delta_{y_2} = 0{,}00005$; por lo tanto,

$$\delta_{y_2} = 0{,}00005/0{,}0002 = 0{,}25 = 25\%.$$

De los ejemplos dados se ve que para pequeños valores del argumento el cálculo inmediato por la fórmula $y = 1 - \cos x$ da el error relativo del orden de 25%. Para $x = 80°$ el error relativo constituye sólo 0,006%.

Cambiemos el esquema de cálculo y para hallar los valores de la función $y = 1 - \cos x$, siempre que los valores del argumento sean pequeños, hagamos uso de la fórmula $y = 1 - \cos x = 2 \operatorname{sen}^2 (x/2)$. Designemos $A = \operatorname{sen} 0°30' \approx 0{,}0087$. Entonces $\Delta_a = 0{,}00005$; $\delta_a = 0{,}5/87 = 0{,}58\%$. Pero

$$y_2 = 2 \cdot 0{,}0087^2 = 0{,}000151;$$
$$\delta_{y_2} = 1{,}2\%$$

(véase a continuación el § 1.7). Como resultado obtenemos

$$\Delta_{y_2} = y_2\delta_{y_2} = 0{,}000151 \cdot 0{,}012 = 0{,}000002$$

(antes hemos tenido $\Delta_{y_1} = 0{,}00005$). Ahora bien, la simple transformación de la fórmula de cálculo ha permitido, con los mismos datos iniciales, obtener un resultado más exacto. ▲

Sin embargo, no siempre es posible transformar el esquema de cálculo. Por eso al sustraer números próximos uno a otro es necesario tomarlos con una cantidad suficiente de cifras justas de reserva (si esto es posible). Si se conoce que los primeros m cifras significativas pueden desaparecer y el resultado debe obtenerse con n cifras significativas justas, los datos iniciales han de tomarse con $m + n$ cifras significativas justas, como hemos hecho en el ejemplo 5.

§ 1.7. Error del producto. Número de cifras justas del producto

Error del producto. Examinemos dos números exactos A_1 y A_2 y sus valores aproximados a_1 y a_2. Sea $A = A_1A_2$ y $a = a_1a_2$. Planteemos el problema: conociendo los errores relativos δ_{a_1} y δ_{a_2}, estimar el error relativo del producto δ_a.

Representemos los valores exactos A_1 y A_2 en la forma

$$A_1 = a_1 + \Delta_1, \quad A_2 = a_2 + \Delta_2, \tag{1}$$

donde las incógnitas Δ_1 y Δ_2 satisfacen las desigualdades

$$|\Delta_1| \leqslant |a_1|\,\delta_{a_1}, \quad |\Delta_2| \leqslant |a_2|\,\delta_{a_2}. \tag{2}$$

Multiplicando los miembros segundos y primeros de las relaciones (1), obtenemos

$$A_1A_2 = a_1a_2 + \Delta_1a_2 + \Delta_2a_1 + \Delta_1\Delta_2.$$

Pasando a los valores absolutos de los miembros segundo y primero de esta relación y utilizando las propiedades de los valores absolutos, encontramos

$$|A_1A_2 - a_1a_2| \leqslant |\Delta_2a_1| + |\Delta_1a_2| + |\Delta_1\Delta_2|. \tag{3}$$

Suprimamos el último sumando del segundo miembro en virtud de su pequeñez y dividamos los miembros segundo y primero de la desigualdad por $|a| = |a_1 a_2|$. Entonces, en vista de las relaciones (2), tenemos

$$\left|\frac{A-a}{a}\right| \leqslant \delta_{a_1} + \delta_{a_2}. \tag{4}$$

De la relación obtenida se deduce que por error relativo del producto $a = a_1 a_2$ se puede tomar la suma de errores relativos de los factores

$$\delta_a = \delta_{a_1} + \delta_{a_2}. \tag{5}$$

La desigualdad (5) se extiende fácilmente al producto de unos cuantos factores, así que si $A = A_1 A_2 \ldots A_n$ y $a = a_1 a_2 \ldots a_n$, se puede tomar que

$$\delta_a = \delta_{a_1} + \delta_{a_2} + \ldots + \delta_{a_n}. \tag{6}$$

En el caso en que todos los factores, a excepción de uno, son números exactos, de la fórmula (6) se desprende que el error relativo del producto coincide con el error relativo del factor aproximado. Ahora bien, si de número aproximado sirve sólo el valor del factor a_i, entonces

$$\delta_a = \delta_{a_i}. \tag{7}$$

Observación. Al multiplicar el número aproximado a por el factor exacto k el error relativo del producto es igual al error relativo del número aproximado a y el error absoluto es $|k|$ veces mayor que el error absoluto del número aproximado.

En efecto, sea $a = ka_1$, donde k es el factor exacto, distinto del cero. Entonces, según la fórmula (7) tenemos $\delta_a = \delta_{a_1}$, o bien

$$\Delta_a = |a|\delta_a = |a|\delta_{a_1} = |ka_1| \frac{\Delta_{a_1}}{|a_1|} = |k|\,\Delta_{a_1},$$

o sea,

$$\Delta_a = |k|\Delta_{a_1}. \tag{8}$$

Conociendo el error relativo δ_a del producto a, se puede determinar su valor absoluto por la fórmula $\Delta_a = |a|\delta_a$.

Si el error relativo del producto de números aproximados se halla con ayuda de la fórmula (6), al multiplicar los números de diferente error relativo no conviene conservar las cifras superfluas en los números de menor error relativo. Se suele proceder del modo siguiente:

1) se separa el número con cantidad mínima de cifras significativas justas;

2) los factores quedados se redondean de un modo tal que contengan una cifra significativa más que la cantidad de cifras significativas justas en el número separado;

3) se conservan en el producto tantas cifras significativas cuantas cifras significativas justas tiene el menos exacto de los factores (número separado).

Ejemplo 1. Hallar el producto de los números aproximados $x_1 = 3{,}6$ y $x_2 = 84{,}489$ todas las cifras de los cuales son justas.

△ En el primer número hay dos cifras significativas justas y en el segundo, cinco. Por eso redondeamos el segundo número hasta tres cifras significativas. Una vez realizado el redondeo, tenemos $x_1 = 3{,}6$; $x_2 = 84{,}5$. De aquí

$$x_1x_2 = 3{,}6 \cdot 84{,}5 = 304{,}20 \approx 3{,}0 \cdot 10^2.$$

Como resultado hemos conservado dos cifras significativas, o sea, tantas cuantas tenía el factor con cantidad mínima de cifras significativas justas. ▲

Ejemplo 2. Determinar el producto de los números aproximados $x_1 = 12{,}4$ y $x_2 = 65{,}54$ y el número de cifras justas contenidas en el mismo si todas las cifras escritas en los factores son justas.

△ En el primero de los números hay tres cifras significativas justas y en el segundo, cuatro; se puede multiplicar los números sin redondearlos previamente; $x_1x_2 = 12{,}4 \cdot 65{,}54 = 812{,}696$. Es necesario conservar tres cifras significativas, ya que el menos exacto de los factores tiene tanta cantidad de cifras significativas justas; ahora bien, $a = 813$. Calculemos el error:

$$\delta_a = \delta_{x_1} + \delta_{x_2} = \frac{0{,}05}{12{,}4} + \frac{0{,}005}{65{,}54} = 0{,}0041.$$

Entonces $\Delta_a = 813 \cdot 0{,}0041 \approx 3{,}4$. Por lo tanto, el producto tiene dos cifras justas y debe escribirse así $A = 813 \pm 4$. ▲

Número de cifras justas del producto. Supongamos que se da el producto de k factores $(k \leqslant 10)$ $a = a_1a_2 \ldots a_k$, donde $a_i \neq 0$. Cada uno de los factores contiene n cifras justas $(n > 1)$, por lo menos. Supongamos que cada uno de los factores tiene la forma

$$\begin{aligned} a_i = {} & \alpha_i \cdot 10^{l_i} + \beta_i \cdot 10^{l_i - 1} + \\ & + \gamma_i \cdot 10^{l_i - 2} + \ldots \ (i = 1,\ 2,\ \ldots,\ k), \end{aligned} \tag{9}$$

donde α_i son las primeras cifras significativas de los factores aproximados escritos en el sistema decimal de numeración.

Para el error relativo del número aproximado que tiene n cifras justas utilicemos la fórmula

$$\delta_{a_i} = \frac{0{,}5}{\alpha_i \cdot 10^{n-1}} \ (i = 1,\ 2,\ \ldots,\ k).$$

Entonces el error relativo del producto de k números aproximados, cada uno de los cuales tiene n cifras significativas justas, es igual a

$$\begin{aligned} \delta_a & = \delta_{a_1} + \delta_{a_2} + \ldots + \delta_{a_k} = \\ & = \frac{0{,}5}{10^{n-1}} \left(\frac{1}{\alpha_1} + \frac{1}{\alpha_2} + \ldots + \frac{1}{\alpha_k} \right). \end{aligned} \tag{10}$$

Teniendo en cuenta que la cantidad de factores es no más de 10 ($k \leqslant 10$), obtenemos

$$\frac{1}{\alpha_1}+\frac{1}{\alpha_2}+\cdots+\frac{1}{\alpha_k}\leqslant 10$$

y, por consiguiente,

$$\delta_a \leqslant \frac{0,5}{10^{n-2}}.$$

Ahora bien, si todos los factores tienen n cifras significativas y la cantidad de factores no es más de 10, entonces la cantidad de cifras justas del producto es una o dos unidades menor que n. En el caso en que los factores tienen diferente exactitud, por n es necesario entender la cantidad de cifras justas contenidas en el menos exacto de los factores.

Observación. Si hay una gran cantidad de factores ($k > 10$) es cómodo hacer uso de la estimación estadística que tiene en cuenta la compensación parcial de los errores de signos opuestos. Si todos los números a_i ($i = 1, 2, \ldots, k$) tienen, aproximadamente, el mismo error relativo δ, el error relativo del producto se toma igual a

$$\delta_a = \sqrt{n}\,\delta. \tag{11}$$

Ejemplo 3. Determinar el error relativo y la cantidad de cifras justas del producto $a = 84,76 \cdot 8,436$, donde todas las cifras de los factores son justas.

△ Aquí $a_1 = 84,36$; $a_2 = 8,436$; $n_1 = n_2 = 4$. El mismo producto $a = 715,03 \ldots$ comienza con la cifra 7, es decir su $\alpha_1 = 7$. Ahora por la fórmula (10) tenemos

$$\delta_a = \frac{0,5}{10^{4-1}}\left(\frac{1}{8}+\frac{1}{8}\right)=\frac{0,5}{4\cdot 10^{4-1}}.$$

Comparando este resultado con el segundo miembro de la fórmula (4) del § 1.5, hallamos

$$\frac{0,5}{(7+1)\,10^{4-1}}<\frac{0,5}{4\cdot 10^{4-1}}<\frac{0,5}{(7+1)\,10^{3-1}}.$$

Por consiguiente, el producto tiene tres cifras justas, por lo menos.

Verifiquemos si es así. Determinemos el error absoluto con ayuda de la fórmula $\Delta_a = |\,a\,|\,\delta_a$; obtenemos $\Delta_a = 715 \cdot 0,125 \cdot 10^{-3} \approx \approx 0,09$. De aquí se desprende que el valor aproximado del producto tiene tres cifras justas y teniendo en cuenta el error de redondeo del resultado se puede escribir $A = 715 \pm 0,2$. ▲

Ejemplo 4. Determinar el error relativo del producto $a = 145,35 \cdot 1,24386$ y la cantidad de cifras justas contenidas en el mismo si los números se dan con cifras justas.

△ Aquí $a_1 = 145,35$, $n_1 = 5$, $a_2 = 1,24386$, $n_2 = 6$. Los números dados tienen diferente cantidad de cifras significativas justas; elegimos $n = 5$. Por la fórmula (10) obtenemos

$$\delta_a = \frac{0,5}{10^{5-1}}\left(\frac{1}{1}+\frac{1}{3}\right)=\frac{0,5}{7,5\cdot 10^{4-1}}.$$

Al calcular el producto $a = a_1 \cdot a_2 = 770{,}43\ \ldots$, comparemos la magnitud δ_a con el segundo miembro de la fórmula (4) del § 1.5:

$$\frac{0{,}5}{(7+1)\cdot 10^{4-1}} < \frac{0{,}5}{7{,}5\cdot 10^{4-1}} < \frac{0{,}5}{(7+1)\, 10^{3-1}}.$$

Por consiguiente, el producto tiene, como mínimo, tres cifras significativas justas. ▲

Ahora bien, en el caso desfavorable el producto de los números aproximados puede tener $n - 2$ cifras significativas justas (donde n es la cantidad mínima de cifras significativas justas de los factores dados).

§ 1.8. Error del cociente. Número de cifras justas del cociente

Error del cociente. Consideremos los números exactos A_1, A_2 y sus valores aproximados a_1, a_2 con errores aproximados Δ_{a_1}, Δ_{a_2}. Planteemos el problema: estimar el error relativo del valor aproximado del cociente $a = a_1/a_2$ para el valor exacto $A = A_1/A_2$. Sea $a_1 \neq 0$, $a_2 \neq 0$. Representemos los valores exactos A_1 y A_2 en la forma

$$A_2 = a_1 + \Delta_1, \quad A_2 = a_2 + \Delta_2, \tag{1}$$

donde las incógnitas Δ_1 y Δ_2 satisfacen las desigualdades

$$|\Delta_1| \leqslant \Delta_{a_1}, \quad |\Delta_2| \leqslant \Delta_{a_2}. \tag{2}$$

Consideremos ahora la diferencia

$$A - a = \frac{a_1 + \Delta_1}{a_2 + \Delta_2} - \frac{a_1}{a_2} = \frac{a_2\Delta_1 - a_1\Delta_2}{a_2\,(a_2 + \Delta_2)}.$$

Dividiendo los miembros segundo y primero por a, examinemos sus valores absolutos:

$$\left|\frac{A-a}{a}\right| = \left|\frac{a_2\Delta_1 - a_1\Delta_2}{a_1\,(a_2 + \Delta_2)}\right| = \left|\frac{a_2}{a_2 + \Delta_2}\right| \left|\frac{\Delta_1}{a_1} - \frac{\Delta_2}{a_2}\right|.$$

Teniendo en cuenta que Δ_2 es pequeño en comparación con a_2, pongamos aproximadamente que $a_2/(a_2 + \Delta_2) \approx 1$. Entonces, utilizando las propiedades de los valores absolutos y las desigualdades (2), obtenemos

$$\left|\frac{A-a}{a}\right| = \left|\frac{\Delta_1}{a_1} - \frac{\Delta_2}{a_2}\right| \leqslant \left|\frac{\Delta a_1}{|a_1|} + \frac{\Delta a_2}{|a_2|} = \delta_{a_1} + \delta_{a_2}.\right.$$

Ahora bien, por error relativo del cociente $a = a_1/a_2$ se puede tomar la suma de errores relativos del dividiendo y del divisor:

$$\delta_a = \delta_{a_1} + \delta_{a_2}. \tag{3}$$

Al utilizar la fórmula (3) para la estimación del error relativo del cociente hacen la aportación principal en este error los números menos exactos (que tienen el máximo error relativo). Por eso al dividir los números de diferente error relativo se suele proceder del modo siguiente:

1) se separa el número menos exacto, o sea, el número que tiene la cantidad mínima de cifras exactas;

2) se redondea el segundo número, conservando una cifra significativa más que las tiene el número separado;

3) en el cociente se conservan tantas cifras significativas cuantas las había en el número exacto menor.

Conociendo el error relativo del cociente, es fácil determinar su error absoluto con ayuda de la fórmula

$$\Delta_a = |a|\,\delta_a = \left|\frac{a_1}{a_2}\right|(\delta_{a_1} + \delta_{a_2}). \tag{4}$$

Ejemplo 1. Calcular el cociente $a = x/y$ de los números aproximados $x = 5{,}735$ e $y = 1{,}23$ si todas las cifras del dividiendo y del divisor son justas. Determinar los errores relativo y absoluto.

1) Primeramente calculemos el cociente. Puesto que el dividiendo $x = 5{,}735$ contiene cuatro cifras significativas justas y el divisor las contiene tres, se puede realizar la división sin previo redondeo; tenemos $a = 5{,}735{:}\ 1{,}23 = 4{,}66$. Como resultado hemos conservado tres cifras significativas, puesto que el número exacto mínimo (divisor) contiene tres cifras significativas justas.

2) Calculemos el error relativo del cociente por la fórmula (3), teniendo en cuenta que $\Delta_x = 0{,}0005$; $\Delta_y = 0{,}005$:

$$\delta_a = \delta_x + \delta_y = \frac{0{,}0005}{5{,}735} + \frac{0{,}005}{1{,}23} = 0{,}00009 + 0{,}0041 = 0{,}0042 \approx 0{,}5\%.$$

3) Determinemos el error absoluto

$$\Delta_a = |a|\,\delta_a = 4{,}66 \cdot 0{,}0042 = 0{,}02.$$

El resultado final, teniendo en cuenta el error de redondeo del producto (0,005), ha de escribirse así: $A = 4{,}66 \pm 0{,}03$.

Nótese que la cifra de centésimas partes es dudosa, puesto que $0{,}03 > 0{,}005$. Si queremos dejar en el resultado sólo cifras justas, es necesario redondear el resultado y tener en cuenta el error de redondeo. El número aproximado $a_1 = 0{,}5$ satisface este requisito, puesto que $\Delta_{a_1} = \Delta_a + \Delta^1_{\text{red}} = 0{,}02 + 0{,}4 = 0{,}42 < 0{,}5$.

Al mismo tiempo es imposible dejar en el número aproximado a dos cifras justas, ya que en este caso obtendríamos $a_2 = 4{,}7$ y $\Delta_{a_2} = \Delta_a + \Delta^2_{\text{red}} = 0{,}02 + 0{,}05 = 0{,}07 > 0{,}05$. ▲

Cantidad de cifras justas del cociente. Supongamos que los números aproximados

$$a_1 = \alpha_1 \cdot 10^{l_1} + \alpha_2 \cdot 10^{l_1 - 1} + \ldots;$$
$$a_2 = \beta_1 \cdot 10^{l_2} + \beta_2 \cdot 10^{l_2 - 1} + \ldots$$

tienen n cifras significativas justas cada uno. Entonces, utilizando las desigualdades

$$\delta_{a_1} = \frac{0{,}5}{\alpha_1 \cdot 10^{n-1}}, \quad \delta_{a_2} = \frac{0{,}5}{\beta_1 \cdot 10^{n-1}},$$

hallamos el error relativo del cociente $a = a_1/a_2$:

$$\delta_a = \delta_{a_1} + \delta_{a_2} = \frac{0,5}{\alpha_1 \cdot 10^{n-1}} + \frac{0,5}{\beta_1 \cdot 10^{n-1}} = \frac{0,5}{10^{n-1}} \left(\frac{1}{\alpha_1} + \frac{1}{\beta_1} \right). \quad (5)$$

Por consiguiente, si $\alpha_1 \geqslant 2$ y $\beta_1 \geqslant 2$, entonces el cociente tiene por lo menos $n - 1$ cifras significantes justas. Si $\alpha_1 = 1$ o bien $\beta_1 = 1$, entonces el cociente puede tener $n - 2$ cifras significantes justas.

Ejemplo 2. Calcular el cociente $a = 39,356 : 2,21$ y determinar cuántas cifras significativas justas se contienen en el mismo si en el dividiendo y en el divisor todas las cifras son justas.

△ 1) Puesto que en el divisor hay tres cifras significativas justas y en el dividiendo las hay cinco, redondeamos el dividiendo hasta cuatro cifras significativas y realizamos la división; tenemos $a = 39,36 : 2,21 = 17,81 \approx 17,8$ (como resultado conservamos tantas cifras significativas cuantas las hay en el número con menor cantidad de cifras significativas justas).

2) Determinemos el error relativo con ayuda de la fórmula (5) donde $n = 3$, ya que el número de exactitud menor contiene tres cifras justas; $\alpha_1 = 3$, $\beta_1 = 2$. Por lo tanto,

$$\delta_a = \frac{0,5}{10^2} \left(\frac{1}{3} + \frac{1}{2} \right) = \frac{5}{12} \cdot 10^{-2} = 0,42\%.$$

Comparando la magnitud δ_a con el segundo miembro de la fórmula (4) del § 1.5, obtenemos

$$\frac{0,5}{(1+1)\,10^{3-1}} < \frac{0,5}{1,2 \cdot 10^{3-1}} < \frac{0,5}{(1+1)\,10^{2-1}}.$$

Por lo tanto el cociente tiene, por lo menos, dos cifras significativas justas, o sea, una cifra significante menos que en el número aproximado (divisor) con menor cantidad de cifras significativas justas. ▲

Ejemplo 3. Determinar el error relativo del cociente $a = 15,834 : 1,72$ y la cantidad de cifras justas contenidas en el mismo si el dividiendo y el divisor contienen las cifras significadas justas.

△ El número de exactitud menor contiene tres cifras significativas justas. Determinemos el error relativo con ayuda de la fórmula (5):

$$\delta_a = \frac{0,5}{10^{3-1}} \left(\frac{1}{1} + \frac{1}{1} \right) = \frac{0,5 \cdot 2}{10^{3-1}}.$$

Calculando el cociente $a = 9,20 \ldots$ y comparando la magnitud δ_a con el segundo miembro de la fórmula (4) del § 1.5, obtenemos ($\alpha_1 = 9$)

$$\frac{0,5}{10 \cdot 10^{2-1}} < \frac{0,5 \cdot 2}{10^{3-1}} < \frac{0,5}{10 \cdot 10^{1-1}}.$$

Por consiguiente, podemos garantizar en el cociente sólo una cifra significativa justa lo que es dos cifras significativas menor que en el divisor, el menos exacto de los dos. ▲

§ 1.9. Errores de la potencia y de la raíz

Consideremos el número aproximado a_1 que tiene el error relativo δ_{a_1}. Supongamos que se necesita estimar el error relativo de la potencia $a = a_1^m$. Es evidente que

$$a = a_1^m = \underbrace{a_1 a_1 \ldots a_1}_{m \text{ factores}}.$$

El error relativo del producto es

$$\delta_a = \underbrace{\delta_{a_1} + \delta_{a_1} + \ldots + \delta_{a_1}}_{m \text{ sumandos}} = m\delta_{a_1}. \quad (1)$$

Ahora bien, al elevar el número aproximado a a la potencia m el error relativo de este número aumenta m veces. En los cálculos prácticos al elevar a potencia un número aproximado como resultado se conservan tantas cifras significativas cuantas se contenían en el mismo número aproximado.

Ejemplo 1. El lado del cuadrado $a = 36{,}5$ cm (con precisión hasta 1 mm). Hallar el área del cuadrado, los errores relativo y absoluto, así como la cantidad de cifras justas del resultado.

△ 1) Calculemos el área del cuadrado

$$S = a^2 = 36{,}5^2 = 1332{,}25 \approx 1{,}33 \cdot 10^3 \text{ cm}^2.$$

2) Determinemos el error relativo del área

$$\delta_s = 2\delta_a = 2 \cdot \frac{0{,}1}{36{,}5} \approx 0{,}0055 = 0{,}55\%.$$

3) Determinemos el error absoluto del área

$$\Delta_s = S\delta_s = 1{,}33 \cdot 10^3 \cdot 0{,}0055 \approx 7{,}4 \text{ cm}^2.$$

La respuesta definitiva puede ser escrita así:

$$S = (1{,}33 \pm 0{,}01) \cdot 10^3 \text{ cm}^2.$$

Ahora bien, el resultado tiene dos cifras significativas justas. ▲

Consideremos el número aproximado a_1 que tiene el error relativo δ_{a_1}. Se puede mostrar que el error relativo del número $a = \sqrt[m]{a_1}$ es m veces menor que el error relativo del número a_1:

$$\delta_a = \frac{1}{m}\delta_{a_1}. \quad (2)$$

En los problemas prácticos al extraer la raíz del número aproximado como resultado se conservan tantas cifras significativas cuantas se contenían en el número subradical.

Ejemplo 2. Determinar, con qué error relativo y con cuántas cifras significativas justas se puede hallar el lado del cuadrado si su área $S = 16{,}45$ cm² con precisión hasta 0,01.

△ Tenemos $a = \sqrt{S} = 4{,}056$ cm;

$$\delta_a = \frac{1}{2}\,\delta_s = \frac{1}{2}\cdot\frac{0{,}01}{16{,}45} = 0{,}0003 = 0{,}03\%;$$

$$\Delta_a = 4{,}056 \cdot 0{,}0003 = 1{,}3 \cdot 10^{-3}.$$

Ahora bien, teniendo en cuenta el redondeo del resultado, $A = 4{,}056 \pm 0{,}002$ cm y la cantidad de cifras significativas justas es igual a 3. ▲

§ 1.10. Reglas de cómputo de las cifras

Durante los cálculos, siempre que no se realice un cómputo estricto de errores, se recomienda usar las reglas de cómputo de las cifras. Estas reglas indican cómo hay que efectuar el redondeo de todos los resultados para que, en primer lugar, se garantice la precisión prefijada del resultado definitivo y, en el segundo lugar, no se lleven a cabo los cálculos con cifras superfluas que no ejercen influencia en las cifras justas del resultado.

Vamos a exponer las **reglas de cómputo de las cifras.**

1°. Al adicionar y sustraer los números aproximados, en el resultado conviene conservar tantas cifras decimales cuantas se contienen en el número aproximado dado con menor cantidad de cifras decimales.

2°. Al multiplicar y dividir, en el resultado es necesario conservar tantas cifras significativas cuantas se contienen en el número aproximado dado con menor cantidad de cifras significativas justas.

3°. Al elevar un número aproximado al cuadrado o al cubo en el resultado han de conservarse tantas cifras significativas cuantas se contiene en la base de la potencia.

4°. Al extraer la raíz cuadrada o cúbica de un número aproximado, en el resultado es necesario conservar tantas cifras significativas cuantas se contienen en el número subradical.

5°. Al calcular los resultados intermedios conviene conservar una cifra más que se recomienda por las reglas 1°—4°. En el resultado final esta cifra «de reserva» se suprime.

6°. Si ciertos datos tienen más cifras decimales (durante la adición y la sustracción) o más cifras significativas (durante otras operaciones) que otros, es necesario redondearlos previamente, conservando sólo una cifra «de reserva».

7°. Al calcular una expresión monomial con ayuda de los logaritmos se recomienda computar la cantidad de cifras significativas en el número aproximado dado con menor cantidad de cifras significativas y hacer uso de la tabla de logaritmos con cantidad de cifras decimales mayor en una unidad. En el resultado final la última cifra significativa se suprime.

8°. Si los datos pueden tomarse con precisión arbitraria, entonces para obtener el resultado con m cifras justas los datos iniciales han de tomarse con tanta cantidad de cifras que, según las reglas precedentes, garanticen $m+1$ cifras en el resultado.

Estas reglas se dan suponiendo que los componentes de operaciones contienen únicamente cifras justas y la cantidad de operaciones no es grande.

Ejemplo 1. Calcular $X=\frac{A^3\sqrt{B}}{C^2}$, donde $A=7{,}45\pm0{,}01$, $B=50{,}46\pm0{,}02$, $C=15{,}4\pm0{,}03$. Determinar el error del resultado.

△ Al calcular los resultados intermedios, conservaremos una cifra «de reserva», o sea, si según la regla general es necesario dejar n cifras significativas, entonces en los resultados intermedios conservaremos $n+1$ cifras. Tenemos

$$a^3=413{,}5;\ \sqrt{b}=7{,}1035;\ c^2=237{,}2;\ x=\frac{413{,}5\cdot7{,}1035}{237{,}2}=12{,}4.$$

En el resultado hemos dejado tres cifras significativas, ya que en los factores la cantidad menor de cifras significativas es igual a 3.

Vamos a computar los errores del resultado:

$$\delta_x=3\delta_a+\frac{1}{2}\,\delta_b+2\delta_c=3\cdot\frac{0{,}01}{7{,}45}+\frac{1}{2}\cdot\frac{0{,}02}{50{,}46}+$$
$$+2\cdot\frac{0{,}03}{15{,}4}\approx0{,}0041+0{,}0002+0{,}004\approx0{,}009;$$
$$\Delta_x=12{,}4\cdot0{,}009\approx0{,}12.$$

Así pues, obtenemos la respuesta: $X=12{,}4\pm0{,}2$; $\delta_x=0{,}9\%$ ▲.

Ejemplo 2. Calcular $X=\frac{(A+B)\,M}{(C-D)^2}$, donde $A=2{,}754\pm0{,}001$; $B=11{,}7\pm0{,}04$; $M=0{,}56\pm0{,}05$; $C=10{,}536\pm0{,}002$; $D=6{,}32\pm0{,}008$. Determinar los errores del resultado.

△ Hallamos

$$a+b=2{,}75+11{,}7=14{,}45;$$
$$\Delta_{a+b}=\Delta_a+\Delta_b+\Delta_{\text{red}}=0{,}001+0{,}04+0{,}004=0{,}045;$$
$$c-d=10{,}536-6{,}32=4{,}216;\ \Delta_{c-d}=0{,}002+0{,}008=0{,}010.$$

Por eso

$$x=\frac{14{,}45\cdot0{,}56}{4{,}216^2}=\frac{14{,}45\cdot0{,}56}{17{,}75}=0{,}456\approx0{,}46=4{,}6\cdot10^{-1};$$
$$\delta_x=\frac{0{,}045}{14{,}45}+\frac{0{,}005}{0{,}56}+2\cdot\frac{0{,}01}{4{,}216}=0{,}00311+0{,}00894+$$
$$+0{,}00474=0{,}02=2\%.$$

Por consiguiente,

$$\Delta_x = 0{,}46 \cdot 0{,}02 = 0{,}01.$$

Así pues, obtenemos la respuesta: $X = 0{,}46 \pm 0{,}01$; $\delta_x = 2\%$. ▲

Ejemplo 3. Haciendo uso de las reglas de cómputo de las cifras, calcular

$$v = \pi h^2 \left(r - \frac{h}{3}\right),$$

donde $h = 11{,}8$, $\pi = 3{,}142$, $r = 23{,}67$.

△ Hallamos

$$v = 3{,}142 \cdot 11{,}8^2 (23{,}67 - 3{,}933) = 3{,}142 \cdot 11{,}8^2 \cdot 19{,}737 =$$
$$= 3{,}142 \cdot 139{,}2 \cdot 19{,}737 = 437{,}37 \cdot 19{,}737 = 8630 \approx$$
$$\approx 8{,}63 \cdot 10^3.$$ ▲

Ejercicios

1. Ejecutar los redondeos sucesivos de los números siguientes: a) 2,75464; b) 3,14159; c) 0,56453; d) 4,1945; e) 0,60653.

2. Redondeando los números siguientes hasta tres cifras significativas, determinar los errores absoluto Δ_a y relativo (en tanto por ciento) δ_a de las aproximaciones obtenidas: a) 1,1426; b) 0,01015; c) 0,1245; d) 921,55; e) 0,002462.

3. Determinar el error absoluto Δ_x de los siguientes números absolutos basándose en su error relativo: a) $x = 2{,}52$; $\delta_x = 0{,}7\%$; b) $x = 0{,}986$; $\delta_x = 10\%$; c) $x = 46{,}75$; $\delta_x = 1\%$; d) $x = 199{,}1$; $\delta_x = 0{,}01$; e) $x = 0{,}86341$; $\delta_x = 0{,}0004$.

4. Determinar la cantidad de cifras significativas justas para los siguientes números aproximados: a) $39{,}285 \pm 0{,}034$; b) $1{,}2785 \pm \pm 0{,}0007$; c) $183{,}3 \pm 0{,}1$; d) $0{,}056 \pm 0{,}0003$; e) $84{,}17 \pm 0{,}0073$.

5. Determinar, cuál de las igualdades es más exacta: a) $6/25 \approx \approx 1{,}4$ o bien $1/3 \approx 0{,}333$; b) $1/9 \approx 0{,}1$ o bien $1/3 \approx 0{,}33$; c) $15/7 \approx \approx 2{,}14$ o bien $1/9 \approx 0{,}11$, d) $6/7 \approx 0{,}86$ ó $\pi \approx 22/7$; e) $\pi \approx 3{,}142$ o bien $\sqrt{10} \approx 3{,}1623$.

Indicación. Hallar previamente los errores relativos. Más exacta es la igualdad cuyo error relativo es menor.

6. Redondear las cifras dudosas del número $A = 47{,}453 \pm 0{,}024$, conservando en éste las cifras justas.

7. Redondear las cifras dudosas del número $A = 46{,}3852 \pm \pm 0{,}0031$, conservando en éste las cifras justas.

8. Redondear las cifras dudosas del número aproximado $a = = 3{,}2873$ si $\delta_a = 0{,}1\%$, conservando en el mismo las cifras justas.

9. Hallar los errores absolutos y relativos de los números aproximados si éstos tienen sólo las cifras justas: a) $a = 0{,}7538$; b) $a = = 17{,}354$.

Indicación. Utilizar la fórmula (3) del § 1.5.

10. Calcular las siguientes expresiones y dar las estimaciones de sus errores. En la respuesta conservar todas las cifras justas y una dudosa. Todos los números se dan con cifras justas.

a) $y=\frac{3,07\cdot 326}{36,4\cdot 323}$; b) $y=\frac{36\cdot 245\cdot 85}{975\cdot 642}$;

c) $y=\frac{37,2+458,67}{36,5\cdot 246}$; d) $y=\frac{96,891-4,25}{33,3+0,426}$.

11. Utilizándo las reglas de cómputo de las cifras, calcular:

a) $s=\frac{h^2}{18}\cdot\frac{a^2+4ab+b^2}{(a+b)^2}$, donde $h=2,73$, $a=0,152$, $b=0,328$;

b) $s=\frac{1}{4}\pi(a^2-b^2)$, donde $a=0,937$, $b=0,0630$.

CAPÍTULO II

Álgebra matricial y algunas nociones de la teoría de los espacios vectoriales lineales

§ 2.1. Matrices y vectores. Operaciones principales con matrices y vectores

Una tabla rectangular que se compone de elementos (en este caso, de números) y tiene m filas (líneas) y n columnas se llama *matriz* de tamaño $m \times n$. Los elementos de la matriz se designan con a_{ij}, donde i es el número de la fila y j, el número de la columna en el cruce de los cuales está este elemento.

Por ejemplo,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

es la matriz de tamaño $m \times n$ que tiene m filas y n columnas.

La notación abreviada tiene la forma $A = [a_{ij}]$, donde $i = 1, 2, \dots, m$; $j = 1, 2, \dots, n$, o bien $[a_{ij}]_{m \times n}$.

Si en la matriz la cantidad de filas no es igual a la de columnas, $m \neq n$, la matriz se denomina *rectangular*.

La matriz que tiene una sola fila, o sea, $m = 1$ se llama *matriz fila* (o *vector fila*), por ejemplo,

$$A = [a_{11} a_{12} \dots a_{1n}] \text{ o bien } A = [1\ 2\ 3\ 4].$$

La matriz que tiene una sola columna, o sea, $n = 1$ se denomina *matriz columna* (o *vector columna*), por ejemplo,

$$A = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} \quad \text{o bien} \quad A = \begin{bmatrix} 1 \\ 0 \\ 4 \\ 5 \end{bmatrix}.$$

A continuación la matriz fila o la matriz columna llamarémoslas *vector* y designaremos con $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ o bien $\mathbf{x} = [x_1\ x_2, \dots, x_n]$.

Los números $x_1, x_2, \ldots, x_n$ se llaman *coordenadas* (o *elementos*) del vector $\mathbf{x}$. Puesto que la cantidad de coordenadas del vector es, por definición, su dimensión, el vector $\mathbf{x}$ es n-dimensional.

Si en la matriz la cantidad de filas es igual a la de columnas, o sea, $m = n$ la matriz se define como *cuadrada*. Tal matriz puede escribirse en la forma

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{bmatrix}.$$

Para una matriz cuadrada la cantidad total de filas o columnas se denomina *orden* de la matriz.

Se llama *diagonal principal* de una matriz cuadrada a la que pasa por los ángulos superior izquierdo e inferior derecho, o sea, el conjunto de los elementos que tiene la forma a_{ii}, donde $i = 1, 2, \ldots, n$.

La matriz cuadrada en la cual todos los elementos dispuestos fuera de la diagonal principal son iguales a cero se denomina matriz *diagonal*. Esta matriz tiene la forma siguiente:

$$A = \begin{bmatrix} a_{11} & 0 & 0 & \ldots & 0 \\ 0 & a_{22} & 0 & \ldots & 0 \\ 0 & 0 & a_{33} & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & a_{nn} \end{bmatrix}.$$

La matriz diagonal en la cual todos los elementos dispuestos en la diagonal principal son iguales a la unidad se denomina matriz *unidad*. La matriz unidad se designa con el símbolo E y tiene la forma

$$E_n = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix}.$$

El índice n señala el orden de la matriz unidad. En el cálculo matricial la matriz unidad desempeña el papel de unidad.

La matriz cuadrada en la cual todos los elementos se disponen simétricamente respecto a la diagonal principial se define como *simétrica*. En caso de la matriz simétrica tiene lugar la igualdad $a_{ij} = a_{ji}$ $(i \neq j)$.

Por ejemplo,

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 2 & 3 & 5 \\ 4 & 5 & 6 \end{bmatrix}$$

es una matriz simétrica.

La matriz en la cual todos los elementos son iguales a cero se denomina *nula* y se designa con O. Si es necesario indicar el número de filas y de columnas, se escribe

$$O_{m\times n}=\begin{bmatrix} 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 \end{bmatrix}.$$

Dos matrices $A=[a_{ij}]$ y $B=[b_{ij}]$ se llaman *iguales* si: 1) son del mismo tamaño, o sea, la cantidad de filas de la matriz A es igual a la de filas de la matriz B y la cantidad de columnas de la matriz A es igual a la de columnas de la matriz B; 2) los elementos correspondientes de estas matrices son iguales entre sí. Ahora bien, si

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

y $a_{ij}=b_{ij}$ $(i=1, 2, \dots, m;\ j=1, 2, \dots, n)$, entonces $A=B$.

Se llama *suma* de dos matrices del mismo tamaño $A+B=[a_{ij}]+[b_{ij}]$ $(i=1, 2, \dots, m;\ j=1, 2, \dots, n)$ a la matriz $C=[c_{ij}]$ del mismo tamaño en la cual los elementos c_{ij} con iguales a las sumas de los elementos correspondientes a_{ij} y b_{ij} de las matrices A y B, o sea, $c_{ij}=a_{ij}+b_{ij}$.

La *diferencia* de las matrices se define de un modo análogo a la suma, no obstante, en los elementos de la matriz a sustraer el signo cambia por el opuesto: $D=A-B$; $d_{ij}=a_{ij}-b_{ij}$ $(i=1, 2, \dots, m;\ j=1, 2, \dots, n)$.

Se llama *producto de la matriz* $A=[a_{ij}]$ por el número α la matriz cuyos elementos se obtienen por la multiplicación de todos los elementos de la matriz A por el número α:

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

La matriz $-A=(-1)A$ es *opuesta* a la matriz A.

La adición de las matrices se subordina a las leyes siguientes: 1°) $A+(B+C)=(A+B)+C$; 2°) $A+B=B+A$; 3°) $A+O=A$; 4°) $A+(-A)=O$.

El producto de la matriz por un número se subordina a las leyes siguientes:

1°) $1\cdot A=A$; 2°) $0\cdot A=O$; 3°) $\alpha(\beta A)=(\alpha\beta)A$.

Ejemplo 1. Sea

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

Entonces

$$C=A+B=\begin{bmatrix} 2 & 1 & 5 & -3 \\ 2 & 3 & 2 & 1 \\ -6 & 0 & 1 & 1 \\ 3 & 2 & 8 & -2 \end{bmatrix}; \quad D=A-B=\begin{bmatrix} 0 & 3 & 1 & 11 \\ -6 & -3 & 6 & 9 \\ -8 & 12 & 1 & 3 \\ -1 & -4 & 0 & 8 \end{bmatrix}.$$

Ejemplo 2. El producto de la matriz

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

por el número 2 es la matriz

$$2A=\begin{bmatrix} 2 & 6 & -8 \\ 0 & 2 & -4 \\ -6 & 2 & 10 \end{bmatrix}.$$

Se llama *producto* AB *de dos matrices*

$$A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \text{ y } B=\begin{bmatrix} b_{11} & b_{12} & \dots & b_{1q} \\ b_{21} & b_{22} & \dots & b_{2q} \\ \dots & \dots & \dots & \dots \\ b_{n1} & b_{n2} & \dots & b_{nq} \end{bmatrix}$$

que tienen los tamaños $m \times n$ y $n \times q$, respectivamente,a la matriz

$$C=\begin{bmatrix} c_{11} & c_{12} & \dots & c_{1q} \\ c_{12} & c_{22} & \dots & c_{2q} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mq} \end{bmatrix}$$

de tamaño $m \times q$. Nótese que la matriz $C = AB$ está definida sólo cuando la cantidad de columnas de la matriz A es igual a la de filas de la matriz B.

Los elementos de la matriz C se calculan con ayuda de la **regla** siguiente: *para hallar el elemento c_{ij} dispuesto en la i-ésima fila y en j-ésima columna del producto de dos matrices es necesario los elementos de la i-ésima fila de la primera matriz multiplicar por los elementos correspondientes de la j-ésima columna de la segunda y sumar los productos obtenidos*:

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj}$$
$$(i = 1, 2, \dots, m; \quad j = 1, 2, \dots, q).$$

Por ejemplo, $c_{23} =$
$= a_{21}b_{13} + a_{22}b_{23} + \ldots + a_{2n}b_{n3}$, $c_{41} = a_{41}b_{11} +$
$+ a_{42}b_{21} + \ldots + a_{4n}b_{n1}$, etc.

Ejemplo 3. $$AB = \begin{bmatrix} 3 & 2 & 8 & 1 \\ 1 & -4 & 0 & 3 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 1 & -3 \\ 0 & 1 \\ 3 & 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 3\cdot 2 + 2\cdot 1 + 8\cdot 0 + 1\cdot 3 & 3(-1) + 2\cdot(-3) + 8\cdot 1 + 1\cdot 1 \\ 1\cdot 2 + (-4)\cdot 1 + 0\cdot 0 + 3\cdot 3 & 1(-1) + (-4)\cdot(-3) + 0\cdot 1 + 3\cdot 1 \end{bmatrix} =$$

$$= \begin{bmatrix} 11 & 0 \\ 7 & 14 \end{bmatrix}.$$

Aquí $AB = [a_{ij}]_{2\times 4}\cdot[b_{ij}]_{4\times 2} = C = [c_{ij}]_{2\times 2}$.

Ejemplo 4. $$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot 2 + 3\cdot 3 \\ 4\cdot 1 + 5\cdot 2 + 6\cdot 3 \\ 7\cdot 1 + 8\cdot 2 + 9\cdot 3 \end{bmatrix} = \begin{bmatrix} 14 \\ 32 \\ 50 \end{bmatrix}.$$

Aquí $AB = [a_{ij}]_{3\times 3}[b_{ij}]_{3\times 1} = [c_{ij}]_{3\times 1}$.

Ejemplo 5. $$AB = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} = \begin{bmatrix} 1\cdot 5 + 2\cdot 7 & 1\cdot 6 + 2\cdot 8 \\ 3\cdot 5 + 4\cdot 7 & 3\cdot 6 + 4\cdot 8 \end{bmatrix} =$$

$$= \begin{bmatrix} 19 & 22 \\ 43 & 50 \end{bmatrix}, \quad BA = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 5\cdot 1 + 6\cdot 3 & 5\cdot 2 + 6\cdot 4 \\ 7\cdot 1 + 8\cdot 3 & 7\cdot 2 + 8\cdot 4 \end{bmatrix} =$$

$$= \begin{bmatrix} 23 & 34 \\ 31 & 46 \end{bmatrix},$$ o sea, $AB \neq BA$.

Ejemplo 6. $$AB = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \begin{bmatrix} 3 & 2 & 4 \\ 2 & 1 & 3 \\ 4 & 3 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 1\cdot 3 + 2\cdot 2 + 3\cdot 4 & 1\cdot 2 + 2\cdot 1 + 3\cdot 3 & 1\cdot 4 + 2\cdot 3 + 3\cdot 0 \\ 4\cdot 3 + 5\cdot 2 + 6\cdot 4 & 4\cdot 2 + 5\cdot 1 + 6\cdot 3 & 4\cdot 4 + 5\cdot 3 + 6\cdot 0 \end{bmatrix} =$$

$$= \begin{bmatrix} 19 & 13 & 10 \\ 46 & 31 & 31 \end{bmatrix}; \quad BA = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 1 & 3 \\ 4 & 3 & 0 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$$ no existe, ya que la cantidad de columnas de la matriz A no es igual a la de filas de la matriz B.

El producto de las matrices se subordina a las leyes siguientes:
1°) $A(BC) = (AB)C$; 2°) $\alpha(AB) = (\alpha A)B$;
3°) $(A + B)C = AC + BC$; 4°) $EA = A$.

Nótese que $AB \neq BA$, o sea, el producto de dos matrices no posee, en el caso general, la propiedad de conmutatividad. La matriz unidad constituye la excepción $AE = EA = A$. Así pues, en el caso general no se puede cambiar de lugar los factores sin cambiar su producto. Si cambia el orden de factores puede resultar que, en general, es imposible realizar la multiplicación de matrices (véase el ejemplo 6).

Hablando del producto AB de dos matrices A y B, suponemos que la matriz B se multiplica por la A a la izquierda o bien que la matriz A se multiplica por la B a la derecha.

El producto de unas cuantas matrices ABC es necesario entender así: la matriz A se multiplica a la derecha por la B y la matriz obtenida se multiplica a la derecha por la C, etc. La cantidad de matrices a multiplicar puede ser cualquiera siempre que se pueda multiplicar cada dos matrices dispuestas una junto a la otra.

La matriz A^n se llama n-ésima *potencia* de la matriz A. Si A es una matriz cuadrada y n, un número positivo entero, entonces

$$A^n = \underbrace{A \cdot A \cdot A \ldots A}_{n \text{ factores}}.$$

Las operaciones de adición y de multiplicación por un número sobre las matrices columna y las matrices fila (o sea, sobre los vectores) se llevan a cabo análogamente a las operaciones correspondientes sobre las matrices cuadradas. Así, se llama suma *de dos vectores* $\mathbf{x} = [x_1 x_2 \ldots x_n]$ e $\mathbf{y} = [y_1 y_2 \ldots y_n]$ el vector $\mathbf{z} = [z_1 z_2 \ldots z_n]$ que tiene por coordenadas $z_1 = x_1 + y_1$, $z_2 = x_2 + y_2$, $\ldots$, $z_n = x_n + y_n$ y se llama *producto del vector* $\mathbf{x} = [x_1 x_2 \ldots x_n]$ por el número α el vector $\mathbf{z} = \alpha\mathbf{x} = [\alpha x_1 \alpha x_2 \ldots \alpha x_n]$.

Ejemplo 7. La suma de los vectores $\mathbf{x} = [1\ 2\ 3]$ e $\mathbf{y} = [-5\ \ -2\ \ 4]$ es el vector $\mathbf{z} = [-4\ \ 0\ \ 7]$; el producto del vector $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}$ por el número $\alpha = 2$ es el vector $\mathbf{z} = \begin{bmatrix} 2 \\ 4 \\ -6 \end{bmatrix}$.

§ 2.2. Matriz transpuesta

Si en la matriz

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

de tamaño $m \times n$ se reemplazan las filas por las columnas correspondientes, se obtiene la matriz

$$A^t = \begin{bmatrix} a_{11} & a_{21} & \ldots & a_{m1} \\ a_{12} & a_{22} & \ldots & a_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ a_{1n} & a_{2n} & \ldots & a_{mn} \end{bmatrix}$$

de tamaño $n \times m$ que se denomina *transpuesta* respecto a la matriz A.

Ejemplo 1. Para la matriz

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

de tamaño 3×4 de matriz transpuesta sirve

$$A^t = \begin{bmatrix} 1 & 2 & 1 \\ 2 & 0 & 2 \\ -3 & 4 & 3 \\ 5 & 2 & 4 \end{bmatrix}$$

de tamaño 4×3.

Ejemplo 2. Para la matriz fila $B = [1\ 2\ 3\ 4]$ de matriz transpuesta sirve la matriz columna

$$B^t = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}.$$

Nótense las siguientes propiedades de la operación de transposición.

1°. Si sobre la matriz A se realiza dos veces la operación de transposición, la matriz no cambia:

$$(A^t)^t = A.$$

2°. La matriz transpuesta de la suma de dos matrices es igual a la suma de las matrices transpuestas:

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

lo que se deduce de la definición de la suma de dos matrices.

3°. La matriz transpuesta del producto de dos matrices es igual al producto de las matrices transpuestas tomadas en el orden inverso:

$$(AB)^t = B^t A^t.$$

La matriz $(AB)^t$ está obtenida multiplicando los elementos de las filas de la matriz A por los elementos de las columnas de la matriz B, reemplazando sucesivamente las filas por las columnas. El mismo resultado puede obtenerse multiplicando los elementos de las columnas de la matriz B (de las filas de la B^t) por los elementos de la matriz A (de las columnas de la A^t).

§ 2.3. Determinante de la matriz. Propiedades del determinante y reglas de su cálculo

Sea A la matriz cuadrada arbitraria de orden n

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}.$$

Con la matriz A está ligado el *determinante* el cual se designa con d, D, $\det A$ o bien $|A|$:

$$d = D = \det A = |A| = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}.$$

El determinante de la matriz es el número que se calcula por ciertas reglas que consideraremos a continuación.

En el determinante se distinguen dos diagonales: principal y secundaria. Al igual que en la matriz cuadrada, la *diagonal principal* se compone de elementos a_{ii}, donde $i = 1, 2, \dots, n$. La *diagonal secundaria* pasa perpendicularmente a la principal a partir del ángulo derecho superior del determinante al izquierdo inferior. El orden del determinante corresponde al de la matriz de determinante de la cual él sirve.

Si el orden de al matriz es igual a la unidad, o sea, esta matriz consta de un solo elemento a_{11}, entonces el número igual a este elemento se llama *determinante de primer orden* correspondiente a tal matriz. Sea dada la matriz cuadrada de segundo orden

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Se denomina *determinante de segundo orden*, correspondiente a esta matriz, el número

$$\det A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}. \qquad (2)$$

La fórmula (2) expresa la regla de cálculo del determinante de segundo orden: *el determinante de segundo orden es igual al producto de los elementos de la diagonal principal menos el producto de los elementos de la diagonal secundaria.*

Ejemplo 1. Calcular el determinante de la matriz $A = \begin{bmatrix} 1 & 2 \\ 4 & 5 \end{bmatrix}$.

△ $\det A = \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} = 1 \cdot 5 - 4 \cdot 2 = -3$. ▲

Se denomina *determinante de tercer orden* el número

$$\det A = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} =$$

$$= a_{11}a_{22}a_{33} + a_{13}a_{21}a_{32} + a_{12}a_{23}a_{31} -$$

$$- a_{13}a_{22}a_{31} - a_{21}a_{12}a_{33} - a_{32}a_{23}a_{11}. \qquad (3)$$

Ahora bien, *cada término del determinante de tercer orden no es sino el producto de tres sus elementos tomados uno a uno de cada fila y de cada columna. Estos productos se toman con signos determinados: con signo más se toman tres términos compuestos por los elementos de la diagonal* *principal y por los dispuestos en los vértices de los triángulos isósceles cuyas bases son paralelas a la diagonal principal* (fig. 2.1); *con signo menos se toman tres términos situados de un modo análogo respecto a la diagonal secundaria* (fig. 2.2). Dicha regla se llama **regla de los triángulos**.

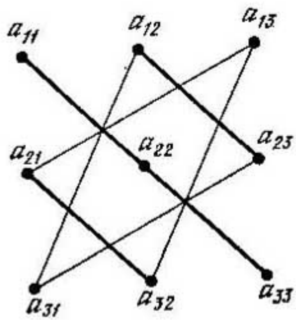


Fig. 2.1

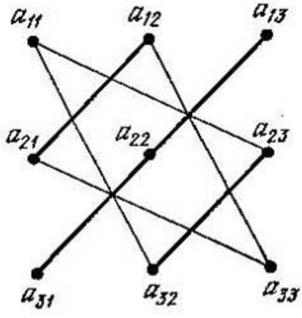


Fig. 2.2

Ejemplo 2. Calcular el determinante de la matriz

$$A=\begin{bmatrix} 1 & 2 & 3 \\ -4 & 5 & -1 \\ 2 & 1 & 2 \end{bmatrix}.$$

$\triangle$ $\det A=\begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & -1 \\ 2 & 1 & 2 \end{vmatrix}=1\cdot5\cdot2+$

$+3\cdot1\cdot(-4)+2\cdot(-1)\cdot2-2\cdot5\cdot3-1\cdot(-1)\cdot1-2\cdot(-4)\cdot2=-19.$ $\blacktriangle$

Consideremos ahora el determinante de todo orden n, donde $n\geqslant 2$. Para calcular tal determinante es necesario introducir el concepto de menor y de complemento algebraico.

Se llama *menor del elemento* a_{ij} del determinante de n-ésimo orden (1) el determinante de orden $(n-1)$ obtenido del inicial borrando la i-ésima fila y la j-ésima columna (de aquella fila y de aquella columna en cuya intersección está el elemento a_{ij}).

El menor del elemento a_{ij} se designa M_{ij}. Aquí el primer subíndice designa el número de la fila y el segundo, el número de la columna las cuales se borran.

Por ejemplo, en el determinante de tercer orden

$$d=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

de menor del elemento a_{12} sirve el determinante de segundo orden

$$M_{12}=\begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix}.$$

Se denomina *complemento algebraico* del elemento a_{ij} del determinante de n-ésimo orden (1) el número

$$A_{ij}=(-1)^{i+j}M_{ij}.$$

Es evidente que si la suma $i+j$ es par, el complemento algebraico tiene el mismo signo que el menor; en cambio, si la suma $i+j$ es impar, el signo cambia por el contrario.

Teorema 1. *El determinante es igual a la suma de los productos de toda fila (columna) suya por los complementos algebraicos correspondientes*

$$\det A=\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix}=$$

$$=a_{i1}A_{i1}+a_{i2}A_{i2}+\dots+a_{in}A_{in}=\sum_{j=1}^{n} a_{ij}A_{ij}\ (i=1,\ 2,\ \dots,\ n) \quad (4)$$

o bien

$$\det A=a_{1j}A_{1j}+a_{2j}A_{2j}+\dots+a_{nj}A_{nj}=$$
$$=\sum_{i=1}^{n} a_{ij}A_{ij}\quad (j=1,\ 2,\ \dots,\ n). \quad (5)$$

La fórmula (4) se llama *desarrollo del determinante según los elementos de la i-ésima fila*, y la fórmula (5), *desarrollo del determinante según los elementos de la j-ésima columna.*

Al desarrollar el determinante de segundo orden según los elementos de toda fila (columna) se obtiene la fórmula (2) citada anteriormente; al desarrollar el determinante de tercer orden según los elementos de toda fila (columna) se obtiene la fórmula (3) (regla de los triángulos).

Ejemplo 3. Calcular el determinante $d=\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix}$, desarrollándolo según los elementos de la primera fila.

△ En consonancia con la fórmula (4) tenemos

$$d=a_{11}A_{11}+a_{12}A_{12}.$$

Puesto que $A_{11}=(-1)^{1+1}\cdot 4=4$; $A_{12}=(-1)^{1+2}\cdot 3=-3$, entonces $d=1\cdot 4+2(-3)=-2$. ▲

Ejemplo 4. Calcular el determinante $d=\begin{vmatrix} 3 & 2 & 1 \\ 2 & 5 & 3 \\ 3 & 4 & 2 \end{vmatrix}$, desarrollándolo según los elementos de la segunda columna.

△ Con ayuda de la fórmula (5) obtenemos

$$d=a_{12}A_{12}+a_{22}A_{22}+a_{32}A_{32}.$$

Luego hallamos

$$A_{12}=(-1)^{1+2}\begin{vmatrix}2 & 3\\ 3 & 2\end{vmatrix}=-(4-9)=5;$$

$$A_{22}=(-1)^{2+2}\begin{vmatrix}3 & 1\\ 3 & 2\end{vmatrix}=6-3=3;$$

$$A_{32}=(-1)^{3+2}\begin{vmatrix}3 & 1\\ 2 & 3\end{vmatrix}=-(9-2)=-7;$$

de donde $d=2\cdot 5+5\cdot 3+4(-7)=-3$. ▲

Teorema 2 (corolario del teorema 1). *Si todos los elementos de la i-ésima fila (columna) del determinante d, a excepción de uno, por ejemplo, de a_{ik}, son iguales a cero, el determinante es igual al producto del elemento a_{ik} por su complemento algebraico:*

$$d=a_{ik}A_{ik}. \tag{6}$$

Ejemplo 5. Calcular el determinante de cuarto orden

$$d=\begin{bmatrix}1 & 1 & 3 & 4\\ 1 & 0 & 0 & 4\\ 3 & 0 & 0 & 2\\ 0 & 0 & -5 & -11\end{bmatrix}$$

desarrollándolo según los elementos de la segunda columna.

△ Puesto que $a_{22}=a_{32}=a_{42}=0$, entonces por la fórmula (6) obtenemos

$$d=a_{12}A_{12}=1\cdot(-1)^{1+2}\begin{vmatrix}1 & 0 & 4\\ 3 & 0 & 2\\ 0 & -5 & -11\end{vmatrix}$$

de donde, volviendo a desarrollar el determinante obtenido de tercer orden según los elementos de la segunda columna, hallamos

$$d=-(-5)(-1)^{3+2}\begin{vmatrix}1 & 4\\ 3 & 2\end{vmatrix}=-5(2-12)=50. \ ▲$$

Teorema 3. *La suma de los productos de los elementos de una fila o de una columna cualquiera del determinante por los complementos algebraicos de los elementos correspondientes de la fila (o de la columna) paralela es igual a cero.*

Así, para el determinante de tercer orden

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

son válidas las igualdades $a_{21}A_{11} + a_{22}A_{12} + a_{23}A_{13} = 0$, $a_{31}A_{21} + a_{32}A_{22} + a_{33}A_{23} = 0$, etc.

Enumeremos ahora las propiedades del determinante.

1°. *El determinante no cambia en caso de la transposición*:

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = \begin{vmatrix} a_{11} & a_{21} & \dots & a_{n1} \\ a_{12} & a_{22} & \dots & a_{n2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{nn} \end{vmatrix}.$$

Esto quiere decir que las filas y las columnas del determinante son equitativas.

De esta propiedad se desprende que el *determinante de la matriz A es igual al de la matriz transpuesta A^t*.

Por ejemplo,

$$\det A = \begin{vmatrix} 1 & 2 & 3 \\ -4 & 5 & -1 \\ 2 & 1 & 2 \end{vmatrix} = \begin{vmatrix} 1 & -4 & 2 \\ 2 & 5 & 1 \\ 3 & -1 & 2 \end{vmatrix} = 1\cdot 5\cdot 2 +$$

$$+3\cdot(-4)\cdot 1+2\cdot(-1)\cdot 2-3\cdot 5\cdot 2-1\cdot(-1)\cdot 1-2\cdot(-4)\cdot 2 = -19.$$

2°. *Si una de las filas o una de las columnas del determinante se compone de ceros, el determinante es igual a cero.*

Por ejemplo,

$$\det A = \begin{vmatrix} 0 & 1 & 2 \\ 0 & 3 & 4 \\ 0 & 10 & 15 \end{vmatrix} = 0\cdot 3\cdot 15 + 0\cdot 1\cdot 4 + 0\cdot 2\cdot 10 -$$

$$-0\cdot 3\cdot 2 - 0\cdot 1\cdot 15 - 0\cdot 4\cdot 10 = 0.$$

3°. *Al permutar dos filas o dos columnas el determinante cambia sólo el signo.*

Por ejemplo,

$$\det A = \begin{vmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ -4 & 5 & -1 \end{vmatrix} = 1\cdot 1\cdot(-1) + 2\cdot 2\cdot(-4) + 2\cdot 5\cdot 3 -$$

$$-3\cdot 1\cdot(-4) - 5\cdot 2\cdot 1 - 2\cdot 2\cdot(-1) = 19$$

(compárese con el ejemplo que ilustra la propiedad 1°).

4°. *El determinante que contiene dos filas iguales o dos columnas iguales es igual a cero.*

Por ejemplo,

$$\det A = \begin{vmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 2 & 1 & 2 \end{vmatrix} = 1\cdot 2\cdot 2 + 2\cdot 3\cdot 2 + 1\cdot 1\cdot 3 - 2\cdot 2\cdot 3 -$$

$$-1\cdot 2\cdot 2 - 1\cdot 1\cdot 3 = 0.$$

5°. *Si todos los elementos de cierta fila o de cierta columna del determinante se multiplican por el número $k \neq 0$, entonces el mismo determinante se multiplicará por este número.*

De otro modo esta propiedad puede ser enunciada así: *el factor de todos los elementos de cierta fila o de cierta columna puede ser sacado del signo del determinante.*

Por ejemplo,

$$\det A = \begin{vmatrix} 3 & 6 & 9 \\ 2 & 1 & 2 \\ -4 & 5 & -1 \end{vmatrix} = 3 \begin{vmatrix} 1 & 2 & 3 \\ 2 & 1 & 2 \\ -4 & 5 & -1 \end{vmatrix} = 3 \cdot 19 = 57$$

(compare con el ejemplo que ilustra la propiedad 3°).

6°. *El determinante que contiene dos filas proporcionales es igual a cero.*

Por ejemplo,

$$\det A = \begin{vmatrix} 3 & 6 & 9 \\ 1 & 2 & 3 \\ 2 & 1 & 2 \end{vmatrix} = 3 \begin{vmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \\ 2 & 1 & 2 \end{vmatrix} = 3 \cdot 0 = 0.$$

7°. *Si todos los elementos de la i-ésima fila del determinante de n-ésimo orden están representados en forma de la suma de dos sumandos: $a_{ij} = b_{ij} + c_{ij}$ $(j = 1, 2, \ldots, n)$, el determinante dado es igual a la suma de dos determinantes en los cuales todas las filas, a excepción de la i-ésima, son las mismas que en el determinante asignado, y la i-ésima fila en uno de los sumandos consta de los elementos b_{ij}, mientras que en el otro, de los elementos c_{ij},* o sea,

$$\det A =$$

$$= \begin{vmatrix} a_{11} & a_{12} & a_{13} & \ldots a_{1n} \\ a_{21} = b_{21} + c_{21} & a_{22} = b_{22} + c_{22} & a_{23} = b_{23} + c_{23} & \ldots a_{2n} = b_{2n} + c_{2n} \\ a_{31} & a_{32} & a_{33} & \ldots a_{3n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & a_{n3} & \ldots a_{nn} \end{vmatrix} =$$

$$= \begin{vmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ b_{21} & b_{22} & b_{23} & \ldots & b_{2n} \\ a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nn} \end{vmatrix} + \begin{vmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ c_{21} & c_{22} & c_{23} & \ldots & c_{2n} \\ a_{31} & a_{32} & a_{33} & \ldots & a_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nn} \end{vmatrix}.$$

Por ejemplo,

$$\det A = \begin{vmatrix} 2 & 7 & 9 \\ 1 & 3 & 4 \\ 1 & 0 & 2 \end{vmatrix} = \begin{vmatrix} 1 & 3 & 4 \\ 1 & 3 & 4 \\ 1 & 0 & 2 \end{vmatrix} + \begin{vmatrix} 1 & 4 & 5 \\ 1 & 3 & 4 \\ 1 & 0 & 2 \end{vmatrix} = 0 - 1 = -1.$$

8°. *Si una de las filas del determinante representa la suma de cualesquiera otras filas o la suma de los productos de cualesquiera otras filas del determinante por el número k, el determinante es igual a cero.* (Esto se deduce de las propiedades 6° y 7° del determinante.)

Por ejemplo,

$$\det A=\begin{vmatrix}1&2&3\\3&2&7\\1&0&2\end{vmatrix}=\begin{vmatrix}1&2&3\\1&2&3\\1&0&2\end{vmatrix}+2\begin{vmatrix}1&2&3\\1&0&2\\1&0&2\end{vmatrix}=0+2\cdot 0=0.$$

9°. *El determinante no cambiará si a los elementos de una de sus filas (columnas) se les añaden los elementos correspondientes de otra fila (columna) multiplicados por un mismo número.*

Por ejemplo, $\det A=\begin{vmatrix}2&7&9\\1&3&4\\1&0&2\end{vmatrix}=-1$ (véase el ejemplo que ilustra la propiedad 7°). Multipliquemos la tercera fila por 3 y sumemos lo obtenido a la segunda fila, entonces

$$\det A=\begin{vmatrix}2&7&9\\4&3&10\\1&0&2\end{vmatrix}=2\cdot 3\cdot 2+1\cdot 7\cdot 10+4\cdot 0\cdot 9-1\cdot 3\cdot 9-$$

$$-4\cdot 7\cdot 2-2\cdot 0\cdot 10=-1.$$

Aplicando las propiedades mencionadas de los determinantes, se puede simplificar el problema de calcular los determinantes de n-ésimo orden. Las transformaciones que no varían la magnitud del determinante suele denominarles *elementales.*

Ejemplo 6. Calcular el determinante

$$d=\begin{vmatrix}1&1&1&1\\-1&2&3&4\\-2&3&6&10\\3&4&10&20\end{vmatrix},$$

desarrollándolo según los elementos de la primera columna.

△ En consonancia con la fórmula (5) $d=a_{11}A_{11}+a_{21}A_{21}+$ $+a_{31}A_{31}+a_{41}A_{41}$. Hallamos los complementos algebraicos. Tenemos

$$A_{11}=(-1)^{1+1}=\begin{vmatrix}2&3&4\\3&6&10\\4&10&20\end{vmatrix}=2\cdot 2\begin{vmatrix}2&3&2\\3&6&5\\2&5&5\end{vmatrix}=$$

$$=4\left(2\begin{vmatrix}6&5\\5&5\end{vmatrix}-3\begin{vmatrix}3&2\\5&5\end{vmatrix}+2\begin{vmatrix}3&2\\6&5\end{vmatrix}\right)=4(2\cdot 5-3\cdot 5+2\cdot 3)=4.$$

Hemos sacado del signo del determinante de A_{11} el factor común 2 de la tercera fila y el factor común 2 de la tercera columna y luego hemos desarrollado el determinante obtenido según los elementos de la primera columna.

Calculamos

$$A_{21}=(-1)^{2+1}\begin{vmatrix}1&1&1\\3&6&10\\4&10&20\end{vmatrix}=-2\begin{vmatrix}1&1&1\\3&6&10\\2&5&10\end{vmatrix}=$$

$$=-2\left(\begin{vmatrix}6&10\\5&10\end{vmatrix}-\begin{vmatrix}3&10\\2&10\end{vmatrix}+\begin{vmatrix}3&6\\2&5\end{vmatrix}\right)=-2(10-10+3)=-6.$$

Hemos sacado el factor común 2 a partir de la tercera fila y desarrollado el determinante según los elementos de la primera fila.

Análogamente obtenemos

$$A_{31}=(-1)^{3+1}\begin{vmatrix}1&1&1\\2&3&4\\4&10&20\end{vmatrix}=2\begin{vmatrix}1&1&1\\2&3&4\\2&5&10\end{vmatrix}=$$

$$=2\left(\begin{vmatrix}3&4\\5&10\end{vmatrix}-\begin{vmatrix}2&4\\2&10\end{vmatrix}+\begin{vmatrix}2&3\\2&5\end{vmatrix}\right)=2(10-12+4)\ 4;$$

$$A_{41}=(-1)^{4+1}\begin{vmatrix}1&1&1\\2&3&4\\3&6&10\end{vmatrix}=$$

$$=-\left(\begin{vmatrix}3&4\\6&10\end{vmatrix}-\begin{vmatrix}2&4\\3&10\end{vmatrix}+\begin{vmatrix}2&3\\3&6\end{vmatrix}\right)=-6+8-3=-1.$$

Desarrollando el determinante de A según los elementos de la primera columna hallamos definitivamente

$$d=1\cdot4+(-1)\cdot(-6)+(-2)\cdot4+3\cdot(-1)=-1. \ \blacktriangle$$

El cálculo del determinante se simplifica considerablemente si, utilizando las propiedades del mismo, éste se transforma de un modo tal que al calcularlo se puede aplicar la fórmula (6).

Ejemplo 7. Calcular el determinante

$$d=\begin{vmatrix}1&1&1&1\\-1&2&3&4\\-2&3&6&10\\3&4&10&20\end{vmatrix}$$

con ayuda de las transformaciones elementales.

$\triangle$ Aplicando las transformaciones elementales del determinante, hagamos ceros todos los elementos de la primera fila, a excepción de $a_{11} = 1$. Para esto, conservando la primera columna sin variar, multipliquemos todos sus elementos por (-1) y adicionemos lo obtenido sucesivamente a las columnas 2, 3 y 4. Tenemos

$$d=\begin{vmatrix} 1 & 1 & 1 & 1 \\ -1 & 2 & 3 & 4 \\ -2 & 3 & 6 & 10 \\ 3 & 4 & 10 & 20 \end{vmatrix}=\begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 3 & 4 & 5 \\ 1 & 5 & 8 & 12 \\ 1 & 1 & 7 & 17 \end{vmatrix}=1\cdot(-1)^{1+1}\begin{vmatrix} 3 & 4 & 5 \\ 5 & 8 & 12 \\ 1 & 7 & 17 \end{vmatrix}.$$

En el determinante de tercer orden obtenido hagamos nulos también todos los elementos de la tercera fila, salvo el primero. Para esto, conservando sin variar la primera columna, multipliquémosla sucesivamente por (-7) y a por (-17) y adicionemos lo obtenido a las columnas 2 y 3, respectivamente; desarrollemos el determinante obtenido según los elementos de la tercera fila:

$$d=\begin{vmatrix} 3 & -17 & -46 \\ 5 & -27 & -73 \\ 1 & 0 & 0 \end{vmatrix}=1\cdot(-1)^{3+1}\begin{vmatrix} -17 & -46 \\ -27 & -73 \end{vmatrix}=-1. \ \blacktriangle$$

§ 2.4. Matriz inversa

Una matriz cuadrada se llama *inversa* respecto a la matriz cuadrada dada si la multiplicación de aquélla tanto a la derecha como a la izquierda por la matriz dada arroja la matriz unidad. Para la matriz A la matriz inversa se designa A^{-1}. Por definición.

$$AA^{-1} = A^{-1}A = E. \tag{1}$$

La matriz cuadrada se considera *regular* (no especial o *no degenerada*) si su determinante no es igual a cero. En cambio, si el determinante de la matriz es igual a cero, ella se dice *singular* (o degenerada).

Teorema. *Para que la matriz cuadrada A tenga la matriz inversa es necesario y suficiente que el determinante de la matriz A sea distinto del cero, o sea, que la matriz A sea regular.*

La determinación de la matriz inversa se denomina *inversión* de la matriz dada.

Consideremos el proceso de inversión de una matriz. Sea

$$A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \tag{2}$$

una matriz cuadrada regular de n-ésimo orden cuyo determinante $d \neq 0$. Formemos la matriz a partir de los complementos algebraicos de los elementos de la matriz dada y luego transpongámosla. La matriz obtenida se llama *adjunto* (o *asociada*) respecto a la matriz A y se designa $\widetilde{A}$:

$$\widetilde{A}=\begin{bmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{bmatrix}. \tag{3}$$

Calculando los productos $A\widetilde{A}$ y $\widetilde{A}A$ por las reglas de multiplicación de las matrices, obtenemos

$$A\widetilde{A}=\widetilde{A}A=dE. \tag{4}$$

□ Vamos a demostrar la validez de las igualdades (4) citando como ejemplo una matriz de tercer orden. Sea

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \text{ y } \widetilde{A}=\begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}.$$

Entonces

$$A\widetilde{A}=\begin{bmatrix} a_{11}A_{11}+a_{12}A_{12}+ & a_{11}A_{21}+a_{12}A_{22}+ & a_{11}A_{31}+a_{12}A_{32}+ \\ +a_{13}A_{13} & +a_{13}A_{23} & +a_{13}A_{33} \\ a_{21}A_{11}+a_{22}A_{12}+ & a_{21}A_{21}+a_{22}A_{22}+ & a_{21}A_{31}+a_{22}A_{32}+ \\ +a_{23}A_{13} & +a_{23}A_{23} & +a_{23}A_{33} \\ a_{31}A_{11}+a_{32}A_{12}+ & a_{31}A_{21}+a_{32}A_{22}+ & a_{31}A_{31}+a_{32}A_{32}+ \\ +a_{33}A_{13} & +a_{33}A_{23} & +a_{33}A_{33} \end{bmatrix}.$$

De acuerdo con el teorema 3 del § 2.3 todos los elementos del producto $A\widetilde{A}$, a excepción de los diagonales, son iguales a cero. Por consiguiente,

$$A\widetilde{A}=\begin{bmatrix} d & 0 & 0 \\ 0 & d & 0 \\ 0 & 0 & d \end{bmatrix}=d\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}=dE.$$

Análogamente se puede mostrar que $\widetilde{A}A=dE$. ■

Puesto que $A\widetilde{A}=\widetilde{A}A=dE$, con la particularidad de que $d\neq 0$, entonces

$$A\frac{\widetilde{A}}{d}=\frac{\widetilde{A}}{d}A=E.$$

Por otro lado, según la definición de la matriz inversa tenemos

$$AA^{-1}=A^{-1}A=E.$$

Comparando las últimas igualdades matriciales, obtenemos la fórmula para hallar la matriz inversa

$$A^{-1}=\widetilde{A}/d=\begin{bmatrix} A_{11}/d & A_{21}/d & A_{31}/d \\ A_{12}/d & A_{22}/d & A_{32}/d \\ A_{13}/d & A_{23}/d & A_{33}/d \end{bmatrix}.$$

En la forma general para una matriz cuadrada regular de n-ésimo orden la matriz inversa se calcula con ayuda de la fórmula

$$A^{-1}=\begin{bmatrix} A_{11}/d & A_{21}/d & \dots & A_{n1}/d \\ A_{12}/d & A_{22}/d & \dots & A_{n2}/d \\ A_{1n}/d & A_{2n}/d & \dots & A_{nn}/d \end{bmatrix}, \tag{5}$$

o sea, los elementos de la matriz inicial e inversa están ligados por la relación $a_{ij}^{-1} = A_{ij}/d$.

Ejemplo 1. Hallar la matriz inversa A^{-1} para la matriz

$$A=\begin{bmatrix} 2 & 3 & 3 \\ 1 & 2 & 3 \\ 2 & 4 & 5 \end{bmatrix}.$$

△ 1) Calculamos el determinante para la matriz dada A:

$$d=\begin{vmatrix} 2 & 3 & 3 \\ 1 & 2 & 3 \\ 2 & 4 & 5 \end{vmatrix}=2\cdot 2\cdot 5+3\cdot 3\cdot 2+1\cdot 4\cdot 3-2\cdot 2\cdot 3-$$

$$-1\cdot 3\cdot 5-4\cdot 3\cdot 2=-1.$$

Puesto que $d\neq 0$, la matriz inversa A^{-1} existe.

2) Hallamos los complementos algebraicos de los elementos de la matriz A:

$$A_{11}=\begin{vmatrix} 2 & 3 \\ 4 & 5 \end{vmatrix}=-2;\quad A_{21}=-\begin{vmatrix} 3 & 3 \\ 4 & 5 \end{vmatrix}=-3;\quad A_{31}=\begin{vmatrix} 3 & 3 \\ 2 & 3 \end{vmatrix}=3;$$

$$A_{12}=-\begin{vmatrix} 1 & 3 \\ 2 & 5 \end{vmatrix}=1;\quad A_{22}=\begin{vmatrix} 2 & 3 \\ 2 & 5 \end{vmatrix}=-4;\quad A_{32}=-\begin{vmatrix} 2 & 3 \\ 1 & 3 \end{vmatrix}=-3;$$

$$A_{13}=\begin{vmatrix} 1 & 2 \\ 2 & 4 \end{vmatrix}=0;\quad A_{23}=-\begin{vmatrix} 2 & 3 \\ 2 & 4 \end{vmatrix}=-2;\quad A_{33}=\begin{vmatrix} 2 & 3 \\ 1 & 2 \end{vmatrix}=1.$$

3) Componemos la matriz adjunta

$$\widetilde{A}=\begin{bmatrix} A_{11} & A_{21} & A_{31} \\ A_{12} & A_{22} & A_{32} \\ A_{13} & A_{23} & A_{33} \end{bmatrix}=\begin{bmatrix} -2 & -3 & 3 \\ 1 & 4 & -3 \\ 0 & -2 & 1 \end{bmatrix}.$$

4) Calculamos la matriz inversa

$$A^{-1}=\widetilde{A}/d=\begin{bmatrix} 2 & 3 & -3\\ -1 & -4 & 3\\ 0 & 2 & -1\end{bmatrix}.$$

Verificación: $AA^{-1}=\begin{bmatrix} 2 & 3 & 3\\ 1 & 2 & 3\\ 2 & 4 & 5\end{bmatrix}\begin{bmatrix} 2 & 3 & -3\\ -1 & -4 & 3\\ 0 & 2 & -1\end{bmatrix}=\begin{bmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{bmatrix}$. ▲

Ejemplo 2. Invertir la matriz

$$A=\begin{bmatrix} 1 & 2 & 4 & 2\\ 3 & 1 & 1 & -3\\ -2 & 3 & -1 & 1\\ -1 & 2 & 0 & 1\end{bmatrix}.$$

△ 1) Calculamos

$$d=\begin{vmatrix} 1 & 2 & 4 & 2\\ 3 & 1 & 1 & -3\\ -2 & 3 & -1 & 1\\ -1 & 2 & 0 & 1\end{vmatrix},$$

para lo cual desarrollamos el determinante según los elementos de la primera fila. Tenemos

$$A_{11}=\begin{vmatrix} 1 & 1 & -3\\ 3 & -1 & 1\\ 2 & 0 & 1\end{vmatrix}=-1+2-6-3=-8;$$

$$A_{12}=-\begin{vmatrix} 3 & 1 & -3\\ -2 & -1 & 1\\ -1 & 0 & 1\end{vmatrix}=-(-3-1+3+2)=-1;$$

$$A_{13}=\begin{vmatrix} 3 & 1 & -3\\ -2 & 3 & 1\\ -1 & 2 & 1\end{vmatrix}=9-1+12-9+2-6=7;$$

$$A_{14}=-\begin{vmatrix} 3 & 1 & 1\\ -2 & 3 & -1\\ -1 & 2 & 0\end{vmatrix}=-(1-4+3+6)=-6.$$

Por lo tanto,

$$d=1\cdot(-8)+2\cdot(-1)+4\cdot 7+2\cdot(-6)=6,$$

es decir, la matriz inversa existe.

2) Calculamos los demás complementos algebraicos de los elementos de la matriz A:

$$A_{21} = -\begin{vmatrix} 2 & 4 & 2 \\ 3 & -1 & 1 \\ 2 & 0 & 1 \end{vmatrix} = -(-2+8+4-12) = 2;$$

$$A_{22} = \begin{vmatrix} 1 & 4 & 2 \\ -2 & -1 & 1 \\ -1 & 0 & 1 \end{vmatrix} = -1-4-2+8 = 1;$$

$$A_{23} = -\begin{vmatrix} 1 & 2 & 2 \\ -2 & 3 & 1 \\ -1 & 2 & 1 \end{vmatrix} = -(3-2-8+6+4-2) = -1;$$

$$A_{24} = \begin{vmatrix} 1 & 2 & 4 \\ -2 & 3 & -1 \\ -1 & 2 & 0 \end{vmatrix} = 2-16+12+2 = 0;$$

$$A_{31} = \begin{vmatrix} 2 & 4 & 2 \\ 1 & 1 & -3 \\ 2 & 0 & 1 \end{vmatrix} = 2-24-4-4 = -30;$$

$$A_{32} = -\begin{vmatrix} 1 & 4 & 2 \\ 3 & 1 & -3 \\ -1 & 0 & 1 \end{vmatrix} = -(1+12+2-12) = -3;$$

$$A_{33} = \begin{vmatrix} 1 & 2 & 2 \\ 3 & 1 & -3 \\ -1 & 2 & 1 \end{vmatrix} = 1+6+12+2+6-6 = 21;$$

$$A_{34} = -\begin{vmatrix} 1 & 2 & 4 \\ 3 & 1 & 1 \\ -1 & 2 & 0 \end{vmatrix} = -(-2+24+4-2) = -24;$$

$$A_{41} = -\begin{vmatrix} 2 & 4 & 2 \\ 1 & 1 & -3 \\ 3 & -1 & 1 \end{vmatrix} = -(2-36-2-6-6-4) = 52;$$

$$A_{42} = \begin{vmatrix} 1 & 4 & 2 \\ 3 & 1 & -3 \\ -2 & -1 & 1 \end{vmatrix} = 1+24-6+4-12-3 = 8;$$

$$A_{43}=-\begin{vmatrix} 1 & 2 & 2 \\ 3 & 1 & -3 \\ -2 & 3 & 1 \end{vmatrix}=-(1+12+18+4+9-6)=-38;$$

$$A_{44}=\begin{vmatrix} 1 & 2 & 4 \\ 3 & 1 & 1 \\ -2 & 3 & -1 \end{vmatrix}=-1-4+36+8-3+6=42.$$

3) Componemos la matriz adjunta

$$\tilde{A}=\begin{bmatrix} A_{11} & A_{21} & A_{31} & A_{41} \\ A_{12} & A_{22} & A_{32} & A_{42} \\ A_{13} & A_{23} & A_{33} & A_{43} \\ A_{14} & A_{24} & A_{34} & A_{44} \end{bmatrix}=\begin{bmatrix} -8 & 2 & -30 & 52 \\ -1 & 1 & -3 & 8 \\ 7 & -1 & 21 & -38 \\ -6 & 0 & -24 & 42 \end{bmatrix}.$$

4) Dividimos todos los elementos de la matriz adjunta por $d=6$ y obtenemos la matriz inversa A^{-1}:

$$A^{-1}=\begin{bmatrix} -8/6 & 2/6 & -30/6 & 52/6 \\ -1/6 & 1/6 & -3/6 & 8/6 \\ 7/6 & -1/6 & 21/6 & -38/6 \\ -6/6 & 0 & -24/6 & 42/6 \end{bmatrix}.$$

Verificación:

$$AA^{-1}=\begin{bmatrix} 1 & 2 & 4 & 2 \\ 3 & 1 & 1 & -3 \\ -2 & 3 & -1 & 1 \\ -1 & 2 & 0 & 1 \end{bmatrix}\begin{bmatrix} -8/6 & 2/6 & -30/6 & 52/6 \\ -1/6 & 1/6 & -3/6 & 8/6 \\ 7/6 & -1/6 & 21/6 & -38/6 \\ -6/6 & 0 & -24/6 & 42/6 \end{bmatrix}=$$

$$=\begin{bmatrix} \frac{-8-2+28-12}{6} & \frac{2+2-4+0}{6} & \frac{-30-6+84-48}{6} & \frac{52+16-152+84}{6} \\ \frac{-24-1+7+18}{6} & \frac{6+1-1+0}{6} & \frac{-90-3+21+72}{6} & \frac{156+8-38-126}{6} \\ \frac{16-3-7-6}{6} & \frac{-4+3+1+0}{6} & \frac{60-9-21-24}{6} & \frac{-104+24+38+42}{6} \\ \frac{8-2+0-6}{6} & \frac{-2+2+0+0}{6} & \frac{30-6+0-24}{6} & \frac{-52+16+0+42}{6} \end{bmatrix}=$$

$$=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}=E. \blacktriangle$$

Si el orden de la matriz A es grande, con el método indicado de inversión de la matriz se necesita realizar un trabajo de cálculo complicado. Existen otros métodos de inversión de la matriz que examinaremos más abajo.

La determinación de la matriz inversa A^{-1} tiene importancia primordial para la resolución de los sistemas de ecuaciones lineales.

§ 2.5. Resolución de las ecuaciones matriciales

Consideremos tres tipos de las ecuaciones matriciales:

$$AX=B, \tag{1}$$

$$XA=B, \tag{2}$$

$$AXB=C, \tag{3}$$

donde

$$A=\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \cdot & \cdot\cdot\cdot\cdot & \cdot \\ a_{n1} & \dots & a_{nn}\end{bmatrix},\quad B=\begin{bmatrix} b_{11} & \dots & b_{1n} \\ \cdot & \cdot\cdot\cdot\cdot & \cdot \\ b_{n1} & \dots & b_{nn}\end{bmatrix},\quad C=\begin{bmatrix} c_{11} & \dots & c_{1n} \\ \cdot & \cdot\cdot\cdot\cdot & \cdot \\ c_{n1} & \dots & c_{nn}\end{bmatrix}$$

son las matrices cuadradas dadas del mismo tamaño y

$$X=\begin{bmatrix} x_{11} & \dots & x_{1n} \\ \cdot & \cdot\cdot\cdot\cdot & \cdot \\ x_{n1} & \dots & x_{nn}\end{bmatrix}$$

es la matriz cuadrada del mismo tamaño cuyos elementos son desconocidos.

Vamos a resolver cada una de las ecuaciones (1)—(3).

Para resolver la ecuación (1) multipliquemos a la izquierda ambos miembros suyos por A^{-1} (suponiendo que la matriz inversa A^{-1} existe):

$$A^{-1}AX = A^{-1}B.$$

Pero el producto $A^{-1}A = E$; por consiguiente, $EX = A^{-1}B$, de donde

$$X = A^{-1}B. \tag{4}$$

Ejemplo 1. Resolver la ecuación matricial

$$\underbrace{\begin{bmatrix} 2 & 5 \\ -1 & 9\end{bmatrix}}_{A} X = \underbrace{\begin{bmatrix} -2 & 3 \\ 1 & -1\end{bmatrix}}_{B}.$$

△ 1) $\det A=\begin{vmatrix} 2 & 5 \\ -1 & 9\end{vmatrix}=23\neq 0.$

2) Hallamos A^{-1}. Puesto que $A_{11}=9$, $A_{21}=-5$, $A_{12}=1$, $A_{22}=2$, entonces

$$\widetilde{A}=\begin{bmatrix} 9 & -5 \\ 1 & 2\end{bmatrix},\text{ de donde } A^{-1}=\frac{1}{23}\begin{bmatrix} 9 & -5 \\ 1 & 2\end{bmatrix}.$$

3) Con ayuda de la fórmula (4) obtenemos

$$X=A^{-1}B=\frac{1}{23}\begin{bmatrix}9 & -5\\1 & 2\end{bmatrix}\begin{bmatrix}-2 & 3\\1 & -1\end{bmatrix}=\frac{1}{23}\begin{bmatrix}-23 & 22\\0 & 5\end{bmatrix}. \blacktriangle$$

En la práctica se encuentra frecuentamente la ecuación del tipo (1), donde **x** y **b** son los vectores columna del mismo tamaño que la matriz A.

Ejemplo 2. Resolver la ecuación matricial

$$\underbrace{\begin{bmatrix}1 & 1 & 2\\2 & -1 & 2\\4 & 1 & 4\end{bmatrix}}_{A}\underbrace{\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}}_{\mathbf{x}}=\underbrace{\begin{bmatrix}-1\\-4\\-2\end{bmatrix}}_{\mathbf{b}}.$$

$\triangle$ 1) $\det A=\begin{vmatrix}1 & 1 & 2\\2 & -1 & 2\\4 & 1 & 4\end{vmatrix}=-4+8+4+8-2-8=6\neq 0.$

2) Hallamos A^{-1}. Tenemos $A_{11}=-6$, $A_{21}=-2$, $A_{31}=4$; $A_{12}=0$, $A_{22}=-4$; $A_{32}=2$; $A_{13}=6$; $A_{23}=3$, $A_{33}=-3$, o sea,

$$\widetilde{A}=\begin{bmatrix}-6 & -2 & 4\\0 & -4 & 2\\6 & 3 & -3\end{bmatrix};\quad A^{-1}=\begin{bmatrix}-1 & -1/3 & 2/3\\0 & -2/3 & 1/3\\1 & 1/2 & -1/2\end{bmatrix}.$$

3) Con ayuda de la fórmula (4) obtenemos

$$x=A^{-1}\mathbf{b}=\begin{bmatrix}-1 & -1/3 & 2/3\\0 & -2/3 & 1/3\\1 & 1/2 & -1/2\end{bmatrix}\begin{bmatrix}-1\\-4\\-2\end{bmatrix}=\begin{bmatrix}1\\2\\-2\end{bmatrix}. \blacktriangle$$

Para resolver la ecuación (2) multipliquemos a la derecha ambos miembros suyos por A^{-1} (suponiendo que la matriz inversa A^{-1} existe):

$$XAA^{-1}=BA^{-1},$$

Por lo tanto, $XE=BA^{-1}$, de donde

$$X=BA^{-1}. \tag{5}$$

Ejemplo 3. Resolver la ecuación matricial

$$X\underbrace{\begin{bmatrix}1 & 1 & -1\\2 & -1 & 1\\1 & 0 & 1\end{bmatrix}}_{A}=\underbrace{\begin{bmatrix}6 & 2 & -1\\6 & 1 & 1\\8 & -1 & 4\end{bmatrix}}_{B}.$$

△ Hallamos

$$\det A=\begin{vmatrix}1 & 1 & -1\\ 2 & -1 & 1\\ 1 & 0 & 1\end{vmatrix}=-3;\ A_{11}=-1;\ A_{21}=-1;\ A_{31}=0;$$

$$A_{12}=-1;\ A_{22}=2;\ A_{32}=-3;\ A_{13}=1;\ A_{23}=1;\ A_{33}=-3;$$

$$\widetilde{A}=\begin{bmatrix}-1 & -1 & 0\\ -1 & 2 & -3\\ 1 & 1 & -3\end{bmatrix};\quad A^{-1}=\begin{bmatrix}1/3 & 1/3 & 0\\ 1/3 & -2/3 & 1\\ -1/3 & -1/3 & 1\end{bmatrix}.$$

Con ayuda de la fórmula (5) obtenemos

$$X=BA^{-1}=\begin{bmatrix}6 & 2 & -1\\ 6 & 1 & 1\\ 8 & -1 & 4\end{bmatrix}\begin{bmatrix}1/3 & 1/3 & 0\\ 1/3 & -2/3 & 1\\ -1/3 & -1/3 & 1\end{bmatrix}=\begin{bmatrix}3 & 1 & 1\\ 2 & 1 & 2\\ 1 & 2 & 3\end{bmatrix}.\ \blacktriangle$$

Para resolver la ecuación (3) multipliquemos ambos miembros suyos: a la izquierda por A^{-1} y a la derecha por B^{-1} (suponiendo que las matrices inversas indicadas existen); entonces obtenemos

$$A^{-1}AXBB^{-1}=A^{-1}CB^{-1} \text{ o bien } EXE=A^{-1}CB^{-1},$$

de donde $X=A^{-1}CB^{-1}$. (6)

Ejemplo 4. Resolver la ecuación matricial

$$\underbrace{\begin{bmatrix}1 & -3 & 2\\ 3 & -4 & 0\\ 2 & -5 & 3\end{bmatrix}}_{A}X\underbrace{\begin{bmatrix}3 & 1 & 1\\ 2 & 1 & 2\\ 1 & 2 & 3\end{bmatrix}}_{B}=\underbrace{\begin{bmatrix}8 & -4 & -4\\ 18 & 5 & 10\\ 17 & -3 & -1\end{bmatrix}}_{C}$$

△ Hallamos

$$A^{-1}=\begin{bmatrix}-12 & -1 & 8\\ -9 & -1 & 6\\ -7 & -1 & 5\end{bmatrix},\quad B^{-1}=\begin{bmatrix}1/4 & 1/4 & -1/4\\ 1 & -2 & 1\\ -3/4 & 5/4 & -1/4\end{bmatrix}$$

(realice los cálculos de por sí). Luego tenemos

$$A^{-1}C=\begin{bmatrix}-12 & -1 & 8\\ -9 & -1 & 6\\ -7 & -1 & 5\end{bmatrix}\begin{bmatrix}8 & -4 & -4\\ 18 & 5 & 10\\ 17 & -3 & -1\end{bmatrix}=\begin{bmatrix}22 & 19 & 30\\ 12 & 13 & 20\\ 11 & 8 & 13\end{bmatrix}.$$

Ahora con ayuda de la fórmula (6) obtenemos

$$X=A^{-1}CB^{-1}=\begin{bmatrix}22 & 19 & 30\\ 12 & 13 & 20\\ 11 & 8 & 13\end{bmatrix}\begin{bmatrix}1/4 & 1/4 & -1/4\\ 1 & -2 & 1\\ -3/4 & 5/4 & -1/4\end{bmatrix}=\begin{bmatrix}2 & 5 & 6\\ 1 & 2 & 5\\ 1 & 3 & 2\end{bmatrix}.\ \blacktriangle$$

§ 2.6. Matrices triangulares. Desarrollo de la matriz en producto de dos matrices triangulares

La matriz cuadrada se llama *triangluar* si los elementos dispuestos más arriba o más abajo de la diagonal principal son iguales a cero. Si son iguales a cero los elementos dispuestos más arriba de la diagonal principal, la matriz se dice *triangular inferior*; tal es, por ejemplo, la matriz

$$T_1 = \begin{vmatrix} t_{11} & 0 & \dots & 0 \\ t_{21} & t_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ t_{n1} & t_{n2} & \dots & t_{nn} \end{vmatrix}.$$

En cambio si son iguales a cero los elementos dispuestos más abajo de la diagonal principal, la matriz se denomina *triangular superior*; tal es, por ejemplo, la matriz

$$T_2 = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & r_{nn} \end{bmatrix}.$$

El determinante de la matriz trangular es igual al producto de sus elementos diagonales. Si $T = [t_{ij}]$ es una matriz triangular, entonces

$$\det T = t_{11}t_{22} \dots t_{nn}.$$

La matriz inversa de una matriz regular es también la matriz triangular del mismo tamaño y de la misma estructura.

Si una matriz cuadrada

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}$$

tiene *menores diagonales* (así se llaman menores del determinante de la matriz en los cuales sobre las diagonales principales están los elementos diagonales de la matriz) distintos de cero, o sea

$$a_{11} \neq 0, \quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0, \dots, \det A \neq 0,$$

se puede desarrollarla en producto de dos matrices triangulares (superior e inferior). Este desarrollo es único si a los elementos diagonales de una de las matrices triangulares se les asignan de antemano valores distintos del cero (por ejemplo, ponerlos iguales a la unidad).

Sea

$$A = CB,$$

donde C es la matriz triangular inferior y B, la matriz triangular superior con los elementos diagonales iguales a 1.

Tomando a título de ejemplo una matriz de cuarto orden, obtenemos las fórmulas que expresan la dependencia entre los elementos de las matrices A, B y C.:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} = \begin{bmatrix} c_{11} & 0 & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix} \begin{bmatrix} 1 & b_{12} & b_{13} & b_{14} \\ 0 & 1 & b_{23} & b_{24} \\ 0 & 0 & 1 & b_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Realicemos la multiplicación e igualemos los elementos obtenidos de la matriz CB a los elementos correspondientes de la matriz A:

$$c_{11} = a_{11}, \tag{1}$$
$$c_{21} = a_{21}, \tag{2}$$
$$c_{31} = a_{31}, \tag{3}$$
$$c_{41} = a_{41}, \tag{4}$$
$$c_{11}b_{12} = a_{12}, \tag{5}$$
$$c_{21}b_{12} + c_{22} = a_{22}, \tag{6}$$
$$c_{31}b_{12} + c_{32} = a_{32}, \tag{7}$$
$$c_{41}b_{12} + c_{42} = a_{42}, \tag{8}$$
$$c_{11}b_{13} = a_{13}, \tag{9}$$
$$c_{21}b_{13} + b_{22}b_{23} = a_{23} \tag{10}$$
$$c_{31}b_{13} + c_{32}b_{23} + c_{33} = a_{33}, \tag{11}$$
$$c_{41}b_{13} + c_{42}b_{23} + c_{43} = a_{43}, \tag{12}$$
$$c_{11}b_{14} = a_{14}, \tag{13}$$
$$c_{21}b_{14} + c_{22}b_{24} = a_{24}, \tag{14}$$
$$c_{31}b_{14} + c_{32}b_{24} + c_{33}b_{34} = a_{34}, \tag{15}$$
$$c_{41}b_{14} + c_{42}b_{24} + c_{43}b_{34} + c_{44} = a_{44}. \tag{16}$$

De las ecuaciones (1). . .(16) obtenemos los elementos b_{ij} y c_{ij} $(i = 1, 2, 3, 4;\ j = 1, 2, 3, 4)$ de las matrices triangulares B y C en el orden siguiente:

I. La primera columna de la matriz C [fórmulas (1). . .(4)]:

$$c_{i1} = a_{i1}; \quad i = 1, 2, 3, 4.$$

II. La primera fila de la matriz B [fórmulas (5), (9), (13)]:

$$b_{1j} = a_{1j}/c_{11}; \quad j = 2, 3, 4.$$

III. La segunda columna de la matriz C [fórmulas (6), (7), (8)]:

$$c_{i2}=a_{i2}-c_{i1}b_{12};\quad i=2,\ 3,\ 4.$$

IV. La segunda fila de la matriz B [fórmulas (10), (14)]:

$$b_{2j}=(a_{2j}-c_{21}b_{1j})/c_{22};\quad j=3,\ 4.$$

V. La tercera columa de la matriz C [fórmulas (11), (12)]:

$$c_{i3}=a_{i3}-c_{i1}b_{13}-c_{i2}b_{23};\quad i=3,\ 4.$$

VI. La tercera fila de la matriz B [fórmula (15)]:

$$b_{34}=(a_{34}-c_{31}b_{14}-c_{32}b_{24})/c_{33}.$$

VII. La cuarta columna de la matriz C [fórmula (16)]:

$$c_{44}=a_{44}-c_{41}b_{14}-c_{42}b_{24}-c_{43}b_{34}.$$

Sucesión de determinación de los elementos b_{ij} y c_{ij}

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}=\begin{bmatrix} c_{11} & 0 & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}\begin{bmatrix} 1 & b_{12} & b_{13} & b_{14} \\ 0 & 1 & b_{23} & b_{24} \\ 0 & 0 & 1 & b_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix};$$

$$\begin{array}{l|llll}
c_{11}=a_{11} & b_{12}=a_{12}/c_{11}; & b_{13}=a_{13}/c_{11}; & b_{14}=a_{14}/c_{11} & \rightarrow \mathrm{II} \\
c_{21}=a_{21} & c_{22} & b_{23} & b_{24} & \rightarrow \mathrm{IV} \\
c_{31}=a_{31} & c_{32} & c_{33} & b_{34} & \rightarrow \mathrm{VI} \\
c_{41}=a_{41} & c_{42} & c_{43} & c_{44} & \\
\downarrow & \downarrow & \downarrow & \downarrow & \\
\mathrm{I} & \mathrm{III} & \mathrm{V} & \mathrm{VII} &
\end{array}$$

Las cifras romanas de las flechas indican la sucesión en la cual han de determinarse los elementos c_{ij} y b_{ij}. Con tal desarrollo primero se determinan las columnas y luego, las filas.

Para que los cálculos sean cómodos el desarrollo de la matriz A en producto de dos matrices triangulares C y B es: conveniente realizarlo utilizando la tabla 2.1.

Esta tabla se llena del modo siguiente:

1. Basándose en las fórmulas anteriormente indicadas, en la columna 1 de la matriz C se escriben los elementos de la columna 1 de la matriz A y en la fila 1 de la matriz B, los elementos dela fila 1 de la matriz A divididos por c_{11}.

2. Los elementos situados debajo de la línea escalonada se determinan así: se toma el elemento correspondiente de la matriz A y de éste se sustrayen los productos de los elementos que están en la misma fila, a la izquierda, y en la misma columna, más arriba, que el elemento a calcular; en este caso el primer elemento de la fila se

Tabla 2.1

a_{11}		a_{12}		a_{13}		a_{14}	
a_{21}		a_{22}		a_{23}		a_{24}	
a_{31}		a_{32}		a_{33}		a_{34}	
a_{41}		a_{42}		a_{43}		a_{44}	
$c_{11}=a_{11}$	1	$b_{12}=a_{12}/c_{11}$		$b_{13}=a_{13}/c_{11}$		$b_{14}=a_{14}/c_{11}$	
$c_{21}=a_{21}$		c_{22}	1	b_{23}		b_{24}	
$c_{31}=a_{31}$		c_{32}		c_{33}	1	b_{34}	
$c_{41}=a_{41}$		c_{42}		c_{43}		c_{44}	1

multiplica por el primer elemento de la columna, el segundo elemento de la fila se multiplica por el segundo elemento de la columna, etc.

Por ejemplo, $c_{33} = a_{33} - c_{31}b_{13} - c_{32}b_{23}$.

3. En cambio, al calcular los elementos situados por encima de la línea escalonada se procede de la misma manera que en el subp. 2, pero el resultado obtenido se divide por el elemento diagonal c_{ii} $(i = 2, 3)$ que está en la misma fila que el elemento a determinar. Por ejemplo, $b_{34} = \dfrac{a_{34} - c_{31}b_{14} - b_{24}}{c_{33}}$.

Análogamente, se puede desarrollar en producto de dos matrices triangulares una matriz cuadrada de todo orden n. Anteriormente hemos indicado la regla de transformación de la matriz en producto de dos matrices triangulares para el caso de $b_{ii} = 1$. Si $c_{ii} = 1$, en el primer lugar es necesario calcular los elementos de las filas de la matriz B con ayuda de las fórmulas

$$b_{ij} = a_{ij} - \sum_{k=1}^{i-1} c_{ik}c_{kj} \, (i \leqslant j) \tag{17}$$

y luego los elementos de las columnas de la matriz C con ayuda de las fórmulas

$$c_{ij} = \frac{a_{ij} - \sum\limits_{k=1}^{i-1} c_{ik}b_{kj}}{b_{ii}} \, (i > j). \tag{18}$$

La matriz A fue representada en forma del producto CB de dos matrices triangulares, donde C es la matriz triangular inferior y B, la superior. Sin embargo, tal orden de factores no es obligatorio, es decir, se puede representar la matriz A en forma del producto BC y obtener las fórmulas análogas para los elementos de las matrices triangulares B y C.

Ejemplo. Desarrollar la matriz

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

en producto CB, donde

$$C=\begin{bmatrix} c_{11} & 0 & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix}, \quad B=\begin{bmatrix} 1 & b_{12} & b_{13} & b_{14} \\ 0 & 1 & b_{23} & b_{24} \\ 0 & 0 & 1 & b_{24} \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

△ La solución se da en la tabla 2.2.

Verificación:

$$CB=\begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 4 & 0 & 0 \\ 2 & 0 & -1 & 0 \\ 4 & -5 & -5/4 & -5/4 \end{bmatrix}\begin{bmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 7/4 & 1/4 \\ 0 & 0 & 1 & 10 \\ 0 & 0 & 0 & 1 \end{bmatrix}=$$

$$=\begin{bmatrix} 1 & 2 & 3 & 4 \\ -1 & 2 & 4 & -3 \\ 2 & 2 & 5 & -2 \\ 4 & 3 & 2 & 1 \end{bmatrix}=A. \quad ▲$$

Tabla 2.2

1	2	3	4	a_{ij}
-1	2	4	-3	
2	4	5	-2	
4	3	2	1	
1	$1 \quad \frac{2}{1}=2$	$\frac{3}{1}=3$	$\frac{4}{1}=4$	b_{ij}
-1	$2-(-1)\cdot 2=4 \quad 1$	$\frac{4-(-1)\cdot 3}{4}=\frac{7}{4}$	$\frac{-3-(-1)\cdot 4}{4}=\frac{1}{4}$	
2	$4-2\cdot 2=0$	$5-2\cdot 3-0\cdot\frac{7}{4}=-1 \quad 1$	$\frac{-2-2\cdot 4-0\cdot\frac{1}{4}}{-1}=10$	
4	$3-4\cdot 2=-5$	$2-4\cdot 3-(-5)\cdot\frac{7}{4}=-\frac{5}{4}$	$1-4\cdot 4-(-5)\times\frac{1}{4}-\left(-\frac{5}{4}\right)\times 10=-\frac{5}{4} \quad 1$	

§ 2.7. Inversión de la matriz con ayuda de su desarrollo en producto de dos matrices triangulares

De la definición de la matriz inversa se deduce que si

$$A = CB, \tag{1}$$

donde todas las matrices son regulares, la matriz inversa puede ser hallada con ayuda de la fórmula

$$A^{-1} = B^{-1}C^{-1}. \tag{2}$$

□ Para demostrar la validez de la fórmula (2) realicemos las transformaciones siguientes.

Multipliquemos ambos miembros de la igualdad (1) a la izquierda por la matriz C^{-1};

$$C^{-1}A = C^{-1}CB \text{ o bien } C^{-1}A = B. \tag{3}$$

Multipliquemos ambos miembros de la igualdad (3) a la izquierda por la matriz B^{-1}:

$$B^{-1}C^{-1}A = B^{-1}B \text{ o bien } (B^{-1}C^{-1})\,A = E. \tag{4}$$

Multipliquemos ambos miembros de la igualdad (4) a la derecha por la matriz A^{-1}:

$$(B^{-1}C^{-1})\,AA^{-1} = A^{-1} \text{ o bien } A^{-1} = B^{-1}C^{-1}. \ \blacksquare$$

En la fórmula (2) la matriz inversa A^{-1} se expresa como producto de las matrices inversas B^{-1} y C^{-1}. No obstante, si las matrices B y C son triangulares, para calcular la matriz A^{-1} no hay necesidad de invertir las matrices B y C.

Vamos a deducir las fórmulas para calcular los elementos de la matriz inversa A^{-1} citando como ejemplo una matriz de cuarto orden. Designemos los elementos de la matriz A^{-1} con α_{ij} y los de las matrices inversas B^{-1} y C^{-1} con β_{ij} y γ_{ij}, respectivamente. Las matrices B y C son del mismo tipo que en el § 2.6. Multipliquemos la igualdad (2) a la derecha por la matriz C y a la izquierda, por la matriz B:

$$A^{-1}C = B^{-1}C^{-1}C \text{ o bien } A^{-1}C = B^{-1}; \tag{5}$$

$$BA^{-1} = BB^{-1}C^{-1} \text{ o bien } BA^{-1} = C^{-1}. \tag{6}$$

Las matrices C^{-1} y B^{-1} son las matrices triangulares del mismo tipo que las matrices C y B. Escribamos las igualdades (5) y (6) del modo siguiente:

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \beta_{43} & \alpha_{44} \end{bmatrix} \begin{bmatrix} c_{11} & 0 & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{bmatrix} = \begin{bmatrix} 1 & \beta_{12} & \beta_{13} & \beta_{14} \\ 0 & 1 & \beta_{23} & \beta_{24} \\ 0 & 0 & 1 & \beta_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{7}$$

$$\begin{bmatrix} 1 & b_{12} & b_{13} & b_{14} \\ 0 & 1 & b_{23} & b_{24} \\ 0 & 0 & 1 & b_{34} \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} & \alpha_{14} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\ \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \end{bmatrix} = \begin{bmatrix} \gamma_{11} & 0 & 0 & 0 \\ \gamma_{21} & \gamma_{22} & 0 & 0 \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & 0 \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} \end{bmatrix}. \tag{8}$$

Realizaremos la multiplicación sólo de aquellas filas y columnas donde los elementos de los productos en las matrices B^{-1} y C^{-1} son iguales a los ceros y las unidades, respectivamente.

Vamos a considerar primero el producto $A^{-1}C = B^{-1}$:

$$\beta_{11} = \alpha_{11}c_{11} + \alpha_{12}c_{21} + \alpha_{13}c_{31} + \alpha_{14}c_{41} = 1, \tag{9}$$

$$\beta_{21} = \alpha_{21}c_{11} + \alpha_{22}c_{21} + \alpha_{23}c_{31} + \alpha_{24}c_{41} = 0, \tag{10}$$

$$\beta_{31} = \alpha_{31}c_{11} + \alpha_{32}c_{21} + \alpha_{33}c_{31} + \alpha_{34}c_{41} = 0, \tag{11}$$

$$\beta_{41} = \alpha_{41}c_{11} + \alpha_{42}c_{21} + \alpha_{43}c_{31} + \alpha_{44}c_{41} = 0, \tag{11}$$

$$\beta_{22} = \alpha_{21}\cdot 0 + \alpha_{22}c_{22} + \alpha_{23}c_{32} + \alpha_{24}c_{42} = 1, \tag{13}$$

$$\beta_{32} = \alpha_{31}\cdot 0 + \alpha_{32}c_{22} + \alpha_{33}c_{22} + \alpha_{34}c_{42} = 0, \tag{14}$$

$$\beta_{42} = \alpha_{41}\cdot 0 + \alpha_{42}c_{22} + \alpha_{43}c_{32} + \alpha_{44}c_{42} = 0, \tag{15}$$

$$\beta_{33} = \alpha_{31}\cdot 0 + \alpha_{32}\cdot 0 + \alpha_{33}c_{33} + \alpha_{34}c_{43} = 1, \tag{16}$$

$$\beta_{43} = \alpha_{41}\cdot 0 + \alpha_{42}\cdot 0 + \alpha_{43}c_{33} + \alpha_{44}c_{43} = 0, \tag{17}$$

$$\beta_{44} = \alpha_{41}\cdot 0 + \alpha_{42}\cdot 0 + \alpha_{43}\cdot 0 + \alpha_{44}c_{44} = 1. \tag{18}$$

Consideremos ahora el producto $BA^{-1} = C^{-1}$:

$$\gamma_{12} = 1\cdot\alpha_{12} + b_{12}\alpha_{22} + b_{13}\alpha_{32} + b_{14}\alpha_{42} = 0, \tag{19}$$

$$\gamma_{13} = 1\cdot\alpha_{13} + c_{12}\alpha_{23} + b_{13}\alpha_{33} + b_{14}\alpha_{43} = 0, \tag{20}$$

$$\gamma_{14} = 1\cdot\alpha_{14} + b_{12}\alpha_{24} + b_{13}\alpha_{34} + b_{14}\alpha_{44} = 0, \tag{21}$$

$$\gamma_{23} = 0\cdot\alpha_{13} + 1\cdot\alpha_{23} + b_{23}\alpha_{33} + b_{24}\alpha_{43} = 0, \tag{22}$$

$$\gamma_{24} = 0\cdot\alpha_{14} + 1\cdot\alpha_{24} + b_{23}\alpha_{34} + b_{24}\alpha_{44} = 0, \tag{23}$$

$$\gamma_{34} = 0\cdot\alpha_{14} + 0\cdot\alpha_{24} + 1\cdot\alpha_{34} + b_{34}\alpha_{44} = 0. \tag{24}$$

Es necesario determinar 16 elementos desconocidos α_{ij} de las igualdades (9). . .(24). Tenemos:

I. $\alpha_{44} = 1/c_{44}$ [de la (18)];

$\alpha_{43} = (1/c_{33})(-\alpha_{44}c_{43})$ [de la (17)];

$\alpha_{42} = (1/c_{22})(-\alpha_{43}c_{32} - \alpha_{44}c_{42})$ [de la (15)];

$\alpha_{41} = (1/c_{11})(-\alpha_{42}c_{21} - \alpha_{43}c_{31} - \alpha_{44}c_{41})$ [de la (12)].

II. $\alpha_{34} = -b_{34}\alpha_{44}$ [de la (24)];

$\alpha_{24} = -b_{23}\alpha_{34} - b_{24}\alpha_{44}$ [de la (23)];

$\alpha_{14} = -b_{12}\alpha_{24} - b_{13}\alpha_{34} - b_{14}\alpha_{44}$ [de la (21)].

III. $\alpha_{33} = (1/c_{33})(1 - \alpha_{34}c_{43})$ [de la (16)];

$$\alpha_{32}=(1/c_{22})\,(-\alpha_{33}c_{32}-\alpha_{34}c_{42}) \qquad \text{[de la (14)];}$$

$$\alpha_{31}=(1/c_{11})\,(-\alpha_{32}c_{21}-\alpha_{33}c_{31}-\alpha_{34}c_{41}) \qquad \text{[de la (11)].}$$

IV. $$\alpha_{23}=-b_{23}\alpha_{33}-b_{24}\alpha_{43} \qquad \text{[de la (22)];}$$

$$\alpha_{13}=-b_{12}\alpha_{23}-b_{13}\alpha_{33}-b_{14}\alpha_{43} \qquad \text{[de la (20)].}$$

V. $$\alpha_{22}=(1/c_{22})\,(1-\alpha_{23}c_{32}-\alpha_{24}c_{42}) \qquad \text{[de la (13)];}$$

$$\alpha_{21}=(1/c_{11})\,(-\alpha_{22}c_{21}-\alpha_{23}c_{31}-\alpha_{24}c_{41}) \qquad \text{[de la (10)];}$$

VI. $$\alpha_{12}=-b_{12}\alpha_{22}-b_{13}\alpha_{32}-b_{14}\alpha_{42} \qquad \text{[de la (19)];}$$

VII. $$\alpha_{11}=(1/c_{11})\,(1-\alpha_{12}c_{21}-\alpha_{13}c_{31}-\alpha_{14}c_{41}) \qquad \text{[de la (9)].}$$

Utilizando las igualdades obtenidas en las etapas I. . .VII, llegamos al esquema siguiente.

Esquema de sucesión de cálculos de los elementos de la matriz inversa

$$\begin{array}{lcccc}
 & & \mathrm{VI} & \mathrm{IV} & \mathrm{II} \\
 & \alpha_{11} & \alpha_{12}\uparrow & \alpha_{13}\uparrow & \alpha_{14}\uparrow \\
\mathrm{VII}\longleftarrow & & & & \\
 & \alpha_{21} & \alpha_{22} & \alpha_{23} & \alpha_{24} \\
\mathrm{V}\longleftarrow & & & & \\
 & \alpha_{31} & \alpha_{32} & \alpha_{33} & \alpha_{34} \\
\mathrm{III}\longleftarrow & & & & \\
 & \alpha_{41} & \alpha_{42} & \alpha_{43} & \alpha_{44} \\
\mathrm{I}\longleftarrow & & & &
\end{array}$$

Con ayuda de este esquema se puede calcular la matriz inversa A^{-1} para la matriz A de todo orden si está desarrollada en producto CB de dos matrices triangulares, donde C y B son las matrices triangulares del tipo anteriormente indicado.

Con tal desarrollo se obtienen las siguientes fórmulas para los elementos de la matriz A^{-1}:

$$\left.\begin{aligned}
\alpha_{ij}&=-\sum_{k=i+1}^{n} b_{ik}\alpha_{kj}\ (i<j);\\
\alpha_{ij}&=\frac{1}{c_{ij}}\left(1-\sum_{k=i+1}^{n}\alpha_{ik}c_{ki}\ (i=j),\right.\\
\alpha_{ij}&=-\frac{\sum\limits_{k=j+1}^{n}\alpha_{ik}c_{kj}}{c_{jj}}\ (i>j).
\end{aligned}\right| \qquad (25)$$

Ejemplo. Con ayuda del desarrollo de la matriz

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

	3	-2	2	0	
	2	1	1	-2	A
	3	-1	2	1	
	1	2	-1	-1	
	3	1 | $-2/3$	$2/3$	0	
	2	$1-2\left(-\frac{2}{3}\right)=\frac{7}{3}$	1 | $\frac{1-2\cdot(2/3)}{7/3}=-\frac{1}{7}$	$\frac{-2-2\cdot 0}{7/3}=-\frac{6}{7}$	
C	3	$-1-3\cdot\left(-\frac{2}{3}\right)=1$	$2-3\cdot\frac{2}{3}-1\times$ $\times\left(-\frac{1}{7}\right)=\frac{1}{7}$	1 | $\frac{1-3\cdot 0-1\cdot(-6/7)}{1/7}=13$	B
	1	$2-1\cdot\left(-\frac{2}{3}\right)=\frac{8}{3}$	$-1-1\cdot\frac{2}{3}-\frac{8}{3}\times$ $\times\left(-\frac{1}{7}\right)=-\frac{9}{7}$	$-1-1\cdot 0-\frac{8}{3}\times$ $\times\left(-\frac{6}{7}\right)-\left(-\frac{8}{7}\right)\times$ $\times 13=18$ | 1	
	$1/3$	$-4/18$	0	$8/18$	
	$-12/18$	$5/18$	$1/2$	$-1/18$	A^{-1}
	$-12/18$	$11/18$	$1/2$	$-13/18$	
	$-1/3$	$-5/18$	$1/2$	$1/18$	

△ Componemos la tabla 2.3.

Realizamos el cálculo de los elementos de la matriz inversa en el orden anteriormente indicado (véase el esquema), utilizando las fórmulas (25).

I. $\alpha_{44}=1/c_{44}=1/18;$

$$\alpha_{43}=-\frac{\alpha_{44}c_{43}}{c_{33}}=-\frac{(1/18)\cdot(-9/7)}{1/7}=\frac{1}{2};$$

$$\alpha_{42}=-\frac{\alpha_{43}c_{32}+\alpha_{44}c_{42}}{c_{22}}=-\frac{(1/2)\cdot 1+(1/18)\cdot(8/3)}{7/3}=-\frac{5}{18};$$

$$\alpha_{41}=-\frac{\alpha_{42}c_{21}+\alpha_{43}c_{31}+\alpha_{44}c_{41}}{c_{11}}=$$

$$=-\frac{(-5/18)\cdot 2+(1/2)\cdot 3+(1/18)\cdot 1}{3}=-\frac{1}{3}.$$

II. $\alpha_{34}=-b_{34}\alpha_{44}=-13/18;$

$$\alpha_{24}=-(b_{23}\alpha_{34}+b_{24}\alpha_{44})=$$

$$=-\left[\left(-\frac{1}{7}\right)\cdot\left(-\frac{13}{18}\right)+\left(-\frac{6}{7}\right)\cdot\frac{1}{18}\right]=-\frac{1}{18};$$

$$\alpha_{14}=-(b_{12}\alpha_{24}+b_{13}\alpha_{34}+b_{14}\alpha_{44})=$$

$$=-\left[\left(-\frac{2}{3}\right)\left(-\frac{1}{18}\right)+\frac{2}{3}\cdot\left(-\frac{13}{18}\right)+0\cdot\frac{1}{18}\right]=\frac{8}{18};$$

III. $\alpha_{33}=\frac{1}{c_{33}}\,(1-\alpha_{34}c_{43})=\frac{1}{1/7}\left[1-\left(-\frac{13}{18}\right)\cdot\left(-\frac{9}{7}\right)\right]=\frac{1}{2};$

$$\alpha_{32}=-\frac{\alpha_{33}c_{32}+{}_{34}c_{42}}{c_{22}}=-\frac{(1/2)\cdot 1+(-13/18)\cdot(8/3)}{7/3}=\frac{11}{18};$$

$$\alpha_{31}=-\frac{\alpha_{32}c_{21}+\alpha_{33}c_{31}+\alpha_{34}c_{41}}{c_{11}}=$$

$$=-\frac{(11/18)\cdot 2+(1/2)\cdot 3+(-13/18)\cdot 1}{3}=-\frac{12}{18}.$$

IV. $\alpha_{23}=-(b_{23}\alpha_{33}+b_{24}\alpha_{43})=$

$$=-\left[\left(-\frac{1}{7}\right)\cdot\frac{1}{2}+\left(-\frac{6}{7}\right)\cdot\frac{1}{2}\right]=\frac{1}{2};$$

$$\alpha_{13}=-(b_{12}\alpha_{23}+b_{13}\alpha_{33}+b_{14}\alpha_{43})=$$

$$=-\left[\left(-\frac{2}{3}\right)\cdot\frac{1}{2}+\frac{2}{3}\cdot\frac{1}{2}+0\cdot\frac{1}{2}\right]=0.$$

V. $\alpha_{22}=\frac{1}{c_{22}}[1-(\alpha_{23}c_{32}+\alpha_{24}c_{42})]=$

$$=\frac{1-\left[(1/2)\cdot 1+\left(-\frac{1}{18}\right)\cdot(8/3)\right]}{7/3}=\frac{5}{18};$$

$$\alpha_{21}=-\frac{\alpha_{22}c_{21}+\alpha_{23}c_{31}+\alpha_{24}c_{41}}{c_{11}}=$$

$$=-\frac{(5/18)\cdot 2+(1/2)\cdot 3+(-1/18)\cdot 1}{3}=-\frac{12}{18}.$$

VI. $\alpha_{12}=-(b_{12}\alpha_{22}+b_{13}\alpha_{32}+b_{14}\alpha_{42})=$

$$=-\left[\left(-\frac{2}{3}\right)\cdot\frac{5}{18}+\frac{2}{3}\cdot\frac{11}{18}+0\cdot\left(-\frac{5}{18}\right)\right]=-\frac{4}{18}.$$

VII. $\alpha_{11}=\frac{1}{c_{11}}[1-(\alpha_{12}c_{21}+\alpha_{13}c_{31}+\alpha_{14}c_{41})]=$

$$=\frac{1}{3}\left[1+\left(-\frac{4}{18}\right)\cdot 2+0\cdot 3+\frac{8}{18}\cdot 1\right]=\frac{1}{3}.$$

Respuesta: $A^{-1}=\frac{1}{18}\begin{bmatrix} 6 & -4 & 0 & 8 \\ -12 & 5 & 9 & -1 \\ -12 & 11 & 9 & -13 \\ -6 & -5 & 9 & 1 \end{bmatrix}$. ▲

§ 2.8. Matrices celulares y operaciones con ellas

Al ejecutar los cálculos con matrices de altos órdenes es conveniente partirlas previamente en células (bloques) con ayuda de los tabiques horizontales y verticales que van a lo largo y a través de toda la matriz. Ahora bien, cada matriz se divide en matrices de órdenes menores sobre las cuales las operaciones de cálculo se realizan mucho más sencillamente.

La matriz partida en células se llama *celular* (o *en bloques*).

Por ejemplo: 1) la matriz A está dividida en cuatro células:

$$A=\left[\begin{array}{cc|cc} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right]=\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right],$$

donde de células sirven las matrices cuadradas

$$A_{11}=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \quad A_{12}=\begin{bmatrix} a_{13} & a_{14} \\ a_{23} & a_{24} \end{bmatrix}, \quad A_{21}=\begin{bmatrix} a_{31} & a_{32} \\ a_{41} & a_{42} \end{bmatrix},$$

$$A_{22}=\begin{bmatrix} a_{33} & a_{34} \\ a_{43} & a_{44} \end{bmatrix};$$

2) La matriz de n-ésimo orden C está dividida en cuatro células;

$$C=\left[\begin{array}{c|c} C_{n-1} & \mathbf{u}_{n-1} \\ \hline \mathbf{v}_{n-1} & c_{nn} \end{array}\right],$$

donde C_{n-1} es la matriz cuadrada de orden $(n-1)$; $\mathbf{u}_{n-1}$, el vector columna de orden $(n-1)$; $\mathbf{v}_{n-1}$, el vector fila de orden $(n-1)$; c_{nn}, el número.

Tal partición se denomina *orladura* y las matrices, *orladas*.

Con matrices celulares pueden ejecutarse las operaciones de adición y de multiplicación, considerando las células de la matriz celular como elementos de una matriz usual

Sean

$$A=\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right] \text{ y } B=\left[\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right]$$

las matrices celulares del mismo orden y de la misma partición. Entonces

$$A+B=\left[\begin{array}{c|c} A_{11}+B_{11} & A_{12}+B_{12} \\ \hline A_{21}+B_{21} & A_{22}+B_{22} \end{array}\right],$$

$$AB=\left[\begin{array}{c|c} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}+A_{12}B_{22} \\ \hline A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22} \end{array}\right].$$

Ejemplo 1. Las matrices

$$A=\begin{bmatrix} 5 & 7 & -3 & -4 \\ 7 & 6 & -4 & -5 \\ 6 & 4 & -3 & -2 \\ 8 & 5 & -6 & -1 \end{bmatrix} \quad \text{y} \quad B=\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{bmatrix}$$

han de partirse en células cuadradas y calcular $A+B$ y AB.

△ 1) Vamos a partir las matrices A y B en células cuadradas del modo siguiente;

$$A=\left[\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right]=\left[\begin{array}{cc|cc} 5 & 7 & -3 & -4 \\ 7 & 6 & -4 & -5 \\ \hline 6 & 4 & -3 & -2 \\ 8 & 5 & -6 & -1 \end{array}\right],$$

$$B=\left[\begin{array}{c|c} B_{11} & B_{12} \\ \hline B_{21} & B_{22} \end{array}\right]=\left[\begin{array}{cc|cc} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ \hline 1 & 3 & 5 & 7 \\ 2 & 4 & 6 & 8 \end{array}\right].$$

Encontramos

$$A+B=\left[\begin{array}{c|c} A_{11}+B_{11} & A_{12}+B_{12} \\ \hline A_{21}+B_{21} & A_{22}+B_{22} \end{array}\right]=\left[\begin{array}{cc|cc} 6 & 9 & 0 & 0 \\ 9 & 9 & 0 & 0 \\ \hline 7 & 7 & 2 & 5 \\ 10 & 9 & 0 & 7 \end{array}\right].$$

3) Tenemos

$$AB=\left[\begin{array}{c|c} A_{11}B_{11}+A_{12}B_{21} & A_{11}B_{12}+A_{12}B_{22} \\ \hline A_{21}B_{11}+A_{22}B_{21} & A_{21}B_{12}+A_{22}B_{22} \end{array}\right].$$

Sucesivamente obtenemos

$$A_{11}B_{11}=\begin{bmatrix} 5 & 7 \\ 7 & 6 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}=\begin{bmatrix} 19 & 31 \\ 19 & 32 \end{bmatrix}.$$

$$A_{12}B_{21}=\begin{bmatrix} -3 & -4 \\ -4 & -5 \end{bmatrix}\begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix}=\begin{bmatrix} -11 & -25 \\ -14 & -32 \end{bmatrix},$$

$$A_{11}B_{11}+A_{12}B_{21}=\begin{bmatrix}8 & 6\\ 5 & 0\end{bmatrix};$$

$$A_{21}B_{11}=\begin{bmatrix}6 & 4\\ 8 & 5\end{bmatrix}\begin{bmatrix}1 & 2\\ 2 & 3\end{bmatrix}=\begin{bmatrix}14 & 24\\ 18 & 31\end{bmatrix},$$

$$A_{22}B_{21}=\begin{bmatrix}-3 & -2\\ -6 & -1\end{bmatrix}\begin{bmatrix}1 & 3\\ 2 & 4\end{bmatrix}=\begin{bmatrix}-7 & -17\\ -8 & -22\end{bmatrix},$$

$$A_{21}B_{11}+A_{22}B_{21}=\begin{bmatrix}7 & 7\\ 10 & 9\end{bmatrix};$$

$$A_{11}B_{12}=\begin{bmatrix}5 & 7\\ 7 & 6\end{bmatrix}\begin{bmatrix}3 & 4\\ 4 & 5\end{bmatrix}=\begin{bmatrix}43 & 55\\ 45 & 58\end{bmatrix},$$

$$A_{12}B_{22}=\begin{bmatrix}-3 & -4\\ -4 & -5\end{bmatrix}\begin{bmatrix}5 & 7\\ 6 & 8\end{bmatrix}=\begin{bmatrix}-39 & -53\\ -50 & -68\end{bmatrix},$$

$$A_{11}B_{12}+A_{12}B_{22}=\begin{bmatrix}4 & 2\\ -5 & -10\end{bmatrix};$$

$$A_{21}B_{12}=\begin{bmatrix}6 & 4\\ 8 & 5\end{bmatrix}\begin{bmatrix}3 & 4\\ 4 & 5\end{bmatrix}=\begin{bmatrix}34 & 44\\ 44 & 57\end{bmatrix},\ A_{22}B_{22}=\begin{bmatrix}-3 & -2\\ -6 & -1\end{bmatrix}\begin{bmatrix}5 & 7\\ 8 & 8\end{bmatrix}=$$

$$=\begin{bmatrix}-27 & -27\\ -36 & -50\end{bmatrix},\ A_{21}B_{12}+A_{22}B_{22}=\begin{bmatrix}7 & 7\\ 8 & 7\end{bmatrix}.$$

Ahora bien,

$$AB=\left[\begin{array}{cc|cc}8 & 6 & 4 & 2\\ 5 & 0 & -5 & -10\\ \hline 7 & 7 & 7 & 7\\ 10 & 9 & 8 & 7\end{array}\right].\ \blacktriangle$$

Sean

$$A=\left[\begin{array}{c|c}A_{n-1} & \mathbf{u}_{n-1}\\ \hline \mathbf{v}_{n-1} & a_{nn}\end{array}\right]\ \text{ y }\ B=\left[\begin{array}{c|c}B_{n-1} & \mathbf{y}_{n-1}\\ \hline \mathbf{x}_{n-1} & b_{nn}\end{array}\right]$$

las matrices orladas celulares de n-ésimo orden y de la misma partición. Entonces

$$A+B=\left[\begin{array}{c|c}A_{n-1}+B_{n-1} & \mathbf{u}_{n-1}+\mathbf{y}_{n-1}\\ \hline \mathbf{v}_{n-1}+\mathbf{x}_{n-1} & a_{nn}+b_{nn}\end{array}\right],$$

donde $A_{n-1}+B_{n-1}$ es la matriz cuadrada de orden $n-1$; $\mathbf{u}_{n-1}+\mathbf{y}_{n-1}$, el vector columna de orden $n-1$; $\mathbf{v}_{n-1}+\mathbf{x}_{n-1}$, el vector fila de orden $n-1$; $a_{nn}+b_{nn}$, el número;

$$AB=\left[\begin{array}{c|c}A_{n-1}B_{n-1}+\mathbf{u}_{n-1}\mathbf{x}_{n-1} & A_{n-1}\mathbf{y}_{n-1}+\mathbf{u}_{n-1}b_{nn}\\ \hline \mathbf{v}_{n-1}B_{n-1}+a_{nn}\mathbf{x}_{n-1} & \mathbf{v}_{n-1}\mathbf{y}_{n-1}+a_{nn}b_{nn}\end{array}\right],$$

donde $A_{n-1}B_{n-1}+\mathbf{u}_{n-1}\mathbf{x}_{n-1}$ es la matriz cuadrada de orden $n-1$; $A_{n-1}\mathbf{y}_{n-1}+\mathbf{u}_{n-1}b_{nn}$, el vector columna de orden $n-1$; $\mathbf{v}_{n-1}B_{n-1}+a_{nn}\mathbf{x}_{n-1}$, el vector fila de orden $n-1$; $\mathbf{v}_{n-1}\mathbf{y}_{n-1}+a_{nn}\ b_{nn}$, el número.

Ejemplo 2. Las matrices

$$A=\begin{bmatrix}5&8&-4\\6&9&-5\\4&7&-3\end{bmatrix}\ \text{y}\ B=\begin{bmatrix}3&2&5\\4&-1&3\\9&6&5\end{bmatrix}$$

han de partirse en células por orladura y se debe calcular $A+B$ y AB.

△ 1) Vamos a partir las matrices A y B con ayuda de la orladura:

$$A=\left[\begin{array}{cc|c}5&8&-4\\6&9&-5\\\hline 4&7&-3\end{array}\right]=\left[\begin{array}{c|c}A_2&\mathbf{u}_2\\\hline \mathbf{v}_2&a_{33}\end{array}\right],\quad B=\left[\begin{array}{cc|c}3&2&5\\4&-1&3\\\hline 9&6&5\end{array}\right]=\left[\begin{array}{c|c}B_2&\mathbf{y}_2\\\hline \mathbf{x}_2&b_{23}\end{array}\right]$$

2)Hallamos

$$A+B=\begin{bmatrix}A_2+B_2&\mathbf{u}_2+\mathbf{y}_2\\\mathbf{v}_2+\mathbf{x}_2&a_{33}+b_{33}\end{bmatrix}=\left[\begin{array}{cc|c}8&10&1\\10&8&-2\\\hline 13&13&2\end{array}\right]$$

3) Tenemos

$$AB=\begin{bmatrix}A_2B_2+\mathbf{u}_2\mathbf{x}_2&A_2\mathbf{y}_2+\mathbf{u}_2b_{33}\\\mathbf{v}_2B_2+a_{33}\mathbf{x}_2&\mathbf{v}_2\mathbf{y}_2+a_{33}b_{33}\end{bmatrix}.$$

Sucesivamente obtenemos

$$A_2B_2=\begin{bmatrix}5&8\\6&9\end{bmatrix}\begin{bmatrix}3&2\\4&-1\end{bmatrix}=\begin{bmatrix}47&2\\54&3\end{bmatrix},$$

$$\mathbf{u}_2\mathbf{x}_2=\begin{bmatrix}-4\\-5\end{bmatrix}[9\ \ 6]=\begin{bmatrix}-36&-24\\-45&-30\end{bmatrix},$$

$$A_2B_2+\mathbf{u}_2\mathbf{x}_2=\begin{bmatrix}11&-22\\9&-27\end{bmatrix};$$

$$A_2\mathbf{y}_2=\begin{bmatrix}5&8\\6&9\end{bmatrix}\begin{bmatrix}5\\3\end{bmatrix}=\begin{bmatrix}49\\57\end{bmatrix},$$

$$\mathbf{u}_2b_{33}=\begin{bmatrix}-4\\-5\end{bmatrix}5=\begin{bmatrix}-20\\-25\end{bmatrix},\quad A_2\mathbf{y}_2+\mathbf{u}_2b_{33}=\begin{bmatrix}29\\32\end{bmatrix};$$

$$\mathbf{v}_2B_2=[4\ \ 7]\begin{bmatrix}3&2\\4&-1\end{bmatrix}=[40\ \ 1],\quad a_{33}\mathbf{x}_2=(-3)\cdot[\,9\ \ 6\,]=[\,-27-18\,],$$

$$\mathbf{v}_2B_2+a_{33}\mathbf{x}_2=[13-17];$$

$$\mathbf{v}_2\mathbf{y}_2 = |4\ \ 7| \begin{bmatrix} 5 \\ 3 \end{bmatrix} = 41, \quad a_{33}b_{33} = (-3)\cdot 5 = -15, \quad \mathbf{v}_2\mathbf{y}_2 + a_{33}b_{33} = 26.$$

Ahora bien,

$$\left[\begin{array}{cc|c} 11 & -33 & 29 \\ 9 & -27 & 32 \\ \hline 13 & -17 & 26 \end{array}\right]. \ \blacktriangle$$

§ 2.9. Inversión de las matrices con ayuda de la partición en células

Para hallar la matriz inversa se puede utilizar el **método de partición en células.** Sea $A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$ la matriz regular celular de n-ésimo orden en la cual A_{11} y A_{22} son las células cuadradas de órdenes p y q (donde $p+q=n$). Se necesita hallar la matriz inversa $A^{-1} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix}$ en la cual B_{11} y B_{22} son las matrices cuadradas también de órdenes p y q.

Según la definición de la matriz inversa, $AA^{-1} = E_n$. En el caso dado la matriz unidad también se partirá en células del modo análogo, o sea, $E_n = \begin{bmatrix} E_p & 0 \\ 0 & E_q \end{bmatrix}$, donde E_p y E_q son las matrices unidad de los órdenes p y q, respectivamente. Entonces

$$A^{-1}A = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} = \begin{bmatrix} E_p & 0 \\ 0 & E_q \end{bmatrix},$$

de donde después de la multiplicación obtenemos cuatro ecuaciones matriciales:

$$\begin{cases} B_{11}A_{11} + B_{12}A_{21} = E_p. \\ B_{11}A_{12} + B_{12}A_{22} = 0. \\ B_{21}A_{11} + B_{22}A_{21} = 0. \\ B_{21}A_{12} + B_{22}A_{22} = E_q. \end{cases} \tag{1}$$

Para hallar las células de la matriz A^{-1} es necesario resolver el sistema de ecuaciones matriciales (1). Con este fin utilicemos el método de eliminación de las incógnitas. Multipliquemos a la derecha la primera ecuación del sistema (1) por $A_{11}^{-1}A_{12}$ y sustrayamos del resultado de la multiplicación la segunda ecuación de este sistema; obtenemos

$$B_{12}(A_{21}A_{11}^{-1}A_{12} - A_{22}) = A_{11}^{-1}A_{12}.$$

De aquí hallamos

$$B_{12} = -A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1},$$
$$B_{11} = A_{11}^{-1} - B_{12}A_{21}A_{11}^{-1}.$$

Análogamente de las ecuaciones tres y cuatro del sistema (1).

$$B_{22}=(A_{22}-A_{21}A_{11}^{-1}A_{12})^{-1},$$
$$B_{21}=-B_{22}A_{21}A_{11}^{-1}.$$

Esto es posible a condición de que las operaciones respectivas tengan sentido. Introduzcamos las designaciones siguientes:

$$X=A_{11}^{-1}A_{12};\quad Y=A_{21}A_{11}^{-1};\quad \Theta=A_{22}-A_{21}X=A_{22}-YA_{12}.$$

Entonces las fórmulas para las células de la matriz inversa A^{-1} pueden escribirse en la forma siguiente:

$$\begin{gathered}B_{11}=A_{11}^{-1}+X\Theta^{-1}Y;\quad B_{12}=-X\Theta^{-1};\\ B_{21}=-\Theta^{-1}Y;\quad B_{22}=\Theta^{-1}.\end{gathered}\tag{2}$$

Las fórmulas (2) son válidas a condición de que A_{11}^{-1} y θ^{-1} existan.

Es cómodo disponer los cálculos en la forma del esquema siguiente:

	A_{21}	A_{22}
$X=A_{11}^{-1}A_{12}$	A_{11}^{-1}	A_{12}
θ^{-1}	$Y=A_{21}A_{11}^{-1}$	$\theta=A_{22}-YA_{12}$

La matriz inversa tiene la forma

$$A^{-1}=\left[\begin{array}{c|c}A_{11}^{-1}+X\Theta^{-1}Y & -X\Theta^{-1}\\ \hline -\Theta^{-1}Y & \Theta^{-1}\end{array}\right].$$

Ejemplo 1. Con ayuda de la partición en células invertir la matriz

$$A=\left[\begin{array}{rr|rr}1&0&1&2\\-1&2&3&1\\ \hline 4&0&-2&1\\0&2&1&2\end{array}\right].$$

△ Designemos

$$A_{11}=\begin{bmatrix}1&0\\-1&2\end{bmatrix};\quad A_{12}=\begin{bmatrix}1&2\\3&1\end{bmatrix};\quad A_{21}=\begin{bmatrix}4&0\\0&2\end{bmatrix};\quad A_{22}=\begin{bmatrix}-2&1\\1&2\end{bmatrix}.$$

Realizamos los cálculos necesarios:

$$\det A_{11}=2;\quad A_{11}^{-1}=\frac{1}{2}\begin{bmatrix}2&0\\1&1\end{bmatrix};$$

$$X=A_{11}^{-1}A_{12}=\frac{1}{2}\left[\begin{array}{cc|cc}2&0&1&2\\1&1&3&1\end{array}\right]=\frac{1}{2}\begin{bmatrix}2&4\\4&3\end{bmatrix}.$$

$$Y=A_{21}A_{11}^{-1}=\frac{1}{2}\begin{bmatrix}4&0\\0&2\end{bmatrix}\begin{bmatrix}2&0\\1&1\end{bmatrix}=\begin{bmatrix}4&0\\1&1\end{bmatrix};$$

$$\Theta=A_{22}-YA_{12}=\begin{bmatrix}2&1\\1&2\end{bmatrix}-\begin{bmatrix}4&0\\1&1\end{bmatrix}\begin{bmatrix}1&2\\3&1\end{bmatrix}=\begin{bmatrix}-6&-7\\-3&-1\end{bmatrix},$$

$$\det\Theta=-15;\quad \Theta^{-1}=\frac{1}{15}\begin{bmatrix}1&-7\\-3&6\end{bmatrix}.$$

Escribamos los datos iniciales y los resultados de los cálculos en el esquema:

	$A_{21}=\begin{bmatrix}4&0\\0&2\end{bmatrix}$	$A_{22}=\begin{bmatrix}-2&1\\1&2\end{bmatrix}$
$X=\frac{1}{2}\begin{bmatrix}2&4\\4&3\end{bmatrix}$	$A_{11}^{-1}=\frac{1}{2}\begin{bmatrix}2&0\\1&1\end{bmatrix}$	$A_{12}=\begin{bmatrix}1&2\\3&1\end{bmatrix}$
$\theta^{-1}=\frac{1}{15}\begin{bmatrix}1&-7\\-3&6\end{bmatrix}$	$Y=\begin{bmatrix}4&0\\1&1\end{bmatrix}$	$\theta=\begin{bmatrix}-6&-7\\-3&-1\end{bmatrix}$

Luego tenemos

$$X\Theta^{-1}=\frac{1}{2}\begin{bmatrix}2&4\\4&3\end{bmatrix}\cdot\frac{1}{15}\begin{bmatrix}1&-7\\-3&6\end{bmatrix}=\frac{1}{30}\begin{bmatrix}-10&10\\-5&-10\end{bmatrix};$$

$$X\Theta^{-1}Y=\frac{1}{30}\begin{bmatrix}-10&-10\\-5&-10\end{bmatrix}\begin{bmatrix}4&0\\1&1\end{bmatrix}=\frac{1}{30}\begin{bmatrix}-30&10\\-30&-10\end{bmatrix};$$

$$\Theta^{-1}Y=\frac{1}{15}\begin{bmatrix}1&-7\\-3&6\end{bmatrix}\begin{bmatrix}4&0\\1&1\end{bmatrix}=\frac{1}{15}\begin{bmatrix}-3&-7\\-6&6\end{bmatrix};$$

$$X\Theta^{-1}Y=\frac{1}{2}\begin{bmatrix}2&4\\4&3\end{bmatrix}\cdot\frac{1}{15}\begin{bmatrix}-3&-7\\-6&6\end{bmatrix}=\frac{1}{30}\begin{bmatrix}-30&10\\-30&-10\end{bmatrix}.$$

Para el control hemos calculado el producto $X\Theta^{-1}Y$ por dos métodos: $X\Theta^{-1}Y=(X\Theta^{-1})Y$ y $X\Theta^{-1}Y=X(\Theta^{-1}Y)$. Definitivamente obtenemos

$$A^{-1}=\begin{bmatrix}A_{11}^{-1}+X\Theta^{-1}Y & -X\Theta^{-1}\\ -\Theta^{-1}Y & \Theta^{-1}\end{bmatrix}=$$

$$=\begin{bmatrix}\frac{1}{30}\begin{bmatrix}0&10\\-15&5\end{bmatrix} & \frac{1}{30}\begin{bmatrix}10&-10\\5&10\end{bmatrix}\\ \frac{1}{15}\begin{bmatrix}3&7\\-6&6\end{bmatrix} & \frac{1}{15}\begin{bmatrix}1&-7\\-3&6\end{bmatrix}\end{bmatrix}=$$

$$=\frac{1}{30}\begin{bmatrix}0&10&10&-10\\-15&5&5&10\\6&14&2&-14\\12&-12&6&12\end{bmatrix}.\ \blacktriangle$$

De caso particular del método expuesto de inversión de las matrices celulares sirve el **de orladura sucesiva**.

Supongamos que se da la matriz regular cuadrada de n-ésimo orden

$$A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix},$$

para la cual se necesita hallar la matriz inversa A^{-1}.

Compongamos la sucesión de las matrices:

$$A_1=[a_{11}];$$

$$A_2=\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix};$$

$$A_3=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}=\left[\begin{array}{cc|c} \multicolumn{2}{c|}{\multirow{2}{*}{A_2}} & a_{13} \\ & & a_{23} \\ \hline a_{31} & a_{32} & a_{33} \end{array}\right];$$

$$A_4=\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}=\left[\begin{array}{ccc|c} & & & a_{14} \\ & A_3 & & a_{24} \\ & & & a_{34} \\ \hline a_{41} & a_{42} & a_{43} & a_{44} \end{array}\right];$$

$$A_2^{-1}=B_2=\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix};$$

$$A_3^{-1}=C_3=\left[\begin{array}{cc|c} & & c_{13} \\ & C_2 & c_{23} \\ \hline c_{31} & c_{32} & c_{33} \end{array}\right];$$

$$A_4^{-1}=D_4=\left[\begin{array}{ccc|c} & & & d_{14} \\ & D_3 & & d_{24} \\ & & & d_{34} \\ \hline d_{41} & d_{42} & d_{43} & d_{44} \end{array}\right],$$

etc. Cada matriz siguiente está obtenida de la precedente mediante la orladura.

La matriz inversa a la segunda de estas matrices $A_2^{-1} = B_2$ se halla inmediatamente:

$$A_2^{-1}=B_2=\begin{bmatrix} a_{22}/|A_2| & -a_{12}/|A_2| \\ -a_{21}/|A_2| & a_{11}/|A_2| \end{bmatrix},$$

donde

$$|A_2|=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}=a_{11}a_{22}-a_{12}a_{21}.$$

Con ayuda de la matriz A_2^{-1}, aplicando a A_3 el esquema de cálculo citado anteriormente, se puede obtener A_3^{-1}, luego con ayuda de A_3^{-1} obtener análogamente A_4^{-1}, etc, por fín hallar $A_n^{-1} = A^{-1}$.

Ejemplo 2. Valiéndose del método de orladura sucesiva, invertir la matriz

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

△ Tenemos

$$A=\left[\begin{array}{cc|c|c} -3 & -4 & 5 & 0 \\ 2 & -3 & 1 & 1 \\ \hline 3 & -5 & -1 & 2 \\ \hline 3 & -1 & 4 & 1 \end{array}\right].$$

$$1)\quad A_2=\begin{bmatrix} 3 & -4 \\ 2 & -3 \end{bmatrix};\ \det A_2=-9+8=-1;$$

$$A_2^{-1}=B_2=\begin{bmatrix} 3 & -4 \\ 2 & -3 \end{bmatrix}.$$

2) El esquema de cálculo de $A_3^{-1} = C_3$ tiene la forma siguiente:

	$[a_{31} a_{32}]$	a_{33}
X	A_2^{-1}	$\begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}$
Θ^{-1}	Y	Θ

Realizamos los cálculos:

$$X=A_2^{-1}\begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}=\begin{bmatrix} 3 & -4 \\ 2 & -3 \end{bmatrix}\begin{bmatrix} 5 \\ 1 \end{bmatrix}=\begin{bmatrix} 11 \\ 7 \end{bmatrix};$$

$$Y=[a_{31}\ \ a_{32}]\,A_2^{-1}=[3\ \ -5]\begin{bmatrix} 3 & -4 \\ 2 & -3 \end{bmatrix}=[-1\ \ 3];$$

$$\Theta=a_{33}-Y\begin{bmatrix} a_{13} \\ a_{23} \end{bmatrix}=-1-[-1\ \ 3]\begin{bmatrix} 5 \\ 1 \end{bmatrix}=-1+2=1;$$

$$\Theta^{-1}=1;\quad X\Theta^{-1}Y=\begin{bmatrix} 11 \\ 7 \end{bmatrix}\cdot 1\,[-1\ \ 3]=\begin{bmatrix} -11 & 33 \\ -7 & 21 \end{bmatrix}.$$

Llenamos el esquema:

	3	−5	−1
11 7	3 2	−4 −3	5 1
1	−1	3	1

Obtenemos los elementos de la matriz inversa $A_3^{-1} = C$ después de las siguientes adiciones y multimplicaciones marticiales:

$$c_{33} = \Theta^{-1} = 1; \quad [c_{31}\ c_{32}] = -\Theta^{-1}Y = [1\ \ -3];$$

$$\begin{bmatrix} c_{13} \\ c_{23} \end{bmatrix} = -X\Theta^{-1} = \begin{bmatrix} -11 \\ -7 \end{bmatrix};$$

$$C_2 = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} = A_2^{-1} + X\Theta^{-1}Y = \begin{bmatrix} 3 & -4 \\ 2 & -3 \end{bmatrix} + \begin{bmatrix} -11 & 33 \\ -7 & 21 \end{bmatrix} =$$

$$= \begin{bmatrix} -8 & 29 \\ -5 & 18 \end{bmatrix}.$$

Así pues,

$$A_3^{-1} = C_3 = \begin{bmatrix} -8 & 29 & -11 \\ -5 & 18 & -7 \\ 1 & -3 & 1 \end{bmatrix}.$$

3) Para calcular $A_4^{-1} = D_4 = A^{-1}$ hacemos el esquema:

	$[a_{41} a_{42} a_{43}]$	a_{44}
X	A_3^{-1}	$\begin{bmatrix} a_{14} \\ a_{24} \\ a_{34} \end{bmatrix}$
Θ^{-1}	Y	Θ

Realizamos los cálculos:

$$X = \begin{bmatrix} -8 & 29 & -11 \\ -5 & 18 & -7 \\ 1 & -3 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = \begin{bmatrix} 7 \\ 4 \\ -1 \end{bmatrix};$$

$$Y = [3\ \ -1\ \ 4] \begin{bmatrix} -8 & 29 & -11 \\ -5 & 18 & -7 \\ 1 & -3 & 1 \end{bmatrix} = [-15\ \ 57\ \ -22];$$

$$\Theta = 1 - [-15\ \ 57\ \ -22] \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} = 1 - 13 = -12; \quad \Theta^{-1} = -1/12.$$

Llenamos el esquema:

	3	−1	4	1
7	−8	29	11	0
4	−5	18	−7	1
−1	1	−3	1	2
−1/12	−15	57	−22	−12

Luego tenemos:

$$d_{44} = -\frac{1}{12}; \quad [d_{41}\ d_{42}\ d_{43}] = \frac{1}{12}[-15\ 57\ -22];$$

$$\begin{bmatrix} d_{14} \\ d_{24} \\ d_{34} \end{bmatrix} = \frac{1}{12}\begin{bmatrix} 7 \\ 4 \\ -1 \end{bmatrix};$$

$$D_3 = A_3^{-1} + X\Theta^{-1}Y =$$

$$= \begin{bmatrix} -8 & 29 & 11 \\ -5 & 18 & -7 \\ 1 & -3 & 1 \end{bmatrix} - \frac{1}{12}\begin{bmatrix} 7 \\ 4 \\ -1 \end{bmatrix}[-15\ 57\ -22] =$$

$$= \frac{1}{12}\begin{bmatrix} -96 & 348 & -132 \\ -60 & 216 & -84 \\ 12 & -36 & 12 \end{bmatrix} + \frac{1}{12}\begin{bmatrix} 105 & -399 & 154 \\ 60 & -228 & 88 \\ -15 & 57 & -22 \end{bmatrix} =$$

$$= \frac{1}{12}\begin{bmatrix} 9 & -51 & 22 \\ 0 & -12 & 4 \\ -3 & 21 & 10 \end{bmatrix}.$$

Así pues,

$$A^{-1} = \frac{1}{12}\begin{bmatrix} 9 & -51 & 22 & 7 \\ 0 & -12 & 4 & 4 \\ -3 & 21 & -10 & -1 \\ -15 & 57 & -22 & -1 \end{bmatrix}. \ \blacktriangle$$

El método de orladura puede emplearse sólo en el caso en que todas las matrices intermedias $A_2, A_3, \ldots, A_{n-1}$ son regulares.

§ 2.10. Valor absoluto y norma de la matriz

Por *valor absoluto* (*módulo*) *de la matriz* $A = [a_{ij}]$ se entiende la matriz $|A| = [|a_{ij}|]$ donde todos los elementos $|a_{ij}|$ son los módulos de los elementos de la matriz A.

Sean A y B las matrices para las cuales las operaciones $A + B$ y AB tienen sentido. Entonces

1°) $|A+B| \leqslant |A| + |B|$;

2°) $|AB| \leqslant |A| \cdot |B|$;

3°) $|\alpha A| = |\alpha| \cdot |A|$, donde α es un número.

Por la *norma* de la matriz $A = |a_{ij}|$ se entiende un número real $\|A\|$ que satisface las condiciones siguientes:

1°) $\|A\| \geqslant 0$ (con la particularidad de que $\|A\| = 0$ si y sólo si $A = 0$);

2°) $\|\alpha A\| = |\alpha| \cdot \|A\|$, donde α es un número (con la particularidad de que $\|-A\| = \|A\|$);

3°($\|A+B\| \leqslant \|A\| + \|B\|$;

4°) $\|AB\| \leqslant \|A\| \cdot \|B\|$;

5°) $\|A-B\| \geqslant | \|B\| - \|A\| |$, donde A y B son las matrices para las cuales las operaciones correspondientes tienen sentido.

Para la matriz $A = [a_{ij}]$ de tamaño arbitrario consideremos las tres normas siguientes que se calculan fácilmente:

$\|A\|_1 = \max\limits_i \sum\limits_j |a_{ij}|$, o sea, *la suma máxima de los módulos de los elementos de la matriz, tomados en filas*;

$\|A\|_2 = \max\limits_j \sum\limits_i |a_{ij}|$, o sea, *la suma máxima de los módulos de los elementos de la matriz, tomados en columnas*;

$\|A\|_3 = \sqrt{\sum\limits_{i,j} |a_{ij}|^2}$, o sea, *la raíz cuadrada extraída de la suma de los cuadrados de los módulos de todos los elementos de la matriz.*

Ejemplo 1. Para la matriz

$$A = \begin{bmatrix} 2 & -1 & 4 \\ 5 & 3 & 2 \\ 6 & -7 & 3 \end{bmatrix}$$

calcular $\|A\|_1$, $\|A\|_2$ y $\|A\|_3$.

△ Hallamos

$\|A\|_1 = \max(2+1+4,\ 5+3+2,\ 6+7+3) = \max(7,\ 10,\ 16) = 16$;

$\|A\|_2 = \max(2+5+6,\ 1+3+7,\ 4+2+3) = \max(13,\ 11,\ 9) = 13$;

$\|A\|_3 = \sqrt{2^2+1^2+4^2+5^2+3^2+2^2+6^2+7^2+3^2} = \sqrt{153} = 12{,}2$. ▲

Para el vector $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ estas normas se calculan por las fórmulas siguientes:

$\|\mathbf{x}\|_1 = \max |x_i|$ es *la máxima de las coordenadas del vector, tomada en módulo*;

$\|\mathbf{x}\|_2 = |x_1| + |x_2| + \ldots + |x_n|$ es *la suma de los módulos de las coordenadas del vector*:

$\|x\|_3 = \sqrt{|x_1|^2 + |x_2|^2 + \ldots + |x_n|^2}$ es *la raíz cuadrada extraída de la suma de los cuadrados de los módulos de las coordenadas del vector.*

La norma $\|\mathbf{x}\|_3$ se llama *valor absoluto del vector.*

Ejemplo 2. Para el vector $\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ 3 \\ -5 \end{bmatrix}$ calcular $\|\mathbf{x}\|_1$, $\|\mathbf{x}\|_2$ y $\|\mathbf{x}\|_3$.

△ Tenemos

$\|\mathbf{x}\|_1 = \text{máx}\ (1,\ 2,\ 3,\ 5) = 5;\ \|\mathbf{x}\|_2 = 1 + 2 + 3 + 5 = 11;$

$$\|\mathbf{x}\|_3 = \sqrt{1^2 + 2^2 + 3^2 + 5^2} = \sqrt{39} = 6{,}2 \ \blacktriangle$$

§ 2.11. Rango de la matriz y métodos de su cálculo

Supongamos que se da la matriz rectangular

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}.$$

Si en esta matriz se eligen de un modo arbitrario k filas y k columnas donde $k \leqslant \text{mín}\ (m,\ n)$, entonces los elementos que están en la intersección de estas filas y columnas forman una matriz cuadrada de orden k cuyo determinante se llama *menor* de k-ésimo orden de la matriz A.

Por ejemplo, en la intersección de las filas 1 y 2 con las columnas 1 y 2 de la matriz A se halla la matriz A de segundo orden $\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$ cuyo determinante es el menor de segundo orden de la matriz A. Designémoslo M_2:

$$M_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}.$$

Se llama *rango* de la matriz A el orden máximo de un menor distinto del cero de estra matriz. De la definición del rango se deduce que si el rango de una matriz es igual a r, en la matriz hay al menos un menor de r-ésimo orden no igual a cero, mientras que todos los menores de orden $r + 1$ y de órdenes más altos son iguales a cero. Nótese que el rango de la matriz nula es igual a cero y el de una matriz fila (columna) no nula es igual a la unidad.

Para una matriz rectangular de tamaño $m \times n$ la diferencia entre el número menor de los números m y n y el rango de la matriz se denomina *defecto* de la matriz. Para una matriz cuadrada de tamaño $n \times n$ el defecto es igual a $n - r$. Si el defecto es igual a cero, el rango de la matriz es el mayor de los posibles para el tamaño dado.

Consideremos uno de los métodos de cálculo del rango de una matriz, fundado sobre la misma definición de este último. En este caso es necesario hallar el menor de orden máximo distinto del cero. Al calcular el rango de una matriz con ayuda de este método se pasa de los menores de órdenes menores (comenzando con los menores de primer orden, o sea, con los elementos de la matriz) a los de órdenes mayores, apoyándose en la regla siguiente: supongamos que hemos hallado el menor de r-ésimo orden M_r distinto del cero; entonces hace falta calcular solamente los menores de orden $r + 1$ que orlan el menor dado M_r. Si todos estos menores son iguales a cero, el rango de la matriz es igual a r; en cambio, si al menos uno de ellos es distinto del cero, esta operación ha de aplicarse a este último, con la particularidad de que en este caso el rango de la matriz es a ciencia cierta mayor que r. Este método de cálculo lleva el nombre de **método de orladura.**

Ejemplo 1. Con ayuda del método de orladura hallar el rango de la matriz

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

△ 1) Elijamos el menor de primer orden no igual a cero: $M_1^1 = a_{22} = 4 \neq 0$ (el índice superior designa el número de orden durante el cálculo y el índice inferior, el orden del menor).

2) Hallemos el menor orlante de segundo orden no igual a cero:

$$M_2^1 = \begin{bmatrix} 4 & 1 \\ -2 & 3 \end{bmatrix} \neq 0.$$

3) Consideremos todos los menores de tercer orden que orlan el menor M_2^1, para lo cual compongamos los menores M_3^1 y M_3^2 de las filas 2, 3 y 4:

$$M_3^1 = \begin{vmatrix} 2 & 4 & 1 \\ -1 & -2 & 3 \\ 0 & 0 & 7 \end{vmatrix} = 0,$$

$$M_3^2 = \begin{vmatrix} 4 & 1 & 1 \\ -2 & 3 & 1 \\ 0 & 7 & 2 \end{vmatrix} = 0$$

ya que en estos menores la tercera fila es igual a la suma de la primera fila y la segunda doblada.

Los menores orlantes de las filas 1, 2 y 3 también son iguales a cero, ya que las filas 1 y 2 son iguales. Si todos los menores de tercer orden son iguales a cero, entonces todos los menores de órdenes superiores (comenzando con $3 + 1 = 4$) asimismo son iguales a cero. Por consiguiente, el rango de la matriz es igual a 2 y el defecto $4-2 = 2$. ▲

Puesto que la cantidad de determinantes de diferentes órdenes generados por la matriz suede ser grande, el cálculo del rango por medio de la orladura cuesta mucho trabajo.

Estos cálculos pueden ser simplificados si el rango se halla con ayuda de las *transformaciones elementales de la matriz*.

Entre las transformaciones elementales figuran:

1) permutación de dos filas (columnas);

2) multiplicación de la fila (columna) por cierto número k ($k \neq 0$);

3) adición a una fila (columna) de la otra multiplicada por cualquier número k no igual a cero;

4) eliminación de la fila (columna) compuesta por cero a partir de la matriz;

5) eliminación de la fila (columna) que es una combinación lineal de otras filas (columnas) a partir de la matriz;

Las transformaciones elementales no cambian el rango de la matriz, o sea, como resultado de tales transformaciones se obtiene una nueva matriz que no es igual a la inicial sino es equivalente a ésta (los rangos de estas matrices son iguales). Para designar la equivalencia de las matrices se usa el signo $\sim$.

Las transformaciones de la matriz pueden realizarse en el orden siguiente:

1. Si el elemento situado en el ángulo superior izquierdo de la matriz $a_{11} = 0$ y en la matriz hay elementos no nulos, entonces mediante la permutación de las filas o de las columnas coloquemos en su lugar un elemento no nulo y con ayuda de las transformaciones elementales anulemos todos los demás elementos de la primera fila y luego también los de la primera columna. A continuación la columna 1 y la fila 1 quedan sin variar (podemos sólo permutarlas). Si todos los demás elementos de la matriz transformada son iguales a cero, el rango es igual a la unidad, o sea, $r = 1$.

2. En cambio, si en la matriz transformada $a_{22} = 0$, pero hay elementos distintos del cero, entonces, colocando uno de ellos en el lugar de intersección de la segunda fila y de la segunda columna y aplicando las transformaciones elementales, anulemos primero todos los demás elementos de la segunda fila, luego los de la segunda columna y continuemos las transformaciones de un modo análogo.

En resumidas cuentas, se puede obtener una matriz de cuyos elementos sirven sólo las unidades que están el la diagonal principal, con la particularidad de que su número es igual al rango de la matriz, mientras que los demás elementos son iguales a cero.

Ejemplo 2. Aplicando las transformaciones elementales, determinar el rango de la matriz

$$A=\begin{bmatrix} 1 & 2 & 3 & -1 \\ -1 & 3 & 2 & 0 \\ 3 & 1 & 4 & -2 \\ -3 & 4 & 1 & 1 \end{bmatrix}.$$

△ Realicemos sucesivamente las siguientes transformaciones elementales de la matriz:

a) Primero adicionemos la primera columna a la cuarta y luego, multiplicándola sucesivamente por (—2) y por (—3), adicionémosla a las columnas 2 y 3, respectivamente.

b) eliminemos las columnas 2 y 3, ya que se obtienen de la cuarta columna multiplicándola por (—5).

Entonces obtenemos

$$A=\begin{bmatrix} 1 & 2 & 3 & -1 \\ -1 & 3 & 2 & 0 \\ 3 & 2 & 4 & -2 \\ -3 & 4 & 1 & 1 \end{bmatrix} \overset{a)}{} \begin{bmatrix} 1 & 0 & 0 & 0 \\ -1 & 5 & 5 & -1 \\ 3 & -5 & -5 & 1 \\ -3 & 10 & 10 & -2 \end{bmatrix}$$

$$\overset{b)}{} \begin{bmatrix} 1 & 0 \\ -1 & -1 \\ 3 & 1 \\ -3 & -2 \end{bmatrix}.$$

Es vidente que el rango de la última matriz es igual a 2. No puede ser mayor que 2, ya que si $r = 2$, el defecto es igual a cero $(n - r = 0)$, ni puede ser menor que 2, puesto que en la matriz hay un menor $M_2 \neq 0$.

La matriz obtenida es equivalente a la A y, por lo tanto, $r(A) = 2$.

Llegaremos el mismo resultado si continuamos a transformar la matriz:

c) Primero adicionemos la primera fila a la segunda y luego, multiplicándola sucesivamente por (—3) y por (3), adicionemos a las filas 3 y 4, respectivamente

d) Primero adicionemos la tercera fila a la segunda y luego, multiplicándola por 2, adicionemos a la cuarta fila.

e) Eliminemos las filas nulas.

Tenemos:

$$A \sim \begin{bmatrix} 1 & 0 \\ -1 & -1 \\ 3 & 1 \\ -3 & -2 \end{bmatrix} \overset{c)}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \overset{d)}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \overset{e)}{\sim} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

La última matriz no desempeña el papel de matriz unidad, está obtenida como resultado de transformaciones elementales y es equivalente a la matriz A. La cantidad de unidades presentes en la diagonal principal de la matriz obtenida es igual a 2. Por consiguiente, $r(A) = 2$. ▲

§ 2.12. Concepto de espacio lineal (vectorial). Dependencia lineal de los vectores

Se llama *espacio lineal* (*vectorial*) el conjunto U de los elementos $\mathbf{x}, \mathbf{y}, \mathbf{z}, \ldots$, para el cual están definidas las operaciones de adición de los elementos y de su multiplicación por un número real, operaciones que no hacen salir fuera de los límites de U y satisfacen los siguientes axiomas:

1°) $\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$;

2°) $(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$;

3°) existe un elemento $\mathbf{0} \in U$ tal que $\mathbf{x} + \mathbf{0} = \mathbf{x}$;

4°) para cada $\mathbf{x}$ existe el elemento contrario $-\mathbf{x} \in U$ tal que $\mathbf{x} + (-\mathbf{x}) = \mathbf{0}$;

5°) $1 \cdot \mathbf{x} = \mathbf{x}$;

6°) $\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$;

7°) $(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$;

8°) $\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$,

donde $\mathbf{x}, \mathbf{y}, \mathbf{z} \in U$; α y β son los números reales.

Nótese que $0 \cdot \mathbf{x} = \mathbf{0}$ y $(-1)\mathbf{x} = -\mathbf{x}$.

Ejemplo 1. El conjunto de todos los vectores n-dimensionales con operaciones ordinarias de adición de los vectores y de multiplicación del vector por un número real α forma un espacio lineal, ya que las operaciones indicadas satisfacen los axiomas 1°. . .8°.

De suma de dos vectores $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ e $\mathbf{y} = (y_1, y_2, \ldots \ldots y_n)$ sirve también el vector $\mathbf{z} = \mathbf{x} + \mathbf{y} = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$; el producto del vector $\mathbf{x}$ por el número α es el vector $\alpha\mathbf{x} = (\alpha x_1, \alpha x_2 \ldots, \alpha x_n)$ (véase el § 2.1). En este caso de vector nulo sirve el vector $\mathbf{0} = (0, 0, \ldots, 0)$; de vector contrario a $\mathbf{x}$ sirve el vector $-\mathbf{x} = (-x_1, -x_2, \ldots, -x_n)$. Es evidente que se cumplen todos los axiomas del espacio lineal.

Ejemplo 2. El conjunto de los vectores de diferente dimensión no es un espacio lineal, ya que para ellos no está definida la operación de adición.

Ejemplo 3. Se puede mostrar que las matrices cuadradas de igual orden forman un espacio lineal, mientras que las matrices cuadradas de diferentes órdenes no forman un espacio lineal, ya que para estas matrices la suma no está definida.

En el espacio lineal (vectorial) tiene gran importancia el concepto de dependencia lineal de los vectores.

Si $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ son los vectores del espacio U, entonces el vector $\mathbf{z} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_n\mathbf{x}_n = \sum\limits_{i=1}^{n} c_i\mathbf{x}_i$ se llama *combinación lineal* de los vectores $\mathbf{x}_i$.

Si el vector $\mathbf{z} = \sum_{i=1}^{n} c_i\mathbf{x}_i = 0$, pero entre los números c_1, c_2, . . ., c_n hay, al menos, uno distinto del cero, los vectores $\mathbf{x}_i$ se denominan *linealmente dependientes.*

Los vectores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ se llaman *linealmente independientes* en el caso en que su combinación lineal $\mathbf{z} = \sum_{i=1}^{n} c_i\mathbf{x}_i$ es igual al vector nulo si y sólo si todos los vectores $c_i = 0$ $(i = 1, 2, \ldots, n)$.

Sobre el plano existen no más de dos vectores linealmente independientes y en el espacio tridimensional, no más de tres.

Teorema. *Si los vectores* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ *que pertenecen a un espacio lineal* U *son linealmente dependientes, entonces al menos uno de ellos es la combinación lineal de los demás.*

□ Puesto que los vectores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ son linealmente dependientes, entonces $c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_n\mathbf{x}_n = \mathbf{0}$, con la particularidad de que al menos uno de los números c_i $(i = 1, 2, \ldots, n)$ es distinto del cero. Sea $c_k \neq 0$ $(1 \leqslant k \leqslant n)$. Entonces el vector $\mathbf{x}_k$ es combinación lineal de los vectores $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{k-1}, \mathbf{x}_{k+1}, \ldots, \mathbf{x}_n$:

$$\mathbf{x}_k = -\frac{c_1}{c_k}\mathbf{x}_1 - \frac{c_2}{c_k}\mathbf{x}_2 - \ldots$$
$$\ldots - \frac{c_{k-1}}{c_k}\mathbf{x}_{k-1} - \frac{c_{k+1}}{c_k}\mathbf{x}_{k+1} - \ldots - \frac{c_n}{c_k}\mathbf{x}_n. \quad \blacksquare$$

Ejemplo 4. Se da un sistema de n vectores unitarios n-dimensionales:

$$\mathbf{e}_1 = (1, 0, 0, \ldots, 0),$$
$$\mathbf{e}_2 = (0, 1, 0, \ldots, 0),$$
$$\mathbf{e}_n = (0, 0, 0, \ldots, 1).$$

¿Es linealmente dependiente este sistema?

△ 1) Hagamos la composición lineal de estos vectores e igualémosla a cero:

$$c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \ldots + c_n\mathbf{e}_n = 0.$$

2) Escribamos esta igualdad en las coordenadas:

$$\begin{cases} c_1\cdot 1 + c_2\cdot 0 + \ldots + c_n\cdot 0 = 0, \\ c_1\cdot 0 + c_2\cdot 1 + \ldots + c_n\cdot 0 = 0, \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ c_1\cdot \quad + c_2\cdot 0 + \ldots + c_n\cdot 1 = 0. \end{cases}$$

De aquí $c_1 = c_2 = c_3 = \ldots = c_n = 0$. Esto quiere decir que el sistema de n vectores unitarios n-dimensionales es linealmente independiente. Es evidente que linealmente independiente es también toda parte de este sistema de vectores. ▲

Ejemplo 5. Se dan los vectores $\mathbf{x}_1 = (1, 2, 3)$ y $\mathbf{x}_2 = (3, 6, 7)$. ¿Son linealmente dependientes estos vectores?

△ 1) Hagamos la combinación lineal de estos vectores e igualémosla a cero:

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 = 0. \qquad (*)$$

2) La igualdad vectorial (*) es equivalente al sistema homogéneo de tres ecuaciones lineales con dos incógnitas:

$$\begin{cases} 1\cdot c_1 + 3c_2 = 0, \\ 2c_1 + 6c_2 = 0, \\ 3c_1 + 7c_2 = 0. \end{cases} \qquad (**)$$

3) Dividamos ambos miembros de la segunda ecuación del sistema (**) por 2:

$$\begin{cases} 1\cdot c_1 + 3c_2 = 0, \\ 1\cdot c_1 + 3c_2 = 0, \\ 3c_1 + 7c_2 = 0. \end{cases} \qquad (***)$$

En el sistema (***) hay dos ecuaciones iguales una de las cuales suprimimos.

4) Resolviendo ahora el sistema

$$\begin{cases} 1\cdot c_1 + 3c_2 = 0, \\ 3c_1 + 7c_2 = 0, \end{cases}$$

obtenemos $c_1 = 0$, $c_2 = 0$. Así pues, el sistema dado de los vectores es linealmente independiente. ▲

Ejemplo 6. Se dan los vectores: $\mathbf{x}_1 = (1, 3)$; $\mathbf{x}_2 = (0, 2)$, $\mathbf{x}_3 = (5, 7)$. Resolver la cuestión referente a la dependencia lineal del sistema dado de los vectores.

△ 1) Hagamos la composición lineal de estos vectores e igualémosla a cero:

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = 0. \qquad (*)$$

Escribemos la igualdad (*) en las coordenadas:

$$\begin{cases} 1\cdot c_1 + 0\cdot c_2 + 5c_3 = 0, \\ 3c_1 + 2c_2 + 7c_3 = 0. \end{cases} \qquad (**)$$

Hemos obtenido el sistema de dos ecuaciones lineales con tres incógnitas.

2) Resolvamos este sistema por el método de sustitución. Sustituyamos la expresión $c_1 = -5c_3$, obtenida de la primera ecuación, en la segunda ecuación: $-15c_3 + 2c_2 + 7c_3 = 0$. De aquí $c_2 = 4c_3$. Demos a c_3 un valor numérico arbitrario, no igual a cero, por ejemplo, $c_3 = 1$. Entonces $c_1 = -5$; $c_2 = 4$.

3) Sustituyamos estos valores en la igualdad (*):

$$-5\mathbf{x}_1 + 4\mathbf{x}_2 + \mathbf{x}_3 = 0, \text{ o bien } \mathbf{x}_3 = 5\mathbf{x}_1 - 4\mathbf{x}_2,$$

o sea, el vector $\mathbf{x}_3$ es una combinación lineal de los vectores $\mathbf{x}_1$ y $\mathbf{x}_2$ y esto quiere decir que los vectores $\mathbf{x}_1$, $\mathbf{x}_2$ y $\mathbf{x}_3$ son linealmente dependientes. ▲

§ 2.13. Base de un espacio

El espacio lineal (vectorial) U se llama *n-dimensional* si en éste hay n vectores linealmente independientes y no existe una cantidad mayor de vectores linealmente independientes. El número n se denomina *dimensión del espacio*, y el mismo espacio se llama *espacio de dimensión finita*. (Se designa U_n).

Si en el espacio U existe toda cantidad de vectores linealmente independientes, éste se considera de *dimensión infinita*.

Todo conjunto de n vectores linealmente independientes de un espacio n-dimensional se llama *base* de este espacio.

En cada espacio hay un conjunto infinito de las bases. Una de ellas es el sistema de *vectores unitarios* (*versores*):

$$\begin{aligned}
\mathbf{e}_1 &= (1,\ 0,\ 0,\ \ldots,\ 0),\\
\mathbf{e}_2 &= (0,\ 1,\ 0,\ \ldots,\ 0),\\
&\ldots\ldots\ldots\ldots\\
\mathbf{e}_n &= (0,\ 0,\ 0,\ \ldots,\ 1).
\end{aligned}$$

Teorema 1. *Cada vector del espacio n-dimensional U_n puede ser representado y, además, de un modo único, en forma de la combinación lineal de los vectores de la base.*

□ Sea $\mathbf{x} \in U_n$ y $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$, la base del espacio U_n. Los vectores $\mathbf{x}, \mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ son linealmente dependientes (su cantidad es igual a $n+1$ y supera la dimensión del espacio), o sea,

$$c_0\mathbf{x} + c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \ldots + c_n\mathbf{e}_n = 0, \tag{1}$$

donde cierto coeficiente $c_j \neq 0$ $(0 \leqslant j \leqslant n)$. En la igualdad (1) el coeficiente $c_0 \neq 0$, ya que en el caso contrario

$$c_1\mathbf{e}_1 + c_2\mathbf{e}_2 + \ldots + c_n\mathbf{e}_n = 0,$$

donde $c_j \neq 0$ $(j \geqslant 1)$ lo que contradice la independencia lineal de los vectores $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$. Por consiguiente, la igualdad (1) puede ser resuelta respecto a $\mathbf{x}$:

$$\mathbf{x} = \gamma_1\mathbf{e}_1 + \gamma_2\mathbf{e}_2 + \ldots + \gamma_n\mathbf{e}_n, \tag{2}$$

donde $\gamma_1 = -c_1/c_0$, $\gamma_2 = -c_2/c_0$, $\ldots$, $\gamma_n = -c_n/c_0$.

Ahora bien, todo vector $\mathbf{x}$ del espacio U_n es una combinación lineal de los vectores de la base.

Vamos a demostrar que el desarrollo (2) es único. Supongamos que hay un otro desarrollo

$$\mathbf{x} = \gamma_2'\mathbf{e}_1 + \gamma_2'\mathbf{e}_2 + \ldots + \gamma_n'\mathbf{e}_n \tag{2'}$$

distinto del primero. Sustrayendo de la igualdad (2) la (2′), obtenemos

$$0 = (\gamma_1 - \gamma_1')\,\mathbf{e}_1 + (\gamma_2 - \gamma_2')\,\mathbf{e}_2 - \ldots - (\gamma_n - \gamma_n')\,\mathbf{e}_n. \tag{3}$$

Puesto que los vectores básicos $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$ son linealmente independientes, deben cumplirse las igualdades

$$(\gamma_1 - \gamma_1') = 0, \quad (\gamma_2 - \gamma_2') = 0, \ldots, (\gamma_n - \gamma_n') = 0,$$

de donde se desprende que $\gamma_1 = \gamma_1', \gamma_2 = \gamma_2', \ldots, \gamma_n = \gamma_n'$, o sea, el desarrollo según la base dada es único. ■

En la igualdad (2) los números $\gamma_1, \gamma_2, \ldots, \gamma_n$ se llaman *coordenadas del vector* $\mathbf{x}$ respecto a la base $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n$.

Supongamos que se da un sistema de m vectores en el espacio n-dimensional:

$$\begin{aligned}
\mathbf{x}_1 &= a_{11}\mathbf{e}_1 + a_{21}\mathbf{e}_2 + \ldots + a_{n1}\mathbf{e}_n,\\
\mathbf{x}_2 &= a_{12}\mathbf{e}_1 + a_{22}\mathbf{e}_2 + \ldots + n_{n2}\mathbf{e}_n,\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
\mathbf{x}_m &= a_{1m}\mathbf{e}_1 + a_{2m}\mathbf{e}_2 + \ldots + a_{nm}\mathbf{e}_n.
\end{aligned}$$

Escribamos la matriz compuesta por las coordenadas de estos vectores de un modo tal que los elementos de la j-ésima columna sean iguales a las coordenadas del j-ésimo vector ($j = 1, 2, \ldots, m$):

$$A = \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nm} \end{pmatrix}.$$

Supongamos que el rango de esta matriz es igual a r. Entonces el menor de r-ésimo orden, distinto del cero, se llama *menor básico* de la matriz A. Las filas y columnas en cuya intersección está el menor básico se denominan *filas básicas* y *columnas básicas*.

Teorema 2 (sobre el menor básico). *Las columnas (filas) básicas de la matriz A son linealmente independientes, y toda columna (fila) de esta matriz es una combinación lineal de sus columnas (filas) básicas.*

Ejemplo. Se da el sistema de los vectores: $\mathbf{x}_1 = (-2, 4, 3, 5)$, $\mathbf{x}_2 = (0, 1, 2, -1)$, $\mathbf{x}_3 = (-2, 7, 9, 2)$. Se necesita determinar la dependencia lineal del sistema dado de los vectores; determinar la base del sistema; representar los vectores del sistema en forma de una combinación lineal de los vectores de la base.

△ 1) Compongamos la matriz a partir de las coordenadas de los vectores:

$$A = \begin{bmatrix} -2 & 0 & -2 \\ 4 & 1 & 7 \\ 3 & 2 & 9 \\ 5 & -1 & 2 \end{bmatrix}.$$

Determinemos el rango de la matriz A, aplicando el método de orladura. Tomemos el menor $M_2 = \begin{bmatrix} -2 & 0 \\ 4 & 1 \end{bmatrix} \neq 0$ y consideremos el menor que le orla

$$M_3^1 = \begin{vmatrix} -2 & 0 & -2 \\ 4 & 1 & 7 \\ 3 & 2 & 9 \end{vmatrix}.$$

Para calcular este último sustrayamos la primera columna de la tercera; obtenemos

$$M_3^1 = \begin{vmatrix} -2 & 0 & 0 \\ 4 & 1 & 3 \\ 3 & 2 & 6 \end{vmatrix} = 0,$$

ya que hay dos columnas proporcionales.

El segundo menor orlante

$$M_3^2 = \begin{vmatrix} -2 & 0 & -2 \\ 4 & 1 & 7 \\ 5 & -1 & 2 \end{vmatrix} = \begin{vmatrix} -2 & 0 & 0 \\ 4 & 1 & 3 \\ 5 & -1 & -3 \end{vmatrix} = 0,$$

puesto que después de una transformación análoga éste también contiene dos columnas proporcionales.

Así pues, todos los menores cuyo orden es superior al segundo son iguales a cero, por eso el rango de la matriz A es igual a 2 y el menor $M_2 = \begin{vmatrix} -2 & 0 \\ 4 & 1 \end{vmatrix}$ es básico. Por consiguiente, los vectores $\mathbf{x}_1$ y $\mathbf{x}_2$ son linealmente independientes y forman la base del sistema, mientras que el vector x_3 es su combinación lineal. Ahora bien, el sistema dado de los vectores es linealmente dependiente.

2) Planteemos la igualdad

$$c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + c_3\mathbf{x}_3 = 0, \qquad (*)$$

de donde obtenemos el sistema de ecuaciones

$$\begin{cases} c_1(-2) + c_2 \cdot 0 + c_3(-2) = 0, \\ c_1 \cdot 4 + c_2 \cdot 1 + c_3 \cdot 7 = 0. \end{cases}$$

Expresemos de la primera ecuación $c_1 = -c_3$ y sustituyamos lo obtenido en la segunda ecuación:

$$-4c_3 + c_2 + 7c_3 = 0; \quad c_2 = -3c_3.$$

Pongamos $c_3 = 1$; entonces $c_1 = -1$; $c_2 = -3$. Sustituyendo los valores de c_1, c_2 y c_3 en la igualdad (*), tenemos

$$-\mathbf{x}_1 + 3\mathbf{x}_2 + \mathbf{x}_3 = 0 \text{ o bien } \mathbf{x}_3 = \mathbf{x}_1 + 3\mathbf{x}_2.$$

Ahora bien, el vector $\mathbf{x}_3$ está representado en forma de una combinación lineal de los vectores de la base $\mathbf{x}_1$ y $\mathbf{x}_2$. Este desarrollo es único. Los números (1, 3) son las coordenadas del vector $\mathbf{x}_3$ en la base $(\mathbf{x}_1, \mathbf{x}_2)$. ▲

§ 2.14. Transformación de las coordenadas del vector al cambiar la base

Sean $\{e\} = (e_1, e_2, \ldots, e_n)$ y $\{f\} = (f_1, f_2, \ldots, f_n)$ dos bases de un mismo espacio lineal U_n. Cada vector de la nueva base $\{f\}$ tiene en la vieja base $\{e\}$ las coordenadas $s_{1j}, s_{2j}, \ldots, s_{nj}$, con la particularidad de que al designar las coordenadas s_{ij} $(s = 1, 2, \ldots, n)$ en el primer lugar se indica el número del vector básico viejo respectivo y en el segundo lugar, el del nuevo vector básico. Por lo tanto,

$$f_j = s_{1j}e_1 + s_{2j}e_2 + \ldots s_{nj}e_n \quad (j = 1, 2, \ldots, n) \tag{1}$$

o bien

$$\left\{\begin{array}{l} f_1 = s_{11}e_1 + s_{21}e_2 + \ldots + s_{n1}e_n, \\ f_2 = s_{12}e_1 + s_{22}e_2 + \ldots + s_{n2}e_n, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ f_n = s_{1n}e_1 + s_{2n}e_2 + \ldots + s_{nn}e_n, \end{array}\right. \tag{2}$$

donde

$$S = \begin{bmatrix} s_{11} & s_{21} & \ldots & s_{n1} \\ s_{12} & s_{22} & \ldots & s_{n2} \\ \cdot & \cdot & \cdot & \cdot \\ s_{1n} & s_{2n} & \ldots & s_{nn} \end{bmatrix}$$

es una matriz regular, ya que $\det S \neq 0$ (en el caso contrario las filas de esta matriz y, por lo tanto, también los vectores $f_1, f_2, \ldots, f_n$ serían linealmente dependientes).

La matriz S se llama *matriz de paso* de la vieja base a la nueva.

Sea x el vector arbitrario del espacio lineal dado U_n. Designemos con x_i las coordenadas de este vector en la base vieja y con y_i, sus coordenadas en la nueva base. Es evidente que

$$x = y_1f_1 + y_2f_2 + \ldots + y_nf_n = x_1e_1 + x_2e_2 + \ldots + x_ne_n$$

o bien

$$x = \sum_{i=1}^{n} x_ie_i = \sum_{j=1}^{n} y_jf_j.$$

Teniendo en cuenta las igualdades (2), tenemos $\sum\limits_{j=1}^{n} f_j = \sum\limits_{i=1}^{n} s_{ij}e_i$, de donde, sustituyendo la expresión para f_j, obtenemos

$$x = \sum_{i=1}^{n} x_ie_i = \sum_{j=1}^{n} y_j \sum_{i=1}^{n} s_{ij}e_i = \sum_{i=1}^{n} e_i \sum_{j=1}^{n} s_{ij}y_j. \tag{3}$$

Por consiguiente, en virtud de la independencia lineal de los vectores $e_1, e_2, \ldots, e_n$ hallamos

$$x_i = \sum_{j=1}^{n} s_{ij}y_j \ (i = 1, 2, \ldots, n),$$

o bien

$$\begin{cases} x_1 = s_{11}y_1 + s_{21}y_2 + \ldots + s_{n1}y_n, \\ x_2 = s_{12}y_1 + s_{22}y_2 + \ldots + s_{n2}y_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = s_{1n}y_1 + s_{2n}y_2 + \ldots + s_{nn}y_n. \end{cases} \tag{4}$$

En la forma matricial la relación (4) se escribirá así:

$$\mathbf{x} = S\mathbf{y}, \tag{5}$$

o sea, el vector en las viejas coordenadas (base) es igual a la matriz de paso S multiplicada por el vector en las nuevas coordenadas. Multipliquemos a la izquierda la igualdad (5) por la matriz inversa S^{-1}:

$$S^{-1}\mathbf{x} = \mathbf{y} \text{ o bien } \mathbf{y} = S^{-1}\mathbf{x}. \tag{6}$$

La matriz S^{-1} tiene la forma

$$S^{-1} = \begin{bmatrix} S_{11}/d & S_{12}/d & \ldots & S_{1n}/d \\ S_{21}/d & S_{22}/d & \ldots & S_{2n}/d \\ \ldots & \ldots & \ldots & \ldots \\ S_{n1}/d & S_{n2}/d & \ldots & S_{nn}/d \end{bmatrix},$$

donde d es el determinante de la matriz S ($d \neq 0$, ya que S es una matriz regular) y s_{ij} son los complementos algebraicos de los elementos del determinante d. Entonces

$$\begin{cases} y_1 = \dfrac{S_{11}}{d} x_1 + \dfrac{S_{12}}{d} x_2 + \ldots + \dfrac{S_{1n}}{d} x_n, \\ y_2 = \dfrac{S_{21}}{d} x_1 + \dfrac{S_{22}}{d} x_2 + \ldots + \dfrac{S_{2n}}{d} x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = \dfrac{S_{n1}}{d} x_1 + \dfrac{S_{n2}}{d} x_2 + \ldots + \dfrac{S_{nn}}{d} x_n. \end{cases} \tag{7}$$

Las relaciones (7) son las fórmulas de paso de las coordenadas x_i del vector $\mathbf{x}$ en la vieja base a las coordenadas y_i en la nueva base ($s = 1, 2, \ldots, n$). El paso de las coordenadas x_i a las y_i se lleva a cabo con ayuda de la matriz S^{-1} que es inversa a la matriz de paso de la vieja base a la nueva.

Ejemplo. Hallar las coordenadas del vector $\mathbf{x} = (0, 0, 0, 1)$ en la base $\mathbf{e}_1 = (1, 1, 0, 1)$; $\mathbf{e}_2 = (2, 1, 3, 1)$; $\mathbf{e}_3 = (1, 1, 0, 0)$; $\mathbf{e}_4 = (0, 1, -1, -1)$.

$\triangle$ 1) Compongamos la matriz de paso S de la vieja base a la nueva:

$$S = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 3 & 0 & -1 \\ 1 & 1 & 0 & -1 \end{bmatrix}.$$

2) Hallamos la matriz inversa

$$S^{-1}=\frac{1}{\det S}\begin{bmatrix} S_{11} & S_{21} & S_{31} & S_{41} \\ S_{12} & S_{22} & S_{32} & S_{42} \\ S_{13} & S_{23} & S_{33} & S_{43} \\ S_{14} & S_{24} & S_{34} & S_{44} \end{bmatrix}.$$

Tenemos

$$\det S=\begin{vmatrix} 1 & 2 & 1 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 3 & 0 & -1 \\ 1 & 1 & 0 & -1 \end{vmatrix}=\begin{vmatrix} 1 & 2 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 3 & 0 & -1 \\ 1 & 1 & 0 & -1 \end{vmatrix}=$$

$$=\begin{vmatrix} 0 & -1 & 1 \\ 0 & 3 & -1 \\ 1 & 1 & -1 \end{vmatrix}=\begin{vmatrix} -1 & 1 \\ 3 & -1 \end{vmatrix}=-2.$$

Hemos realizado las siguientes transformaciones del determinante S: primero, de la segunda fila hemos sustraído la primera; luego, hemos desarrollado el determinante obtenido según los elementos de la tercera columna; por último, hemos desarrollado el determinante obtenido según los elementos de la primera columna.

Calculamos los complementos algebraicos:

$$S_{11}=\begin{vmatrix} 1 & 1 & 1 \\ 3 & 0 & -1 \\ 1 & 0 & -1 \end{vmatrix}=2; \quad S_{12}=-\begin{vmatrix} 1 & 1 & 1 \\ 0 & 0 & -1 \\ 1 & 0 & -1 \end{vmatrix}=1;$$

$$S_{13}=\begin{vmatrix} 1 & 1 & 1 \\ 0 & 3 & -1 \\ 1 & 1 & -1 \end{vmatrix}=\begin{vmatrix} 1 & 1 & 1 \\ 0 & 3 & -1 \\ 0 & 0 & -2 \end{vmatrix}=-6;$$

$$S_{14}=-\begin{vmatrix} 1 & 1 & 1 \\ 0 & 3 & 0 \\ 1 & 1 & 0 \end{vmatrix}=-\begin{vmatrix} 0 & 3 \\ 1 & 1 \end{vmatrix}=3;$$

$$S_{21}=-\begin{vmatrix} 2 & 1 & 0 \\ 3 & 0 & -1 \\ 1 & 0 & -1 \end{vmatrix}=\begin{vmatrix} 3 & -1 \\ 1 & -1 \end{vmatrix}=-2;$$

$$S_{22}=\begin{vmatrix} 1 & 1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & -1 \end{vmatrix}=-\begin{vmatrix} 0 & -1 \\ 1 & -1 \end{vmatrix}=-1;$$

$$S_{23}=\begin{vmatrix} 1 & 2 & 0 \\ 0 & 3 & -1 \\ 1 & 1 & -1 \end{vmatrix}=-\begin{vmatrix} 1 & 2 & 0 \\ 0 & 3 & -1 \\ 0 & -1 & -1 \end{vmatrix}=-\begin{vmatrix} 3 & -1 \\ -1 & -1 \end{vmatrix}=4;$$

$$S_{24}=\begin{vmatrix}1 & 2 & 1\\ 0 & 3 & 0\\ 1 & 1 & 0\end{vmatrix}=\begin{vmatrix}0 & 3\\ 1 & 1\end{vmatrix}=-3;$$

$$S_{31}=\begin{vmatrix}2 & 1 & 0\\ 1 & 1 & 1\\ 1 & 0 & -1\end{vmatrix}=0 \quad \text{(la fila 1ª es igual a la suma de las filas 2 y 3)}$$

$$S_{32}=-\begin{vmatrix}1 & 1 & 0\\ 1 & 1 & 1\\ 1 & 0 & -1\end{vmatrix}=-\begin{vmatrix}1 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & -1\end{vmatrix}=\begin{vmatrix}0 & 1\\ 1 & -1\end{vmatrix}=-1;$$

$$S_{33}=\begin{vmatrix}1 & 2 & 0\\ 1 & 1 & 1\\ 1 & 1 & -1\end{vmatrix}=\begin{vmatrix}1 & 2 & 0\\ 1 & 1 & 1\\ 2 & 2 & 0\end{vmatrix}=-\begin{vmatrix}1 & 2\\ 2 & 2\end{vmatrix}=2;$$

$$S_{34}=-\begin{vmatrix}1 & 2 & 1\\ 1 & 1 & 1\\ 1 & 1 & 0\end{vmatrix}=-\begin{vmatrix}1 & 2 & 1\\ 0 & -1 & 0\\ 1 & 1 & 0\end{vmatrix}=-\begin{vmatrix}0 & -1\\ 1 & 1\end{vmatrix}=-1;$$

$$S_{41}=-\begin{vmatrix}2 & 1 & 0\\ 1 & 1 & 1\\ 3 & 0 & -1\end{vmatrix}=-\begin{vmatrix}2 & 1 & 0\\ 1 & 1 & 1\\ 4 & 1 & 0\end{vmatrix}=\begin{vmatrix}2 & 1\\ 4 & 1\end{vmatrix}=-2;$$

$$S_{42}=\begin{vmatrix}1 & 1 & 0\\ 1 & 1 & 1\\ 0 & 0 & -1\end{vmatrix}=0 \quad \text{(hay dos columnas iguales);}$$

$$S_{43}=-\begin{vmatrix}1 & 2 & 0\\ 1 & 1 & 1\\ 0 & 3 & -1\end{vmatrix}=-\begin{vmatrix}1 & 2 & 0\\ 1 & 4 & 0\\ 0 & 3 & -1\end{vmatrix}=\begin{vmatrix}1 & 2\\ 1 & 4\end{vmatrix}=2;$$

$$S_{44}=\begin{vmatrix}1 & 2 & 1\\ 1 & 1 & 1\\ 0 & 3 & 0\end{vmatrix}=0.$$

Por lo tanto,

$$S^{-1}=-\frac{1}{2}\begin{bmatrix}2 & -2 & 0 & -2\\ 1 & -1 & -1 & 0\\ -6 & 4 & 2 & 2\\ 3 & -3 & -1 & 0\end{bmatrix}.$$

3) Con ayuda de la fórmula (6) hallamos las coordenadas del vector $\mathbf{x}$ en la nueva base:

$$\mathbf{y}=S^{-1}\mathbf{x}=-\frac{1}{2}\begin{bmatrix} 2 & -2 & 0 & -2 \\ 1 & -1 & -1 & 0 \\ -6 & 4 & 2 & 2 \\ 3 & -3 & -1 & 0 \end{bmatrix}\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}=$$

$$=-\frac{1}{2}\begin{bmatrix} -2 \\ 0 \\ 2 \\ 0 \end{bmatrix}=\begin{bmatrix} 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}. \blacktriangle$$

Ejercicios

1. Calcular AB si

a) $A=\begin{bmatrix} 1 & -2 & 3 \\ 3 & -1 & 2 \\ 4 & -2 & 1 \end{bmatrix}$; $B=\begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 3 \\ 2 & 1 & 3 \end{bmatrix}$;

b) $A=\begin{bmatrix} 1 & -3 & 2 \\ 3 & -4 & 1 \\ 2 & -5 & 3 \end{bmatrix}$; $B=\begin{bmatrix} 2 & 5 & 6 \\ 1 & 2 & 5 \\ 1 & 3 & 2 \end{bmatrix}$.

2. Calcular $2(A+B)(2B-A)$ si:

a) $A=\begin{bmatrix} 2 & 3 & -1 \\ 4 & 5 & 2 \\ -1 & 0 & 7 \end{bmatrix}$; $B=\begin{bmatrix} -1 & 0 & 5 \\ 0 & 1 & 3 \\ 2 & -2 & 4 \end{bmatrix}$;

b) $A=\begin{bmatrix} 4 & 5 & -2 \\ 3 & -1 & 0 \\ 4 & 2 & 7 \end{bmatrix}$; $B=\begin{bmatrix} 2 & 1 & -1 \\ 0 & 1 & 3 \\ 5 & 7 & 3 \end{bmatrix}$.

3. Hallar el producto XY si:

a) $X=\begin{bmatrix} 5 \\ 7 \\ -3 \\ 2 \end{bmatrix}$, $Y=[1\ \ 2\ \ -2\ \ 3]$; b) $X=\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$;

$Y=[4\ \ 5]$;

c) $X=[10\ \ 17\ \ 8\ \ 5\ \ 11]$, $Y=\begin{bmatrix} 12 \\ 7 \\ 5 \\ 4 \\ 10 \end{bmatrix}$.

4. Hallar el producto AX si:

a) $A=\begin{bmatrix} 2 & -1 & -3 & 0 & 4 \\ 7 & 2 & 5 & 5 & 2 \\ 3 & 4 & 1 & -7 & -1 \end{bmatrix}, \quad X=\begin{bmatrix} -1 \\ 1 \\ 2 \\ 3 \\ 0 \end{bmatrix};$

b) $A=\begin{bmatrix} 1 & 2 & -4 \\ 3 & 0 & -2 \\ 2 & 2 & -3 \end{bmatrix}, \quad X=\begin{bmatrix} -2 \\ -3 \\ -4 \end{bmatrix}.$

5. Calcular los determinantes:

a) $d=\begin{vmatrix} 1 & 3 & -4 \\ 0 & 1 & 1 \\ 2 & -5 & 3 \end{vmatrix}; \quad$ b) $d=\begin{vmatrix} 1 & 1 & -2 & 3 \\ 7 & 8 & 4 & 1 \\ 2 & 4 & 6 & -3 \\ 5 & 6 & 8 & -4 \end{vmatrix};$

c) $d=\begin{vmatrix} 1{,}6 & 5{,}4 & -7{,}7 & -3{,}1 \\ 8{,}2 & 1{,}4 & -2{,}3 & 0{,}2 \\ 5{,}3 & -5{,}9 & 2{,}7 & -7{,}9 \\ 0{,}7 & 1{,}9 & -8{,}5 & 4{,}8 \end{vmatrix}.$

6. Calcular A^{-1} para las matrices siguientes:

a) $A=\begin{bmatrix} 1 & -3 & 2 \\ 3 & -4 & 0 \\ 2 & -5 & 3 \end{bmatrix}; \quad$ b) $A=\begin{bmatrix} 2 & 7 & 1 & 4 \\ 5 & 2 & 0 & -1 \\ 3 & 4 & 2 & 1 \\ 6 & 8 & 4 & 3 \end{bmatrix};$

c) $A=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 2 & -3 & 4 & 0 \\ 3 & 2 & -1 & 3 \end{bmatrix}.$

7. Hallar AB, donde

$$A=\begin{bmatrix} 1 & 4 & 1 & 3 \\ 0 & -1 & 3 & -1 \\ 3 & 1 & 0 & 2 \\ 1 & -2 & 5 & 1 \end{bmatrix}, \quad B=\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 4 & 2 & 3 \\ 1 & 10 & 3 & 6 \\ 7 & 10 & 1 & 4 \end{bmatrix},$$

por dos métodos: a) partiendo A y B en células cuadradas; b) partiendo A y B en células por medio de la orladura.

8. Calcular A^{-1} aplicando la partición en células cuadradas y la or adura, si:

$$\text{a) } A=\begin{bmatrix} 1 & 2 & 3 & -2 \\ 2 & -1 & -2 & -3 \\ 3 & 2 & -1 & 2 \\ 2 & -3 & 2 & 1 \end{bmatrix}; \quad \text{b) } A=\begin{bmatrix} 3 & -2 & 2 & 0 \\ 2 & 1 & 1 & -2 \\ 3 & -1 & 2 & 1 \\ 1 & 2 & -1 & -1 \end{bmatrix}.$$

9. Desarrollar las matrices dadas en el ejercicio 8 en producto de dos matrices triangulares e invertirlas, aplicando el desarrollo de las matrices en producto de dos matrices triangulares.

10. Resolver las ecuaciones matriciales:

$$\text{a) } X\begin{bmatrix} 0 & 3 & -1 \\ 2 & -1 & 2 \\ -3 & 1 & 4 \end{bmatrix}=\begin{bmatrix} 7 & 6 & -3 \\ -8 & 3 & 6 \\ 11 & 9 & 13 \end{bmatrix};$$

$$\text{b) } \begin{bmatrix} 7 & 6 & -3 \\ -8 & 3 & 6 \\ 11 & 9 & 13 \end{bmatrix}X=\begin{bmatrix} -3 & -10 & -4 \\ 21 & 14 & -10 \\ 48 & 2 & 30 \end{bmatrix}.$$

11. Calcular los rangos de las matrices siguientes:

$$\text{a) } A=\begin{bmatrix} 2 & -1 & 3 & -2 & 4 \\ 4 & -2 & 5 & 1 & 7 \\ 2 & -1 & 1 & 8 & 2 \end{bmatrix}; \quad \text{b) } B=\begin{bmatrix} 2 & 4 & 1 & 0 \\ 2 & 4 & 1 & 0 \\ -1 & -2 & 3 & 1 \\ 5 & 10 & 6 & 1 \\ 0 & 0 & 7 & 2 \end{bmatrix}.$$

12. Determinar si son linealmente dependientes o linealmente independientes los siguientes sistemas de los vectores: a) $\mathbf{x}_1 = (5, 4, 3)$, $\mathbf{x}_2 = (3, 3, 2)$, $\mathbf{x}_3 = (8, 1, 3)$; b) $\mathbf{x}_1 = (1, 0, 0, 2, 5)$; $\mathbf{x}_2 = (0, 1, 0, 3, 4)$; $\mathbf{x}_3 = (0, 0, 1, 4, 7)$, $\mathbf{x}_4 = (2, -3, 4, 11, 12)$.

13. Para el sistema de los vectores $\mathbf{x}_1 = (5, 2, -3, 1)$; $\mathbf{x}_2 = (4, 1, -2, 3)$, $\mathbf{x}_3 = (1, 1, -1, -2)$; $\mathbf{x}_4 = (3, 4, -1, 2)$ hallar la base y expresar los demás vectores por los básicos.

14. Hallar las coordenadas del vector $\mathbf{x} = (1, 2, 1, 1)$ en la base $\mathbf{e}_1 = (1, 1, 1, 1)$, $\mathbf{e}_2 = (1, 1, -1, -1)$, $\mathbf{e}_3 = (1, -1, 1, -1)$, $\mathbf{e}_4 = (1, -1, -1, 1)$.

CAPÍTULO III

Resolución de los sistemas de ecuaciones lineales

§ 3.1. Sistemas de ecuaciones lineales

En la forma general *un sistema de m ecuaciones lineales con n incógnitas* se escribe así:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1j}x_j + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2j}x_j + \ldots + x_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{ij}x_j + \ldots + a_{in}x_n = b_i, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mj}x_j + \ldots + a_{mn}x_n = b_m. \end{cases} \tag{1}$$

Las ecuaciones del sistema se consideran numeradas: la primera, la segunda, . . ., la m-ésima. Los números $x_1, x_2, \ldots, x_n$ se llaman *incógnitas* del sistema, $a_{11}, a_{12}, \ldots, a_{mn}$, *coeficientes* de las incógnitas del sistema.

En la i-ésima ecuación el coeficiente de la incógnita x_{ij} se designa con a_{ij}, donde el primer índice i señala el número de la ecuación en la cual está el coeficiente dado y el segundo índice j señala el número de la incógnita junto a la cual se halla el coeficiente dado. Por ejemplo, el coeficiente a_{23} está en la segunda ecuación del sistema, adjunto a la incógnita x_3.

Los números $b_1, b_2, \ldots, b_m$ se denominan *términos independientes* del sistema.

Brevemente el sistema (1) puede ser escrito así:

$$\sum_{j=1}^{n} a_{ij}x_j = b_i \quad (i = 1, 2, \ldots, m). \tag{1'}$$

Se llama *solución* del sistema de ecuaciones lineales (1) todo conjunto de los números $\alpha_1, \alpha_2, \ldots, \alpha_n$ que, puestos en lugar de las incógnitas $x_1, x_2, \ldots, x_n$ en las ecuaciones del sistema dado convierten todas estas ecuaciones en identidades.

El sistema de ecuaciones lineales (1) se denomina *compatible* si tiene una solución. Si el sistema de ecuaciones lineales no tiene solución, se considera *incompatible* (o *contradictorio*).

Un sistema compatible de ecuaciones lineales puede tener una o varias soluciones y se llama *determinado* si tiene una sola solución e *indeterminado* si tiene más que una solución.

Dos sistemas de ecuaciones lineales con una misma cantidad de incógnitas se llaman *equivalentes* si son ambos incompatibles o son ambos compatibles y tienen unas mismas soluciones.

Se denominan *transformaciones elementales de un sistema de ecuaciones lineales* los siguientes tres tipos de transformaciones:

1) permutación de dos ecuaciones del sistema;

2) multiplicación de ambos miembros de la ecuación del sistema por todo número distinto del cero;

3) adición (sustracción) a ambos miembros de una ecuación de los miembros respectivos de otra ecuación, multiplicados por todo número.

Se puede demostrar que las transformaciones elementales hacen pasar el sistema dado de ecuaciones a un otro, equivalente. El cumplimiento de las transformaciones elementales es equivalente a la expresión de una incógnita por medio de otras.

El sistema en el cual los términos independientes $b_1, b_2, \ldots, b_n$ son iguales a cero se nombra *homogéneo*.

§ 3.2. Teorema de Kronecker—Capelli

Supongamos que se da un sistema de ecuaciones lineales

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m. \end{cases} \tag{10}$$

Para determinar las condiciones de compatibilidad de este sistema es necesario introducir el concepto de matriz de un sistema y el de matriz ampliada de este último.

Se llama *matriz del sistema* (1) la matriz compuesta por los coeficientes de las incógnitas de este sistema:

$$A = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}.$$

Si la columna de los términos independientes se adjunta a la matriz A, se obtiene la matriz $\overline{A}$ que se denomina *matriz ampliada del sistema* (1):

$$\overline{A} = \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} b_2 \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mn} b_m \end{bmatrix}.$$

De la definición de la matriz del sistema A y de la matriz ampliada $\overline{A}$ se pone claro que sus rangos $r(A)$ y $r(\overline{A})$ son iguales entre sí o el rango $r(\overline{A})$ es en una unidad mayor que el $r(A)$.

La cuestión referente a la compatibilidad del sistema (1) se resuelve con ayuda del **teorema de Kronecker—Capelli:** *un sistema de ecuaciones lineales* (1) *es compatible si y sólo si el rango de la matriz ampliada* $\bar{A}$ *es igual al de la matriz* A, *o sea, si* $r(\bar{A}) = r(A)$.

Ejemplo. Investigar si es compatible o no el siguiente sistema de ecuaciones:

$$\begin{cases} 7x_1 + 3x_2 = 2, \\ x_1 - 2x_2 = -3, \\ 4x_1 + 9x_2 = 11. \end{cases}$$

△1) Compongamos la matriz del sistema dado y calculemos su rango:

$$A = \begin{bmatrix} 7 & 3 \\ 1 & -2 \\ 4 & 9 \end{bmatrix}; \ r(A) = 2, \text{ puesto que } M_2^1 = \begin{vmatrix} 7 & 3 \\ 1 & -2 \end{vmatrix} \neq 0.$$

2) Compongamos la matriz ampliada del sistema

$$\bar{A} = \begin{bmatrix} 7 & 3 & 2 \\ 1 & -2 & -3 \\ 4 & 9 & 11 \end{bmatrix}.$$

Puesto que $M_2^1 \neq 0$ y el menor que lo orla

$$M_3 = \det A = \begin{vmatrix} 7 & 3 & 2 \\ 1 & -2 & -3 \\ 4 & 9 & 11 \end{vmatrix} = 0$$

(la primera fila es igual a la suma de la segunda, multiplicada por 3, y de la tercera), entonces $r(\bar{A}) = 2$. Así pues, $r(A) = r(\bar{A}) = 2$, o sea, el sistema dado es compatible. ▲

Corolario 1. *Si el sistema* (1) *es compatible y el rango de la matriz del sistema* $r(A) = r$ *es igual a la cantidad de incógnitas* n, *entonces el sistema tiene una sola solución.*

Corolario 2. *Si el sistema* (1) *es compatible y el rango de la matriz del sistema* $r(A) = r$ *es menor que la cantidad de incógnitas* n, *el sistema tiene un conjunto infinito de soluciones.*

§ 3.3. Resolución de los sistemas de n ecuaciones lineales con n incógnitas valiéndose de las fórmulas de Cramer

Supongamos que se da un sistema de ecuaciones lineales en el cual la cantidad de ecuaciones es igual a la de incógnitas:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ \dots \ + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ \dots \ + a_{2n}x_n = b_2, \\ \dots\dots\dots\dots\dots\dots\dots\dots \\ a_{n1}x_1 + a_{n2}x_2 + \ \dots \ + a_{nn}x_n = b_n \end{cases} \tag{1}$$

y

$$A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad \mathbf{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad \mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

son la matriz del sistema, la columna de los términos independientes y la columna de las incógnitas, respectivamente. Supongamos que el determinante del sistema

$$d=\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} \neq 0.$$

Si ahora reemplazamos sucesivamente en el determinante d las columnas de los coeficientes de las incógnitas x_j $(j = 1, 2, \dots, n)$ por la columna de los términos independientes b_i, obtendremos, respectivamente, los determinantes

$$d_1=\begin{vmatrix} b_1 & a_{12} & \dots & a_{1,n-1} & a_{1n} \\ b_2 & a_{22} & \dots & a_{2,n-1} & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ b_n & a_{n2} & \dots & a_{n,n-1} & a_{nn} \end{vmatrix};$$

$$d_2=\begin{vmatrix} a_{11} & b_1 & \dots & a_{1,n-1} & a_{1n} \\ a_{21} & b_2 & \dots & a_{2,n-1} & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & b_n & \dots & a_{n,n-1} & a_{nn} \end{vmatrix};$$

$$\dots\dots\dots\dots\dots$$

$$d_{n-1}=\begin{vmatrix} a_{11} & a_{12} & \dots & b_1 & a_{1n} \\ a_{21} & a_{22} & \dots & b_2 & a_{2n} \\ a_{n1} & a_{n2} & \dots & b_n & a_{nn} \end{vmatrix};$$

$$d_n=\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1,n-1} & b_1 \\ a_{21} & a_{22} & \dots & a_{2,n-1} & b_2 \\ a_{n1} & a_{n2} & \dots & a_{n,n-1} & b_n \end{vmatrix}.$$

Teorema de Cramer. *Un sistema de n ecuaciones lineales con n incógnitas cuyo determinante se distingue del cero es siempre compatible y tiene una sola solución que se calcula por las fórmulas*

$$x_1 = d_1/d; \; x_2 = d_2/d; \; \dots ; \; x_{n-1} = d_{n-1}/d; \quad x_n = d_n/d. \tag{2}$$

las fórmulas (2) se llaman *fórmulas de Cramer*.

Ejemplo 1. Resolver con ayuda de las fórmulas de Cramer el sistema de ecuaciones lineales

$$\begin{cases} x_1+x_2+2x_3=-1, \\ 2x_1-x_2+2x_3=-4, \\ 4x_1+x_2+4x_3=-2. \end{cases}$$

△ 1) Calculamos el determinante del sistema

$$d=\begin{vmatrix} 1 & 1 & 2 \\ 2 & -1 & 2 \\ 4 & 1 & 4 \end{vmatrix}=2\begin{vmatrix} 1 & 1 & 1 \\ 2 & -1 & 1 \\ 4 & 1 & 2 \end{vmatrix}=2(-2+4+2+4-4-1)=6.$$

2) Calculamos los determinantes compuestos por los coeficientes de las incógnitas x_1, x_2, x_3:

$$d_1=\begin{vmatrix} -1 & 1 & 2 \\ -4 & -1 & 2 \\ -2 & 1 & 4 \end{vmatrix}=2\begin{vmatrix} -1 & 1 & 1 \\ -4 & -1 & 1 \\ -2 & 1 & 2 \end{vmatrix}=2(2-2-4-2+1+8)=6;$$

$$d_2=\begin{vmatrix} 1 & -1 & 2 \\ 2 & -4 & 2 \\ 4 & -2 & 4 \end{vmatrix}=2\begin{vmatrix} 1 & -1 & 1 \\ 2 & -4 & 1 \\ 4 & -2 & 2 \end{vmatrix}=4\begin{vmatrix} 1 & -1 & 1 \\ 2 & -4 & 1 \\ 2 & -1 & 1 \end{vmatrix}=$$

$$=4(-4-2-2+8+2+1)=12;$$

$$d_3=\begin{vmatrix} 1 & 1 & -1 \\ 2 & -1 & -4 \\ 4 & 1 & -2 \end{vmatrix}=(2-16-2-4+4+4)=-12.$$

3) Utilizando las fórmulas de Cramer (2), hallamos la solución del sistema: $x_1=d_1/d=6/6=1$; $x_2=d_2/d=12/6=2$; $x_3=$ $=-12/6=-2$. ▲

Ejemplo 2. Resolver con ayuda de las fórmulas de Cramer el sistema de ecuaciones lineales

$$\begin{cases} 2x_1-x_2+x_3+3x_4=-1, \\ x_1+x_2-x_3-4x_4=6, \\ 3x_1-x_2+x_3+x_4=4, \\ x_1-3x_2+3x_4=-5. \end{cases}$$

△ Hallamos los determinantes d, d_1, d_2, d_3 y d_4, desarrollándolos en menores según los elementos de la última fila y luego aplicando la regla de los triángulos:

$$d=\begin{vmatrix} 2 & -2 & 1 & 3 \\ 1 & 1 & -1 & -4 \\ 3 & -1 & 1 & 1 \\ 1 & -3 & 0 & 3 \end{vmatrix}=-\begin{vmatrix} -1 & 1 & 3 \\ 1 & -1 & -4 \\ -1 & 1 & 1 \end{vmatrix}-3\begin{vmatrix} 2 & 1 & 3 \\ 1 & -1 & -4 \\ 3 & 1 & 1 \end{vmatrix}+$$

$$+3\begin{vmatrix} 2 & -2 & 1 \\ 1 & 1 & -1 \\ 3 & -1 & 1 \end{vmatrix}=(-1)\cdot 0-3\cdot 5+3\cdot 0=-15$$

(los determinantes primero y tercero son iguales a cero, ya que tienen las columnas proporcionales);

$$d_1=\begin{vmatrix}-1&-1&1&3\\6&6&-1&-4\\4&-1&1&1\\-5&-3&0&3\end{vmatrix}=5\begin{vmatrix}-1&1&3\\1&-1&-4\\-1&1&1\end{vmatrix}-3\begin{vmatrix}-1&1&3\\6&-1&-4\\4&1&1\end{vmatrix}+$$

$$+3\begin{vmatrix}-1&-1&1\\6&1&-1\\4&-1&1\end{vmatrix}=5\cdot 0-3\cdot 5+3\cdot 0=-15;$$

$$d_2=\begin{vmatrix}2&-1&1&3\\1&6&-1&-4\\3&1&1&1\\1&-5&0&3\end{vmatrix}=-\begin{vmatrix}-1&1&3\\6&-1&-4\\4&1&1\end{vmatrix}-5\begin{vmatrix}2&1&3\\1&-1&-4\\3&1&1\end{vmatrix}+$$

$$+3\begin{vmatrix}2&-1&1\\1&6&-1\\3&4&1\end{vmatrix}=-1\cdot 5-5\cdot 5+3\cdot 10=0;$$

$$d_3=\begin{vmatrix}2&-1&-1&3\\1&1&6&-4\\3&-1&4&1\\1&-3&-5&3\end{vmatrix}=-\begin{vmatrix}-1&-1&3\\1&6&-4\\-1&4&1\end{vmatrix}-3\begin{vmatrix}2&-1&3\\1&6&-4\\3&4&1\end{vmatrix}+$$

$$+5\begin{vmatrix}2&-1&3\\1&1&-4\\3&-1&1\end{vmatrix}+3\begin{vmatrix}2&-1&1\\1&1&6\\3&-1&4\end{vmatrix}=(1-)\cdot 5-3\cdot 15+5\cdot(-5)+$$

$$+3\cdot 10=-45;$$

$$d_4=\begin{vmatrix}2&-1&1&-1\\1&1&-1&6\\3&-1&1&4\\1&-3&0&-5\end{vmatrix}=-\begin{vmatrix}-1&1&-1\\1&-1&6\\-1&1&4\end{vmatrix}-3\begin{vmatrix}2&1&-1\\1&-1&6\\3&1&4\end{vmatrix}-$$

$$-5\begin{vmatrix}2&-1&1\\1&1&-1\\3&-1&1\end{vmatrix}=(-1)\cdot 0-3(-10)-5\cdot 0=30.$$

Ahora con ayuda de las fórmulas de Cramer (2) obtenemos la solución del sistema:

$$x_1=d_1/d=(-15)/(-15)=1;\ x_2=d_2/d=0/(-15)=0;$$
$$x_3=d_3/d=(-45)/(-15)=3;\quad x_4=d_4/d=30/(-15)=-2.\blacktriangle$$

Nótese que la resolución de un sistema de ecuaciones lineales con ayuda de las fórmulas de Cramer cuesta mucho trabajo. En la práctica tales sistemas suelen resolverse por otros métodos.

§ 3.4. Resolución de los sistemas arbitrarios de ecuaciones lineales

Sea

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{cases} \tag{1}$$

un sistema arbitrario de ecuaciones lineales donde el número m de ecuaciones del sistema no es igual al número n de incógnitas ($m \neq n$).

Supongamos que el sistema (1) es compatible, o sea, $r(A) = r(\bar{A}) = r$ y $r \leqslant \text{mín}\{m, n\}$. Entonces en las matrices A y $\bar{A}$ del sistema dado hay r filas linealmente independientes y las demás $m - r$ filas resultan sus combinaciones lineales. Permutando las ecuaciones, se puede alcanzar que estas r filas linealmente independientes ocupen los primeros r lugares.

De aquí se desprende que toda de las últimas $m - r$ ecuaciones del sistema (1) puede representarse como suma de las primeras r ecuaciones (que se llaman *linealmente independientes* o *básicas*) tomadas con ciertos coeficientes. Entonces el sistema (1) es equivalente al siguiente sistema de r ecuaciones con n incógnitas:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{r1}x_1 + a_{r2}x_2 + \ldots + a_{rn}x_n = b_r \end{cases} \tag{2}$$

Supongamos que el menor de r-ésimo orden compuesto por los coeficientes de las primeras r incógnitas es distinto del cero:

$$M_r = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1r} \\ a_{21} & a_{22} & \ldots & a_{2r} \\ \cdot & \cdot & \cdot & \cdot \\ a_{r1} & a_{r2} & \ldots & a_{rr} \end{vmatrix} \neq 0,$$

o sea, es un menor básico (véase el § 2.13). En este caso las incógnitas cuyos coeficientes constituyen el menor básico se nombran *incógnitas básicas* y las demás $n - r$ incógnitas se llaman *independientes*.

En cada una de las ecuaciones del sistema (2) transpongamos al segundo miembro todos los términos con incógnitas independientes

$x_{r+1}, {}_{r+2}, \ldots, x_n$. Entonces obtenemos el sistema

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1r}x_r = b_1 - a_{1,r+1}x_{r+1} - \\ \qquad - a_{1,r+2}x_{r+2} - \ldots - a_{1n}x_n, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2r}x_r = b_2 - a_{2,r+1}x_{r+1} - \\ \qquad - a_{2,r+2}x_{r+2} - \ldots - a_{2n}x_n \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{r1}x_1 + a_{r2}x_2 + \ldots + a_{rr}x_r = b_r - a_{r,r+1}x_{r+1} - \\ \qquad - a_{r,r+1}x_{r+2} - \ldots - a_{rn}x_n \end{cases} \tag{3}$$

el cual contiene r ecuaciones con r incógnitas básicas. Puesto que el determinante del sistema (3) es el menor básico $M_r \neq 0$, entonces el sistema (3) tiene una sola solución respecto a las incógnitas básicas $x_1, x_2, \ldots, x_r$ la cual puede ser hallada con ayuda de las fórmulas de Cramer. Asignado a las incógnitas independientes $x_{r+1}, x_{r+2}, \ldots, x_n$ los valores numéricos arbitrarios $x_{r+1} = c_1$, $x_{r+2} = c_2, \ldots, x_n = c_{n-r}$, obtenemos la solución general del sistema inicial (1).

Ejemplo. Resolver el sistema de tres ecuaciones con cuatro incógnitas

$$\begin{cases} 3x_1 - 2x_2 + 5x_3 + 4x_4 = 2, \\ 6x_1 - 4x_2 + 4x_3 + 3x_4 = 3, \\ 9x_1 - 6x_2 + 3x_3 + 2x_4 = 4. \end{cases} \tag{*}$$

△ Investiguemos si es compatible o no el sistema. Compongamos las matrices A y $\bar{A}$:

$$A = \begin{bmatrix} 3 & -2 & 5 & 4 \\ 6 & -4 & 4 & 3 \\ 9 & -6 & 3 & 2 \end{bmatrix}, \quad \bar{A} = \begin{bmatrix} 3 & -2 & 5 & 4 & 2 \\ 6 & -4 & 4 & 3 & 3 \\ 9 & -6 & 3 & 2 & 4 \end{bmatrix}.$$

Determinemos rangos de estas matrices, aplicando las transformaciones elementales:

a) las columnas 1 y 2 son proporcionales, suprimimos una de ellas (la segunda);

b) multipliquemos la primera columna por 1/3;

c) multipliquemos la primera columna sucesivamente por (—5) y por (—4) y adicionemos lo obtenido a las columnas 2 y 3, respectivamente;

d) en la matriz obtenida dos columnas (2 y 3) son proporcionales; eliminemos una de ellas (la tercera) y multipliquemos la segunda por —1/6;

e) multipliquemos la primera fila sucesivamente por (—2) y por (—3) y adicionemos el resultado a las filas 2 y 3, respectivamente;

f) multipliquemos la segunda fila por (—2), adicionemos el resultado a la tercera y eliminemos la fila nula obtenida.

Tenemos

$$A=\begin{bmatrix} 3 & -2 & 5 & 4 \\ 6 & -4 & 4 & 3 \\ 9 & -6 & 3 & 2 \end{bmatrix} \underset{\sim}{a)} \begin{bmatrix} 3 & 5 & 4 \\ 6 & 4 & 3 \\ 9 & 3 & 2 \end{bmatrix} \underset{\sim}{b)} \begin{bmatrix} 1 & 5 & 4 \\ 2 & 4 & 3 \\ 3 & 3 & 2 \end{bmatrix}$$

$$\underset{\sim}{c)} \begin{bmatrix} 1 & 0 & 0 \\ 2 & -6 & -5 \\ 3 & -12 & -10 \end{bmatrix} \underset{\sim}{d)} \begin{bmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 2 \end{bmatrix} \underset{\sim}{e)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 2 \end{bmatrix} \underset{\sim}{f)} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Es evidente que el rango de la última matriz es igual a 2: $r(A) = 2$.

Transformemos de un modo análogo la matriz $\overline{A}$:

$$\overline{A}=\begin{bmatrix} 3 & -2 & 5 & 4 & 2 \\ 6 & -4 & 4 & 3 & 3 \\ 9 & -6 & 3 & 2 & 4 \end{bmatrix} \sim \begin{bmatrix} 3 & 5 & 4 & 2 \\ 6 & 4 & 3 & 3 \\ 9 & 3 & 2 & 4 \end{bmatrix} \sim \begin{bmatrix} 1 & 5 & 4 & 2 \\ 2 & 4 & 3 & 3 \\ 3 & 3 & 2 & 4 \end{bmatrix} \sim$$

$$\sim \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & -6 & -5 & -1 \\ 3 & -12 & -10 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 2 & -1 \\ 3 & -2 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & -1 \\ 0 & -2 \end{bmatrix} \sim$$

$$\sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \sim \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Por lo tanto, $r(\overline{A}) = 2$.

Ahora bien, $r(A) = r(\overline{A}) = 2$, o sea, el sistema dado es compatible.

Puesto que el rango del sistema es igual a 2, el orden máximo del menor distinto del cero es igual a 2 y el sistema tiene dos incógnitas básicas. Determinemos cualquier menor de segundo orden distinto del cero. Tal, por ejemplo, es el menor $M_2 = \begin{vmatrix} 5 & 4 \\ 4 & 3 \end{vmatrix} \neq 0$ formado por los coeficientes de las incógnitas x_3 y x_4. Por consiguiente, las incógnitas x_3 y x_4 pueden considerarse básicas, y las incógnitas x_1 y x_2, independientes.

El sistema (*) es equivalente al siguiente:

$$\begin{cases} 3x_1 - 2x_2 + 5x_3 + 4x_4 = 2, \\ 6x_1 - 4x_2 + 4x_3 + 3x_4 = 3. \end{cases} \qquad (**)$$

Transponemos las incógnitas independientes al segundo miembro:

$$\begin{cases} 5x_3 + 4x_4 = 2 - 3x_1 + 2x_2, \\ 4x_3 + 3x_4 = 3 - 6x_1 + 4x_2. \end{cases} \qquad (***)$$

Resolvemos el sistema (3) con ayuda de las fórmulas de Cramer:

$$d=\begin{vmatrix} 5 & 4 \\ 4 & 3 \end{vmatrix}=15-16=-1;$$

$$d_3=\begin{vmatrix} 2-3x_1+2x_2 & 4 \\ 3-6x_1+4x_2 & 3 \end{vmatrix}=3(2-3x_1+2x_2)-4(3-6x_1+4x_2)=$$
$$=-6+15x_1-10x_2;$$

$$d_4=\begin{vmatrix} 5 & 2-3x_1+2x_2 \\ 4 & 3-6x_1+4x_2 \end{vmatrix}=5(3-6x_1+4x_2)-$$
$$-4(2-3x_1+2x_2)=7-18x_1+12x_2;$$

$$x_3=d_3/d=6-15x_1+10x_2,\quad x_4=d_4/d=-7+18x_1-12x_2.$$

La solución obtenida en la cual las incógnitas básicas x_3 y x_4 están expresadas por medio de las x_1 y x_2 independientes es la solución general del sistema (*). Sustituyendo en ésta los valores arbitrarios para las incógnitas independientes, obtenemos distintas soluciones particulares. Por ejemplo, si $x_1 = 0$, $x_2 = 0$, entonces $x_3 = 6$, $x_4 = -7$; si $x_1 = 1$, $x_2 = 2$, entonces $x_3 = 11$, $x_4 = 13$, etc. Las colecciones de los números (0, 0, 6, —7), (1, 2, 11, 13), etc. son soluciones particulares del sistema (*). ▲

§ 3.5. Sistema homogéneo de ecuaciones lineales

Supongamos que se dé un sistema homogéneo de m ecuaciones lineales con n incógnitas

$$\begin{cases} a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n=0, \\ a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n=0, \\ \dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x_n=0. \end{cases}$$

Puesto que la adición de la columna de ceros no cambia el rango de la matriz del sistema, entonces, en virtud del teorema de Kronecker—Capelli, este sistema es siempre compatible y tiene, por lo menos, la solución nula ($x_1 = x_2 = \ldots = x_n = 0$). Si el determinante del sistema (1) es distinto del cero y la cantidad de ecuaciones del sistema es igual a la cantidad de incógnitas, entonces según el teorema de Cramer la solución nula es la única.

En el caso en que el rango de la matriz del sistema (1) es menor que el número de incógnitas, o sea, $r(A) < n$, el sistema dado tendrá, además de la solución nula, también soluciones no nulas. Para encontrar estas soluciones en el sistema (1) separamos r ecuaciones linealmente independientes y suprimimos las demás. En el primer miembro de las ecuaciones separadas dejamos r incógnitas básicas y transponemos las demás $n - r$ incógnitas independientes al segundo miembro. Entonces llegamos al sistema

$$\left\{\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1r}x_r = \\ \qquad = -a_{1,r+1}x_{r+1} - a_{1,r+2}x_{r+2} - \ldots - a_{1n}x_n, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2r}x_r = \\ \qquad = -a_{2,r+1}x_{r+1} - a_{2,r+2}x_{r+2} - \ldots - a_{2n}x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{r1}x_1 + a_{r2}x_2 + \ldots a_{2r}x_r = \\ \qquad = -a_{r,r+1}x_{r+1} - a_{r,r+2}x_{r+2} - \ldots - a_{rn}x_n, \end{array}\right. \qquad (2)$$

resolviendo el cual, expresamos según las formulas de Cramer, r incógnitas básicas $x_1, x_2, \ldots, x_r$ por medio de $n - r$ incógnitas independientes $x_{r+1}, x_{r+2}, \ldots, x_n$.

El sistema (1) tiene un conjunto infinito de soluciones. Entre este conjunto hay soluciones que son linealmente independientes entre sí.

Se llama *sitema fundamental de soluciones* a $n - r$ soluciones linealmente independientes de un sistema homogéneo de ecuaciones.

Ejemplo. Se da un sistema homogéneo de ecuaciones

$$\begin{cases} 2x_1 - 4x_2 + 5x_3 + 3x_4 = 0, \\ 3x_1 - 6x_2 + 4x_3 + 2x_4 = 0, \\ 4x_1 - 8x_2 + 17x_3 + 11x_4 = 0 \end{cases}$$

Hallar su solución general y el sistema fundamental de soluciones.

△ 1) Aquí la cantidad de incógnitas $n = 4$, la cantidad de ecuaciones $m = 3$. Calculemos el rango de la matriz del sistema, utilizando las transformaciones lineales:

a) suprimimos la segunda columna, ya que es proporcional a la primera;

b) multipliquemos primero la tercera columna por (—2) y adicionemos lo obtenido a la segunda; luego multipliquemos la tercera columna por (—3) y sumemos el resultado con la primera multiplicada por 2;

c) suprimimos la primera columna, ya que es proporcional a la segunda;

d) multipliquemos por 3 la primera columna y adicionémosla a la segunda;

e) multipliquemos la primera fila por 5 y adicionémosla a la cuarta;

f) suprimimos la tercera fila y dividimos la primera fila por (—1) y la segunda por 2.

Tenemos

$$A \sim \begin{bmatrix} 2 & -4 & 5 & 3 \\ 3 & -6 & 4 & 2 \\ 4 & -8 & 17 & 11 \end{bmatrix} \underset{\sim}{a)} \begin{bmatrix} 2 & 5 & 3 \\ 3 & 4 & 2 \\ 4 & 17 & 11 \end{bmatrix} \underset{\sim}{b)} \begin{bmatrix} -5 & -1 & 3 \\ 0 & 0 & 2 \\ -25 & -5 & 11 \end{bmatrix} \underset{\sim}{c)} \sim$$

$$\sim\begin{bmatrix}-1 & 3\\ 0 & 2\\ -5 & 11\end{bmatrix} \underset{\sim}{d)} \begin{bmatrix}-1 & 0\\ 0 & 2\\ -5 & -4\end{bmatrix} \underset{\sim}{e)} \begin{bmatrix}-1 & 0\\ 0 & 2\\ 0 & -4\end{bmatrix} \underset{\sim}{f)} \begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}.$$

Puesto que $r(A) = 2$, o sea, $r \leqslant \min\{m, n\}$, el sistema dado tiene el sistema fundamental de soluciones cuya cantidad es $n - 2 = 4 - 2 = 2$.

2) Vamos a buscar la solución general del sistema. Determinemos el menor básico, o sea, el menor de segundo orden distinto del cero. De tal menor sirve, por ejemplo, el menor compuesto por los coeficientes de las x_3 y x_4 en las ecuaciones primera y segunda del sistema: $M_2 = \begin{vmatrix}5 & 3\\ 4 & 2\end{vmatrix} = 2 \neq 0$. Dejando las incógnitas básicas x_3 y x_4 en el primer miembro y transponiendo las incógnitas independientes x_1 y x_2 al segundo miembro, llegamos al sistema

$$\begin{cases} 5x_3 + 3x_4 = -2x_1 + 4x_2,\\ 4x_3 + 2x_4 = -3x_1 + 6x_2.\end{cases}$$

Su solución, hallada con ayuda de las fórmulas de Cramer, tiene la forma

$$x_3 = -2{,}5x_1 + 5x_2,$$
$$x_4 = 3{,}5x_1 - 7x_2.$$

3) Para obtener el sistema fundamental de soluciones hace falta hallar cualesquiera dos soluciones linealmente independientes del sistema dado (ya que $n - r = 2$). Suponiendo primero $x_1 = 1, x_2 = 0$, tenemos $x_3 = -2{,}5$, $x_4 = 3{,}5$; suponiendo luego $x_1 = 0$, $x_2 = 1$, obtenemos $x_3 = 5$, $x_4 = -7$. Ahora bien, el sistema fundamental de soluciones tiene la forma

$$R_1 = \begin{pmatrix}0\\ 1\\ -2{,}5\\ 3{,}5\end{pmatrix}; \quad R_2 = \begin{pmatrix}0\\ 1\\ 5\\ -7\end{pmatrix}$$

y la solución general $R = c_1R_1 + c_2R_2$, donde c_1 y c_2 son los números arbitrarios.

Asignando a c_1 y c_2 distintos valores, se puede obtener toda solución del sistema dado.

Sea, por ejemplo, $x_1 = 1$, $x_2 = 2$, entonces $x_3 = -2{,}5\cdot 1 + 5\cdot 2 = 7{,}5$, $x_4 = 3{,}5\cdot 1 - 7\cdot 2 = -10{,}5$. La solución particular obtenida

$$R = \begin{pmatrix}1\\ 2\\ 7{,}5\\ -10{,}5\end{pmatrix}$$

es una combinación lineal de soluciones que forman el sistema fundamental para $c_1 = 1$, $c_2 = 2$; $R = R_1 + 2R_2$.▲

§ 3.6. Resolución de un sistema de ecuaciones lineales con ayuda del método de eliminación sucesiva de las incógnitas (por el método de Gauss)

El método más difundido de resolución de los sistemas de ecuaciones lineales es el de **eliminación sucesiva de las incógnitas (método de Gauss).**

Consideremos este método citando un ejemplo del sistema de cuatro ecuaciones con cuatro incógnitas. Supongamos que se da el sistema

$$\begin{cases} a_{11}x_1+a_{12}x_2+a_{13}x_3+a_{14}x_4=a_{15}, \\ a_{21}x_1+a_{22}x_2+a_{23}x_3+a_{24}x_4=a_{25}, \\ a_{31}x_1+a_{32}x_2+a_{33}x_3+a_{34}x_4=a_{35}, \\ a_{41}x_1+a_{42}x_2+a_{43}x_3+a_{44}x_4=a_{45}, \end{cases} \tag{1}$$

Vamos a eliminar la incógnita x_1 de todas las ecuaciones del sistema (1), menos de la primera. Llamemos x_1 *incógnita guía* y coeficiente a_{11} *el coeficiente guía*. Dividiendo la primera ecuación por a_{11} (esto es posible si $a_{11} \neq 0$), obtenemos

$$x_1+\frac{a_{12}}{a_{11}}x_2+\frac{a_{13}}{a_{11}}x_3+\frac{a_{14}}{a_{11}}x_4=\frac{a_{15}}{a_{11}}.$$

Designemos $a_{12}/a_{11}=b_{12}$, $a_{13}/a_{11}=b_{13}$, $a_{14}/a_{11}=b_{14}$, $a_{15}/a_{11}=b_{15}$ y en general $b_{1j}=a_{1j}/a_{11}$ $(j>1)$. Entonces la ecuación en cuestión tendrá la forma

$$x_1+b_{12}x_2+b_{13}x_3+b_{14}x_4=b_{15} \tag{2}$$

o bien

$$x_1=b_{15}-b_{12}x_2-b_{13}x_3-b_{14}x_4.$$

Para eliminar la incógnita x_1 de las ecuaciones del sistema (1) realicemos las transformaciones siguientes.

1) De la segunda ecuación del sistema (1) sustrayamos la ecuación (2) multiplicada por a_{21}:

$$\begin{array}{r} a_{21}x_1+a_{22}x_2+a_{23}x_3+a_{24}x_4=a_{25} \\ -a_{21}x_1-a_{21}b_{12}x_2-a_{21}b_{13}x_3-a_{21}b_{14}x_4=-a_{21}b_{15} \\ \hline (a_{22}-a_{21}b_{12})\,x_2+(a_{23}-a_{21}b_{13})\,x_3+(a_{24}-a_{21}b_{14})\,x_4=(a_{25}-a_{21}b_{15}) \end{array}.$$

Designemos

$$a_{22}-a_{21}b_{12}=a_{22}^{(1)};\ a_{23}-a_{21}b_{13}=a_{23}^{(1)},$$
$$a_{24}-a_{21}b_{14}=a_{24}^{(1)};\ a_{25}-a_{21}b_{15}=a_{25}^{(1)}$$

y reescribiremos la ecuación obtenida en la forma

$$a_{22}^{(1)}x_2+a_{23}^{(1)}x_3+a_{24}^{(1)}x_4=a_{25}^{(1)}.$$

2) De la tercera ecuación del sistema (1) sustrayamos la ecuación (2) multiplicada por a_{31}:

$$\begin{array}{r} a_{31}x_1+a_{32}x_2+a_{33}x_3+a_{34}x_4=a_{35} \\ -a_{31}x_1-a_{31}b_{12}x_2-a_{31}b_{13}x_3-a_{31}b_{14}x_4=-a_{31}b_{15} \\ \hline (a_{32}-a_{31}b_{12})\,x_2+(a_{33}-a_{31}b_{13})\,x_3+(a_{34}-a_{31}b_{14})\,x_4=a_{35}-a_{31}b_{15} \end{array}.$$

Designando $a_{32} - a_{31}b_{12} = a_{32}^{(1)}$; $a_{33} - a_{31}b_{13} = a_{33}^{(1)}$, etc., escribamos la ecuación obtenida en la forma

$$a_{32}^{(1)}x_2 + a_{33}^{(1)}x_3 + a_{34}^{(1)}x_4 = a_{35}^{(1)}.$$

3) De la cuarta ecuación del sistema (1) sustrayamos la ecuación (2) multiplicada por a_{41}. Utilizando las designaciones análogas, obtenemos la ecuación siguiente:

$$a_{42}^{(1)}x_2 + a_{43}^{(1)}x_3 + a_{44}^{(1)}x_4 = a_{45}^{(1)}.$$

Como resultado de las transformaciones elementales realizadas tenemos un sistema de tres ecuaciones con tres incógnitas

$$\left\{\begin{array}{l} a_{22}^{(1)}x_2 + a_{23}^{(1)}x_3 + a_{24}^{(1)}x_4 = a_{25}^{(1)}, \\ a_{32}^{(1)}x_2 + a_{33}^{(1)}x_3 + a_{34}^{(1)}x_4 = a_{35}^{(1)}, \\ a_{42}^{(1)}x_2 + a_{43}^{(1)}x_3 + a_{44}^{(1)}x_4 = a_{45}^{(1)}, \end{array}\right. \tag{1'}$$

donde los coeficientes a_{ij} $(i, j \geqslant 2)$ se calculan por la fórmula $a_{ij}^{(1)} = a_{ij} - a_{i1}b_{1j}$ (por ejemplo, $a_{23}^{(1)} = a_{23} - a_{21}b_{12}$).

Dividiendo luego los coeficientes de la primera ecuación del sistema (1′) por el coeficiente guía $a_{22}^{(1)} \neq 0$, obtenemos la primera ecuación en la forma

$$x_2 + \frac{a_{23}^{(1)}}{a_{22}^{(1)}}x_3 + \frac{a_{24}^{(1)}}{a_{22}^{(1)}}x_4 = \frac{a_{25}^{(1)}}{a_{22}^{(1)}}.$$

Designemos

$$a_{23}^{(1)}/a_{22}^{(1)} = b_{23}^{(1)};\ a_{24}^{(1)}/a_{22}^{(1)} = b_{24}^{(1)};\ a_{25}^{(1)}/a_{22}^{(1)} = b_{25}^{(1)}$$

y en general $a_{2j}^{(1)}/a_{22}^{(1)} = b_{2j}^{(1)}$ $(j > 2)$. Entonces la primera ecuación del sistema (1′) tendrá la forma

$$x_2 + b_{23}^{(1)}x_3 + b_{24}^{(1)}x_4 = b_{25}^{(1)} \tag{2'}$$

o bien

$$x_2 = b_{25}^{(1)} - b_{23}^{(1)}x_3 - b_{24}^{(1)}x_4.$$

Eliminando ahora x_2 de todas las ecuaciones del sistema (1′), salvo la primera, haciendo uso del mismo método que hemos utilizado para eliminar x_1, llegamos al siguiente sistema de dos ecuaciones con dos incógnitas:

$$\left\{\begin{array}{r} a_{33}^{(2)}x_3 + a_{34}^{(2)}x_4 = a_{35}^{(2)}, \\ a_{43}^{(2)}x_3 + a_{44}^{(1)}x_4 = a_{45}^{(1)}, \end{array}\right. \tag{1''}$$

donde $a_{ij}^{(2)} = a_{ij}^{(1)} - a_{i2}^{(1)}\, b_{2j}^{(1)}$ $(i, j \geqslant 3)$. Por ejemplo, $a_{34}^{(2)} = a_{34}^{(1)} - a_{32}^{(1)}b_{24}^{(1)}$. Dividiendo los coeficientes de la primera ecuación del sistema (1″) por el coeficiente guía $a_{33}^{(2)} \neq 0$, obtenemos

$$x_3 + b_{34}^{(1)}x_4 = b_{35}^{(2)}, \tag{2''}$$

donde $b_{3j}^{(2)} = a_{3j}^{(2)}/a_{33}^{(2)}$ $(j > 3)$, o sea,

$$x_3 = b_{35}^{(2)} - b_{34}^{(2)}x_4.$$

Eliminando ahora x_3 del sistema (1″), al utilizar el método análogo, hallamos

$$a_{44}^{(3)}x_4 = a_{45}^{(3)}, \tag{1‴}$$

donde $a_{ij}^{(3)} = a_{ij}^{(2)} - a_{i3}^{(2)}b_{3j}^{(2)}$ $(i, j \geqslant 4)$. De aquí

$$x_4 = a_{45}^{(3)}/a_{44}^{(3)}. \tag{2‴}$$

Las demás incógnitas del sistema se obtienen sucesivamente de las ecuaciones (2″), (2′) y (2):

$$x_3 = b_{35}^{(2)} - b_{34}^{(2)}x_4,$$
$$x_2 = b_{25}^{(1)} - b_{24}^{(1)}x_4 - b_{23}^{(1)}x_3,$$
$$x_1 = b_{15} - b_{14}x_4 - b_{13}x_3 - b_{12}x_2.$$

Ahora bien, el proceso de resolución de un sistema de ecuaciones lineales con ayuda del método de Gauss se reduce a la construcción de un sistema equivalente de ecuaciones (2), (2′), (2″), (2‴). El método de Gauss es aplicable a condición de que todos los coeficientes guía sean distintos del cero.

Para la comodidad los cálculos se llevan a cabo según el esquema llamado **esquema de división única.** El cálculo de los elementos b_{ij} se llama *paso directo*, y el cálculo de los valores de las incógnitas se denomina *paso invertido*, ya que primeramente se determina el valor de la última incógnita.

El esquema de división única (esquema de Gauss) se hace del modo siguiente.

En la parte I del esquema (véase la tabla 3.1) se escriben los coeficientes de las incógnitas (en las columnas de las incógnitas respectivas), los términos independientes y para cada fila «las sumas de control» (columna Σ_2) iguales a la suma de los elementos a_{ij} en la fila dada (aquí $i = 1, 2, 3, 4$; $j = 1, 2, 3, 4, 5$); la última fila de la parte I, compuesta por 1 y por los elementos b_{ij}, se obtiene dividiendo la primera fila de la parte por el coeficiente guía a_{11}.

Los elementos de la parte II del esquema son iguales a los elementos correspondientes de la parte 1 menos el producto $a_{i1}b_{1j}$ $(i, j \geqslant 2)$; por ejemplo, $a_{23}^{(1)} = a_{23} - a_{21}b_{13}$. La última fila de la parte II, compuesta por 1 y por los elementos $b_{2j}^{(1)}$, se obtiene dividiendo la primera fila de la parte por el coeficiente guía $a_{22}^{(1)}$.

Análogamente se calculan los elementos de las partes III y IV del esquema. Las partes I, II, III y IV, que se terminan con el cálculo de los elementos $b_{ij}^{(i-1)}$ $(i = 1, 2, 3, 4;\ j = 2, 3, 4, 5)$, constituyen el **paso directo** de los cálculos del esquema.

El paso invertido comienza con el cálculo de la última incógnita del sistema de ecuaciones lineales x_4 y termina con el cálculo de la primera incógnita x_1. Durante el paso invertido se utilizan sólo las filas del paso directo que contienen las unidades y los elementos correspondientes b_{ij} (llamemos estas filas «marcadas»).

El elemento $b_{45}^{(3)}$ de la última fila «marcada» y de la columna de los términos independientes da el valor x_4. Luego, las demás incóg-

x_1	x_2	x_3	x_4	Términos independientes
$\boxed{a_{11}}$ a_{21} a_{31} a_{41}	a_{12} a_{22} a_{32} a_{42}	a_{13} a_{23} a_{33} a_{43}	a_{14} a_{24} a_{34} a_{44}	a_{15} a_{25} a_{35} a_{45}
$1=\frac{a_{11}}{a_{11}}$	$b_{12}=\frac{a_{12}}{a_{11}}$	$b_{13}=\frac{a_{13}}{a_{11}}$	$b_{14}=\frac{a_{14}}{a_{11}}$	$b_{15}=\frac{a_{15}}{a_{11}}$
	$\boxed{a_{22}^{(1)}}$ $a_{32}^{(1)}$ $a_{42}^{(1)}$	$a_{23}^{(1)}$ $a_{33}^{(1)}$ $a_{43}^{(1)}$	$a_{24}^{(1)}$ $a_{34}^{(1)}$ $a_{44}^{(1)}$	$a_{25}^{(1)}=a_{25}-a_{21}b_{15}$ $a_{35}^{(1)}=a_{35}-a_{31}b_{15}$ $a_{45}^{(1)}=a_{45}-a_{41}b_{15}$
	$1=\frac{a_{22}^{(1)}}{a_{22}^{(1)}}$	$b_{23}^{(1)}=\frac{a_{23}^{(1)}}{a_{22}^{(1)}}$	$b_{24}^{(1)}=\frac{a_{24}^{(1)}}{a_{22}^{(1)}}$	$b_{25}^{(1)}=\frac{a_{25}^{(1)}}{a_{22}^{(1)}}$
		$\boxed{a_{33}^{(2)}}$ $a_{43}^{(2)}$	$a_{34}^{(2)}$ $a_{44}^{(2)}$	$a_{35}^{(2)}=a_{35}^{(1)}-a_{32}^{(1)}b_{25}^{(1)}$ $a_{45}^{(2)}=a_{45}^{(1)}-a_{42}^{(1)}b_{25}^{(1)}$
		$1=\frac{a_{33}^{(2)}}{a_{33}^{(2)}}$	$b_{34}^{(2)}=\frac{a_{34}^{(2)}}{a_{33}^{(2)}}$	$b_{35}^{(2)}=\frac{a_{35}^{(2)}}{a_{33}^{(2)}}$
			$\boxed{a_{44}^{(3)}}$	$a_{45}^{(3)}=a_{45}^{(2)}-a_{43}^{(2)}b_{35}^{(2)}$
			$1=\frac{a_{44}^{(3)}}{a_{44}^{(3)}}$	$b_{45}^{(3)}=\frac{a_{45}^{(3)}}{a_{44}^{(3)}}$
1	1	1	1	$x_4=b_{45}^{(3)}$ $x_3=b_{35}^{(2)}-b_{34}^{(2)}x_4$ $x_2=b_{25}^{(1)}-b_{24}^{(1)}x_4-b_{23}^{(1)}x_3$ $x_1=b_{15}-b_{14}x_4-b_{13}x_3-$ $-b_{12}x_2$

Tabla 3.1

Σ_1	Σ_2	Partes del esquema	
	$a_{16}=a_{11}+a_{12}+a_{13}+a_{14}+a_{15}$ $a_{26}=a_{21}+a_{22}+a_{23}+a_{24}+a_{25}$ $a_{36}=a_{31}+a_{32}+a_{33}+a_{34}+a_{35}$ $a_{46}=a_{41}+a_{42}+a_{43}+a_{44}+a_{45}$	I	Paso directo
$b_{16}=\dfrac{a_{16}}{a_{11}}$	$b_{16}=1+b_{12}+b_{13}+b_{14}+b_{15}$		
$a_{26}^{(1)}=a_{26}-a_{21}b_{16}$ $a_{36}^{(1)}=a_{36}-a_{31}b_{16}$ $a_{46}^{(1)}=a_{46}-a_{41}b_{16}$	$a_{26}^{(1)}=a_{22}^{(1)}+a_{23}^{(1)}+a_{24}^{(1)}+a_{25}^{(1)}$ $a_{36}^{(1)}=a_{32}^{(1)}+a_{33}^{(1)}+a_{34}^{(1)}+a_{35}^{(1)}$ $a_{46}^{(1)}=a_{42}^{(1)}+a_{43}^{(1)}+a_{44}^{(1)}+a_{45}^{(1)}$	II	
$b_{26}^{(1)}=\dfrac{a_{26}^{(1)}}{a_{22}^{(1)}}$	$b_{26}^{(1)}=1+b_{23}^{(1)}+b_{24}^{(1)}+b_{25}^{(1)}$		
$a_{36}^{(2)}=a_{36}^{(1)}-a_{32}^{(1)}b_{26}^{(1)}$ $a_{46}^{(2)}=a_{46}^{(1)}-a_{42}^{(1)}b_{26}^{(1)}$	$a_{36}^{(2)}=a_{33}^{(2)}+a_{34}^{(2)}+a_{35}^{(2)}$ $a_{46}^{(2)}=a_{43}^{(2)}+a_{44}^{(2)}+a_{45}^{(2)}$	III	
$b_{36}^{(2)}=\dfrac{a_{36}^{(2)}}{a_{33}^{(2)}}$	$b_{36}^{(2)}=1+b_{34}^{(2)}+b_{35}^{(2)}$		
$a_{46}^{(3)}=a_{46}^{(2)}-a_{43}^{(2)}b_{36}^{(2)}$	$a_{46}^{(3)}=a_{44}^{(3)}+a_{45}^{(3)}$	IV	
$b_{46}^{(3)}=\dfrac{a_{46}^{(3)}}{a_{44}^{(3)}}$	$b_{46}^{(3)}=1+b_{45}^{(3)}$		
$\bar{x}_4=b_{46}^{(3)}$ $\bar{x}_3=b_{36}^{(2)}-b_{34}^{(2)}\bar{x}_4$ $\bar{x}_2=b_{26}^{(1)}-b_{24}^{(1)}\bar{x}_4-b_{23}^{(1)}\bar{x}_3$ $\bar{x}_1=b_{16}-b_{14}\bar{x}_4-b_{13}\bar{x}_3-$ $-b_{12}\bar{x}_2$	$\bar{x}_4=1+x_4$ $\bar{x}_3=1+x_3$ $\bar{x}_2=1+x_2$ $\bar{x}_1=1+x_1$	V	Paso directo

nitas x_3, x_2 y x_1 se hallan sustrayendo del término independiente de la fila «marcada» la suma de los productos de sus coeficientes por los valores correspondientes de las incógnitas antes encontradas; por ejemplo, $x_3 = b_{35}^{(2)} - b_{34}^{(2)} x_4$.

Los valores de las incógnitas se ponen sucesivamente en la parte V. Las unidades allí colocadas ayudan a encontrar para x_j los coeficientes respectivos en las filas «marcadas».

Para controlar los cálculos se utilizan las así llamadas sumas de control:

$$a_{i6} = \sum_{j=1}^{5} a_{ij} \quad (i = 1, 2, 3, 4)$$

y

$$b_{i6} = \sum_{j=1}^{5} b_{ij} + 1 \quad (i = 1, 2, 3, 4)$$

colocadas en la columna Σ_2.

En la columna Σ_1 de las partes II, III, y IV en cada fila se hacen con las sumas de control las mismas operaciones que con los demás elementos de esta fila. Si no hay errores en los cálculos los elementos de las columnas Σ_1 y Σ_2 son iguales. Ahora bien, se controla el paso directo del esquema.

Para controlar el paso invertido $\bar{x}_4$ se halla en la última fila «marcada» de la columna Σ_1, o sea, $\bar{x}_4 = b_{46}^{(3)}$ y las demás incógnitas de esta columna $\bar{x}_j$ ($j = 3, 2, 1$) se calculan en las mismas filas y con ayuda de las mismas fórmulas que las incógnitas x_j, con una sola particularidad de que en las fórmulas se sustituyen las $\bar{x}_j$ correspondientes. En resumidas cuentas los números $\bar{x}_j$ deben coincidir con los números $x_j + 1$ de la columna Σ_2.

Ejemplo 1. Con ayuda del esquema de división única resolver el sistema de ecuaciones

$$\begin{cases} 2x_1 + 2x_2 - x_3 + x_4 = 4, \\ 4x_1 + 3x_2 - x_3 + 2x_4 = 6, \\ 8x_1 + 5x_2 - 3x_3 + 4x_4 = 12, \\ 3x_1 + 3x_2 - 2x_3 + 2x_4 = 6. \end{cases}$$

△ En la parte I de la tabla 3.2 escribimos la matriz del sistema, sus términos independientes y las sumas de control. Luego calculamos la fila «marcada» de esta parte dividiendo la primera fila por $a_{11} = 2$. Por ejemplo, $b_{12} = a_{12}/a_{11} = 2/2 = 1$.

Calculamos los elementos de la parte II con ayuda de la regla siguiente: cada elemento de esta parte es igual al elemento correspondiente de la parte I menos el producto del primer elemento de su fila por el elemento de la fila «marcada» en su columna. Escribimos el resultado obtenido en el lugar respectivo de la parte II. Por ejemplo:

$$a_{23}^{(1)} = a_{23} - a_{21}b_{13} = -1 - 4\,(-0{,}5) = 1,$$
$$a_{33}^{(1)} = a_{33} - a_{31}b_{13} = -3 - 8\,(-0{,}5) = 1.$$

Tabla 3.2

x_1	x_2	x_3	x_4	Términos independientes	Σ_1	Σ_2	Partes del esquema
$\boxed{2}$	2	−1	1	4	8		I
4	3	−1	2	6	14		
8	5	−3	4	12	26		
3	3	−2	2	6	12		
1	1	−0,5	0,5	2	4	4	
	$\boxed{-1}$	1	0	−2	−2	−2	II
	−3	1	0	−4	−6	−6	
	0	−0,5	0,5	0	0	0	
	1	−1	0	2	2	2	
		$\boxed{-2}$	0	2	0	0	III
		−0,5	0,5	0	0	0	
		1	0	−1	0	0	
			$\boxed{0,5}$	−0,5	0	0	IV
			1	−1	0	0	
			1	$x_4 = -1$	$\bar{x}_4 = 0$	0	V
		1		$x_3 = -1$	$\bar{x}_3 = 0$	0	
	1			$x_2 = 1$	$\bar{x}_2 = 2$	2	
1				$x_1 = 1$	$\bar{x}_1 = 2$	2	

Obtenemos los elementos de la fila «marcada» de la parte II dividiendo su primera fila por el coeficiente guía $a_{22}^{(1)} = -1$. Por ejemplo,

$$b_{23}^{(1)} = a_{23}^{(1)}/a_{22}^{(1)} = 1/(-1) = -1.$$

Análogamente se calculan los elementos de las partes III y IV. Por ejemplo:

$$a_{44}^{(2)} = a_{44}^{(1)} - a_{42}^{(1)}b_{24}^{(1)} = 2 - 3\cdot 0,5 = 0,5,$$
$$a_{45}^{(3)} = a_{45}^{(2)} - a_{43}^{(2)}b_{35}^{(2)} = 0 - (-0,5\,(-1) = -0,5.$$

Para calcular los elementos de la parte V, o sea, para hallar las incógnitas utilizamos las filas «marcadas» comenzando con la última.

La incógnita x_4 no es sino el término independiente de la última fila «marcada»: $x_4 = b_{45}^{(3)} = 1$ y las demás incógnitas x_3, x_2 y x_1 se obtienen sucesivamente sustrayendo de los términos independientes de las filas «marcadas» la suma de los productos de los coeficientes correspondientes $b_{ij}^{(i-1)}$ por los valores antes hallados de las incógnitas.

El control se lleva a cabo con ayuda de las columnas Σ_1 y Σ_2. Con la columna Σ_1 se realizan las mismas operaciones que con las demás columnas (véanse las tablas 3.1 y 3.2) y en resumen la suma de los elementos de cada fila del esquema (sin la columna Σ_1) debe ser igual al elemento de esta fila tomado de la columna Σ_2. Los números $\bar{x}_j$ de la columna Σ_1 han de ser iguales a los números $1 + x_j$ de la columna Σ_2.

Como resultado obtenemos $x_1 = 1$, $x_2 = 1$, $x_3 = -1$, $x_4 = -1$. ▲

Ejemplo 2. Con ayuda del esquema de división única resolver el siguiente sistema con precisión hasta 0,0001:

$$\begin{cases} 0{,}63x_1 + 1{,}00x_2 + 0{,}71x_3 + 0{,}34x_4 = 2{,}08, \\ 1{,}17x_1 + 0{,}18x_2 - 0{,}65x_3 + 0{,}71x_4 = 0{,}17, \\ 2{,}71x_1 - 0{,}75x_2 + 1{,}17x_3 - 2{,}35x_4 = 1{,}28, \\ 3{,}58x_1 + 0{,}28x_2 - 3{,}45x_3 - 1{,}18x_4 = 0{,}05. \end{cases}$$

△ La resolución del sistema se da en la tabla 3.3. La respuesta final: $x_1 = 0{,}4026$, $x_2 = 1{,}5016$, $x_3 = 0{,}5862$, $x_4 = -0{,}2678$. ▲

Si los valores aproximados de las incógnitas, obtenidos con ayuda del esquema de Gauss, son suficientemente exactos, o sea, sus correcciones son pequeñas en valor absoluto, se puede no realizar una precisión.

En caso de necesidad de precisar los valores aproximados de las incógnitas se procede del modo siguiente:

1) para cada ecuación del sistema se calculan las *discrepancias*, o sea, las diferencias entre los miembros segundo y primero del sistema las cuales se obtienen después de sustituir en las ecuaciones los valores aproximados de las incógnitas; si los valores aproximados de las incógnitas se designan con $x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}$, las discrepancias con $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n$ y los términos independientes con $b_1, b_2, \ldots, b_n$, entonces

$$\varepsilon_1 = b_1 - \sum_{j=1}^{n} a_{1j}x_j^{(0)},$$
$$\varepsilon_2 = b_2 - \sum_{j=1}^{n} a_{2j}x_j^{(0)},$$
$$\cdots\cdots\cdots\cdots$$
$$\varepsilon_n = b_n - \sum_{j=1}^{n} a_{nj}x_j^{(0)};$$

2) se escriben las discrepancias ε_i ($i = 1, 2, \ldots, n$) en una columna especial del esquema de Gauss (ε) y se realizan con ellas las mismas operaciones que con otras columnas del esquema;

3) considerando la columna ε como columna de los términos independientes, se obtienen las correcciones δ_i de las incógnitas;

Tabla 3.3

x_1	x_2	x_3	x_4	Términos independientes	Σ_1	Σ_2
$\boxed{0,63}$	1,00	0,71	0,34	2,08	4,76	
1,17	0,18	−0,65	0,71	0,17	1,58	
2,71	−0,75	1,17	−2,35	1,28	2,06	
3,58	0,21	−3,45	−1,18	0,05	−0,79	
1	1,587	1,127	0,539	3,302	7,555	
	$\boxed{-1,6768}$	−1,9686	0,0794	−3,6933	−7,2593	−7,2593
	−5,0508	−1,8842	−3,8107	−7,6684	−18,4141	−18,4141
	−5,4715	−7,4847	−3,1096	−11,7712	−27,8370	−27,8370
	1	1,17402	−0,04735	2,20259	4,32926	4,32926
		$\boxed{4,04554}$	−4,04986	3,45644	3,45212	3,45212
		−1,06105	−3,36868	0,28027	−4,14946	−4,14946
		1	−1,00106	0,85438	0,85332	0,85332
			$\boxed{-4,43085}$	1,18681	−3,24404	−3,24404
			1	−0,26785	0,73215	0,73215
			1	$x_4 = -0,26785$	$\bar{x}_4 = 0,73215$	0,73215
		1		$x_3 = 0,58625$	$\bar{x}_3 = 1,58625$	1,58625
	1			$x_2 = 1,50164$	$\bar{x}_2 = 2,50164$	2,50164
1				$x_1 = 0,40257$	$\bar{x}_1 = 1,40257$	1,40257

4) se hallan los valores precisados de las incógnitas, adicionando a los valores aproximados de las incógnitas $x_i^{(0)}$ las correcciones correspondientes δ_i:

$$x_i = x_1^{(0)} + \delta_1;\ x_2 = x_2^{(0)} + \delta_2;\ \ldots;\ x_n = x_n^{(0)} + \delta_n.$$

Ejemplo 3. Haciendo uso del método de Gauss, resolver con tres cifras decimales el sistema

$$\begin{cases} 7,09x_1 + 1,17x_2 - 2,23x_3 = -4,75, \\ 0,43x_1 + 1,4x_2 - 0,62x_3 = -1,05, \\ 3,21x_1 - 4,25x_2 + 2,13x_3 = 5,06 \end{cases}$$

y precisar los valores aproximados obtenidos de las incógnitas hasta 10^{-4}.

△ Con ayuda del esquema de Gauss calculamos $x_1^{(0)}$, $x_2^{(0)}$ y $x_3^{(0)}$ con tres cifras significativas (tabla 3.4).

Tabla 3.4

x_1	x_2	x_3	Términos independientes	Σ	ε
$\boxed{7,09}$	1,17	−2,23	−4,75	1,28	0,00097
0,43	1,4	−0,62	−1,05	0,16	0,00087
3,21	−4,25	2,13	5,06	6,15	−0,00295
1	0,1650	−0,3145	−0,6700	0,1805	0,00014
	$\boxed{1,3290}$	−0,4847	−0,7619	0,0824	0,00081
	−4,7796	3,1395	7,2107	5,5706	−0,00340
	1	−0,3647	−0,5733	0,0620	0,00061
		$\boxed{1,3964}$	4,4705	5,8669	−0,00048
		1	3,2015	4,2015	−0,00035
		1	3,2015	4,2015	−0,00035
	1		0,5943	1,5943	0,00048
1			0,2388	1,2388	−0,0005

Ahora bien, $x_1^{(0)} = 0,239$, $x_2^{(0)} = 0,594$, $x_3^{(0)} = 3,202$.

Para hallar la corrección $\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \end{bmatrix}$, es necesario resolver el sistema dado con la misma matriz

$$A = \begin{bmatrix} 7,09 & 1,17 & -2,23 \\ 0,43 & 1,4 & -0,62 \\ 3,21 & -4,25 & 2,13 \end{bmatrix}$$

y con nuevos términos independientes ε_i (discrepancias) que se calculan del modo siguiente.

1. Calculemos las discrepancias para lo cual sustituimos en las ecuaciones del sistema dado los valores de $x_1^{(0)}$, $x_2^{(0)}$, $x_3^{(0)}$

$$7,09 \cdot 0,239 + 1,17 \cdot 0,594 - 2,23 \cdot 3,202 = -4,75097;$$
$$0,43 \cdot 0,239 + 1,4 \cdot 0,594 - 0,62 \cdot 3,202 = -1,05087;$$
$$3,21 \cdot 0,239 - 4,25 \cdot 0,594 - 2,13 \cdot 3,202 = 5,06295.$$

Las discrepancias son, respectivamente, iguales a

$$\varepsilon_1 = -4,75 - (-4,75097) = 0,00097;$$
$$\varepsilon_2 = -1,05 - (-1,05087) = 0,00087;$$
$$\varepsilon_3 = 5,06 - 5,06295 = -0,00295.$$

2. Resolvemos el sistema dado, haciendo uso del esquema de Gauss, con términos independientes $\varepsilon_1 = 0{,}00097$, $\varepsilon_2 = 0{,}00087$, $\varepsilon_3 = 0{,}00295$. Con exactitud hasta 10^{-4} obtenemos los valores respectivos de las correcciones $\delta_1 = -0{,}0004$; $\delta_2 = 0{,}0005$; $\sigma_3 = -0{,}0001$. Ahora precisemos las incógnitas:

$$x_1 = x_1^{(0)} + \delta_1 = 0{,}239 - 0{,}0004 = 0{,}2386;$$
$$x_2 = x_2^{(0)} + \delta_2 = 0{,}594 + 0{,}0005 = 0{,}5945;$$
$$x_3 = x_3^{(0)} + \delta_3 = 3{,}202 - 0{,}0001 = 3{,}2019. \ \blacktriangle$$

§ 3.7. Cálculo de los determinantes con ayuda del esquema de Gauss

El método de Gauss puede ser utilizado al calcular los determinantes:

$$\begin{vmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{vmatrix} = a_{11} a_{22}^{(1)} a_{33}^{(2)} \dots a_{nn}^{(n-1)},$$

donde $a_{11}, a_{22}^{(1)}, a_{33}^{(2)}, \dots, a_{nn}^{(n-1)}$ son los elementos guía del esquema de división única.

Tabla 3.5

Columnas 1	2	3	4	Σ_1	Σ_2
$\boxed{1}$	1	2	3	7	
3	−1	−1	−2	−1	
2	3	−1	−1	3	
1	2	3	−1	5	
1	1	2	3	7	7
	$\boxed{-4}$	−7	−11	−22	−22
	1	−5	−7	−11	−11
	1	1	−4	−2	−2
	1	7/4	11/4	22/4	22/4
		$\boxed{-27/4}$	−39/4	−66/4	−66/4
		−3/4	−27/4	−30/4	−30/4
		1	13/9	22/9	22/9
			$\boxed{-17/3}$	−17,3	−17/3

Ejemplo 1. Con ayuda del esquema de división única calcular el determinante

$$d = \begin{vmatrix} 1 & 1 & 2 & 3 \\ 3 & -1 & -1 & -2 \\ 2 & 3 & -1 & -1 \\ 1 & 2 & 3 & -1 \end{vmatrix}.$$

△ La resolución se da en la tabla 3.5. Ahora bien, $d = 1 \cdot (-4) \times \times (-27/4) \cdot (-17/3) = -153$. ▲

Ejemplo 2. Con ayuda del esquema de división única calcular el determinante, con exactitud hasta 0,001,

$$d = \begin{vmatrix} 1,00 & 0,42 & 0,54 & 0,66 \\ 0,42 & 1,00 & 0,32 & 0,44 \\ 0,54 & 0,32 & 1,00 & 0,22 \\ 0,66 & 0,44 & 0,22 & 1,00 \end{vmatrix}.$$

△ La solución se da en la tabla 3.6. Finalmente tenemos $d = 1 \cdot 0,8236 \cdot 0,6978 \cdot 0,4979 = 0,286$. ▲

Tabla 3.6

Columnas				Σ_1	Σ_2
1	2	3	4		
$\boxed{1}$	0,42	0,54	0,66	2,62	
0,42	1,00	0,32	0,44	2,18	
0,54	0,32	1,00	0,22	2,08	
0,66	0,44	0,22	1,00	2,32	
1	0,42	0,54	0,66	2,62	
	$\boxed{0,8236}$	0,0932	0,1628	1,0796	1,0796
	0,0932	0,7084	0,1364	0,6652	0,6652
	0,1628	−0,1364	0,5644	0,5908	0,5908
	1	0,1135	0,1973	1,3108	1,3108
		$\boxed{0,6978}$	−0,1548	0,5430	0,5430
		−0,1549	0,5323	0,3774	0,3774
		1	−0,2219	0,7782	0,7781
			$\boxed{0,4979}$	0,4979	0,4979

§ 3.8. Inversión de una matriz con ayuda del esquema de Gauss

Supongamos que se da la matriz regular $A = [a_{ij}]$ $(i, j = 1, 2, \ldots, n)$. Para hallar la matriz inversa $A^{-1} = [x_{ij}]$ se utiliza la relación fundamental $AA^{-1} = E$, donde E es la matriz unidad de n-ésimo orden.

Así, para la matriz de cuarto orden, multiplicando

$$AA^{-1} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \\ x_{41} & x_{42} & x_{43} & x_{44} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

obtenemos 4 sistemas de ecuaciones respecto a 16 incógnitas x_{ij} $(i, j = 1, 2, 3, 4)$.

En el caso general tienen lugar las relaciones

$$\sum_{k=1}^{n} a_{ik} \cdot x_{kj} = \delta_{ij} \quad (i, j = 1, 2, \ldots, n),$$

donde

$$\delta_{ij} = \begin{cases} 1 & \text{para } i = j, \\ 0 & \text{para } i \neq j; \end{cases}$$

δ_{ij} se llama *símbolo de Kronecker*.

Los n sistemas obtenidos de ecuaciones lineales para $j = 1, 2, \ldots, n$ tienen una misma matriz A y distintos términos independientes que forman la matriz unidad. Por eso estos sistemas pueden ser resueltos con ayuda del esquema de Gauss.

Las soluciones de x_{ij} halladas con ayuda del esquema de división única serán precisamente los elementos de la matriz inversa A^{-1}.

Ejemplo 1. Invertir con ayuda del esquema de Gauss la matriz

$$A = \begin{bmatrix} 1 & 0 & 1 & 2 \\ -1 & 2 & 3 & 1 \\ 4 & 0 & -2 & 1 \\ 0 & 2 & 1 & 2 \end{bmatrix}.$$

△ La resolución se da en la tabla 3.7.

Ahora bien,

$$A^{-1} = \begin{bmatrix} 0 & 1/3 & 1/3 & -1/3 \\ -1/2 & 1/6 & 1/6 & 1/3 \\ 1/5 & 7/15 & 1/15 & -7/15 \\ 2/5 & -1/5 & -1/5 & 2/5 \end{bmatrix}. \quad ▲$$

Ejemplo 2. Invertir la matriz

$$A = \begin{bmatrix} 1{,}00 & 0{,}47 & -0{,}11 & 0{,}55 \\ 0{,}42 & 1{,}00 & 0{,}35 & 0{,}17 \\ -0{,}25 & 0{,}67 & 1{,}00 & 0{,}36 \\ 0{,}54 & -0{,}32 & -0{,}74 & 1{,}00 \end{bmatrix}$$

Tabla 3.7

x_{1j}	x_{2j}	x_{3j}	x_{4j}	$j=1$	$j=2$	$j=3$	$j=4$	Σ_1	Σ_2
$\boxed{1}$	0	1	2	1	0	0	0	5	
−1	2	3	1	0	1	0	0	0	
4	0	−2	1	0	0	1	0	4	
0	2	1	2	0	0	0	1	6	
1	0	1	2	1	0	0	0	5	5
	$\boxed{2}$	4	3	1	1	0	0	11	11
	0	−6	−7	−4	0	1	0	−16	−16
	2	1	2	0	0	0	1	6	6
	1	2	3/2	1/2	1/2	0	0	11/2	11/2
		$\boxed{-6}$	−7	−4	0	1	0	−16	−16
		3	−7	−1	−1	0	1	−5	−5
		1	7/6	4/6	0	−1/6	0	16/6	16/6
			$\boxed{5/2}$	1	−1	−1/2	1	3	3
			1	2/5	−2/5	−1/5	2/5	6/5	6/5
			1	2/5	−2/5	−1/5	2/5	6/5	6/5
		1		1/5	7/15	1/15	−7/15	19/15	19/15
	1			−1/2	1/6	1/6	1/3	7/6	7/6
1				0	1/3	1/3	−1/3	4/3	4/3

con ayuda del esquema de Gauss; llevar a cabo todos los cálculos con cuatro cifras decimales. Redondear la respuesta hasta cifras decimales

△ La resolución se da en la tabla 3.8.

Por consiguiente,

$$A^{-1}=\begin{bmatrix} 1,9759 & -1,2017 & -0,0120 & -0,8781 \\ -1,2883 & 2,1003 & -0,4869 & 0,5268 \\ 1,4921 & -1,7239 & 1,0873 & -0,9189 \\ -0,3750 & 0,0453 & 0,6553 & 0,9626 \end{bmatrix} \approx$$

$$\approx \begin{bmatrix} 1,976 & -1,202 & -0,012 & -0,878 \\ -1,288 & 2,100 & -0,487 & 0,527 \\ 1,492 & -1,724 & 1,087 & -0,919 \\ -0,375 & 0,045 & 0,655 & 0,963 \end{bmatrix} \quad ▲$$

Tabla 3.8

x_{1j}	x_{2j}	x_{3j}	x_{4j}	$j=1$
$\boxed{1,00}$ 0,42 −0,25 0,54	0,47 1,00 0,67 −0,32	−0,11 0,35 1,00 −0,74	0,55 0,17 0,36 1,00	1 0 0 0
1	0,47	−0,11	0,55	1
	$\boxed{0,8026}$ 0,7875 −0,5738	0,3962 0,9725 −0,6806	−0,0610 0,4975 0,7030	−0,4200 0,2500 −0,5400
	1	0,4936	−0,0760	−0,5233
		$\boxed{0,5838}$ −0,3974	0,5573 0,6594	0,6621 −0,8403
		1	0,9546	1,1341
			$\boxed{1,0388}$	−0,3896
			1	−0,3750
1 0 1 0 0	1 0 0 1 0	1 0 0 0 1	1 2,91 2,94 2,78 1,48	$x_{41}=-0,3750$ $x_{31}=1,4921$ $x_{21}=-1,2883$ $x_{11}=1,9759$
0	0	0	2,91	2,91
1 0 0	0 1 0	0 0 1	1,7178 3,5075 −0,0914	1,7178 3,5075 −0,0914
1,2460	0	0	2,1403	2,1403
−0,9812 0,7150	1 0	0 1	1,8220 1,1367	1,8220 1,1367

$j=2$	$j=3$	$j=4$	Σ_1	Σ_2
−1,6807	1,7129	0	3,1209	3,1209
0,0471	0,6807	1	2,3770	2,3770
0,0453	0,6553	0,9626	2,2882	2,2882
$x_{42}=$ 0,0453 $x_{32}=$ −1,7239 $x_{22}=$ 2,1003 $x_{12}=$ −1,2017	$x_{43}=$ 0,6553 $x_{33}=$ 1,0873 $x_{23}=$ −0,4869 $x_{13}=$ −0,0120	$x_{44}=$ 0,9626 $x_{34}=$ −0,9189 $x_{24}=$ 0,5268 $x_{14}=$ −0,8781	2,2882 0,9366 1,8519 0,8841	2,2882 0,9366 1,8519 0,8841

§ 3.9. Método de elementos principales

Supongamos que se da un sistema de n ecuaciones lineales con n incógnitas:

$$\left\{\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1q}x_q + \ldots + a_{1n}x_n = a_{1,n+1}, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2q}x_q + \ldots + a_{2n}x_n = a_{2,n+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{p1}x_1 + a_{p2}x_2 + \ldots + a_{pq}x_q + \ldots\ a_{pn}x_n = a_{p,n+1}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nq}x_q + \ldots + a_{nn}x_n = a_{n,n+1}. \end{array}\right. \quad (1)$$

Compongamos la matriz ampliada del sistema (1):

$$\overline{A} = \begin{bmatrix} a_{11}a_{12}a_{13}\ \ldots\ a_{1q}\ \ldots\ a_{1n}a_{1,n+1} \\ a_{21}a_{22}a_{23}\ \ldots\ a_{2q}\ \ldots\ a_{2n}a_{2,n+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{p1}a_{p2}a_{p3}\ \ldots\ \boxed{a_{pq}}\ \ldots\ a_{pn}a_{p,n+1} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{n1}a_{n2}a_{n3}\ \ldots\ a_{nq}\ \ldots\ a_{nn}a_{n,n+1} \end{bmatrix}.$$

Elijamos un elemento no nulo a_{pq} de la matriz $\overline{A}$ el cual sea máximo en módulo y no forme parte de la columna de los términos independientes. Este elemento se llama *principal*. Luego calculemos los factores $m_i = -a_{iq}/a_{pq}$ para todas las filas con números $i \neq p$ (la p-ésima fila que contiene el elemento principal se denomina *principal*).

Luego adicionemos a cada i-ésima fila no principal la fila principal multiplicada por el factor m_i correspondiente para esta fila.

Por ejemplo:

$$b_{11}^{(1)} = a_{11} - a_{p1}\frac{a_{1q}}{a_{pq}};$$

$$b_{n2}^{(1)} = a_{n2} - a_{p2}\frac{a_{nq}}{a_{pq}}, \text{ etc.}$$

Como resultado obtenemos una nueva matriz en la cual todos los elementos de la q-ésima columna, salvo a_{pq}, se componen de ceros:

$$a_{1q} - a_{pq}\frac{a_{1q}}{a_{pq}} = 0,$$
$$a_{2q} - a_{pq}\frac{a_{2q}}{a_{pq}} = 0,$$
$$\dots\dots\dots\dots$$
$$a_{nq} - a_{pq}\frac{a_{nq}}{a_{pq}} = 0.$$

Suprimiendo esta columna y la p-ésima fila principal, obtenemos una nueva matriz $B^{(1)}$ en la cual el número de filas y columnas es en una unidad menor. Repetimos las mismas operaciones con la matriz $B^{(1)}$ después de lo cual obtenemos la matriz $B^{(2)}$, etc.

Ahora bien, vamos a construir la sucesión de las matrices $\bar{A}$, $B^{(1)}$, $B^{(2)}$, . . ., $B^{(n-1)}$ la última de las cuales es una matriz fila (fila principal) de dos términos. Para determinar las incógnitas x_i reunimos en el sistema todas las filas principales comenzando con la última.

El método expuesto de resolución de un sistema de ecuaciones lineales con n incógnitas se denomina **método de elementos principales**. La condición indispensable de su aplicación consiste en que $\det A \neq 0$.

Ejemplo 1. Haciendo uso del método de elementos principales resolver el sistema

$$\begin{cases} 3x_1 + 2x_2 + x_3 = 5, \\ 2x_1 + 5x_2 + x_3 = -3, \\ 2x_1 + x_2 + 3x_3 = 11. \end{cases}$$

$\triangle$ 1) Componemos la matriz ampliada $\bar{A}$ del sistema dado, hallamos el elemento principal y calculamos el factor m_i:

$$\bar{A} = \begin{bmatrix} 3 & 2 & 1 & 5 \\ 2 & \boxed{5} & 1 & -3 \\ 2 & 1 & 3 & 11 \end{bmatrix};$$

$a_{22} = 5$ es elemento principal; $m_1 = -2/5$; $m_3 = -1/5$.

2) Escribimos la matriz $B^{(1)}$. Tenemos

$$b_{11}^{(1)}=3-2\cdot\frac{2}{5}=\frac{11}{5};$$

$$b_{12}^{(1)}=1-1\cdot\frac{2}{5}=\frac{3}{5};$$

$$b_{13}^{(1)}=5+3\cdot\frac{2}{5}=\frac{31}{5};$$

$$b_{21}^{(1)}=2-2\cdot\frac{1}{5}=\frac{8}{5};\quad b_{22}^{(1)}=3-1\cdot\frac{1}{5}=\frac{14}{5};$$

$$b_{23}^{(1)}=11+3\cdot\frac{1}{5}=\frac{58}{5};$$

$$B^{(1)}=\begin{bmatrix} 1/5 & 3/5 & 31/5 \\ 8/5 & \boxed{14/5} & 58/5 \end{bmatrix}.$$

Aquí $b_{22}^{(1)} = 14/5$ es el elemento principal; $m_1 = -1/14$.

3) Hallamos $B^{(2)}$. Tenemos

$$b_{11}^{(2)}=\frac{11}{5}-\frac{8}{5}\cdot\frac{3}{14}=\frac{13}{7};\quad b_{12}^{(2)}=\frac{31}{5}-\frac{58}{5}\cdot\frac{3}{14}=\frac{26}{7};$$

$$B^{(2)}=[13/7\ \ 26/7].$$

4) Utilizando las filas principales, llegamos al sistema

$$\begin{cases}2x_1+5x_2+x_3=-3\\ (8/5)\,x_1+(14/5)\,x_3=58/5, \quad \text{o bien}\\ (13/7)\,x_1=26/7\end{cases}$$

$$\begin{cases}2x_1+5x_2+x_3=-3,\\ 4x_1+7x_3=29,\\ 13x_1=26.\end{cases}$$

Con ayuda del paso invertido obtenemos $x_1 = 2$, $x_3 = 3$, $x_2 = -2$. ▲

El método de Gauss es un caso particular del método de elementos principales si en calidad de elemento principal se elige el elemento superior izquierdo, distinto del cero, de la matriz correspondiente del sistema.

Para resolver los sistemas de ecuaciones lineales con ayuda del método de elementos principales se puede utilizar el esquema dado en la tabla 3.9 (en este esquema los elementos principales, elegidos arbitrariamente para el ejemplo, están encuadrados y las filas principales están numeradas con cifras romanas I a IV).

Ejemplo 2. Haciendo uso del método de elementos principales, resolver el sistema

$$\begin{cases} x_1+2x_2+3x_3+4x_4=5,\\ 2x_1+x_2+2x_3+3x_4=1,\\ 3x_1+2x_2+x_3+2x_4=1,\\ 4x_1+3x_2+2x_3+x_4=-5.\end{cases}$$

Tabla 3.9

m_i	Columnas de la matriz del sistema				Términos indepen-dientes	$\sum$	Filas princi-pales	Matri-ces
	x_1	x_2	x_3	x_4				
m_1	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	Σa_{1j}		
	a_{21}	a_{22}	$\boxed{a_{23}}$	a_{24}	a_{25}	Σa_{2j}	II	$\overline{A}$
m_3	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	Σa_{3j}		
m_4	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	Σa_{4j}		
$m_3^{(1)}$	$b_{11}^{(1)}$	$\boxed{b_{12}^{(1)}}$		$b_{14}^{(1)}$	$b_{15}^{(1)}$	$\Sigma b_{1j}^{(1)}$		
$m_4^{(1)}$	$b_{31}^{(1)}$	$b_{32}^{(1)}$		$b_{34}^{(1)}$	$b_{35}^{(1)}$	$\Sigma b_{3j}^{(1)}$	I	$B^{(1)}$
	$b_{41}^{(1)}$	$b_{42}^{(1)}$		$b_{44}^{(1)}$	$b_{45}^{(1)}$	$\Sigma b_{4j}^{(1)}$		
$m_4^{(2)}$	$\boxed{b_{31}^{(2)}}$			$b_{34}^{(2)}$	$b_{35}^{(2)}$	$\Sigma b_{3j}^{(2)}$	III	$B^{(2)}$
	$b_{41}^{(2)}$			$b_{44}^{(2)}$	$b_{45}^{(2)}$	$b_{4j}^{(2)}$		
				$\boxed{b_{44}^{(3)}}$	$b_{45}^{(3)}$	$b_{4j}^{(3)}$	IV	$B^{(3)}$

Tabla 3.10

m_i	Columnas de la matriz del sistema				Términos indepen-dientes	Σ	Filas princi-pales	Matrices
	x_1	x_2	x_3	x_4				
	1	2	3	$\boxed{4}$	5	15		
−3/4	2	1	2	3	1	9	I	$\overline{A}$
−1/2	3	2	1	2	1	9		
−1/4	4	3	2	1	−5	5		
−1/3	5/4	−1/2	−1/4		−11/4	−9/4		
−2/3	5/2	1	−1/2		−3/2	1	III	$B^{(1)}$
	$\boxed{15/4}$	5/2	5/4		−25/4	5/4		
		$\boxed{-4/3}$	−2/3		−2/3	−8/3		
−1/2		−2/3	−4/3		8/3	2/3	I	$B^{(2)}$
			−1		3	2	IV	$B^{(3)}$

△ La resolución se da en la tabla 3.10.
Así pues, obtenemos el sistema

$$\begin{cases} & & -x_3 & = 3, \\ & -(4/3)\,x_2 - (2/3)\,x_3 & & = -\ 2/3, \\ (15/4)\,x_1 + (5/2)\,x_2 + (5/4)\,x_3 & & & = -25/4, \\ x_1 + \quad 2x_2 + \quad 3x_3 + 4x_4 = & & & 5, \end{cases}$$

de donde determinamos $x_3 = -3$, $x_2 = 2$, $x_1 = -2$, $x_4 = 3$. ▲

§ 3.10. Esquema de Jaletski

Supongamos que un sistema de ecuaciones lineales se da en la forma matricial

$$Ax = b, \tag{1}$$

donde $A = [a_{ij}]$ es la matriz cuadrada de orden n, y $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}$,

$\mathbf{b} = \begin{bmatrix} a_{1,\,n+1} \\ a_{2,\,n+1} \\ \vdots \\ a_{n,\,n+1} \end{bmatrix}$ son los vectores columnas.

Representemos la matriz A en forma del producto de la matriz triangular inferior $C = [c_{ij}]$ y de la matriz triangular superior $B = [b_{ij}]$ con la diagonal unitaria, o sea

$$A = CB, \tag{2}$$

donde

$$C = \begin{bmatrix} c_{11} & 0 & 0 & \dots & 0 \\ c_{21} & c_{22} & 0 & \dots & 0 \\ c_{31} & c_{32} & c_{33} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & c_{n3} & \dots & c_{nn} \end{bmatrix},$$

$$B = \begin{bmatrix} 1 & b_{12} & b_{13} & \dots & b_{1n} \\ 0 & 1 & b_{23} & \dots & b_{2n} \\ 0 & 0 & 1 & \dots & b_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix},$$

con la particularidad de que los elementos c_{ij} y b_{ij} se determinan con ayuda de las fórmulas (17) y (18) del § 2.6:

$$c_{i1} = c_{i1}, \quad c_{ij} = a_{ij} - \sum_{k=1}^{j-1} c_{ik} b_{kj} \text{ para } 1 < j \leqslant i, \tag{3}$$

$$b_{ij} = \frac{a_{1j}}{c_{11}}, \quad b_{ij} = \frac{a_{ij} - \sum\limits_{k=1}^{i-1} c_{ik} b_{kj}}{c_{ii}} \text{ para } 1 < i < i. \tag{4}$$

La ecuación (1) puede ser escrita en la forma siguiente:

$$C B \mathbf{x} = \mathbf{b}. \tag{5}$$

El producto $B\mathbf{x}$ de la matriz B por el vector columna $\mathbf{x}$ es un vector columna que designemos con $\mathbf{y}$:

$$B\mathbf{x} = \mathbf{y}. \tag{6}$$

Entonces escribamos la ecuación (5) en la forma

$$C\mathbf{y} = \mathbf{b} \tag{7}$$

o bien

$$\begin{bmatrix} c_{11} & 0 & 0 & \ldots & 0 \\ c_{21} & c_{22} & 0 & \ldots & 0 \\ c_{31} & c_{32} & c_{33} & \ldots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ c_{n1} & c_{n2} & c_{n3} & \ldots & c_{nn} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} a_{1,\,n+1} \\ a_{2,\,n+1} \\ a_{3,\,n+1} \\ \ldots \\ a_{n,\,n+1} \end{bmatrix}. \tag{7'}$$

Aquí los elementos c_{ij} $(i, j = 1, 2, \ldots, n)$ están conocidos, ya que la matriz A del sistema (1) se considera ya desarrollada en producto de dos matrices triangulares C y B [fórmulas (3) y (4)].

Multiplicando las matrices dadas en el primer miembro de la igualdad (7′), obtenemos el sistema de ecuaciones

$$\left\{\begin{aligned} c_{11}y_1 &= a_{1,\,n+1}, \\ c_{2}y_1 + c_{22}y_2 &= a_{2,\,n+1}, \\ c_{31}y_1 + c_{32}y_2 + c_{33}y_3 &= a_{3,\,n+1}, \\ c_{n1}y_1 + c_{n2}y_2 + c_{n3}y_3 + \ldots + c_{nn}y_n &= a_{n,\,n+1}, \end{aligned}\right. \tag{8}$$

de donde obtenemos las siguientes fórmulas para determinar las incógnitas:

$$y_1 = \frac{a_{1,\,n+1}}{c_{11}}; \quad y_i = \frac{a_{i,\,n+1} - \sum\limits_{k=1}^{j-1} c_{ik} y_k}{c_{ii}}, \quad i > 1. \tag{9}$$

Es cómodo calcular las incógnitas y_i junto con los elementos b_{ij}.

Determinados todos los elementos y_i $(i = 1, 2, \ldots, n)$ con ayuda de las fórmulas (9), los sustituimos en la ecuación (6):

$$\begin{bmatrix} 1 & b_{12} & b_{13} & \ldots & b_{1n} \\ 0 & 1 & 0 & \ldots & b_{2n} \\ 0 & 0 & 1 & \ldots & b_{3n} \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \ldots & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}.$$

Multiplicando, obtenemos el sistema

$$\begin{cases} x_1 + b_{12}x_2 + b_{13}x_3 + \ldots + b_{1n}x_n = y_1, \\ \qquad x_2 + b_{23}x_3 + \ldots + b_{2n}x_n = y_2, \\ \qquad\qquad x_3 + \ldots + b_{3n}x_n = y_3, \\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot\ \cdot \\ \qquad\qquad\qquad x_n = y_n. \end{cases} \tag{10}$$

Puesto que los coeficientes b_{ij} están determinados [véase la fórmula (4)], calculamos los valores de las incógnitas, comenzando con la última, por las fórmulas siguientes:

$$x_n = y_n, \; x_i = y_i - \sum_{k=j+1}^{n} b_{ik}x_k, \quad i < n. \tag{11}$$

Este método ha recibido el nombre de **esquema de Jaletski.** En él se aplica el control ordinario con ayuda de las sumas.

Al resolver los sistemas con ayuda del esquema de Jaletski es cómodo hacer uso de la tabla 3.11 y determinar y_i junto con los coeficientes b_i.

Tabla 3.11

	x_1	x_2	x_3	x_4	Términos independientes	Σ	x_1	x_2	x_3	x_4	Términos independientes	Σ
I	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	1	2	−1	2	4	8
	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}	a_{26}	2	3	−1	4	6	14
	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}	a_{36}	4	5	−3	8	12	26
	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}	a_{46}	2	3	−2	3	6	12
II	c_{11} 1	b_{12}	b_{13}	b_{14}	$y_1 = b_{15}$	b_{16}	1 1	2	−1	2	4	8
	c_{21}	c_{22} 1	b_{23}	b_{24}	$y_2 = b_{25}$	b_{26}	2	−1 1	−1	0	2	2
	c_{31}	c_{32}	c_{33} 1	b_{34}	$y_3 = b_{35}$	b_{36}	4	−3	−2 1	0	−1	0
	c_{41}	c_{42}	c_{43}	c_{44} 1	$y_4 = b_{45}$	b_{46}	2	−1	−1	−1 1	1	2
III	x_1	x_2	x_3	x_4			−1	1	−1	1		

Ejemplo. Utilizando el esquema de Jaletski, resolver el sistema

$$\begin{cases} 3x_1 - 2x_2 - 5x_3 + x_4 = 3, \\ 2x_1 - 3x_2 + x_3 + 5x_4 = -3, \\ x_1 + 2x_2 \qquad - 4x_4 = -3, \\ x_1 - x_2 - 4x_3 + 9x_4 = 22. \end{cases}$$

△ La resolución se da en la tabla 3.11. En la parte I de la tabla escribamos la matriz de los coeficientes, los términos independientes de la misma y las sumas de control.

Luego llenamos la parte II según la regla indicada en el § 2.6, o sea, primero hallamos la primera fila de la matriz C, luego la segunda fila de la matriz B, la segunda columna de la matriz C, la segunda fila de la matriz B, etc.

En la parte III determinamos x_i.

El control corriente se lleva a cabo con ayuda de la columna Σ con la cual se realizan las mismas operaciones que con la columna de los términos independientes.

1) Los elementos de la primera columna de la matriz C los determinamos con ayuda de la fórmula

$$c_{i1} = a_{i1} \ (i = 1,\ 2,\ 3,\ 4).$$

Luego escribimos la primera columna de la parte I en la primera columna de la parte II:

$$c_{11} = a_{11} = 1,\ c_{21} = a_{21} = 2,\ c_{31} = a_{31} = 4,\ \ c_{41} = a_{41} = 2.$$

2) Determinamos los elementos de la primera fila de la matriz B con ayuda de la fórmula

$$\begin{gathered} b_{ij} = a_{1j}/c_{11}\ (j = 2,\ 3,\ 4,\ 5,\ 6), \text{ o sea,} \\ b_{12} = a_{12}/c_{11} = 2;\ b_{13} = a_{13}/c_{11} = -1; \\ b_{14} = a_{14}/c_{11} = 2;\ y_1 = b_{15} = a_{15}/c_{11} = 4; \\ b_{16} = a_{16}/c_{11} = 8;\ b_{16} = 1 + b_{12} + b_{13} + b_{14} + b_{15} = \\ = 1 + 2 - 1 + 2 + 4 = 8. \end{gathered}$$

3) Obtenemos los elementos de la segunda columna de la matriz C con ayuda de la fórmula

$$\begin{gathered} c_{i2} = a_{i2} - c_{i1}b_{12}\ (i = 2,\ 3,\ 4), \text{ o sea,} \\ c_{22} = a_{22} - c_{21}b_{12} = 3 - 2\cdot 2 = -1; \\ c_{32} = a_{32} - c_{31}b_{12} = 5 - 4\cdot 2 = -3; \\ c_{42} = a_{42} - c_{41}b_{12} = 3 - 2\cdot 2 = -1. \end{gathered}$$

4) Hallamos los elementos de la segunda fila de la matriz B con ayuda de la fórmula

$$b_{2j} = \frac{a_{2j} - c_{21} - b_{1j}}{c_{22}} \quad (j = 3,\ 4,\ 5,\ 6),$$

o sea,

$$b_{23} = \frac{a_{23} - c_{21}b_{13}}{c_{22}} = \frac{-1-2\cdot(-1)}{-1} = -1;$$

$$b_{24} = \frac{a_{24} - c_{21}b_{14}}{c_{22}} = \frac{4-2\cdot 2}{-1} = 0;$$

$$y_2 = b_{25} = \frac{a_{25} - c_{21}b_{15}}{c_{22}} = \frac{6-2\cdot 4}{-1} = 2;$$

$$b_{26} = \frac{a_{26} - c_{21}b_{16}}{c_{22}} = \frac{14-2\cdot 8}{-1} = 2;$$

$$b_{26} = 1 + b_{23} + b_{24} + b_{25} = 1 - 1 + 0 + 2 = 2.$$

5) Encontramos los elementos de la tercera columna de la matriz C con ayuda de la fórmula

$$c_{i3} = a_{i3} - c_{i1}b_{13} = c_{i2}b_{23} \ (i = 3,\ 4),$$

o sea,

$$c_{33} = a_{33} - c_{31}b_{13} - c_{32}b_{23} = 3 - 4\cdot(-1) - (-3)\,(-1) = -2;$$

$$c_{43} = a_{43} - c_{41}b_{13} - c_{42}b_{23} = -2 - 2\cdot(-1) - (-1)\cdot(-1) = 1.$$

6) Determinamos los elementos de la tercera fila de la matriz B con ayuda de la fórmula

$$b_{3j} = \frac{a_{3j} - c_{31}b_{1j} - c_{32}b_{2j}}{c_{33}} \ \ (j = 4,\ 5,\ 6),$$

o sea,

$$b_{34} = \frac{a_{34} - c_{31}b_{14} - c_{32}b_{24}}{c_{33}} = \frac{8-4\cdot 2-(-3)\cdot 0}{-2} = 0;$$

$$y_3 = b_{35} = \frac{a_{35} - c_{31}b_{15} - c_{32}b_{25}}{c_{33}} = \frac{12-4\cdot 4-(-3)\cdot 2}{-2} = -1;$$

$$b_{36} = \frac{a_{36} - c_{31}b_{16} - c_{32}b_{26}}{c_{33}} = \frac{26-4\cdot 8-(-3)\cdot 2}{-2} = 0;$$

$b_{36} = 1 + b_{34} + b_{35} = 1 + 0 - 1 = 0.$

7) Encontramos los elementos de la cuarta columna de la matriz C con ayuda de la fórmula

$$c_{44} = a_{44} - c_{41}b_{14} - c_{42}b_{24} - c_{43}b_{34},$$

o sea,

$$c_{44} = 3 - 2\cdot 2 - (-1)\cdot 0 - 4\cdot 0 = -1.$$

8) Hallamos los elementos de la cuarta fila de la matriz B con ayuda de la fórmula

$$b_{4j} = \frac{a_{4j} - c_{41}b_{1j} - c_{42}b_{2j} - c_{43}b_{3j}}{c_{44}} \ \ (j = 5,\ 6),$$

o sea,

$$y_4 = b_{45} = \frac{a_{45} - c_{41}b_{15} - c_{42}b_{25} - c_{43}b_{35}}{c_{44}} =$$

$$= \frac{6-2\cdot4-(-1)\cdot2-(-1)(-1)}{-1} = 1;$$

$$b_{46} = \frac{a_{46} - c_{41}b_{16} - c_{42}b_{26} - c_{43}b_{36}}{c_{44}} = \frac{12-2\cdot8-(-1)\cdot2-4\cdot0}{-1} = 2;$$

$$b_{46} = 1 + b_{45} = 1 + 1 = 2.$$

9) Calculamos x_i con ayuda de la fórmula

$$x_i = y_i - \sum_{k=i+1}^{n} b_{ik}x_k; \quad i = 1,\ 2,\ 3,\ 4,$$

donde $y_1 = 4$, $y_2 = 2$, $y_3 = -1$, $y_4 = 1$. Tenemos

$x_4 = y_4 = 1,$
$x_3 = y_3 - b_{34}x_4 = -1 - 0\cdot1 = -1;$
$x_2 = y_2 - b_{23}x_3 - b_{24}x_4 = 2 - (-1)(-1) - 0\cdot1 = 1;$
$x_1 = y_1 - b_{12}x_2 - b_{13}x_3 - b_{14}x_4 = 4 - 2\cdot1 - (-1)(-1) - 2\cdot1 = -1.$

Así pues,

$$x_1 = -1;\ x_2 = 1;\ x_3 = -1;\ x_4 = 1. \ \blacktriangle$$

§ 3.11. Método de iteraciones (método de aproximaciones sucesivas)

Los métodos aproximados de resolución de los sistemas de ecuaciones lineales permiten obtener los valores de las raíces del sistema con exactitud asignada en forma del límite de la sucesión de ciertos vectores. El proceso de construcción de esta sucesión se denomina proceso *iterativo*.

La efectividad de aplicación de los métodos aproximados depende de la elección del vector inicial y de la rapidez de la convergencia del proceso.

En este párrafo vamos a considerar el **método de iteraciones (método de aproximaciones sucesivas)**. Sea dado un sistema de n ecuaciones lineales con n incógnitas:

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1, \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2, \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ a_{n1}x_1 + a_{n2}x_2 + \ldots + a_{nn}x_n = b_n. \end{cases} \tag{1}$$

Escribamos el sistema (1) en la forma matricial:

$$A\mathbf{x} = \mathbf{b}, \tag{2}$$

donde

$$A=\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad \mathbf{x}=\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b}=\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Suponiendo que los elementos diagonales $a_{ii} \neq 0$ $(i = 1, 2, \dots, n)$, expresemos x_1 por la primera ecuación del sistema, expresemos x_2 por la segunda ecuación, etc. Como resultado obtenemos un sistema equivalente al sistema (1):

$$\left\{\begin{aligned} x_1 &= \frac{b_{11}}{a_{11}} - \frac{a_{12}}{a_{11}} x_2 - \frac{a_{13}}{a_{11}} x_3 - \dots - \frac{a_{1n}}{a_{11}} x_n; \\ x_2 &= \frac{b_2}{a_{22}} - \frac{a_{21}}{a_{22}} x_1 - \frac{a_{23}}{a_{22}} x_3 - \dots - \frac{a_{2n}}{a_{22}} x_n; \\ & \dots\dots\dots\dots\dots\dots\dots\dots \\ x_n &= \frac{b_n}{a_{nn}} - \frac{a_{n1}}{a_{nn}} x_1 - \frac{a_{n2}}{a_{nn}} x_2 - \dots - \frac{a_{n,\,n-1}}{a_{nn}} x_{n-1}. \end{aligned}\right. \tag{3}$$

Designemos $b_i/a_{ii} = \beta_i$; $-a_{ij}/a_{ii} = \alpha_{ij}$, donde $i = 1, 2, \dots, n$; $j = 1, 2, \dots, n$. Entonces el sistema (3) se escribe del modo siguiente:

$$\left\{\begin{aligned} x_1 &= \beta_1 + \alpha_{12}x_2 + \alpha_{13}x_3 + \dots + \alpha_{1n}x_n, \\ x_2 &= \beta_2 + \alpha_{21}x_1 + \alpha_{23}x_3 + \dots + \alpha_{2n}x_n, \\ & \dots\dots\dots\dots\dots\dots\dots\dots \\ x_n &= \beta_n + \alpha_{n1}x_1 + \alpha_{n2}x_n + \dots + \alpha_{n,\,n-1}x_{n-1}. \end{aligned}\right. \tag{3'}$$

El sistema (3′) se llama sistema reducido *a la forma normal*. Introduciendo las designaciones

$$\alpha=\begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{bmatrix}, \quad \beta=\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \ \end{bmatrix},$$

vamos a escribir el sistema (3′) en la forma matricial:

$$\mathbf{x} = \beta + \alpha\mathbf{x},$$

o bien

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}. \tag{4}$$

Vamos a resolver el sistema (4) con ayuda del método de sucesiones aproximadas. Por aproximación nula tomemos la columna de los términos independientes

$$\begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} \quad \text{(la aproximación nula).}$$

Luego vamos a construir las matrices columna

$$\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ \vdots \\ x_n^{(0)} \end{bmatrix} \quad \text{(la primera aproximación);}$$

$$\begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ \vdots \\ x_n^{(2)} \end{bmatrix} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix} + \begin{bmatrix} \alpha_{11} & \alpha_{12} & \dots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \dots & \alpha_{2n} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \dots & \alpha_{nn} \end{bmatrix} \begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ \vdots \\ x_n^{(1)} \end{bmatrix} \quad \text{(la segunda aproximación),}$$

etc.

En general, toda $k+1$-ésima aproximación se calcula con ayuda de la fórmula

$$\mathbf{x}^{(k+1)} = \beta + \alpha\mathbf{x}^{(k)} \quad (k = 0, 1, \dots, n). \tag{5}$$

Si la sucesión de aproximaciones $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \dots, \mathbf{x}^{(k)}$ tiene el límite $\mathbf{x} = \lim\limits_{k\to\infty} \mathbf{x}^{(k)}$, este límite es solución del sistema (3), puesto que en virtud de la propiedad del límite

$$\lim_{x\to\infty} \mathbf{x}^{(k+1)} = \beta + \alpha \lim_{k\to\infty} \mathbf{x}^{(k)}, \text{ o sea, } \mathbf{x} = \alpha + \beta\mathbf{x}.$$

Ejemplo 1. Haciendo uso del método de iteraciones, con exactitud hasta 10^{-1} resolver el sistema

$$\begin{cases} 8x_1 + x_2 + x_3 = 26, \\ x_1 + 5x_2 - x_3 = 7, \\ x_1 - x_2 + 5x_3 = 7. \end{cases}$$

$\triangle$ 1) Reduzcamos el sistema a la forma normal:

$$\begin{cases} x_1 = 3{,}25 - 0{,}125x_2 - 0{,}125x_3, \\ x_2 = 1{,}4 - 0{,}2x_1 + 0{,}2x_3, \\ x_3 = 1{,}4 - 0{,}2x_1 + 0{,}2x_2; \end{cases}$$

$$\alpha = \begin{bmatrix} 0 & -0{,}125 & -0{,}125 \\ -0{,}2 & 0 & 0{,}2 \\ -0{,}2 & 0{,}2 & 0 \end{bmatrix}; \quad \beta = \begin{bmatrix} 3{,}25 \\ 1{,}4 \\ 1{,}4 \end{bmatrix}.$$

2) Construimos las aproximaciones sucesivas. La aproximación nula es la siguiente:

$$\begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{bmatrix} = \begin{bmatrix} 3,25 \\ 1,4 \\ 1,4 \end{bmatrix}.$$

La primera aproximación:

$$\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} 3,25 \\ 1,4 \\ 1,4 \end{bmatrix} + \begin{bmatrix} 0 & -0,125 & -0,125 \\ -0,2 & 0 & 0,2 \\ -0,2 & 0,2 & 0 \end{bmatrix} \begin{bmatrix} 3,25 \\ 1,4 \\ 1,4 \end{bmatrix} = \begin{bmatrix} 2,9 \\ 1,03 \\ 1,03 \end{bmatrix}.$$

La segunda aproximación:

$$\begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{bmatrix} = \begin{bmatrix} 3,25 \\ 1,4 \\ 1,4 \end{bmatrix} + \begin{bmatrix} 0 & -0,125 & -0,125 \\ -0,2 & 0 & 0,2 \\ -0,2 & 0,2 & 0 \end{bmatrix} \begin{bmatrix} 2,9 \\ 1,03 \\ 1,03 \end{bmatrix} = \begin{bmatrix} 2,992 \\ 1,026 \\ 1,026 \end{bmatrix}$$

La tercera aproximación:

$$\begin{bmatrix} x_1^{(3)} \\ x_2^{(3)} \\ x_3^{(3)} \end{bmatrix} = \begin{bmatrix} 3,25 \\ 1,4 \\ 1,4 \end{bmatrix} + \begin{bmatrix} 0 & -0,125 & -0,125 \\ -0,2 & 0 & 0,2 \\ -0,2 & 0,2 & 0 \end{bmatrix} \begin{bmatrix} 2,992 \\ 1,026 \\ 1,026 \end{bmatrix} = \begin{bmatrix} 2,99 \\ 1,01 \\ 1,01 \end{bmatrix}.$$

Ahora bien, $x_1 = 2{,}99$, $x_2 = 1{,}01$, $x_3 = 1{,}01$ y con exactitud hasta 10^{-1} obtenemos $x_1 = 3,0$, $x_2 = 1,0$, $x_3 = 1,0$. ▲

Ejemplo 2. Haciendo uso del método de iteraciones, con exactitud hasta 10^{-3} resolver el sistema

$$\begin{cases} 7{,}6x_1 + 0{,}5x_2 + 2{,}4x_3 = 1{,}9, \\ 2{,}2x_1 + 9{,}1x_2 + 4{,}4x_3 = 9{,}7, \\ -1{,}3x_1 + 0{,}2x_2 + 5{,}8x_3 = -1{,}4. \end{cases} \qquad (*)$$

△ 1) Reduzcamos el sistema a la forma normal:

$$\begin{cases} x_1 = \dfrac{1,9}{7,6} - \dfrac{0,5}{7,6} x_2 - \dfrac{2,4}{7,6} x_3, \\ x_2 = \dfrac{9,7}{9,1} - \dfrac{2,2}{9,1} x_1 - \dfrac{4,4}{9,1} x_3, \quad \text{o bien} \\ x_3 = \dfrac{-1,4}{5,8} + \dfrac{1,3}{5,8} x_1 - \dfrac{0,2}{5,8} x_2, \end{cases}$$

$$\begin{cases} x_1 = 0,25 - 0,065x_2 - 0,3158x_3, \\ x_2 = 1,0659 - 0,2418x_1 - 0,4847x_2, \\ x_3 = -0,2414 + 0,2241x_1 - 0,3448x_2; \end{cases}$$

$$\alpha = \begin{bmatrix} 0 & -0,065 & -0,3158 \\ -0,2418 & 0 & -0,4847 \\ 0,2241 & -0,3448 & 0 \end{bmatrix}; \qquad \beta = \begin{bmatrix} 0,25 \\ -1,0659 \\ -0,2414 \end{bmatrix}.$$

Nótese que el sistoma lineal puede ser reducido a la forma normal también del modo siguiente: escribir los coeficientes de las x_1, x_2 y x_3 de las ecuaciones correspondientes del sistema (*) en la forma kx, donde k es un número próximo al coeficiente de la incógnita respectiva y por el cual es fácil dividir los coeficientes de las incógnitas y de los términos independientes.

Por ejemplo:

$$10x_1 = 7,6x_1 + 2,4x_1 \text{ (en la primera ecuación)},$$
$$10x_2 = 9,1x_2 + 0,9x_2 \text{ (en la segunda ecuación)},$$
$$10x_3 = 5,8x_3 + 4,2x_3 \text{ (en la tercera ecuación)}.$$

Escribamos el sistema (*) así:

$$\begin{cases} 10x_1 = 1,9 + 2,4x_1 - 0,5x_2 - 2,4x_3, \\ 10x_2 = 9,7 - 2,2x_1 + 0,9x_2 - 4,4x_3, \\ 10x_3 = -1,4 + 1,3x_1 - 0,2x_2 + 4,2x_3; \end{cases}$$

$$\begin{cases} x_1 = 0,19 + 0,24x_1 - 0,05x_2 - 0,24x_3, \\ x_2 = 0,97 - 0,22x_1 + 0,09x_2 - 0,44x_3, \\ x_3 = -0.14 + 0,13x_1 - 0,02x_2 + 0,42x_3. \end{cases}$$

La matriz α y el vector β toman la forma

$$\alpha = \begin{bmatrix} 0,24 & -0,05 & -0,24 \\ -0,22 & 0,09 & -0,44 \\ 0,13 & -0,02 & 0,42 \end{bmatrix}, \qquad \beta = \begin{bmatrix} 0,19 \\ 0,97 \\ -0,14 \end{bmatrix}.$$

2) Sucesivamente hallamos

$$\begin{bmatrix} x_1^{(0)} \\ x_2^{(0)} \\ x_3^{(0)} \end{bmatrix} = \begin{bmatrix} 0,19 \\ 0,97 \\ -0,14 \end{bmatrix};$$

$$\begin{bmatrix} x_1^{(1)} \\ x_2^{(1)} \\ x_3^{(1)} \end{bmatrix} = \begin{bmatrix} 0,19 \\ 0,97 \\ -0,14 \end{bmatrix} + \begin{bmatrix} 0,24 & -0,05 & -0,24 \\ -0,22 & 0,09 & -0,44 \\ 0,13 & -0,02 & 0,42 \end{bmatrix} \begin{bmatrix} 0,19 \\ 0,97 \\ -0,14 \end{bmatrix} = \begin{bmatrix} 0,2207 \\ 1,0771 \\ -0,1935 \end{bmatrix};$$

$$\begin{bmatrix} x_1^{(2)} \\ x_2^{(2)} \\ x_3^{(2)} \end{bmatrix} = \begin{bmatrix} 0,19 \\ 0,97 \\ -0,14 \end{bmatrix} + \begin{bmatrix} 0,24 & -0,05 & -0,24 \\ -0,22 & 0,09 & -0,44 \\ 0,13 & -0,02 & 0,42 \end{bmatrix} \begin{bmatrix} 0,2207 \\ 1,0771 \\ -0,1935 \end{bmatrix} = \begin{bmatrix} 0,2359 \\ 1,1034 \\ -0,2141 \end{bmatrix}.$$

Ahora bien, con la exactitud hasta 10^{-3} obtenemos

$$x_1 = 0,236; \; x_2 = 1,103; \; x_3 = -0,214. \; \blacktriangle$$

§ 3.12. Condiciones de convergencia del proceso iterativo

Supongamos que se da un sistema de ecuaciones lineales $\mathbf{x} = \beta + \alpha\mathbf{x}$ reducido a la forma normal. La condición de convergencia del proceso iterativo consiste en lo siguiente: *si la suma de los módulos de los elementos de las filas o la suma de los módulos de los elementos de las columnas es menor que la unidad, para el sistema dado el proceso de iteración converge a la solución única independientemente de la elección del vector inicial.*

Por consiguiente, la condición de convergencia puede escribirse así:

$$\sum_{j=1}^{n} |\alpha_{ij}| < 1 \quad (i = 1, 2, \ldots, n) \text{ o bien}$$

$$\sum_{i=1}^{n} |\alpha_{ij}| < 1 \quad (j = 1, 2, \ldots, n).$$

Ejemplo. Para el sistema

$$\begin{cases} 8x_1 + x_2 + x_3 = 26, \\ x_1 + 5x_2 - x_3 = 7, \\ x_1 - x_2 + 5x_3 = 7 \end{cases} \text{ o bien } \begin{cases} x_1 = 3,25 - 0,125x_2 - 0,125x_3, \\ x_2 = 1,4 - 0,2x_1 + 0,2x_3, \\ x_3 = 1,4 - 0,2x_1 + 0,2x_2 \end{cases}$$

el proceso iterativo converge, ya que

$$\alpha = \begin{bmatrix} 0 & -0,125 & -0,125 \\ -0,2 & 0 & 0,2 \\ -0,2 & 0,2 & 0 \end{bmatrix}$$

y

$$|\alpha_{11}|+|\alpha_{21}|+|\alpha_{31}|=0,2+0,2=0,4<1;$$
$$|\alpha_{12}|+|\alpha_{22}|+|\alpha_{32}|=0,125+0,2=0,325<1;$$
$$|\alpha_{13}|+|\alpha_{23}|+|\alpha_{33}|=0,125+0,2=0,325<1.$$

De un modo análogo se podría verificar el cumplimiento de la condición de convergencia, tomando las sumas de los módulos de los elementos de las filas.

El proceso de iteración converge a ciencia cierta si los elementos de la matriz α satisfacen la desigualdad $|\alpha_{ij}|<1/n$, donde n es la cantidad de incógnitas del sistema dado. En el ejemplo dado $n=3$ y todos los elementos $|\alpha_{ij}|<1/3$.

La convergencia del proceso iterativo está ligada con las normas de la matriz α por las relaciones siguientes. *Si se cumple una de las condiciones*

$$\|\alpha\|_1=\max_i \sum_{j=1}^{n} |\alpha_{ij}|<1$$

o

$$\|\alpha\|_2=\max_j \sum_{i=1}^{n} |\alpha_{ij}|<1$$

o bien

$$\|\alpha\|_3=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} |\alpha_{ij}|^2}<1,$$

el proceso de iteración de un sistema lineal converge a la solución única.

Así, en el ejemplo anteriormente considerado

$$\|\alpha\|_2=\text{máx}\,(0,4;\ 0,325;\ 0,325)=0,4<1,$$

o sea, el proceso iterativo converge.

§ 3.13. Estimación del error del proceso aproximado del método de iteraciones

Si se da el error admisible de cálculos ε y $\mathbf{x}_i$ es el vector de los valores exactos de las incógnitas de un sistema lineal, siendo $\mathbf{x}_i^{(k)}$ la k-ésima aproximación de los valores de las incógnitas calculada con ayuda del método de iteración, entonces para estimar el error $\|\mathbf{x}_i-\mathbf{x}_i^{(k)}\|\leqslant\varepsilon$ del método se utiliza la fórmula

$$\|\mathbf{x}_i-\mathbf{x}_i^{(k)}\|\leqslant\frac{\|\alpha\|^{k+1}}{1-\|\alpha\|}\cdot\|\beta\| \tag{1}$$

donde $\|\alpha\|$ es una de las tres normas de la matriz α; $\|\beta\|$ es la misma norma del vector β; k, la cantidad de iteraciones necesaria para alcanzar la exactitud deseada. En este caso se supone que las aproximaciones sucesivas $\mathbf{x}_i^{(j)}$ (donde $j=0, 1, \ldots; k$, $i=1, 2, \ldots, n$) se calculan exactamente y en ellas no hay errores de redondeo.

Ejemplo. Mostrar que para el sistema

$$\begin{cases} 9{,}9x_1 - 1{,}5x_2 + 2{,}6x_3 = 0, \\ 0{,}4x_1 + 13{,}6x_2 - 4{,}2x_3 = 8{,}2, \\ 0{,}7x_1 + 0{,}4x_2 + 7{,}1x_3 = -1{,}3 \end{cases}$$

el proceso iterativo converge y determinar cuántas iteraciones han de cumplirse para hallar las incógnitas con exactitud hasta 10^{-4}.

△ 1) Reduzcamos el sistema a la forma normal

$$\begin{cases} 10x_1 = 0{,}1x_1 + 1{,}5x_2 - 2{,}6x_3, \\ 20x_2 = -0{,}4x_1 + 6{,}4x_2 + 4{,}2x_3 + 8{,}2, \\ 10x_3 = -0{,}7x_1 - 0{,}4x_2 + 2{,}9x_3 - 1{,}3 \end{cases}$$

o bien

$$\begin{cases} x_1 = 0{,}01x_1 + 0{,}15x_2 - 0{,}26x_3 \\ x_2 = -0{,}02x_1 + 0{,}32x_2 + 0{,}21x_3 + 0{,}41, \\ x_3 = -0{,}07x_1 - 0{,}04x_2 + 0{,}29x_3 - 0{,}13. \end{cases}$$

2) La matriz del sistema

$$\alpha = \begin{bmatrix} 0{,}01 & 0{,}15 & -0{,}26 \\ -0{,}02 & 0{,}32 & 0{,}21 \\ -0{,}07 & -0{,}04 & 0{,}29 \end{bmatrix}.$$

Utilizando la norma $\|\alpha\|_2$, obtenemos $\|\alpha\|_2 = \text{máx}\,(0{,}1;\ 0{,}51;\ 0{,}76) = 0{,}76 < 1$. Por lo tanto, el proceso iterativo para el sistema dado converge.

3) Tenemos $\beta = \begin{bmatrix} 0 \\ 0{,}41 \\ -0{,}13 \end{bmatrix}$, $\|\beta\|_2 = 0 + 0{,}41 + 0{,}13 = 0{,}54$.

4) Aplicando la fórmula (1) hallamos

$$\|\mathbf{x} - \mathbf{x}^{(k)}\| \leqslant \frac{\|\alpha\|_2^{k+1} \cdot \|\beta\|_2}{1 - \|\alpha\|_2} = \frac{0{,}76^{k+1} \cdot 0{,}54}{0{,}46} \leqslant 10^{-4};$$

$$0{,}76^{k+1} \cdot 0{,}54 \leqslant 10^{-4} \cdot 0{,}46;\quad 0{,}76^{k+1} \leqslant \frac{10^{-4} \cdot 46}{54};$$

$$(k+1)\log 0{,}76 \leqslant \log 46 - \log 54 - 4;$$

$$-(k+1) \cdot 0{,}1192 \leqslant 1{,}6628 - 1{,}7324 - 2 = -4{,}0696;$$

$$(k+1) > \frac{4{,}0696}{0{,}1192} = 32{,}9;\quad k > 32{,}9;\quad k = 33.$$

La estimación teórica de la cantidad de iteraciones necesarias para asegurar la exactitud prefijada resulta prácticamente elevada. ▲

§ 3.14. Método de Seydel. Condiciones de convergencia del proceso de Seydel

El método de Seydel es la modificación del método de aproximaciones sucesivas. Al calcular la $(k+1)$-ésima aproximación de la incógnita x_i $(i > 1)$ en el método de Seydel tienen en cuenta las $(k+1)$-ésimas aproximaciones antes halladas de las incógnitas $x_1, x_2, \ldots, x_{i-1}$.

Supongamos que se da un sistema lineal reducido a la forma normal:

$$\left\{\begin{array}{l} x_1 = \beta_1 + \alpha_{11}x_1 + \alpha_{12}x_2 + \ldots + \alpha_{1n}x_n, \\ x_2 = \beta_2 + \alpha_{21}x_1 + \alpha_{22}x_2 + \ldots + \alpha_{2n}x_n, \\ x_3 = \beta_3 + \alpha_{31}x_1 + \alpha_{32}x_2 + \ldots + \alpha_{3n}x_n, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n = \beta_n + \alpha_{n1}x_1 + \alpha_{n2}x_2 + \ldots + \alpha_{nn}x_n. \end{array}\right. \tag{1}$$

Elegimos arbitrariamente las aproximaciones lineales de las incógnitas $x_1^{(0)}, x_2^{(0)}, \ldots, x_n^{(0)}$ y las sustituimos en la primera ecuación del sistema (1):

$$x_1^{(1)} = \beta_1 + \alpha_{11}x_1^{(0)} + \alpha_{12}x_2^{(0)} + \ldots \alpha_{1n}x_n^{(0)};$$

sustituimos la primera aproximación obtenida $x_1^{(1)}$ en la segunda ecuación del sistema (1):

$$x_2^{(1)} = \beta_2 + \alpha_{21}x_1^{(1)} + \alpha_{22}x_2^{(0)} + \ldots + \alpha_{2n}x_n^{(0)};$$

sustituimos las primeras aproximaciones obtenidas $x_1^{(1)}$ y $x_2^{(1)}$ en la tercera ecuación del sistema (1):

$$x_3^{(1)} = \beta_3 + \alpha_{31}x_1^{(1)} + \alpha_{32}x_2^{(1)} + \ldots + \alpha_{3n}x_n^{(0)},$$

etc. Por último,

$$x_n^{(1)} = \beta_n + \alpha_{n1}x_1^{(1)} + \alpha_{n2}x_2^{(1)} + \ldots + \alpha_{n,n-1}x_{n-1}^{(1)} + \alpha_{nn}x_n^{(0)}.$$

De un modo análogo construimos las iteraciones dos, tres, etc.

Ahora bien, suponiendo que las k-ésimas aproximaciones $x_i^{(k)}$ están conocidas, con ayuda del método de Seydel construimos las $k+1$-ésimas aproximaciones al hacer uso de las fórmulas siguientes:

$$\begin{array}{l} x_1^{(k+1)} = \beta_1 + \sum\limits_{j=1}^{n} \alpha_{1j}x_j^{(k)}, \\ x_2^{(k+1)} = \beta_2 + \alpha_{21}x_1^{(k+1)} + \sum\limits_{j=2}^{n} \alpha_{2j}x_j^{(k)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ x_n^{(k+1)} = \beta_n + \sum\limits_{j=1}^{n-1} \alpha_{nj}x_j^{(k+1)} + \alpha_{nn}x_n^{(k)}, \end{array} \tag{2}$$

donde $k = 0, 1, 2, \ldots, n$.

Ejemplo 1. Haciendo uso del método de Seydel, con precisión hasta 10^{-3}, resolver el sistema

$$\begin{cases} 7{,}6x_1+0{,}5x_2+2{,}4x_3=1{,}9, \\ 2{,}2x_1+9{,}1x_2+4{,}4x_3=9{,}7, \\ -1{,}3x_1+0{,}2x_2+5{,}8x_3=-1{,}4. \end{cases}$$

△ 1) Reduzcamos el sistema a la forma normal:

$$\begin{cases} 10x_1= \quad 1{,}9+2{,}4x_1-0{,}5x_2-2{,}4x_3, \\ 10x_2= \quad 9{,}7-2{,}2x_1+0{,}9x_2-4{,}4x_3, \\ 10x_3=-1{,}4+1{,}3x_1-0{,}2x_2+4{,}2x_3; \end{cases}$$

o bien

$$\begin{cases} x_1= \quad 0{,}19+0{,}24x_1-0{,}05x_2-0{,}24x_3, \\ x_2= \quad 0{,}97-0{,}22x_1+0{,}09x_2-0{,}44x_3, \\ x_3=-0{,}14+0{,}13x_1-0{,}02x_2+0{,}42x_3. \end{cases}$$

2) Por aproximaciones nulas tomemos los valores correspondientes de los términos independientes: $x_0^{(1)}=0{,}19$; $x_2^{(0)}=0{,}97$; $x_3^{(0)}=$ $=-0{,}14$.

3) Construimos las iteraciones con ayuda del método de Seydel. Las primeras aproximaciones:

$$x_1^{(1)}=0{,}19+0{,}24\cdot0{,}19-0{,}05\cdot0{,}97-0{,}24\cdot(-0{,}14)=0{,}2207,$$
$$x_2^{(1)}=0{,}97-0{,}22\cdot0{,}2207+0{,}09\cdot0{,}97-0{,}44\cdot(-0{,}14)=1{,}0703,$$
$$x_3^{(1)}=-0{,}14+0{,}13\cdot0{,}2207-0{,}02\cdot1{,}0703+0{,}42\cdot(-0{,}14)= \\ =-0{,}1915.$$

Las segundas aproximaciones:

$$x_1^{(2)}=0{,}19+0{,}24\cdot0{,}2207-0{,}05\cdot1{,}0703-0{,}24\cdot(-0{,}1915)= \\ =0{,}2354,$$
$$x_2^{(2)}=0{,}97-0{,}22\cdot0{,}2354+0{,}09\cdot1{,}0703-0{,}44\,(-0{,}1915)= \\ =1{,}0988,$$
$$x_3^{(2)}=-0{,}14+0{,}13\cdot0{,}2354-0{,}02\cdot1{,}0988+0{,}42\cdot(-0{,}1915)= \\ =-0{,}2118, \text{ etc.}$$

La resolución de este ejemplo se da en la tabla 3.12. La construcción de las iteraciones termina cuando, con el grado de

Tabla 3.12

Nº de la iteración	x_1	x_2	x_3	Nº de la iteración	x_1	x_2	x_3
0	0,19	0,97	−0,14	5	0,2467	1,1138	−0,2237
1	0,2207	1,0703	−0,1915	6	0,2472	1,1143	−0,2241
2	0,2354	1,0988	−0,2118	7	0,2474	1,1145	−0,2243
3	0,2424	1,1088	−0,2196	8	0,2475	1,1145	−0,2243
4	0,2454	1,1124	−0,2226				

exactitud prefijado, obtenemos iguales valores en dos iteraciones seguidas. En el ejemplo dado són las iteraciones 7 y 8.

La respuesta final: $x_1 \approx 0{,}248$; $x_2 \approx 1{,}114$; $x_3 \approx -0{,}224$. ▲

Para un sistema lineal $\mathbf{x} = \beta + \alpha\mathbf{x}$ el proceso de Seydel, al igual que el de aproximaciones sucesivas, *converge a la única solución en toda elección de la aproximación inicial si al menos una de las normas de la matriz α es menor que la unidad, o sea, si*

$$\|\alpha\|_1 = \max_i \sum_{j=1}^{n} |\alpha_{ij}| < 1$$

o

$$\|\alpha\|_2 = \max_j \sum_{i=1}^{n} |\alpha_{ij}| < 1$$

o bien

$$\|\alpha\|_3 = \sqrt{\sum_{i,\,j} |\alpha_{ij}|^2} < 1.$$

El proceso de Seydel converge a la única solución más rápidamente que el de simple iteración.

Ejemplo 2. Verificar si converge o no el proceso de Seydel para el sistema considerado en el ejemplo 1.

△ 1) Reducido el sistema a la forma normal (véase la pág. 144), obtenemos la matriz

$$\alpha = \begin{bmatrix} 0{,}24 & -0{,}05 & -0{,}24 \\ -0{,}22 & 0{,}09 & -0{,}44 \\ 0{,}13 & -0{,}02 & 0{,}42 \end{bmatrix}.$$

2) Hallamos $\|\alpha\|_1 = \max\limits_i \sum\limits_j |\alpha_{ij}| = \max(0{,}53;\ 0{,}75;\ 0{,}57) =$ $= 0{,}75 < 1$. Por lo tanto, para el sistema dado el proceso de iteración converge a la única solución, a pesar de que

$$\|\alpha\|_2 = \max_j \sum_{i=1}^{n} |\alpha_{ij}| = \max(0{,}59;\ 0{,}16;\ 1{,}1) = 1{,}1 > 1. \ ▲$$

§ 3.15. Estimación del error del proceso de Seydel

Supongamos que se da un sistema lineal $\mathbf{x} = \beta + \alpha\mathbf{x}$. Si $\mathbf{x}_i$ es el valor exacto de las incógnitas del sistema lineal y $\mathbf{x}_i^{(k)}$ es la k-ésima aproximación, calculada con ayuda del método de Seydel, entonces para estimar el error de este método se aplica la fórmula

$$\|\mathbf{x} - \mathbf{x}^{(k)}\|_1 \leqslant \frac{\|\alpha\|_1^{(k)}}{1 - \|\alpha\|_1} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|_1. \qquad (1)$$

Ejemplo. Calcular cuántas iteraciones han de cumplirse con ayuda del método de Seydel para que con exactitud hasta 10^{-4} se halle la solución del sistema

$$\begin{cases} 9,9x_1 - 1,5x_2 + 2,6x_3 = 0, \\ 0,4x_1 + 13,6x_2 - 4,2x_3 = 8,2, \\ 0,7x_1 + 0,4x_2 + 7,1x_3 = -1,3. \end{cases}$$

△ 1) Reduzcamos el sistema a la forma normal (véase la pág. 142):

$$\begin{cases} x_1 = 0,01x_1 + 0,15x_2 - 0,26x_3, \\ x_2 = 0,41 - 0,02x_1 + 0,32x_2 + 0,21x_3, \\ x_3 = -0,13 - 0,07x_1 - 0,04x_2 + 0,29x_3. \end{cases}$$

2) Tomemos por aproximaciones nulas la columna de los términos independientes $x_1^{(0)} = 0$; $x_2^{(0)} = 41$; $x_3^{(0)} = -0,13$ y calculemos las primeras aproximaciones:

$$x_1^{(1)} = 0,01\cdot 0 + 0,15\cdot 0,41 - 0,26\cdot(-0,13) = 0,0953;$$

$$x_2^{(1)} = 0,41 - 0,02\cdot 0,0953 + 0,32\cdot 0,41 + 0,21\cdot(-0,13) = 0,5120;$$

$$x_3^{(1)} = -0,13 - 0,07\cdot 0,0953 - 0,04\cdot 0,5120 + 0,29 \times (-0,13) = -0,1948.$$

3) La matriz

$$\alpha = \begin{bmatrix} 0,01 & 0,15 & -0,26 \\ -0,02 & 0,32 & 0,21 \\ -0,07 & -0,04 & 0,29 \end{bmatrix}.$$

Así pues, $\|\alpha\|_1 = \text{máx}\,(0,42;\ 0,55;\ 0,40) = 0,55$. Puesto que

$$\mathbf{x}^{(0)} = \begin{bmatrix} 0 \\ 0,41 \\ -0,13 \end{bmatrix}, \quad \text{y} \quad \mathbf{x}^{(1)} = \begin{bmatrix} 0,0953 \\ 0,5120 \\ -0,1948 \end{bmatrix},$$

tenemos

$$\mathbf{x}^{(1)} - \mathbf{x}^{(0)} = \begin{bmatrix} 0,0953 \\ 0,1120 \\ -0,0648 \end{bmatrix}, \text{ o sea, } \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\| = 0,1120.$$

4) Con ayuda de la fórmula (1) determinamos k:

$$10^{-4} \leqslant \frac{0,55^k}{0,45}\cdot 0,1120;\ 10^{-4}\cdot 0,45 \leqslant 0,55^k\cdot 0,1120;$$

$$-4\log 10 + \log 0,45 \leqslant k\log 0,55 + \log 0,1120;$$

$$-4 - 0,3468 \leqslant k(-0,2596 - 0,9508);\quad k \geqslant \frac{4,3468}{1,2104} = 3,59;\ k = 4.$$

De un modo análogo se puede estimar el método de Seydel por la norma $\|\alpha\|_2$. ▲

§ 3.16. Reducción de un sistema de ecuaciones lineales a la forma cómoda para iteraciones

Los procesos de aproximaciones sucesivas y de Seydel para un sistema lineal $\mathbf{x} = \beta + \alpha\mathbf{x}$ convergen a la única solución sin depender de la elección del vector inicial si

$$\sum_{j=1}^{n} |\alpha_{ij}| < 1 \ (i = 1, 2, \ldots, n)$$

o bien

$$\sum_{i=1}^{n} |\alpha_{ij}| < 1 \ (j = 1, 2, \ldots, n).$$

Ahora bien, para la convergencia de los procesos iterativos anteriormente mencionados es suficiente que los valores de los elementos α_{ij} de la matriz α, cuando $i \neq j$, sean pequeños en módulo. Esto es equivalente al hecho de que si para un sistema lineal $A\mathbf{x} = \mathbf{b}$ los módulos de los coeficientes diagonales de cada ecuación del sistema son más que la suma de los módulos de todos los demás coeficientes (sin tener en cuenta los términos independientes), los procesos iterativos para este sistema convergen, o sea, si se da un sistema $\sum_{j=1}^{n} a_{ij}x_j = b_i$ $(i = 1, 2, \ldots, n)$, con la particularidad de que $|a_{ii}| > \sum_{j \neq i} |a_{ij}|$, para este sistema los procesos de aproximaciones sucesivas y de Seydel convergen.

Realizando las transformaciones lineales, el sistema lineal $A\mathbf{x} = \mathbf{b}$ puede ser sustituido por un tal sistema equivalente $\mathbf{x} = \beta + \alpha\mathbf{x}$ para el cual las condiciones de convergencia estarán cumplidas.

Ejemplo. Reducir a una forma cómoda para iteraciones el sistema de ecuaciones lineales

$$\begin{cases} 0{,}9x_1 + 2{,}7x_2 - 3{,}8x_3 = 2{,}4, & \text{(A)} \\ 2{,}5x_1 + 5{,}8x_2 - 0{,}5x_3 = 3{,}5, & \text{(B)} \\ 4{,}5x_1 - 2{,}1x_2 + 3{,}2x_3 = -1{,}2. & \text{(C)} \end{cases}$$

△ 1) Del sistema dado separamos las ecuaciones con los coeficientes cuyos módulos son más que la suma de los módulos de los demás coeficientes del sistema. Cada ecuación separada la escribimos en tal fila del nuevo sistema que el coeficiente mayor en módulo resulte ser diagonal.

En la ecuación (B) el coeficiente de x_2 es mayor en módulo que la suma de los módulos de los demás coeficientes. Tomamos la ecuación (B) por segunda ecuación del sistema nuevo:

$$2{,}5x_1 + 5{,}8x_2 - 0{,}5x_3 = 3{,}5. \qquad \text{(II)}$$

2) De las ecuaciones quedadas no utilizadas del sistema hacemos combinaciones linealmente independientes entre sí. Así, por primera ecuación del nuevo sistema se puede tomar la combinación lineal

(2B) + (A), entonces tenemos

$$9{,}9x_1 - 1{,}5x_2 + 2{,}6x_3 = 0. \tag{I}$$

Por tercera ecuación del nuevo sistema se puede tomar la combinación lineal (2A) . . . (B), es decir,

$$0{,}7x_1 + 0{,}4x_2 + 7{,}1x_3 = -1{,}3. \tag{III}$$

3) En resumen obtenemos un sistema transformado de las ecuaciones (I), (II), (III) que es equivalente al inicial y que satisface la condición de convergencia del proceso iterativo:

$$\begin{cases} 9{,}9x_1 - 1{,}5x_2 + 2{,}6x_3 = 0, \\ 2{,}5x_1 + 5{,}8x_2 - 0{,}5x_3 = 3{,}5, \\ 0{,}7x_1 + 0{,}4x_2 + 7{,}1x_3 = -1{,}3. \end{cases}$$

Reduciendo este sistema a la forma normal, tenemos

$$\begin{cases} x_1 = 0{,}1x_1 + 0{,}15x_2 - 0{,}26x_3 \\ x_2 = 0{,}35x_1 - 0{,}21x_2 + 0{,}42x_3 + 0{,}05 \\ x_3 = -0{,}13x_1 - 0{,}07x_2 - 0{,}04x_3 + 0{,}29; \end{cases}$$

$$\alpha = \begin{bmatrix} 0{,}10 & 0{,}15 & -0{,}26 \\ 0{,}35 & -0{,}21 & 0{,}42 \\ -0{,}13 & -0{,}07 & -0{,}04 \end{bmatrix};$$

$$\| \alpha \|_2 = \text{máx}\,(0{,}58;\ 0{,}43;\ 0{,}72) = 0{,}72 < 1.$$

Nos queda por resolver el sistema haciendo uso de uno de los métodos iterativos. ▲

Ejercicios

1. Resolver los sistemas homogéneos de ecuaciones:

a) $\begin{cases} x_1 + 3x_2 + 2x_3 = 0, \\ 2x_1 - x_2 + 3x_3 = 0, \\ 3x_1 - 5x_2 + 4x_3 = 0, \\ x_1 + 17x_2 + 4x_3 = 0; \end{cases}$ b) $\begin{cases} 3x_1 + 4x_2 - 5x_3 + 7x_4 = 0, \\ 2x_1 - 3x_2 + 3x_3 - 2x_4 = 0, \\ 4x_1 + 11x_2 - 13x_3 + 16x_4 = 0, \\ 7x_1 - 2x_2 + x_3 + 3x_4 = 0. \end{cases}$

2. Investigar la compatibilidad y hallar la solución general y una solución particular de los siguientes sistemas de ecuaciones:

a) $\begin{cases} 2x_1 + 7x_2 + 3x_3 + x_4 = 6, \\ 3x_1 + 5x_2 + 2x_3 + 2x_4 = 4, \\ 9x_1 + 4x_2 + x_3 + 7x_4 = 2; \end{cases}$ b) $\begin{cases} 2x_1 - 3x_2 + 5x_3 + 7x_4 = 1, \\ 4x_1 - 6x_2 + 2x_3 + 3x_4 = 2, \\ 2x_1 - 3x_2 - 11x_3 - 15x_4 = 1. \end{cases}$

3. Haciendo uso de las fórmulas de Cramer, resolver los siguientes sistemas de ecuaciones lineales:

a) $\begin{cases} 2x_1 + x_2 + 4x_3 = 7, \\ 2x_1 - x_2 - x_3 = -5, \\ 3x_1 + 4x_2 - 5x_3 = -14; \end{cases}$ b) $\begin{cases} 11x + 3y - z = 15, \\ 2x + 5y + 5z = -11, \\ x + y + z = 1. \end{cases}$

4. Con ayuda del esquema de Gauss resolver los siguientes sistemas:

a) $\begin{cases} x_1-4x_2 \qquad - x_4=6, \\ x_1+x_2+2x_3+3x_4=-1, \\ 2x_1+3x_2-x_3-x_4=-1, \\ x_1+2x_2+3x_3-x_4=3; \end{cases}$ b) $\begin{cases} 2x_1-x_3-2x_4=-8, \\ x_2+2x_3-x_4=-1, \\ x_1-x_2-x_4=-6, \\ -x_1+3x_2-2x_3=7. \end{cases}$

5. Haciendo uso del esquema de Gauss, resolver con exactitud hasta 0,001 los siguientes sistemas:

a) $\begin{cases} 1{,}14x_1-2{,}15x_2-5{,}11x_3=2{,}05, \\ 0{,}42x_1-1{,}13x_2+7{,}05x_3=0{,}80, \\ -0{,}71x_1+0{,}81x_2-0{,}02x_3=-1{,}07; \end{cases}$

b) $\begin{cases} 0{,}61x+0{,}71y-0{,}05z=-0{,}16, \\ -1{,}03x-2{,}05y+0{,}87z=0{,}50, \\ 2{,}5x-3{,}12y+5{,}03z=0{,}95. \end{cases}$

6. Con ayuda del esquema de Gauss calcular los determinantes:

a) $d=\begin{vmatrix} 1 & 4 & 1 & 3 \\ 0 & -1 & 3 & -1 \\ 3 & 1 & 0 & 2 \\ 1 & -2 & 5 & 1 \end{vmatrix}$; b) $d=\begin{vmatrix} -1{,}6 & 5{,}4 & -7{,}7 & 3{,}1 \\ 8{,}2 & 1{,}4 & -2{,}3 & 0{,}2 \\ 5{,}3 & -5{,}9 & 2{,}7 & -7{,}9 \\ 0{,}7 & 1{,}9 & -8{,}5 & 4{,}8 \end{vmatrix}$.

7. Haciendo uso del esquema de Gauss, invertir las matrices siguientes:

a) $A=\begin{bmatrix} 1 & 2 & 2 & -1 \\ 2 & 7 & 6 & -1 \\ 0 & 3 & 1 & 4 \\ 0 & 0 & 1 & -1 \end{bmatrix}$;

b) $A=\begin{bmatrix} 0{,}32 & 0{,}52 & -0{,}42 & 0{,}23 \\ 0{,}44 & -0{,}25 & 0{,}36 & -0{,}51 \\ -1{,}06 & 0{,}74 & -0{,}83 & 0{,}48 \\ 0{,}96 & 0{,}82 & 0{,}55 & 0{,}36 \end{bmatrix}$.

Llevar a cabo los cálculos con tres cifras decimales, redondear la respuesta hasta dos cifras decimales.

8. Determinada previamente la cantidad necesaria de iteraciones, por el método de aproximaciones sucesivas resolver con exacti-

tud hasta 0,01 los siguientes sistemas de ecuaciones lineales:

$$\text{a)}\quad \begin{cases} 8{,}7x_1-3{,}1x_2+1{,}8x_3-2{,}2x_4=-9{,}7, \\ 2{,}1x_1+6{,}7x_2-2{,}2x_3 \qquad\qquad =13{,}1, \\ 3{,}2x_2-1{,}8x_2-9{,}5x_3-1{,}9x_4=6{,}9, \\ 1{,}2x_1+2{,}8x_2-1{,}4x_3-9{,}9x_4=25{,}1; \end{cases}$$

$$\text{b)}\quad \begin{cases} 6{,}1x+0{,}7y-0{,}05z=6{,}97, \\ -1{,}3x-2{,}05y+0{,}87z=0{,}10, \\ 2{,}5x-3{,}12y-5{,}03z=2{,}04. \end{cases}$$

9. Determinada previamente la cantidad necesaria de iteraciones, con ayuda del método de Seydel resolver los sistemas de ecuaciones lineales del ejercicio 8.

10. Para las matrices

$$\text{a)}\ A=\begin{bmatrix} -0{,}3 & 1{,}2 & -0{,}2 \\ -0{,}1 & -0{,}2 & 1{,}6 \\ -1{,}5 & -0{,}3 & 0{,}1 \end{bmatrix};\quad \text{b)}\ A=\begin{bmatrix} 0{,}2 & 0{,}44 & 0{,}81 \\ 0{,}58 & -0{,}29 & 0{,}05 \\ 0{,}05 & 0{,}34 & 0{,}1 \end{bmatrix}$$

calcular las normas $\| A \|_1$, $\| A \|_2$ y $\| A \|_3$.

11. Haciendo uso del método de elementos principales, resolver los siguientes sistemas:

$$\text{a)}\quad \begin{cases} 3x_1-2x_2-5x_3+\ x_4=-5, \\ 2x_1-3x_2+\ x_3+5x_4=7, \\ x_1+2x_2 \qquad\quad -4x_4=-1, \\ x_1-x_2-3x_3 \quad +9x_4=-4; \end{cases}$$

$$\text{b)}\quad \begin{cases} 4x_1-3x_2+\ x_3-5x_4=7, \\ 7x_1-2x_2-3x_3-2x_4=6, \\ 3x_1-2x_2+5x_3-2x_4=0, \\ 2x_1+3x_2+5x_3+4x_4=-5. \end{cases}$$

12. Con ayuda del esquema de Jaletski resolver los sistemas de ecuaciones lineales:

$$\text{a)}\quad \begin{cases} x_1+2x_2+3x_3+4x_4=5, \\ 2x_1+\ x_2+2x_3+3x_4=1, \\ 3x_1+2x_2+\ x_3+2x_4=1, \\ 4x_1+3x_2+2x_3+\ x_4=-5; \end{cases}\qquad \text{b)}\quad \begin{cases} x_1+2x_2+3x_3=0, \\ 2x_1+\ x_2+2x_3=1, \\ 3x_1+2x_2+\ x_3=4. \end{cases}$$

CAPITULO IV

Cálculo de los valores de las funciones elementales

En la práctica de cálculos se necesita, bastante frecuentemente, resolver un problema de cálculo de cierta función en el punto dado. En este caso conviene tener en cuenta que matemáticamente las expresiones equivalentes no siempre resultan ser unívocas desde el punto de vista del cálculo de sus valores. Por ejemplo, para calcular el primer miembro de la identidad

$$a^2+2ab+b^2=(a+b)^2$$

es necesario cumplir cuatro operaciones de multiplicación y dos operaciones de adición, mientras que para calcular el segundo miembro hace falta hacer una sola adición y una sola multiplicación.

Ahora bien, llegamos a un problema importante referente a la forma óptima de representar la función para construir el algoritmo de cálculo de los valores de la misma. Lo óptimo puede entenderse como minimización de la cantidad total de las operaciones aritméticas o del tiempo indispensable para calcular los valores de la función.

§ 4.1. Cálculo de los valores de los polinomios algebraicos

Se llama *polinomio algebraico de n-ésimo grado* la expresión que tiene la forma

$$P(x)=a_0x^n+a_1x^{n-1}+\ldots+a_{n-1}x+a_n, \tag{1}$$

donde los coeficientes $a_0, a_1, \ldots, a_n$ son los números reales, con la particularidad de que a_0 se considera distinto del cero. El coeficiente a_n se denomina *término independiente* del polinomio (1).

El cálculo de los valores del polinomio algebraico es un ejemplo característico concerniente a la minimización de la cantidad de operaciones de un procedimiento de cálculo. Este problema resulta ser importante desde el punto de vista práctico no sólo en su sentido directo sino también porque está estrechamente vinculado con el problema de división del polinomio por un binomio lineal (polinomio de primer grado), es decir, con el problema de determinación de las raíces del polinomio.

Se denomina *raíz (cero) del polinomio* (1) todo número ξ que anula este polinomio al sustituir en el mismo ξ en vez de x, o sea, $P(\xi)=0$.

La raíz ξ del polinomio (1) se llama *raíz de multiplicidad* s (*o raíz de s-ésima multiplicidad*) si el mismo polinomio y todas sus derivadas hasta el orden $s-1$ se anulan en el punto $x=\xi$, mientras que la s-ésima derivada no se anula en este punto:

$$P(\xi)=P'(\xi)=\ldots=P^{(s-1)}(\xi)=0;\quad P^{(s)}(\xi)\neq 0.$$

Si la raíz de un polinomio tiene la multiplicidad $s=1$, esta raíz se considera *simple*.

Ejemplo 1. Determinar la multiplicidad de la raíz $x=-2$ del polinomio

$$P(x)=x^4+7x^3+18x^2+20x+8.$$

△ Hallamos la derivada primera $P'(x)=4x^3+21x^2+36x+20$; calculamos $P'(-2)=0$. Determinamos la segunda derivada $P''(x)=12x^2+42x+36$; tenemos $P''(-2)=0$. Encontramos la tercera derivada $P'''(x)=24x+42$; tenemos $P'''(-2)=-6\neq 0$.

Ahora bien, el mismo polinomio y sus derivadas primera y segunda se anulan en el punto $x=-2$, mientras que la tercera derivada es distinta del cero en este punto; por lo tanto, la raíz $\xi=-2$ tiene la multiplicidad $s=3$. ▲

Las propiedades fundamentales de las raíces del polinomio algebraico se dan en el § 5.8.

Consideremos el problema de cálculo del polinomio (1) en cierto punto $x=x^*$. Para resolver el problema planteado representemos este polinomio en la forma siguiente:

$$P_n(x)=a_n+x(a_{n-1}+x(a_{n-2}+\ldots+x(a_1+xa_0)\ldots)). \tag{2}$$

Para hallar el valor de $P_n(x^*)$ calculemos sucesivamente

$$\begin{aligned} b_0&=a_0,\\ b_1&=a_1+x^*b_0,\\ b_2&=a_2+x^*b_1,\\ &\ldots\ldots\ldots\\ b_n&=a_n+x^*b_{n-1}.\end{aligned} \tag{3}$$

De las expresiones dadas para b_i se ve que cada coeficiente siguiente b_i se obtiene adicionando el coeficiente respectivo a_i al producto del coeficiente precedente b_{i-1} por x^*.

Para calcular $P_n(x^*)$ y los coeficientes b_i «a mano» es cómodo hacer uso de la tabla siguiente.

En la primera fila se escriben los coeficientes del polinomio $P_n(x)$ (los coeficientes negativos se toman con signo «—», delante de los coeficientes positivos el signo «+» puede omitirse). En la tercera fila se pone inmediatamente $b_0=a_0$. Luego cada coeficiente b_i se multiplica por x^* y se escribe debajo del coeficiente siguiente

Tabla 4.1

a_0	a_1	a_2	a_3	...	a_{n-1}	a_n
	x^*b_0	x^*b_1	x^*b_2	...	x^*b_{n-2}	x^*b_{n-1}
b_0	b_1	b_2	b_3	...	b_{n-1}	b_n

a_{i+1}. Los números contenidos en las filas 1 y 2 se suman y el resultado b_{i+1} se escribe en la tercera fila.

En virtud de construcción del proceso de cálculo es evidente que $b_n = P_n(x^*)$.

No es difícil computar la cantidad de operaciones que ha de cumplirse para calcular $P_n(x^*)$: por la fórmula (1) hace falta efectuar $2n - 1$ multiplicaciones y n adiciones, mientras que por la fórmula 2 deben realizarse n adiciones y n multiplicaciones.

Ahora bien, la fórmula (2) permite economizar casi dos veces al realizar la operación de multiplicación. Además, si tenemos en cuenta que para la operación de multiplicación se gasta un tiempo varias veces mayor que es necesario para realizar la operación de adición, la aplicación del esquema de cálculo (2) resulta bastante eficaz.

El esquema considerado (2) . . . (3) lleva el nombre de **esquema de Horner**.

Ejemplo 1. Haciendo uso del esquema de Horner, calcular el valor del polinomio $P_5(x) = x^5 + 3x^4 - 2x^3 + x^2 - x + 1$ para $x = 3$.

△ Componemos el esquema de Horner para el polinomio dado

1	3	−2	1	−1	1
	3	18	48	147	438
1	6	16	49	146	$439 = P_5(3)$

Ahora bien, $P_5(3) = 439$. ▲

Resulta que al realizar el esquema de Horner no sólo calculamos $b_n = P_n(x^*)$, sino determinamos también los coeficientes b_i $(i = 0, 1, \ldots, n-1)$ del polinomio

$$P^{(1)}_{n-1}(x) = b_0 x^{n-1} + b_1 x^{n-2} + \ldots + b_{n-1} \tag{4}$$

que es el cociente obtenido de la división del polinomio $P_n(x)$ por el binomio $x - x^*$:

$$P_n(x) = (x - x^*) P_{n-1}^{(1)}(x) + P_n(x^*). \tag{5}$$

En efecto, abriendo paréntesis en el segundo miembro de la relación (5), reduciendo los términos semejantes y teniendo en cuenta que $P_n(x^*) = b_n$, obtenemos

$$P_n(x) = b_0x^n + (b_1 - x^*b_0)x^{n-1} + (b_2 - x^*b_1)x^{n-2} + \ldots$$
$$\ldots + (b_{n-1} - x^*b_{n-2})x + b_n - x^*b_{n-1}.$$

De aquí

$$a_0 = b_0,$$
$$a_1 = b_1 - x^*b_0,$$
$$a_2 = b_2 - x^*b_1,$$
$$a_n = b_n - x^*b_{n-1}$$

lo que está en plena concordancia con las igualdades (3) y de este modo demuestra la representación (5).

Ahora bien, llegamos al teorema siguiente.

Teorema del resto (de Bezout). *El resto quedado de la división del polinomio $P_n(x)$ por binomio $x - x^*$ es igual al valor de este polinomio par $x = x^*$.*

Ejemplo 2. Calcular el valor del polinomio $P_4(x) = 12x^4 + 19x^3 - 4$ en el punto $x^* = -2$ y determinar los coeficientes del polinomio $P_3^{(1)}(x)$ que es el cociente obtenido de la división de $P_4(x)$ por el binomio $x + 2$.

△ Hacemos el esquema de Horner para el polinomio dado:

12	19	0	0	−4
	−24	10	−20	40
12	−5	10	−20	36

Ahora bien, $P_4(-2) = 36$ y $P_3^{(1)}(x) = 12x^3 - 5x^2 + 10x - 20$. ▲

De otro ejemplo de utilización del esquema de Horner sirve el problema de cambio de la variable en el polinomio.

Supongamos que en el polinomio (1) se necesita pasar a una nueva variable y ligada con la variable x por la relación lineal $x = y + x^*$ o bien $y = x - x^*$.

Esto quiere decir que es necesario hallar los coeficientes del polinomio

$$P_n(y + x^*) = A_0y^n + A_1y^{n-1} + \ldots + A_{n-1}y + A_n. \tag{6}$$

Se puede mostrar que

$$A_n = P_n(x^*),$$
$$A_{n-1} = P^{(1)}_{n-1}(x^*),$$
$$A_{n-2} = P^{(2)}_{n-2}(x^*),$$
$$\dots\dots\dots$$
$$A_0 = P^{(n)}_0,$$

donde $P_n(x)$ es el polinomio dado (1) y los demás polinomios

$$P^{(k)}_{n-k}(x) = b^{(k)}_0 x^{n-k} + b^{(k)}_1 x^{n-k-1} + \dots + b^{(k)}_{n-k}$$
$$(k = 1,\ 2,\ \dots,\ n)$$

se determinan por las relaciones

$$P_n(x) = (x - x^*)\,P^{(1)}_{n-1}(x) + P_n(x^*),$$
$$P^{(1)}_{n-1}(x) = (x - x^*)\,P^{(2)}_{n-2}(x) + P^{(1)}_{n-1}(x^*),$$
$$\dots\dots\dots\dots\dots\dots\dots\dots$$
$$P^{(n-1)}_1(x) = (x - x^*)\,P^{(n)}_0 + P^{(n-1)}_1(x^*).$$

Ahora bien, se obtiene el siguiente algoritmo simple de determinación de A_j $(j = n,\ n-1,\ \dots,\ 1,\ 0)$:

1°. Con ayuda del esquema de Horner se calcula el coeficiente $A_n = P_n(x^*)$ y junto con éste los coeficientes $b^{(1)}_i$ $(i = 0,\ 1,\ \dots,\ n-1)$ del polinomio $P^{(1)}_{n-1}(x)$.

2°. Aplicando el esquema de Horner a los polinomios $P^{(1)}_{n-1}(x)$, $P^{(2)}_{n-2}(x)$, $\dots$, $P^{(k)}_{n-k}(x)$, se calculan los coeficientes $A_{n-k} = P^{(k)}_{n-k}(x^*)$ y junto con éstos los coeficientes $b^{(k+1)}_i$ $(i = 0,\ 1,\ \dots,\ n-k-1)$ del polinomio $P^{(k+1)}_{n-k-1}(x)$. Este proceso se ejecuta hasta que k llegue a ser igual a $n-1$. Entonces se supone $A_0 = P^{(n)}_0 = b^{(n)}_0 = a_0$ y todos los coeficientes A_j están determinados.

El algoritmo considerado suele llamarse **esquema generalizado de Horner.**

Ejemplo 3. En el polinomio $P_4(x) = 12x^4 + 19x^3 - 4$ pasar a una nueva variable $y = x + 2$.

△ Utilizando las relaciones (5) y (6), hacemos el esquema

12	19	0	0	−4	
	−24	10	−20	40	
12	−5	10	−20	36	$= A_4$
	−24	58	−136		
12	−29	68	−156 $= A_3$		
	−24	106			
12	−53	174 $= A_2$			
	−24				
12	−77 $= A_1$				
$12 = A_0$					

Así, pues el polinomio $Q_4(y) = P_4(y-2)$ tiene la forma

$$Q_4(y) = 12y^4 - 77y^3 + 174y^2 - 156y + 36.\ ▲$$

El cálculo de los valores de una función analítica se basa frecuentemente en su representación en forma de la serie de Taylor rápidamente convergente la cual en muchos casos es un instrumento cómodo para calcular los valores de esta función en los puntos pertenecientes al dominio de convergencia de la serie.

Supongamos que se necesita calcular el valor que la función $f(x)$, analítica sobre el segmento $[a, b]$, tiene en el punto $x = x^*$, $x^* \in [a, b]$ con el valor absoluto dado ε máximamente admisible.

La fórmula de Taylor con el término residual en la forma de Lagrange para la función $f(x)$ en el entorno del punto $x = c$, $c \in \in [a, b]$ tiene la forma

$$f(x) = f(c) + (x-c)\frac{f'(c)}{1!} + \dots$$

$$\dots + (x-c)^n \frac{f^{(n)}(c)}{n!} + (x-c)^{n+1}\frac{f^{(n+1)}(\bar{c})}{(n+1)!},$$

donde $\bar{c} = c + \theta(x - c)$; $0 < \theta < 1$.

Por lo tanto, el problema de cálculo de la $f(x^*)$ se reduce al cálculo de

$$f(x^*) = S_n(x^*) + R_n(x^*),$$

donde $S_n(x^*)$ es la n-ésima suma parcial de la serie:

$$S_n(x^*) = \sum_{i=0}^{n} (x^* - c)^i \frac{f^{(i)}(c)}{i!}$$

$$(0! = 1,\ \ f^{(0)}(c) = f(c))$$

y $R_n(x^*)$, el valor del término residual $R_n(x)$ para $x = x^*$:

$$R_n(x^*) = (x^* - c)^{n+1}\frac{f^{(n+1)}(\bar{c}^*)}{(n+1)!},$$

$$\bar{c}^* = c + \theta^*(x^* - c), \quad 0 < \theta^* < 1.$$

El algoritmo de resolución del problema planteado consiste en lo siguiente:

1°. Sobre el segmento $[a, b]$ se elige un punto $x = c$ que sea, en la medida de lo posible, próximo al punto $x = x^*$ y tal que la misma función $f(x)$ y sus derivadas puedan ser calculadas fácilmente cuando $x = c$.

2°. Se representa ε en la forma de la suma:

$$\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3, \tag{1}$$

donde ε_1 es el error residual (error del método); ε_2, el error absoluto, máximamente admisible, de cálculo de $S_n(x^*)$; ε_3, el error absoluto, máximamente admisible, de redondeo del resultado. Hablando en general, ε_1, ε_2 y ε_3 pueden ser números positivos arbitrarios que satisfagan la condición (1).

Sin embargo, en la práctica ε se asigna de ordinario en la forma $\varepsilon = 10^{-m}$, donde m es un número entero. En este caso suelen tomarse $\varepsilon_3 = 0{,}5 \cdot 10^{-m}$ y $\varepsilon_1 = \varepsilon_2 = 0{,}25 \cdot 10^{-m}$. Si el error de redondeo final falta, se toma $\varepsilon_1 = \varepsilon_2 = 0{,}5 \cdot 10^{-m}$; $\varepsilon_3 = 0$.

3°. Se elige una cantidad de sumandos en S_n de un modo tal que se cumpla la desigualdad

$$|f(x^*) - S_n(x^*)| = |R_n(x^*)| \leqslant \varepsilon_1.$$

4°. Se calcula cada sumando de S_n de un modo tal que el valor aproximado $\overline{S}_n$ se distinga del valor exacto S_n no más que en ε_2. Para esto se calcula habitualmente cada sumando de $\overline{S}_n$ con error absoluto $\varepsilon_2/(n+1)$.

5°. Se redondea la suma aproximada $\overline{S}_n$, obtenida en el subp. 4°, (si $\varepsilon_3 \neq 0$) hasta la magnitud $\overline{\overline{S}}_n$.

6°. Se escribe la solución del problema planteado en la forma $f(x^*) = \overline{\overline{S}}_n \pm \varepsilon$.

Ejemplo. Calcular $e^{2,25}$ con exactitud hasta $\varepsilon = 0{,}01$.

△ La fórmula de Taylor con el término residual en la forma de Lagrange para la función e^x en el entorno del punto $x = 0$ tiene la forma

$$e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \ldots + \frac{x^n}{n!} + e^{\theta x}\frac{x^{n+1}}{(n+1)!}, \quad 0 < \theta < 1.$$

Puesto que al ser grandes x la serie de Taylor de la función e^x converge más lentamente, es racional calcular el valor de $e^{2,25}$ en la forma del producto $e^2 \cdot e^{0,25}$.

La magnitud e^2 puede ser fácilmente calculada con todo grado de exactitud y prácticamente consideramos que el error de su cálculo es igual a cero.

Tomemos ahora que los errores de redondeo y de cálculo de $e^2 \cdot e^{0,25}$ son iguales a 0,005. Entonces el error de cálculo de $e^{0,25}$ constituirá $\bar{\varepsilon} = 0{,}005/e^2 \approx 0{,}0006$.

Ahora bien, hemos reducido el cálculo de $e^{2,25}$ con la exactitud hasta $\varepsilon = 0{,}01$ al cálculo

$$e^{0,25} = 1 + 0{,}25^2 + \frac{0{,}25^2}{2!} + \ldots + \frac{0{,}25^n}{n!} + e^{0,25\theta}\frac{0{,}25^{n+1}}{(n+1)!}$$

con exactitud hasta $\bar{\varepsilon} = 0{,}0006$.

Sea luego $\varepsilon_1 = \varepsilon_2 = 0{,}0003$. Determinemos n, utilizando el término residual de la fórmula de Taylor para e^x si $x = 0{,}25$:

$$e^{0,25\theta}\frac{0{,}25^{n+1}}{(n+1)!} \leqslant 0{,}0003; \quad 0 < \theta < 1.$$

Teniendo en cuenta que $e^{0,25\theta} < e^{0,25} < 1{,}5$, obtenemos $n \geqslant 3$. Por consiguiente, debemos calcular la suma

$$S_3 = 1 + 0{,}25 + \frac{0{,}25^2}{2!} + \frac{0{,}25^3}{3!}$$

con error absoluto 0,0003. Puesto que los dos primeros sumandos ya están calculados con exactitud absoluta, es suficiente calcular los sumandos dos y tres con error absoluto máximo 0,0001, cada uno. Realizando los cálculos necesarios, hallamos

$$\overline{S}_3 = 1 + 0{,}25 + 0{,}0312 + 0{,}0026 = 1{,}2838.$$

Multiplicando el resultado obtenido por e^2, obtenemos $e^2\overline{S}_3 = 9{,}4860\ldots$ Por último, redondeando hasta las centésimas partes, tenemos $e^{2,25} = 9{,}49 \pm 0{,}01$. ▲

De un modo análogo se calculan los valores de las funciones trigonométricas (del seno y del coseno). Las fórmulas respectivas de Taylor con el término residual en la forma de Lagrange se escriben así:

$$\operatorname{sen} x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots$$

$$\ldots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + (-1)^n \frac{x^{2n+1}}{(2n+1)!}\cdot\theta,\ 0<\theta<1; \tag{2}$$

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \ldots$$

$$\ldots + (-1)^{n-1}\frac{x^{2n-2}}{(2n-2)!} + (-1)^n \frac{x^{2n}}{(2n)!}\cdot\theta,\ 0<\theta<1. \tag{3}$$

En las fórmulas (2) y (3) el argumento debe ser expresado en radianes y pertenecer al segmento $[0, \pi/4]$. En este caso la convergencia de las series respectivas será bastante rápida. Hacer que el argumento pertenezca al segmento indicado se puede con ayuda de las fórmulas de reducción y de la relación conocida entre las fórmulas trigonométricas.

Así, por ejemplo, para calcular el valor de sen 2,53 hace falta utilizar la fórmula de reducción $\operatorname{sen} 2{,}53 = \operatorname{sen}(\pi - 2{,}53)$; el argumento $\pi - 2{,}53$ pertenece al segmento $[0, \pi/4]$. En cambio, para calcular el valor de cos 1,27 conviene hacer uso de la fórmula $\cos 1{,}27 = \operatorname{sen}(\pi/2 - 1{,}27)$ y calcular el valor obtenido del seno, puesto que $(\pi/2 - 1{,}27) \in [0, \pi/4]$.

Al calcular el valor del seno, la condición para la elección de la cantidad de sumandos n se escribirá así:

$$\frac{x^{2n+1}}{(2n+1)!} \leqslant \varepsilon_1;\ x \in [0, \pi/4], \tag{4}$$

y al calcular el valor del coseno, dicha condición se escribirá:

$$\frac{x^{2n}}{(2n)!} \leqslant \varepsilon_1;\ x \in [0, \pi/4]. \tag{5}$$

Para calcular los valores de la función logarítmica la fórmula de desarrollo en serie

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \ldots +$$

$$+ (-1)^{n-1}\frac{x^n}{n} + \ldots;\ -1 < x \leqslant 1 \tag{6}$$

es poco útil debido a la convergencia lenta cuando $|x|$ es próximo a la unidad. Además, esta fórmula no permite calcular los logaritmos de números mayores que dos.

Para acelerar la convergencia y ampliar el dominio de aplicación transformemos la fórmula (6) del modo siguiente. Reemplazando en sus miembros segundo y primero x por $-x$, tenemos

$$\ln(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n} - \ldots$$

Sustrayendo de la igualdad obtenida la inicial, hallamos

$$\ln\frac{1-x}{1+x} = -2\left(x + \frac{x^3}{3} + \frac{x^5}{5} + \ldots + \frac{x^{2n-1}}{2n-1} + \ldots\right).$$

Suponiendo $(1-x)/(1+x) = z$ y teniendo en cuenta que en este caso $x = (1-z)/(1+z)$, obtenemos la fórmula inicial para calcular el logaritmo natural de todo número z contenido en el intervalo $(0, \infty)$:

$$\ln z = -2\left[\frac{1-z}{1+z} + \frac{1}{3}\left(\frac{1-z}{1+z}\right)^3 + \frac{1}{5}\left(\frac{1-z}{1+z}\right)^5 + \ldots\right]. \qquad (7)$$

En los cálculos prácticos el número positivo x cuyo logaritmo natural ha de determinarse se representa cómodamente en la forma siguiente:

$$x = 2^m \cdot z, \qquad (8)$$

donde m es cierto número entero y

$$0{,}5 \leqslant z < 1. \qquad (9)$$

Entonces

$$\ln x = m \ln 2 + \ln z.$$

El primer sumando se calcula fácilmente si se conoce m y $\ln 2 = 0{,}69314718\ \ldots$, y el segundo sumando se puede hallar con ayuda de la fórmula (7). En este caso en virtud de la desigualdad (9) la magnitud

$$\xi = (1-z)/(1+z) \qquad (10)$$

varía dentro de los límites

$$0 < \xi \leqslant 1/3 \qquad (11)$$

lo que facilita la convergencia rápida de la serie (7).

Así pues, finalmente tenemos

$$\ln x = m\ln 2 - 2\left(\xi + \frac{\xi^3}{3} + \frac{\xi^5}{5} + \ldots + \frac{\xi^{2n-1}}{2n-1}\right) - R_n. \qquad (12)$$

Estimemos el resto:

$$R_n = 2\left(\frac{\xi^{2n+1}}{2n+1} + \frac{\xi^{2n+3}}{2n+3} + \ldots\right) < \frac{2}{2n+1}\cdot\frac{\xi^{2n+1}}{1-\xi^2}.$$

Utilizando la desigualdad (11), esta estimación puede ser transformada en una forma cómoda para compararla con los términos correspondientes de la serie (7):

$$R_n < \frac{9}{4}\,\frac{\xi^{2n+1}}{2n+1} \leqslant \frac{\xi^{2n-1}}{4\,(2n+1)}. \tag{13}$$

Para calcular el valor del logaritmo decimal es necesario hacer uso de la fórmula

$$\log x = M \ln x,$$

donde $M = \log e = 0{,}434294481903252\ldots$

§ 4.3. Método iterativo de cálculo de los valores de las funciones

Consideremos un procedimiento artificial más para calcular el valor de una función continua sobre el segmento $[a, b]$

$$y = f\,(x) \tag{1}$$

en el punto $x = x^*$; $x^* \in [a, b]$.

Este procedimiento está basado en la utilización del método de Newton para resolver las ecuaciones algebraicas y trascendentes y consiste en lo siguiente:

1°. La función (1) se escribe en la forma implícita y se sustituye en vez de x su valor x^* en la expresión obtenida:

$$F\,(x^*,\ y) = 0. \tag{2}$$

De solución de esta ecuación sirve precisamente el valor buscado de la función $y^* = f\,(x^*)$.

2°. La ecuación (2) se resuelve por el método de Newton para lo cual se elige la aproximación inicial y_0 de un modo tal que se cumpla la condición

$$F\,(x^*,\ y_0)\, F''_{yy}\,(x^*,\ y_0) > 0 \tag{3}$$

y cada valor aproximado sucesivo y_n $(n = 1, 2, 3, \ldots)$ se calcula con ayuda de la fórmula

$$y_n = y_{n-1} - \frac{F\,(x^*,\ y_{n-1})}{F'_y\,(x^*,\ y_{n-1})}\ (n = 1,\ 2,\ 3,\ \ldots). \tag{4}$$

La convergencia del proceso iterativo (4) quedará asegurada si la función $F\,(x^*, y)$ satisface las condiciones de convergencia del método de Newton.

Nótese que la función (1) puede representarse en la forma implícita (2) por un conjunto infinito de métodos. Entre todos estos métodos ha de elegirse tal que el proceso iterativo (4) sea convergente y la velocidad de convergencia sea bastante grande.

Como ejemplo importante de utilización del procedimiento indicado sirve el cálculo del valor de la función $y = \sqrt[m]{x}$ $(m = 2, 3, \ldots)$ en el intervalo $(0, \infty)$.

Para esta función en calidad de $F(x^*, y)$ es conveniente utilizar la expresión $y^m - x$. Entonces la condición (3) se transforma en la forma

$$y_0 > \sqrt[m]{x^*} \tag{3'}$$

y el proceso iterativo (4), en la forma

$$y_n = y_{n-1} - \frac{y_{n-1}^m - x^*}{m y_{n-1}^{m-1}}. \tag{4'}$$

Nótese que en virtud de las propiedades de la función $y^m - x^*$ el proceso iterativo (4′) converge no sólo en caso de cumplimiento de la condición (3′), sino también en caso de toda aproximación inicial positiva ($y_0 > 0$, con la particularidad de que para todas las aproximaciones sucesivas se cumplirá la condición $y_n > \sqrt[m]{x^*}$ ($n = 1, 2, \ldots$).

El error del valor aproximado y_n puede ser estimado del modo siguiente:

$$\Delta y = y_n - \sqrt[m]{x^*} < \sqrt{\frac{y_{n-1}^m}{x^*}}\,(y_{n-1} - y_n)$$

o bien

$$\Delta y_n = y_n - \sqrt[m]{x^*} < \frac{m-1}{2}\cdot\frac{y_{n-1}^{m-1}}{x^*}\,(y_{n-1} - y_n)^2. \tag{5}$$

Puesto que el caso $m = 2$ es mucho más frecuente que los demás, demos para él una estimación más exacto del error.

Representemos x^* en la forma $x^* = 2^k \bar{x}$, donde k es un número entero y $\bar{x}^* \in [0{,}5;\ 1)$. Entonces, suponiendo $y_0 = 2^{E(k/2)}$, obtenemos

$$\Delta y_n = y_n - \sqrt{x^*} < \frac{25}{12}\, y_1 \left(\frac{1}{5}\right)^{2^n} \leqslant \frac{25}{8}\, y_0 \left(\frac{1}{5}\right)^{2^n}.$$

Ejemplo. Calcular $\sqrt[3]{34}$ con exactitud hasta $\varepsilon = 10^{-4}$.

△En consonancia con la desigualdad (3′) elegimos $y_0 = 3{,}4 > \sqrt[3]{34}$. Luego, utilizando la fórmula (4′), para $x^* = 34$ e $y_0 = 3{,}4$ hallamos sucesivamente y_n y con ayuda de la fórmula (5) calculamos Δy_n:

$$y_1 = 3{,}4 - \frac{3{,}4^3 - 34}{3\cdot 3{,}4^2} = 3{,}247;\ \ \Delta y_1 = 0{,}01;$$

$$y_2 = 3{,}247 - \frac{3{,}247^3 - 34}{3\cdot 3{,}247^2} = 3{,}23964;\ \ \Delta y_2 = 2\cdot 10^{-5}.$$

Así pues, $\sqrt[3]{34} = 3{,}2396 \pm 0{,}0001$. ▲

Ejercicios

1. Haciendo uso del esquema de Horner, dividir el polinomio $P_5(x) = x^5 + 3x^4 - 2x^3 - x + 1$ por el binomio $x - 3$.

2. ¿Es $\xi = 1$ la raíz de la ecuación $x^3 + 2x^2 - 3 = 0$?

3. Haciendo uso del desarrollo en serie de potencias, calcular con exactitud hasta $\varepsilon = 0{,}001$: a) sen 25°; b) cos 20°; c) ln 4; d) $\sqrt[5]{e}$; e) cos 36°; f) sen 18°.

4. Haciendo uso del método iterativo, calcular con exactitud hasta $\varepsilon = 0{,}0005$: a) $\sqrt{12}$; b) $\sqrt{56}$; c) $\sqrt{42}$.

CAPITULO V

Métodos de resolución de las ecuaciones no lineales

§ 5.1. Ecuaciones algebraicas y trascendentes

En los problemas prácticos se necesita con frecuencia resolver las ecuaciones. Toda ecuación con una incógnita puede ser representada en la forma

$$\varphi(x) = g(x), \tag{1}$$

donde $\varphi(x)$ y $g(x)$ son las funciones dadas, definidas sobre cierto conjunto numérico X llamado *dominio de valores admisibles de la ecuación*.

La ecuación con una incógnita puede escribirse en la forma

$$f(x) = 0. \tag{2}$$

En efecto, transponiendo $g(x)$ al primer miembro de la ecuación (1), tenemos la ecuación $\varphi(x) - g(x) = 0$ equivalente a (1). Si designamos el primer miembro de la última ecuación con $f(x)$, obtenemos la ecuación (2).

El conjunto de los valores de la variable x con los cuales la ecuación (1) se transforma en identidad se llama *solución* de esta ecuación y cada valor x de este conjunto se denomina *raíz* de la ecuación.

Por ejemplo, la ecuación $x^2 = 2 - x$ tiene las raíces $x_1 = -2$ y $x_2 = 1$. Sustituyendo -2 y 1 en la ecuación dada en vez de x, obtenemos las identidades: $(-2)^2 = 2 - (-2)$, o sea, $4 \equiv 4$; $1^2 = 2 - 1$, o sea, $1 \equiv 1$.

Resolver una ecuación quiere decir hallar el conjunto de todas las raíces de esta ecuación. Este conjunto puede ser finito o infinito. Así, la ecuación recién considerada tiene dos raíces. La ecuación $\operatorname{sen} x = 0$ tiene la solución $x = \pi n$ ($n = 0, \pm 1, \pm 2, \ldots$). Asignando a n distintos valores, obtenemos un conjunto infinito de raíces.

El conjunto de varias ecuaciones con varias incógnitas se llama *sistema de ecuaciones* (la incógnita designada con la misma letra en cada una de las ecuaciones debe denotar la misma magnitud incógnita).

Se llama *solución de un sistema de ecuaciones* con varias incógnitas el conjunto de los valores de estas incógnitas el cual convierte en identidad cada ecuación del sistema.

Por ejemplo, el sistema

$$\begin{cases} x^2 + y = 5, \\ x + y^2 = 3 \end{cases}$$

tiene la solución $x = 2$, $y = 1$, ya que con estos valores de las incógnitas las ecuaciones del sistema se convierten en identidades: $4 + 1 \equiv 5$, $2 + 1 \equiv 3$.

Resolver un sistema de ecuaciones significa hallar el conjunto de todas sus soluciones o mostrar que este sistema no tiene soluciones.

De todo el conjunto de las ecuaciones se destaca con frecuencia la clase de ecuaciones algebraicas que poseen varias particularidades características que se utilizan al resolverlas. Se denomina ecuación algebraica con una incógnita a la ecuación que tiene la forma

$$a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \dots + a_{n-1}x + a_n = 0. \quad (3)$$

Aquí n es el número positivo entero llamado grado de la ecuación.

De otro tipo frecuente de ecuaciones sirven las ecuaciones irracionales en las cuales las variables incógnitas se hallan bajo el signo de radical.

Por ejemplo: $3x^2 - 4x + \sqrt[3]{x - 1} = 0$.

Al mismo tiempo la ecuación $\sqrt{1 + \sqrt{5}}\, x^2 + x\sqrt{2} - 3\sqrt{7} = 0$ no es irracional, puesto que x no está bajo el signo de radical.

Si la ecuación (3) está obtenida transformando cualquier otra ecuación, por ejemplo la irracional, es necesario tener en cuenta que las funciones que forman parte de la ecuación inicial pueden ser definidas no en el todo eje numérico como esto tiene lugar en la ecuación algebraica.

Por ejemplo, la ecuación

$$\sqrt{x - 2} + \sqrt{6 - x} = 3$$

después de suprimir la irracionalidad tomará la forma

$$4x^2 - 32x - 73 = 0.$$

Sin embargo, la ecuación inicial está definida no en todo el eje numérico sino para las x que pertenecen al segmento $[2, 6]$.

Los números $a_0, a_1, \dots, a_n$ se denominan *coeficientes* de la ecuación (3), pueden ser tanto reales como complejos. A continuación se considerarán ecuaciones algebraicas de la forma (3) sólo en caso de tener coeficientes reales.

Las ecuaciones que contienen funciones trascendentes: función exponencial a^x, función logarítmica $\log_a x$, funciones trigonométricas $\operatorname{sen} x$, $\cos x$, $\operatorname{tg} x$, $\operatorname{ctg} x$, funciones trigonométricas inversas y otras, se reúnen bajo el nombre común de ecuaciones trascendentes las cuales junto con las ecuaciones algebraicas cuyo grado supera el primero suelen llamarse ecuaciones no lineales.

La resolución de ecuación con una incógnita consiste en la determinación de las raíces, o sea, los valores de x que convierten la ecuación en identidad. Las raíces de una ecuación pueden ser reales y no reales (complejas).

Se puede hallar los valores exactos de las raíces de una ecuación sólo en los casos excepcionales, por lo general, cuando existe cualquier fórmula simple para calcular los valores de las raíces, tal que permita expresarlos por las magnitudes conocidas.

Así, para determinar las raíces de la ecuación cuadrática que tiene la forma $x^2 + px + q = 0$ se utiliza la fórmula

$$x_{1,2} = -\frac{p}{2} \pm \sqrt{\frac{p^2}{4} - q}. \tag{4}$$

Para resolver la ecuación cúbica que tiene la forma $x^3 + px + q = 0$ se emplea la fórmula

$$x = \sqrt[3]{-\frac{q}{2} + \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}} + \sqrt[3]{-\frac{q}{2} - \sqrt{\frac{q^2}{4} + \frac{p^3}{27}}}. \tag{5}$$

No obstante, la aplicación práctica de esta fórmula presenta muchas dificultades y se necesita utilizar números complejos.

Para resolver una ecuación de cuarto grado también existe el algoritmo, pero es tan complicado que prácticamente no se emplea y no vamos a considerarlo.

El matemático noruego Abel demostró que para $n \geqslant 5$ no existe ninguna fórmula que exprese la solución de la ecuación algebraica (3) a través de sus coeficientes con ayuda de las operaciones aritméticas y de la extracción de las raíces. Únicamente para algunos casos particulares de las ecuaciones algebraicas cuyo grado es más de cuatro pueden existir las fórmulas de solución.

Además, los coeficientes de algunas ecuaciones son números aproximados y, por consiguiente, la cuestión acerca de la determinación de las raíces exactas no puede ser planteada de ningún modo.

Por eso adquieren gran importancia los métodos de cálculo aproximado de las raíces de la ecuación $f(x) = 0$.

Al resolver muchos problemas prácticos la solución exacta de una ecuación no siempre es necesaria. El problema de determinación de las raíces se considera resuelto si las raíces están calculadas con el grado prefijado de precisión.

Entonces ¿cómo es necesario entender la afirmación «la raíz está calculada con el grado prefijado de precisión»? Sea ξ la raíz de la ecuación y $\bar{x}$, su valor aproximado con exactitud hasta ε; esto quiere decir que $|\xi - \bar{x}| \leqslant \varepsilon$. Si se ha determinado que la raíz buscada ξ está encerrada entre los números a y b, o sea $a < \xi < b$, con la particularidad de que $b - a \leqslant \varepsilon$, entonces los números a y b son valores aproximados de la raíz ξ, con defecto y con exceso, respectivamente, con exactitud hasta ε, ya que $|\xi - a| < b - a \leqslant \varepsilon$ y $|\xi - b| < b - a \leqslant \varepsilon$. Por valor aproximado de la raíz ξ con exactitud hasta ε se puede tomar todo número comprendido entre a y b.

En la práctica, siempre que esté determinado que la raíz ξ pertenece al segmento $[a, b]$, por su valor aproximado se toma el centro de este segmento y en este caso el error no superará la mitad de la longitud del segmento $\xi = \frac{a+b}{2} \pm \frac{b-a}{2}$.

Así, por ejemplo, si la raíz ξ está encerrada entre 23,07 y 23,10, o sea $23{,}07 \leqslant \xi \leqslant 23{,}10$, por valor aproximado de la raíz se toma $\frac{23{,}07+23{,}10}{2} = 23{,}085$ y el error de este valor aproximado $\frac{23{,}10-23{,}07}{2} = 0{,}015$. Así que $\xi = 23{,}085 \pm 0{,}015$.

En este capítulo consideraremos los métodos de resolución aproximada de las ecuaciones y de los sistemas de ecuaciones. Algunos de estos métodos son igualmente aplicable a la determinación de las raíces tanto de las ecuaciones trascendentes como algebraicas. Otros métodos son aplicables únicamente a las ecuaciones algebraicas.

§ 5.2. Separación de las raíces

El proceso de determinación de los valores aproximados de las raíces de ecuaciones se subdivide en dos etapas: 1) separación de las raíces; 2) determinación más exacta de las raíces hasta el grado prefijado de precisión.

En este párrafo consideramos la primera etapa, o sea, la separación de las raíces.

La raíz ξ de la ecuación $f(x) = 0$ se considera *separada* sobre el segmento $[a, b]$ si en este segmento la ecuación $f(x) = 0$ no tiene otras raíces.

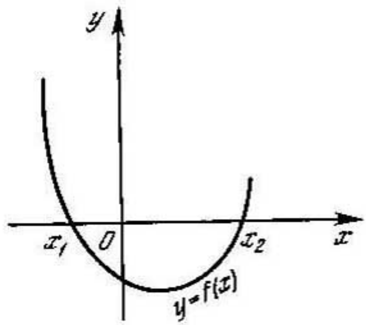


Fig. 5.1

Separar las raíces quiere decir partir todo el dominio de los valores admisibles en segmentos en cada uno de los cuales se contiene una sola raíz. La separación de las raíces puede realizarse por dos métodos: el gráfico y el analítico.

Método gráfico de separación de las raíces. *Procedimiento I.* Las raíces se separan fácilmente si está construido el gráfico de la función $y = f(x)$. Los puntos de intersección del gráfico con el eje Ox dan los valores de la raíz y por el gráfico no es difícil determinar dos números a y b entre los cuales está encerrada una sola raíz (fig. 5.1).

Procedimiento II. Todos los términos de la ecuación se dividen en dos grupos uno de los cuales se escribe en el primer miembro de la ecuación y el otro, en el segundo miembro, o sea, la representan en la forma de $\varphi(x) = g(x)$. Luego se trazan los gráficos de dos funciones $y = \varphi(x)$ e $y = g(x)$. Las abscisas de los puntos de intersección de estas dos funciones sirven precisamente de raíces de la ecuación dada. Supongamos que el punto de intersección de los gráficos tiene la abscisa x_0, las ordenadas de ambos gráficos en este punto son iguales entre sí, o sea, $\varphi(x_0) = g(x_0)$. De esta igualdad se deduce que x_0

es la raíz de la ecuación (fig. 5.2). Por el gráfico se determinan dos números a y b entre los cuales está encerrada la raíz.

Ejemplo 1. Determinar gráficamente entre qué números enteros están encerradas las raíces de la ecuación $x^3 - 3x - 1 = 0$.

△ *Procedimiento I.* Construyamos el gráfico de la función $y = x^3 - 3 \quad - 1$ (fig. 5.3) y determinemos las abscisas de los puntos de intersección de este gráfico con el eje Ox. La curva corta el eje Ox en tres puntos; por lo tanto, la ecuación tiene tres raíces reales (nótese que la ecuación algebraica de tercer grado tiene ora una raíz

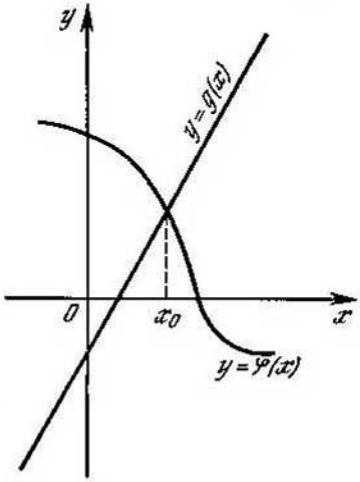


Fig. 5.2

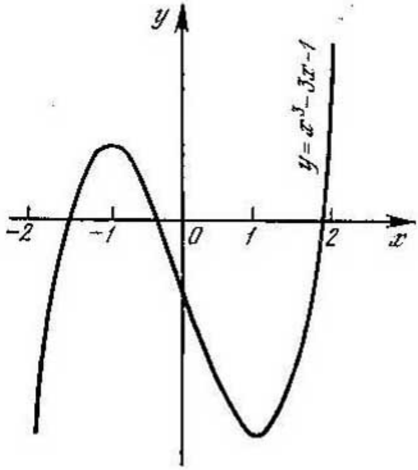


Fig. 5.3

real ora tres raíces reales). Del dibujo se ve que las raíces pertenecen a los segmentos $[-2, -1]$, $[-1, 0]$, $[1, 2]$.

Procedimiento II. Representemos la ecuación dada en la forma $x^3 = 3x + 1$ y tracemos los gráficos de las funciones $y = x^3$ e $y = 3x + 1$ (fig. 5.4). Las abscisas de los puntos de intersección de los gráficos de estas funciones son raíces de la ecuación dada. Del dibujo es fácil determinar los intervalos de aislamiento de las raíces. ▲

Ejemplo 2. Separar gráficamente las raíces de la ecuación $\log x - 3x + 5 = 0$.

△ Escribamos la ecuación del modo siguiente: $\log x = 3x - 5$. Las funciones en los miembros primero y segundo de la ecuación tienen el dominio común de definición: el intervalo $0 < x < +\infty$. Por eso buscaremos las raíces precisamente en este intervalo.

Trazamos los gráficos de las funciones $y = \log x$ e $y = 3x - 5$ (fig. 5.5). La recta $y = 3x - 5$ corta la curva logarítmica en dos puntos. En el dibujo es difícil mostrar la intersección de los gráficos de estas funciones en el primer punto; no obstante, teniendo en cuenta que la rama inferior de la curva logarítmica se aproxima ilimitadamente al eje Oy, se puede suponer que los gráficos se cortarán cerca del punto de intersección del gráfico de la función $y = 3x - 5$ y el eje Oy. Así pues, las raíces se sitúan sobre los segmentos $[0; 0{,}5]$ y $[1, 2]$. ▲

Ejemplo 3. Separar gráficamente las raíces de la ecuación $x - \cos x = 0$.

△ Escribamos la ecuación en la forma $x = \cos x$ y construyamos los gráficos de las funciones $y = x$ e $y = \cos x$ en el intervalo

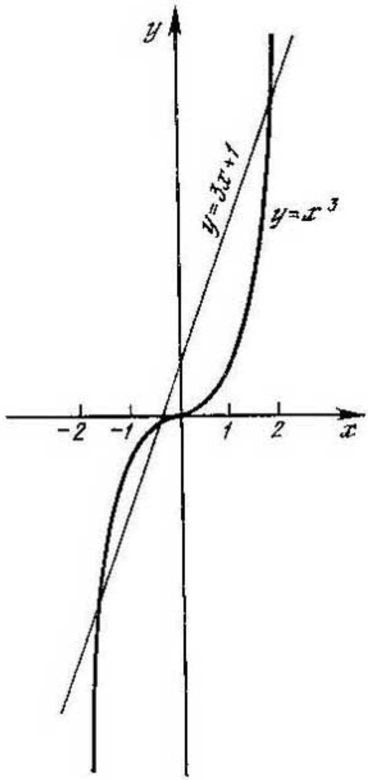


Fig. 5.4

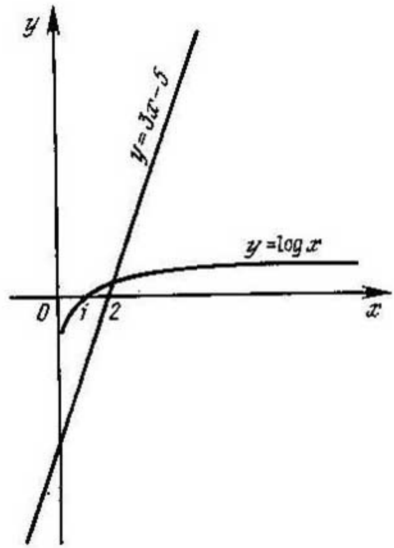


Fig. 5.5

$-\pi/2 \leqslant x \leqslant \pi/2$. Los gráficos de las funciones se cortan en un punto; en el intervalo dado la ecuación tiene la única raíz. Teniendo en cuenta las propiedades de las funciones $y = x$ e $y = \cos x$, se puede cerciorarse de que fuera de este intervalo la ecuación dada no tiene raíces. En caso de una construcción más exacta del dibujo determinamos que la raíz se sitúa en el segmento $[0{,}6; 0{,}8]$ (fig. 5.6). ▲

Ejemplo 4. Separar gráficamente por dos procedimientos las raíces de la ecuación $2^x - 5x - 3 = 0$.

△ *Procedimiento I.* Construyamos el gráfico de la función $y = 2^x - 5x - 3$ y determinemos las abscisas de los puntos de intersección de este gráfico con el eje Ox. La curva corta el eje Ox en dos puntos; por consiguiente, la ecuación dada tiene dos raíces reales. Del dibujo se ve que las raíces se sitúan en los segmentos $[-1, 0]$ y $[4, 5]$ (fig. 5.7).

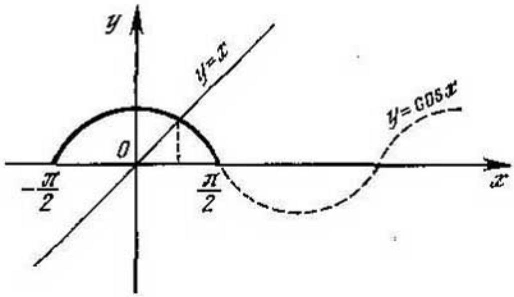


Fig. 5.6

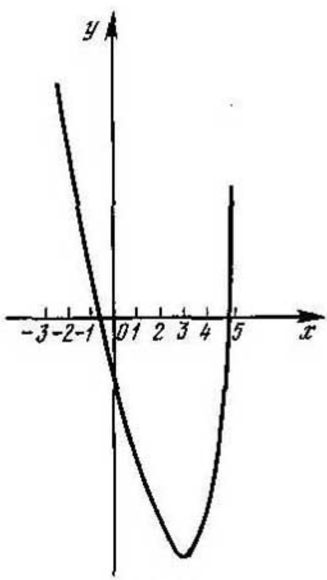


Fig. 5.7

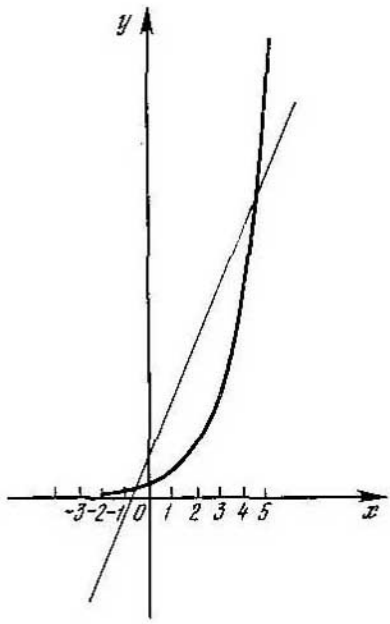


Fig. 5.8

Procedimiento II. Representemos la ecuación dada en la forma $2^x = 5x + 3$ y tracemos los gráficos de las funciones $y = 2^x$ e $y = 5x + 3$. Determinemos las abscisas de los puntos de intersección de los gráficos de estas funciones; estas abscisas son las raíces de la ecuación dada. Obtenemos los mismos intervalos de aislamiento de las raíces $[-1, 0]$ y $[4, 5]$ (fig. 5.8). ▲

Observación. Supongamos que el gráfico de la función $y = f(x)$ tiene la forma representada en la fig. 5.9. La curva corta tres veces el eje de abscisas; por lo tanto, la ecuación tiene tres raíces simples.

En cambio, si la curva toca el eje de abscisas (fig. 5.10), la ecuación tiene una raíz doble. Por ejemplo, la ecuación $x^3 - 3x + 2 = 0$ tiene tres raíces: $x_1 = -2$; $x_2 = x_3 = 1$ (fig. 5.11).

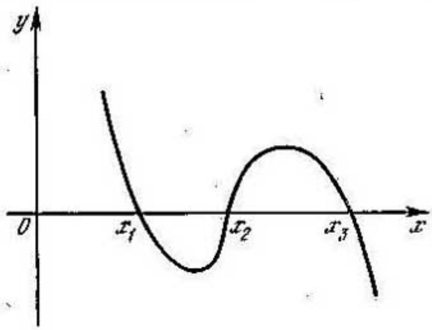


Fig. 5.9

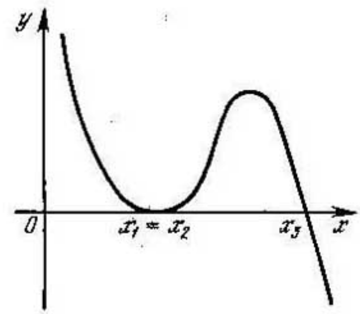


Fig. 5.10

No obstante, si la ecuación tiene una raíz real triple, en el lugar de contacto con el eje la curva $y = f(x)$ tiene un punto de inflexión (fig. 5.12). Por ejemplo, la ecuación $x^3 - 3x^2 + 3x - 1 = 0$ tiene la raíz triple igual a la unidad (fig. 5.13).

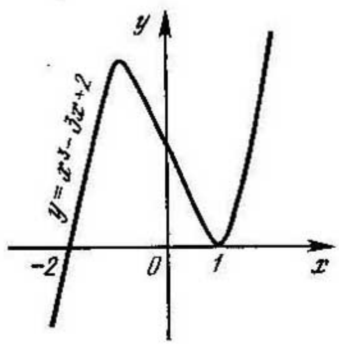


Fig. 5.11

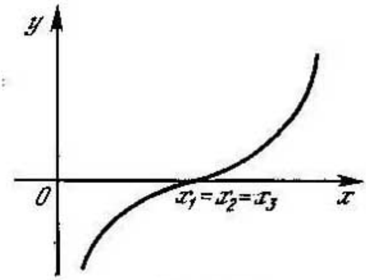


Fig. 5.12

El método gráfico de separación de las raíces no posee gran precisión. Ofrece la posibilidad de determinar aproximadamente los intervalos de aislamiento de las raíces. Luego las raíces se precisan, haciendo uso de los procedimientos indicados posteriormente.

Método analítico de separación de las raíces. Analíticamente las raíces de la ecuación $f(x) = 0$ pueden ser separadas utilizando algunas propiedades de las funciones, propiedades que se estudian en el curso de análisis matemático.

Exponemos ahora algunas nociones contenidas en el análisis matemático las cuales necesitaremos posteriormente.

Si la función $f(x)$ se da analíticamente, entonces se llama *dominio de existencia* (*dominio de definición*) *de la función* el conjunto de todos aquellos valores reales del argumento con los cuales la expresión analítica que define la función no pierde sentido numérico y toma sólo los valores reales.

La función $y = f(x)$ se denomina monótona en el intervalo dado si para todos valores $x_2 > x_1$ contenidos en este intervalo satisface la condición $f(x_2) \geqslant f(x_1)$ (función no decreciente) o bien $f(x_2) \leqslant f(x_1)$ (*función no creciente*). En cambio, si para todos valores $x_2 > x_1$ pertenecientes al intervalo dado $f(x_2) > f(x_1)$ (función creciente) o bien $f(x_2)\ f(x_1)$ (función decreciente), entonces la función $f(x)$ se denomina estrictamente monótona.

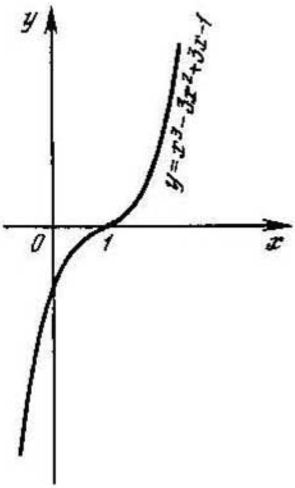


Fig. 5.13

Si la función continua $y = f(x)$ tiene la derivada en todos los puntos interiores del intervalo dado, la condición necesaria y suficiente de monotonía de la función en este intervalo es el cumplimiento de la desigualdad $f'(x) \geqslant 0$ o $f'(x) \leqslant 0$.

Supongamos que sobre el segmento $[a, b]$ la función $f(x)$ es continua y toma en los extremos del segmento los valores de signos contrarios, y la derivada $f'(x)$ mantiene un signo constante sobre el intervalo (a, b). Entonces *si en todos los puntos del intervalo* (a, b) *la primera derivada es positiva, o sea,* $f'(x) > 0$, *la función* $f(x)$ *en este intervalo crece*: $f(x_2) > f(x_1)$ *para* $x_2 > x_1$ (fig. 5.14, *a*, *c*).

En cambio, *si en todos los puntos del intervalo* (a, b) *la primera derivada es negativa*, o sea $f'(x) < 0$, *la función en este intervalo decrece*: $f(x_2) < f(x_1)$ para $x_2 > x_1$ (fig. 5.14, *b*, *d*). De raíz de la función sirve la abscisa del punto de intersección del gráfico de la función $f(x)$ con el eje Ox.

Supongamos que sobre el segmento $[a, b]$ la función $f(x)$ tiene la derivada de segundo orden que conserva un signo constante en todo el segmento. Entonces *si* $f''(x) > 0$, *el gráfico de la función es convexo hacia abajo* (fig. 5.14, *a*, *d*); en cambio, *si* $f''(x) < 0$, *el gráfico de la función es convexo hacia arriba* (fig. 5.14, *b*, *c*).

Los puntos en los cuales la primera derivada de la función es igual a cero, así como también tales en los cuales ella no existe (por ejemplo, se convierte en infinito), pero la función conserva su continuidad, se llaman *críticos* (este criterio es criterio necesario del extremo).

Si la función $f(x)$ es continua sobre el segmento $[a, b]$, entonces en este segmento siempre hay puntos en los cuales ella toma los valores máximo y mínimo. La función alcanza estos valores ora en los puntos críticos ora en los extremos del segmento. Por consiguiente,

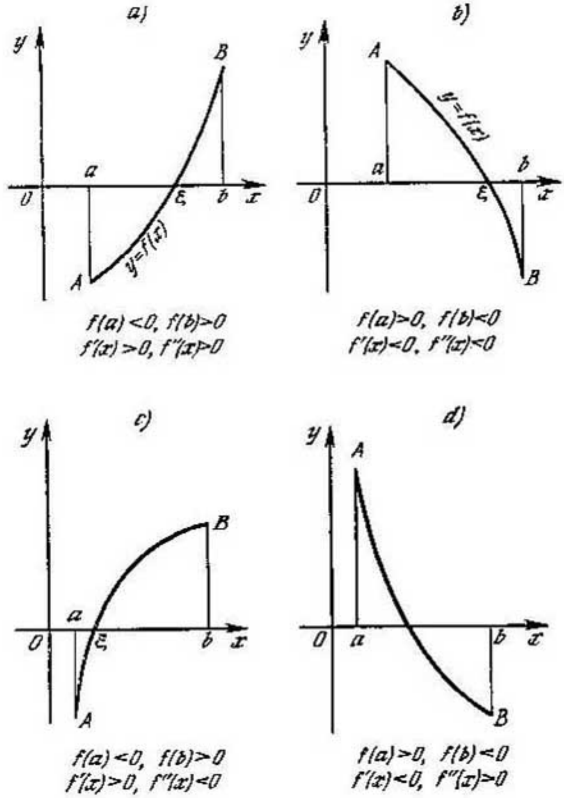


Fig. 5.14

para determinar los valores máximo y mínimo de la función sobre el segmento es necesario: 1) determinar los puntos críticos de la función; 2) calcular los valores de la función en los puntos críticos y en los extremos del segmento $[a, b]$; 3) el mayor de los valores hallados en el subpárr. 2 será valor máximo de la función en el segmento y el menor será valor mínimo de la misma.

Enunciemos, sin demostrarlos, los teoremas cuyo conocimiento es necesario al separar las raíces.

Teorema 1. *Si la función $f(x)$ es continua sobre el segmento $[a, b]$ y toma en los extremos de este segmento los valores de signos opuestos,* *dentro del segmento* $[a, b]$ *existe, al menos, una raíz de la ecuación* $f(x) = 0$.

Teorema 2. *Si la función* $f(x)$ *es continua y estrictamente monótona sobre el segmento* $[a, b]$ *y toma en los extremos del mismo los valores de signos opuestos, entonces dentro del segmento* $[a, b]$ *hay una raíz de la ecuación* $f(x) = 0$ *y esta raíz es la única.*

Teorema 3. *Si la función* $f(x)$ *es continua sobre el segmento* $[a, b]$ *y toma en sus extremos los valores de signos opuestos y la derivada* $f'(x)$ *existe y conserva un signo constante dentro del segmento, o sea* $f'(x) > 0$ *o bien* $f'(x) < 0$, *entonces dentro del segmento existe una raíz de la ecuación* $f(x) = 0$ *y esta raíz es la única.*

De acuerdo con lo expuesto anteriormente se puede recomendar el orden siguiente de operaciones para separar las raíces valiéndose del método analítico:

1) se halla $f'(x)$, o sea, la primera derivada;

2) se hace la tabla de signos de la función $f(x)$ suponiendo x igual: a) a los valores críticos (raíces) de la derivada o a los valores próximos a estos últimos; b) a los valores de frontera (partiendo del dominio de los valores admisibles de la incógnita);

3) se determinan los intervalos en cuyos extremos la función toma los valores de signos contrarios. Dentro de estos intervalos se contiene, cada vez, una y una sola raíz.

Ejemplo 5. Separar las raíces de la ecuación $2^x - 5x - 3 = 0$ con ayuda del método analítico.

△ Designemos $f(x) = 2^x - 5x - 3$. El dominio de definición de la función $f(x)$: todo el eje numérico. Determinemos la primera derivada $f'(x) = 2^x \ln 2 - 5$.

Igualamos a cero esta derivada y calculamos la raíz:

$$2^x \ln 2 - 5 = 0; \quad 2^x \ln 2 = 5; \quad 2^x = 5/\ln 2;$$

$$x \log 2 = \log 5 - \log \ln 2;$$

$$x = \frac{\log 5 - \log \ln 2}{\log 2} = \frac{0,6990 + 0,1592}{0,3010} = \frac{0,8582}{0,3010} = 2,85.$$

Hacemos la tabla de signos de la función $f(x)$ suponiendo x igual: a) a los valores críticos (raíces de la derivada) o a los valores próximos a estos últimos; b) a los valores de frontera (partiendo del dominio de los valores admisibles de la incógnita):

x	$-\infty$	2	3	$+\infty$
Signo de $f(x)$	+	−	−	+

La ecuación tiene dos raíces, ya que ocurren dos cambios de signo de la función. Hagamos una nueva tabla con intervalos más pequeños de aislamiento de la raíz:

x	-1	0	1	2	3	4	5
Signo de $f(x)$	+	−	−	−	−	−	+

Las raíces de la ecuación están encerradas en los intervalos $(-1, 0)$ y $(4, 5)$. ▲

§ 5.3. Determinación máx exacta de las raíces. Método de pruebas

El párrafo precedente ha sido dedicado a la primera de las etapas de resolución aproximada de las ecuaciones algebraicas y trascendentes, o sea, a la separación de las raíces.

La segunda etapa consiste en determinar más exactamente las raíces, es decir, en llevarlas al grado prefijado de precisión. En este párrafo consideramos algunos procedimientos para precisar las raíces los cuales se emplean al resolver las ecuaciones algebraicas y tras-

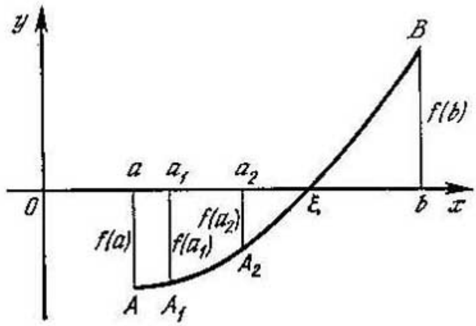


Fig. 5.15

cendentes. No obstante, existen los procedimientos que son aplicables sólo para la resolución de las ecuaciones algebraicas. Vamos a examinarlos a continuación.

Supongamos que se da la ecuación $f(x) = 0$, donde $f(x)$ es la función continua. Se necesita hallar la raíz de esta ecuación ξ con exactitud hasta ε, donde ε es cierto número positivo suficientemente pequeño.

Vamos a suponer que la raíz ξ está separada y se encuentra sobre el segmento $[a, b]$, o sea, tiene lugar la desigualdad $a \leqslant \xi \leqslant b$. En este caso por valor aproximado de la raíz se toma el centro del segmento $\frac{a+b}{2}$ y el error de este valor aproximado es igual a $\frac{b-a}{2}$. Por eso, si $b - a \leqslant 2\varepsilon$, la exactitud requerida está alcanzada y se toma $\xi = \frac{a+b}{2} \pm \varepsilon$. En cambio, si $b - a > 2\varepsilon$, es necesario estrechar el intervalo sobre el cual está la raíz ξ, o sea, elegir tales números $\bar{a}$ y $\bar{b}$ que se cumplan las desigualdades siguientes:

$$\bar{a} \leqslant \xi \leqslant b \quad \text{y} \quad \bar{b} - \bar{a} \leqslant 2\varepsilon.$$

Supongamos que la raíz ξ de la ecuación $f(x) = 0$ [$f(x)$ es una función continua] está separada sobre el segmento $[a, b]$, o sea $f(a) \cdot f(b) < 0$, con la particularidad de que $b - a < 2\varepsilon$. Se necesita hallar el valor de la raíz ξ con exactitud hasta ε (fig. 5.15).

Sobre el segmento $[a, b]$ elegimos de un modo arbitrario el punto a_1 que lo divide en dos segmentos $[a, a_1]$ y $[a_1, b]$. De estos dos segmentos hay que elegir aquél en cuyos extremos la función toma los valores contrarios en signo. En el caso en cuestión $f(a) \cdot f(a_1) > 0$, $f(a_1) \cdot f(b) < 0$, por eso conviene elegir el segmento $[a_1, b]$. Luego, sobre este segmento estrechado volvamos a tomar arbitrariamente el punto a_2 y determinemos los signos de los productos $f(a_1) \cdot f(a_2)$ y $f(a_2) \cdot f(b)$. Puesto que $f(a_2) \cdot f(b) < 0$, escogemos el segmento $[a_2, b]$. Continuamos este proceso hasta que la longitud del segmento, sobre el cual está la raíz, llegue a ser menor que 2ε. Obtenemos la

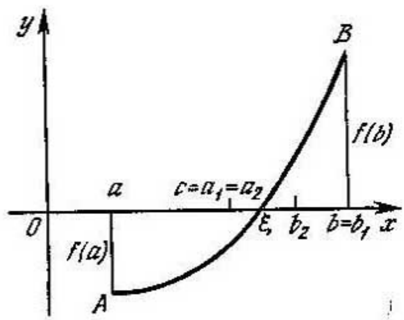


Fig. 5.16

raíz ξ como media aritmética de los extremos del segmento determinado, con la particularidad de que el error de la raíz no supera ε.

El método considerado se llama método de pruebas.

En la forma descrita anteriormente el método de pruebas no se emplea en los ordenadores. Para hacer programas y cálculos en los ordenadores éste se aplica en la forma del llamado **método de bisección**.

Supongamos que la raíz ξ de la ecuación $f(x) = 0$ está separada y se halla sobre el segmento $[a, b]$, o sea, $f(a) \cdot f(b) < 0$, con la particularidad de que $b - a > 2\varepsilon$ [aquí $f(x)$ es una función continua]. Al igual que antes, tomemos sobre el segmento $[a, b]$ un punto intermedio, pero no de un modo arbitrario sino así que éste sea el centro del segmento $[a, b]$, o sea, $c = (a + b)/2$. Entonces el segmento $[a, b]$ se divide por el punto c en dos segmentos iguales $[a, c]$ y $[c, b]$ cuyas longitudes son iguales a $(b - a)/2$ (fig. 5.16). Si $f(c) = 0$, entonces c es la raíz exacta de la ecuación $f(x) = 0$. En cambio, si $f(c) \neq 0$, entonces de dos segmentos obtenidos $[a, c]$ y $[c, b]$ elegimos aquél en cuyos extremos la función $f(x)$ toma los valores de signos opuestos. Designémoslo $[a_1, b_1]$. Luego bisecamos también el segmento $[a_1, b_1]$ y razonamos de un modo análogo. Obtenemos el segmento $[a_2, b_2]$ cuya longitud es igual a $(b - a)/2^2$. El proceso de bisección del segmento se lleva a cabo hasta que en cierta n-ésima etapa el centro del segmento resulte ser la raíz de la ecuación (caso muy raro en la práctica) o bien se obtenga el segmento $[a_n, b_n]$

tal que $b_n - a_n = (b-a)/2^n \leqslant 2\varepsilon$ y $a_n \leqslant \xi \leqslant b_n$ (el número n indica la cantidad de bisecciones efectuadas). Los números a_n y b_n son las raíces de la ecuación $f(x) = 0$ con exactitud hasta 2ε. Según hemos señalado anteriormente por valor aproximado de la raíz es necesario tomar $\xi = (a_n + b_n)/2$, con la particularidad de que el error no supera $(b-a)/2^{n+1} \leqslant \varepsilon$.

Ejemplo 1. Haciendo uso del método de pruebas, precisar hasta $\varepsilon = 10^{-3}$ la menor raíz de la ecuación $x^3 + 3x^2 - 3 = 0$.

△ Separamos las raíces de la ecuación analíticamente. La función $f(x) = x^3 + 3x^2 - 3$ está definida sobre todo el eje numérico. Igualamos a cero $f'(x)$ y calculamos la raíz de la derivada:

$$f'(x) = 3x^2 + 6x; \quad 3x^2 + 6x = 0; \quad x(x+2) = 0; \quad x_1 = 0; \quad x_2 = -2.$$

Hacemos la tabla de los signos de la función:

x	$-\infty$	-2	-1	0	1	$+\infty$
Signo de $f(x)$	$-$	$+$	$-$	$-$	$+$	$+$

Vemos que la primera raíz se contiene en el intervalo $(-\infty, -2)$. Tomemos para la prueba $x = -3$ y determinemos $f(-3) = -3$;

x	-3	-2	-1	0	1
Signo de $f(x)$	$-$	$+$	$-$	$-$	$+$

Por lo tanto, las raíces de la ecuación $x^3 + 3x^2 - 3 = 0$ se contienen en los intervalos $(-3; -2)$, $(-2, -1)$, $(0, 1)$.

Precisemos la menor raíz que está en el intervalo $(-3, -2)$ valiéndonos del método de bisección. Para la comodidad de los cálculos hagamos la tabla (véase la tabla 5.1). Los signos «$-$» y «$+$» presentes en los índices superiores a_n y b_n significan que $f(a_n) < 0$ y $f(b_n) > 0$.

Tabla 5.1

n	a_n^-	b_n^+	$x_n = \frac{a_n + b_n}{2}$	x_n^3	$3x_n^2$	$f(x_n)$
0	-3	-2	$-2,500$	$-15,625$	18,750	0,125
1	-3	$-2,500$	$-2,750$	$-20,800$	22,689	$-1,111$
2	$-2,750$	$-2,500$	$-2,625$	$-17,990$	20,670	$-0,320$
3	$-2,625$	$-2,500$	$-2,563$	$-16,840$	19,701	$-0,139$
4	$-2,563$	$-2,500$	$-2,532$	$-16,230$	19,233	0,003
5	$-2,563$	$-2,532$	$-2,548$	$-16,540$	19,479	$-0,071$
6	$-2,548$	$-2,532$	$-2,540$	$-16,390$	19,356	$-0,034$
7	$-2,540$	$-2,532$	$-2,536$	$-16,310$	19,293	$-0,014$
8	$-2,536$	$-2,532$	$-2,534$	$-16,270$	19,263	$-0,007$
9	$-2,534$	$-2,532$				

Así pues, la raíz de la ecuación $x \approx -2{,}533$. ▲

Ejemplo 2. Separar gráficamente la raíz de la ecuación $x^2 \times \log_{0,5}(x+1) = 1$. Haciendo uso del método de pruebas, calcular esta raíz con exactitud $\varepsilon = 10^{-2}$.

△ Representemos la ecuación en la forma $\log_{0,5}(x+1) = 1/x^2$ y construyamos los gráficos de las funciones $y = \log_{0,5}(x+1)$

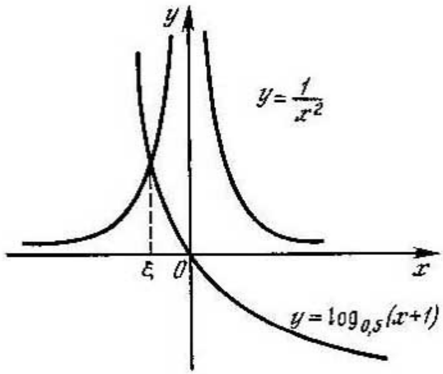


Fig. 5.17

e $y = 1/x^2$. De la fig. 5.17 se ve que la ecuación tiene una sola raíz $x_1 \approx -0{,}7$. Determinemos los signos de la función a la izquierda y a la derecha de x_1:

x	$-0{,}8$	$-0{,}5$
Signo de $f(x)$	$+$	$-$

Para la comodidad de los cálculos pasemos a los logaritmos decimales

$$f(x) = x^2 \frac{\log(x+1)}{\log 0{,}5} - 1 = x^2 \frac{\log(x+1)}{-0{,}301} - 1.$$

Hagamos la siguiente tabla (tabla 5.2):

Tabla 5.2

n	a_n^+	b_n^-	$x_n = \frac{a_n + b_n}{2}$	x_n^2	$\log(x_n + 1)$	$f(x_n)$
0	$-0{,}8$	$-0{,}5$	$-0{,}65$	$0{,}4225$	$-0{,}4559$	$-0{,}360$
1	$-0{,}8$	$-0{,}65$	$-0{,}73$	$0{,}5329$	$-0{,}5686$	$-0{,}0067$
2	$-0{,}73$	$-0{,}65$	$-0{,}69$	$0{,}4761$	$-0{,}5086$	$-0{,}196$
3	$-0{,}73$	$-0{,}69$	$-0{,}71$	$0{,}5041$	$-0{,}5376$	$-0{,}099$
4	$-0{,}73$	$-0{,}71$				

Así pues, $x_1 \approx -0{,}72$. ▲

§ 5.4. Método de las cuerdas

El **método de las cuerdas** es uno de los más difundidos de resolución de las ecuaciones algebraicas y trascendentes. En los manuales se utilizan también otros sus nombres: «método de posición falsa» (*regula falsi*), «método de interpolación lineal» y «método de las partes proporcionales».

Supongamos que se da la ecuación $f(x) = 0$, donde $f(x)$ es una función continua que tiene en el intervalo (a, b) las derivadas de primer orden y de segundo orden. La raíz se considera separada y se halla sobre el segmento $[a, b]$, o sea, $f(a) \cdot f(b) < 0$.

La idea del método de las cuerdas consiste en que al ser suficientemente pequeño el intervalo $[a, b]$ el arco de la curva $y = f(x)$ se

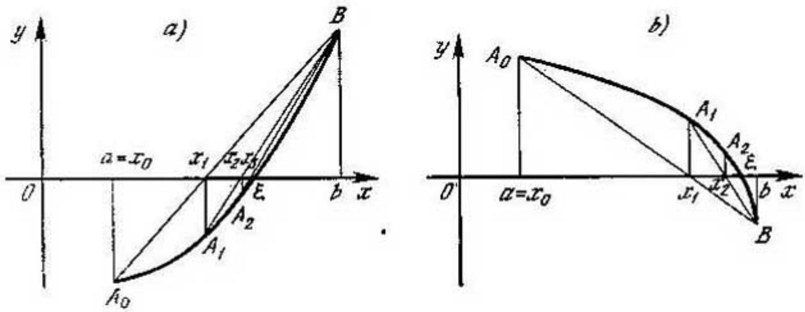


Fig. 5.18

reemplaza por la cuerda que lo subtiende. En calidad de valor aproximado de la raíz se toma el punto de intersección de la cuerda con el eje Ox.

Antes hemos examinado diferentes variantes de disposición del arco de la curva en función de los signos de las derivadas primera y segunda.

Caso I. Las derivadas primera y segunda tienen los mismos signos, o sea $f'(x) \cdot f''(x) > 0$. Supongamos, por ejemplo, que $f(a) < 0$, $f(b) > 0$, $f'(x) > 0$, $f''(x) > 0$ (fig. 5.18, *a*). El gráfico de la función pasa por los puntos $A_0\,(a;\, f(a))$, $B\,(b;\, f(b))$. La raíz buscada de la ecuación $f(x) = 0$ es la abscisa del punto de intersección del gráfico $y = f(x)$ con el eje Ox. Desconocemos este punto, pero en vez de él tomamos el punto x_1 de intersección de la cuerda A_0B con el eje Ox. Éste es precisamente el valor aproximado de la raíz.

La ecuación de la cuerda que atraviesa los puntos A_0 y B tiene la forma

$$\frac{y - f(a)}{f(b) - f(a)} = \frac{x - a}{b - a}.$$

Determinemos el valor $x = x_1$ para el cual $y = 0$:

$$x_1 = a - \frac{f(a)(b - a)}{f(b) - f(a)}. \qquad (1)$$

Ahora la raíz ξ se halla dentro del segmento $[x_1, b]$. Si el valor de la raíz x_1 no nos satisface, podemos precisarlo, aplicando el método de las cuerdas al segmento $[x_1, b]$. Unimos el punto A_1 $(x_1; f(x_1))$ con el punto B $(b; f(b))$ y hallamos x_2, o sea, el punto de intersección de la cuerda A_1B con el eje Ox:

$$x_2 = x_1 - \frac{f(x_1)(b-x_1)}{f(b)-f(x_1)}.$$

Continuando este proceso, tenemos

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

y en general

$$x_{n+1} = x_n - \frac{f(x_n)(b-x_n)}{f(b)-f(x_n)}. \tag{2}$$

El proceso continúa hasta que obtengamos la raíz aproximada con el grado prefijado de precisión.

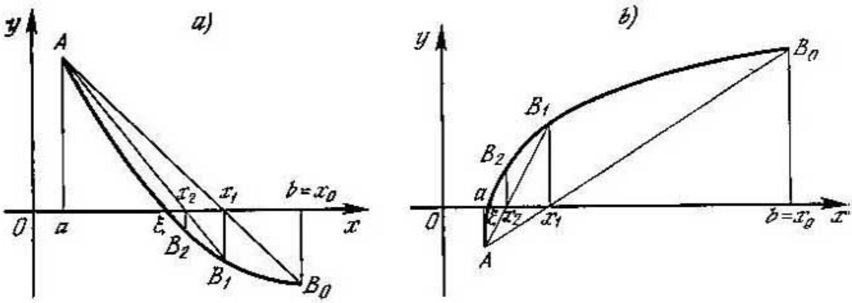


Fig. 5.19

Con ayuda de las fórmulas recién dadas se calculan también las raíces cuando $f(a) > 0$, $f(b) < 0$, $f'(x) < 0$, $f''(x) < 0$ (fig. 5.18,*b*).

Caso II. Las derivadas primera y segunda tienen los signos opuestos, o sea, $f'(x) \cdot f''(x) < 0$. Sea, por ejemplo, $f(a) > 0$, $f(b) < 0$, $f'(x) < 0$, $f''(x) > 0$ (fig. 5.19, *a*). Unimos los puntos A $(a; f(a))$ y B_0 $(b; f(b))$ y escribimos la ecuación de la cuerda que pasa por A y B_0:

$$\frac{y-f(b)}{f(b)-f(a)} = \frac{x-b}{b-a}.$$

Determinemos x_1 como punto de intersección de la cuerda con el eje Ox, suponiendo $y = 0$:

$$x_1 = b - \frac{f(b)(b-a)}{f(b)-f(a)}. \tag{3}$$

La raíz ξ está encerrada ahora dentro del segmento $[a, x_1]$. Aplicando el método de las cuerdas al segmento $[a, x_1]$, obtenemos

$$x_2 = x_1 - \frac{f(x_1)(x_1-a)}{f(x_1)-f(a)}. \tag{4}$$

y en general

$$x_{n+1}=x_n-\frac{f(x_n)(x_n-a)}{f(x_n)-f(a)}. \tag{4}$$

Con ayuda de estas mismas fórmulas se determina el valor aproximado de la raíz también cuando $f(a)<0$, $f(b)>0$, $f'(x)>0$, $f''(x)<0$ (fig. 5.19, *b*).

Por lo tanto, si $f'(x)\cdot f''(x)>0$, la raíz aproximada se calcula por las fórmulas (1) y (2); en cambio, si $f'(x)\cdot f''(x)<0$, se calcula por las fórmulas (3) y (4).

Sin embargo, la elección de unas u otras fórmulas se puede llevar a cabo valiéndose de una **regla** simple: *como extremo fijo del segmento sirve aquél para el cual el signo de la función coincide con el signo de la segunda derivada.*

Si $f(b)\cdot f''(x)>0$, es fijo el extremo b y en calidad de aproximación inicial hace falta tomar el extremo a [fórmulas (1) y (2)]. En cambio, si $f(a)\cdot f''(x)>0$, es fijo el extremo a y en calidad de aproximación inicial es necesario tomar el extremo b [fórmulas (3) y (4)].

Ahora bien, como resultado de aplicación reiterada de las fórmulas (2) ó (4) obtenemos la sucesión monótona $x_1, x_2, x_3, \ldots, x_n$ que converge hacia el valor de la raíz ξ.

Al estimar el error de aproximación se puede utilizar la fórmula

$$|\xi-x_n|<|x_n-x_{n-1}|, \tag{5}$$

donde ξ es el valor exacto de la raíz y x_{n-1} y x_n son las aproximaciones a éste obtenidas en los pasos $n-1$ y n. Se puede hacer uso de ella si se cumple la condición

$$M\leqslant 2m, \quad \text{donde} \quad M=\max_{[a,\,b]}|f'(x)|, \quad m=\min_{[a,\,b]}|f'(x)|. \tag{6}$$

Ejemplo 1. Con ayuda del método de las cuerdas precisar hasta $\varepsilon=0{,}001$ la menor raíz de la ecuación $x^3+3x^2-3=0$. Las raíces de la ecuación están separadas y la menor raíz se contiene sobre el segmento $[-3, -2]$ (véase el ejemplo 1, § 5.3).

$\triangle$ Verificamos el cumplimiento de la condición (6):

$$|f'(x)|=|3x^2+6x|;$$

$$M=\max_{[-3,\,-2]}|f'(x)|=|27-18|=9;$$

$$m=\min_{[-3,\,-2]}|f'(x)|=|12-12|=0;\quad M\not\leqslant 2m.$$

Tomamos el centro del segmento $[-3, -2]$, o sea, el punto $x=-2{,}5$ y eligimos el intervalo $[-3; -2{,}5]$. Volvemos a verificar el cumplimiento de la condición (6):

$$M=\max_{[-3,\,-2{,}5]}|f'(x)|=9,\quad m=\min_{[-3\,-2{,}5]}|f'(x)|=3{,}75$$

$$M\not\leqslant 2m.$$

Tomamos ahora el centro del segmento $[-3; -2{,}5]$, o sea, el punto $x = -2{,}75$; tenemos $f(-2{,}75) < 0$, $f(-2{,}5) > 0$, $f(-3) < 0$; elegimos el segmento $[-2{,}75; -2{,}5]$. Encontramos $M = \max\limits_{[-2{,}75;\,-2{,}5]} |f'(x)| = 6{,}189$; $m = \min\limits_{[-2{,}75;\,-2{,}5]} |f'(x)| = 3{,}75$, o sea, en este caso está cumplida la condición $M < 2m$.

Ahora bien, para estimar el error de la raíz que está sobre el segmento $[-2{,}75; -2{,}5]$ se puede hacer uso de la fórmula (5): $|\xi - x_n| < |x_n - x_{n-1}|$, o sea, el proceso de aproximación sucesiva a la raíz ha de continuarse hasta que quede cumplida la condición $|x_n - x_{n-1}| \leqslant \varepsilon$.

Vamos a determinar el signo de la segunda derivada y con ayuda de qué fórmula hace falta llevar a cabo los cálculos. Hallamos $f''(x) = 6x + 6$; sobre el segmento $[-2{,}75; -2{,}5]$ tienen lugar las desigualdades $f(-2{,}75) < 0$, $f(-2{,}75) \cdot f''(x) > 0$. Esto quiere decir que por extremo fijo del segmento hay que tomar $x = -2{,}75$. Entonces es necesario realizar los cálculos valiéndose de las fórmulas (3) y (4):

$$x_1 = b - \frac{f(a)(b-a)}{f(b) - f(a)}; \quad x_{n+1} = x_n - \frac{f(x_n)(x_n - a)}{f(x_n) - f(a)},$$

donde $a = -2{,}75$ y $f(a) = -1{,}111$. Si la última expresión se representa en la forma

$$x_{n+1} - x_n = -\frac{f(x_n)(x_n - a)}{f(x_n) - f(a)},$$

inmediatamente se puede obtener la diferencia entre dos aproximaciones sucesivas y llevar a cabo la verificación del fin de los cálculos, o sea, comprobar el cumplimiento de la desigualdad $|x_{n+1} - x_n| \leqslant \varepsilon$.

Es cómodo realizar todos los cálculos en la tabla siguiente:

Tabla 5.3

n	x_n	x_n^3	x_n^2	$3x_n^2$	$f(x_n) = x_n^3 + 3x_n^2 - 3$	$x_n - a$	$-\frac{f(x)(x-a)}{f(x_n) - f(a)}$
0	$-2{,}5$	$-15{,}625$	$6{,}250$	$18{,}75$	$0{,}125$	$0{,}25$	$-0{,}025$
1	$-2{,}525$	$-16{,}098$	$6{,}3756$	$19{,}1268$	$0{,}0288$	$0{,}225$	$-0{,}006$
2	$-2{,}531$	$-16{,}213$	$6{,}4060$	$19{,}2180$	$0{,}0050$	$0{,}219$	$-0{,}0009$
3	$-2{,}5319$						

De la tabla 5.3. se ve que $|x_3 - x_2| = 0{,}0009$, por eso, redondeando hasta las milésimas partes, obtenemos $\xi = -2{,}532$. ▲.

Ejemplo 2. Haciendo uso del método de las cuerdas, precisar hasta $\varepsilon = 0{,}001$ la raíz de la ecuación $x - \operatorname{sen} x = 0{,}25$, encerrada en el segmento $[0, \pi/2]$.

△ Escribamos la ecuación en la forma $x - \operatorname{sen} x - 0{,}25 = 0$ y determinemos $f'(x) = 1 - \cos x$. Para verificar el cumplimiento de la condición (6) hagamos una tabla auxiliar en la cual las primeras dos columnas indican el comienzo del intervalo elegido de aislamiento de la raíz y su fin.

Tabla 5.4

n	b	Signos		M	m	Se cumple o no la condición $M \leqslant 2m$	$\frac{a+b}{2}$	$f\left(\frac{a+b}{2}\right)$
		$f(a)$	$f(b)$					
0,00	1,57	—	+	1,00	0	no	0,785	—
0,785	1,57	—	+	1,00	0,29251	no	1,178	+
0,785	1,178	—	+	0,6172	0,2925	no	0,982	—
0,982	1,178	—	+	0,6172	0,4446	sí		

De la última fila de la tabla 5.4 se ve que sobre el segmento [0,982; 1,178] la condición $M \leqslant 2m$ se cumple. Por consiguiente, al estimar el error del valor aproximado de la raíz con ayuda del método de las cuerdas se puede hacer uso de la desigualdad $|x_n - x_{n-1}| < \varepsilon$. La raíz de la ecuación $x - \operatorname{sen} x - 0{,}25 = 0$ está sobre el segmento [0,982; 1,178]. Vamos a determinar el signo de la segunda derivada dentro del segmento:

$$f'(x) = 1 - \cos x; \quad f''(x) = \operatorname{sen} x > 0.$$

Si retornamos a las anteriores designaciones, $a = 0{,}982$; $b = 1{,}178$. El signo de la segunda derivada coincide con el de la función en el punto b. Por lo tanto, este extremo del segmento es fijo y todas las aproximaciones a la raíz están por el lado del extremo a. Para calcular la raíz utilizamos las fórmulas (1) y (2):

$$x_1 = a - \frac{f(a)(b-a)}{f(b)-f(a)}; \quad x_{n+1} = x_n - \frac{f(x_n)(b-x_n)}{f(b)-f(x_n)},$$

donde $b = 1{,}178$; $f(b) = 0{,}00416$. Hagamos la tabla siguiente:

Tabla 5.5

n	x_n	$-\operatorname{sen} x_n$	$f(x_n) = x_n - \operatorname{sen} x_n - 0{,}25$	$b - x_n$	$-\frac{f(x_n)(b-x_n)}{f(b)-f(x_n)}$
0	0,982	−0,83161	−0,09961	0,196	0,189
1	1,171	−9,92114	−0,00014	0,007	0,0002
2	1,1712				

Así pues, $x = 1{,}171$ con exactitud hasta $\varepsilon = 0{,}001$. ▲

§ 5.5. Método de Newton (método de las tangentes)

De otro método iterativo sirve el **de Newton**.

Supongamos que la raíz de la ecuación $f(x) = 0$ está separada sobre el segmento $[a, b]$ con la particularidad de que $f'(x)$ y $f''(x)$

son continuas y conservan signos constantes en todo el segmento $[a, b]$.

El sentido geométrico del método de Newton consiste en que el arco de la curva $y = f(x)$ se reemplaza por la tangente a esta curva (de aquí proviene precisamente el segundo nombre de este método, es decir, el **método de las tangentes**).

Caso I. Las derivadas primera y segunda tienen los mismos signos. Sea $f(a) < 0$, $f(b) > 0$, $f'(x) > 0$, $f''(x) > 0$ (fig. 5.20, a) o bien $f(a) > 0$, $f(b) < 0$, $f'(x) < 0$, $f''(x) < 0$ (fig. 5.20, b).

Tracemos la tangente a la curva $y = f(x)$ en el punto B_0 $(b; f(b))$ y determinemos la abscisa del punto de intersección de la tangente

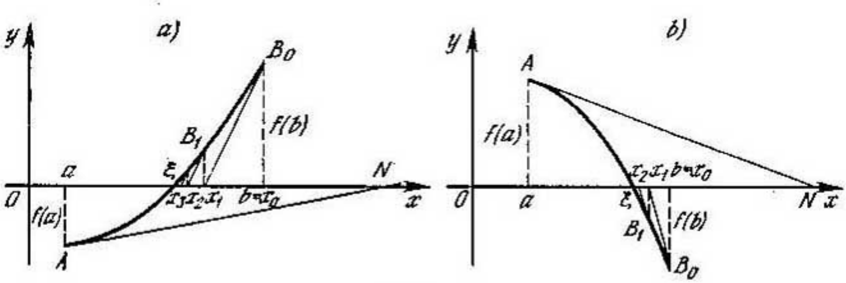


Fig. 5.20

con el eje Ox. Es sabido que la ecuación de la tangente en el punto B_0 $(b; f(b))$ tiene la forma

$$y - f(b) = f'(b)(x - b).$$

Suponiendo $y = 0$, $x = x_1$, obtenemos

$$x_1 = b - \frac{f(b)}{f'(b)}. \tag{1}$$

Ahora la raíz de la ecuación está sobre el segmento $[a, x_1]$. Volviendo a aplicar el método de Newton, tracemos la tangente a la curva en el punto B_1 $(x_1; f(x_1))$ y obtenemos

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

y en general

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{2}$$

Obtenemos la sucesión de los valores aproximados $x_1, x_2, \ldots$ $\ldots, x_n, \ldots$, en la cual cada término sucesivo es más próximo a la raíz ξ que el precedente. No obstante, todas las x_n quedan mayores que la raíz verdadera ξ, o sea, x_n es el valor aproximado de la raíz ξ con exceso.

Caso II. Las derivadas primera y segunda tienen los signos contrarios. Sea $f(a) < 0$, $f(b) > 0$, $f'(x) > 0$, $f''(x) < 0$ (fig. 5.21, *a*) o bien $f(a) > 0$, $f(b) < 0$, $f'(x) < 0$, $f''(x) > 0$ (fig. 5.21, *b*). Si volvemos a trazar la tangente a la curva $y = f(x)$ en el punto B, ésta cortará el eje de abscisas en un punto que no pertenece al segmento $[a, b]$. Por eso tracemos la tangente en el punto $A_0\ (a;\ f(a))$ y escribamos su ecuación para el caso dado:

$$y - f(a) = f'(a)(x - a).$$

Suponiendo $y = 0$, $x = x_1$, encontramos

$$x_1 = a \frac{f(a)}{f'(a)}. \tag{3}$$

Ahora la raíz ξ está sobre el segmento $[x_1, b]$. Volviendo a aplicar

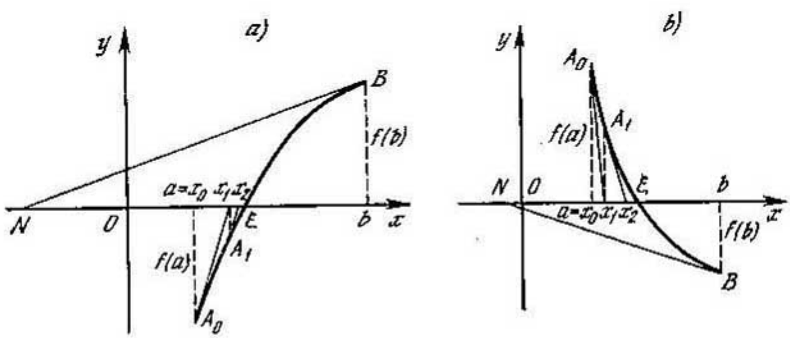


Fig. 5.21

el método de Newton, tracemos la tangente en el punto $A_1\ (x_1;\ f(x_1))$ y obtenemos

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

y en general

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}. \tag{4}$$

Se obtiene la sucesión de los valores aproximados $x_1, x_2, \ldots, x_n, \ldots$ en la cual cada término sucesivo es más próximo a la raíz verdadera ξ que el precedente, o sea x_n es el valor aproximado de la raíz ξ con defecto.

Comparando estas fórmulas con las anteriormente deducidas, notamos que se distinguen unas de otras sólo por la elección de la aproximación inicial: en el primer caso por x_0 hemos tomado el extremo b del segmento y en el segundo, el extremo a.

Al elegir la aproximación inicial de la raíz es necesario guiarse por la **regla** siguiente: *por punto inicial ha de tomarse aquel extremo* *del segmento* $[a, b]$ *en el cual el signo de la función coincide con el de la segunda derivada.* En el primer caso $f(b) \cdot f''(x) > 0$ y el punto inicial $b = x_0$, en el segundo caso $f(a) \cdot f''(x) > 0$ y en calidad de la aproximación inicial tomamos $a = x_0$.

Para estimar el error se puede usar la fórmula general

$$|\xi - x_n| \leqslant \frac{|f(x_n)|}{m}, \text{ donde } m = \min_{[a,\, b]} |f'(x)| \tag{5}$$

(esta fórmula es útil también para el método de las cuerdas).

En el caso en que el segmento $[a, b]$ es tan pequeño que sobre él se cumple la condición $M_2 < 2m_1$, donde $M_2 = \max\limits_{[a,\, b]} |f''(x)|$

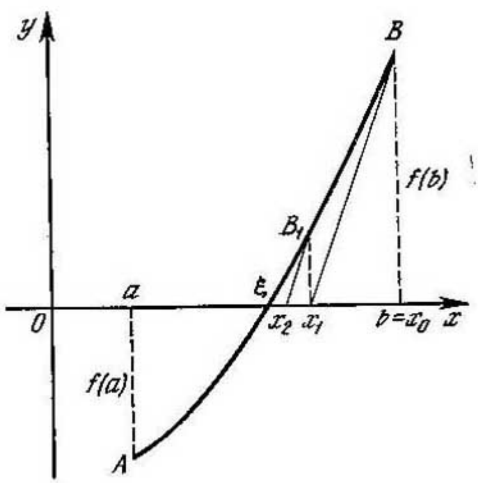


Fig. 5.22

y $m_1 = \min\limits_{[a,\, b]} |f'(x)|$, la exactitud de la aproximación en el n-ésimo paso se estima del modo siguiente: si $|x_n - x_{n-1}| < \varepsilon$, entonces $|\xi - x_n| < \varepsilon^2$.

Si la derivada $f'(x)$ poco varía sobre el segmento $[a, b]$, para simplificar los cálculos se puede utilizar la fórmula

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_0)}, \tag{6}$$

o sea, el valor de la derivada en el punto inicial basta calcularlo una sola vez. Geométricamente esto significa que las tangentes en los puntos $B_n(x_n;\ f(x_n))$ se reemplazan por las rectas paralelas a la tangente trazada a la curva $y = f(x)$ en el punto $B_0(x_0;\ f(x_0))$ (fig. 5.22).

Ejemplo 1. Haciendo uso del método de las tangentes, precisar hasta $\varepsilon = 0{,}001$ la raíz de la ecuación $x^3 + 3x^2 - 3 = 0$, situada sobre el segmento $[-2{,}75;\ -2{,}5]$.

△ Antes hemos determinado que $f(-2,75)\cdot f''(x) > 0$ (véase el ejemplo 1 del § 5.4). Por eso para servirse del método de las tangentes es necesario elegir $x_0 = -2,75$. Vamos a realizar los cálculos con ayuda de la fórmula (6). Hallamos

$$f'(x) = 3x^2 + 6x; \quad f'(x_0) = f'(-2,75) = 6,1875.$$

Para comodidad demos todos los cálculos en la tabla 5.6.

Tabla 5.6

n	x_n	x_n^3	x_n^2	$3x_n^2$	$f(x_n)$	$-\frac{f(x_n)}{6,1875}$
0	−2,75	−20,797	7,5625	22,6875	−1,111	0,179
1	−2,571	−16,994	6,6100	19,8300	−0,164	0,026
2	−2,545	−16,484	6,4770	19,431	−0,053	0,008
3	−2,537	−16,329	6,4364	19,309	0,020	0,003
4	−2,534	−16,271	6,4212	19,2636	0,007	0,001
5	−2,533					

De la tabla se ve que $|x_5 - x_4| = 0,001$; por eso $\xi = 2,533$. ▲

Ejemplo 2. Haciendo uso del método de las tangentes, precisar hasta $\varepsilon = 0,0001$ la raíz de la ecuación $x - \operatorname{sen} x = 0,25$, situada sobre el segmento $[0,982;\ 1,178]$.

△ Aquí $a = 0,982$; $b = 1,178$. Hallamos $f'(x) = 1 - \cos x$; $f''(x) = \operatorname{sen} x > 0$ sobre el segmento $[0,982;\ 1,178]$; $f(1,178) \times \times f''(x) > 0$. Por lo tanto, $x_0 = 1,178$. Vamos a realizar los cálculos con ayuda de las fórmulas (1) y (2) y démoslos en la tabla 5.7.

Tabla 5.7

n	x_n	$-\operatorname{sen} x_n$	$f(x_n) = x_n - \operatorname{sen} x_n - 0.25$	$f'(x_n) = 1 - \cos x_n$	$-\frac{f(x_n)}{f'(x_n)}$
0	1,178	−0,92384	0,00416	0,61723	−0,0065
1	1,1715	−0,92133	0,00017	0,61123	−0,0002
2	1,1713	−0,92127	0,00003	0,61110	−0,00005
3	1,17125				

De la tabla se ve que $|x_3 - x_2| < 0,0001$. Así pues, $\xi \approx 1,1712$. ▲

§ 5.6. Método combinado de las cuerdas y de las tangentes

El método de las cuerdas y el de las tangentes dan las aproximaciones de la raíz por los lados opuestos. Por eso se emplean frecuentemente combinados uno con otro y la raíz se precisa con mayor rapidez.

Supongamos que se da la ecuación $f(x) = 0$, la raíz ξ está separada y se halla sobre el segmento $[a, b]$. Apliquemos el **método combinado de las cuerdas y las tangentes** teniendo en cuenta el tipo del gráfico de la función.

Si $f'(x) \cdot f''(x) > 0$, *el método de las cuerdas da las aproximaciones de la raíz con defecto y el de las tangentes, con exceso* (fig. 5.23, a, b).

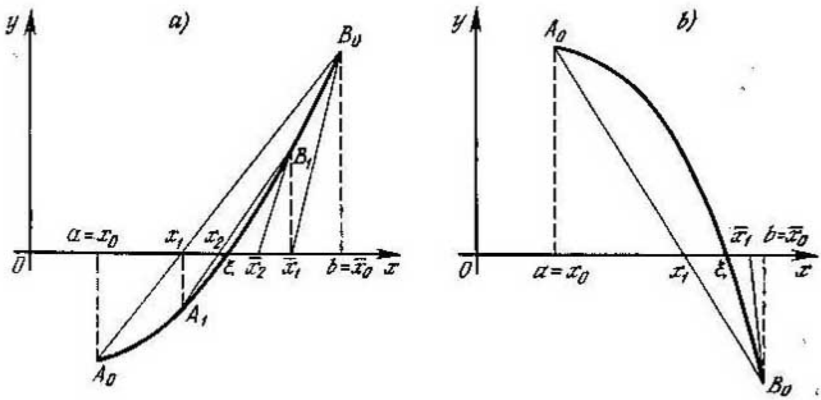


Fig. 5.23

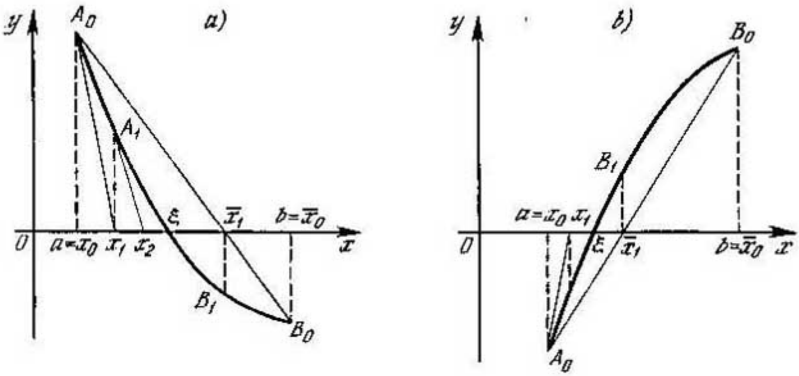


Fig. 5.24

En cambio, *si* $f'(x) \cdot f''(x) < 0$, *por el método de las cuerdas obtenemos los valores de la raíz con exceso y por el de las tangentes, con defecto* (fig. 5.24, a, b).

Sin embargo, en todos los casos la raíz verdadera está encerrada entre los valores aproximados de las raíces, obtenidos por el método de las cuerdas y por el de las tangentes, o sea, se cumple la desigualdad $a < \bar{x}_n < \xi < \bar{\bar{x}}_n < b$, donde $\bar{x}_n$ es el valor aproximado de la raíz tomado con defecto y $\bar{\bar{x}}_n$, tomado con exceso.

Los cálculos han de realizarse en el orden siguiente. Si $f'(x) \times \times f''(x) > 0$, en calidad de aproximación inicial para el método debe tomarse el extremo a y para el de las tangentes, el extremo b; entonces

$$a_1 = a - \frac{f(a)(b-a)}{f(b)-f}, \quad b_1 = b - \frac{f(b)}{f'(b)}. \tag{1}$$

Ahora la raíz verdadera está sobre el segmento $[a_1, b_1]$. Aplicando a este segmento el método combinado, obtenemos

$$a_2 = a_1 - \frac{f(a_1)(b_1-a_1)}{f(b_1)-f(a_1)}, \quad b_2 = b_1 - \frac{f(b_1)}{f'(b_1)}$$

y en general

$$a_{n+1} = a_n - \frac{f(a_n)(b_n-a_n)}{f(b_n)-f(a_n)}, \quad b_{n+1} = b_n - \frac{f(b_n)}{f'(b_n)} \tag{2}$$

(véase la fig. 5.23, a, b).

En cambio, si $f'(x) \cdot f''(x) < 0$, en calidad de aproximación inicial para el método de las cuerdas es necesario tomar el extremo b y para el de las tangentes, el extremo a. Entonces

$$a_1 = a - \frac{f(a)}{f'(a)}, \quad b_1 = b - \frac{f(b)(b-a)}{f(b)-f(a)}. \tag{3}$$

Aplicando al segmento $[a_1, b_1]$ el método combinado, obtenemos

$$a_2 = a_1 - \frac{f(a_1)}{f'(a_1)}, \quad b_2 = b_1 - \frac{f(b_1)(b_1-a_1)}{f(b_1)-f(a_1)}$$

y en general

$$a_{n+1} = a_n - \frac{f(a_n)}{f'(a_n)}, \quad b_{n+1} = b_n - \frac{f(b_n)(b_n-a_n)}{f(b_n)-f(a_n)} \tag{4}$$

(véase la fig. 5.24, a, b).

El método combinado es muy cómodo para estimar el error de cálculos. El proceso de cálculos cesa tan pronto como se cumpla la desigualdad $|\bar{\bar{x}}_n - \bar{x}_n| < 2\varepsilon$. Por valor aproximado de la raíz ha de tomarse

$$\xi = \frac{1}{2}(\bar{x}_n + \bar{\bar{x}}_n), \tag{5}$$

donde $\bar{x}_n$ y $\bar{\bar{x}}_n$ son los valores aproximados de la raíz con defecto y con exceso, respectivamente.

Ejemplo. Con ayuda del método combinado de las cuerdas y las tangentes precisar hasta 0,001 las raíces de la ecuación $x^3 + 3x^2 - - 24x + 1 = 0$.

△ 1) Separamos las raíces analíticamente. Tenemos

$$f(x) = x^3 + 3x^2 - 24x + 1; \quad f'(x) = 3x^2 + 6x - 24,$$

o sea, las raíces de la derivada $x_1 = -4$; $x_2 = 2$. Hagamos la tabla de los signos de la función:

x	$-\infty$	-4	2	$+\infty$
Signo de $f(x)$	$-$	$+$	$-$	$+$

La ecuación dada tiene tres raíces reales: $x_1 \in (-\infty, -4)$, $x_2 \in (-4, 2)$, $x_3 \in (2, +\infty)$. Disminuyamos los intervalos de determinación de las raíces hasta la longitud igual a 1:

x	-7	-6	-5	-4	-3	-2	-1	0	1	2	3	4
Signo de $f(x)$	$-$	$+$	$+$	$+$	$+$	$+$	$+$	$+$	$-$	$-$	$-$	$+$

Así pues, $x_1 \in (-7, -6)$ $x_2 \in (0, 1)$, $x_3 \in (3, 4)$.

2) Utilizando el método combinado de las cuerdas y las tangentes precisemos la raíz que esta en el intervalo $(-7, -6)$. Tenemos $f(-7) = -27 < 0, f(-6) = 37 > 0, f'(x) = 3x^2 + 6x - 24 > 0$, $f''(x) = 6x + 6 < 0$, $f'(x) \cdot f''(x) < 0$.

Para los cálculos utilizamos las fórmulas (4):

$$a_{n+1} = a_n - \frac{f(a_n)}{f'(a_n)}, \quad b_{n+1} = b_n - \frac{f(b_n)(b_n - a_n)}{f(b_n) - f(a_n)},$$

o sea,

$$a_{n+1} = a_n + \Delta a_n, \text{ donde } \Delta a_n = -\frac{f(a_n)}{f'(a_n)};$$

$$b_{n+1} = b_n + \Delta b_n, \text{ donde } \Delta b_n = -\frac{f(b_n)(b_n - a_n)}{f(b_n) - f(a_n)}$$

(a_n y b_n son los valores aproximados de la raíz tomados con defecto y con exceso, respectivamente). Aquí $a_0 = a = -7$, $b_0 = b = -6$.

Es cómodo realizar los cálculos con ayuda de la tabla 5.8

Tabla 5.8

n	a_n / b_n	$b_n - a_n$	a_n^2 / b_n^2	a_n^3 / b_n^3	$f(a_n)$ / $f(b_n)$	$f'(a_n)$	$f(b_n) - f(a_n)$	Δa_n / Δb_n	a_{n+1} / b_{n+1}
0	-7	1	49	-343	-27	81	64	0,3333	$-6,6667$
	-6		36	-216	37			$-0,5781$	$-6,5781$
1	$-6,6667$	0,0886	44,4449	$-296,3007$	$-1,9652$	69,3345	6,01002	0,0283	$-6,6384$
	$-6,5781$		43,2714	$-284,6436$	4,0450			$-0,0596$	$-6,6377$
2	$-6,6384$	0,0007							
	$-0,6377$								

Teniendo en cuenta que $|b_2 - a_2| = 0{,}0007 < 0{,}002$, conviene terminar los cálculos y por valor aproximado de la raíz ξ_1 tomar

$$\xi_1 = \frac{1}{2}(-6{,}6384 - 6{,}6377) = -6{,}638.$$

3) Determinemos la raíz aproximada para el intervalo (0, 1). Tenemos $(0) > 0$, $f(1) < 0$, $f'(x) = 3x^2 + 6x - 24 < 0$, $f''(x) = 6x + 6 > 0$, $f'(x) \cdot f''(x) < 0$. Al igual que en el primer caso hagamos uso de las fórmulas 4) para $a_0 = a = 0$; $b_0 = b = 1$.

Hagamos la tabla siguiente

Tabla 5.9

n	a_n / b_n	$b_n - a_n$	a_n^2 / b_n^2	a_n^3 / b_n^3	$f(a_n)$ / $f(b_n)$	$f'(a_n)$	$f(b_n) - f(a_n)$	Δa_n / Δb_n	a_{n+1} / b_{n+1}
0	0	1	0	0	1	−24	−20	0,0417	0,0417
	1		1	1	−19			−0,9500	0,0500
1	0,0417	0,0083	0,0017	0,00007	0,0045	−23,7446	−0,1969	0,0002	0,0419
	0,0500		0,0025	0,0001	−0,1924			−0,0081	0,0419
2	0,0419	0,0000							
	0,0419								

Ahora bien, con la exactitud hasta 0,001 se puede tomar $\xi_2 = 0{,}042$.

4) Determinemos la raíz aproximada contenida en el intervalo (3, 4). Tenemos $f(3) = -17 < 0$, $f(4) = 17 > 0$, $f'(x) = 3x^2 + 6x - 24 > 0$, $f''(x) = 6x + 6 > 0$, $f'(x) \cdot f''(x) > 0$.

Vamos a realizar los cálculos con ayuda de las fórmulas (2):

$$a_{n+1} = a_n - \frac{f(a_n)(b_n - a_n)}{f(b_n) - f(a_n)}, \quad b_{n+1} = b_n - \frac{f(b_n)}{f'(b_n)},$$

o sea,

$$a_{n+1} = a_n + \Delta a_n, \text{ donde } \Delta a_n = -\frac{f(a_n)(b_n - a_n)}{f(b_n) - f(a_n)},$$

$$b_{n+1} = b_n + \Delta b_n, \text{ donde } \Delta b_n = -\frac{f(b_n)}{f'(b_n)}.$$

Aquí $a_0 = a = 3$, $b_0 = b = 4$.

Realicemos los cálculos con ayuda de la tabla (véase la tabla 5.10 en la pág. 190).

Ahora bien, con la exactitud hasta 0,001 se puede tomar $\xi_3 = 3{,}596$. ▲

Tabla 5.10

n	a_n	$b_n - a_n$	a_n^2	a_n^3	$f(a_n)$	$f'(b_n)$	$f(b_n) - f(a_n)$	Δa_n	a_{n+1}
	b_n		b_n^2	b_n^3	$f(b_n)$			Δb_n	b_{n+1}
0	3	1	9	27	−17	48	34	0,5000	3,5000
	4		16	64	17			−0,3542	3,6458
1	3,5000	0,1458	12,2500	42,875	−3,3750	37,7505	5,2110	0,0944	3,5944
	3,6458		13,2919	48,4595	1,8360			−0,0486	3,5972
2	3,5944	0,0028	12,9197	46,4386	−0,0679	36,4026	0,1017	0,0019	3,5963
	3,5972		12,9398	46,5472	0,0338			−0,0009	3,5963
3	3,5963	0,0000							
	3,5963								

§ 5.7. Método de iteraciones

Esencia del método. Uno de los métodos más importantes de la matemática de cálculos es el **método de iteraciones (método de aproximaciones sucesivas)**. La ventaja principal de este método consiste en que en cada paso se cumplen las operaciones del mismo tipo lo que facilita en grado considerable la programación para los ordenadores basada en los algoritmos iterativos.

La esencia del método de iteraciones consiste en lo siguiente. Consideremos la ecuación

$$f(x) = 0. \tag{1}$$

Supongamos que $f(x)$ es la función, continua sobre el segmento $[a, b]$, la cual se anula dentro de este intervalo, al menos, en un punto ξ. Se necesita precisar, como mínimo, una de sus raíces reales situados sobre $[a, b]$.

Reemplacemos la ecuación (1) por otra, equivalente, cuyas raíces son las mismas y que tiene la forma

$$x = \varphi(x). \tag{2}$$

Elijamos en calidad de aproximación inicial cualquier valor $x_0 \in [a, b]$, por ejemplo, $x_0 = (a + b)/2$. Luego calculemos $\varphi(x_0)$ y tomemos el número obtenido $x_1 = \varphi(x_0)$ por primer valor aproximado de la raíz ξ. Sustituyendo x_1 en vez de x en el segundo miembro de la ecuación (2), obtenemos un nuevo número $x_2 = \varphi(x_1)$. Continuando luego este proceso, llegamos a la sucesión de los números $x_0, x_1, x_2, \ldots, x_n, \ldots$, que se definen por las relaciones siguientes:

$$x_0 = (a + b)/2; \quad x_n = \varphi(x_{n-1}) \quad (n = 1, 2, \ldots). \tag{3}$$

Si existe $\lim\limits_{n\to\infty} x_n = \xi$ y la función $\varphi(x)$ es continua, entonces, pasando al límite en la igualdad (3), obtenemos

$$\xi = \varphi(\xi), \tag{4}$$

o sea, el límite ξ es la raíz de la ecuación (2) y, por consiguiente, de la ecuación (1). En virtud de la convergencia del proceso (3) esta raíz puede ser calculada, al ser bastante grande n, con toda exactitud prefijada.

Nótese que la representación (2) [o sea, la forma de la función $\varphi(x)$] puede realizarse por un conjunto infinito de procedimientos. Esto tiene gran importancia, ya que la forma de la función $\varphi(x)$ ejerce una influencia sustancial tanto en el mismo hecho de convergencia como también en la velocidad de esta última (si la convergencia está determinada).

Las condiciones de convergencia del proceso iterativo (3) se definen por el teorema siguiente.

Teorema. *Supongamos que se cumplen las condiciones:*

1°) *la función $\varphi(x)$ está definida y es derivable sobre el segmento $[a, b]$;*

2°) *todos los valores de $\varphi(x) \in [a, b]$ para $x \in [a, b]$;*

3°) *existe tal número $q < 1$ que para $x \in [a, b]$*

$$|\varphi'(x)| \leqslant q < 1. \tag{5}$$

Entonces el proceso iterativo (3) converge independientemente de la elección de la aproximación inicial $x_0 \in [a, b]$ y $\lim\limits_{n\to\infty} x_n$ es la raíz única y simple de la ecuación $x = \varphi(x)$ sobre el segmento $[a, b]$.

□ Planteemos las diferencias siguientes:

$$|x_1 - x_0| = |x_1 - x_0|,$$
$$|x_2 - x_1| = |\varphi(x_1) - \varphi(x_0)| = |\varphi'(c_1)|\,|x_1 - x_0| \leqslant$$
$$\leqslant q\,|x_1 - x_0|,$$
$$|x_3 - x_2| = |\varphi(x_2) - \varphi(x_1)| = |\varphi'(c_2)| \cdot |x_2 - x_1| \leqslant$$
$$\leqslant q^2\,|x_1 - x_0|,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$|x_n - x_{n-1}| = |\varphi(x_{n-1}) - \varphi(x_{n-2})| =$$
$$= |\varphi'(c_n)| \cdot |x_{n-1} - x_{n-2}| \leqslant q^{n-1}\,|x_1 - x_0|.$$

Aquí $c_k \in [x_{k-1}, x_k]$ y $x_k \in [a, b]$ en virtud de la condición 2°.

Consideremos la serie con las sumas parciales siguientes: $S_{n+1} = x_0 + (x_1 - x_0) + (x_2 - x_1) + \ldots + (x_n - x_{n-1} \ldots) + \ldots$ Es evidente que la suma $S_{n+1} = x_n$. La serie $x_0 + \sum\limits_{i=0}^{\infty} (x_{i+1} -$

$- x_i)$ es convergente, puesto que todos sus términos, comenzando con $x_1 - x_0$, no superan en valor absoluto los términos de la progresión geométrica con el denominador $q < 1$. Así pues, existe

$$\lim_{n\to\infty} S_{n+1} = \lim_{n\to\infty} x_n = \xi.$$

En virtud de la continuidad de la función $\varphi(x)$ tenemos $\lim_{n\to\infty} \varphi(x_{n-1}) = \varphi(\xi)$ y ya que $x_n = \varphi(x_{n-1})$, entonces $\varphi(\xi) = \xi$, o sea $\xi = \lim_{n\to\infty} x_n$ es la raíz de la ecuación (2).

Demostremos que esta raíz es la única. Sean ξ_1 y ξ_2 dos raíces de la ecuación (2), es decir, $\xi_1 = \varphi(\xi_1)$, $\xi_2 = \varphi(\xi_2)$. Entonces

$$|\xi_1 - \xi_2| = |\varphi(\xi_1) - \varphi(\xi_2)| = |\varphi'(c)| \cdot |\xi_1 - \xi_2|, \tag{6}$$

donde $c \in (\xi_1, \xi_2)$. Transformemos la relación (6) de modo que tenga la forma

$$|\xi_1 - \xi_2| (1 - |\varphi'(c)|) = 0$$

y de la condición 3° se deduce que $\xi_1 = \xi_2$, o sea, no existen dos raíces diferentes.

Demostremos, por último, que la raíz obtenida es simple. Para esto basta demostrar que $x - \varphi(x)$ tiene la derivada que no se anula en ningún punto del segmento $[a, b]$. En efecto, $(x - \varphi(x))' = 1 - \varphi'(x)$ y en virtud de la condición 3° es evidente que esta expresión es positiva sobre el segmento $[a, b]$. ■

Estimación de error. Consideremos la diferencia entre los valores exacto y aproximado de la raíz ξ:

$$|\xi - x_n| = |\varphi(\xi) - \varphi(x_{n-1})| \leqslant q|\xi - x_{n-1}| =$$
$$= q|\xi - x_n + x_n - x_{n-1}| \leqslant q|\xi - x_n| + q|x_n - x_{n-1}|.$$

De aquí

$$|\xi - x_n| \leqslant \frac{q}{1-q}|x_n - x_{n-1}| \leqslant \frac{q^n}{1-q}|x_1 - x_0|. \tag{7}$$

La relación (7) permite, inmediatamente después de la primera iteración (7) permite, inmediatamente después de la primera iteración, hallar el número máximo de iteraciones $N(\xi)$ necesario para calcular la raíz con el grado de precisión prefijado ε. En efecto, para que $|\xi - x_n|$ sea no más que ε es suficiente que

$$\frac{q^n}{1-q}|x_1 - x_0| \leqslant \varepsilon,$$

de donde

$$N(\varepsilon) \geqslant \frac{\log \dfrac{\varepsilon(1-q)}{|x_1 - x_0|}}{\log q}. \tag{8}$$

Para $q \leqslant 1/2$ la estimación del error (7) se simplifica y toma la forma siguiente:

$$|\xi - x_n| \leqslant |x_n - x_{n-1}|. \tag{9}$$

Según hemos señalado anteriormente, la forma de la ecuación $x = \varphi(x)$ no es indiferente para la convergencia del método de iteraciones. Ahora vamos a mostrar un procedimiento bastante general de construcción de la función $\varphi(x)$ para la cual está asegurado el cumplimiento de la condición 3°) del teorema.

Consideremos la ecuación inicial $f(x) = 0$. Supongamos que sobre el segmento $[a, b]$ existe la única raíz ξ de la ecuación, con la particularidad de que para $x \in [a, b]$ la derivada $f'(x)$ existe y conserva su signo así que

$$0 < m_1 \leqslant f'(x) \leqslant M_1, \text{ donde } M_1 = \max_{[a, b]} f'(x); \quad m_1 = \min_{[a, b]} f'(x). \tag{10}$$

(sin alterar la generalidad, se puede considerar que $f'(x) > 0$).

Reemplacemos la ecuación $f(x) = 0$ por la ecuación equivalente

$$x = x - \lambda f(x) \tag{11}$$

y elijamos la constante λ de modo que se asegure el cumplimiento de la condición 3°).

Para la función $\varphi(x) \equiv x - \lambda f(x)$ la condición 3°) se escribe del modo siguiente:

$$|1 - \lambda f'(x)| < 1.$$

Vamos a resolver esta desigualdad; tenemos $-1 < 1 - \lambda f'(x) < 1$, de la segunda desigualdad obtenemos que $\lambda > 0$ y de la primera, que $\lambda < 2/f'(x)$, o sea,

$$0 < \lambda < 2/f'(x), \quad x \in [a, b],$$

o bien

$$0 < \lambda < 2/M_1.$$

De ordinario en calidad de λ se toma $1/M_1$. Ahora bien, obtenemos la ecuación

$$x = x - \frac{f(x)}{M_1} \tag{12}$$

y el proceso iterativo correspondiente tiene la forma

$$x_{n+1} = x_n - \frac{f(x_n)}{M_1}. \tag{13}$$

Con ayuda de los razonamientos análogos se puede mostrar fácilmente que en el caso de que $f'(x) < 0$ y $0 < m_1 \leqslant |f'(x)| \leqslant M_1$ obtenemos la ecuación

$$x = x + \frac{f(x)}{M_1} \tag{12'}$$

y el proceso iterativo correspondiente tomará la forma

$$x_{n+1} = x_n + \frac{f(x_n)}{M_1}. \tag{13'}$$

Supongamos ahora que, además de la condición (10) tiene lugar la relación

$$M_1 \leqslant 3m_1. \tag{14}$$

Entonces se puede exigir que se cumpla la desigualdad $|1 - \lambda f'(x)| \leqslant 1/2$. Resolviéndola, obtenemos para λ las restricciones siguientes:

$$\frac{1}{2m_1} \leqslant \lambda \leqslant \frac{3}{2M_1}. \tag{15}$$

Interpretación geométrica. Supongamos que se da la ecuación $f(x) = 0$ [$f(x)$ es la función continua]. Reduzcamos esta ecuación

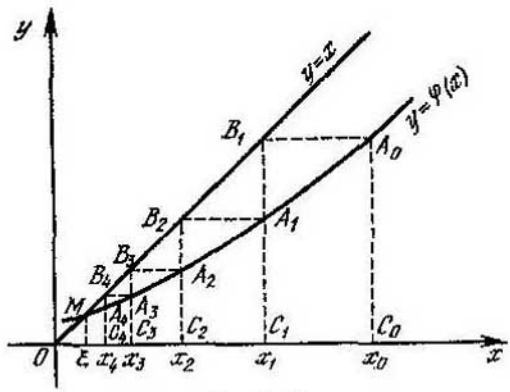


Fig. 5.25

a la forma $x = \varphi(x)$ y construyamos los gráficos de las funciones $y = x$ e $y = \varphi(x)$. La abscisa del punto de intersección de estas funciones es precisamente la raíz verdadera ξ (fig. 5.25).

Elijamos $x_0 \in [a, b]$ y determinemos $\varphi(x_0)$. Convengamos en designar la sucesión de los puntos que están sobre la curva $y = \varphi(x)$ con A_i $(i = 0, 1, 2, \ldots)$ y la sucesión de los puntos que están sobre la recta $y = x$, con B_i $(i = 1, 2, 3, \ldots)$. Del punto $A_0(x_0; \varphi(x_0))$ tracemos la recta paralela al eje Ox hasta que se interseque con la recta $y = x$; entonces obtenemos el punto $B_1(x_1; \varphi(x_0))$.

En efecto, $A_0C_0 = \varphi(x_0) = B_1C_1$, ya que $A_0B_1 \parallel OC_0$ y $B_1C_1 \parallel A_0C_0$. Pero $OC_1 = B_1C_1$ ($\triangle\, OC_1B_1$ es rectángulo e isósceles, puesto que la recta $y = x$ es la bisectriz del ángulo del cuadriculado). Por lo tanto, $x_1 = \varphi(x_0)$.

Tracemos $A_1B_2 \parallel OC_1$ y, repitiendo los razonamientos anteriormente expuestos, nos convencemos de que $x_2 = \varphi(x_1)$.

En la fig. 5.25 está representado el proceso iterativo convergente. La curva corta la bisectriz $y = x$ en el punto M con la abscisa ξ y para $x > \xi$ está debajo de la bisectriz; $\varphi'(x)$ satisface la condición

$0 < \varphi'(x) < 1$. Las aproximaciones sucesivas $x_0, x_1, \ldots, x_n, \ldots$ (las abscisas comunes de los puntos de los gráficos de ambas funciones) **decrecen monótonamente**. Cada aproximación sucesiva x_n es más próxima a la raíz verdadera que cada precedente x_{n-1}. La línea quebrada $A_0B_1A_1B_2A_2 \ldots$ tiene la forma de «escalera».

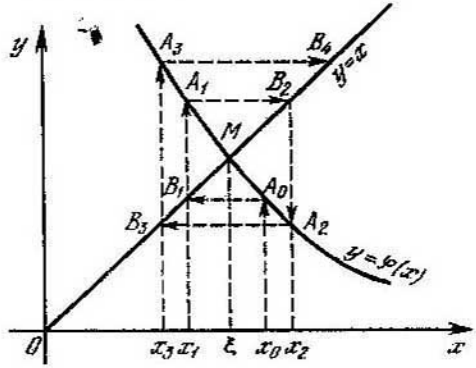


Fig. 5.26

En la fig. 5.26 la derivada $\varphi'(x) < 0$, pero en valor absoluto es menor que la unidad, o sea, $|\varphi'(x)| < 1$. El proceso iterativo converge, pero las aproximaciones **oscilan** cerca del valor exacto de la raíz. La quebrada $A_0B_1B_2A_2 \ldots$ tiene la forma de «espiral».

Ahora bien, si en cierto entorno (a, b) de la raíz ξ de la ecuación $x = \varphi(x)$ la derivada $\varphi'(x)$ conserva el signo constante y está cumplida la desigualdad $|\varphi'(x)| \leqslant q < 1$, con la particularidad de que $\varphi'(x) > 0$, entonces las aproximaciones sucesivas $x_n = \varphi(x_{n-1})$ $(n = 1, 2, \ldots)$, $x_0 \in [a\ b]$ convergen monótonamente a la raíz. En cambio, en el caso en que $\varphi'(x) < 0$, las aproximaciones sucesivas oscilan alrededor de la raíz ξ.

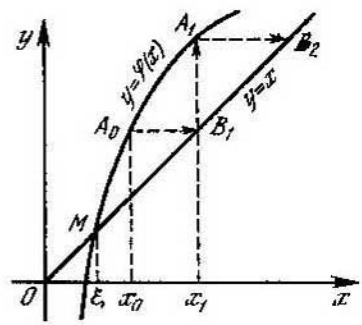


Fig. 5.27

En la fig. 5.27 se muestra el proceso iterativo divergente. Aquí $\varphi'(x) > 1$. La curva corta la bisectriz $y = x$ en el punto M y para $x > \xi$ está por encima de la bisectriz.

En la fig. 5.28 se representa el proceso iterativo divergente para el caso $|\varphi'(x)| > 1$. Las «aproximaciones» sucesivas se alejan del valor exacto de la raiz ξ.

Ejemplo 1. Con ayuda del método de iteraciones precisar hasta 10^{-4} la raíz de la ecuación $5x^3 - 20x + 3 = 0$, encerrada en el segmento $[0, 1]$.

△ La ecuación dada ha de reducirse a la forma $x = \varphi(x)$. Se puede hacer esto por varios procedimientos, por ejemplo:

$$x = x + (5x^3 - 20x + 3), \quad \text{entonces} \quad \varphi_1(x) = 5x^3 - 19x + 3;$$
$$x = \sqrt[3]{(20x - 3)/5}, \quad \text{entonces} \quad \varphi_2(x) = \sqrt[3]{(20x - 3)/5};$$
$$x = (5x^3 + 3)/20, \quad \text{entonces} \quad \varphi_3(x) = (5x^3 + 3)/20.$$

Determinemos cuál de las funciones obtenidas $\varphi(x)$ debemos utilizar para calcular las aproximaciones sucesivas. Recuérdese

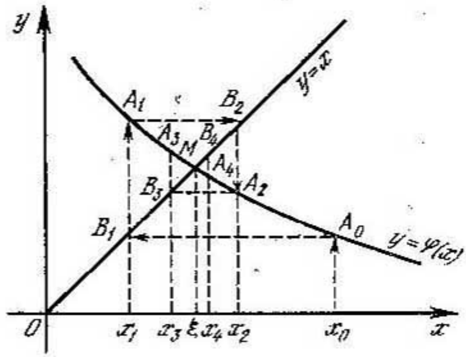


Fig. 5.28

que si $\varphi(x)$ satisface sobre el segmento $[a, b]$ la condición $|\varphi'(x)| \leqslant \leqslant q < 1$, el proceso iterativo converge. Hallamos

$$|\varphi_1'(x)| = |15x^2 - 19| > 1 \quad \text{sobre} \quad [0, 1],$$
$$|\varphi_3'(x)| = 15x^2/20 = 3x^2/4 < 1 \quad \text{sobre} \quad [0, 1].$$

Por consiguiente, se puede utilizar la función $\varphi_3(x)$ y buscar las aproximaciones sucesivas por el método de iteraciones valiéndose de la fórmula $x_n = (5x_{n-1}^3 + 3)/20$. Tomemos por aproximación inicial máx $\varphi'(x)$ sobre $[0, 1]$, o sea, $x_0 = 0{,}75$. Haciendo uso de la fórmula (7) determinemos cuál ha de ser la diferencia entre dos aproximaciones sucesivas para que quede alcanzada la exactitud prefijada:

$$|x_n - x_{n-1}| \leqslant \frac{0{,}0001 \cdot (1 - 0{,}75)}{0{,}75} = \frac{0{,}0001 \cdot 0{,}25}{0{,}75} = 0{,}00003.$$

Ahora bien, cuando el valor absoluto de la diferencia $|x_n - x_{n-1}|$ no supere 0,00003, será necesario cesar el proceso iterativo y suponer que la exactitud prefijada está alcanzada.

Es cómodo realizar el cálculo con ayuda de la tabla 5.11

Aquí se puede terminar el proceso iterativo y considerar $\xi = 0{,}1514$. ▲

Ejemplo 2. Calcular con exactitud hasta $\varepsilon = 10^{-5}$ la raíz de la ecuación $e^x - x^2 = 0$.

Tabla 5.11

n	x_n	x_n^3	$\varphi(x_n) = x_{n+1}$
0	0,75	0,42188	0,25547
1	0,2555	0,016777	0,154144
2	0,1541	0,005652	0,151413
3	0,1514	0,005443	0,151361
4	0,15136	0,005442	0,151361

△ Escribamos la ecuación en la forma $e^x = x^2$ y separemos gráficamente las raíces. Construyamos los gráficos de las funciones $y = e^x$ e $y = x^2$ (fig. 5.29). Del dibujo se ve que la ecuación $e^x - x^2 = 0$ tiene una raíz real que está sobre el segmento $[-0,8;\ -0,7]$.

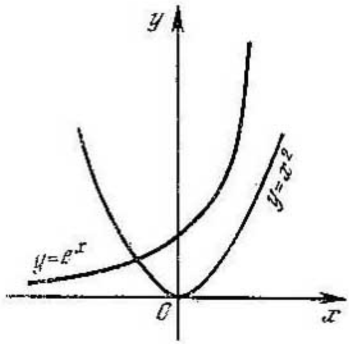


Fig. 5.29

Verifiquemos si verdaderamente es así. Separemos $f(-0,8)$ y $f(-0,7)$; tenemos $f(-0,8) = 0,44933 - 0,64 = -0,19067 < 0$, $(f - 0,7) = 0,49659 - 0,49 = 0,00659 > 0$. Puesto que los signos de la función $f(x) = e^x - x^2$ son opuestos en los extremos del segmento $[-0,8;\ -0,7]$, la raíz de la ecuación está dentro de este segmento.

Probemos hacer más estrecho el segmento, aplicando el método de pruebas. Encontramos $f(-0,75) = 0,49237 - 0,56250 < 0$ y $f(-0,7) > 0$. Por lo tanto, la raíz se halla sobre el segmento $[-0,75;\ 0,7]$. Estrechemos este segmento una vez más. Tenemos $f(-0,725) = 0,48432 - 0,52562 = -0,4130 < 0$ y $f(-0,7) > 0$. Por consiguiente la raíz está sobre el segmento $[-0,725;\ -0,7]$.

De la ecuación $e^x = x^2$ obtenemos que $x = -\sqrt{e^x}$ (delante del radical tomamos el signo «—», ya que conocemos que la raíz es negativa). Escribamos esta ecuación en la forma $x = e^{x/2}$ y comprobemos cuál será el proceso iterativo: convergente o divergente, o sea, se cumple o no la desigualdad $|\varphi'(x)| < 1$. En el ejemplo dado

$$\varphi(x) = -e^{x/2}, \quad \varphi'(x) = (1/2)\, e^{x/2}, \quad |\varphi'(-0,725)| = 0,34727;$$
$$|\varphi'(x)| = |\varphi'(-0,7)| = 0,35230.$$

Puesto que $|\varphi'(x)| < 1$, el proceso iterativo converge. En la fórmula (7) tomemos el número q igual a 0,36. Puesto que $\varepsilon = 10^{-5}$, entonces

$$|x_n - x_{n-1}| \leqslant \frac{0,00001\,(1-0,36)}{0,36} = 0,000018.$$

Ahora bien, la exactitud requerida puede ser alcanzada si se cumple la desigualdad $|x_n - x_{n-1}| \leqslant 0{,}000018$. Por aproximación nula se puede tomar cada uno de los extremos del segmento $[-0{,}725; -0{,}7]$ y todo punto dentro de este último. Tomemos $x_0 = -0{,}7$.
Hagamos los cálculos con ayuda de la tabla siguiente:

Tabla 5.12

n	x_n	$x_n/2$	$e^{x_n/2}$
0	−0,7	−0,35	−0,70460
1	−0,70460	−0,35230	−0,70307
2	−0,70307	−0,35154	−0,70360
3	−0,70360	−0,35180	−0,70342
4	−0,70342	−0,35171	−0,70348
5	−0,70348	−0,35174	−0,70346
6	−0,70346	−0,35173	−0,70347
7	−0,70347		

Puesto que $|x_7 - x_6| = |-0{,}70347 - (-0{,}70346)| = 0{,}00001$, la exactitud requerida de cálculos está alcanzada y $\xi \approx -70347$. ▲

Ejemplo 3. Haciendo uso del método de iteraciones, calcular con exactitud hasta 0,001 la raíz de la ecuación $x^3 + 3x^2 - 3 = 0$, situada sobre el segmento $[-2{,}75; -2{,}5]$ (véanse el ejemplo 1 del § 5.4 y el ejemplo 1 del § 5.5).

△ Encontramos $f'(x) = 3x^2 + 6x$. Por lo tanto,

$$M_1 = \max_{[-2,75;\ -2,5]} |f'(x)| = 6{,}189;\ m_1 = \min_{[-2,75;\ -2,5]} |f'(x)| = 3{,}75;$$

$$q = 1 - \frac{m_1}{M_1} = 1 - \frac{3{,}75}{6{,}189} < \frac{1}{2}.$$

Puesto que $q < 1/2$, para estimar el error se puede utilizar la fórmula (9). Tomemos $M_1 = 6$, entonces $\lambda = 1/6$ y

$$\varphi(x) = x - \lambda f(x) = x - \frac{1}{6}(x^3 + 3x^2 - 3).$$

El proceso iterativo correspondiente tiene la forma

$$x_{n+1} = x_n - \frac{1}{6}(x_n^3 + 3x_n^2 - 3);$$

entonces $|x_{n+1} - x_n| = \frac{1}{6}(x_n^3 + 3x_n^2 - 3)$. Los cálculos han de cesarse tan pronto como $|x_{n+1} - x_n| < \varepsilon$.

Realicemos los cálculos con ayuda de la tabla siguiente:

Tabla 5.13

n	x_n	x_n^3	$3x_n^2$	$\frac{1}{6}(x^3+3x^2-3)$
0	−2,5	−15,625	18,75	0,02
1	−2,52	−16,0030	19,0512	0,0080
2	−2,5280	−16,1559	19,1724	0,0028
3	−2,5308	−16,2096	19,2148	0,0008
4	−2,5316			

Así pues, por valor aproximado de la raíz con exactitud hasta 0,001 se puede tomar $\xi = -2{,}532$. ▲

§ 5.8. Propiedades generales de las ecuaciones algebraicas. Determinación de la cantidad de raíces reales de una ecuación algebraica

Propiedades generales de las ecuaciones algebraicas. Escribamos una ecuación algebraica de n-ésimo grado:

$$P_n(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n = 0, \qquad (1)$$

donde $P_n(x)$ es el polinomio de n-ésimo grado; n, el grado superior de la incógnita; $a_0, a_1, \ldots, a_n$ son los coeficientes reales.

Es sabido que todo número ξ que anula el polinomio, o sea tal que $P_n(\xi) = 0$, se llama raíz del polinomio.

El número ξ es la raíz del polinomio $P_n(x)$ si y sólo si $P_n(x)$ se divide exactamente por $x - \xi$. Recuérdese que si en este caso $P_n(x)$ se divide exactamente por $(x - \xi)^k$ $(k \geqslant 1)$, pero no se divide ya por $(x - \xi^{k+1})$, ξ se denomina raíz de k-ésimo grado de multiplicidad (o raíz de multiplicidad k) del polinomio $P_n(x)$. Las raíces de multiplicidad $k = 1$ se llaman raíces simples del polinomio.

Surge la pregunta ¿todo polinomio tiene obligatoriamente raíces? La respuesta a esta pregunta se da por el teorema siguiente que citamos sin demostración.

Teorema 1 (teorema fundamental del álgebra). *Todo polinomio con todos coeficientes reales el grado del cual no sea menor que la unidad tiene, al menos, una raíz que en el caso general es compleja.*

De este teorema se deduce un corolario importante que dice: *todo polinomio $P_n(x)$ de grado n $(n \geqslant 1)$ con cualesquiera coeficientes reales tiene exactamente n raíces, reales o complejas, si cada una de las raíces se cuenta tantas veces cuanta es su multiplicidad.*

Ahora bien, las raíces de la ecuación algebraica (1) pueden ser tanto reales como complejas.

Las raíces complejas de la ecuación (1) poseen la propiedad de *conjugación par*, o sea, si la ecuación (1) tiene una raíz compleja

$\xi = \alpha + \beta i$ (donde α y β son los números reales) de multiplicidad k, esta ecuación tiene asimismo la raíz compleja $\bar{\xi} = \alpha - \beta i$ también de multiplicidad k. Los módulos de estas raíces son iguales: $|\xi| = |\bar{\xi}| = \sqrt{\alpha^2 + \beta^2}$.

Si la ecuación (1) tiene raíces complejas, su número es par. Por eso toda ecuación algebraica de grado impar con coeficientes reales tiene, como mínimo, una raíz real.

Antes de calcular las raíces de una ecuación algebraica es necesario primeramente: a) determinar la cantidad de raíces que posee la ecuación dada; b) hallar el dominio de existencia de las raíces (determinar las cotas superior e inferior de situación de las raíces). Luego se puede comenzar a separar las raíces y precisarlas.

Determinación de la cantidad de raíces reales de las ecuaciones algebraicas. La cantidad de raíces reales positivas de la ecuación algebraica (1)

$$P_n(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n = 0$$

puede ser determinada aproximadamente con ayuda de la **regla de Descartes:** *la cantidad de raíces positivas reales de la ecuación algebraica $P_n(x) = 0$ con coeficientes reales (raíces cada una de las cuales se cuenta tantas veces cuanta es su multiplicidad) ora es igual al número de cambios del signo en la sucesión de los coeficientes de la ecuación $P_n(x) = 0$ ora es en un número par menor (los coeficientes iguales a cero no se tienen en cuenta).*

La cantidad de raíces negativas es igual al número de cambios del signo en la sucesión de los coeficientes $P_n(-x)$ o en un número par es menor.

Si la ecuación es completa, la cantidad de sus raíces positivas es igual al número de cambios del signo en la sucesión de los coeficientes o en un número par es menor y la cantidad de raíces negativas es igual al número de constancias del signo o en un número par es menor.

Ejemplo 1. Determinar la cantidad de raíces positivas y negativas de la ecuación $x^5 - 17x^4 + 12x^3 + 7x^2 - x + 1 = 0$.

△ Según el teorema fundamental del álgebra esta ecuación tiene cinco raíces (de ellas una raíz, como mínimo, es real).

La ecuación es completa, la sucesión de signos de los coeficientes es tal: $+, -, +, +, -, +$. El signo cambia cuatro veces, por consiguiente, hay cuatro o dos raíces positivas o bien ninguna raíz es positiva.

El número de constancias del signo es igual a 1, por lo tanto, la ecuación dada tiene una raíz negativa. ▲

Ejemplo 2. Determinar la cantidad de raíces reales positivas y negativas de la ecuación $x^6 - 3x^4 + x^3 + x^2 - 1 = 0$.

△ La ecuación dada tiene seis raíces; la sucesión de los signos $+, -, +, -$. Tiene lugar tres cambios del signo; por consiguiente, hay tres raíces positivas o una raíz positiva. Luego, para el polinomio

$$P_n(-x) = x^6 - 3x^4 - x^3 + x^2 - 1$$

la sucesión de los signos es tal: $+$, $-$, $-$, $+$, $-$. Aquí también tenemos tres cambios del signo, por eso hay tres raíces negativas o una raíz negativa. ▲

La cantidad de raíces de una ecuación algebraica se puede determinar más exactamente con ayuda del teorema de Sturm.

Puesto que las raíces múltiplas de la ecuación pueden siempre ser separadas como raíces generales de las ecuaciones $P_n(x) = 0$ y $P'_n(x) = 0$, sin limitar la generalidad se puede considerar que la ecuación dada $P_n(x) = 0$ tiene sólo raíces simples.

Supongamos que para la ecuación algebraica dada $P_n(x) = 0$ hemos determinado, por un procedimiento cualquiera, que todas sus raíces reales están en el intervalo (a, b) (a y b son los números reales y no son raíces de la ecuación; $a < b$). Determinemos la primera derivada $P'_n(x)$ y dividamos por ésta el polinomio $P_n(x)$. Tomemos el resto de la división de $P_n(x)$ por $P'_n(x)$ con el signo contrario y designémoslo con $R_1(x)$.

Luego, de un modo exactamente igual dividamos $P'_n(x)$ por $R_1(x)$; tomemos con el signo contrario el resto obtenido y designémoslo con $R_2(x)$. Al dividir $R_1(x)$ por $R_2(x)$ y al volver a tomar el resto con el signo opuesto, obtenemos $R_3(x)$. Continuamos el proceso de división hasta que se obtenga un resto que sea una magnitud constante. Tomemos esta magnitud también con el signo contrario.

Obtenemos una sucesión de funciones

$$P_n(x),\ P'_n(x),\ R_1(x),\ R_2(x),\ \ldots,\ R_{m-1}(x),\ R_m = \text{const}$$

que se llama *sistema de Sturm*. En este sucesión sustituimos en vez de x primero a, luego b y contamos el número de cambios del signo en ambos casos [designemos los números obtenidos con $W(a)$ y $W(b)$, respectivamente].

Teorema 2 (teorema de Sturm). *Si los números reales a y b ($a < b$) no son raíces del polinomio $P_n(x)$, que no tiene raíces múltiplas, entonces $W(a) \geqslant W(b)$ y la diferencia $W(a) - W(b)$ es igual a la cantidad de raíces reales del polinomio $P_n(x)$, encerradas entre a y b.*

Con ayuda del teorema de Sturm se puede determinar la cantidad de raíces negativas de la ecuación $P_n(x) = 0$, [o sea, la cantidad de raíces reales de la ecuación $P_n(x) = 0$ en el intervalo $(-\infty, 0)$], así como la cantidad de raíces positivas [en el intervalo $(0, +\infty)$]. El teorema de Sturm se aplica también para separar las raíces. Las funciones que forman parte del sistema de Sturm pueden multiplicarse y dividirse por números positivos arbitrarios. Esto simplifica mucho el trabajo cuando se realiza la división inexacta.

Ejemplo 3. Determinar la cantidad de raíces reales de la ecuación $5x^3 - 20x + 3 = 0$, así como separar estas raíces haciendo uso del teorema de Sturm.

△ Formamos el sistema de funciones de Sturm. Tenemos $P_n(x) = 5x^3 - 20x + 3$; $P'_n(x) = 15x^2 - 20$. Para determinar

$R_1(x)$ multipliquemos $P_n(x)$ por 3 y dividamos el resultado por $P'_n(x)$

$$\begin{array}{r|l} 15x^3-60x+9 & 15x^2-20 \\ \cline{2-2} \mp 15x^3 \pm 20x & x \\ \cline{1-1} -40x+9 & \end{array}$$

Por lo tanto, $R_1(x)=40x-9$ (hemos tomado el resto con el signo contrario). Multipliquemos $P'_n(x)$ por 8 y dividamos este producto por $R_1(x)$:

$$\begin{array}{r|l} 120x^2-160 & 40x-9 \\ \cline{2-2} \mp 120x^2 \pm 27x & 3x+27 \\ \cline{1-1} 40\,(27x-160) & \\ \hline 40\cdot 27x-40\cdot 160 & \\ \mp 40\cdot 27x \pm 9\cdot 27 & \\ \hline - & \end{array}$$

Puesto que el último resto es un número constante con el signo «—» (en este caso nos interesa precisamente el signo del resto constante), lo cambiamos por el opuesto, es decir, por «+».

Componemos la siguiente tabla de los signo de funciones que forman parte del sistema de Sturm:

x	$P_n(x)$	$P'_n(x)$	$R_1(x)$	R_2	$W(x)$
$-\infty$	$-$	$+$	$-$	$+$	3
0	$+$	$-$	$-$	$+$	2
$+\infty$	$+$	$+$	$+$	$+$	0

De la tabla se ve que en el intervalo $(-\infty, +\infty)$ hay tres raíces reales [ya que $W(-\infty)-W(+\infty)=3-0=3$]. Una de ellas es negativa [$W(-\infty)-W(0)=3-2=1$] y dos son positivas [$W(0)-W(+\infty)=2-0=2$].

Utilizando este teorema de Sturm, separamos las raíces al abreviar los intervalos hasta la longitud igual a 1:

x	$P_n(x)$	$P'_n(x)$	$R_1(x)$	$R_2(x)$	$W(x)$
—	$-$	$+$	$-$	$+$	3
-3	$-$	$+$	$-$	$+$	3
-2	$+$	$+$	$-$	$+$	2
-1	$+$	$-$	$+$	$+$	2
0	$+$	$-$	$-$	$+$	2
1	$-$	$-$	$+$	$+$	1
2	$+$	$+$	$+$	$+$	0

De la última tabla se ve que las raíces están situadas en los intervalos $(-3, -2)$, $(0, 1)$, $(1, 2)$. ▲

§ 5.9. Determinación del dominio de existencia de las raíces de una ecuación algebraica

Regla del anillo. *Sea dada una ecuación algebraica* $P_n(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_{n-1}x + a_n = 0$, *donde* $a_0, a_1, \ldots, a_n$ *son los coeficientes reales y sea* $A = \text{máx}\{|a_1|, |a_2|, \ldots, |a_n|\}$, $B = \text{máx}\{|a_0|, |a_1|, \ldots, |a_{n-1}|\}$.
Entonces las raíces de la ecuación están encerradas en el anillo circular $r < |x| < R$, *donde*

$$r = \frac{1}{1+B/|a_n|}; \quad R = 1 + \frac{A}{|a_0|}.$$

En este caso r es la cota inferior y R, la cota superior de las raíces positivas de la ecuación algebraica $P_n(x) = 0$ y $-R$, $-r$ son las cotas inferior y superior, respectivamente, de las raíces negativas (fig. 5.30).

Ejemplo 1. Determinar las cotas de las raíces de la ecuación $5x^3 - 20x + 3 = 0$.

△ Aquí $|a_0| = 5$, $A = 20$, $|a_n| = 3$, $B = 20$, o sea,

$$R = 1 + \frac{A}{|a_0|} = 1 + \frac{20}{5} = 5;$$

$$r = \frac{1}{1+B/|a_n|} = \frac{1}{1+20/3} = \frac{3}{23} \approx 0{,}013.$$

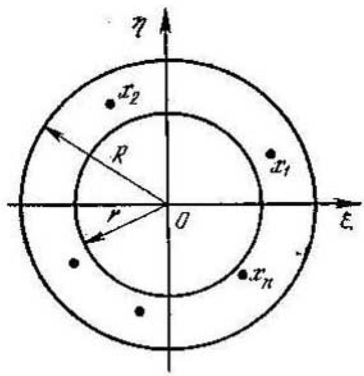


Fig. 5.30

Entonces si las raíces reales de la ecuación $5x^3 - 20x + 3 = 0$ existen (y éstas existen obligatoriamente, ya que la ecuación es de grado impar), se disponen en el intervalo $(-5; 5)$; en este caso las raíces negativas están en el intervalo $(-5; -0{,}013)$ y las positivas, en el intervalo $(0{,}013; 5)$. ▲

Al resolver las ecuaciones es cómodo determinar primero las cotas de las raíces y luego aplicar el teorema de Sturm. Según la regla del anillo estas cotas se determinan muy aproximadamente.

Vamos a mostrar el procedimiento de una determinación más exacta de las raíces reales de la ecuación algebraica $P_n(x) = 0$. Si R_1 es la cota superior de las raíces positivas de $P_n(x)$, R_2 es la cota superior de las raíces positivas de $P_n(-x)$, $R_3 > 0$ es la cota superior de las raíces positivas de $x^nP_n(1/x)$ y R_4 es la cota superior de las raíces positivas de $x^nP_n(-1/x)$, entonces todas las raíces reales, distintas del cero, de la ecuación $P_n(x) = 0$ (si éstas existen) están dentro de los intervalos $(-R_2, -1/R_4)$ y $(1/R_3, R_1)$.

Para determinar la cota superior de las raíces positivas de una ecuación algebraica se puede utilizar los métodos de Lagrange o de Newton.

Método de Lagrange. *Si los coeficientes del polinomio*

$$P_n(x) = a_0x^n + a_1x^{n-1} + a_2x^{n-2} + \ldots + a_n$$

satisfacen las condiciones $a_0 > 0$, $a_1, a_2, \ldots, a_{m-1} \geqslant 0$, $a_m < 0$, *la cota superior de las raíces positivas de las ecuaciones* $P_n(x) = 0$ *se determina con ayuda de la fórmula* $R = 1 + \sqrt[m]{B/a_0}$, *donde* B *es el máximo entre los valores absolutos de los coeficientes negativos de* $P_n(x)$.

Ejemplo 2. Haciendo uso del método de Lagrange, determinar las cotas de las raíces positivas y negativas de la ecuación

$$8x^4 - 8x^2 - 32x + 1 = 0.$$

$\triangle$ Aquí $a_0 = 8 > 0$; $a_1 = 0$; $a_2 = -8 < 0$; $a_3 = -32$; $a_4 = 1$, $m = 2$ (número del primero entre los coeficientes negativos), $B = 32$. Por consiguiente, $R_1 = 1 + \sqrt{32/8} = 3$.

Consideremos el polinomio

$$P_n(-x) = 8x^4 - 8x^2 + 32x + 1.$$

De un modo análogo encontramos que para este polinomio como cota superior de las raíces positivas sirve $R_2 = 1 + \sqrt{8/8} = 2$.

Luego, para el polinomio

$$x^4P_n(1/x) = x^4 - 32x^3 - 8x^2 + 8$$

tenemos $a_0 = 1 > 0$; $a_1 = -32 < 0$, o sea, $m = 1$; $B = 32$; $R_3 = 1 + 32 = 33$.

Por último, para el polinomio

$$x^4P_n(-1/x) = x^4 + 32x^3 - 8x^2 + 8$$

tenemos $a_0 = 1 > 0$; $a_1 = 32$; $a_2 = -8$; $a_3 = 0$; $a_4 = 8$, o sea, $m = 2$. Por eso $R_4 = 1 + \sqrt{8} = 1 + 2\sqrt{2} = 3{,}828$.

Por lo tanto, si la ecuación $8x^4 - 8x^2 - 32x + 1 = 0$ tiene raíces reales, éstas se hallan obligatoriamente en los intervalos $(-2, -1/3{,}828)$ y $(1/33, 3)$. $\blacktriangle$

Método de Newton. *Si para* $x = c$ *el polinomio*

$$P_n(x) = a_0x^n + a_1x^{n-1} + \ldots + a_n$$

y sus derivadas $P_n'(x)$, $P_n''(x)$, ... *toman los valores positivos, entonces* c *es la cota superior de las raíces positivas de la ecuación* $P_n(x) = 0$.

Ejemplo 3. Haciendo uso del método de Newton, determinar la cota superior de las raíces positivas de la ecuación $8x^4 - 8x^2 - 32x + 1 = 0$.

$\triangle$ Hallamos

$$P(x) = 8x^4 - 8x^2 - 32x + 1; \quad P'(x) = 32x^3 - 16x - 32;$$
$$P''(x) = 96x^2 - 16; \quad P'''(x) = 192x; \quad P^{\mathrm{IV}}(x) = 192.$$

Han de verificarse los valores de $x > 0$. Para $x = c = 1$ tenemos $P(1) < 0$. Por consiguiente, luego no debe realizarse la verificación para $x = 1$. Comprobemos el valor $x = c = 2$: $P(2) > 0$; $P'(2) > 0$; $P''(2) > 0$; $P'''(2) > 0$; $P^{IV} > 0$. Ahora bien, de cota superior de las raíces positivas sirve el número 2, o sea, $R = 2$. En calidad de la cota inferior se puede tomar el número inverso a R, o sea, $r = 1/2$. ▲

§ 5.10. Método de Horner para precisar las raíces reales de una ecuación algebraica

Sea dada la ecuación algebraica

$$P(x) = x^n + a_1 x^{n-1} + a_2 x^{n-2} + \ldots + a_n = 0, \qquad (1)$$

donde $a_1, a_2, \ldots, a_n$ son los coeficientes reales del polinomio.

Representemos la raíz buscada de la ecuación, escrita en el sistema decimal, en la forma

$$X = c_0 \cdot 10^m + c_1 \cdot 10^{m-1} + c_2 \cdot 10^{m-2} + \ldots + c_k \cdot 10^{m-k} + \ldots$$

El método de Horner consiste en determinar sucesivamente las cifras de la raíz c_0, c_1, ... con ayuda de transformaciones especiales de la ecuación (1).

Si se emplea la sustitución $x = 10^m \xi$ (para $c_0 > 0$) o bien $x = -10^m \xi$ (para $c_0 < 0$), la ecuación (1) se transforma en ecuación

$$f_1(\xi) = \xi^n + a_1' \xi^{n-1} + a_2' \xi^{n-2} + \ldots + a_n' = 0$$

cuya raíz está encerrada en el intervalo (0, 10). Por eso a continuación vamos a examinar precisamente este caso particular. Esto simplifica el proceso de cálculo de la raíz, aunque no es obligatorio para la aplicación del método de Horner.

Puesto que $0 < X < 10$, la raíz de la ecuación (1) se puede escribir en la forma

$$X = c_0 + \frac{c_1}{10} + \frac{c_2}{10^2} + \frac{c_3}{10^3} + \ldots = c_0,\ c_1 c_2 c_3 \ldots$$

Es fácil hallar la primera cifra de la raíz c_0 con ayuda del método de tabla o utilizando el intervalo $[a, b]$ del aislamiento de la raíz.

Luego, aplicando a la ecuación (1) la transformación

$$x - c_0 = y, \qquad (2)$$

obtenemos $x = y + c_0$, de donde

$$P(x) = P(y + c_0) = -\varphi(y) = 0. \qquad (3)$$

Como raíz de la última ecuación sirve, evidentemente, el número $y = 0,\ c_1 c_2 c_3 \ldots$ Aplicando a la ecuación (3) la sustitución

$$y = Y/10, \qquad (4)$$

llegamos a la ecuación

$$P_1(Y) = 0, \qquad (3')$$

que tiene la raíz $Y = 10y = c_1, c_2c_3 \ldots$ De la ecuación (3′) determinamos la cifra c_1 que es la segunda cifra de la raíz de la ecuación (1). Luego se puede volver a aplicar la sustitución del tipo (2) ó (4), pero ahora a la ecuación (3′). Como resultado obtenemos la ecuación

$$P_2(Z) = 0, \tag{5}$$

de cuya raíz sirve el número $z = c_2, c_3, \ldots$ Después se determina la cifra c_2.

El proceso descrito puede ser repetido hasta que se obtenga la cantidad requerida de cifras. El método de Horner puede aplicarse en combinación con cualquier otro método, por ejemplo, con el de las cuerdas, con el de las tangentes o bien con el combinado de las cuerdas y las tangentes. Primero se determinan algunas cifras de la raíz con ayuda del método de Horner y luego la raíz se precisa por otros métodos.

Cuando se necesita obtener una pequeña cantidad de cifras de la raíz, se puede utilizar el procedimiento siguiente.

Si una de las raíces de la ecuación es mucho menor que las demás, esta raíz puede ser calculada aproximadamente por dividir el término independiente, tomado con el signo «—», por el coeficiente de la x de primer grado.

¿Cómo realmente se realizan las sustituciones (2) y (4)?

Para ejecutar la sustitución (4) conviene multiplicar los coeficientes $a_1, a_2, \ldots, a_n$ de la ecuación inicial por $10, 10^2, 10^3, \ldots \ldots, 10^n$, respectivamente.

La sustitución (2) se lleva a cabo después de desarrollar previamente el polinomio $P(x)$ según las potencias de $x - c_0$. Tal desarrollo puede ser efectuado por la fórmula de Taylor o bien utilizando la siguiente serie de igualdades:

$$\begin{aligned} P(x) &= q_1(x)(x - c_0) + r_0, \\ q_1(x) &= q_2(x)(x - c_0) + r_1, \\ q_2(x) &= q_3(x)(x - c_0) + r_2, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ q_{n-2}(x) &= q_{n-1}(x)(x - c_0) + r_{n-2}, \\ q_{n-1}(x) &= (x - c_0) + r_{n-1}, \end{aligned} \tag{6}$$

donde $q_i(x)$ y r_{i-1} son el cociente y el resto, respectivamente, que se obtienen al dividir $q_{i-1}(x)$ por $(x - c_0)$. Eliminando sucesivamente $q_i(x)$ de la igualdad (6), llegamos a la identidad

$$\begin{aligned} P(x) = (x - c_0)^n + r_{n-1}(x - c_0)^{n-1} + r_{n-2}(x - c_0)^{n-2} + \ldots \\ \ldots + r_1(x - c_0) + r_0, \end{aligned} \tag{7}$$

o sea, obtenemos el desarrollo de $f(x)$ según las potencias de $x - c_0$. El cálculo de cada uno de los restos r_i $(i = 0, 1, \ldots, n - 1)$, o sea de los coeficientes del desarrollo, se lleva a cabo con ayuda del esquema de Horner.

El sistema de igualdades (6) corresponde a la tabla

c_0

1	a_1	a_2	a_3	$\ldots$	a_{n-1}	a_n
1	b_1	b_2	b_3	$\ldots$	b_{n-1}	r_0
1	e_1	e_2	e_3	$\ldots$	r_1	
.	..	..	..	..	..	..
1	r_{n-1}					

donde $b_1 = 1\cdot c_0$, $b_2 = b_1c_0$, $b_3 = b_2c_0$, . . ., $e_1 = 1\cdot c_0$, $e_2 = e_1c_0$, $e_3 = e_2c_0$, . . . Si el coeficiente de x^n es igual a a_0, en la primera columna en vez de la unidad debe ser a_0.

Ejemplo. Haciendo uso del método de Horner hallar la menor raíz de la ecuación $x^3 + 3x^2 - 3 = 0$ con seis cifras significativas; las raíces de la ecuación están separadas y la menor de ellas se halla sobre el segmento $[-3, -2]$.

△ 1) Puesto que la raíz es negativa, transformemos la ecuación inicial con ayuda de la sustitución $x = -\tilde{x}$:

$$-\tilde{x}^3 + 3\tilde{x}^2 - 3 = 0 \quad \text{o bien} \quad \tilde{x}^3 - 3\tilde{x}^2 + 3 = 0.$$

La raíz buscada de esta ecuación $\tilde{x} \in [2, 3]$. Por lo tanto, la primera cifra de la ecuación transformada es 2. Entonces las sustituciones (2) y (4) tienen la forma $\tilde{x} - 2 = y$ e $y = Y/10$. Realizamos la primera sustitución con ayuda del esquema de Horner:

$c_0 = 2$

1	-3	0	3
1	-1	-2	$\boxed{-1}$
1	1	$\boxed{0}$	
1	$\boxed{3}$		

De esta tabla según la fórmula (7) obtenemos $\varphi(y) = y^3 + 3y - 1$ y $P_1(Y) = Y^3 + 30Y^2 - 1000 = 0$.

A continuación vamos a escribir inmediatamente con ayuda de la tabla ambas transformaciones juntas, conservando la designación y, es decir, en vez de la última ecuación vamos a escribir $P_1(y) = y^3 + 30y^2 - 1000 = 0$.

Vamos a determinar los valores del polinomio $P_1(y)$ para ciertos valores de $y \in [0, 10]$ utilizando el esquema de Horner:

y	1	30	0	-1000
5	1	35	175	-125
6	1	36	216	296

Puesto que $P_1(5) < 0$ y $P_1(6) > 0$, entonces $y \in [5, 6]$. De aquí se deduce que $c_1 = 5$.

3) Para determinar la cifra siguiente planteemos la ecuación, cuyos coeficientes obtenemos con ayuda del esquema de Horner, haciendo uso de las sustituciones (2) y (4) para $c_1 = 5$:

$c_1 = 5$

1	30	0	−1000
1	35	175	$\boxed{-125}$
1	40	$\boxed{375}$	
1	$\boxed{45}$		

Ahora bien, $P_2(y) = y^3 + 450y^2 + 37\,500y - 125\,000 = 0$.

Determinemos los valores de $P_2(y)$ para ciertos valores de $y \in [0, 10]$ con ayuda del esquema de Horner:

y	1	450	37 500	−125 000
3	1	453	38 859	−8 423
4	1	454	39 316	32 264

Puesto que $P_2(3) < 0$ y $P_2(4) > 0$, entonces $y \in [3, 4]$. De aquí resulta que $c_2 = 3$.

3) Planteemos la ecuación para determinar las cifras sucesivas, en la cual se puede hallar los coeficientes utilizando el esquema de Horner:

$c_2 = 3$

1	450	37 500	−125 000
1	453	38 859	$\boxed{-8\,423}$
1	456		
1	$\boxed{459}$		

Por lo tanto, $P_3(y) = y^3 + 4590y^2 + 4\,022\,700y - 8\,423\,000 = 0$.

A la ecuación obtenida se puede aplicarle el procedimiento especial descrito anteriormente. Como resultado de la triple sustitución (4) todas las raíces, a excepción de la buscada, han aumentado 10^3 veces, aproximadamente. Entonces la raíz de la última ecuación es igual aproximadamente a $8\,423\,000/4\,022\,700 \approx 2{,}09$. Esto quiere decir que las cifras 2, 0 y 9 sirven de cifras decimales posteriores de la raíz de la ecuación inicial. Como resultado hallamos la raíz de la ecuación dada: $X = -2{,}53209$. ▲

Ejercicios

1. Separar las raíces analíticamente y precisarlas hasta 0.001 con ayuda del método de pruebas:

a) $x^3 - x + 1 = 0$; b) $x^3 + 2x - 4 = 0$; c) $x^4 + 5x - 3 = 0$; d) $2{,}2x - 2^x = 0$; e) $2^x - 2x^2 - 1 = 0$; f) $2^x - 4x = 0$.

2. Separar las raíces gráficamente y precisarlas hasta 0,001 con ayuda del método de las cuerdas:

a) $x^3 + x - 3 = 0$; b) $x^3 + 8x - 6 = 0$; c) $x^3 + 10x - 9 = 0$; d) $x^2 - \cos \pi x = 0$; e) $x^2 - \operatorname{sen} \pi x = 0$; f) $\log x - \frac{1}{x^2} = 0$.

3. Haciendo uso del método de las tangentes, hallar con exactitud hasta 0,001 las raíces de las ecuaciones:

a) $x^3 - 6x^2 + 9x - 3 = 0$; b) $x^3 - 12x - 8 = 0$;

c) $x^3 + 4x^2 - 6 = 0$; d) $2 \log x - \frac{x}{2} + 1 = 0$;

e) $x^2 - 20 \operatorname{sen} x = 0$; e) $x - \cos x = 0$.

4. Utilizando el método combinado de las cuerdas y las tangentes determinar con exactitud hasta 0,001 las raíces de las ecuaciones:

a) $x^3 + 6x - 5 = 0$; b) $x^3 - 2x + 7 = 0$; c) $x^3 - 2x^2 + x + 1 = 0$; d) $1{,}8x^2 - \operatorname{sen} 10x = 0$; e) $\log x - \frac{7}{2x+6} = 0$; f) $2x \log x - 1 = 0$.

5. Haciendo uso del teorema de Sturm, separar las raíces de las ecuaciones y precisarlas hasta 0,001 con ayuda del método de iteraciones:

a) $x^3 + 4x - 3 = 0$; b) $x^4 - 2x - 1 = 0$;
c) $x^5 - 5x + 2 = 0$; d) $x^4 + x - 3 = 0$.

6. Utilizando el método de iteraciones, hallar las raíces de las ecuaciones con exactitud hasta 0,001:

a) $\ln x + (x + 1)^3 = 0$; b) $\sqrt{x + 1} = 1/x$; c) $x - \cos x = 0$;
d) $3x - \cos x - 1 = 0$; e) $x + \log x = 0{,}5$.

CAPÍTULO VI

Determinación de los valores propios de una matriz y de sus vectores propios

§ 6.1. Polinomio característico

Sean dadas la matriz cuadrada A y el vector columna no nulo $\mathbf{x}$

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Multiplicando la matriz A por el vector $\mathbf{x}$, obtenemos el vector columna

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

o sea,

$$\mathbf{y} = A\mathbf{x}. \tag{1}$$

Si resulta que las coordenadas y_i $(i = 1, 2, \ldots, n)$ del vector $\mathbf{y}$ son proporcionales a las coordenadas correspondientes x_i del vector dado $\mathbf{x}$ con el coeficiente de proporcionalidad λ, o sea, si $y_i = \lambda x_i$ y, por consiguiente,

$$\mathbf{y} = \lambda\mathbf{x}, \tag{2}$$

el vector columna no nulo $\mathbf{x}$ se llama *vector propio* de la matriz A y el coeficiente de proporcionalidad λ, *valor propio* (o *número característico*) de la matriz A. Puesto que $\mathbf{y} = A\mathbf{x}$ e $\mathbf{y} = \lambda\mathbf{x}$, es evidente que

$$A\mathbf{x} = \lambda\mathbf{x}. \tag{3}$$

Ahora bien, si se cumple la condición (3), el vector $\mathbf{x}$ es el vector propio de la matriz A el cual corresponde al valor propio λ de esta última.

Ejemplo 1. Sea

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 6 & -2 \\ 3 & 4 & -1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Entonces

$$Ax = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 6 & -2 \\ 3 & 4 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 6 \\ 6 \\ 6 \end{bmatrix} = 6 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Por lo tanto, el número $\lambda = 6$ es el valor propio de la matriz A ya que se cumple la igualdad (3):

$$A_x = \lambda_x = 6 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

y el vector $\mathbf{x} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ es el vector propio de la matriz A el cual corresponde al valor propio $\lambda = 6$.

Escribamos la relación (3) en la forma $A\mathbf{x} - \lambda\mathbf{x} = \mathbf{0}$ o bien

$$(A - \lambda E)\,\mathbf{x} = \mathbf{0}, \tag{4}$$

donde E es la matriz unidad del mismo tamaño que la matriz A y $\mathbf{0}$, el vector columna nulo. Es evidente que sin el factor E de λ la ecuación (4) no habría tenido sentido.

Puesto que

$$\lambda E = \begin{vmatrix} \lambda & 0 & 0 & \dots & 0 \\ 0 & \lambda & 0 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \dots & \lambda \end{vmatrix}, \quad \mathbf{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

la ecuación (4) se puede escribir en la forma

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ v_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{5}$$

La relación (5) es un sistema homogéneo lineal de ecuaciones el cual tiene soluciones no nulas si y sólo si su determinante es igual a cero, o sea, cuando se cumple la condición

$$\det (A - \lambda E) = 0. \tag{6}$$

La ecuación (6) se llama *ecuación característica* de la matriz A y su primer miembro, *polinomio característico* (o *determinante característico*) de la matriz A.

En la forma desarrollada la ecuación característica se escribe así:

$$\begin{vmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22}-\lambda & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn}-\lambda \end{vmatrix} = 0. \tag{7}$$

Si se desarrolla el determinante en el primer miembro de la ecuación (7), se obtiene el polinomio de n-ésimo grado respecto a λ:

$$D(\lambda) = \det(A - \lambda E) =$$
$$= (-1)^n [\lambda^n - p_1\lambda^{n-1} + p_2\lambda^{n-2} - \dots + (-1)^{(n)}p_n]. \tag{8}$$

La magnitud λ, que se determina de la ecuación (8), toma n valores $\lambda_1, \lambda_2, \dots, \lambda_n$ entre los cuales pueden ser también iguales. Para hallar los vectores propios $\mathbf{x}^{(i)}$ correspondientes a los valores propios λ_i ($i = 1, 2, \dots, n$) es necesario resolver el sistema homogéneo lineal de ecuaciones (5) para cada valor λ_i.

Ejemplo 2. Hallar los valores propios y los vectores propios de la matriz $A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

△ 1) Escribamos el polinomio característico de la matriz A y determinemos λ. Tenemos

$$\det(A - \lambda E) = \begin{vmatrix} 2-\lambda & 1 \\ 1 & 2-\lambda \end{vmatrix} = \lambda^2 - 4\lambda + 3.$$

La ecuación característica $\lambda^2 - 4\lambda + 3 = 0$ tiene dos raíces $\lambda_1 = 1$ y $\lambda_2 = 3$ que son precisamente los valores propios de la matriz A.

2) Determinemos el vector propio $\mathbf{x}^{(1)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ correspondiente al valor $\lambda_1 = 1$. La ecuación matricial

$$(A - \lambda_1 E)\,\mathbf{x}^{(1)} = 0 \quad \text{o bien} \quad \begin{bmatrix} 2-1 & 1 \\ 1 & 2-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

es equivalente al sistema $\begin{cases} x_1 + x_2 = 0, \\ x_1 + x_2 = 0 \end{cases}$ el cual tiene un conjunto infinito de soluciones en la forma $x_1 = -x_2$. Suponiendo $x_1 = c$ (c es todo número), obtenemos $x_2 = -c$. Entonces el vector propio buscado se escribirá así: $\mathbf{x}^{(1)} = c\begin{bmatrix} 1 \\ -1 \end{bmatrix}$.

3) Determinemos el vector propio $\mathbf{x}^{(2)} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ correspondiente al segundo valor propio $\lambda_2 = 3$. Tenemos

$$(A - \lambda_2 E)\,\mathbf{x}^{(2)} = 0 \quad \text{o bien} \quad \begin{bmatrix} 2-3 & 1 \\ 1 & 2-3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

Esta ecuación matricial conduce al sistema $\begin{cases} -x_1+x_2=0 \\ x_1-x_2=0 \end{cases}$ de donde $x_1 = x_2$. Suponiendo $x_1 = c$, obtenemos $x_2 = c$; por lo tanto, el segundo vector propio tiene la forma $\mathbf{x}^{(2)} = c\begin{bmatrix}1\\1\end{bmatrix}$. ▲

Al determinar los valores propios de las matrices y sus vectores propios se resuelve uno de dos problemas: 1) determinación de todos los valores propios y de los vectores propios de las matrices los cuales les pertenecen o bien 2) determinación de uno o de algunos valores propios y de los vectores propios que les pertenecen.

El primer problema consiste en desarrollar el determinante característico en polinomio de n-ésimo grado (o sea, en determinar los coeficientes $p_1, p_2, \ldots, p_n$), en calcular luego los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$ y, por último, en determinar las coordenadas del vector propio $\mathbf{x}^{\mathsf{T}} = (x_1, x\ , \ldots, x_n)$.

El segundo problema consiste en determinar los valores propios de λ (uno o varios) con ayuda del método iterativo sin desarrollar preliminarmente el determinante característico.

Los métodos del primer problema son **exactos**, o sea, si se emplean para las matrices cuyos elementos se dan exactamente (por los números racionales) y si se realizan exactamente los cálculos (por las reglas de operaciones con las fracciones ordinarias), como resultado se obtendrá el valor exacto de los coeficientes del polinomio característico y las coordenadas de los vectores propios resultarán expresadas por las fórmulas exactas con ayuda de los valores propios.

De ordinario se logra determinar los vectores propios de la matriz, utilizando los resultados intermedios de cálculos realizados para determinar los coeficientes del polinomio característico. Desde luego, para determinar el vector propio que pertenece a uno u otro valor, este valor propio debe ya estar calculado.

Los métodos de resolución del segundo problema son **iterativos**, aquí los valores propios se obtienen como límites de ciertas sucesiones numéricas, al igual que las coordenadas de los vectores propios que les pertenecen. Puesto que estos métodos no requieren calcular los coeficientes del polinomio característico, son menos fatigosos. A continuación se consideran algunos métodos de desarrollo del determinante característico y los métodos iterativos de determinación de los valores propios de la matriz.

§ 6.2. Método de desarrollo inmediato

Tomando como ejemplo la matriz de tercer orden, consideremos cómo se determinan los coeficientes del polinomio característico por el desarrollo inmediato del determinante característico. Sea

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad \det(A-\lambda E)=\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{33} & a_{33}-\lambda \end{vmatrix}$$

Calculemos el determinante por la regla de los triángulos:

$$\det (A-\lambda E)=\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix}=$$

$$=(a_{11}-\lambda)(a_{22}-\lambda)(a_{33}-\lambda)+a_{12}a_{23}a_{31}+$$
$$+a_{13}a_{32}a_{21}-a_{13}a_{31}(a_{22}-\lambda)-a_{12}a_{21}(a_{33}-\lambda)-$$
$$-a_{23}a_{32}(a_{11}-\lambda)=-\lambda^3+\lambda^2(a_{11}+a_{22}+a_{33})-$$
$$-\lambda[(a_{11}a_{22}-a_{12}a_{21})+(a_{22}a_{33}-a_{23}a_{32})+(a_{11}a_{33}-a_{13}a_{31})]+$$
$$+(a_{11}a_{22}a_{33}+a_{12}a_{23}a_{31}+a_{13}a_{32}a_{21}-a_{13}a_{22}a_{31}-a_{12}a_{21}a_{33}-a_{23}a_{32}a_{11})=$$
$$=(-1)^3\cdot\left[\lambda^3-\lambda^2(a_{11}+a_{22}+a_{33})+\lambda\left(\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{vmatrix}+\right.\right.$$
$$\left.\left.+\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{vmatrix}+\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33}\end{vmatrix}\right)-\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{vmatrix}\right),$$

o bien

$$\det (A-\lambda E)=(-1)^3(\lambda^3-p_1\lambda^2+p_2\lambda-p_3)=0.$$

Aquí el coeficiente p_1 es la suma de los elementos diagonales de la matriz A; se llama *traza* de la matriz y se designa $\operatorname{Sp} A$:

$$p_1=\operatorname{Sp} A=a_{11}+a_{22}+a_{33};$$

el coeficiente p_2 es la suma de todos los menores diagonales de segundo orden de la matriz A:

$$p_2=\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22}\end{vmatrix}+\begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33}\end{vmatrix}+\begin{vmatrix} a_{11} & a_{13} \\ a_{31} & a_{33}\end{vmatrix}$$

(recuérdese que se llaman menores diagonales de segundo, tercero, ..., n-ésimo orden los menores cuyos elementos de las diagonales principales son elementos de la diagonal principal del determinante $\det A$); el coeficiente

$$p_3=\det A=\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33}\end{vmatrix}.$$

En general, si se necesita desarrollar el determinante $\det (A-\lambda E)$ en polinomio de n-ésimo grado:

$$D(\lambda)=(-1)^n[\lambda^n-p_1\lambda^{n-1}+p_2\lambda^{n-2}-\ldots+(-1)^n p_n],$$

los coeficientes $p_1, p_2, \ldots, p_n$ se calculan por las fórmulas siguientes:

$p_1=\sum\limits_{i=1}^{n} a_{ii}=\operatorname{Sp} A$ *es la suma de todos los elementos diagonales de la matriz* A;

$p_2 = \sum\limits_{\alpha<\beta} \begin{vmatrix} a_{\alpha\alpha} & a_{\alpha\beta} \\ a_{\beta\alpha} & a_{\beta\beta} \end{vmatrix}$ *es la suma de todos los menores diagonales de segundo orden de la matriz* A;

$p_3 = \sum\limits_{\alpha<\beta<\gamma} \begin{vmatrix} a_{\alpha\alpha} & a_{\alpha\beta} & a_{\alpha\gamma} \\ a_{\beta\alpha} & a_{\beta\beta} & a_{\beta\gamma} \\ a_{\gamma\alpha} & a_{\gamma\beta} & a_{\gamma\gamma} \end{vmatrix}$ *es la suma de todos los menores diagonales de tercer orden de la matriz* A;

. . . .

$p_n = \det A$ *es el determinante de la matriz* A.

La cantidad de menores diagonales de k-ésimo orden de la matriz es igual a

$$C_n^k = \frac{n(n-1)(n-2)\ldots(n-k+1)}{k!} \quad (k = 1, 2, \ldots, n)$$

El método de desarrollo inmediato cuesta mucho trabajo y se emplea al determinar los polinomios característicos para las matrices de pequeños órdenes.

Ejemplo 1. Haciendo uso del método de desarrollo inmediato, hallar el polinomio característico de la matriz

$$A = \begin{bmatrix} -4 & -3 & 1 & 1 \\ 2 & 0 & 4 & -1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix}.$$

1) Determinamos

$$p_1 = \operatorname{Sp} A = a_{11} + a_{22} + a_{33} + a_{44} = -4 + 0 + 2 - 1 = -3.$$

2) Tenemos $p_2 = \sum\limits_{\alpha<\beta} \begin{vmatrix} a_{\alpha\alpha} & a_{\alpha\beta} \\ a_{\beta\alpha} & a_{\beta\beta} \end{vmatrix}$. La cantidad de los menores diagonales de segundo orden en la matriz de cuarto orden es igual a $C_4^2 = \frac{43}{1\cdot 2} = 6$. Escribiendo todos estos menores y sumándolos, obtenemos

$$p_2 = \underbrace{\begin{vmatrix} -4 & -3 \\ 2 & 0 \end{vmatrix}}_{\alpha=1;\ \beta=2} + \underbrace{\begin{vmatrix} -4 & 1 \\ 1 & 2 \end{vmatrix}}_{\alpha=1;\ \beta=2} + \underbrace{\begin{vmatrix} -4 & 1 \\ 1 & -1 \end{vmatrix}}_{\alpha=1;\ \beta=4} + \underbrace{\begin{vmatrix} 0 & 4 \\ 1 & 2 \end{vmatrix}}_{\alpha=2;\ \beta=3} +$$

$$+ \underbrace{\begin{vmatrix} 0 & -1 \\ 1 & -1 \end{vmatrix}}_{\alpha=2;\ \beta=4} + \underbrace{\begin{vmatrix} 2 & -2 \\ -1 & -1 \end{vmatrix}}_{\alpha=3;\ \beta=4} = -7.$$

3) Tenemos $p_3 = \sum\limits_{\alpha<\beta<\gamma} \begin{vmatrix} a_{\alpha\alpha} & a_{\alpha\beta} & a_{\alpha\gamma} \\ a_{\beta\alpha} & a_{\beta\beta} & a_{\beta\gamma} \\ a_{\gamma\alpha} & a_{\gamma\beta} & a_{\gamma\gamma} \end{vmatrix}$. La cantidad de menores diagonales de tercer orden en la matriz de cuarto orden es igual a $C_4^3 = \frac{4\cdot3\cdot2}{1\cdot2\cdot3} = 4$. Por consiguiente,

$$p_3 = \underbrace{\begin{vmatrix} -4 & -3 & 1 \\ 2 & 0 & 4 \\ 1 & 1 & 2 \end{vmatrix}}_{\alpha=1;\ \beta=2;\ \gamma=3} + \underbrace{\begin{vmatrix} -4 & -3 & 1 \\ 2 & 0 & -1 \\ 1 & 1 & -1 \end{vmatrix}}_{\alpha=1;\ \beta=2;\ \gamma=4} + \underbrace{\begin{vmatrix} -4 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & -1 & -1 \end{vmatrix}}_{\alpha=1;\ \beta=3;\ \gamma=4} +$$

$$+ \underbrace{\begin{vmatrix} 0 & 4 & -1 \\ 1 & 2 & -2 \\ 1 & -1 & -1 \end{vmatrix}}_{\alpha=2;\ \beta=3;\ \gamma=4} = 24.$$

4) por último, hallamos

$$p_4 = \det A = \begin{vmatrix} -4 & -3 & 1 & 1 \\ 2 & 0 & 4 & -1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{vmatrix} = -15.$$

5) Ahora bien, finalmente obtenemos

$$D(\lambda) = \lambda^4 - p_1\lambda^3 + p_2\lambda^2 - p_3\lambda + p_4 =$$
$$= \lambda^4 + 3\lambda^3 - 7\lambda^2 - 24\lambda - 15. \ \blacktriangle$$

Ejemplo 2. Desarrollar el polinomio característico de la matriz

$$A = \begin{bmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{bmatrix}$$

y determinar cualquier valor propio y el vector propio correspondiente.

△ 1) Tenemos

$$p_1 = \operatorname{Sp} A = 3 - 1 - 2 = 0;$$

$$p_2 = \sum_{\alpha<\beta} \begin{vmatrix} a_{\alpha\alpha} & a_{\alpha\beta} \\ a_{\beta\alpha} & a_{\beta\beta} \end{vmatrix} = \underbrace{\begin{vmatrix} 3 & 1 \\ -4 & -1 \end{vmatrix}}_{\alpha=1;\ \beta=2} + \underbrace{\begin{vmatrix} 3 & 0 \\ 4 & -2 \end{vmatrix}}_{\alpha=1;\ \beta=3} + \underbrace{\begin{vmatrix} -1 & 0 \\ -8 & -2 \end{vmatrix}}_{\alpha=2;\ \beta=3} =$$
$$= 1 - 6 + 2 = -3;$$

$$p_3 = \det A = \begin{vmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{vmatrix} = -2.$$

Por lo tanto, $D(\lambda) = (-1)^3 (\lambda^3 - 3\lambda + 2)$; $\lambda^3 - 3\lambda + 2 = 0$; uno de los valores propios de la matriz A es $\lambda = 1$.

2) Hallamos el vector propio $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$ correspondiente a $\lambda = 1$.

La ecuación matricial

$$(A - \lambda E)\,\mathbf{x} = 0 \quad \text{o bien} \quad \begin{bmatrix} 3-1 & 1 & 0 \\ -4 & -1-1 & 0 \\ 4 & -8 & -2-1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

es equivalente al sistema

$$\begin{cases} 2x_1 + x_2 = 0, \\ -4x_1 - 2x_2 = 0, \\ 4x_1 - 8x_2 - 3x_3 = 0. \end{cases}$$

Aquí las ecuaciones primera y segunda son proporcionales, por eso, suprimiendo la segunda ecuación, obtenemos el sistema

$$\begin{cases} 2x_1 + x_2 = 0, \\ 4x_1 - 8x_2 - 3x_3 = 0. \end{cases}$$

El menor $d = \begin{vmatrix} 2 & 1 \\ 4 & -8 \end{vmatrix} = -20$ es básico; x_1 y x_2 son las incógnitas básicas; x_3 es la incógnita independiente. Entonces con ayuda de las fórmulas de Cramer hallamos

$$x_1 = \frac{d_1}{d} = \frac{\begin{vmatrix} 0 & 1 \\ 3x_3 & -8 \end{vmatrix}}{-20} = \frac{3x_3}{20},$$

$$x_2 = \frac{d_2}{d} = \frac{\begin{vmatrix} 2 & 0 \\ 4 & 3x_3 \end{vmatrix}}{-20} = \frac{6x_3}{-20}.$$

Suponiendo $x_3 = 20$, tenemos $x_1 = 3$; $x_2 = -6$. Así pues $\mathbf{x}^{(1)} = c \begin{bmatrix} 3 \\ -6 \\ 20 \end{bmatrix}$ es el vector propio buscado. ▲

§ 6.3. Método de Krylov para desarrollar el determinante característico

El método de Krylov se funda en propiedad de la matriz cuadrada de anular su polinomio característico.

Según la **identidad de Hamilton—Cayley** toda la *matriz cuadrada es raíz de su polinomio característico y, por lo tanto, lo anula.*

Sea

$$D(\lambda) = \det(A - \lambda E) =$$
$$= (-1)^n (\lambda^n + p_1\lambda^{n-1} + p_2\lambda^{n-2} + \ldots + p_n) \qquad (1)$$

el polinomio característico de la matriz A. Reemplazando en la igualdad (1) la magnitud λ por $A = [a_{ij}]$ (donde $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$), obtenemos

$$A^n + p_1A^{n-1} + p_2A^{n-2} + \ldots + p_nE = 0. \qquad (2)$$

Tomemos un vector no nulo arbitrario

$$\mathbf{y}^{(0)} = \begin{bmatrix} y_1^{(0)} \\ y_2^{(0)} \\ \vdots \\ y_n^{(0)} \end{bmatrix} \qquad (3)$$

y multipliquemos ambos miembros de la igualdad (2) a la derecha por $\mathbf{y}^{(0)}$:

$$A^n\mathbf{y}^{(0)} + p_1A^{n-1}\mathbf{y}^{(0)} + p_2A^{n-2}\mathbf{y}^{(0)} + \ldots + p_n\mathbf{y}^{(0)} = 0. \qquad (4)$$

Pongamos ahora

$$A\,\mathbf{y}^{(k-1)} = \mathbf{y}^{(k)} \quad (k = 1, 2, \ldots, n), \qquad (5)$$

o sea,

$$\begin{aligned} \mathbf{y}^{(1)} &= A\,\mathbf{y}^{(0)}, \\ \mathbf{y}^{(2)} &= A\,\mathbf{y}^{(1)} = A^2\mathbf{y}^{(0)}, \\ &\ldots\ldots\ldots\ldots \\ \mathbf{y}^{(n)} &= A\,\mathbf{y}^{n-1} = A^n\mathbf{y}^{(0)}. \end{aligned}$$

Entonces la igualdad (4) se toma la forma siguiente:

$$\mathbf{y}^{(n)} + p_1\mathbf{y}^{(n-1)} + p_2\mathbf{y}^{(n-2)} + \ldots + p_n\mathbf{y}^{(0)} = 0 \qquad (6)$$

o bien

$$p_1\mathbf{y}^{(n-1)} + p_2\mathbf{y}^{(n-2)} + \ldots + p_n\mathbf{y}^{(0)} = -\mathbf{y}^{(n)},$$

o bien

$$p_1\begin{bmatrix} y_1^{(n-1)} \\ y_2^{(n-1)} \\ \vdots \\ y_n^{(n-1)} \end{bmatrix} + p_2\begin{bmatrix} y_1^{(n-2)} \\ y_2^{(n-2)} \\ \vdots \\ y_n^{(n-2)} \end{bmatrix} + \ldots + p_n\begin{bmatrix} y_1^{(0)} \\ y_2^{(0)} \\ \vdots \\ y_n^{(0)} \end{bmatrix} = -\begin{bmatrix} y_1^{(n)} \\ y_2^{(n)} \\ \vdots \\ y_n^{(n)} \end{bmatrix},$$

o sea,

$$\left\{ \begin{array}{l} p_1y_1^{(n-1)} + p_2y_1^{(n-2)} + \ldots + p_ny_1^{(0)} = -y_1^{(n)}, \\ p_1y_2^{(n-1)} + p_2y_2^{(n-2)} + \ldots + p_ny_2^{(0)} = -y_2^{(n)}, \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ p_1y_n^{(n-1)} + p_2y_n^{(n-2)} + \ldots + p_ny_n^{(0)} = -y_n^{(n)} \end{array} \right. \qquad (7)$$

o bien, por último, en la forma matricial

$$\begin{bmatrix} y_1^{(n-1)} & y_1^{(n-2)} & \dots & y_1^{(0)} \\ y_2^{(n-1)} & y_2^{(n-2)} & \dots & y_2^{(0)} \\ \dots & \dots & \dots & \dots \\ y_n^{(n-1)} & y_n^{(n-2)} & \dots & y_n^{(0)} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = - \begin{bmatrix} y_1^{(n)} \\ y_2^{(n)} \\ \vdots \\ y_n^{(n)} \end{bmatrix}. \tag{8}$$

Los vectores $\mathbf{y}_i^{(1)}$, $\mathbf{y}_i^{(2)}$, ..., $\mathbf{y}_i^{(n)}$ se calculan por las fórmulas

$$\begin{gathered} \mathbf{y}_i^{(1)} = \sum_{j=1}^{n} a_{ij}\mathbf{y}_j^{(0)} = A\mathbf{y}^{(0)}, \\ \mathbf{y}_i^{(2)} = \sum_{j=1}^{n} a_{ij}\mathbf{y}_j^{(1)} = A\mathbf{y}^{(1)}, \\ \dots\dots\dots\dots\dots\dots\dots \\ \mathbf{y}_i^{(n)} = \sum_{j=1}^{n} a_{ij}\mathbf{y}_j^{(n-1)} = A\mathbf{y}^{(n-1)} \quad (i = 1, 2, \dots, n), \end{gathered} \tag{9}$$

con la particularidad de que las coordenadas del vector inicial (3) se toman arbitrariamente. Si el sistema lineal (7) tiene la única solución, sus raíces p_1, p_2, ..., p_n son coeficientes del polinomio característico (1). Esta solución puede ser hallada por el método de Gauss.

Ejemplo 1. Haciendo uso del método de Krylov, desarrollar el determinante característico de la matriz

$$A = \begin{bmatrix} -4 & -3 & 1 & 1 \\ 2 & 0 & 4 & -1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix}.$$

△ 1) Elegimos el vector inicial $\mathbf{y}^{(0)} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$.

2) Con ayuda de las fórmulas (9) determinamos las coordenadas de los vectores $\mathbf{y}^{(k)} = A\,\mathbf{y}^{(k-1)}$ $(k = 1, 2, 3, 4)$:

$$\mathbf{y}^{(1)} = A\mathbf{y}^{(0)} = \begin{bmatrix} -4 & -3 & 1 & 1 \\ 2 & 0 & 4 & -1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -4 \\ 2 \\ 1 \\ 1 \end{bmatrix},$$

$$\mathbf{y}^{(2)} = A\mathbf{y}^{(1)} = \begin{bmatrix} 4 & -3 & 1 & 1 \\ 2 & 0 & 4 & -1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} -4 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 12 \\ -5 \\ -2 \\ -4 \end{bmatrix},$$

$$\mathbf{y}^{(3)} = A\mathbf{y}^{(2)} = \begin{bmatrix} -4 & -3 & 1 & 1 \\ 2 & 0 & 4 & -1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 12 \\ -5 \\ -2 \\ -4 \end{bmatrix} = \begin{bmatrix} -39 \\ 20 \\ 11 \\ 13 \end{bmatrix},$$

$$\mathbf{y}^{(4)} = A\mathbf{y}^{(3)} = \begin{bmatrix} -4 & -3 & 1 & 1 \\ 2 & 0 & 4 & -1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} -39 \\ 20 \\ 11 \\ 13 \end{bmatrix} = \begin{bmatrix} 120 \\ -47 \\ -23 \\ -43 \end{bmatrix}.$$

3) Planteamos la ecuación matricial

$$\begin{bmatrix} -39 & 12 & -4 & 1 \\ 20 & -5 & 2 & 0 \\ 11 & -2 & 1 & 0 \\ 13 & -4 & 1 & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = -\begin{bmatrix} 120 \\ -47 \\ -23 \\ -43 \end{bmatrix}.$$

Tabla 6.1

p_1	p_2	p_3	p_4	Términos independientes	Σ_1	Σ_2
$\boxed{-39}$	12	−4	1	−120	−150	
20	−5	2	0	47	64	
11	−2	1	0	23	33	
13	−4	1	0	43	53	
1	−4/13	4/39	−1/39	40/13	50/13	
	$\boxed{15/13}$	−2/39	20/39	−189/13	−168/13	−168/13
	18/13	−5/39	11/39	−141/13	−121/13	−121/13
	0	−1/3	−1/3	3	3	3
	1	−2/45	4/9	−53/5	−56/5	−56/5
		$\boxed{-1/15}$	−1/3	33/5	31/5	31/5
		−1/3	1/3	3	3	3
		1	5	−99	−93	−93
			$\boxed{2}$	−30	−28	−28
			1	−15	−14	−14
			1	$p_4 = -15$	$\bar{p}_4 = -14$	−14
		1		$p_3 = -24$	$\bar{p}_3 = -23$	−23
	1			$p_2 = -7$	$\bar{p}_2 = -6$	−6
1				$p_1 = 3$	$\bar{p}_1 = 4$	4

Escribimos el sistema que tiene la forma (7):

$$\begin{cases} -39p_1+12p_2-4p_3+p_4=-120, \\ 20p_1-5p_2+2p_3 \qquad =47, \\ 11p_1-2p_2+p_3 \qquad =23, \\ 13p_1-4p_2+p_3 \qquad =43. \end{cases}$$

Resolvemos este sistema según el esquema de Gauss (véase la tabla 6.1).

Ahora bien,

$$D(\lambda) = \det(A-\lambda E) = \lambda^4 + p_1\lambda^3 + p_2\lambda^2 + p_3\lambda + p_4 = \\ = \lambda^4 + 3\lambda^3 - 7\lambda^2 - 24\lambda - 15. \blacktriangle$$

En cambio, si el sistema lineal obtenido (7) no tiene la solución única, es necesario cambiar el vector inicial.

Ejemplo 2. Haciendo uso del método de Krylov, desarrollar el determinante característico de la matriz

$$A=\begin{bmatrix} 1 & -1 & -1 & 2 \\ 2 & 3 & 0 & -4 \\ 1 & 1 & -2 & -2 \\ 1 & 1 & 0 & -1 \end{bmatrix}.$$

△ 1) Tomemos en calidad de vector inicial $\mathbf{y}^{(0)}=\begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix}$.

Entonces obtenemos

$$\mathbf{y}^{(1)}=A\mathbf{y}^{(0)}=\begin{bmatrix} 1 \\ 2 \\ 1 \\ 1 \end{bmatrix}; \quad \mathbf{y}^{(2)}=A\mathbf{y}^{(1)}=\begin{bmatrix} 0 \\ 4 \\ -1 \\ 2 \end{bmatrix};$$

$$\mathbf{y}^{(3)}=A\mathbf{y}^{(2)}=\begin{bmatrix} 1 \\ 4 \\ 2 \\ 2 \end{bmatrix}; \quad \mathbf{y}^{(4)}=A\mathbf{y}^{(3)}=\begin{bmatrix} -1 \\ 6 \\ -3 \\ 3 \end{bmatrix}.$$

2) Planteamos la ecuación matricial

$$\begin{bmatrix} 1 & 0 & 1 & 1 \\ 4 & 4 & 2 & 0 \\ 2 & -1 & 1 & 0 \\ 2 & 2 & 1 & 0 \end{bmatrix}\begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = -\begin{bmatrix} 1 \\ 6 \\ -3 \\ 3 \end{bmatrix},$$

de donde obtenemos el sistema de ecuaciones

$$\begin{cases} p_1 + p_3 + p_4 = 1, \\ 4p_1 + 4p_2 + 2p_3 = -6, \\ 2p_1 - p_2 + p_3 = 3, \\ 2p_1 + 2p_2 + p_3 = -3. \end{cases}$$

Resolvemos este sistema con ayuda del esquema de única división

Tabla 6.2

p_1	p_2	p_3	p_4	Términos independientes	Σ_1	Σ_2
$\boxed{1}$	0	1	1	1	4	
4	4	2	0	−6	4	
2	−1	1	0	3	5	
2	2	1	0	−3	2	
1	0	1	1	1	4	
	$\boxed{4}$	−2	−4	−10	−12	−12
	−1	−1	−2	1	−3	−3
	2	−1	−2	−5	−6	−6
	1	−0,5	−1	−3	−3,5	−3,5
		$\boxed{-1,5}$	−3	−2	−6,5	−6,5
		0	0	1	1	1
		1	2	1,333	4,333	4,333
			$\boxed{0}$	1	1	1

Puesto que el elemento guía es igual a cero, no es posible continuar los cálculos con ayuda de este esquema.

3) Para obtener la única solución cambiamos el vector inicial; suponiendo $\mathbf{y}^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$, hallamos

$$\mathbf{y}^{(1)} = A\mathbf{y}^{(0)} = \begin{bmatrix} 2 \\ -4 \\ -2 \\ -1 \end{bmatrix}; \quad \mathbf{y}^{(2)} = A\mathbf{y}^{(1)} = \begin{bmatrix} 6 \\ -4 \\ 4 \\ -1 \end{bmatrix};$$

$$\mathbf{y}^{(3)} = A\mathbf{y}^{(2)} = \begin{bmatrix} 4 \\ 4 \\ -4 \\ 3 \end{bmatrix}; \quad \mathbf{y}^{(4)} = A\mathbf{y}^{(3)} = \begin{bmatrix} 10 \\ 8 \\ 10 \\ 5 \end{bmatrix}.$$

Tabla 6.3

p_1	p_2	p_3	p_4	Términos independientes	Σ_1	Σ_2
$\boxed{2}$	3	1	0	−5	1	
1	−1	−1	0	2	−3	
−2	2	−1	0	−5	−6	
3	−1	−1	1	−5	−3	
1	1,5	0,5	0	−2,5	0,5	
	$\boxed{-2,5}$	−1,5	0	0,5	−3,5	−3,5
	5	0	0	−10	−5	−5
	−5,5	−2,5	1	2,5	−4,5	−4,5
	1	0,6	0	−0,2	1,4	1,4
		$\boxed{-3}$	0	−9	−12	−12
		0,8	1	1,4	3,2	3,2
		1	0	3	4	4
			$\boxed{1}$	−1	0	0
			1	$p_4=-1$	$\bar{p}_4=0$	0
		1		$p_3=3$	$\bar{p}_3=4$	4
	1			$p_2=-2$	$\bar{p}_2=-1$	−1
1				$p_1=-1$	$\bar{p}_1=0$	0

La ecuación matricial

$$\begin{bmatrix} 4 & 6 & 2 & 0 \\ 4 & -4 & -4 & 0 \\ -4 & 4 & -2 & 0 \\ 3 & -1 & -1 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{bmatrix} = - \begin{bmatrix} 10 \\ 8 \\ 10 \\ 5 \end{bmatrix}$$

conduce al sistema

$$\left\{ \begin{aligned} 4p_1+6p_2+2p_3 &= -10, \\ 4p_1-4p_2-4p_3 &= -8, \\ -4p_1+4p_2-2p_3 &= -10, \\ 3p_1-p_2-p_3+p_4 &= -5, \end{aligned} \right. \quad \text{o bien} \quad \left\{ \begin{aligned} 2p_1+3p_2+p_3 &= -5, \\ p_1-p_2-p_3 &= -2, \\ -2p_1+2p_2-p_3 &= -5, \\ 3p_1-p_2-p_3+p_4 &= -5, \end{aligned} \right.$$

el cual se resuelve con ayuda del esquema de única división (véase la tabla 6.3).

Por lo tanto,

$$D(\lambda) = \lambda^4 - \lambda^3 - 2\lambda^2 + 3\lambda - 1. \ \blacktriangle$$

§ 6.4. Cálculo de los vectores propios con ayuda del método de Krylov

Si se conocen los coeficientes $p_1, p_2, \ldots, p_n$ y las raíces λ_1, $\lambda_2, \ldots, \lambda_4$ del polinomio característico, el método de Krylov ofrece la posibilidad de determinar los vectores propios respectivos valiéndose de la fórmula siguiente:

$$\mathbf{x}^{(i)} = \mathbf{y}^{(n-1)} + q_{1i}\mathbf{y}^{(i-2)} + \ldots + q_{-1,i}\,\mathbf{y}^{(0)} \quad (i = 1, 2, \ldots, n). \tag{1}$$

Aquí $\mathbf{y}^{(n-1)}$, $\mathbf{y}^{(n-2)}, \ldots, \mathbf{y}^{(0)}$ son los vectores utilizados al hallar los coeficientes $p_1, p_2, \ldots, p_n$ con ayuda del método de Krylov y los coeficientes q_{ji} $(j = 1, 2, \ldots, n-1;\ i = 1, 2, \ldots, n)$ se determinan por el esquema de Horner:

$$q_{0i} = 1; \quad q_{ji} = \lambda_i q_{j-1,i} + p_j. \tag{2}$$

Ejemplo. Haciendo uso del método de Krylov, hallar los vectores propios de la matriz

$$A = \begin{bmatrix} 1 & -1 & -1 & 2 \\ 2 & 3 & 0 & -4 \\ 1 & 1 & -2 & -2 \\ 1 & 1 & 0 & -1 \end{bmatrix}.$$

△ El polinomio característico de la matriz A está conocido:

$$\det(A - \lambda E) = \lambda^4 - \lambda^3 - 2\lambda^2 + 3\lambda - 1$$

(véase el ejemplo 2 en el § 6.3), y los valores propios son tales: $\lambda_1 = \lambda_2 = 1$; $\lambda_3 = 0{,}618$; $\lambda_4 = -1{,}618$. Para determinar los vectores propios utilicemos la fórmula (1)

$$\mathbf{x}^{(i)} = \mathbf{y}^{(3)} + q_{1i}\mathbf{y}^{(2)} + q_{2i}\mathbf{y}^{(1)} + q_{3i}\mathbf{y}^{(0)}.$$

Aquí $q_{0i} = 1$ y los coeficientes q_{ji} $(j = 1, 2, 3;\ i = 1, 2, 3, 4)$ se calculan con ayuda del esquema de Horner (véase la tabla 6.4).

Tabla 6.4

λ_i	$p_0 = 1$	$p_1 = 1$	$p_2 = -2$	$p_3 = 3$
$\lambda_1 = 1$	$q_{01} = 1$	$q_{11} = 0$	$q_{21} = -2$	$q_{31} = 1$
$\lambda_2 = 1$	$q_{02} = 1$	$q_{12} = 0$	$q_{22} = -2$	$q_{32} = 1$
$\lambda_3 = 0{,}618$	$q_{03} = 1$	$q_{13} = -0{,}382$	$q_{23} = -2{,}236$	$q_{33} = 1{,}618$
$\lambda_4 = -1{,}618$	$q_{04} = 1$	$q_{14} = -2{,}618$	$q_{24} = 2{,}236$	$q_{34} = -0{,}618$

Utilizando las expresiones para los vectores $\mathbf{y}^{(0)}$, $\mathbf{y}^{(1)}$, $\mathbf{y}^{(2)}$, $\mathbf{y}^{(3)}$ hallados en el ejemplo 2 del § 6.3, obtenemos

$$\mathbf{x}^{(1)}=\mathbf{x}^{(2)}=\begin{bmatrix}4\\4\\-4\\3\end{bmatrix}+0\begin{bmatrix}6\\-4\\4\\-1\end{bmatrix}-2\begin{bmatrix}2\\-4\\-2\\-1\end{bmatrix}+1\begin{bmatrix}0\\0\\0\\1\end{bmatrix}=$$

$$=\begin{bmatrix}0\\12\\0\\6\end{bmatrix}=\begin{bmatrix}0\\1\\0\\0,5\end{bmatrix};$$

$$\mathbf{x}^{(3)}=\begin{bmatrix}4\\4\\-4\\3\end{bmatrix}-0,382\begin{bmatrix}6\\-4\\4\\-1\end{bmatrix}-2,236\begin{bmatrix}2\\-4\\-2\\-1\end{bmatrix}+$$

$$+1,618\begin{bmatrix}0\\0\\0\\1\end{bmatrix}=\begin{bmatrix}-2,764\\14,472\\-1,056\\7,236\end{bmatrix}=\begin{bmatrix}0,19\\1\\-0,07\\0,50\end{bmatrix};$$

$$\mathbf{x}^{(4)}=\begin{bmatrix}4\\4\\-4\\3\end{bmatrix}-2,618\begin{bmatrix}6\\-4\\4\\-1\end{bmatrix}+2,236\begin{bmatrix}2\\-4\\-2\\-1\end{bmatrix}-$$

$$-0,618\begin{bmatrix}0\\0\\0\\1\end{bmatrix}=\begin{bmatrix}-7,236\\5,528\\-18,944\\2,764\end{bmatrix}=\begin{bmatrix}-0,38\\0,29\\1\\0,15\end{bmatrix}. \blacktriangle$$

§ 6.5. Método de Le Verrier—Faddéev

Este método fue propuesto por Le Verrier y luego simplificado por el matemático soviético Faddéev. El **método de Le Verrier** se funda en las fórmulas de Newton para las sumas de los grados de las raíces de una ecuación algebraica y consiste en lo siguiente. Sea

$$\det(A-\lambda E)=\lambda^n+p_1\lambda^{n-1}+\ldots+p_n \tag{1}$$

el polinomio característico de la matriz $A=[a_{ij}]$ (donde $i=1, 2, \ldots, n$; $j=1, 2, \ldots, n$) y $\lambda_1, \lambda_2, \ldots, \lambda_n$ es el conjunto completo de las raíces del polinomio (1). Consideremos las sumas

$$S_k=\lambda_1^k+\lambda_2^k+\ldots+\lambda_n^k \quad (k=1, 2, \ldots, n),$$

o sea,

$$S_1=\lambda_1+\lambda_2+\dots+\lambda_n=\operatorname{Sp} A,$$
$$S_2=\lambda_1^2+\lambda_2^2+\dots+\lambda_n^2=\operatorname{Sp} A^2,$$
$$\dots\dots\dots\dots\dots\dots$$
$$S_n=\lambda_1^n+\lambda_2^n+\dots+\lambda_n^n=\operatorname{Sp} A^n$$

(cada suma S_k es la traza de la matriz A^k). Entonces para $k\leqslant n$ son válidas las *fórmulas de Newton*

$$S_k+p_1S_{k-1}+\dots+p_{k-1}S_1=-kp_k,$$

de donde obtenemos

$$\begin{aligned}&\text{para } k=1 \quad p_1=-S_1,\\ &\text{para } k=2 \quad p_2=-\frac{1}{2}(S_2+p_1S_1),\\ &\dots\dots\dots\dots\dots\dots\dots\dots\\ &\text{para } k=n \quad p_n=-\frac{1}{n}(S_n+p_1S_{n-1}+p_2S_{n-2}+\dots+p_{n-1}S_1).\end{aligned} \tag{2}$$

Por consiguiente, los coeficientes del polinomio característico $p_1, p_2, \dots, p_n$ pueden ser determinados fácilmente si están conocidas las sumas $S_1, S_2, \dots, S_n$.

Ahora bien, el esquema de desarrollo del determinante característico por el método de Le Verrier consiste en lo siguiente:

1) se calculan los grados de $A^k=A^{k-1}A$ $(k=1, 2, \dots, n)$;

2) se determinan S_k, o sea, las sumas de los elementos de las diagonales principales de las matrices A^k;

3) Haciendo uso de las fórmulas (2), se hallan los coeficientes p_i $(i=1, 2, \dots, n)$.

El método modificado de Le Verrier, propuesto, por Faddéev, consiste en calcular la sucesión de las matrices $A_1, A_2, \dots, A_n$ según el esquema siguiente:

$$\begin{aligned}&A_1=A; \quad \operatorname{Sp} A_1=q_1; \quad B_1=A_1-q_1E;\\ &A_1=AB_1; \quad \frac{\operatorname{Sp} A_2}{2}=q_2; \quad B_2=A_2-q_2E;\\ &\dots\dots\dots\dots\dots\dots\dots\dots\\ &A_{n-1}=AB_{n-2}; \quad \frac{\operatorname{Sp} A_{n-1}}{n-1}=q_{n-1}; \quad B_{n-1}=A_{n-1}-q_{n-1}E;\end{aligned} \tag{3}$$

$$A_n=AB_{n-1}; \quad \frac{\operatorname{Sp} A_n}{n}=q_n; \quad B_n=A_n-q_nE \ (B_n \text{ es la matriz nula});$$

$$q_1=-p_1;\ q_2=-p_2;\ \dots;\ q_{n-1}=-p_{n-1},\ q_n=-p_n.$$

Ejemplo 1. Utilizando el método de Le Verrier—Faddéev, desarrollar el determinante característico de la matriz

$$A=\begin{bmatrix}1 & -1 & -1 & 2\\ 2 & 3 & 0 & -4\\ 1 & 1 & -2 & -2\\ 1 & 1 & 0 & -1\end{bmatrix}.$$

△ Sucesivamente hallamos

$$1)\ A_1=A=\begin{bmatrix}1 & -1 & -1 & 2\\ 2 & 3 & 0 & -4\\ 1 & 1 & -2 & -2\\ 1 & 1 & 0 & -1\end{bmatrix};\quad q_1=\operatorname{Sp}A_1=1+3-2-1=1;$$

$$B_1=A_1-q_1E=\begin{bmatrix}0 & -1 & -1 & 2\\ 2 & 2 & 0 & -4\\ 1 & 1 & -3 & -2\\ 1 & 1 & 0 & -2\end{bmatrix};$$

$$2)\ A_2=AB_1=\begin{bmatrix}1 & -1 & 1 & 2\\ 2 & 3 & 0 & -4\\ 1 & 1 & -2 & -2\\ 1 & 1 & 0 & -1\end{bmatrix}\begin{bmatrix}0 & -1 & -1 & 2\\ 2 & 2 & 0 & -4\\ 1 & 1 & 3 & -2\\ 1 & 1 & 0 & -2\end{bmatrix}=$$

$$=\begin{bmatrix}-1 & -2 & 2 & 4\\ 2 & 0 & -2 & 0\\ -2 & -3 & 5 & 6\\ 1 & 0 & -1 & 0\end{bmatrix};\quad q_2=\frac{\operatorname{Sp}A_2}{2}=\frac{-1+0+5+0}{2}=2;$$

$$B_2=A_2-q_2E=\begin{bmatrix}-3 & -2 & 2 & 4\\ 2 & -2 & -2 & 0\\ -2 & -3 & 3 & 6\\ 1 & 0 & -1 & -2\end{bmatrix}.$$

$$3)\ A_3=AB_2=\begin{bmatrix}1 & -1 & -1 & 2\\ 2 & 3 & 0 & -4\\ 1 & 1 & -2 & -2\\ 1 & 1 & 0 & -1\end{bmatrix}\begin{bmatrix}-3 & -2 & 2 & 4\\ 2 & -2 & -2 & 0\\ -2 & -3 & 3 & 6\\ 1 & 0 & -1 & -2\end{bmatrix}=$$

$$=\begin{bmatrix}-1 & 3 & -1 & -6\\ -4 & -10 & 2 & 16\\ 1 & 2 & -4 & -4\\ -2 & -4 & 1 & 6\end{bmatrix};\quad q_3=\frac{\operatorname{Sp}A_3}{3}=\frac{-1-10-4+6}{3}=-3;$$

$$B_3=A_3-q_3E=\begin{bmatrix}2 & 3 & -1 & -6\\ -4 & -7 & 2 & 16\\ 1 & 2 & -1 & -4\\ -2 & -4 & 1 & 9\end{bmatrix};$$

4) $A_4 = AB_3 = \begin{bmatrix} 1 & -1 & -1 & 2 \\ 2 & 3 & 0 & -4 \\ 1 & 1 & -2 & -2 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & -1 & -6 \\ -4 & -7 & 2 & 16 \\ 1 & 2 & -1 & -4 \\ -2 & -4 & 1 & 9 \end{bmatrix} =$

$$= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}; \quad q_4 = \frac{\operatorname{Sp} A_4}{4} = \frac{1+1+1+1}{4} = 1;$$

$$B_4 = A_4 - q_4 E = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = O_4.$$

5) Ahora bien, $p_1 = -q_1 = -1$; $p_2 = -q_2 = -2$; $p_3 = -q_3 = 3$; $p_4 = -q_4 = -1$ y $D(\lambda) = \lambda^4 - \lambda^3 - 2\lambda^2 + 3\lambda - 1$. ▲

La modificación del método de Le Verrier, propuesta por Faddéev, permite determinar la matriz inversa A^{-1}. De las fórmulas (3) tenemos $A_n = AB_{n-1}$, $B_n = A_n - q_nE = O_n$, de donde $A_n = q_nE$,

$$AB_{n-1} = q_nE. \tag{4}$$

Multiplicando a la izquierda la igualdad (4) por A^{-1}, obtenemos $A^{-1}AB_{n-1} = A^{-1}q_nE$, de donde

$$A^{-1} = \frac{B_{n-1}}{q_n} \text{ o bien } A^{-1} = \frac{B_{n-1}}{-p_n}. \tag{5}$$

Ejemplo 2. Calcular la matriz inversa para la matriz A dada en el ejemplo 1.

△ Utilizando la fórmula (5), obtenemos

$$A^{-1} = \frac{B_3}{-p_4} = 1 \begin{bmatrix} 2 & 3 & -1 & -6 \\ -4 & -7 & 2 & 16 \\ 1 & 2 & -1 & -4 \\ -2 & -4 & 1 & 9 \end{bmatrix}.$$

Verificación:

$$AA^{-1} = \begin{bmatrix} 1 & -1 & -1 & 2 \\ 2 & 3 & 0 & -4 \\ 1 & 1 & -2 & -2 \\ 1 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & -1 & -6 \\ -4 & -7 & 2 & 16 \\ 1 & 2 & -1 & -4 \\ -2 & -4 & 1 & 9 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \; ▲$$

§ 6.6. Cálculo de los vectores propios con ayuda del método de Le Verrier—Faddéev

Si se conocen las matrices $B_1, B_2, \ldots, B_{-1}$, obtenidas por el método de Le Verrier — Faddéev, y las raíces $\lambda_1, \lambda_2, \ldots, \lambda_n$ del polinomio característico $D(\lambda)$, los vectores propios $\mathbf{x}^{(i)}$ se pueden calcular con ayuda de la fórmula

$$\mathbf{x}^{(i)} = \lambda_i^{(n-1)}\mathbf{e} + \lambda_i^{(n-2)}\mathbf{b}_1 + \lambda_i^{(n-3)}\mathbf{b}_2 + \ldots + \mathbf{b}_{n-1},$$

donde $\mathbf{e}$ es cualquier vector unitario y $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_{n-1}$ son los vectores columna de las matrices $B_1, B_2, \ldots, B_{n-1}$, homónimos con $\mathbf{e}$.

Ejemplo. Calcular los vectores propios de la matriz

$$A=\begin{bmatrix} 1 & -1 & -1 & 2 \\ 2 & 3 & 0 & -4 \\ 1 & 1 & -2 & -2 \\ 1 & 1 & 0 & -1 \end{bmatrix},$$

si se conocen las matrices B_1, B_2, B_3 (véase el ejemplo 1 en el § 6.5) y los valores propios $\lambda_1 = \lambda_2 = 1$; $\lambda_3 = 0{,}618$; $\lambda_4 = -1{,}168$ del polinomio característico $D(\lambda) = \lambda^4 - \lambda^3 - 2\lambda^2 + 3\lambda - 1$.

△ Tomemos $\mathbf{e}=\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$; entonces $\mathbf{b}_1=\begin{bmatrix} 2 \\ -4 \\ -2 \\ -2 \end{bmatrix}$; $\mathbf{b}_2=\begin{bmatrix} 4 \\ 0 \\ 6 \\ -2 \end{bmatrix}$;

$$\mathbf{b}_3=\begin{bmatrix} -6 \\ 16 \\ -4 \\ 9 \end{bmatrix}$$

(las columnas 4 de las matrices B_1, B_2, B_3). Por la fórmula $\mathbf{x}^{(i)} = \lambda_i^3\mathbf{e} + \lambda_i^2\mathbf{b}_1 + \lambda_i\mathbf{b}_2 + \mathbf{b}_3$ hallamos:

$$\mathbf{x}^{(1)} = \mathbf{x}^{(2)} = 1\cdot\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + 1\cdot\begin{bmatrix} 2 \\ -4 \\ -2 \\ -2 \end{bmatrix} + 1\cdot\begin{bmatrix} 4 \\ 0 \\ 6 \\ -2 \end{bmatrix} + \begin{bmatrix} -6 \\ 16 \\ -4 \\ 9 \end{bmatrix} = \begin{bmatrix} 0 \\ 12 \\ 0 \\ 6 \end{bmatrix};$$

$$\mathbf{x}^{(3)} = 0{,}618^3\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + 0{,}618^2\begin{bmatrix} 2 \\ -4 \\ -2 \\ -2 \end{bmatrix} + 0{,}618\begin{bmatrix} 4 \\ 0 \\ 6 \\ -2 \end{bmatrix} +$$

$$+\begin{bmatrix} -6 \\ 16 \\ -4 \\ 9 \end{bmatrix} = \begin{bmatrix} -2{,}764 \\ 14{,}472 \\ -1{,}056 \\ 7{,}236 \end{bmatrix};$$

$$\mathbf{x}^{(4)} = (-1{,}618)^3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} + (-1{,}618)^2 \begin{bmatrix} 2 \\ -4 \\ -2 \\ -2 \end{bmatrix} + (-1{,}618) \begin{bmatrix} 4 \\ 0 \\ 6 \\ -2 \end{bmatrix} + \begin{bmatrix} -6 \\ 16 \\ -4 \\ 9 \end{bmatrix} = \begin{bmatrix} -7{,}236 \\ 5{,}528 \\ -18{,}944 \\ 2{,}764 \end{bmatrix}.$$

Escribimos los resultados de los cálculos en la tabla siguiente:

λ_i	I	II	III	IV	V	VI
$\lambda_1=\lambda_2=1$	0 0 0 1	2 −4 −2 −2	4 0 6 −2	−6 16 −4 6	0 12 0 6	0 1 0 0,5
$\lambda_3=0{,}618$	0 0 0 0,236	0,764 −1,528 −0,764 −0,764	2,472 0 3,708 −1,236	−6 16 −4 9	−2,764 14,472 −1,056 7,236	0,19 1 −0,07 0,50
$\lambda_4=-1{,}618$	0 0 0 −4,236	5,236 −10,472 −5,236 −5,236	−6,472 0 −9,708 3,236	−6 16 −4 9	−7,236 5,528 −18,944 2,764	−0,38 0,29 1 0,15

En las columnas II, III, IV están escritas las coordenadas de la columna 4 de las matrices B_i, multiplicadas por los grados correspondientes de λ_i y en la columna 1, las coordenadas del vector $\lambda_i^3 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$. La columna V contiene las coordenadas de los vectores $\mathbf{x}^{(i)}$ y la columna VI, sus coordenadas después de la normación. ▲

§ 6.7. Método de Danilevski

Dos matrices A y B se llaman *semejantes* si una se obtiene de la otra por medio de la transformación con ayuda de cierta matriz regular, o sea, se cumple la igualdad

$$B = S^{-1}AS.$$

Si la matriz B es semejante a la A, se escribe $B \backsim A$.

Al construir el esquema de cálculo, en el **método de Danilevski** se utiliza la propiedad fundamental de las matrices semejantes: *las matrices semejantes tienen iguales polinomios característicos.*

Si se reduce la matriz dada

$$A=\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ a_{31} & a_{32} & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix} \tag{1}$$

con ayuda de las transformaciones de semejanza a la llamada *forma de Frobenius*

$$F=\begin{bmatrix} f_{11} & f_{12} & f_{13} & \dots & f_{1,n-1} & f_{1n} \\ 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & 0 \end{bmatrix} \tag{2}$$

y luego se desarrolla el determinante

$$\det(F-\lambda E)=\begin{vmatrix} f_{11}-\lambda & f_{12} & f_{13} & \dots & f_{1,n-1} & f_{1n} \\ 1 & -\lambda & 0 & \dots & 0 & 0 \\ 0 & 1 & -\lambda & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -\lambda \end{vmatrix} \tag{3}$$

según los elementos de la primera fila, obtenemos

$$D(\lambda)=\det(F-\lambda E)=(f_{11}-\lambda)(-\lambda)^{n-1}-f_{12}(-\lambda)^{n-2}+ \\ +f_{13}(-\lambda)^{n-3}-\dots+(-1)^{n-1}f_{1n},$$

o bien

$$D(\lambda)=\det(F-\lambda E)=(-1)^n(\lambda^n-p_1\lambda^{n-1}-p_2\lambda^{n-2}- \\ -p_3\lambda^{n-3}-\dots p_n). \tag{4}$$

Aquí $p_1=f_{11}$, $p_2=f_{12}$, $p_3=f_{13}$, $\dots$, $p_n=f_{1n}$ son los coeficientes del polinomio característico de la matriz F los cuales en virtud de semejanza de las matrices F y A son coeficientes del polinomio característico de la matriz dada A.

Según el método de Danilevski el paso de la matriz A a la matriz semejante de Frobenius F se lleva a cabo con ayuda de $n-1$ transformaciones de semejanza que transforman sucesivamente las filas de la matriz A, comenzando con la última, en filas correspondientes de la matriz F.

Esquema de transformación de la matriz A en matriz semejante de Frobenius F. 1°. Supongamos que nos hace falta convertir la fila $a_{n1}a_{n2}\dots a_{n,n-1}a_{nn}$ en fila $0\ 0\dots 1\ 0$. Admitiendo que $a_{n,n-1}\neq$

$\neq 0$, dividamos todos los elementos de la columna $n-1$ de la matriz A por $a_{n,\,n-1}$. Entonces su n-ésima fila toma la forma

$$a_{n1}a_{n2}\ldots\frac{a_{n,\,n-1}}{a_{n,\,n-1}}a_{nn} \quad \text{o bien } a_{n1}a_{n2}\ldots 1\ a_{nn}.$$

2°. Sustrayamos la columna $n-1$ de la matriz transformada, multiplicada, correspondientemente, por los números $a_{n1}, a_{n2}, \ldots$ $\ldots, a_{nn}$, de todas las demás columnas. Para la n-ésima fila obtenemos

$$a_{n1}-a_{n1}a_{n2}-a_{n2}\ \ldots\ 1\ a_{nn}-a_{nn}, \text{ o bien } 0\ 0\ \ldots\ 1\ 0.$$

3°. En calidad de matriz regular tomamos la matriz M_{n-1} que se obtiene de la matriz unidad después de las mismas transformaciones:

$$M_{n-1}=\begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ m_{n-1,\,1} & m_{n-1,\,2} & & m_{n-1,\,n-1} & m_{n-1,\,n} \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix},$$

donde

$$m_{n-1,\,i}=-\frac{a_{ni}}{a_{n,\,n-1}}, \quad m_{n-1,\,n-1}=\frac{1}{a_{n,\,n-1}}. \tag{5}$$

Las operaciones realizadas son equivalentes a la multiplicación, a la derecha, de la matriz M_{n-1} por la A:

$$B=AM_{n-1}=\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1,\,n-1} & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2,\,n-1} & a_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{n-1,\,1} & a_{n-1,\,2} & \ldots & a_{n-1,\,n-1} & a_{n-1,\,n} \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}\times$$

$$\times\begin{bmatrix} 1 & 0 & \ldots & 0 & 0 \\ 0 & 1 & \ldots & 0 & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ m_{n-1,\,1} & m_{n-1,\,2} & \ldots & m_{n-1,\,n-1} & m_{n-1,\,n} \\ 0 & 0 & \ldots & 0 & 1 \end{bmatrix}=$$

$$=\begin{bmatrix} b_{11} & b_{12} & \ldots & b_{1,\,n-1} & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2,\,n-1} & b_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ b_{n-1,\,1} & b_{n-1,\,2} & \ldots & b_{n-1,\,n-1} & b_{n-1,\,n} \\ 0 & 0 & \ldots & 1 & 0 \end{bmatrix}.$$

Los elementos de la matriz B se calculan con ayuda de las fórmulas

$$b_{ij}=a_{ij}+a_{i,\,n-1}m_{n-1,\,j}; \quad b_{i,n-1}=a_{i,n-1}m_{n-1,\,n-1}. \tag{6}$$

Sin embargo, la matriz construida $B = AM_{n-1}$ no será semejante a la matriz A.

4°. Para obtener la transformación de semejanza es necesario multiplicar a la izquierda la matriz inversa M_{n-1}^{-1} por la matriz B:

$$M_{n-1}^{-1}AM_{n-1} = M_{n-1}^{-1}B.$$

La matriz inversa M_{n-1}^{-1} tiene la forma

$$M_{n-1}^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{n,\,n-1} & a_{nn} \\ 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

Suponemos $M_{n-1}^{-1}AM_{n-1} = C$; por consiguiente, $C = M_{n-1}^{-1}B$. La multiplicación de la matriz M_{n-1}^{-1} a la izquierda por la matriz B no cambia la fila transformada de esta última y la matriz C tiene la forma

$$C = M_{n-1}^{-1}B = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{n,\,n-1} & a_{nn} \\ 0 & 0 & & 0 & 1 \end{bmatrix} \times$$

$$\times \begin{bmatrix} b_{11} & b_{12} & \dots & b_{1,\,n-1} & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2,\,n-1} & b_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ b_{n-1,\,1} & b_{n-1,\,2} & \dots & b_{n-1,\,n-1} & b_{n-1,\,n} \\ 0 & 0 & & 1 & 0 \end{bmatrix} =$$

$$= \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1,\,n-1} & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2,\,n-1} & c_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ c_{n-1,\,1} & c_{n-1,\,2} & \dots & c_{n-1,\,n-1} & c_{n-1,\,n} \\ 0 & 0 & & 1 & 0 \end{bmatrix}.$$

En efecto, multiplicando las matrices M_{n-1}^{-1} y B, cambiamos sólo la fila $(n-1)$-ésima de la matriz B, ya que $c_{ij} = b_{ij}$ para todas las demás filas. Los elementos de esta fila se determinan con ayuda de las fórmulas

$$c_{n-1,\,j} = \sum_{k=1}^{n} a_{nk}b_{kj} \quad (j = 1,\ 2,\ \dots,\ n). \tag{7}$$

La matriz obtenida C es semejante a la A y tiene una fila reducida.

5°. Luego, si $c_{n-1\ n-2} \neq 0$, entonces con la matriz C repetimos las operaciones análogas, tomando por fundamental la fila $n-2$. En este caso, utilizando la matriz intermedia $D = CM_{n-2}$, como resultado obtenemos la matriz $E = M_{n-2}^{-1}D = M_{n-2}^{-1}CM_{n-2}$ con dos filas reducidas. Con la matriz E hacemos las mismas operaciones, etc. hasta que se obtenga la matriz de Frobenius.

Todas estas transformaciones se formalizan en forma esquemática de cálculo, el proceso de composición de la cual consideremos en un ejemplo.

Tabla 6.6

Filas	M^{-1}	Columnas 1	2	3	4	Σ	Σ'
1		−4	−3	1	1	−5	
2		2	0	4	−1	5	
3		1	1	2	−2	2	
4		1	1	$\boxed{-1}$	−1	0	
I	M_3 / M_3^{-1}	1	1	−1	−1	0	
5	1	−3	−2	−1	0	−6	−5
6	1	6	4	−4	−5	1	5
7	−1	3	3	−2	−4	0	2
8	−1	0	0	1	0	1	0
7′		0	$\boxed{-1}$	−4	−1	−6	
II	M_2 / M_2^{-1}	0	−1	−4	−1	−6	
9	0	−3	2	7	2	8	6
10	−1	6	−4	−20	−9	−27	−23
11	−4	0	1	0	0	1	0
12	−1	0	0	1	0	1	0
10′		$\boxed{-6}$	0	19	9	22	
III	M_1 / M_1^{-1}	0,167−1	0	3,167	1,500	3,667	
13	−6	0,500	2,000	−2,500	−2,500	−2,500	−3
14	0	1	0	0	0	1	0
15	19	0	1	0	0	1	0
16	9	0	0	1	0	1	0
13′		−3	7	24	15	43	

Ejemplo 1. Haciendo uso del método de Danilevski, desarrollar el determinante característico

$$A=\begin{bmatrix} -4 & -3 & 1 & 1 \\ 2 & 0 & 4 & -1 \\ 1 & 1 & 2 & -2 \\ 1 & 1 & -1 & -1 \end{bmatrix}.$$

△ *Etapa I.* Reduzcamos la matriz A a la forma de Frobenius. Para los cálculos hacemos la tabla (véase la tabla 6.6).

1) En las filas 1 — 4 de la tabla de cálculo colocamos los elementos a_{ij} (i, j = 1, 2, 3, 4) de la matriz dada A y las sumas de control $a_{i5} = \sum\limits_{j=1}^{4} a_{ij}$ (i = 1, 2, 3, 4), o sea, la columna Σ. Marcamos el elemento $a_{43} = -1$ perteneciente a la tercera columna (columna marcada).

2) En la fila 1 escribimos los elementos de la tercera fila de la matriz $M_{n-1} = M_3$ los cuales se calculan con ayuda de las fórmulas (5):

$$m_{31} = -\frac{a_{41}}{a_{43}} = -\frac{1}{-1} = 1; \quad m_{32} = -\frac{a_{42}}{a_{43}} = -\frac{1}{-1} = 1;$$
$$m_{33} = \frac{1}{a_{43}} = \frac{1}{-1} = -1; \quad m_{34} = -\frac{a_{44}}{a_{43}} = -\frac{-1}{-1} = -1;$$
$$m_{35} = -\frac{a_{45}}{a_{43}} = \frac{0}{-1} = 0.$$

El número 0 debe coincidir con la suma de los elementos de la fila 1 después de sustituir el valor obtenido del elemento m_{33} por -1, pero en el ejemplo dado $m_{33} = -1$. (Para comodidad el número -1 suele escribirse junto al elemento m_{33} y se separa de este último por la raya).

3) En las filas 5...8 en la columna M^{-1} escribimos la tercera fila de la matriz M_3^{-1} la cual debe coincidir con la cuarta fila de la matriz inicial A.

4) En las filas 5...8 en las columnas correspondientes escribimos los elementos de la matriz $B = A \cdot M_3$ los cuales se calculan por las fórmulas (6) para las columnas no marcadas.

La primera columna:

$$b_{11} = a_{11} + a_{13}m_{31} = -4 + 1\cdot 1 = -3;$$
$$b_{21} = a_{21} + a_{23}m_{31} = 2 + 4\cdot 1 = 6;$$
$$b_{31} = a_{31} + a_{33}m_{31} = 1 + 2\cdot 1 = 3;$$
$$b_{41} = a_{41} + a_{43}m_{31} = 1 + (-1)\cdot 1 = 0.$$

La segunda columna:

$$b_{12} = a_{12} + a_{13}m_{32} = -3 + 1\cdot 1 = -2;$$
$$b_{22} = a_{22} + a_{23}m_{32} = 0 + 4\cdot 1 = 4;$$
$$b_{32} = a_{32} + a_{33}m_{32} = 1 + 2\cdot 1 = 3;$$
$$b_{42} = a_{42} + a_{43}m_{32} = 1 + (-1)\cdot 1 = 0.$$

La cuarta columna:

$$b_{14} = a_{14} + a_{13} \cdot m_{34} = 1 + 1 \cdot (-1) = 0;$$
$$b_{24} = a_{24} + a_{23} m_{34} = -1 + 4 \cdot (-1) = -5;$$
$$b_{34} = a_{34} + a_{33} m_{34} = -2 + 2 \cdot (-1) = -4;$$
$$b_{44} = a_{44} + a_{43} m_{34} = -1 + (-1) \cdot (-1) = 0.$$

Los elementos transformados de la tercera columna (marcada) se obtienen con ayuda de la multiplicación de los elementos iniciales por $m_{33} = -1$.

La tercera columna:

$$b_{33} = a_{13} m_{33} = 1 \cdot (-1) = -1; \quad b_{23} = a_{23} \cdot m_{33} = 4 \cdot (-1) = -4;$$
$$b_{33} = a_{33} m_{33} = 2 \cdot (-1) = -2; \quad b_{43} = a_{43} m_{33} = (-1)(-1) = 1.$$

La última fila de la matriz B debe teñer la forma 0 0 1 0.

Para el control completamos la matriz B por los elementos correspondientes de la columna Σ' transformados según las fórmulas análogas de dos términos para $m_{35} = 0$:

$$b_{16} = a_{15} + a_{13} m_{35} = -5 + 1 \cdot 0 = -5;$$
$$b_{26} = a_{25} + a_{23} m_{35} = 5 + 4 \cdot 0 = 5;$$
$$b_{36} = a_{35} + a_{33} m_{35} = 2 + 2 \cdot 0 = 2;$$
$$b_{46} = a_{45} + a_{43} m_{35} = 0 + (-1) \cdot 0 = 0.$$

Escribimos los resultados obtenidos en las filas correspondientes de la columna Σ'. Al adicionar a los elementos de la columna Σ' los elementos respectivos de la tercera columna (marcada), obtenemos las sumas de control para las filas 5...8 ($i = 1, 2, 3, 4$).

La columna Σ:

$$b_{15} = b_{16} + a_{13} = -5 - 1 = -6;$$
$$b_{25} = b_{26} + a_{23} = 5 - 4 = 1,$$
$$b_{35} = b_{36} + a_{33} = 2 - 2 = 0;$$
$$b_{45} = b_{46} + a_{43} = 0 + 1 = 1.$$

Además, los elementos de la columna Σ para el control se calculan con ayuda de la fórmula $b_{i5} = \sum_{j=1}^{4} b_{ij}$ ($i = 1, 2, 3, 4$):

$$b_{15} = b_{11} + b_{12} + b_{13} + b_{14} = -3 - 2 - 1 + 0 = -6;$$
$$b_{25} = b_{21} + b_{22} + b_{23} + b_{24} = 6 + 4 - 4 - 5 = 1;$$
$$b_{35} = b_{31} + b_{32} + b_{33} + b_{34} = 3 + 3 - 2 - 4 = 0;$$
$$b_{45} = b_{41} + b_{42} + b_{43} + b_{44} = 0 + 0 + 1 + 0 = 1.$$

La matriz B tiene la forma siguiente:

$$B = \begin{bmatrix} -3 & -2 & -1 & 0 \\ 6 & 4 & -4 & -5 \\ 3 & 3 & -2 & -4 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

5) La transformación M_3^{-1} realizada en la matriz B, produciendo como resultado la matriz $C = M_3^{-1}B$, cambia sólo la tercera fila de la matriz B, o sea, la séptima fila de la tabla. Los elementos de esta fila transformada 7' no son sino las sumas de los productos pares de los elementos de la columna M_3^{-1}, colocados en las filas 5...8, por los elementos correspondientes de cada una de las columnas de la matriz B [véanse las fórmulas (7)]:

$$c_{31} = 1\cdot(-3) + 1\cdot 6 + (-1)\cdot 3 + (-1)\cdot 0 = 0;$$
$$c_{32} = 1\cdot(-2) + 1\cdot 4 + (-1)\cdot 3 + (-1)\cdot 0 = -1;$$
$$c_{33} = 1\cdot(-1) + 1\cdot(-4) + (-1)\cdot(-2) + (-1)\cdot 1 = -4;$$
$$c_{34} = 1\cdot 0 + 1\cdot(-5) + (-1)(-4) + (-1)\cdot 0 = -1.$$

Realizamos las mismas transformaciones en la columna Σ:

$$c_{35} = 1\cdot(-6) + 1\cdot 1 + (-1)\cdot 0 + (-1)\cdot 1 = -6.$$

Como resultado obtenemos la matriz C compuesta por las filas 5, 6, 7', 8 con las sumas de control en la columna Σ:

$$C = \begin{bmatrix} -3 & -2 & -1 & 0 \\ 6 & 4 & -4 & -5 \\ 0 & \boxed{-1} & -4 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

La matriz C es semejante a la matriz A y tiene una fila reducida. En esto se termina la construcción de la primera transformación semejante: $C = M_2^{-1}AM$.

Etapa II. Tomando por inicial la matriz C, separamos el elemento $c_{32} = -1$ (segunda columna) y continuamos procediendo de un modo análogo.

1) Hallamos los elementos de la matriz $M_{n-2} = M_2$ son ayuda de las fórmulas (5):

$$m_{21} = -\frac{c_{31}}{c_{32}} = -\frac{0}{-1} = 0; \quad m_{22} = \frac{1}{c_{32}} = \frac{1}{-1} = -1;$$
$$m_{23} = -\frac{c_{33}}{c_{32}} = -\frac{-4}{-1} = -4; \quad m_{24} = -\frac{c_{34}}{c_{32}} = -\frac{-1}{-1} = -1;$$
$$m_{25} = -\frac{c_{35}}{c_{32}} = -\frac{-6}{-1} = -6.$$

Vamos a sumar: $0 - 1 - 4 - 1 = -6$ ($m_{22} = -1$; si $m_{22} \neq -1$, sería necesario reemplazar m_{22} por -1).

2) En las filas 9...12 en la columna M^{-1} escribimos la segunda fila de la matriz M_2^{-1} la cual coincide con la tercera fila de la matriz C (véase la tabla 6.6). Hallamos los elementos de la matriz $D = CM_2$.

La primera columna:

$$d_{11} = -3 + (-2)\cdot 0 = -3; \quad d_{21} = 6 + 4\cdot 0 = \quad -6;$$
$$d_{31} = 0 + 0 = 0.$$

La segunda columna (marcada) se obtiene multiplicando los elementos correspondientes de la matriz C por $m_{22} = -1$:

$$d_{12} = c_{12}m_{22} = (-2)\cdot(-1) = 2;\ d_{22} = c_{22}m_{22} = 4\cdot(-1) = -4;$$
$$d_{32} = c_{32}m_{22} = (-1)\cdot(-1) = 1.$$

La tercera columna:

$$d_{13} = c_{13} + c_{12}m_{23} = -1 + (-2)\cdot(-4) = 7;$$
$$d_{23} = c_{23} + c_{22}m_{23} = -4 + 4\cdot(-4) = -20;$$
$$d_{33} = c_{33} + c_{32}m_{23} = -4 + (-1)\cdot(-4) = 0.$$

La cuarta columna:

$$d_{14} = c_{14} + c_{12}m_{24} = 0\cdot 1 + (-2)\cdot(-1) = 2;$$
$$d_{24} = c_{24} + c_{22}m_{24} = -5 + 4\cdot(-1) = -9;$$
$$d_{34} = c_{34} + c_{32}m_{24} = -1 + (-1)\cdot(-1) = 0.$$

La columna Σ':

$$d_{16} = c_{15} + c_{12}m_{25} = -6 + (-2)\cdot(-6) = 6;$$
$$d_{26} = c_{25} + c_{22}m_{25} = 1 + 4\cdot(-6) = -23;$$
$$d_{36} = c_{35} + c_{32}m_{25} = -6 + (-1)\,(-6) = 0.$$

Los elementos de la matriz Σ se obtienen sumando los elementos de la columna Σ' con los elementos correspondientes de la columna marcada:

$$d_{15} = d_{16} + d_{12} = 6 + 2 = 8;$$
$$d_{25} = d_{26} + d_{22} = -23 - 4 = -27;$$
$$d_{36} = d_{36} + d_{32} = 0 + 1 = 1.$$

La matriz D tiene la forma

$$D = \begin{bmatrix} -3 & 2 & 7 & 2 \\ 6 & -4 & -20 & -9 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

La transformación M_2^{-1} que se realiza en la matriz D y ofrece la matriz $E = M_2^{-1}D$ cambia sólo la segunda fila de la matriz D, o sea, la décima fila de la tabla. Los elementos de esta fila transformada 10′ no son sino las sumas de los productos pares de los elementos de la columna M_2^{-1} que están en las filas 9...12:

$$\begin{aligned}
e_{21} &= 0\cdot(-3) + (-1)\cdot 6 + (-4)\cdot 0 + (-1)\cdot 0 = -6;\\
e_{22} &= 0\cdot 2 + (-1)\cdot(-4) + (-4)\cdot 1 + (-1)\cdot 0 = 0;\\
e_{23} &= 0\cdot 7 + (-1)\cdot(-20) + (-4)\cdot 0 + (-1)\cdot 1 = 19;\\
e_{24} &= 0\cdot 2 + (-1)\cdot(-9) + (-4)\cdot 0 + (-1)\cdot 0 = 9;\\
e_{25} &= 0\cdot 8 + (-1)\cdot(-27) + (-4)\cdot 1 + (-1)\cdot 1 = 22;
\end{aligned}$$

$$\Sigma e_{2j} = -6 + 0 + 19 + 9 = 22.$$

En esto termina la construcción de la segunda transformación semejante: $E = M_2^{-1}CM_2$. La matriz $E \backsim C$ contiene dos filas reducidas:

$$E = \begin{bmatrix} -3 & 2 & 7 & 2 \\ \boxed{-6} & 0 & 19 & 9 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Etapa III. Tomamos por inicial la matriz E. En ésta separamos el elemento $e_{21} = -6$ (primera columna) y transformamos la matriz E en semejante matriz de Frobenius F. Continuando de un modo análogo el proceso, con ayuda de las fórmulas (5) determinemos los elementos de la matriz $M_{n-3} = M_1$:

$$m_{11} = \frac{1}{e_{21}} = \frac{1}{-6} = -0,167; \quad m_{12} = -\frac{e_{22}}{e_{21}} = -\frac{0}{-6} = 0;$$

$$m_{13} = -\frac{e_{23}}{e_{21}} = -\frac{19}{-6} = 3,167; \quad m_{14} = -\frac{e_{24}}{e_{21}} = -\frac{9}{-6} = 1,500;$$

$$m_{15} = -\frac{e_{25}}{e_{21}} = -\frac{22}{-6} = 3,667.$$

Para que se obtenga la suma $\Sigma = 3,667$ reemplazamos $m_{11} = -0,167$ por -1;

$$\Sigma = -1 + 0 + 3,167 + 1,500 = 3,667.$$

Vamos a escribir la matriz de Frobenius F en las filas 13...16. Primero construimos $G = EM_1$ y luego $F = M_1^{-1}G$. En la columna M_1^{-1} escribimos la fila 10′ de la matriz E (véase la tabla 6.6).

La primera columna (marcada):

$$\begin{aligned}
g_{11} &= e_{11}m_{11} = (-3)\cdot(-0,167) = 0,500;\\
g_{21} &= e_{21}m_{11} = (-6)\cdot(-0,167) = 1,000.
\end{aligned}$$

La segunda columna:

$$\begin{aligned}
g_{12} &= e_{12} + e_{11}m_{12} = 2 + (-3)\cdot 0 = 2,000;\\
g_{22} &= e_{22} + e_{21}m_{12} = 0 + (-6)\cdot 0 = 0.
\end{aligned}$$

La tercera columna:

$$g_{13} = e_{13} + e_{11}m_{13} = 7 + (-3)\cdot 3{,}167 = -2{,}500;$$
$$g_{23} = e_{23} + e_{21}m_{13} = 19 + (-6)\cdot 3{,}167 = 0.$$

La cuarta columna:

$$g_{14} = e_{14} + e_{11}m_{14} = 2 + (-3)\cdot 1{,}500 = -2{,}500;$$
$$g_{24} = e_{24} + e_{21}m_{14} = 9 + (-6)\cdot 1{,}500 = 0.$$

La columna Σ':

$$g_{16} = e_{15} + e_{11}m_{15} = 8 + (-3)\cdot 3{,}667 = -3;$$
$$g_{26} = e_{25} + e_{21}m_{15} = 22 + (-6)\cdot 3{,}667 = 0.$$

La columna Σ:

$$g_{15} = g_{16} + g_{11} = -3 + 0{,}500 = -2{,}500;$$
$$g_{15} = g_{11} + g_{12} + g_{13} + g_{14} = 0{,}500 + 2{,}000 - 2{,}500 - 2{,}500 = = -2{,}500;$$
$$g_{25} = g_{26} + g_{21} = 0 + 1 = 1;$$
$$g_{25} = g_{21} + g_{22} + g_{23} + g_{24} = 1 + 0 + 0 + 0 = 1.$$

Los elementos de la fila transformada 13′ no son más que las sumas de los productos pares de la columna M_1^{-1} que están en las filas 13...16:

$$f_{11} = (-6)\cdot 0{,}500 + 0\cdot 1 + 19\cdot 0 + 9\cdot 0 = -3;$$
$$f_{12} = (-6)\cdot 2{,}000 + 0\cdot 0 + 19\cdot 1 + 9\cdot 0 = 7;$$
$$f_{13} = (-6)(-2{,}500) + 0\cdot 0 + 19\cdot 0 + 9\cdot 1 = 24;$$
$$f_{14} = (-6)(-2{,}500) + 0\cdot 0 + 19\cdot 0 + 9\cdot 0 = 15.$$
$$\Sigma = (-6)\cdot(-2{,}500) + 0\cdot 1 + 19\cdot 1 + 9\cdot 1 = 43;$$
$$\Sigma = -3 + 7 + 24 + 15 = 43.$$

Ahora bien, la matriz buscada de Frobenius F, semejante a la A, tiene la forma

$$F = \begin{bmatrix} -3 & 7 & 24 & 15 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Vamos a buscar el determinante característico de la matriz F:

$$D(\lambda) = \det(A - \lambda E) = \det(F - \lambda E) = \begin{vmatrix} -3-\lambda & 7 & 24 & 15 \\ 1 & -\lambda & 0 & 0 \\ 0 & 1 & -\lambda & 0 \\ 0 & 0 & 1 & -\lambda \end{vmatrix}.$$

De aquí, desarrollando el determinante $D(\lambda)$ según los elementos de la primera fila, obtenemos

$$D(\lambda)=\begin{vmatrix} -3-\lambda & 7 & 24 & 15 \\ 1 & -\lambda & 0 & 0 \\ 0 & 1 & -\lambda & 0 \\ 0 & 0 & 1 & \lambda \end{vmatrix}=(-3-\lambda)\begin{vmatrix} -\lambda & 0 & 0 \\ 1 & -\lambda & 0 \\ 0 & 1 & -\lambda \end{vmatrix}-$$

$$-7\begin{vmatrix} 1 & 0 & 0 \\ 0 & -\lambda & 0 \\ 0 & 1 & -\lambda \end{vmatrix}+24\begin{vmatrix} 1 & -\lambda & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -\lambda \end{vmatrix}-15\begin{vmatrix} 1 & -\lambda & 0 \\ 0 & 1 & -\lambda \\ 0 & 0 & 1 \end{vmatrix}=$$

$$=(-3-\lambda)(-\lambda^3)-7\lambda^2+24(-\lambda)-15=$$

$$=\lambda^4+3\lambda^3-7\lambda^2-24\lambda-15. \ \blacktriangle$$

Casos excepcionales en el método de Danilevski. Este método es aplicable sin complicaciones algunas si todos los elementos a separar son distintos del cero (al igual que en el ejemplo recién considerado). En cambio, si al transformar la matriz $A=[a_{ij}]$ $(i=1, 2, \ldots, n;\ j=1, 2, \ldots, n)$ en matriz de Frobenius F se obtiene la matriz en la forma

$$D=\begin{bmatrix} d_{11} & d_{12} & \ldots & d_{1l} & \ldots & d_{1,k-1} & d_{1k} & \ldots & d_{1,n-1} & d_{1n} \\ d_{21} & d_{22} & \ldots & d_{2l} & \ldots & d_{2,k-1} & d_{2k} & \ldots & d_{2,n-1} & d_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ d_{l1} & d_{l2} & \ldots & d_{ll} & \ldots & d_{l,k-1} & d_{lk} & \ldots & d_{l,n-1} & d_{ln} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ d_{k1} & d_{k2} & \ldots & d_{kl} & \ldots & d_{k,k-1} & d_{kk} & \ldots & d_{k,n-1} & d_{kn} \\ 0 & 0 & \ldots & 0 & \ldots & 0 & 1 & \ldots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \ldots & 0 & \ldots & 0 & 0 & \ldots & 1 & 0 \end{bmatrix},$$

con la particularidad de que resulta que $d_{k,k-1}=0$, no se puede continuar las transformaciones siguiendo el método de Danilevski. Aquí son posibles dos casos.

Caso I. Supongamos que cualquier elemento de la matriz D, el cual está a la izquierda del elemento nulo $d_{k,k-1}$, es distinto del cero; por ejemplo, $d_{kl}\neq 0$, donde $l<k-1$. Entonces promovemos este elemento al lugar del elemento nulo $d_{k,k-1}$, o sea, permutamos la columna $k-1$ y la columna l de la matriz D y simultáneamente su fila $k-1$ y su fila l. La nueva matriz será semejante a la dada y se puede continuar los cálculos según el método de Danilevski.

Caso II. Sea $d_{kl}=0$ $(l=1, 2, \ldots, k-1)$, es decir, el elemento separado, así como todos los elementos de la matriz que están a la izquierda del separado son iguales a cero. Entonces la matriz D tiene la forma

$$D = \left[\begin{array}{cccc|cccc}
d_{11} & d_{12} & \dots & d_{1,k-1} & d_{1k} & \dots & d_{1,n-1} & d_{1n} \\
d_{21} & d_{22} & \dots & d_{2,k-1} & d_{2k} & \dots & d_{2,n-1} & d_{2n} \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
d_{k-1,1} & d_{k-1,2} & \dots & d_{k-1,k-1} & d_{k-1,k} & \dots & d_{k-1,n-1} & d_{k-1,n} \\
\hline
0 & 0 & \dots & 0 & d_{kk} & \dots & d_{k,n-1} & d_{kn} \\
0 & 0 & \dots & 0 & 1 & \dots & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
0 & 0 & \dots & 0 & 0 & \dots & 1 & 0
\end{array}\right] =$$

$$= \left[\begin{array}{c|c} D_1 & D_2 \\ \hline 0 & D_3 \end{array}\right].$$

Partimos la matriz D en cuatro células de un modo tal que una matriz sea nula. Entonces el determinante característico $\det(D - \lambda E)$ se descompone en dos determinantes:

$$\det(D - \lambda E) = \det(D_1 - \lambda E)\cdot\det(D_3 - \lambda E),$$

pero la matriz D_3 tiene ya la forma de Frobenius, por eso nos queda sólo por reducir a esta forma la matriz D_1.

Tabla 6.7

Filas	M_I	Columnas 1	2	3	4	Σ	Σ'
1		3	−2	1	−1	1	
2		3	−2	1	1	3	
3		5	−4	2	0	3	
4		−1	−1	$\boxed{1}$	1	0	
I	M_3 \ M_3^{-1}	1	1	1 | −1	−1	0	
5	−1	4	−1	1	−2	2	1
6	−1	4	−1	1	0	4	3
7	1	7	−2	2	−2	5	3
8	1	0	0	1	0	1	0
7′		−1	$\boxed{0}$	1	0	0	

Ejemplo 2. Haciendo uso del método de Danilevski, desarrollar el determinante característico de la matriz

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

△ Escribimos los cálculos en la tabla (véase la tabla 6.7). El elemento a separar $c_{32} = 0$; no se puede continuar los cálculos siguiendo el esquema de Danilevski. Puesto que $c_{31} \neq 0$, permutamos las columnas primera y segunda y las filas primera y segunda de la matriz C y continuamos los cálculos (véase la tabla 6.8).

Tabla 6.8

Filas	M^{-1}	Columnas				Σ	Σ'
		1	2	3	4		
5	−1	−1	4	1	0	4	
6	−1	−1	4	1	−2	2	
7′	1	0	$\boxed{-1}$	1	0	0	
8	1	0	0	1	0	1	
II	M_2 / M_2^{-1}	0	−1	1	0	0	
9	0	−1	−4	5	0	0	4
10	−1	−1	−4	5	−2	−2	2
11	1	0	1	0	0	1	0
12	0	0	0	1	0	1	0
10′		$\boxed{1}$	5	−5	2	3	
III	M_1 / M_1^{-1}	$1 \mid -1$	−5	5	−2	−3	
13	1	−1	1	0	2	2	3
14	5	1	0	0	0	1	0
15	−5	0	1	0	0	1	0
16	2	0	0	1	0	1	0
13′		4	−4	2	2	4	

Como resultado obtenemos la matriz de Frobenius $F \backsim A$:

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

de donde hallamos

$$D(\lambda) = \det(A - \lambda E) = \det(F - \lambda E) = \lambda^4 - 4\lambda^3 + \\ + 4\lambda^2 - 2\lambda + 2. \ \blacktriangle$$

Ejemplo 3. Haciendo uso del método de Danilevski, desarrollar el determinante característico de la matriz

$$A = \begin{bmatrix} 0 & 1 & 3 & 2 \\ 1 & 4 & 5 & 0 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

$\triangle$ Escribamos los resultados de cálculos en la tabla (véase la tabla 6.9).

Puesto que el elemento a separar es igual a cero, no se puede continuar los cálculos según el esquema de Danilevski.

La matriz $D \backsim C$ tiene la forma

$$D = \left[\begin{array}{c|ccc} 1 & 0{,}5 & -2{,}5 & 2{,}5 \\ \hline 0 & 6 & 4 & -7 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{array}\right].$$

Partimos la matriz D en cuatro células por orladura y calculamos $D(\lambda)$:

$$D(\lambda) = \det(D - \lambda E) = \left|\begin{array}{c|ccc} 1-\lambda & 0{,}5 & -2{,}5 & 2{,}5 \\ \hline 0 & 6-\lambda & 4 & -7 \\ 0 & 1 & -\lambda & 0 \\ 0 & 0 & 1 & -\lambda \end{array}\right| =$$

$$= (1-\lambda) \begin{vmatrix} 6-\lambda & 4 & -7 \\ 1 & -\lambda & 0 \\ 1 & 1 & -\lambda \end{vmatrix} = (1-\lambda)\,[(6-\lambda)\lambda^2 + 4\lambda - 7] =$$

$$= \lambda^4 - 7\lambda^3 + 2\lambda^2 + 11\lambda - 7. \ \blacktriangle$$

Tabla 6.9

Filas	M^{-1}	Columnas 1	2	3	4	Σ	Σ'
1		0	1	3	2	6	
2		1	4	5	0	10	
3		1	1	2	1	5	
4		1	1	$\boxed{1}$	1	4	
I	M_3 / M_3^{-1}	−1	−1	1 | −1	−1	−4	
5	1	−3	−2	3	−1	−3	−6
6	1	−4	−1	5	−5	−5	−10
7	1	−1	−1	2	−1	−1	−3
8	1	0	0	1	0	1	0
7'		−8	$\boxed{-4}$	11	−7	−8	
II	M_2 / M_2^{-1}	−2	−0,25 | −1	2,75	−1,75	−2	
9	−8	1	0,5	−2,5	2,5	1,5	1
10	−4	−2	0,25	2,25	−3,25	−2,75	−3
11	11	0	1	0	0	1	0
12	−7	0	0	1	0	1	0
10'		$\boxed{0}$	6	4	−7	3	

§ 6.8. Cálculo de los vectores propios con ayuda del método de Danilevski

Sea $\mathbf{y} = (y_1, y_2, \ldots, y_n)$ el vector propio de la matriz de Frobenius F el cual corresponde al valor dado de λ. Entonces $F\mathbf{y} = \lambda\mathbf{y}$, de donde $(F - \lambda E)\,\mathbf{y} = 0$, o bien

$$\begin{bmatrix} f_{11}-\lambda & f_{12} & f_{13} & \ldots & f_{1n} \\ 1 & -\lambda & 0 & \ldots & -0 \\ 0 & 1 & -\lambda & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & & -\lambda \end{bmatrix} \cdot \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = 0. \qquad (1)$$

Realizando la multiplicación, obtenemos el sistema para determinar las coordenadas $y_1, y_2, \ldots, y_n$ del vector propio $\mathbf{y}$:

$$\left\{\begin{array}{l} (f_{11}-\lambda)\, y_1+f_{12}y_2+f_{13}y_3+\ldots+f_{1n}y_n=0, \\ \qquad y_1-\lambda y_2 \qquad\qquad\qquad\qquad\quad =0, \\ \qquad\qquad y_2-\lambda y_3 \qquad\qquad\qquad\quad =0 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ \qquad\qquad\qquad\qquad\qquad y_{n-1}-\lambda y_n=0. \end{array}\right\} \qquad (1')$$

Este sistema de ecuaciones lineales es homogéneo. Con exactitud hasta el coeficiente de proporcionalidad sus soluciones pueden hallarse del modo siguiente. Pongamos $y_n = 1$; entonces obtenemos sucesivamente

$$y_{n-1} = \lambda, \quad y_{n-2} = \lambda y_{n-1} = \lambda^2, \ \ldots, \ y_1 = \lambda^{n-1}.$$

Ahora bien, el vector propio buscado es

$$\mathbf{y}=\begin{bmatrix} \lambda^{n-1} \\ \lambda^{n-2} \\ \vdots \\ 1 \end{bmatrix}. \qquad (2)$$

Puesto que la matriz F es semejante a la matriz A, entonces λ es también el valor propio de la matriz A.

Designemos con $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ el vector propio de la matriz A el cual corresponde al valor de λ. Entonces obtenemos

$$\mathbf{x} = M_{n-1}M_{n-2} \ldots M_2M_1\mathbf{y}, \qquad (3)$$

donde $M_{n-1}, M_{n-2}, \ldots, M_1$ son las matrices unidad transformadas según el método de Danilevski.

Por ejemplo, la transformación M_1 realizada en $\mathbf{y}$ ofrece

$$M_1\mathbf{y}=\begin{bmatrix} m_{11} & m_{12} & \ldots & m_{1n} \\ 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & 1 \end{bmatrix}\cdot\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}=\begin{bmatrix} \sum\limits_{k=1}^{n} m_{1k}y_k \\ y_2 \\ \vdots \\ y_n \end{bmatrix}=$$

$$=\begin{bmatrix} \sum\limits_{k=1}^{n} m_{1k}y_k \\ \lambda^{n-2} \\ \vdots \\ 1 \end{bmatrix}.$$

Por consiguiente, la transformación M_1 cambia sólo la primera coordenada del vector $\mathbf{y}$. Análogamente, la transformación M_2 cambia sólo la segunda coordenada del vector $M_1\mathbf{y}$, etc. Repitiendo este proceso $n-1$ veces, obtenemos el valor propio buscado $\mathbf{x}$ de la matriz A.

Ejemplo. En el ejemplo 1 del § 6.7 hemos mostrado que la matriz

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

con ayuda del método de Danilevski se reduce a la forma de Frrobenius

$$F=\begin{bmatrix} -3 & 7 & 24 & 15 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}.$$

Calcular el vector propio $\mathbf{x}^{(1)} = (x_1, x_2, x_3, x_4)$ si $\lambda_1 = -1$.

△ Utilicemos la fórmula (3), o sea, $\mathbf{x} = M_3M_2M_1\mathbf{y}$, donde

$$\mathbf{y}=\begin{bmatrix} \lambda^3 \\ \lambda^2 \\ \lambda \\ 1 \end{bmatrix}=\begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix},$$

y tomemos las matrices M_3, M_2, M_1 de la tabla 6.6. Sucesivamente hallamos

$$M_1\mathbf{y}=\begin{bmatrix} -0{,}167 & 0 & 3{,}167 & 1{,}500 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}=\begin{bmatrix} -1{,}500 \\ 1 \\ -1 \\ 1 \end{bmatrix},$$

$$M_2M_1\mathbf{y}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & -4 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1{,}5 \\ 1 \\ -1 \\ 1 \end{bmatrix}=\begin{bmatrix} -1{,}5 \\ 2 \\ -1 \\ 1 \end{bmatrix};$$

$$\mathbf{x}^{(1)}=M_3M_2M_1\mathbf{y}=\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} -1{,}5 \\ 2 \\ -1 \\ 1 \end{bmatrix}=\begin{bmatrix} -1{,}5 \\ 2 \\ 0{,}5 \\ 1 \end{bmatrix}.$$

De un modo análogo se puede hallar los vectores propios también para los valores λ_2, λ_3, λ_4. ▲

§ 6.9. Determinación del primer valor propio de la matriz con ayuda del método de iteraciones

Con ayuda de los métodos iterativos se puede determinar el primer valor propio (es decir, el mayor en módulo) de la matriz A sin desarrollar el determinante característico.

Sea

$$det\,(A - \lambda E) = 0 \tag{1}$$

la ecuación característica; $\lambda_1, \lambda_2, \ldots, \lambda_n$, sus raíces que son *los valores propios* de la matriz $A = [a_{ij}]$ (donde $i = 1, 2, \ldots, n$; $j = 1, 2, \ldots, n$). Supongamos que

$$|\lambda_1| > |\lambda_2| \geqslant |\lambda_3| \geqslant \ldots \geqslant |\lambda_n|,$$

o sea, λ_1 es el mayor en módulo número propio.

Entonces para hallar el valor aproximado de la raíz λ_1 se utiliza el esquema siguiente:

1) se elige arbitrariamente el vector inicial $\mathbf{y}$;

2) se plantean las iteraciones sucesivas:

$$\begin{aligned}
&\mathbf{y}^{(1)} = A\mathbf{y},\\
&\mathbf{y}^{(2)} = A \cdot A\mathbf{y} = A^2\mathbf{y},\\
&\mathbf{y}^{(3)} = A \cdot A^2\mathbf{y} = A^3\mathbf{y},\\
&\ldots\ldots\ldots\ldots\ldots\\
&\mathbf{y}^{(m)} = A \cdot A^{m-1}\mathbf{y} = A^{(m)}\mathbf{y},\\
&\mathbf{y}^{(m+1)} = A \cdot A^{(m)}\mathbf{y} = A^{(m+1)}\,\mathbf{y};
\end{aligned}$$

3) se eligen de la sucesión los dos últimos valores $\mathbf{y}^m = A^m\mathbf{y}$ e $\mathbf{y}^{(m+1)} = A^{m+1}\mathbf{y}$; entonces

$$\lambda_1 = \lim_{m\to\infty} \frac{y_i^{(m+1)}}{y_i^{(m)}} \text{ o bien } \lambda_1 \approx \frac{y_i^{(m+1)}}{y_i^{(m)}}\,, \tag{2}$$

donde $y_i^{(m+1)}$ e $y_i^{(m)}$ son las coordenadas correspondientes de los vectores $\mathbf{y}^{(m+1)}$ e $\mathbf{y}^{(m)}$ $(i = 1, 2, \ldots, n)$.

Ahora bien, tomando un número de iteración bastante grande m, se puede con todo grado de precisión calcular la raíz λ_1, mayor en módulo, de la ecuación característica de la matriz. Para determinar esta raíz puede ser utilizada toda coordenada del vector $\mathbf{y}^{(m)}$, en particular, se puede tomar la media aritmética de las relaciones correspondientes para diferentes coordenadas.

Si el vector inicial $\mathbf{y}$ está elegido con poca fortuna la fórmula (2) puede no dar la raíz necesaria e, incluso, en general no tener sentido, es decir, el límite de la relación $y_i^{(m+1)}/y_i^{(m)}$ puede no existir. Lo último se advierte fácilmente por los valores «saltantes» de esta relación. En tales casos es necesario cambiar el vector inicial. En calidad de primer vector propio se puede tomar $\mathbf{y}^{(m+1)}$.

Ejemplo. Hallar el primer valor propio de la matriz

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

(con tres cifras decimales) y el vector propio correspondiente.

△ 1) Elegimos el vector inicial $\mathbf{y}=\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

2) Planteamos $m = 10$ iteraciones:

$$\mathbf{y}^{(1)} = A\mathbf{y}, \quad \mathbf{y}^{(2)} = A^2\mathbf{y}^{(1)}, \ \ldots, \ \mathbf{y}^{(10)} = A^{10}\mathbf{y}.$$

Colocamos los cálculos en la tabla 6.10.

Tabla 6.10

$\mathbf{y}$	$A\mathbf{y}$	$A^2\mathbf{y}$	$A^3\mathbf{y}$	$A^4\mathbf{y}$	$A^5\mathbf{y}$	$A^6\mathbf{y}$	$A^7\mathbf{y}$	$A^8\mathbf{y}$	$A^9\mathbf{y}$	$A^{10}\mathbf{y}$
1	4	17	69	274	1075	4189	16 260	62 973	243 569	941 370
1	5	18	67	253	964	3693	24 193	54 650	210 663	812 585
1	2	17	25	92	345	1309	5 002	19 195	73 845	284 508

3) Terminando las iteraciones en $\mathbf{y}^{(10)} = A^{10}\mathbf{y}$, tenemos para distintas coordenadas:

$$\lambda_1^{(1)} \approx \frac{y_1^{(10)}}{y_1^{(9)}} = \frac{941\,370}{243\,569} = 3{,}865; \quad \lambda_1^{(2)} \approx \frac{y_2^{(10)}}{y_2^{(9)}} = \frac{812\,585}{210\,663} = 3{,}857;$$

$$\lambda_1^{(3)} \approx \frac{y_3^{(10)}}{y_3^{(9)}} = \frac{284\,508}{73\,845} = 3{,}853.$$

4) Calculamos λ_1 como media aritmética $\lambda_1^{(1)}$, $\lambda_1^{(2)}$ y $\lambda_1^{(3)}$

$$\lambda_1 = \frac{\lambda_1^{(1)}+\lambda_1^{(2)}+\lambda_1^{(3)}}{3} = \frac{3{,}865+3{,}857+3{,}853}{3} = 3{,}858.$$

5) En calidad de primer vector propio de la matriz A se puede tomar el vector $\mathbf{y}^{(10)} = A^{10}\mathbf{y} = \begin{bmatrix} 941\ 370 \\ 812\ 585 \\ 284\ 508 \end{bmatrix}$. Normándolo, o sea, dividiendo todos sus coordenadas por la norma del vector, igual a

$$\| \mathbf{y}^{(10)} \|_3 = \sqrt{941\,370^2 + 812\,585^2 + 284\,508^2} = 1{,}28 \cdot 10^6,$$

obtenemos el primer vector propio de la matriz A el cual pertenece al primer valor propio $\lambda_1 = 3{,}858$:

$$\mathbf{x}^{(1)} \begin{bmatrix} 0{,}74 \\ 0{,}64 \\ 0{,}22 \end{bmatrix} \ \blacktriangle$$

§ 6.10. Determinación de los valores propios sucesivos y de los vectores propios que les pertenecen

Supongamos que los valores propios de la matriz A son tales que

$$|\lambda_1| > |\lambda_2| > |\lambda_3| \geqslant \ldots \geqslant |\lambda_n|. \tag{1}$$

Entonces para determinar λ_2 podemos utilizar las llamadas λ-*diferencias* haciendo uso el valor λ_1 que ya tenemos

$$\Delta_{\lambda_1} A^m \mathbf{y} = A^{m+1}\mathbf{y} \quad \lambda_1 A^m \mathbf{y}, \ \Delta_{\lambda_1} A^{m-1}\mathbf{y} = A^m \mathbf{y} \quad \lambda_1 A^{m-1}\mathbf{y}$$

o bien

$$\Delta_{\lambda_1}\mathbf{y}^{(m)} = \mathbf{y}^{(m+1)} - \lambda_1 \mathbf{y}^{(m)}, \quad \Delta_{\lambda_1}\mathbf{y}^{(m-1)} = \mathbf{y}^{(m)} - \lambda_1 \mathbf{y}^{(m-1)}, \tag{2}$$

de donde, pasando a las coordenadas de los vectores, obtenemos

$$\lambda_2 \approx \frac{\Delta_{\lambda_1} y_i^{(m)}}{\Delta_{\lambda_1} y_i^{(m-1)}} = \frac{y_i^{(m+1)} - \lambda_1 y_i^{(m)}}{y_i^{(m)} - \lambda_1 y_i^{(m-1)}} \quad (i = 1, 2, \ldots, n). \tag{3}$$

La fórmula (3) ofrece los valores muy aproximados para λ_2, ya que λ_1 también fue determinado aproximadamente.

Si los módulos de todos los valores propios son distintos entre sí, con ayuda de las fórmulas, análogas a la (3), se puede calcular asimismo los demás valores propios, pero los resultados sucesivos serán todavía menos exactos.

Al calcular λ_2 el número de iteración m ha de tomarse menor que al calcular λ_1 para evitar que se pierda la exactitud durante la sustracción de los números próximos.

Ejemplo. Para la matriz

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

hallar el segundo valor propio λ_2 y el vector propio $\mathbf{x}^{(2)}$ que le pertenece. Realizar los cálculos para $m = 8$ iteraciones con tres cifras decimales.

$\triangle$ Utilicemos la tabla de los valores $A^m\mathbf{y}$ para $m = 7, 8, 9$ (véase la tabla 6.10)

$A^7\mathbf{y}$	$A^8\mathbf{y}$	$A^9\mathbf{y}$
16 260	62 973	243 569
14 193	54 650	210 663
5 002	19 195	73 845

Con ayuda de la fórmula (2) planteamos las λ-diferencias:

$$\Delta_{\lambda_1}\mathbf{y}_i^{(m)} = \mathbf{y}_i^{(m+1)} - \lambda_1 \mathbf{y}_i^{(m)} \quad (i = 1, 2, 3).$$

Para cada una de las filas se toma el propio valor λ_1. $\lambda_1^{(1)} = 3{,}865$; $\lambda_1^{(2)} = 3{,}857$; $\lambda_1^{(3)} = 3{,}853$. Obtenemos la tabla siguiente:

Tabla 6.11

$A^8\mathbf{y}$	$\lambda_1 A^7\mathbf{y}$	$\Delta_1 A^7\mathbf{y}$	$A^9\mathbf{y}$	$\lambda_1 A^8\mathbf{y}$	$\Delta_{\mathbf{y}_1} A^8\mathbf{y}$
62 973	62 845	128	243 569	243 290	179
54 650	54 742	−92	210 663	210 785	−122
19 195	19 272	−77	73 845	73 958	−113

3) Con ayuda de la fórmula (3) para cada fila calculamos λ_2:

$$\lambda_2^{(1)} \approx \frac{\Delta_{\lambda_1} y_1^{(8)}}{\Delta_{\lambda_1}\mathbf{y}_1^{(7)}} = \frac{179}{128} = 1{,}400; \quad \lambda_2^{(2)} \approx \frac{\Delta_{\lambda_1} y_2^{(8)}}{\Delta_{\lambda_1} y_2^{(7)}} = \frac{-122}{-92} = 1{,}326;$$

$$\lambda_2^{(3)} \approx \frac{\Delta_{\lambda_1} y_3^{(8)}}{\Delta_{\lambda_1} y_3^{(7)}} = \frac{-113}{-77} = 1{,}468.$$

4) Determinamos λ_2 como media aritmética de $\lambda_2^{(1)}$, $\lambda_2^{(2)}$ y $\lambda_2^{(3)}$:

$$\lambda_2 = \frac{\lambda_2^{(1)} + \lambda_2^{(2)} + \lambda_2^{(3)}}{3} = \frac{1{,}400 + 1{,}326 + 1{,}468}{3} = 1{,}398.$$

5) En calidad de segundo vector propio tomamos

$$\mathbf{x}^{(2)} = \Delta_{\lambda_1} A^8\mathbf{y} = \begin{bmatrix} 179 \\ -122 \\ -113 \end{bmatrix}.$$

Normándolo, tenemos

$$\mathbf{x}^{(2)} = \begin{bmatrix} 0{,}73 \\ -0{,}50 \\ -0{,}46 \end{bmatrix}.$$

Para la matriz dada se puede hallar el tercer valor propio, conociendo la traza de la matriz: puesto que $\lambda_1 + \lambda_2 + \lambda_3 = \mathrm{Sp}\, A = 3 + 2 + 1 = 6$, entonces $\lambda_3 \approx 6 - 3{,}858 - 1{,}398 = 0{,}744$. ▲

Ejercicios

1. Utilizando el método de desarrollo inmediato, hallar los polinomios característicos para las matrices siguientes:

$$\text{a)}\ A=\begin{bmatrix}2 & 5 & 7\\ 6 & 3 & 4\\ 5 & -2 & -3\end{bmatrix};\quad \text{b)}\ A=\begin{bmatrix}2 & -1 & 3 & -4\\ 3 & -2 & 4 & -3\\ 5 & -3 & -2 & 1\\ 3 & -3 & -1 & 2\end{bmatrix};$$

$$\text{c)}\ A=\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & 1 & -1 & -1\\ 1 & -1 & 1 & -1\\ 1 & -1 & -1 & 1\end{bmatrix}.$$

2. Determinar los números característicos y vectores propios de las matrices

$$\text{a)}\ A=\begin{bmatrix}3 & 4\\ 5 & 2\end{bmatrix},\quad \text{b)}\ A=\begin{bmatrix}5 & 6 & -3\\ -1 & 0 & 1\\ 1 & 2 & -1\end{bmatrix}.$$

3. Con ayuda del método de Krylov desarrollar los determinantes característicos de las matrices siguientes:

$$\text{a)}\ A=\begin{bmatrix}1 & 2 & -4\\ 3 & -1 & 2\\ 1 & 0 & -1\end{bmatrix};\quad \text{b)}\ A=\begin{bmatrix}2 & -2 & 1 & 3\\ 4 & -1 & -2 & -3\\ 1 & 2 & -5 & 4\\ 1 & -1 & 2 & -4\end{bmatrix};$$

$$\text{c)}\ A=\begin{bmatrix}1 & -2 & 1 & -2\\ 2 & -1 & 2 & -1\\ 1 & 1 & -2 & -2\\ -2 & -2 & 1 & 1\end{bmatrix}.$$

4. Haciendo uso del método de Le Verrier — Faddéev, desarrollar los determinantes característicos para las matrices siguientes:

$$\text{a)}\ A=\begin{bmatrix}5 & -4 & 3 & 0\\ 2 & 1 & -2 & 1\\ 1 & 1 & 1 & -1\\ 2 & 1 & -2 & 0\end{bmatrix};\quad \text{b)}\ A=\begin{bmatrix}2 & -4 & 3 & -2\\ 1 & 0 & -5 & -3\\ 1 & 1 & -2 & -2\\ 0 & 1 & -1 & -1\end{bmatrix}.$$

5. Aplicando el método de Le Verrier — Faddéev, hallar las matrices inversas para las dadas en el ejercicio 4.

6. Desarrollar el determinante característico de la matriz

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

y calcular el vector propio con ayuda del método de Le Verrier — Faddéev.

7. Haciendo uso del método de iteraciones, calcular los valores propios primero y segundo de la matriz

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

CAPÍTULO VII

Interpolación y extrapolación

§ 7.1. Función y métodos de su representación

En la actividad práctica nos encontramos constantemente con la necesidad de revelar las formas de relación en los procesos y fenómenos y con la necesidad de describirlas matemáticamente.

Consideremos las formas de relación para las cuales cierta magnitud y que caracteriza el proceso depende de un conjunto de magnitudes no relacionadas entre sí $x_1, x_2, \ldots, x_n$ de un modo tal que a cada colección $(x_1, x_2, \ldots, x_n)$ corresponda el único valor de la magnitud y. Tal correspondencia unívoca de la magnitud y al conjunto de las variables independientes $x_1, x_2, \ldots, x_n$ se llama *dependencia funcional* y la misma variable y, *función* de las variables $x_1, x_2, \ldots, x_n$ lo que se escribe formalmente así:

$$y = f(x_1, x_2, \ldots, x_n).$$

Ahora bien, la expresión $y = x_1^2 + 3\sqrt{x_2} + x_1x_3^2$ es función de tres variables.

Si la magnitud y es función de una sola variable independiente x, esta relación puede ser representada en la forma siguiente:

$$y = f(x).$$

Por ejemplo, el área de un círculo S es función de la variable independiente, o sea, del radio de círculo R, es decir, $S = f(R)$; la forma concreta de esta función $S = \pi R^2$. El volumen de una figura es ya la función de tres dimensiones: $V = f(x_1, x_2, x_3)$ y según la forma de la figura esta relación funcional se concretiza respectivamente.

Del curso de análisis matemático se conocen tres métodos de representación de las dependencias funcionales: 1) analítico; 2) gráfico; 3) tabular.

El método más cómodo de representar la dependencia funcional $y = f(x)$ es el **analítico**, ya que muestra directamente las operaciones, y la sucesión de realizarlas, con la variable independiente x para que se obtenga el valor correspondiente de la magnitud y.

Así, por ejemplo, como resultado del tratamiento matemático se puede obtener la siguiente dependencia analítica de los créditos monetarios en la agricultura para los valores mercantiles y materiales en función de los gastos para el ganado bovino:

$$y = 51{,}0203 + 0{,}1059x,$$

donde y son los créditos para los valores mercantiles y materiales; x, los gastos para el ganado bovino.

Otro ejemplo de la dependencia analítica: en el movimiento uniformemente acelerado el espacio recorrido y el tiempo están ligados por la relación

$$s = vt + 0{,}5at^2.$$

La ventaja del método analítico de representación consiste en la posibilidad de obtener los valores de y para todo argumento fijo x con toda precisión.

Entre los inconvenientes de este método figura la necesidad de realizar toda la sucesión de los cálculos; además, el método analítico no está visualizado.

Las desventajas indicadas del método analítico se eliminan al utilizar el **método gráfico** de representación de la función $y = f(x)$.

Se llama *gráfico* de la función dada $y = f(x)$ el conjunto de los puntos del plano xOy cuyas coordenadas satisfacen la ecuación $y = f(x)$.

El método tabular de representación de las funciones está difundido en la técnica, la física, la economía, las ciencias naturales (y con más frecuencia aparece como resultado de un experimento).

Por ejemplo, supongamos que como resultado de una prueba está obtenida la resistencia óhmica R de una barra de cobre en función de la temperatura t° en forma de la tabla siguiente:

R	77,80	79,75	80,80	82,35	83,90	85,10
t°	25,0	30,1	36,0	40,0	45,1	50,0

En este experimento el valor de la resistencia óhmica de la barra de cobre cambia durante las oscilaciones de la temperatura y es una variable dependiente.

La ventaja del método tabular de representar una función consiste en que para cada valor de la variable independiente colocada en la tabla se puede hallar inmediatamente, sin mediciones o cálculos algunos, el valor correspondiente de la función.

El inconveniente del método tabular consiste en que no es posible representar continuamente toda la función, es decir, siempre hay tales valores de la variable independiente que faltan en la tabla.

§ 7.2. Tablas matemáticas

Entre las funciones que constantemente figuran en las matemáticas hay muchas tales cuyo cálculo cuesta bastante trabajo, a pesar de ser ellas sencillas. Tales funciones se someten a la tabulación, o sea, se representan en forma de las tablas matemáticas.

Tienen la más vasta aplicación las tablas de funciones de una variable. Entre ellas figuran las de los números inversos, de los cuadrados y cubos de los números, de las raíces cuadradas y cúbicas, las tablas de los logaritmos de las funciones trigonométricas, las tablas de la función exponencial y de otras funciones elementales. Hay tablas de las funciones de dos variables y de una cantidad mayor de estas últimas. Como ejemplo de la tabla de las funciones de dos variables puede servir la de los productos de dos números.

La tabla no es sino una colección de los valores de la función para una sucesión de los valores de los argumentos $x_0, x_1, x_2, \ldots, x_n$. Debe contener tal colección de los valores del argumento que para todos valores de éste, distintos de $x_0, x_1, x_2, \ldots, x_n$ se pueda obtener el valor de la función con el grado necesario de precisión.

Las características fundamentales de las tablas son: 1) denominación de las funciones cuyos valores ellas expresan; 2) volumen; 3) paso; 4) cantidad de cifras de la función a tabular; 5) cantidad de entradas.

En calidad de *denominación de la función* cuyos valores numéricos se dan en la tabla sirve la expresión analítica de esta función, por ejemplo, sen x, log x, e^x, etc.

El volumen de la tabla se predetermina por los valores inicial y final del argumento. Así, por ejemplo, el volumen de la tabla $y =$ $=$ sen x contiene los valores del argumento entre $0°0'$ y $90°$.

Casi para todas las funciones a tabular los valores del argumento en la tabla forman la progresión aritmética cuya diferencia h se llama *paso de la tabla*. Ahora bien,

$$h = x_i - x_{i-1} = \text{const} \quad (i = 1, 2, \ldots, n).$$

Entonces

$$x_i = x_0 + ih \quad (i = 0, 1, \ldots, n).$$

A título de ilustración consideremos un fragmento de la tabla de conversión de los radianes en grados y de los grados en radianes (tabla 7.1)

En las primeras dos columnas de la tabla mencionada como variable independiente sirve la medida en radianes, mientras que los grados se consideran su función. Lo mismo es válido también para las columnas 3 y 4. En calidad del paso de la tabla aquí se elige $h = 0{,}01$ radianes.

Comenzando con la columna 5 se examina la función inversa a la dada, donde como variable independiente se eligen grados (o minutos y la medida en radianes es, respectivamente, la función de los grados

Tabla 7.1

Radianes	Grados	Radianes	Grados	Grados	Radianes
(1)	(2)	(3)	(4)	(5)	(6)
0,20	11,459	0,70	40,107	20	0,34907
21	12,032	71	40,680	21	36652
22	12,605	72	41,253	22	38397
23	13,178	73	41,826	23	40143
24	13,751	74	42,399	24	41888

Continuación

Grados	Radianes	Minutos	Radianes	Minutos	Radianes
(7)	(8)	(9)	(10)	(11)	(12)
70	1,22175	20	0,00582	50	0,01454
71	23918	21	00621	51	01484
72	25662	22	00640	52	01513
73	27409	23	00669	53	02542
74	29151	24	00698	54	01571

(o de los minutos). El paso de esta parte de la tabla es igual a un grado (en las columnas 5 y 7) y a un minuto (en las columnas 9 y 11).

En las tablas de guía se utiliza asimismo un paso compuesto de dos niveles. Por la vertical en la columna se dan los valores del argumento con un paso relativamente grande h^*: $x_i = x_0 + ih^*$ $(i = 0, 1, \ldots, n)$ y los valores correspondientes de la función $y_i = f_i = f(x_i)$. Por la horizontal en la primera fila se colocan los valores del argumento con un paso más pequeño h que de ordinario es igual a una décima parte del paso grande: $h = 0{,}1h^*$. En la segunda fila y en las sucesivas se ponen los valores de la función para el argumento que es igual a la suma de los valores x_i que están en la misma fila y en la misma columna en la intersección de las cuales se escribe el valor de la función. Así en la tabla 7.2 se da un fragmento de la tabla de las raíces cúbicas. Del fragmento mencionado de la tabla no es difícil determinar el paso grande por la vertical, igual a $h^* = 1$, y el paso pequeño por la horizontal, igual a $h = 0{,}1$.

Generalmente el paso de la tabla se expresa por una inidad de cualquier orden (más raramente por dos o cinco unidades de un orden determinado). Por ejemplo, en las tablas de los cuadrados y los cubos, en las de las raíces cuadradas y cúbicas y en las de los logaritmos el paso grande $h^* = 1$, en las tablas de los logaritmos naturales y en las de los números inversos el paso grande h^* es igual a 0,1 (véase la tabla 7.2).

Tabla 7.2

	0	1	2	3	4	5	6	7	8	9
60	8,43433	43904	44369	48836	45303	45769	46235	46700	47165	47629
61	8,48093	48556	49018	49481	49942	50403	50954	51324	51784	52243
62	8,52702	53160	53618	54075	54542	54988	55444	55899	56354	56808
63	8,57262	57715	58168	58620	59062	59524	59975	60425	60875	61325
64	8,61774	62222	62670	63118	63566	64012	64459	64904	55350	65795

Tabla 7.3

	0′	6′	12′	18′	24′	30′	36′	42′	48′	54′			1′	2′	3′	4′	5′
65°	0,9063	9070	9078	9085	9092	9100	9107	9114	9121	9128	0,9335	24°	1	2	4	5	6
66°	0,9135	9143	9150	9157	9164	9171	9178	9184	9191	9198	0,9205	23°	1	2	3	5	6
67°	0,9205	9212	9219	9225	9232	9239	9245	9252	9259	9265	0,9272	22°	1	2	3	4	6
68°	0,9272	9278	9285	9291	9298	9304	9311	9317	9323	9330	0,9336	21°	1	2	3	4	5
69°	0,9336	9342	9348	9354	9361	9367	9383	9379	9385	9391	0,9397	20°	1	2	3	4	5

Si examinamos la tabla de los senos (tabla 7.3), veremos que en esta tabla en calidad de paso grande se elige un grado y el paso pequeño es igual a seis minutos.

La siguiente característica de las tablas es la *cantidad de cifras* de la función a tabular, ya que los valores de la función $y = f(x)$ para los valores tabulados del argumento en las tablas matemáticas y los resultados de mediciones en las tablas técnicas son magnitudes aproximadas.

Para los cálculos prácticos manuales las más aplicables son «Tablas matemáticas de cinco cifras decimales» de B. I. Segal y K. A. Semendiáev, «Guía de matemáticas» de I. N. Bronshtein y K. A. Semendiáev, etc.

En las tablas se colocan sólo las cifras justas del valor numérico de la función: esto quiere decir que el error no supera cinco unidades del primer orden suprimido. En este caso los valores de la función para todos los valores de x dados en la tabla se determinan con error absoluto igual. La exactitud con que se dan en la tabla los valores de la función se llama *exactitud de la tabla*. A veces en distintas partes de la tabla la exactitud puede ser diferente.

En algunos casos al trabajar con las tablas es necesario conocer las diferencias de los valores vecinos de la función dados en la tabla, o sea, $y_{i+1} - y_i$. Estas diferencias se llaman *diferencias finitas de primer orden* y se escriben, a veces, en la columna de la función en el intervalo entre los valores de esta última que participan en la formación de la diferencia finita correspondiente. Las diferencias se escriben en las unidades del último orden sin ceros delante de las cifras significativas y sin coma. Por ejemplo, en la tabla

x	sen x	
1,000	0,84147	
		54
1,001	0,84201	

la diferencia finita 0,00054 = 0,84201 — 0,84147 está escrita entre los valores respectivos de la función.

La siguiente característica importante de las tablas es la *cantidad de entradas* en la misma. Es igual al número de argumentos de la función. Así, las tablas para las dependencias funcionales $y = f(x)$ son tablas con una entrada. Entre ellas figuran las tablas 7.1, 7.2 y 7.3 anteriormente dadas.

La tabulación de la función de dos variables $z = f(x, y)$ conduce a la tabla con dos entradas. Entre semejantes tablas tienen amplia aplicación práctica las tablas de multiplicación.

En la tabla 7.4 el multiplicando de tres órdenes está escrito en la columna izquierda y el multiplicador de un orden, en la fila superior de la tabla. El paso de ambas entradas en la tabla es igual a la unidad. Para obtener el producto de un número de tres cifras por un número de una cifra basta encontrar la fila en cuya primera columna está puesto el multiplicando y elegir la columna en la cual está situado el multiplicador. En la intersección de la fila y columna halladas está precisamente el producto buscado.

Tabla 7.4

1	2	3	4	5	6	7	8	9
541	1082	1623	2164	2705	3246	3787	4328	4869
542	1084	1626	2168	2710	3252	3794	4336	4878
543	1085	1629	2172	2715	3258	3801	4344	4887
544	1088	1632	2176	2720	3260	3808	4352	4896
545	1090	1635	2180	2725	3270	3815	4360	4905

Ejemplo 1. Supongamos que se necesita multiplicar 543 por 8. En la tabla 7.4 encontramos la fila que contiene 543 y la columna cuyo número es 8. En su intersección leemos el número 4344 lo que es precisamente el producto buscado.

Para multiplicar números polidígitos el multiplicando se divide en partes que contienen tres cifras, como máximo, y a cada una de estas partes se aplica el método indicado.

Ejemplo 2. Supongamos que se necesita multiplicar 541 544 por 37. Dividimos el número 541 544 en dos partes de tres órdenes:

541 y 544. Sucesivamente multiplicamos cada parte por tres decenas y por 7 unidades y sumamos los productos parciales obtenidos:

$$\begin{array}{r} 541\times 30=16\,230 \\ 541\times 7=\ \ 3\,787 \\ \hline 20\,017 \end{array} \qquad \begin{array}{r} 544\times 30=16\,320 \\ 544\times\ \ 7=\ \ 3\,808 \\ \hline 20\,128 \end{array}$$

El primer producto parcial se pone en tres órdenes a la izquierda respecto al segundo y se realiza la sumación:

$$\begin{array}{r} 20\,017 \\ +\qquad 20\,128 \\ \hline 20\,037\,128 \end{array}$$

Éste es precisamente el resultado buscado.

§ 7.3. Conceptos principales de la teoría de aproximación de las funciones

La teoría y la práctica de *aproximación* se aplican al resolver muchos problemas prácticos.

Supongamos, por ejemplo, que en el proceso de cierto experimento en instantes discretos de tiempo $x_0, x_1, \ldots, x_N$ han sido obtenidos los valores $f_0, f_1, \ldots, f_N$ de cierta variable $f(x)$. Se necesita reconstruir la función $f(x)$ para otros $x \neq x_i$ $(i = 0, 1, \ldots, N)$. Un problema semejante surge al calcular muchas veces en el ordenador una misma función compuesta f en diferentes puntos. En vez de esto con frecuencia puede ser racional calcular la función f en un pequeño número de puntos característicos x_i y en los demás puntos calcular sus valores siguiendo una regla más sencilla al utilizar la información sobre los valores ya conocidos $f_i = f(x_i)$.

En calidad de otros ejemplos difundidos de aproximación de las funciones sirven los problemas de determinación de la derivada $f'(x)$ y de la integral $\int\limits_a^b f(x)\,dx$ según los valores dados f_i.

Por último, al componer los algoritmos de los programas estándares para el cálculo de las funciones elementales y especiales aparece una vez más el problema de aproximación de las funciones.

El enfoque clásico de resolución de semejantes problemas consiste en que, utilizando la información disponible sobre la función f, se considera otra función φ que es próxima, en cierto sentido, a f y permite realizar con ésta la operación correspondiente y obtener la estimación del error de tal «sustitución analítica».

En el proceso de realización numérica de este enfoque es necesario examinar las siguientes cuatro cuestiones fundamentales:

1. La cuestión acerca de la información disponible respecto a la función f, o sea, acerca de la forma en la cual está representada la función f.

2. La cuestión acerca de la clase de las funciones que aproximan, o sea, acerca de qué funciones φ será aproximada la función f.

3. La cuestión acerca de la proximidad de las funciones aproximable y aproximante, o sea, acerca de la elección del criterio de aceptación al cual debe satisfacer la función φ.

4. La cuestión acerca del error, o sea, acerca de la determinación de la diferencia entre los valores exacto y aproximado.

En la cuestión sobre la información respecto a la función f se distinguen dos casos principales: ora la función está representada analíticamente, ora en la forma de una tabla. El método gráfico de representar la función se refiere al primer caso o al segundo según el problema concreto. A continuación consideraremos sobre el segmento $[a, b]$ las funciones continuas $f(x)$, junto con una cantidad suficiente de sus derivadas, determinadas por sus valores $f_i = f(x_i)$ en los nodos x_i de la red dada

$$\Lambda_N: \{a \leqslant x_0 < x_1 < \ldots < x_N \leqslant b\}. \tag{1}$$

En la cuestión sobre la clase de las funciones aproximantes es necesario guiarse por dos factores principales. En primer lugar, la función aproximante debe reflejar las particularidades características de la aproximable y, en segundo lugar, debe ser bastante cómoda en el tratamiento, o sea, al realizar con ella las operaciones necesarias.

En el análisis numérico tienen gran aplicación tres grupos de las funciones aproximantes. El primero está formado por las funciones que tienen la forma $1, x, \ldots, x^n$ cuyas combinaciones lineales generan la clase de todos los polinomios de grado no superior a n. El segundo grupo está constituido por las funciones trigonométricas sen $a_i x$ y cos $a_i x$ que generan las series de Fourier y la integral de Fourier. Por último, el tercer grupo se compone de las funciones exponenciales $e^{a_i x}$ que determinan los fenómenos del tipo descomposición y acumulación los cuales se encuentran frecuentemente en situaciones reales.

A continuación examinaremos con más detalles la aproximación polinomial, o sea, tomaremos en calidad de la función aproximante el polinomio de cierto grado n.

En este caso la función aproximante suele designarse $P_n(x)$ y tiene la forma

$$\varphi(x) \equiv P_n(x) = \sum_{k=0}^{n} a_k x^k. \tag{2}$$

La cuestión sobre el criterio de aceptación consiste de hecho en determinar de cierto modo la «distancia» entre las funciones aproximable y aproximante y luego de toda la clase de las funciones aproximantes elegir aquélla para la cual esta «distancia» es mínima.

Uno de los criterios de aceptación difundidos es el de Chébyshev fundado en el concepto de distancia como magnitud máxima de desviación de la función φ respecto a la función f en los nodos x_i:

$$\rho_1 = \max_{0 \leqslant i \leqslant N} |f(x_i) - \varphi(x_i)|. \tag{3}$$

Presenta mayor interés el caso particular cuando para la función aproximante la distancia $\rho_1 = 0$. Esto quiere decir que para la función tabulada $y = f(x)$ representada por sus valores $y_i = f_i = f(x_i)$:

x_0	x_1	...	x_N
y_0	y_1	...	y_N

se necesita (construir la función aproximante $\varphi(x)$ que coincida en los nodos x_i con los valores de la función dada $y = f(x)$, o sea tal que $\varphi(x_i) = y_i$.

Tal método de aproximación basado en el criterio de coincidencia de f y φ en los nodos x_i se llama *interpolación*. Si el argumento x,

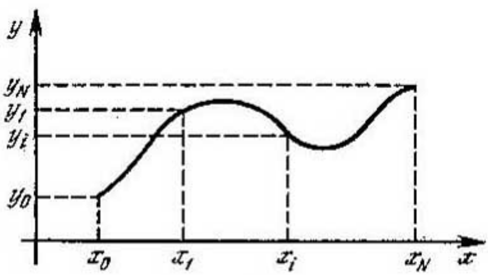


Fig. 7.1

para el cual se determina el valor aproximado de la función, pertenece al segmento $[x_0, x_N]$, el problema de determinación del valor de la función en el punto x se denomina *interpolación en sentido estrecho*. En cambio, si el argumento x está fuera del segmento $[x_0, x_N]$, el problema planteado se nombra *extrapolación*.

El problema de interpolación para la función de una variable $y = f(x)$ significa geométricamente la construcción de la curva que pasa por los puntos con coordenadas $(x_0; y_0)$, $(x_1; y_1)$, . . ., $(x_N; y_N)$ sobre el plano xOy (fig. 7.1).

Vamos a citar un ejemplo más de criterio de aceptación. Introduzcamos el concepto de distancia entre las funciones f y φ como suma de cuadrados de sus desviaciones en los puntos nodales:

$$\rho = \sum_{i=0}^{N} [f(x_i) - \varphi(x_i)]^2. \tag{6}$$

Elijamos ahora en calidad de función aproximante aquélla para la cual ρ es el mínimo. Es racional utilizar este criterio en caso de una gran cantidad de información dada con pequeña exactitud; el método de aproximación fundado en dicho criterio suele llamarse

método de cuadrados mínimos.. Entre las ventajas de este método conviene mencionar la sencillez y la armonía de su teoría matemática.

Por fin, la última cuestión — acerca de la exactitud de la solución que se obtiene — es, en muchos aspectos, la fundamental. En efecto en resumidas cuentas la calidad del método se determina, en primer lugar, por la rapidez con que se obtiene la solución con exactitud requerida o bien, como se dice de otra manera, por la velocidad de convergencia. Por eso es comprensible que la elección de los puntos nodales, de la clase de funciones aproximantes y del criterio de aceptación debe ser subordinada a la única cuestión, o sea, a la de exactitud requerida.

A primera vista la cuestión sobre la exactitud de la solución a obtener parece bastante simple: es necesario que la solución aproximada se distinga de la exacta en valor prefijado ε como máximo. Sin embargo, la cuestión acerca de la posibilidad de aproximar tan exactamente como se quiera la función f, cuestión que depende de los «parámetros» anteriormente citados (nodos x_i, clase de funciones φ, criterio de aceptación de f y φ), en el caso general, queda pendiente y se somete a la investigación para cada proceso de aproximación concreto.

§ 7.4. Interpolación con ayuda de los polinomios

Consideremos más detalladamente el problema de interpolación de la función f con ayuda de los polinomios algebraicos.

En este caso la función aproximante φ suele designarse $P_n(x)$ y tiene la forma

$$\varphi(x) \equiv P_n(x) = \sum_{k=0}^{n} a_k x^{n-k}. \tag{1}$$

La elección del valor concreto n se determina en gran parte por las propiedades de la función aproximable, por la exactitud requerida, así como por los nodos de interpolación. A continuación veremos que en la elección de la magnitud n ejerce influencia esencial también el proceso de cálculo que introduce en el resultado un error adicional.

En calidad de criterio de aceptación se toma, evidentemente, la condición de coincidencia de f y φ en los puntos nodales.

Es natural suponer que para determinar unívocamente $n+1$ coeficientes a_k del polinomio P_n es necesario exigir que coincidan f y P_n en el punto nodal $n+1$:

$$f(x_i) = P_n(x_i) \quad (i = 0, 1, \ldots, n). \tag{2}$$

El polinomio $P_n(x)$ que satisface las condiciones (2) se llama *polinomio interpolador*. Para subrayar que este polinomio depende de la función f, se designa frecuentemente $P_n(f, x)$.

En el caso en que es necesario calcular el valor de la función $f(x)$ en un punto x^*, por *error de interpolación* Δ_1 se entiende el valor absoluto de la diferencia entre los valores exacto y aproximado:

$$\Delta_1 = | f(x^*) - P_n(x^*) |. \tag{3}$$

En cambio, en el caso en que la interpolación se lleva a cabo sobre todo el segmento $[a, b]$, por error se toma la desviación máxima del polinomio P_n respecto a la función f en el segmento dado:

$$\Delta_1 = \max_{[a,b]} |f(x) - P_n(x)|.$$

Ahora bien, considermos el siguiente problema de interpolación En la red $\Lambda_n : a \leqslant x_0 < x_1 < \ldots < x_n \leqslant b$ en los nodos x_i se dan los valores $f_i = f(x_i)$ $(i = 0, 1, \ldots, n)$ de la función f. Se necesita construir el polinomio interpolador P_n que coincida con f en los nodos x_i y estimar el error Δ_1.

La existencia del polinomio interpolador y su unicidad se deducen del teorema siguiente.

Teorema. *Supongamos que:* 1°) *sobre el segmento* $[a, b]$ *se da la red* $\Lambda_n : a \leqslant x_0 \leqslant x_1 < \ldots < x_n \leqslant b$; 2°) *se dan los números arbitrarios* c_i $(i = 0, 1, \ldots, n)$. *Entonces existe el polinomio* P_n *de grado no superior a* n *que toma en los nodos* x_i *los valores dados de* c_i *y este polinomio es único.*

□ De las condiciones para determinar los coeficientes desconocidos a_k del polinomio P_n obtenemos el sistema de ecuaciones algebraicas

$$a_0 x_i^n + a_1 x_i^{n-1} + \ldots + a_n = c_i \ (i = 0, 1, \ldots, n). \tag{4}$$

El determinante de este sistema

$$W = \begin{vmatrix} x_0^n & x_0^{n-1} & \ldots & x_0 & 1 \\ x_1^n & x_1^{n-1} & \ldots & x_1 & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ x_n^n & x_n^{n-1} & \ldots & x_n & 1 \end{vmatrix} \tag{5}$$

es el determinante de Vandermonde que, como se sabe del álgebra, es distinto del cero si se cumple la condición $x_i \neq x_j$ para $i \neq j$. Esta condición se cumple, evidentemente, para la red en cuestión Λ_n. Por consiguiente, el sistema (4) tiene la única solución (la única colección de los coeficientes a_k). ■

Es evidente que si en calidad de números c_i se asignan los valores f_i de la función f en los nodos x_i, obtenemos la afirmación sobre la existencia del polinomio interpolador $P_n(f, x)$ y sobre su unicidad.

Los coeficientes a_k del polinomio interpolador (1) pueden ser determinados poniendo en el sistema (4) $c_i = f_i$ y resolviéndolo, por ejemplo, con ayuda de las fórmulas de Cramer

$$a_k = \Delta_k / W.$$

Aquí Δ_k es el determinante que se obtiene de W reemplazando la columna de los términos que contienen el grado $n - k$ de x_i $(i = 0,$

1, . . ., n) por la columna f_i de los términos independientes del sistema (4):

$$\Delta_k = \begin{vmatrix} x_0^n & \dots & x_0^{n-k+1} & f_0 & x_0^{n-k-1} & \dots & 1 \\ x_1^n & \dots & x_1^{n-k+1} & f_1 & x_1^{n-k-1} & \dots & 1 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ x_n^n & \dots & x_n^{n-k+1} & f_n & x_n^{n-k-1} & \dots & 1 \end{vmatrix}. \tag{7}$$

Sustituyendo los valores obtenidos de a_k en la igualdad (1), llegamos a una nueva forma de representación del polinomio interpolador $P_n(f, x)$:

$$\begin{vmatrix} P_n & 1 & x & \dots & x^n \\ f_0 & 1 & x_0 & \dots & x_0^n \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ f_n & 1 & x_n & \dots & x_n^n \end{vmatrix} = 0. \tag{8}$$

Nótese que en la práctica suelen utilizarse los polinomios interpoladores de primer grado y de segundo grado. En este caso se trata de *interpolación lineal y cuadrática.*

Ejemplo. Con ayuda de los nodos x_0, x_1, x_2 y los valores correspondientes de la función f_0, f_1, f_2 construir el polinomio interpolador representándolo en la forma de la combinación lineal de los valores f_i $(i = 0, 1, 2)$.

△ De acuerdo con la fórmula (8) tenemos

$$\begin{vmatrix} P_2 & 1 & x & x^2 \\ f_0 & 1 & x_0 & x_0^2 \\ f_1 & 1 & x_1 & x_1^2 \\ f_2 & 1 & x_2 & x_2^2 \end{vmatrix} = 0.$$

Desarrollando el determinante según los elementos de la primera columna, obtenemos

$$P_2 \begin{vmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{vmatrix} - f_0 \begin{vmatrix} 1 & x & x^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{vmatrix} + f_1 \begin{vmatrix} 1 & x & x^2 \\ 1 & x_0 & x_0^2 \\ 1 & x_2 & x_2^2 \end{vmatrix} - f_2 \begin{vmatrix} 1 & x & x^2 \\ 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \end{vmatrix} = 0.$$

Teniendo en cuenta que

$$\begin{vmatrix} 1 & x_i & x_i^2 \\ 1 & x_j & x_j^2 \\ 1 & x_k & x_k^2 \end{vmatrix} = (x_j - x_i)(x_k - x_i)(x_k - x_j),$$

finalmente hallamos

$$P_2(x) = f_0 \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} + f_1 \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} +$$
$$+ f_2 \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} \cdot \ ▲$$

§ 7.5. Error de los procesos de interpolación

Supongamos que la función f es aproximada por el polinimio interpolador, o sea,

$$f(x) = P_n(x) + R_n(x), \qquad (1)$$

donde $R_n(x)$ es el término residual de la fórmula de interpolación

$$f(x) \approx P_n(x). \qquad (2)$$

El término residual depende de muchos factores: de las propiedades de la función f, de los parámetros de interpolación y de la posición del punto de interpolación. Por eso el estudio de $R_n(x)$ es un problema difícil. En este caso es necesario, ante todo, responder a la pregunta: ¿qué es la medida numérica del error? Si el punto de interpolación x^* es fijo, por medida del error es natural tomar la magnitud $\Delta_1 = |R_n(x^*)|$. En cambio, si el punto x^* no está conocido de antemano y la interpolación se lleva a cado sobre el segmento $[a, b]$, es racional por medida del error tomar la magnitud

$$\Delta_1 = \max_{[a, b]} |R_n(x)|. \qquad (3)$$

Según el problema concreto pueden ser elegidas también otras medidas del eror.

Por regla general, la estimación de la medida del error no se realiza para una función tomada por separado sino para toda una clase de funciones que poseen ciertas propiedades comunes.

Deduzcamos la expresión explícita para estimar el error (3) de la fórmula de interpolación (2) para la clase de funciones $C^{n+1}(a, b)$ que tienen sobre el segmento $[a, b]$ una derivada continua de orden $n + 1$.

Con este fin vamos a demostrar el teorema siguiente.

Teorema. *Supongamos que*: 1°) *los nodos* x_i ($i = 0, 1, \ldots, n$) *son distintos y junto con* x^* *pertenecen al segmento* $[a, b]$; 2°) *la función* f *tiene sobre* $[a, b]$ *una derivada continua de orden* $n + 1$. *Entonces existe tal punto* $\xi \in (a, b)$ que

$$R_n(x^*) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^{n} (x^* - x_i). \qquad (4)$$

□ Nótese que si x^* coincide con uno de los nodos, la relación (4) se cumple, ya que su primer miembro y su segundo miembro son iguales a cero. Por eso suponemos luego que $x^* \neq x_i$ ($i = 0, 1, \ldots, n$). Examinemos la función auxiliar

$$\Psi(x) = f(x) - P_n(x) - k \prod_{i=0}^{n} (x^* - x_i),$$

donde k es la constante elegida de un modo tal que la función Ψ se anule para $x = x^*$, o sea,

$$\Psi(x^*) = 0 = f(x^*) - P_n(x^*) - k\prod_{i=0}^{n}(x^* - x_i).$$

De aquí

$$k = \frac{f(x^*) - P_n(x^*)}{\prod_{i=0}^{n}(x^* - x_i)} = \frac{R_n(x^*)}{\prod_{i=0}^{n}(x^* - x_i)}. \tag{6}$$

En virtud de tal elección de k la función Ψ sobre el segmento $[a, b]$ se anula $n + 2$ veces, como mínimo, en los puntos $x_0, x_1, \ldots, x_n, x^*$. Entonces, utilizando el teorema de Rolle, se puede afirmar que la derivada Ψ' sobre el intervalo (a, b) se anula, al menos, $n + 1$ veces y la derivada Ψ'' se anula n veces, por lo menos, etc. hasta la derivada $\Psi^{(n+1)}$ que se anulará en un punto, como mínimo. Supongamos que éste es el punto $\xi \in (a, b)$.

Ahora, derivando los miembros segundo y primero de la relación (5) $n + 1$ veces respecto a x y luego suponiendo $x = \xi$, en el primer miembro se obtiene el cero, ya que $\Psi^{(n+1)}(\xi) = 0$. El primer sumando del segundo miembro ofrece el valor de la derivada en el punto ξ: $f^{(n+1)}(\xi)$. El segundo sumando del segundo miembro da el cero como derivada de orden $n + 1$ del polinomio cuyo grado no supera n. El tercer sumando es el producto de la constante k por el polinomio de grado $n + 1$ con el coeficiente superior igual a 1; la derivada de orden $n + 1$ de este polinomio, como se sabe, es igual a $(n + 1)!$. Ahora bien, sumando todo lo dicho, tenemos

$$0 = f^{(n+1)}(\xi) - k(n + 1)!$$

Sustituyendo aquí en vez de k su expresión (6), obtenemos la relación requerida (4). ■

Supongamos ahora que, para precisar,

$$|f^{(n+1)}(x)| \leqslant M_{n+1}; \quad x \in [a, b]. \tag{7}$$

Utilizando esta limitación y el teorema que acabamos de demostrar, llegamos a la siguiente estimación del error para el punto fijo x^*:

$$\Delta_1 = |R_n(x^*)| \leqslant \frac{M_{n+1}}{(n+1)!}\prod_{i=0}^{n}|x^* - x_i|. \tag{8}$$

Ahora no es difícil construir para la red fija Λ_n la estimación $|R_n|$ uniforme en todo el segmento $[a, b]$. Así pues,

$$\Delta_1 = \max_{[a,b]}|R_n(x)| \leqslant \frac{M_{n+1}}{(n+1)!}\max_{[a,b]}|\omega_n(x)|, \tag{9}$$

donde $\omega_n(x) = \prod_{i=0}^{n}(x - x_i)$.

Ejemplo 1. Obtener sobre el segmento $[-1, 1]$ la estimación uniforme de la desviación de la función $f = 1 - \cos(\pi x/2)$ respecto a su polinomio interpolador construido por los nodos $x_i = -1 + \frac{2i}{n}$ $(i = 0, 1, \ldots, n;\ n = 2, 3, 4)$.

△ Nótese, ante todo, que para la función en cuestión sobre el segmento dado $M_{n+1} = (\pi/2)^{n+1}$. Por eso en virtud de la estimación (9) la resolución del problema se reduce a la estimación de la magnitud $\max_{[-1,\,1]} |\,\omega_n(x)\,|$ lo que se puede cumplir con ayuda de las reglas ordinarias de análisis matemático.

1. Consideremos el caso $n = 2$. Entonces

$$\omega_2(x) = (x+1)\,x\,(x-1); \quad \omega_2'(x) = 3x^2 - 1.$$

Las raíces del polinomio $\omega_2'(x)$ son $x_{1,2} = \pm 1/\sqrt{3} \approx \pm 0{,}5774$. Sustituyendo los valores obtenidos en ω_2, tenemos

$$\max_{[-1,\,1]} |\,\omega_2(x)\,| = |\,\omega_2(x_1)\,| = 2/\sqrt{27} \approx 0{,}3849$$

y, por lo tanto,

$$\Delta_1 \leqslant \left(\frac{\pi}{2}\right)^3 \cdot \frac{1}{3!} \cdot \frac{2}{\sqrt{27}} \approx 0{,}25.$$

2. Supongamos ahora que $n = 3$. En este caso

$$\omega_3(x) = (x+1)\left(x+\frac{1}{3}\right)\left(x-\frac{1}{3}\right)(x-1); \quad \omega_3'(x) = 4x^3 - \frac{20}{x}\,x.$$

De raíces del polinomio $\omega_3'(x)$ sirven $x_1 = 0$; $x_{2,\,3} = \pm\sqrt{5/3} \approx \approx \pm 0{,}7454$. Es fácil comprobar que el valor máximo $|\,\omega_3(x)\,|$ se alcanza en los puntos x_1, x_3:

$$\max_{[-1,\,1]} |\,\omega_3(x)\,| = |\,\omega_3(x_2)\,| = 16/81 \approx 0{,}1975.$$

Por eso

$$\Delta_1 \leqslant \left(\frac{\pi}{2}\right)^4 \cdot \frac{1}{4!} \cdot \frac{16}{81} \approx 0{,}05.$$

3. Para $n = 4$ reduzcamos la estimación

$$\omega_4(x) = (x+1)\left(x+\frac{1}{2}\right)x\left(x-\frac{1}{2}\right)(x-1)$$

a la obtenida en el subp. 1. Efectivamente, en virtud de la imparidad de $\omega_4(x)$ se puede limitarse por la determinación del valor máximo de $|\,\omega_4(x)\,|$ sobre el segmento $[0, 1]$. En este caso

$$\max_{[0,\,1]} |\omega_4(x)| < 2\cdot\frac{3}{2}\cdot\max_{[0,\,1]}\left|x\left(x-\frac{1}{2}\right)(x-1)\right|.$$

Reemplazando en el segundo miembro de la desigualdad obtenida la variable con ayuda de la fórmula $x = \frac{1}{2}(y+1)$ y tomando en cuenta los resultados del subp. 1, obtenemos

$$3\max_{[0,\,1]}\left|x\left(x-\frac{1}{2}\right)(x-1)\right| = \frac{3}{8}\max_{[-1,\,1]} |\,(y+1)\,y\,(y-1)\,| = \frac{1}{\sqrt{48}} \approx 0{,}1443.$$

Ahora bien,

$$\max_{[-1,\ 1]} |\omega_4(x)| < 1/\sqrt{48} \approx 0{,}1443$$

y la estimación buscada

$$\Delta_1 < \left(\frac{\pi}{2}\right)^5 \cdot \frac{1}{5!} \cdot \frac{1}{\sqrt{48}} \approx 0{,}012. \ \blacktriangle$$

Parece que el ejemplo citado afirma la suposición acerca de que la estimación (9) es prácticamente conveniente y toma pequeños valores para la mayoría de las funciones al ser n suficientemente grandes. Sin embargo, en muchos casos esto no tiene lugar.

La cosa consiste en que sólo para una clase restringida de funciones (por ejemplo, para las funciones enteras) las derivadas de bastante alto orden son pequeñas. En cambio, para la mayoría de las funciones algunas de las derivadas de orden superior tienen la tendencia a crecer como $n!$. A título de ejemplo consideremos la función $y = \ln x$. Es evidente que para ella $y^{(n)} = (-1)^{n-1} \frac{(n-1)!}{x^n}$. Ahora bien, incluso cerca de los puntos donde la curva $y = \lg x$ parece suave, sus derivadas de órdenes bastante altos llegan a ser muy grandes y se comportan como $n!$

El inconveniente de la aproximación polinomial consiste en que, por lo general, falta el sentido físico el cual conduce de ordinario a generalizaciones útiles.

Por otro lado, la sencillez de la teoría de la aproximación polinomial y su desarrollo profundo en combinación con el mínimo de cálculos contribuyen a que este tipo de aproximación sea un instrumento cómodo al resolver distintos problemas, aún más porque, como muestra la experiencia, en los cálculos prácticos se obtienen buenos resultados en caso de la aproximación por polinomios, aunque el término residual ora, en general, es difícil de estimar, ora su estimación resulta demasiado exagerada.

Así pues, hemos examinado sólo un aspecto de la cuestión sobre el error, o sea, la influencia ejercida por las propiedadas de la función f en la magnitud Δ_1. La cuestión sobre el error en dependencia de la situación de los nodos de la red está estrechamente vinculada con las propiedades del polinomio de Chébyshev, y por eso conviene volver a estudiar con detalles este problema después de considerar estos polinomios. Aquí nos limitamos por una de las posibles estimaciones de la magnitud $|\omega_n(x)|$ sobre la red fija Λ_n. Sea que x esté entre x_k y x_{k+1}. Pongamos $\max\limits_{1 \leqslant i \leqslant n} (x_i - x_{i-1}) = h$, entonces

$$|\omega_n(x)| = \prod_{i=0}^{n} |x - x_i| < (k+1)!\,(n-k)!\,h^{n+1} \leqslant n!\,h^{n+1}. \qquad (10)$$

Por eso la desigualdad (8) puede escribirse en la forma

$$\Delta_1 = \max_{[a;\ b]} |R_n(x)| < \frac{M_{n+1}}{n+1} h^{n+1}.$$

Nótese que la estimación (10) es bastante aproximada y no es difícil mejorarla (proponemos que el lector lo haga de por sí en calidad del ejercicio).

Ejemplo 2. ¿Con qué exactitud se puede calcular $\sqrt{117}$ con ayuda del polinomio interpolador para la función $y = \sqrt{x}$, eligiendo como nodos de interpolación $x_0 = 100$, $x_1 = 121$, $x_2 = 144$?

△ Ante todo, determinamos $M_3 = \max\limits_{[100;\, 144]} |(\sqrt{x})'''|$. Para esto hallamos

$$y' = \frac{1}{2} x^{-1/2};\ y'' = -\frac{1}{4} x^{-3/2};\ y''' = \frac{3}{8} x^{-5/2}.$$

De aquí $M_3 = \frac{3}{8} \cdot 100^{-5/2} = \frac{3}{8} \cdot 10^{-5}$. Por eso en virtud de la relación (8) tenemos

$$\Delta_1 \leqslant \frac{3}{8} \cdot 10^{-5} \cdot \frac{1}{3!} |(117-100)(117-121)(117-144)| \approx$$

$$\approx 0{,}12 \cdot 10^{-2}. \quad ▲$$

§ 7.6. Polinomio interpolador de Lagrange

La determinación inmediata de los coeficientes a_k del polinomio interpolador está ligada con ciertas dificultades de cálculo. Por eso al resolver los problemas prácticos se utilizan tipos especiales del polinomio interpolador.

En este párrafo vamos a examinar la forma del polinomio interpolador que se llama forma de Lagrange y suele designarse $L_n(x)$. Para construir L_n examinemos primero los polinomios auxiliares $l_i(x)$ de grado n que poseen las dos propiedades siguientes:

$$l_i(x_i) = 1 \quad (i = 0, 1, \ldots, n); \tag{1}$$

$$l_i(x_k) = 0 \quad (i \neq k;\ \mathrm{i, k} = 0, 1, \ldots, n). \tag{2}$$

Estas propiedades significan que, por ejemplo, el polinomio $l_0(x)$ toma en el punto x_0 el valor igual a la unidad y en los demás nodos se anula; el polinomio $l_1(x)$ toma en el nodo x_1 el valor igual a la unidad y en los demás puntos se anula, etc. En el caso general el polinomio $l_i(x)$ en el nodo x_i toma el valor igual a la unidad y en los demás nodos se anula. Ahora bien, en virtud de la propiedad (2) y del requisito de que el polinomio $l_i(x)$ tenga el grado n, obtenemos

$$l_i(x) = c_i(x - x_0) \ldots (x - x_{i-1})(x - x_{i+1}) \ldots (x - x_n). \tag{3}$$

Luego, utilizando la propiedad (1), para determinar la constante c_i tenemos la ecuación

$$l_i(x_i) = c_i(x_i - x_0) \ldots (x_i - x_{i-1})(x_i - x_{i+1}) \ldots$$
$$\ldots (x_i - x_n) = 1.$$

De aquí

$$c_i = \frac{1}{(x_i - x_0) \ldots (x_i - x_{i-1})(x_i - x_{i+1}) \ldots (x_i - x_n)}. \tag{4}$$

Por eso la expresión explícita para $l_i(x)$ se puede representar del modo siguiente:

$$l_i(x)=\frac{(x-x_0)\ \dots\ (x-x_{i-1})(x-x_{i+1})\ \dots\ (x-x_n)}{(x_i-x_0)\ \dots\ (x_i-x_{i-1})(x_i-x_{i+1})\ \dots\ (x_i-x_n)}. \qquad (5)$$

Hagamos ahora la siguiente combinación lineal de los polinomios l_i:

$$L_n(x)=\sum_{i=0}^{n} f_i l_i(x). \qquad (6)$$

La expresión (6) es el polinomio de grado no superior a n. En el nodo x_i este polinomio toma el valor f_i, ya que el sumando correspondiente de la suma $f_i l_i(x_i)$ es igual a f_i y los demás sumandos $f_j l_j(x_i)$ son iguales al cero. Ahora bien, está construido el polinomio interpolador para la función $f(x)$. Esta forma precisamente se denomina *polinomio interpolador de Lagrange.*

Teniendo en cuenta que

$$\omega_n(x)=(x-x_0)(x-x_1)\dots(x-x_n),$$

se puede considerar su derivada en el punto x_i:

$$\omega_n'(x_i)=(x_i-x_0)\dots(x_i-x_{i-1})(x_i-x_{i+1})\dots(x_i-x_n)$$

y escribir el polinomio de Lagrange en la forma

$$L_n(x)=\sum_{i=0}^{n} f_i\,\frac{\omega_n(x)}{(x-x_i)\,\omega_n'(x_i)}. \qquad (7)$$

Las magnitudes $l_i(x)$ son, en cierto modo, polinomios ponderales de los nodos correspondientes y suelen llamarse *multiplicadores de Lagrange.* En adición a las propiedades (1) y (2) citemos una propiedad importante más de estos multiplicadores:

$$\sum_{i=0}^{n} l_i(x)=1. \qquad (8)$$

En efecto, sea $f(x)\equiv 1$, entonces todas las $f_i=1$ $(i=0, 1, \dots, n)$. Por otro lado, $f^{(n+1)}(x)\equiv 0$ y en virtud del teorema del § 7.5 $L_n(x)=$ $=f(x)=1$. Sustituyendo lo obtenido en la expresión (6), llegamos a la igualdad (8).

Ejemplo 1. Haciendo uso de los nodos $x_0=0$, $x_1=1/3$, $x_2=1$, construir el polinomio interpolador de Lagrange para la función $f=\operatorname{sen}(\pi x/2)$ y obtener una estimación uniforme del error sobre el segmento $[0, 1]$.

△ Nótese, ante todo, que $f(x_0)=0$, $f(x_1)=1/2$, $f(x_2)=1$. Luego, utilizando la expresión (7) para $n=2$, construimos el polinomio interpolador deseado:

$$L_2(x)=\frac{1}{2}\,\frac{x(x-1)}{\frac{1}{3}\left(\frac{1}{3}-1\right)}+1\cdot\frac{x\left(x-\frac{1}{3}\right)}{1\left(1-\frac{1}{3}\right)}.$$

La estimación de error se obtiene fácilmente de la relación (8) del § 7.5 para $n = 2$:

$$\Delta_1 \leqslant \frac{M_3}{3!} \max_{[0,\,1]} \left| x \left(x - \frac{1}{3} \right) (x - 1) \right|.$$

En el caso dado es evidente que $M_3\ (\pi/2)^3$ y $\max\limits_{[0,\,1]} \ \left| x\left(x - \frac{1}{3}\right) \times \right.$ $\times\ (x - 1) \mid = 0{,}079$ y se determina de un modo análogo al hecho en el ejemplo 1 del § 7.5. Por eso tenemos finalmente

$$\Delta_1 \leqslant \left(\frac{\pi}{2} \right)^3 \cdot \frac{1}{3!} \cdot 0{,}079 \approx 0{,}05. \ \blacktriangle$$

Ejemplo 2. La función $f(x)$ se da en la forma tabular

x	0	1	2	6
y	-1	-3	3	1187

Valiéndose del polinomio interpolador de Lagrange, determinar el valor de esta función en el punto $x = 4$.

$\triangle$ Sustituyendo en la fórmula (7) los valores x_i y f_i para $n = 3$ y $x = 4$, obtenemos

$$L_3(4) = 1 \cdot \frac{(4-1)(4-2)(4-6)}{(-1)(-2)(-6)} - 3 \cdot \frac{4(4-2)(4-6)}{1(1-2)(1-6)} +$$
$$+ 3 \cdot \frac{4(4-1)(4-6)}{2(2-1)(2-6)} + 1187 \cdot \frac{4(4-1)(4-2)}{6(6-1)(6-2)} = 255. \ \blacktriangle$$

Si en el ejemplo considerado añadimos a la tabla un punto más, tendremos que volver a realizar los valores de la función para $x = 4$. Además, del mismo ejemplo se ve que el proceso de obtención del valor aproximado de la función ayuda de la fórmula de interpolación de Lagrange está vinculado con grandes cálculos. Esto origina la necesidad de simplificar el trabajo de cálculos.

Para la comodidad de los cálculos hagamos una tabla auxiliar

$x - x_0$	$x_0 - x_1$	$x_0 - x_2$	$\ldots$	$x_0 - x_n$	k_0
$x_1 - x_0$	$x - x_1$	$x_1 - x_2$	$\ldots$	$x_1 - x_n$	k_1
$x_2 - x_0$	$x_2 - x_1$	$x - x_2$	$\ldots$	$x_2 - x_n$	k_2
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$
$x_n - x_0$	$x_n - x_1$	$x_n - x_2$	$\ldots$	$x - x_n$	k_n

donde $x_0, x_1, \ldots, x_n$ son los nodos de interpolación y x es el valor del argumento para el cual se determina el valor aproximado con ayuda de la fórmula de interpolación de Lagrange. Designemos con k_0 el producto de los elementos de la primera fila

$$k_0 = (x - x_0)(x_0 - x_1) \ldots (x_0 - x_n).$$

En la forma general el producto de los elementos de la i-ésima fila es

$$k_i = (x_i - x_0) \ldots (x_i - x_{i-1})(x - x_i)(x_i - x_{i+1}) \ldots (x_i - x_n).$$

Coloquemos los números $k_0, k_1, \ldots, k_n$ en la columna derecha extrema de la tabla. Adicionalmente calculemos el producto de los elementos situados en la diagonal principal:

$$\omega_n(x) = (x - x_0)(x - x_1) \ldots (x - x_n).$$

Entonces el polinomio interpolador de Lagrange puede ser escrito en la forma

$$L_n(x) = \omega_n(x) \sum_{i=0}^{n} \frac{y_i}{k_i}. \tag{9}$$

Haciendo uso de la fórmula (9), volvamos a resolver el ejemplo 2. Hacemos la tabla

4	−1	−2	−6	−48
1	4−1	1−2	1−6	15
2	2−1	4−2	2−6	−16
6	6−1	6−2	4−6	−240

y hallamos $\omega_3(4) = -48$. Determinamos el valor aproximado de la función en el punto $x = 4$, o sea, $f(4) \approx L_3(4)$, haciendo uso de la fórmula

$$L_3(x) = \omega_3(x) \sum_{i=0}^{3} \frac{y_i}{k_i},$$

o bien

$$L_3(4) = -48\left[\frac{-11}{-48} + \frac{-3}{15} + \frac{3}{-16} + \frac{1187}{-240}\right] = 255. \quad \blacktriangle$$

La fórmula de interpolación de Lagrange se simplifica considerablemente si los nodos de interpolación son *equidistantes*, o sea, $h = x_{i+1} - x_i = \text{const}$, donde h es el paso de interpolación. Introduzcamos la designación $t = (x - x_0)/h$. Según la fórmula (5) tenemos

$$l_i = (x) = \frac{(x - x_0) \ldots (x - x_{i-1})(x - x_{i+1}) \ldots (x - x_n)}{(x_i - x_0) \ldots (x_i - x_{i-1})(x_i - x_{i+1}) \ldots (x_i - x_n)}.$$

Puesto que

$$x-x_0=th,$$
$$x-x_1=th-h=h(t-1).$$
$$\cdots\cdots\cdots\cdots$$
$$x-x_i=th-ih=h(t-i),$$
$$\cdots\cdots\cdots\cdots$$
$$x-x_n=th-nh=h(t-n),$$

entonces

$$l_i=\frac{t(t-1)\ldots(t-i+1)(t-i-1)\ldots(t-n)h^n}{ih(i-1)h\ldots 1h(-1)h\ldots[-(n-i)h]}. \tag{10}$$

Nótese que una parte del producto en el denominador es igual a

$$ih(i-1)h\ldots h=i!h^i$$

y la otra parte es igual a

$$(-h)\ldots[-(n-i)h]=(-1)^{n-i}(n-i)!h^{n-i}.$$

Multiplicando el numerador y el denominador del segundo miembro de la igualdad (10) por $(-1)^{n-i}(t-i)$, obtenemos

$$l_i=\frac{t(t-1)\ldots(t-n)}{(t-i)i!(n-i)!}(-1)^{n-i}=$$
$$=(-1)^{n-i}\frac{C_n^i}{t-i}\,\frac{t(t-1)\ldots(t-n)}{n!},$$

donde $C_n^i=\dfrac{n!}{i!(n-i)!}$.

Entonces el polinomio interpolador de Lagrange para los nodos equidistantes de interpolación se puede escribir en la forma

$$L_n(x)=L_n(x_0+ht)=\frac{t(t-1)\ldots(t-n)}{n!}\sum_{i=0}^{n}(-1)^{n-i}\frac{C_n^i}{t-i}y_i. \tag{11}$$

Ejemplo 3. La función $y=\operatorname{sen}x$ se da en la forma tabular

x	0	$\pi/4$	$\pi/2$
y	0	0,707	1

Haciendo uso del polinomio interpolador de Lagrange, determinar el valor de esta función en el punto $x^*=\pi/6$. Estimar el error Δ_1.

$\triangle$ Ante todo, determinamos $t^*=\left(\frac{\pi}{6}-0\right)\left(\frac{\pi}{4}\right)^{-1}=\frac{2}{3}$. Sustituyendo en la fórmula (11) el valor obtenido de t^* y los valores y_i para $n=2$, tenemos

$$L_2\left(\frac{\pi}{6}\right)=\frac{(2/3)(2/3-1)(2/3-2)}{2!}\times$$
$$\times\left(\frac{2}{2/3-0}\cdot 0-\frac{2}{2/3-1}\cdot 0{,}707+\frac{1}{2/3-2}\cdot 1\right)=0{,}517.$$

Para estimar el error hacemos uso de la fórmula (8) del § 7.5. Aquí $M_3 = \max\limits_{[0,\ \pi/2]} |(\text{sen } x)'''| = 1$, $x^* = \pi/6$; por eso

$$\Delta_1 = \frac{1}{3!} \cdot \frac{\pi}{6} \cdot \frac{\pi}{12} \cdot \frac{\pi}{3} = 0{,}024.$$

Nótese que al calcular el error covniene convertir la medida en grados en medida en radianes.

Ahora bien, redondeando el resultado hasta dos cifras obtenemos sen $(\pi/6) = 0{,}52 \pm 0{,}03$. ▲

§ 7.7. Diferencias finitas

En los párrafos precedentes hemos considerado diferentes cuestiones de la teoría de interpolación sobre la red arbitraria de los nodos. En la práctica la información disponible sobre la función (en forma de sus valores) se da con frecuencia sobre una red uniforme, o sea, para los nodos equidistantes. En este caso no sólo se simplifican las formas de los polinomios interpoladores sino se reduce también el mismo proceso de cálculo lo que es un hecho no menos sustancial desde el punto de vista práctico. Al construir los polinomios interpoladores sobre un red uniforme se utilizan las magnitudes llamadas *diferencias finitas.*

Examinemos una red uniforme con paso h: $x_i = x_0 + ih$ $(i = 0, \pm 1, \pm 2, \ldots)$ en cuyos nodos se dan los valores $f_i = f(x_i)$ de la función $f(x)$.

En las obras matemáticas se utilizan tres tipos de diferencias finitas: *diferencias descendentes* $\Delta^k f_i$ para la interpolación hacia adelante; *diferencias ascendentes* $\nabla^k f_i$ para la interpolación hacia atrás; *diferencias centrales* $\delta^k f_i = f_i^k$ para construir las fórmulas de interpolación centrales.

Se llama *diferencia finita de primer orden* la dirercencia entre los valores de la función en el nodo dado y en el nodo precedente:

$$\begin{aligned} f_1 - f_0 &= \Delta f_0 = \nabla f_1 = \delta f_{1/2} = f^1_{1/2}, \\ f_2 - f_1 &= \Delta f_1 = \nabla f_2 = \delta f_{3/2} = f^1_{3/2}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ f_{i+1} - f_i &= \Delta f_i = \nabla f_{i+1} = \delta f_{i+1/2} = f^1_{i+1/2}. \end{aligned} \tag{1}$$

Esta definición puede escribirse en otra forma:

$$\Delta f_i = f_{i+1} - f_i; \ \nabla f_i = f_i - f_{i-1}; \quad \delta f_i = f_{i+1/2} - f_{i-1/2} \tag{2}$$

Se denomina *diferencia finita de segundo orden* la diferencia entre los valores de la diferencia finita primera en el nodo dado y en el precedente

$$\begin{aligned} \Delta^2 f_i = \Delta f_{i+1} - \Delta f_i; \quad \nabla^2 f_i = \nabla f_i - \nabla f_{i-1}; \\ \delta^2 f_i = f_i^2 = \delta f_{i+1/2} - \delta f_{i-1/2}. \end{aligned} \tag{3}$$

De un modo análogo se definen las diferencias finitas de orden arbitrario k;

$$\Delta^k f_i = \Delta^{k-1} F_{i+1} - \Delta^{k-1} f_i; \quad \nabla^k f_i = \nabla^{k-1} f_i - \nabla^{k-1} f_{i-1}; \quad (4)$$
$$\delta^k f_i = f_i^k = \delta^{k-1} f_{i+1/2} - \delta^{k-1} f_{i-1/2}.$$

En algunas fórmulas de interpolación que se examinan a continuación junto con diferoncias (4) se utilizan las medias aritméticas de las

Tabla 7.5

x	f	Δf	$\Delta^2 f$	$\Delta^3 f$	$\Delta^4 f$
x_0	f_0	Δf_0	$\Delta^2 f_0$	$\Delta^3 f_0$	$\Delta^4 f_0$
x_0+h	f_1	Δf_1	$\Delta^2 f_1$	$\Delta^3 f_1$	...
x_0+2h	f_2	Δf_2	$\Delta^2 f_2$	...	
x_0+3h	f_3	Δf_3	...		
x_0+4h	f_4	...			
...	...	...	...	...	...

Tabla 7.6

x	f	∇f	$\nabla^2 f$	$\nabla^3 f$	$\nabla^4 f$
...	...	...	...	...	...
x_0-4h	f_{-4}	...			
x_0-3h	f_{-3}	∇f_{-3}	...		
x_0-2h	f_{-2}	∇f_{-2}	$\nabla^2 f_{-2}$	...	
x_0-h	f_{-1}	∇f_{-1}	$\nabla^2 f_{-1}$	$\nabla^3 f_{-1}$	...
x_0	f_0	∇f_0	$\nabla^2 f_0$	$\Delta^3 f_0$	$\nabla^4 f_0$

diferencias finitas vecinas del mismo orden:

$$\mu f_i^k = \frac{1}{2}\left(f_{i+1/2}^k + f_{i-1/2}^k\right),\quad \mu f_{i+1/2}^k = \frac{1}{2}\left(f_{i+1}^k + f_i^k\right). \tag{5}$$

La primera de estas magnitudes se utiliza cuando k es impar y la segunda, cuando k es par.

Es cómodo escribir las diferencias finitas de la función f en la forma de tablas. En este caso las diferencias finitas ascendentes y descendentes se escriben en la forma de las tablas horizontales (tablas 7.5 y 7.6) y las diferencias finitas centrales, en la forma de las tablas centrales (tabla 7.7).

Tabla 7.7

x	f	δf	$\delta^2 f$	$\delta^3 f$	$\delta^4 f$
...	...	...	...	...	...
x_0-2h	f_{-2}				
		$\delta f_{-3/2}$			
x_0-h	f_{-1}		$\delta^2 f_{-1}$		
		$\delta f_{-1/2}$		$\delta^3 f_{-1/2}$	
x_0	f_0		$\delta^2 f_0$		$\delta^4 f_0$
		$\delta f_{1/2}$		$\delta^3 f_{1/2}$	
x_0+h	f_1		$\delta^2 f_1$		
		$\delta f_{3/2}$			
x_0+2h	f_2				
...	...	...	...	...	...

Consideremos algunas propiedades de las diferencias finitas.

1°. *Las diferencias descendentes, ascendentes y centrales están ligadas entre sí por las relaciones siguientes*:

$$\Delta^k f_i = \nabla^k f_{i+k} = \delta^k f_{i+k/2} \tag{6}$$

lo que se demuestra fácilmente con ayuda del método de inducción matemática partiendo de la definición de las diferencias finitas (4).

☐ Para $k = 1$ las relaciones (6) son evidentes, ya que en virtud de las igualadades (4) tenemos

$$\Delta f_i = f_{i+1} - f_i;\ \nabla f_{i+1} = f_{i+1} - f_i;\ \delta f_{i+1/2} = f_{i+1} - f_i.$$

Supongamos ahora que las relaciones (6) son justas para todo $k \leqslant m - 1$. Mostremos que en este caso serán justas también para $k = m$ y, por consiguiente, para todos los k. Utilizando las igualdades (4) y la suposición sobre la validez de (6) para $k \leqslant m - 1$, tenemos

$$\Delta^m f_i = \Delta^{m-1} f_{i+1} - \Delta^{m-1} f_i = \nabla^{m-1} f_{i+m} - \nabla^{m-1} f_{i+m-1} = \nabla^m f_{i+m},$$

$$\Delta m f_i = \Delta^{m-1} f_{i+1} - \Delta^{m-1} f_i = \delta^{m-1} f_{i+\frac{m+1}{2}} - \delta^{m-1} f_{i+\frac{m-1}{2}} = \delta^m f_{i+\frac{m}{2}}. \ \blacksquare$$

2°. *La diferencia finita satisface la igualdad*

$$\Delta (af + bg)_i = a\Delta f_i + b\Delta g_i, \tag{7}$$

donde a y b son constantes.

☐ En efecto,

$$\Delta (af + bg)_i = af_{i+1} + bg_{i+1} - (af_i + bg_i) = a\Delta f_i + b\Delta g_i. \ \blacksquare$$

Esta propiedad significa, en particular, que la diferencia finita de la suma de dos funciones o de la diferencia de las mismas es igual a la suma o diferencia de las diferencias finitas de estas funciones, respectivamente, así como que la diferencia finita del producto de la función por el multiplicador constante es igual al producto de este multiplicador por la diferencia finita de la función.

3°. *La diferencia finita está ligada con la derivada correspondiente por la relación*

$$\Delta^k f_i = h^k f^{(k)}(\xi);\quad \xi \in (x_i;\ x_i + kh). \tag{8}$$

Como consecuencia de la igualdad (8) obtenemos que las diferencias finitas de orden n del polinomio de grado n son constantes e iguales a $h^n n! a_0$ y las diferencias finitas de todo orden más alto son iguales a cero (a_0 es el coeficiente del polinomio para el grado superior de x).

4°. *La diferencia finita de orden k puede ser representada en forma de la siguiente combinación lineal de valores f_i:*

$$\Delta^k f_i = \sum_{j=0}^{k} (-1)^j C_k^j f_{i+k-j}, \tag{9}$$

donde $C_k^j = \frac{k!}{j!(k-j)!}$ es el número de combinaciones de k elementos dispuestos j a j (con la particularidad de que $0! = 1$).

☐ Hagamos uso del método de inducción matemática. Para $k = 1$ esta relación es evidente, puesto que no es sino la definición de la diferencia finita primera: $\Delta f_i = f_{i+1} - f_i$.

Supongamos ahora que la igualdad (9) es justa para cierto $k = m$. Entonces

$$\Delta^{m+1}f_i = \Delta^m f_{i+1} - \Delta^m f_i =$$

$$= \sum_{j=0}^{m} (-1)^j C_m^j f_{i+1+m-j} - \sum_{j=0}^{m} (-1)^j C_m^j f_{i+m-j} =$$

$$= C_m^0 f_{i+1+m} + \sum_{j=1}^{m} (-1)^j (C_m^j + C_m^{j-1}) f_{i+1+m-j} +$$

$$+ (-1)^{m+1} C_{m+1}^{m+1} f_i = \sum_{j=0}^{m+1} (-1)^j C_{m+1}^j f_{i+m+1-j}.$$

Hemos utilizado las propiedades de las combinaciones: $C_m^{j-1} = C_{m+1}^j - C_m^j$ y $C_m^0 = C_{m+1}^0 = C_m^m = C_{m+1}^{m+1} = 1$. Ahora bien, si la igualdad (9) es justa para $k = m$, ésta será justa también para $k = m + 1$. ■

Ejemplo 1. Hacer la tabla horizontal de diferencias finitas de la función $y = x^3 + 3x^2 - x - 1$ a partir del valor inicial $x = 0$, tomando el paso $h = 1$.

△ Suponiendo $x_0 = 0$, $x_1 = 1$, $x_2 = 2$, . . ., hallamos los valores correspondientes:

x	0	1	2	3	4	5	...
y	−1	2	17	50	107	194	...

Hallamos las diferencias finitas de primer orden:

$$\begin{aligned}
\Delta y_0 &= y_1 - y_0 = 2 - (-1) = 3;\\
\Delta y_1 &= y_2 - y_1 = 17 - 2 = 15;\\
\Delta y_2 &= y_3 - y_2 = 50 - 17 = 33;\\
\Delta y_3 &= y_4 - y_3 = 107 - 50 = 57;\\
\Delta y_4 &= y_5 - y_4 = 194 - 107 = 87; \ \ldots
\end{aligned}$$

¿Cuales son las diferencias finitas de segundo orden?

$$\begin{aligned}
\Delta^2 y_0 &= \Delta y_1 - \Delta y_0 = 15 - 3 = 12;\\
\Delta^2 y_1 &= \Delta y_2 - \Delta y_1 = 33 - 15 = 18;\\
\Delta^2 y_2 &= \Delta y_3 - \Delta y_2 = 57 - 33 = 24;\\
\Delta^2 y_3 &= \Delta y_4 - \Delta y_3 = 87 - 57 = 30; \ \ldots
\end{aligned}$$

Determinamos las diferencias finitas de tercer orden:

$$\begin{aligned}
\Delta^3 y_0 &= \Delta^2 y_1 - \Delta^2 y_0 = 18 - 12 = 6;\\
\Delta^3 y_1 &= \Delta^2 y_2 - \Delta^2 y_1 = 24 - 18 = 6;\\
\Delta^3 y_2 &= \Delta^2 y_3 - \Delta^2 y_2 = 30 - 24 = 6; \ \ldots
\end{aligned}$$

Vemos que las diferencias finitas terceras $\Delta^3 y_0$, $\Delta^3 y_1$, $\Delta^3 y_2$ son constantes. Esto se explica por el hecho de que la función $f(x)$ es un polinomio de tercer grado. La diferencia finita tercera puede calcularse también con ayuda de la fórmula

$$\Delta^n P_n(x) = n! a_0 h^n,$$

o sea, $\Delta^3 P_3(x) = 3! \cdot 1 \cdot 1^3 = 6$ y las diferencias finitas de cuarto orden son iguales a cero.

Hacemos la tabla de diferencias finitas:

x	y	Δy	$\Delta^2 y$	$\Delta^3 y$	$\Delta^4 y$
0	−1	3	12	6	0
1	2	15	18	6	0
2	17	33	24	6	
3	50	57	30		
4	107	87			
5	194				

Posteriormente durante los cálculos es conveniente apuntar inmediatamente las diferencias finitas en la tabla. ▲

Puesto que los valores iniciales de la función f_i se dan, por lo general, con cierto error ε, que representa en sí errores de redondeo o errores aleatorios, es racional considerar la influencia que estos factores ejercen en el error de las diferencias finitas de órdenes superiores.

Comencemos con la influencia ejercida por los errores aleatorios y los de redondeo. Supongamos que en vez de f_i^k hemos obtenido $f_i^k + \varepsilon$. Entonces la tabla de diferencias finitas tomará la forma representada en la pág. 281 (véase la tabla 7.8).

En virtud de la igualdad (9) esto quiere decir que el error ε en la diferencia de orden k se extiende luego a las diferencias de orden $k + m$ con los coeficientes $(-1)^j C_m^j$.

Sin embargo, si los valores iniciales f_i se dan con el mismo error ε, este error se extiende a las diferencias de orden m con el coeficiente 2^m y aumenta rápidamente con el crecimiento de m ($\Delta(f_i^m) = 2^m \varepsilon$).

Si las derivadas de órdenes bastante altos de la función f quedan restringidas, de la fórmula (8) se deduce que las diferencias finitas respectivas f_j^m decrecen con el aumento de m. Por eso llega naturalmente tal momento cuando los errores de diferencias finitas, introducidos debido al redondeo o debido a la inexactitud de la información inicial, llegarán a ser comparables con los mismos valores de las diferencias finitas o incluso los superarán. Por lo tanto, la información contenida en la tabla de estas diferencias, en realidad, resultará información sobre las diferencias de los errores y no sobre las de la función; en este caso la utilización de esta información se hará no racional. Entonces se dice que el orden de las últimas diferencias

Tabla 7.8

...	...	...	...
f_{i-2}^{k}			
	$f_{i-3/2}^{k+1}$		
f_{i-1}^{k}		$f_{i-1}^{k+2}+\varepsilon$	
	$f_{i-1/2}^{k+1}+\varepsilon$		$f_{i-1/2}^{k+3}-3\varepsilon$
$f_{i}^{k}+\varepsilon$		$f_{i}^{k+2}-2\varepsilon$	
	$f_{i+1/2}^{k+1}-\varepsilon$		$f_{i+1/2}^{k+3}+3\varepsilon$
f_{i+1}^{k}		$f_{i+1}^{k+2}+\varepsilon$	
	$f_{i+3/2}^{k+1}$		
f_{i+2}^{k}			
...			

finitas que todavía es conveniente usar en los cálculos es el *orden de exactitud de la tabla de diferencias finitas.*

Ejemplo 2. Se da la tabla de los valores de la función f = sen x:

x	46°	47°	48°	49°	50°	51°	52°
f	0,7198	0,7314	0,7431	0,7547	0,7660	0,7771	0,7880

Todas las cifras citadas son justas en el sentido estrecho. Hacer la tabla de diferencias finitas y determinar el orden de exactitud de la tabla.

△ Calculamos las diferencias finitas y hacemos la tabla:

Tabla 7.9

x	f	f^1	f^2	f^3	f^4	f^5	f^6
46°	0,7193	121	−4	3	−5	8	−12
47°	0,7314	117	−1	−2	3	−4	
48°	0,7431	116	−3	1	−1		
49°	0,7547	113	−2	0			
50°	0,7660	111	−2				
51°	0,7771	109					
52°	0,7880						
	0,00005	1	2	4	8	16	32

Nótese que las diferencias finitas suelen escribirse en las unidades del último orden de los valores de la función.

En la última fila de la tabla citada están colocados los errores absolutos correspondientes. Es evidente que los valores absolutos de las diferencias finitas terceras son comparables con su error y los valores absolutos de las diferencias sucesivas son considerablemente menores de sus errores. Por eso el orden de exactitud de la tabla 7.9 es igual a 2. ▲

Lo no racional de utilizar las diferencias finitas de un orden superior al segundo en el ejemplo citado puede ser confirmado también del modo siguiente. Puesto que los datos iniciales f_i están dados con error $\varepsilon = 0{,}5\cdot 10^{-4}$, las diferencias finitas que en valor absoluto son menores de este error no tiene sentido tomarlas en consideración. Por otro lado, utilizando la igualdad (8), obtenemos $\Delta^h f_i \approx$ $\approx (\pi/180)^h$, por eso el orden de exactitud de la tabla 7.9 se define por las desigualdades siguientes:

$$(\pi/180)^k > 0{,}5\cdot 10^{-4} \geqslant (\pi/180)^{k+1}$$

que se cumplen, evidentemente, para $k = 2$.

En el caso general, de orden de exactitud de la tabla de diferencias finitas en el sentido mencionado sirve el valor mínimo de k que satisfaga la desigualdad.

$$h^{k+1} M_{k+1} \leqslant \varepsilon, \text{ donde } M_\nu = \max_{[x_i,\, x_{i+\nu}]} |f^{(\nu)}(x)|. \tag{10}$$

Hemos considerado la influencia que un error de la información inicial ejerce en el grado del polinomio interpolador. Además, si los valores f_i se dan aproximadamente o por cualesquiera causas el cálculo del valor del polinomio $P_n(x^*)$ no puede ser realizado con exactitud absoluta, obtenemos de hecho sólo un valor aproximado $\overline{P}_n(x^*)$ para el exacto $P_n(x^*)$. En este caso el error de cálculo

$\Delta_2(\overline{P}_n) = | P_n(x^*) - \overline{P}_n(x^*) |$ se estima según las reglas generales de cálculo del error de una función.

Consideremos, por ejemplo, el polinomio de Lagrange $L_n(x) = \sum\limits_{i=0}^{n} f_i l_i(x)$. Supongamos que se necesita calcular $L_n(x^*)$ para los errores dados f_i y sus errores ε_i. Las magnitudes de los coeficientes de Lagrange $l_i(x)$ están tabuladas para los nodos equidistantes y pueden considerarse números exactos, ya que están obtenidas a partir de los valores exactos de los nodos y a partir del x^* exacto. Por eso para el polinomio de Lagrange tenemos

$$\Delta_2(\overline{L}_n) \leqslant \sum_{i=0}^{n} \varepsilon_i \mid l_i(x^*) \mid .$$

En el caso en que todos los ε_i son los mismos e iguales a ε obtenemos

$$\Delta_2(\overline{L}_n) \leqslant \varepsilon \sum_{i=0}^{n} \mid f_i(x^*) \mid .$$

Análogamente se puede hallar el error de cálculo también para otras formas del polinomio interpolador.

Ejemplo 3. Sobre el segmento $[-1, 1]$ obtener la estimación uniforme del valor de cálculo de los valores del polinomio interpolador de Lagrange construido para la finción $f = \cos(\pi x/2)$ según los nodos $x_0 = -1/2$, $x_1 = 0$, $x_2 = 1/2$.

△ Puesto que $f_0 = f_2 = 1/\sqrt{2} = 0{,}707 \pm 0{,}0002$ y $f_1 = 1$ es un número exacto, el error buscado de cálculo tiene la forma

$$\Delta_2(\overline{L}_2) = 0{,}0002 \left| \frac{x^*\left(x^* - \frac{1}{2}\right)}{-\frac{1}{2}\left(-\frac{1}{2} - \frac{1}{2}\right)} \right| + 0{,}0002 \left| \frac{\left(x^* + \frac{1}{2}\right) x^*}{\left(\frac{1}{2} + \frac{1}{2}\right) \cdot \frac{1}{2}} \right| =$$

$$= 0{,}0004 \left[\left| x^*\left(x^* - \frac{1}{2}\right) \right| + \left| \left(x^* + \frac{1}{2}\right) x^* \right| \right].$$

No es difícil mostrar que sobre el segmento $[-1, 1]$ $\Delta_2(\overline{L}_2)$ toma el valor máximo en los puntos $x^* = \pm 1$ y por eso la estimación buscada es $\Delta_2(\overline{L}_2) = 0{,}0008$. ▲

§ 7.8. Polinomios interpoladores de Stirling y de Bessel

Primero vamos a examinar una cuestión importante, desde el punto de vista del error de interpolación, concerniente a la elección de los nodos de interpolación para un grado fijo del polinomio. Consideremos la expresión para el término residual del polinomio interpolador:

$$\frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^{n} (x^* - x_i); \ \xi \in (x_0, x_n).$$

Puesto que el segmento de interpolación es de ordinario no grande, la derivada $f^{n+1}(x)$ tiene una pequeña gama de medición. Por consiguiente, la gama de medición de la magnitud del error se determina, en lo fundamental, por el producto

$$|\omega_n(x^*)| = \prod_{i=0}^{n} |x^* - x_i|.$$

Esta magnitud será mínima si en calidad de nodos de interpolación se toman los nodos más próximos a x^*. Ahora bien, para un grado par $n = 2k$ del polinomio de interpolación es necesario tomar el nodo más próximo al punto x^* y k nodos a la izquierda y k nodos a la derecha de éste y, siempre que se trate de un grado impar $n = 2k + 1$, $k + 1$ nodos a la izquierda y $k + 1$ nodos a la derecha del punto x^*

Pasemos ahora al problema de construcción del polinomio interpolador para la función f asignada con sus valores f_i en los nodos x_i de la red uniforme con paso h.

Supongamos que el punto x^* está situado en la proximidad de cierto nodo que designamos x_0. Se necesita construir un polinomio interpolador de grado par. En virtud de lo dicho anteriormente en calidad de nodos de interpolación conviene elegir una red simétrica respecto al nodo x_0.

$$\ldots, x_{-k}, \ldots, x_{-1}, x_0, x_1, \ldots, x_k, \ldots$$

Introduzcamos una nueva variable t con cuya ayuda el origen se traslada al punto x_0:

$$t = (x - x_0)/h; \tag{1}$$

en este caso $t^* = (x^* - x_0)/h$.

Construyamos el polinomio interpolador en la forma siguiente:

$$P_{2k}(x) = P_{2k}(x_0 + th) = a_0 + \frac{a_1}{1!} t + \frac{a_2}{2!} t^2 + \ldots$$
$$\ldots + \frac{a_{2k-1}}{(2k-1)!} t(t^2 - 1^2) \ldots (t^2 - (k-1)^2) +$$
$$+ \frac{a_{2k}}{(2k)!} t^2 (t^2 - 1^2) \ldots (t^2 - (k-1)^2). \tag{2}$$

Los coeficientes desconocidos a_i los determinemos de las condiciones de coincidencia del polinomio P_{2k} y de la función f en los nodos x_i. En este caso nótese que la relación (1) hace corresponder al nodo x_i la magnitud $t = i$. Por ejemplo, al nodo x_0 le corresponde $t = 0$ y al nodo x_{-3} le responde $t = -3$.

Ahora bien, para determinar los coeficientes a_i obtenemos el sistema de ecuaciones lineales

$$P_{2k}(x_0 + ih) = f_i \quad (i = 0, \pm 1, \ldots, \pm k). \tag{3}$$

La estructura de este sistema es tal que a_0 se determina inmediatamente de la primera escuación del sistema (3):

$$P_{2k}(x_0) = a_0 = f_0$$

y la determinación de los demás coeficientes se reduce a la resolución sucesiva de los sistemas de dos ecuaciones con dos incógnitas:

$$P_{2k}(x_0+h) = f_0 + a_1 + \frac{a_2}{2} = f_1,$$

$$P_{2k}(x_0-h) = f_0 - a_1 + \frac{a_2}{2} = f_{-1}.$$

De aquí (véanse las designaciones del § 7.7) $a_1 = \mu f_0^1$, $a_2 = f_0^2$.

Continuando este proceso, vemos que en el j — ésimo paso el determinante del sistema de ecuaciones respecto a los coeficientes a_{2j-1} y a_{2j} tiene la forma

$$\begin{vmatrix} 1 & 1/2 \\ -1 & 1/2 \end{vmatrix} = 1 \neq 0.$$

Por consiguiente, todos los coeficientes a_i del polinomio (2) se determinan unívocamente por el sistema de ecuaciones (3) y en virtud del teorema sobre la unicidad el polinomio (2) es interpolador para la función f.

Como resultado de transformaciones poco complicadas llegamos a las siguientes expresiones para los coeficientes:

$$a_0 = f_0, \; a_{2j-1} = \mu f_0^{2j-1}, \; a_{2j} = f_0^{2j} \quad (j = 1, 2, \ldots). \tag{4}$$

Sustituyendo los valores hallados de los coeficientes en la expresión (2), obtenemos el *polinomio interpolador de Stirling* que suele designarse S_{2k}:

$$S_{2k}(t) = f_0 = \mu f_0^1 t + \frac{1}{2!} f_0^2 t^2 + \ldots + \frac{\mu f_0^{2k-1}}{(2k-1)!} t(t^2-1^2) \ldots$$

$$\ldots (t^2-(k-1)^2) + \frac{f_0^{2k}}{(2k)!} t^2(t^2-1^2) \ldots (t^2-(k-1)^2). \tag{5}$$

Puesto que el polinomio de Stirling no es sino una nueva forma del polinomio interpolador de Lagrange construido sobre los nodos $x_{-k}, \ldots, x_k$, entonces, en virtud de la fórmula (4) del § 7.5 su termino residual respecto a la variable t se puede representar en la forma

$$R_{2k} = \frac{f^{(2k+1)}(\xi)}{(2k+1)!} h^{2k+1} \prod_{i=-k}^{k} (t-i); \; \xi \in (x_{-k}, x_k) \tag{6}$$

y la estimación del error del valor aproximado de $P_{2k}(x) = S_{2k}(t)$ (del error del método), en la forma

$$\Delta_1 = |f(x) - S(t)| \leqslant \frac{M_{2k+1}}{(2k+1)!} h^{2k+1} |t(t^2-1^2) \dots (t^2-k^2)|, \quad (7)$$

donde $M_{2k+1} = \max\limits_{[x_{-k}, x_k]} |f^{(2k+1)}(x)|$.

Supongamos ahora que el punto de interpolación x^* está entre los nodos x_0 y x_1 cerca del punto $(x_0 + x_1)/2$. Se necesita construir el polinomio interpolador de grado impar. Entonces, según hemos señalado anteriormente, la red que minimiza el error es simétrica respecto al punto $(x_0 + x_1)/2$, o sea, respecto al punto $t = 1/2$.

Ahora bien, sobre la red

$$\dots, x_{-k}, \dots, x_{-1}, x_0, x_1, x_2, \dots, x_{k+1}, \dots$$

respecto a la variable t, definida por la fórmula (1), construimos el polinomio interpolador en la forma siguiente:

$$P_{2k+1}(x) = P_{2k+1}(x_0 + th) = b_0 + \frac{b_1}{1!}\left(t - \frac{1}{2}\right) + \frac{b_2}{2!} t(t-1) + \dots$$
$$\dots + \frac{b_{2k}}{(2k)!} t(t^2-1^2) \dots (t^2-(k-1)^2)(t-k) +$$
$$+ \frac{b_{2k+1}}{(2k+1)!} t(t^2-1^2) \dots (t^2-(k-1)^2)\left(t - \frac{1}{2}\right)(t-k). \quad (8)$$

Los coeficientes desconocidos b_i se determinan de las condiciones de coincidencia de los valores del polinomio P_{2k+1} con los valores de la función f en los nodos x_i.

Ahora bien, para determinar los coeficientes b_i tenemos un sistema de ecuaciones lineales

$$P_{2k+1}(x_0 + ih) = f_i \ (i = -k, \dots, 0, 1, \dots, k+1). \quad (9)$$

La estructura de este sistema es tal que su resolución se reduce a la resolución sucesiva del sistema de dos ecuaciones con dos incógnitas

$$P_{2k+1}(x_0) = b_0 - \frac{b_1}{2} = f_0,$$

$$P_{2k+1}(x_1) = b_0 + \frac{b_1}{2} = f_1.$$

De aquí $b_0 = \mu f_{1/2}$, $b_1 = f^1_{1/2}$.
Continuando este proceso, vemos que en el j-ésimo paso el determinante del sistema de ecuaciones respecto a los coeficientes b_{2j-2} y b_{2j-1} es distinto del cero lo que se puede mostrar fácilmente de un modo análogo al cómo esto hemos hecho al construir al polinomio de Stirling. Por lo tanto, todos los coeficientes b_i del polinomio (8) se definen unívocamente por el sistema de ecuaciones (9) y en virtud del teorema de unicidad el polinomio (8) es interpolador para la función f.

Las transformaciones poco complicadas ofrecen las siguientes expresiones para los coeficientes:

$$b_{2j-2}=\mu f_{1/2}^{2j-2},\quad b_{2j-1}=f_{1/2}^{2j-1}\quad (j=1,\ 2,\ \ldots). \tag{10}$$

Sustituyendo los valores hallados de los coeficientes en la expresión (8), obtenemos el *polinomio interpolador de Bessel* que suele designarse B_{2k+1}:

$$B_{2k+1}(t)=\mu f_{1/2}+\frac{f_{1/2}^{1}}{1!}\left(t-\frac{1}{2}\right)+$$

$$+\frac{\mu f_{1/2}^{2}}{2!}t(t-1)+\frac{f_{1/2}^{3}}{3!}t(t-1)\left(t-\frac{1}{2}\right)+\ldots$$

$$\ldots+\frac{\mu f_{1/2}^{2k}}{(2k)!}t(t^2-1^2)\ldots(t^2-(k-1)^2(t-k)+$$

$$+\frac{f_{1/2}^{2k+1}}{(2k+1)!}t(t^2-1^2)\ldots$$

$$\ldots(t^2-(k-1)^2)(t-k)\left(t-\frac{1}{2}\right). \tag{11}$$

Puesto que el polinomio de Bessel es una forma más de representar el polinomio interpolador de Lagrange, construido sobre los nodos $x_{-k}, \ldots, x_{k+1}$, entonces, en virtud de la fórmula (4) del § 7.5, su término residual respecto a la variable t puede escribirse en la forma

$$R_{2k+1}=\frac{f^{(2k+2)}(\xi)}{(2k+2)!}h^{2k+2}\prod_{i=-k}^{k+1}(t-i);\ \xi\in(x_{-k};\ x_{k+1}) \tag{12}$$

y la estimación del error del valor aproximado de $P_{2k+1}(x)=$ $=B_{2k+1}(t)$ (del error del método), en la forma

$$\Delta_1=|f(x)-B_{2k+1}(t)|\leqslant\frac{M_{2k+2}}{(2k+2)!}h^{2k+2}\prod_{i=-k}^{k+1}|t-i|, \tag{13}$$

donde $M_{2k+2}=\max\limits_{[x_{-k},\ x_{k+1}]}|f^{(2k+2)}(x)|$.

Ahora bien, hemos considerado dos polinomios interpoladores: el polinomio de Stirling que se utiliza al construir un polinomio de grado par y se construye según el número impar de nodos y el polinomio de Bessel que se utiliza al construir un polinomio de grado impar y se construye según un número par de nodos.

En cambio, si el grado del polinomio está fijado no rígidamente, o sea, puede ser tanto par como impar, es conveniente utilizar el polinomio de Stirling en el caso en que

$$|t^*|=|x^*-x_0|/h\leqslant 0{,}25, \tag{14}$$

o sea cuando el punto de interpolación x^* está situado más próximo al nodo x_0 que al centro entre los nodos. El polinomio de Bessel ha de utilizarse en el caso en que

$$0{,}25 \leqslant t^* \leqslant 0{,}75, \tag{15}$$

o sea cuando el punto de interpolación x^* está situado más próximo al centro entre los nodos x_0 y x_1. Una de las condiciones (14) ó (15) puede ser siempre asegurada por la elección del nodo respectivo en calidad de x_0.

Ejemplo. Utilizando el polinomio interpolador correspondiente, calcular en los puntos $x_1^* = 48{,}63°$ y $x_2^* = 49{,}19°$ los valores de la función $f = \operatorname{sen} x$, representada en la forma de una tabla con paso igual a 1° (véase la tabla 7.9), que contiene los valores f_i con cuatro cifras justas en el sentido estrecho. Estimar el error del resultado.

$\triangle$ En el ejemplo 2 del § 7.7 hemos determinado que el orden de exactitud de la tabla dada es 2. Por eso no es racional construir un polinomio interpolador cuyo grado sea superior al segundo, o sea, conviene construir el polinomio de Stirling de segundo grado o bien el polinomio de Bessel de primer grado.

Puesto que el punto $x_1^* = 48{,}63$ está situado más próximamente al centro entre los nodos de 48° y 49°, entonces, al calcular sen x_1^*, en calidad de x_0 es necesario tomar el nodo de 48° y hacer uso del polinomio de Bessel; en este caso $t_1^* = (x_1^* - x_0) \times h^{-1} = 0{,}63$. El punto $x_2^* = 49{,}19°$ está cerca del ángulo 49°, por eso, al calcular sen x_2^*, en calidad de central hay que elegir el nodo $x_0 = 49°$ y valerse del polinomio de Stirling; en este caso $t_2^* = (x_2^* - x_0)h^{-1} = 0{,}19$.

Ahora bien, utilizando lo dicho anteriormente, las fórmulas (11) y (5), así como los datos de la tabla 7.9, tenemos

$$B_1(0{,}63) = \frac{0{,}7431 + 0{,}7547}{2} + \frac{0{,}0116}{1!} \cdot 0{,}13 = 0{,}750408;$$

$$S_2(0{,}19) = 0{,}7547 + \frac{0{,}0116 + 0{,}0113}{2 \cdot 1!}\, 0{,}19 + \frac{-0{,}0003}{2!} \cdot 0{,}19^2 = \\ = 0{,}7568809.$$

Ahora vamos a estimar los errores del método con ayuda de las fórmulas (13) y (7), respectivamente:

$$\Delta_1(B_1) < \frac{0{,}76}{2!}\left(\frac{\pi}{180}\right)^2 \cdot 0{,}63 \cdot 0{,}37 = 0{,}27 \cdot 10^{-4};$$

$$\Delta_1(S_2) < \frac{1}{3!}\left(\frac{\pi}{180}\right)^3 \cdot 0{,}19\,|0{,}19^2 - 1^2| = 0{,}2 \cdot 10^{-6}.$$

Teniendo en cuenta errores en los valores de la función y de sus diferencias finitas, citados en la tabla 7.9, obtenemos las magnitudes de los errores de cálculo:

$$\Delta_2(B_1) = 0{,}5 \cdot 10^{-4} + 1 \cdot 10^{-4} \cdot 0{,}13 = 0{,}63 \cdot 10^{-4};$$

$$\Delta_2(S_2) = 0{,}5 \cdot 10^{-4} + 1 \cdot 10^{-4} \cdot 0{,}19 + 1 \cdot 10^{-4} \cdot 0{,}19^2 = \\ = 0{,}73 \cdot 10^{-4}.$$

Al redondear los valores de B_1 (0,63) y S_2 (0,19) hasta cuatro cifras, obtenemos los siguientes errores de redondeo: $\Delta_3(B_1) = 0{,}08 \times \times 10^{-4}$; $\Delta_3(S_2) = 0{,}11 \cdot 10^{-4}$.

Uniendo todos los errores hallados tenemos finalmente sen $48{,}63^\circ = 0{,}7509 \pm 0{,}0001$; sen $49{,}19^\circ = 0{,}7569 \pm 0{,}0001$. ▲

Observación. En algunos casos para minimizar el error total resulta racional la utilización de un polinomio interpolador cuyo grado sea superior al orden de exactitud de la tabla de diferencias finitas. Esto se explica por lo siguiente. Con el aumento del grado del polinomio el error de cálculo naturalmente crece. Al mismo tiempo la disminución del error del método puede ser tan brusco que originará también la disminución del error total.

§ 7.9. Primero y segundo polinomios interpoladores de Newton

Si el punto de interpolación x^* está al comienzo de la tabla o a su fin, no siempre es posible elegir una cantidad suficiente de nodos a la izquierda de x^* y a su derecha para construir las diferencias finitas necesarias. En este caso se utilizan formas especiales del polinomio interpolador.

Supongamos que el punto x^* está situado en la proximidad del primer nodo de la red $x_0, x_1, \ldots, x_h, \ldots$ Consideremos la variable t, definida por la relación (1) del § 7.8, y construyamos el polinomio interpolador en la forma siguiente:

$$P_k(x) = P_k(x_0 + th) = a_0 + \frac{a_1}{1!} t + \frac{a_2}{2!} t(t-1) + \ldots$$
$$\ldots + \frac{a_k}{k!} t \ldots (t-k+1). \qquad (1)$$

Los coeficientes desconocidos a_i se determinan de las condiciones de coincidencia del polinomio P_k y de la función f en los nodos x_i. En este caso recuérdese que al nodo x_i le corresponde la magnitud $t = i$. Ahora bien, para determinar los coeficientes a_i obtenemos el sistema de ecuaciones lineales

$$P_k(x_0 + ih) = f_i \ (i = 0, 1, \ldots, k). \qquad (2)$$

La estructura de este sistema es tal que a_0 se determina inmediatamente de la primera ecuación del sistema (2), a_1 se determina de la segunda ecuación para a_0 ya determinado, etc. En efecto, suponiendo $i = 0$, de la primera ecuación del sistema (2) obtenemos

$$P_k(x_0) = a_0 = f_0$$

y de la segunda ecuación para $i = 1$ tenemos

$$P_k(x_0 + h) = f_0 + \frac{a_1}{1!} \cdot 1 = f_1; \ a_1 = \Delta f_0.$$

Continuando este proceso, como resultado de transformaciones poco complicadas obtenemos las siguientes expresiones para los coeficientes:

$$a_0 = f_0;\ a_i = \Delta^i f_0\ (i = 1, 2, \ldots, k). \qquad (3)$$

Sustituyendo los valores hallados de los coeficientes en la igualdad (1), obtenemos el *primer polinomio interpolador de Newton* que suele designarse N_k^1:

$$N_k^1(t) = f_0 + \frac{\Delta f_0}{1!} t + \frac{\Delta^2 f_0}{2!} t(t-1) + \ldots$$

$$\ldots + \frac{\Delta^k f_0}{k!} t(t-1) \ldots (t-k+1). \qquad (4)$$

En virtud de la fórmula (4) del § 7.5 el término residual respecto a la variable t puede ser representado en la forma

$$R_k = \frac{f^{(k+1)}(\xi)}{(k+1)!} h^{k+1} t(t-1) \ldots (t-k);\ \xi \in (x_0, x_k) \qquad (5)$$

y la estimación del error del valor aproximado de $N_k^1(t)$ (el error del método), en la forma

$$\Delta_1 = |f(x) - N_k^1(t)| \leqslant \frac{M_{k+1}}{(k+1)!} h^{k+1} |t(t-1) \ldots (t-k)|, \qquad (6)$$

donde $M_{k+1} = \max\limits_{[x_0, x_k]} |f^{(k+1)}(x)|$.

Supongamos ahora que el punto x^* está situado cerca del último nodo de la red $\ldots x_{-k}, \ldots, x_{-1}, x_0$. Volviendo a utilizar la variable t, definida por la relación (1) del §7.8, para esta red construimos el polinomio interpolador en la forma siguiente:

$$P_k(x) = P_k(x_0 + th) = a_0 + \frac{a_1}{1!} t +$$

$$+ \frac{a_2}{2!} t(t+1) + \ldots + \frac{a_k}{k!} t \ldots (t+k-1). \qquad (7)$$

Los coeficientes desconocidos a_i se determinan de las condiciones de coincidencia del polinomio P_k y de la función f en los nodos x_i. En este caso obsérvese que al nodo x_{-i} le corresponde el valor de $t = -i$. Ahora bien, para hallar los coeficientes obtenemos el sistema de ecuaciones lineales

$$P_k(x_0 - ih) = f_i \quad (i = 0, 1, \ldots, k). \qquad (8)$$

La estructura de este sistema es tal que a_0 se calcula inmediatamente de la primera ecuación del sistema (8) y a_1, de la segunda ecuación para a_0 ya determinado, etc. En efecto, suponiendo $i = 0$, de la primera ecuación del sistema (8) hallamos

$$P_k(x_0) = a_0 = f_0$$

y de la segunda ecuación para $i = 1$ tenemos

$$P_k(x_0 - h) = f_0 + \frac{a_1}{1!} 1 = f_{-1};\ a_1 = \nabla f_0.$$

Continuando este proceso, como resultado de transformaciones poco complicadas obtenemos las siguientes expresiones para los coeficientes:

$$a_0 = f_0;\ a_i = \nabla^i f_0\ (i = 1, 2, \ldots, k). \qquad (9)$$

Sustituyendo los valores hallados de los coeficientes en la igualdad (7), obtenemos el *segundo polinomio interpolador de Newton* que suele designarse N_k^{II}:

$$N_k^{\mathrm{II}}(t) = f_0 + \frac{\nabla f_0}{1!}t + \frac{\nabla^2 f_0}{2!}t(t+1) + \ldots$$
$$\ldots + \frac{\nabla^k f_0}{k!}t(t+1)\ldots(t+k-1) \qquad (10)$$

con el término residual

$$R_k = \frac{f^{(k+1)}(\xi)}{(k+1)!}h^{k+1}t(t+1)\ldots(t+k); \qquad (11)$$
$$\xi \in (x_{-k},\ x_0)$$

y con la estimación del error del valor aproximado

$$\Delta_1 = |f(x) - N^{\mathrm{II}}(t)| \leqslant \frac{M_{k+1}}{(k+1)!}h^{k+1}|t(t+1)\ldots(t+k)|, \quad (12)$$

donde $M_{k+1} = \max\limits_{[x_{-k}, x_0]} |f^{(k+1)}(x)|$.

Las fórmulas (4) y (10) se llaman con frecuencia *fórmulas de interpolación de Newton para la interpolación hacia adelante y hacia atrás*, respectivamente.

Ejemplo. Plantear los polinomios interpoladores respectivos y calcular en los puntos $x^* = 0{,}63$ y $x_2^* = 1{,}35$ los valores de la función $f = 3^x$ representada en forma de la tabla siguiente que contiene los valores f_i con cuatro cifras justas en amplio sentido:

x	0,50	0,75	1,00	1,25	1,50
f	1,732	2,280	3,000	3,948	5,196

Estimar el error del resultado.

△ Vamos a completar la tabla dada por los valores de las diferencias finitas colocando en la última fila los valores de los errores de las diferencias finitas correspondientes.

Puesto que la cuarta diferencia finita coincide con el propio error no conviene aproximar, desde el punto de vista del error de cálculo, la función representada haciendo uso de un polinomio cuyo grado sea superior al tercero.

Luego, puesto que $x_1^* = 0{,}63$ está situado al inicio de la tabla y $x_2^* = 1{,}35$ al fin de ésta, para calcular $f_1^* = 3^{0{,}63}$ es necesario utilizar el primer polinomio interpolador de Newton y para calcular $f_2^* = 3^{1{,}35}$, el segundo.

x	f	f^1	f^2	f^3	f^4
0,50	1,732	548	172	56	16
0,75	2,280	720	228	72	
1,00	3,000	948	300		
1,25	3,948	1248			
1,50	5,196				
	0,001	2	4	8	16

Así pues, suponiendo $x_0 = 0{,}5$, calculamos $t_1^* = (x_0^* - x_0)/h = (0{,}63 - 0{,}5)/0{,}25 = 0{,}52$. Sustituyendo el valor obtenido de t_1^* en la expresión (4) para el primer polinomio interpolador de Newton y utilizando las magnitudes de las diferencias finitas dadas en la tabla 7.10, tenemos

$$N_3^{\mathrm{I}}(0{,}52) = 1{,}732 + \frac{0{,}548}{1!}\cdot 0{,}52 + \frac{0{,}172}{2!}\cdot 0{,}52\,(-0{,}48) + \\ + \frac{0{,}056}{3!}\cdot 0{,}52\,(-0{,}48)\,(-1{,}48) = 1{,}9989420.$$

Análogamente, suponiendo $x_0 = 1{,}50$, calculamos $t_2^* = (1{,}35 - 1{,}50)/0{,}25 = -0{,}60$ y utilizando la expresión (10) para el segundo polinomio interpolador de Newton, obtenemos

$$N_3^{\mathrm{II}}(-0{,}60) = 5{,}196 + \frac{1{,}248}{1!}\cdot(-0{,}60) + \frac{0{,}300}{2!}\cdot(-0{,}60)\cdot 0{,}40 + \\ + \frac{0{,}072}{3!}\cdot(-0{,}60)\cdot 0{,}40\cdot 1{,}40 = 4{,}407168.$$

Vamos a estimar el error del método con ayuda de las fórmulas (6) y (12), respectivamente:

$$\Delta_1(N_3^{\mathrm{I}}) < \frac{4\ln^4 3}{4!}\cdot 0{,}25^4\cdot 0{,}52\cdot 0{,}48\cdot 1{,}48\cdot 2{,}48 = 0{,}0009;$$

$$\Delta_1(N_3^{\mathrm{II}}) < \frac{5{,}2\ln^4 3}{4!}\cdot 0{,}25^4\cdot 0{,}60\cdot 0{,}40\cdot 1{,}40\cdot 2{,}40 = 0{,}001.$$

Teniendo en cuenta los errores de f^k dados en la tabla 7.10, estimamos los errores de cálculo:

$\Delta_2(N_3^{\mathrm{I}}) < 0{,}001 + 0{,}0011 + 0{,}0005 + 0{,}0005 = 0{,}0031;$

$\Delta_2(N_3^{\mathrm{II}}) < 0{,}001 + 0{,}0012 + 0{,}0005 + 0{,}0005 = 0{,}0032.$

Redondeando los valores de $N_3^{\rm I}$ (0,52) y $N_3^{\rm II}$ (−0,60) hasta cuatro cifras, obtenemos los siguientes errores de redondeo:

$$\Delta_3(N_3^{\rm I}) = 0{,}06\cdot 10^{-3}; \quad \Delta_3(N_3^{\rm II}) = 0{,}2\cdot 10^{-3}.$$

Uniendo todos los errores hallados, tenemos finalmente $3^{0,63} = 1{,}999 \pm 0{,}005$; $3^{1,35} = 4{,}407 \pm 0{,}005$. ▲

§ 7.10. Diferencias divididas

En los párrafos precedentes hemos considerado diferentes formas del polinomio interpolador para la red uniforme de nodos. Los coeficientes de los polinomios construidos se han determinado con ayuda de las diferencias finitas. En este párrafo volvemos a examinar el caso en que los valores de la función se dan en los nodos no equidistantes. En este caso en vez de diferencias finitas se consideran las *diferencias divididas* que son, en cierto sentido, un análogo del concepto de derivada y se definen del modo siguiente.

Supongamos que la función $y = f(x)$ se da por sus valores $y_0 = f_0 = f(x_0)$, $y_1 = f_1 = f(x_1)$, $\ldots$, $y_k = f_k = f(x_k)$, $\ldots$ en los nodos x_i de una red arbitraria Λ_n. *Las diferencias divididas de orden nulo de* $f(x_i)$ coinciden con los valores que la función tiene en los puntos x_i. Se llaman *diferencias divididas de primer orden* las relaciones

$$f(x_0;\ x_1) = \frac{f_1 - f_0}{x_1 - x_0}; \quad f(x_1;\ x_2) = \frac{f_2 - f_1}{x_2 - x_1}; \quad \ldots$$

$$\ldots; \quad f(x_i;\ x_{i+1}) = \frac{f_{i+1} - f_i}{x_{i+1} - x_i}; \quad \ldots$$

Se denominan *diferencias divididas de segundo orden* las relacionese

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

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

$$f(x_i;\ x_{i+1};\ x_{i+2}) = \frac{f(x_{i+1};\ x_{i+2}) - f(x_i;\ x_{i+1})}{x_{i+1} - x_i}.$$

En el caso general la *diferencia dividida de k-ésimo orden* se define por la diferencia dividida de $(k-1)$-ésimo orden con ayuda de la fórmula

$$f(x_i;\ x_{i+1};\ \ldots;\ x_{i+k}) = \frac{f(x_{i+1};\ \ldots\ x_{i+k}) - f(x_i;\ \ldots,\ x_{i+k-1})}{x_{i+k} - x_i}.$$

Para las relaciones de diferencia se utiliza también otra designación:

$$f(x_i;\ x_{i+1};\ \ldots;\ x_{i+k}) \equiv [x_i,\ x_{i+1};\ \ldots;\ x_{i+k}].$$

Es cómodo escribir las diferencias divididas en forma de una tabla, análogamente a como lo hemos hecho para las diferencias finitas centrales (véase la tabla 7.11).

Tabla 7.11

x_i	$y_i=f(x_i)$	$[x_i;\ x_{i+1}]$	$[x_i;\ x_{i+1};\ x_{i+2}]$	$[x_i;\ x_{i+1};\ x_{i+2};\ x_{i+3}]$	$[x_i;\ x_{i+1};\ x_{i+2};\ x_{i+3};\ x_{i+4}]$
x_0	y_0				
		$[x_0;\ x_1]$			
x_1	y_1		$[x_0;\ x_1;\ x_2]$		
		$[x_1;\ x_2]$		$[x_0;\ x_1;\ x_2;\ x_3]$	
x_2	y_2		$[x_1;\ x_2;\ x_3]$		$[x_0;\ x_1;\ x_2;\ x_3;\ x_4]$
		$[x_2;\ x_3]$		$[x_1;\ x_2;\ x_3;\ x_4]$	
x_3	y_3		$[x_2;\ x_3;\ x_4]$		
		$[x_3;\ x_4]$			
x_4	y_4				

Ejemplo 1. Hacer la tabla de diferencias divididas para la función $y = f(x)$ representada en forma de la tabla siguiente:

x	0	1	5	10
y	10	20	100	1100

Haciendo uso inmediatamente de la definición hallamos las diferencias divididas de primer orden:

$$[x_0;\ x_1] = \frac{y_1 - y_0}{x_1 - x_0} = \frac{20-10}{1-0} = 10;$$

$$[x_1;\ x_2] = \frac{y_2 - y_1}{x_2 - x_1} = \frac{100-20}{5-1} = \frac{80}{4} = 20;$$

$$[x_2;\ x_3] = \frac{y_3 - y_2}{x_3 - x_2} = \frac{1100-100}{10-5} = \frac{1000}{5} = 200.$$

Análogamente obtenemos las diferencias divididas de segundo orden

$$[x_0;\ x_1;\ x_2]=\frac{[x_1;\ x_2]-[x_0;\ x_1]}{x_2-x_0}=\frac{20-10}{5-0}=2;$$

$$[x_1;\ x_2;\ x_3]=\frac{[x_2;\ x_3]-[x_1;\ x_2]}{x_3-x_1}=\frac{200-20}{10-1}=20.$$

La diferencia dividida de tercer orden

$$[x_0;\ x_1;\ x_2;\ x_3]=\frac{[x_1;\ x_2;\ x_3]-[x_0;\ x_1;\ x_2]}{x_3-x_0}=\frac{20-2}{10}=1{,}8.$$

Representemos los resultados de cálculos en forma de la tabla de diferencias divididas:

x	y	$[x_i;\ x_{i+1}]$	$[x_i;\ x_{i+1};\ x_{i+2}]$	$[x_i;\ x_{i+1};\ x_{i+2};\ x_{i+3}]$
0	10			
		10		
1	20		2	
		20		1,8
5	100		20	
		200		
10	1100			

Vamos a citar algunas propiedades de las diferencias divididas.

1°. *Las diferencias divididas de todos los órdenes son una combinación lineal de los valores $f_i = f(x_i)$, es decir, es válida la fórmula siguiente:*

$$f(x_0;\ \ldots,\ x_k)=\sum_{i=0}^{k}\frac{f_i}{\prod\limits_{\substack{j=0\\ j\neq i}}^{k}(x_i-x_j)}. \tag{2}$$

□ Para $k=0$ obtenemos la identidad $f(x_0)=f_0$; para $k=1$ tenemos $f(x_0;\ x_1)=\frac{f(x_0)-f(x_1)}{x_0-x_1}$ lo que es la definición de la diferencia dividida $f(x_0;\ x_1)$.

Realicemos la demostración posterior haciendo uso de la inducción. Supongamos que la igualdad (2) es válida para todos los $k\leqslant n$. Entonces

$$f(x_0;\ \ldots;\ x_{n+1})=\frac{f(x_0;\ \ldots;\ x_n)-f(x_1;\ \ldots;\ x_{n+1})}{x_0-x_{n+1}}=$$

$$=\frac{1}{x_0-x_{n+1}}\left(\sum_{i=0}^{n}\frac{f_i}{\prod\limits_{\substack{j=0\\ j\neq i}}^{n}(x_i-x_j)}-\sum_{i=1}^{n+1}\frac{f_i}{\prod\limits_{\substack{j=1\\ j\neq i}}^{n+1}(x_i-x_j)}\right)=$$

$$
= \frac{f_0}{\prod\limits_{j=1}^{n+1}(x_0 - x_j)} + \sum_{i=1}^{n} \frac{f_i}{\prod\limits_{\substack{j=0\\ j\neq i}}^{n+1}(x_i - x_j)} +
$$

$$
+ \frac{f_{n+1}}{\prod\limits_{j=0}^{n}(x_{n+1} - x_j)} = \sum_{i=0}^{n+1} \frac{f_i}{\prod\limits_{\substack{j=0\\ j\neq i}}^{n+1}(x_i - x_j)}.
$$

Ahora bien, la relación (2) es justa también para $k = n+1$. ■

2°. *La diferencia dividida es la función simétrica de los propios argumentos, o sea, no cambia para toda permutación de los mismos.*

□ Para toda permutación de los argumentos $x_0, \ldots, x_k$ en el segundo miembro de la relación (2) los sumandos correspondientes cambian sólo de lugar y la suma, como es evidente, no cambia. Por consiguiente, tampoco cambia el primer miembro de la relación (2), o sea, $f(x_0; \ldots; x_k)$. ■

3°. *La diferencia dividida satisface la igualdad*

$(af + bg)(x_0; \ldots; x_k) = af(x_0; \ldots; x_k) + bf(x_0; \ldots; x_k)$, (3)

donde a y b son constantes.

Esto se deduce inmediatamente de la relación (2) en virtud de que su segundo miembro es lineal respecto a f_i.

La segunda propiedad expresa la relación existente entre la diferencia dividida $f(x_0; \ldots; x_k)$ y la derivada $f^{(k)}(x)$.

4°. *Si los nodos $x_0, \ldots, x_k$ pertenecen al segmento $[a, b]$ y la función $f(x)$ tiene sobre $[a, b]$ una derivada continua de k-ésimo orden, entonces existe tal punto $\xi \in (a, b)$ que*

$$
f(x_0; \ldots; x_k) = \frac{1}{k!} f^{(k)}(\xi). \tag{4}
$$

De esta propiedad se desprende un corolario simple. Sea $f(x) = a_0x^k + a_1x^{k-1} + \ldots + a_k$ el polinomio de k-ésimo grado. Entonces, evidentemente, $f^{(k)}(x) = a_0k!$ y la relación (4) da para la diferencia dividida el valor

$$
f(x_0; \ldots x_k) = \frac{1}{k!} a_0k! = a_0. \tag{5}
$$

Así pues, en todo polinomio de k-ésimo grado las diferencias divididas de k-ésimo orden son iguales a una magnitud constante, es decir, al coeficiente del grado superior del polinomio. Las diferencias divididas de órdenes superiores (mayores que k) son, evidentemente, iguales a cero.

Ejemplo 2. Verifiquemos la validez de la propiedad 1° utilizando cuatro valores de la función $y = f(x)$: $y_i = f(x_i)$ $(i = 0, 1, 2, 3)$.

Las diferencias divididas primeras las determinamos así:

$$f(x_0;\ x_1)=\frac{y_0-y_1}{x_0-x_1};\ f(x_1;\ x_2)=\frac{y_1-y_2}{x_1-x_2};$$

$$f(x_2;\ x_3)=\frac{y_2-y_3}{x_2-x_3}.$$

Hallamos las diferencias divididas segundas:

$$f(x_0;\ x_1;\ x_2)=\frac{1}{x_0-x_2}\left(\frac{y_0-y_1}{x_0-x_1}-\frac{y_1-y_2}{x_1-x_2}\right)=$$

$$=\frac{y_0}{(x_0-x_1)(x_0-x_2)}+\frac{y_1}{(x_1-x_0)(x_1-x_2)}+\frac{y_2}{(x_2-x_0)(x_2-x_1)};$$

$$f(x_1;\ x_2;\ x_3)=\frac{1}{x_1-x_3}\left(\frac{y_1-y_2}{x_1-x_2}-\frac{y_2-y_3}{x_2-x_3}\right)=$$

$$=\frac{y_1}{(x_1-x_2)(x_1-x_3)}+\frac{y_2}{(x_2-x_1)(x_2-x_3)}+\frac{y_3}{(x_3-x_1)(x_3-x_2)}$$

lo que está en plena correspondencia con la propiedad 1°. Análogamente se puede mostrar que la diferencia dividida tercera

$$f(x_0;\ x_1;\ x_2;\ x_3)=\frac{y_0}{(x_0-x_1)(x_0-x_2)(x_0-x_3)}+$$

$$+\frac{y_1}{(x_1-x_0)(x_1-x_2)(x_1-x_3)}+\frac{y_2}{(x_2-x_0)(x_2-x_1)(x_2-x_3)}+$$

$$+\frac{y_3}{(x_3-x_0)(x_3-x_1)(x_3-x_2)}.$$

Calculemos la diferencia finita tercera, utilizando los valores de la función y del argumento, dados en el ejemplo 1.

$$[x_0;\ x_1;\ x_2;\ x_3]=\frac{10}{(-1)(-5)(-10)}+\frac{20}{1\cdot(-4)(-9)}+$$

$$+\frac{100}{5\cdot4\cdot(-5)}+\frac{1100}{10\cdot9\cdot5}=-\frac{1}{5}+\frac{5}{9}-1+\frac{22}{9}=1{,}8$$

lo que coincide con el valor obtenido en el ejemplo 1.

Ejemplo 3. Verifiquemos ahora para la función $y=f(x)$ la validez de la propiedad 2°. Mostremos, por ejemplo, que $f(x_0;\ x_1;\ x_2)=$ $=f(x_1;\ x_0;\ x_2)$.

En efecto, según la definición de la diferencia dividida tenemos

$$f(x_1;\ x_0;\ x_2)=\frac{1}{x_1-x_2}\left(\frac{y_1-y_0}{x_1-x_0}-\frac{y_0-y_2}{x_0-x_2}\right)=$$

$$=\frac{y_1}{(x_1-x_0)(x_1-x_2)}+\frac{y_0}{(x_0-x_1)(x_0-x_2)}+\frac{y_2}{(x_2-x_0)(x_2-x_1)}.$$

Hemos obtenido el mismo resultado que en el ejemplo 2 (con el cambio del orden de los sumandos).

Ejemplo 4. Vamos a ilustrar la propiedad 3°. Junto con la función $y=f(x)$ (véase el ejemplo 2) consideremos la función $z=g(x)$ y sus valores $z_i=g(x_i)$ $(i=0,\ 1,\ 2,\ 3)$ dados en los mismos nodos que y_i. Planteemos la combinación lineal $u(x) = af(x) + bg(x)$, donde a y b son las constantes. Calculemos ahora $u(x_0;\ x_1;\ x_2)$. Al igual que en el ejemplo 2, hallamos

$$u(x_0;\ x_1;\ x_2) = \frac{ay_0 + bz_0}{(x_0 - x_1)(x_0 - x_2)} + \frac{ay_1 + bz_1}{(y_1 - x_0)(x_1 - x_2)} + \frac{ay_2 + bz_2}{(x_2 - x_0)(x_2 - x_1)}.$$

Agrupando por separado los primeros y segundos sumandos de cada uno de tres términos de la suma obtenemos

$$u(x_0;\ x_1;\ x_2) = af(x_0;\ x_1;\ x_2) + bg(x_0;\ x_1;\ x_2)$$

que es lo que se necesitaba mostrar.

§ 7.11. Polinomio interpolador de Newton para una red arbitraria de nodos

Utilizando la forma de Lagrange, representemos el polinomio interpolador en la forma siguiente:

$$L_0(x) + (L_1(x) - L_0(x)) + \ldots + (L_n(x) - L_{n-1}(x)).$$

Aquí $L_0(x) = f(x_0)$; $L_k(x)$ $(k = 1, 2, \ldots, n)$ son los polinomios interpoladores en la forma de Lagrange construidos sobre los nodos $x_0, x_1, \ldots, x_k$. Consideremos las diferencias

$$L_k - L_{k-1} = \sum_{i=0}^{k} f_i \frac{\omega_k(x)}{(x - x_i)\,\omega'_k(x_i)} - \sum_{i=0}^{k-1} f_i \frac{\omega_{k-1}(x)}{(x - x_i)\,\omega'_{k-1}(x_i)} = \sum_{i=0}^{k-1} f_i \frac{\omega_{k-1}(x)}{\omega'_k(x_i)} + f_k \frac{\omega_{k-1}(x)}{\omega'_k(x_k)} = \omega_{k-1}(x) \sum_{i=0}^{k} \frac{f_i}{\omega'_k(x_i)}.$$

Ahora bien, utilizando la fórmula (2) del § 7.10, obtenemos

$$L_k - L_{k-1} = \varphi_{k-1}(x)\, f(x_0;\ \ldots;\ x_k) \tag{1}$$

y el polinomio interpolador toma la forma

$$N_n(x) = f_0 + (x - x_0)\, f(x_0;\ x_1) + \ldots + (x - x_0)\ldots(x - x_{n-1})\, f(x_0;\ \ldots;\ x_n). \tag{2}$$

Esta forma se llama *polinomio interpolador de Newton con diferencias divididas.*

La expresión para el error tiene, evidentemente, la misma forma que en el caso del polinomio de Lagrange [véanse las fórmulas (8) y (9) del § 7.5].

Ejemplo 1. Construir sobre los nodos $x_0 = 0$, $x_1 = 1/3$, $x_2 = 1$ el polinomio interpolador de Newton para la función $f = \operatorname{sen}(\pi x/2)$.

△ Teniendo en cuenta que $f_0 = 0$; $f_1 = 0{,}5$; $f_2 = 1$ planteemos las diferencias divididas necesarias:

$$f(x_0;\ x_1) = \frac{0-\frac{1}{2}}{0-\frac{1}{3}} = \frac{3}{2};\quad f(x_1,\ x_2) = \frac{\frac{1}{2}-1}{\frac{1}{3}-1} = \frac{3}{4};$$

$$f(x_0;\ x_1;\ x_2) = \frac{\frac{3}{2}-\frac{3}{4}}{0-1} = -\frac{3}{4}.$$

Sustituyendo los valores obtenidos en la fórmula (2), para $n = 2$ tenemos

$$N_2(x) = 0 + \frac{3}{2}(x-0) - \frac{3}{4}(x-0)\left(x-\frac{1}{3}\right).$$

El polinomio construido debe ser, evidentemente, idéntico al de Lagrange construido en el ejemplo 1 del § 7.6; proponemos que el lector se cerciore de ello de por si. ▲

Nótese que en la forma (2) del polinomio interpolador los nodos x_i llevan impuesta la única condición, es decir, la falta de coincidencia de los mismos. Por eso los nodos pueden ser numerados en un orden arbitrario. Por ejemplo, con índice «0» se designa frecuentemente el último nodo de la tabla, por x_1 se toma el penúltimo nodo designándolo x_{-1}, etc. En este caso el polinomio (2) toma la forma

$$\begin{aligned} N_n(x) &= f_0 + (x-x_0)\, f(x_0;\ x_{-1}) + \ldots \\ &\ldots + (x-x_0)\ldots(x-x_{-n+1})\, f(x_0;\ \ldots;\ x_{-n}) \end{aligned} \qquad (3)$$

y se llama *polinomio de Newton para interpolar hacia atrás.*

A título de ilustración de lo dicho vamos a resolver el ejemplo 1 para $x_0 = 1$, $x_{-1} = 1/3$; $x_{-2} = 0$. Según la fórmula (3) obtenemos

$$N_2(x) = 1 + \frac{3}{4}(x-1) - \frac{3}{4}(x-1)\left(x-\frac{1}{3}\right).$$

Nótese que las diferencias divididas necesarias han sido tomadas del ejemplo 1, con la particularidad de que hemos utilizado las propiedades de su simetría respecto a los propios argumentos.

De la comparación de las formas de Lagrange y de Newton para el polinomio interpolador se deduce que la utilización de la representación en la forma de Lagrange puede recomendarse, en primer lugar, en las investigaciones teóricas, por ejemplo al estudiar la cuestión acerca de la convergencia de $P_n(f, x)$ hacia f cuando $n \to \infty$; en segundo lugar, al interpolar varias funciones sobre una misma red de nodos, puesto que en este caso se puede calcular una vez dos multiplicadores de Lagrange l_i y emplearlos para la interpolación de todas las funciones.

La representación en la forma de Newton resulta más cómoda en los cálculos prácticos. En efecto, el número de nodos a utilizar y el grado del polinomio interpolador con frecuencia no están conocidos de antemano y al pasar de n nodos a $n+1$ nodos en la forma de Newton se añade sólo un término que no es sino una corrección para el valor ya calculado. En cambio, la adición de un sumando más en la forma de Lagrange se acompaña del recálculo completo del resultado antes obtenido. Además, en la práctica de cálculos la interpolación suele realizarse sobre un pequeño segmento cuya longitud $h < 1$. En este caso los sumandos de la forma de Newton tienen el orden $h^0, h^1, h^2, \ldots$, o sea, están situados en el orden de decrecimiento lo que es útil al determinar la exactitud del resultado de interpolación.

§ 7.12. Interpolación práctica en las tablas

Por regla general, al interpolar en las tablas se utiliza la interpolación lineal o la cuadrática.

En caso de la interpolación lineal el valor que la función tiene en un punto, distinto de los nodos de interpolación, se determina por dos valores conocidos de la función a tabular $y_i = f(x_i)$, $y_{i+1} = f(x_{i+1})$ en los nodos de interpolación x_i y x_{i+1} entre los cuales está situado el valor del argumento x que nos interesa.

La fórmula de interpolación de Lagrange para la interpolación lineal tomará la forma

$$L_1(x) = y_i \frac{x - x_{i+1}}{x_i - x_{i+1}} + y_{i+1} \frac{x - x_i}{x_{i+1} - x_i}$$

y la primera fórmula de interpolación de Newton

$$N_1(x) = y_i + \frac{\Delta y_i}{h}(x - x_i),$$

donde $\Delta y_i = y_{i+1} - y_i$ es la diferencia finita primera en el punto x_i y $h = x_{i+1} - x_i$, el paso de interpolación.

Así pues, para obtener el valor aproximado de la función $y(x)$ con ayuda de la fórmula de Newton es suficiente añadir una corrección, igual a $\Delta y_i (x - x_i) h^{-1}$ al valor tabular y_i.

Ejemplo 1. Calcular cuántos grados se contiene en una medida en radianes igual a 0,222.

△ Hagamos uso de la tabla

Radianes	Grados
0,22	12,605
23	13,178
24	13,751

Para la interpolación lineal basta considerar los datos contenidos en las primeras dos filas. Planteemos la diferencia tabular

$$\Delta y_i = y_{i+1} - y_i = 13,178 - 12,605 = 0,573.$$

El paso de la tabla $h = 0,01$; $x - x_i = 0,222 - 0,220 = 0,002$. Calculamos la corrección

$$\frac{\Delta y_i}{h}(x - x_i) = \frac{0,573}{0,01} \cdot 0,002 = 0,1146$$

y añadímosla al valor tabular

$$y = 12,605 + 0,1146 = 12,7196.$$

Vamos a estimar el error del valor obtenido. Utilizando la fórmula (8) del § 7,5 para el error del método, tenemos

$$\Delta_1 = \frac{M_2}{2!}\,|(0,222 - 0,22)(0,222 - 0,23)|.$$

Puesto que la función a interpolar es $y = (180x/\pi)$, entonces $M_2 = 0$ y $\Delta_1 = 0$.

Hallamos el error de cálculo, teniendo en cuenta que el error de los datos iniciales es de 0,0005:

$$\Delta_2 = 0,0005 + \frac{0,001}{0,01} \cdot 0,002 = 0,0007.$$

Redondeando el valor y (0,222) hasta tres cifras después de la coma, obtenemos $\Delta_3 = 0,0004$.

Sumando todos los errores hallados, tenemos finalmente $y\,(0,222) = 12,720 \pm 0,0011$. ▲

En caso de la interpolación cuadrática se utilizan tres valores de la función tabulada $y_{-1} = f(x_{-1})$, $y_0 = f(x_0)$, $y_1 = f(x_1)$. En este caso el polinomio interpolador se construye en la forma de Lagrange (para los nodos no equidistantes) o bien en forma de Newton, cuando el punto de interpolación está situado más próximo a x_{-1} o a x_1 que a x_0, o bien en la forma de Stirling, cuando el punto de interpolación está cerca del nodo x_0.

Ejemplo 2. Para la función $y = \ln x$ según la tabla dada de sus valores

x	2	3	5
y	0,6931	1,0986	1,6094

construir el polinomio interpolador de Newton y obtener la estimación uniforme del error sobre el segmento [2, 5].

△ Ante todo, obtenemos las diferencias divididas:

$$f(2;\ 3) = \frac{1,0986 - 0,6931}{3-2} = 0,4055;$$

$$f(3;\ 5) = \frac{1,6092 - 1,0986}{5-3} = 0,2554;$$

$$f(2;\ 3;\ 5) = \frac{0,2554 - 0,4055}{5-2} = -0,0500.$$

Sustituyendo los valores hallados en la fórmula (2) del § 7.11, para $n = 2$, tenemos

$$N_2(x) = 0{,}6931 + 0{,}4055\,(x-2) - 0{,}0500\,(x-2)(x-3).$$

Utilizando la estimación (8) del § 7.5, para el error del método tenemos

$$\Delta_1 = \frac{M_3}{3!} \max_{[2,\,5]} |(x-2)(x-3)(x-5)|.$$

Luego hallamos

$$M_3 = \max_{[2,\,5]} \left|\frac{2}{x^3}\right| = \frac{1}{4}; \quad \max_{[2,\,5]} |(x-2)(x-3)(x-5)| \approx 2{,}2.$$

Ahora bien, $\Delta_1 = 0{,}1$.

Es evidente que el error de cálculo será despreciablemente pequeño en comparación con el del método. Por eso el error de interpolación máximo posible es de 0,1. ▲

§ 7.13. Método de iteración-interpolación de Aitken

En los casos en que falta la necesidad de obtener una expresión analítica aproximada de la función $f(x)$, representada tabularmente y se requiere sólo determinar el valor de esta función en cierto punto x^* distinto de los nodos de interpolación es conveniente utilizar el **método de iteración-interpolación de Aitken.** Dicho con propiedad este método consiste en interpolación lineal sucesiva. El proceso de cálculo de $f(x^*)$ radica en lo siguiente. Numeremos los nodos de interpolación, por ejemplo, en el orden en que éstos se alejan de x^*, y hagamos la tabla siguiente.

Tabla 7.12

$\vdots$	$\vdots$				
x_4	P_0^4				
		$P_1^{2,\,4}$			
x_2	P_0^2		$P_2^{0\,24}$		
		$P_1^{0,\,2}$		$P_3^{0\,124}$	
x_0	P_0^0		$P_2^{0\,12}$		$\cdots$
		$P_1^{1,\,0}$		$P_3^{0\,123}$	
x_1	P_0^1		$P_2^{0\,13}$		
		$P_1^{1,\,3}$			
x_3	P_0^3				
$\vdots$	$\vdots$				

Aquí:

$$P_0^k = f(x_k);$$

$$P_1^{ij}(x) = f_i\,\frac{x-x_j}{x_i-x_j} + f_j\,\frac{x-x_i}{x_j-x_i} = \frac{1}{x_j-x_i}\begin{vmatrix} x-x_i & P_0^i \\ x-x_j & P_0^j \end{vmatrix}$$

es un polinomio interpolador de grado no superior al primero, construido sobre los nodos x_i y x_j:

$$P_2^{ijk}(x)=\frac{1}{x_k-x_i}\begin{vmatrix} x-x_i & P_1^{ij} \\ x-x_k & P_1^{jk}\end{vmatrix}$$

es un polinomio interpolador de grado no superior al segundo, construido sobre los nodos x_i, x_j, x_k. Continuando este proceso, construimos el polinomio

$$P_n^{ij\ldots km}(x)=\frac{1}{x_m-x_i}\begin{vmatrix} x-x_i & P_{n-1}^{ij\ldots k}(x) \\ x-x_m & P_{n-1}^{j\ldots km}(x)\end{vmatrix}. \tag{1}$$

Mostremos que si $P_{n-1}^{ij\ldots k}(x)$ y $P_{n-1}^{j\ldots km}(x)$ son los polinomios interpoladores construidos sobre los nodos $x_i, x_j, \ldots, x_k$ y $x_j, \ldots, x_k, x_m$, respectivamente, entonces $P_n^{ij\ldots km}(x)$ es el polinomio interpolador construido sobre los nodos $x_i, x_j, \ldots, x_k, x_m$.

Efectivamente, en primer lugar, $P_n^{ij\ldots km}(x)$ es un polinomio cuyo grado no supera n, lo que es evidente de la construcción de la fórmula (1). En segundo lugar, en todos los nodos x_p el polinomio $P_n^{ij\ldots km}(x)$ toma el valor correspondiente:

$$P_n^{ij\ldots km}(x_i)=\frac{-(x_i-x_m)f_i}{x_m-x_i}=f_i \quad (x_p=x_i);$$

$$P_n^{ij\ldots km}(x_m)=\frac{(x_m-x_i)f_m}{x_m-x_i}=f_m \quad (x_p=x_m);$$

$$P_n^{ij\ldots km}(x_p)=\frac{1}{(x_m-x_i)}((x_p-x_i)f_p-(x_p-x_m)f_p)=f_p.$$

Calculando sucesivamente con ayuda de la fórmula (1) los valores de $P_n^{01\ldots n}(x^*)$, los toman por las aproximaciones sucesivas de $f(x^*)$. El proceso de cálculo termina prácticamente cuando el valor absoluto de la diferencia de dos aproximaciones sucesivas llega a ser suficientemente pequeño.

Ejemplo 1. Calcular en el punto $x^*=6$ con exactitud hasta $\varepsilon=0{,}05$ el valor de la función $f=\ln x$ representada en la forma de la tabla:

x	1	2	4	5	8	10
y	0,00	0,69	1,39	1,61	2,08	2,30

△ Numeremos los nodos en el orden siguiente: $x_0=5$, $x_1=8$, $x_2=4$, $x_3=10$, $x_4=2$, $x_5=1$. Con ayuda de la fórmula (1) calculamos los valores de los polinomios interpoladores P_n (6):

$$P_1^{01}=\frac{1}{8-5}\cdot\begin{vmatrix} 6-5 & 1{,}61 \\ 6-8 & 2{,}08\end{vmatrix}=1{,}77;$$

$$P_1^{02}=\frac{1}{5-4}\cdot\begin{vmatrix} 6-4 & 1{,}39 \\ 6-5 & 1{,}61\end{vmatrix}=1{,}83;$$

$$P_2^{012}=\frac{1}{8-4}\cdot\begin{vmatrix} 6-4 & 1{,}83 \\ 6-8 & 1{,}77\end{vmatrix}=1{,}80.$$

Puesto que $|P_1^{02} - P_2^{012}| = 0,03 < 0,05$, cesamos los cálculos y suponemos $\ln 6 = 1,80 \pm 0,03$. ▲

Ejemplo 2. Haciendo uso del esquema de Aitken calcular con exactitud de hasta $0,5\cdot 10^{-4}$ el valor de sen 0,674 para la función $y = \operatorname{sen} x$ representada en la forma tabular:

$x_0=0,66$	$x_1=0,67$	$x_2=0,68$
$y_0=0,61312$	$y_1=0,62099$	$y_2=0,62879$

△ Según la fórmula (1) tenemos

$$P_1^{01}(0,674) = \frac{\begin{vmatrix} 0,674-0,68 & 0,61312 \\ 0,674-0,67 & 0,62090 \end{vmatrix}}{0,67-0,66} = 0,625730;$$

$$P_1^{12} = \frac{1}{0,68-0,67}\cdot\begin{vmatrix} 0,674-0,67 & 0,62099 \\ 0,674-0,68 & 0,625643 \end{vmatrix} = 0,625643;$$

$$P_2^{012} = \frac{1}{0,68-0,66}\cdot\begin{vmatrix} 0,674-0,66 & 0,625730 \\ 0,674-0,68 & 0,625643 \end{vmatrix} = 0,625676.$$

Por lo tanto, sen $0,674 = 0,62568 + 0,00004$. ▲

§ 7.14. Optimización de los nodos de interpolación

Volvamos a considerar la estimación del error, expresada por la fórmula (8) del § 7.5. Supongamos ahora que no hay restricciones en la elección de la red Λ_n. Planteemos el problema acerca de la mejor elección de los nodos de interpolación. En la clase dada de funciones $C^{n+1}([a, b])$, partiendo de la estimación (8) del § 7.5, como mejores nodos de interpolación han de reconocerse los x_i para los cuales la expresión $\max\limits_{[a,b]} |\omega_n(x)|$ es mínima. La determinación de estos nodos se reduce de hecho a la búsqueda de las raíces del polinomio que se desvía mínimamente del cero sobre el segmento $[a, b]$. Según se sabe de la teoría de las funciones, tal polinomio se genera por *polinomios de Chébyshev de primer género* que se definen por las relaciones recurrentes siguientes:

$$T_0 = 1;\quad T_1 = x;\quad T_{n+1} = 2xT_n - T_{n-1},\ n > 0. \tag{1}$$

Consideremos las propiedades fundamentales de los polinomios de Chébyshev.

1°. *El término de mayor grado del polinomio T_{n+1} se obtiene del término de mayor grado del polinomio T_n ($n = 1, 2, \ldots$) multiplicando este último por $2x$.*

□ Tenemos:

$$T_2 = 2x\cdot x - 1 = 2x^2 - 1,$$
$$T_3 = 2x(2x^2 - 1) - x = 4x^3 - 3x.$$

Por eso el término de mayor grado del polinomio T_{n+1} es igual a $2^n x^{n+1}$. La forma general del polinomio T_{n+1} es tal:

$$T_{n+1} = 2^n x^{n+1} + \ldots \ ■$$

2°. Todos los *polinomios* $T_{2n}(x)$ *son funciones pares y* $T_{2n+1}(x)$, *funciones impares.*

□ Para $n = 0$ esto es evidente. Supongamos que esto es válido para cierto n. Entonces la función $2xT_{2n+1}(x)$ es par y, por lo tanto, $T_{2n+2} = 2xT_{2n+1}(x) - T_{2n}(x)$ también es una función par. Luego, la función $2xT_{2n+2}(x)$ es impar y por eso $T_{2n+3}(x) = 2xT_{2n+2}(x) - T_{2n+1}(x)$ es también una función impar. ■

3°. *Si* $x \in]-1, 1[$, *entonces los polinomios de Chébyshev tienen la expresión explícita siguiente*:

$$T_{n+1}(x) = \cos[(n+1)\arccos x]; \quad n \geqslant -1. \tag{2}$$

□ Mostremos que el segundo miembro de la igualdad (2) satisface la definición (1) del polinomio de Chébyshev. En efecto,

$$\{\cos[(n+1)\arccos x]\}_{n=-1} = 1 = T_0;$$
$$\{\cos[(n+1)\arccos x]\}_{n=0} = x = T_1.$$

Para demostrar el cumplimiento de la fórmula recurrente consideremos la relación trigonométrica evidente

$$\cos(n+1)\theta = 2\cos\theta\cos n\theta - \cos(n-1)\theta.$$

Suponiendo $\theta = \arccos x$, de donde $x = \cos\theta$, obtenemos $\cos[(n+1)\arccos x] = 2x\cos[n\arccos x] - \cos[(n-1)\arccos x]$. ■

4°. *Sobre el segmento* $[-1, 1]$ *los polinomios* $T_{n+1}(x)$ *tienen* $n+1$ *raíces diferentes*:

$$x_k = \cos\frac{2k+1}{2(n+1)}\pi \quad (k = 0, 1, \ldots, n). \tag{3}$$

□ Utilizando la expresión (2), para determinar las raíces del polinomio $T_{n+1}(x)$ obtenemos la ecuación

$$(n+1)\arccos x_k = \frac{\pi}{2} + \pi k \ (k = 0, 1, \ldots, n).$$

Despejando en esta ecuación x_k, llegamos a la relación (3). ■

5°. *Sobre el segmento* $[-1, 1]$ *es válida la desigualdad*

$$[T_{n+1}(x)] \leqslant 1. \tag{4}$$

Esto se desprende inmediatamente de la relación (2).

De la misma relación (2) determinemos todos los puntos x_m en los cuales el polinomio $T_{n+1}(x)$ alcanza sus valores extremos ± 1. Para esto es necesario que

$$(n+1)\arccos x_m = \pi m \ (m = 0, 1, \ldots, n+1)$$

y, por consiguiente,

$$x_m = \cos\frac{m}{n+1}\pi \ (m = 0, 1, \ldots, n+1). \tag{5}$$

Sustituyendo estos valores en la igualdad (2), obtenemos

$$T_{n+1}(x_m) - \cos m\pi = (-1)^m. \tag{6}$$

Esto quiere decir que los puntos en los cuales $T_{n+1} = 1$ y $T_{n+1} = -1$ alternan comenzando con $x_0 = 1$, donde $T_{n+1}(1) = 1$. Nótese una vez más que la desigualdad (4) es válida no para todas las x. Si $|x| > 1$, entonces $\arccos x$ no existe sobre el conjunto de los números reales.

Examinemos ahora los polinomios

$$\overline{T}_{n+1}(x) = 2^{-n}T_{n+1}(x) = x^{n+1} + \ldots \qquad (7)$$

Son los polinomios que se desvían mínimamente del cero sobre el segmento $[-1, 1]$ lo que confirma el teorema siguiente.

Teorema. *Sea $P_{n+1}(x)$ el polinomio de grado $n+1$ con el coeficiente superior igual a 1. Entonces*

$$\max_{[-1,1]} |P_{n+1}(x)| \geqslant \max_{[-1,1]} |\overline{T}_{n+1}(x)| = 2^{-n}. \qquad (8)$$

□ Supongamos que la desigualdad (8) no se cumple. Entonces el polinomio $Q_n(x) = \overline{T}_{n+1} - P_{n+1}$, cuyo grado no supera n, habría coincidido en todos $n+2$ puntos extremos x_m del polinomio T_{n+1} con este último en signo y por lo tanto, habría tomado alternadamente en estos puntos una vez el valor positivo, otra vez el valor negativo. Por eso $Q_n(x)$ debería tener $n+1$ raíces diferentes lo que es imposible para un polinomio cuyo grado no supera n. La contradicción obtenida demuestra el teorema. ■

Todo segmento $[a, b]$ puede ser obtenido del segmento $[-1, 1]$ por el reemplazo lineal de las variables

$$x' = \frac{b+a}{2} + \frac{b-a}{2}x. \qquad (9)$$

En este caso el polinomio $\overline{T}_{n+1}(x)$ se transforma en polinomio $\overline{T}_{n+1}\left(\frac{2x-(b+a)}{b-a}\right)$ con el coeficiente superior $\left(\frac{2}{b-a}\right)^{n+1}$.

Por consiguiente,

$$\overline{T}_{n+1}^{[a,b]}(x) = (b-a)^{n+1}2^{-2n-1}T_{n+1}\left(\frac{2x-(b+a)}{b-a}\right) \qquad (10)$$

es un polinomio, con el coeficiente superior igual a 1, el cual se desvía mínimamente del cero sobre $[a, b]$ y para todo polinomio $P_{n+1}(x)$ de grado $n+1$, con el coeficiente superior igual a 1, se cumple la desigualdad

$$\max_{[a,b]} |P_{n+1}(x)| \geqslant \max_{[a,b]} (x) |\overline{T}_{n+1}^{[a,b]}(x)| = 2^{-n}\left(\frac{b-a}{2}\right)^{n+1}. \qquad (11)$$

En virtud del reemplazo lineal (9) las raíces del polinomio $\overline{T}_{n+1}^{[a,b]}(x)$ tienen la forma

$$x_k = \frac{b+a}{2} + \frac{b-a}{2}\cos\frac{2k+1}{2(n+1)}\pi \quad (k = 0, \ldots, n) \qquad (12)$$

y los puntos extremos tienen la forma

$$x_m = \frac{b+a}{2} + \frac{b-a}{2} \cos \frac{m}{n+1} \pi \quad (m=0,\ 1,\ \ldots,\ n+1). \qquad (13)$$

Volvamos ahora a considerar el problema de minimización del error de interpolación Δ_1 sobre el segmento $[a, b]$ para una red arbitraria en la clase de funciones continuamente derivables $n+1$ veces que satisfacen la condición (7) del § 7.5. Vamos a designar esta clase de funciones $C^{n+1}(M_{n+1}, [a, b])$. Para resolver el problema planteado en virtud de la fórmula (8) del § 7.5 es necesario minimizar la magnitud $\max\limits_{[a,b]} |\omega_n(x)|$. Puesto que $\omega_n(x)$ es el polinomio de grado $n+1$ con el coeficiente superior igual a 1, es evidente que la magnitud $\max\limits_{[a,b]} |\omega_n(x)|$ alcanza su valor mínimo para los polinomios de Chébyshev $\overline{T}_{n+1}^{[a,b]}(x)$. Por lo tanto, en calidad de nodos de interpolación conviene elegir los puntos x_k definidos por la expresión (12). En este caso

$$\max_{[a,b]} |\omega_n(x)| = \max_{[a,b]} |\overline{T}_{n+1}^{[a,b]}(x)| = 2^{-n} \left(\frac{b-a}{2}\right)^{n+1} \qquad (14)$$

y la estimación (8) del § 7.5 toma la forma

$$\Delta_1 \leqslant \frac{M_{n+1}}{(n+1)!} 2^{-n} \left(\frac{b-a}{2}\right)^{n+1}. \qquad (15)$$

Esta estimación es no mejorable, ya que en ella tiene lugar el signo de igualdad si en calidad de función $f(x)$ se elige el siguiente polinomio de grado $n+1$:

$$f(x) = \frac{M_{n+1}}{(n+1)!} x^{n+1} + a_n x^n + \ldots$$

y en calidad de nodos de interpolación, los puntos x_k definidos por la expresión (12).

Ejemplo 1. Sobre el segmento $[-1, 1]$ obtener la estimación uniforme de desviación de la función $f(x) = 1 - \cos(\pi x/2)$ respecto a su polinomio interpolador construido sobre los nodos de Chébyshev (3) para $n = 2, 3, 4$.

△ Nótese, ante todo, que para la función en cuestión sobre el intervalo dado $M_{n+1} = (\pi/2)^{n+1}$, $b - a = 2$. Por eso en virtud de la estimación (15) tenemos:

$$n=2,\ \Delta_1 \leqslant \left(\frac{\pi}{2}\right)^3 \frac{1}{3!} \left(\frac{1}{2}\right)^2 \approx 0{,}17;$$

$$n=3,\ \Delta_1 \leqslant \left(\frac{\pi}{2}\right)^4 \frac{1}{4!} \left(\frac{1}{2}\right)^3 \approx 0{,}032;$$

$$n=4,\ \Delta_1 \leqslant \left(\frac{\pi}{2}\right)^5 \frac{1}{5!} \left(\frac{1}{2}\right)^4 \approx 0{,}005.$$

Recomendamos comparar la solución obtenida con la del ejemplo 1 del § 7.5. ▲

§ 7.15. Interpolación con nodos múltiplos

Hasta ahora hemos considerado el problema en el cual los parámetros de interpolación—los coeficientes del polinomio interpolador—se determinaban sólo por los valores de la función a interpolar. Tal problema suele llamarse *problema de interpolación según Lagrange* y el mismo proceso de construcción del polinomio interpolador, *proceso de Lagrange.*

Examinemos ahora un problema algo más amplio: el problema de interpolación respecto a los valores de la función y de sus derivadas o, como se dice de otro modo, el *problema de interpolación múltipla.*

Supongamos que sobre la red Λ_m: $a \leqslant x_0 < x_1 < \ldots < x_m \leqslant b$ en los nodos x_i se dan los valores f_i de cierta función f y las derivadas de esta última $f_i^{(k)}$ $(i = 0, 1, \ldots, m;\ k = 0, 1, \ldots, \alpha_i - 1)$, con la particularidad de que $\sum\limits_{i=0}^{m} \alpha_i = n + 1$. Se necesita construir el polinomio H_n cuyos valores y derivadas hasta el orden $\alpha_i - 1$ en los nodos x_i $(i = 0, 1, \ldots, m)$ coinciden con los valores de f y con las derivadas correspondientes de esta última, así como estimar el error.

Tal tipo de interpolación se llama *interpolación según Hermite* y el polinomio respectivo H_n, *polinomio de Hermite*. Los números α_i se denominan *multiplicidades de los nodos* x_i. En este caso se puede demostrar que el polinomio de Hermite existe y es único.

El término residual de la fórmula de interpolación $f(x) \approx H_n(x)$ se puede representar en la forma siguiente:

$$R_n(x) = f(x) - H_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{i=0}^{m} (x - x_i)^{\alpha_i};\quad \xi \in (a, b). \tag{1}$$

Sea ahora, para precisar,

$$|f^{(n+1)}(x)| \leqslant M_{n+1},\quad x \in [a, b]. \tag{2}$$

Utilizando esta restricción y la fórmula (1), obtenemos la estimación del error para el punto fijo x:

$$\Delta_1 = |f(x) - H_n(x)| \leqslant \frac{M_{n+1}}{(n+1)!} \prod_{i=0}^{m} |x - x_i|^{\alpha_i}. \tag{3}$$

Ahora no es difícil construir una estimación uniforme sobre todo el segmento $[a, b]$ para la red fija Λ_m. En efecto,

$$\Delta_1 = \max_{[a, b]} |R_n(x)| \leqslant \frac{M_{n+1}}{(n+1)!} \max_{[a, b]} |\Omega_n(x)|, \tag{4}$$

donde

$$\Omega_n(x) = \prod_{i=0}^{m} (x - x_i)^{\alpha_i}. \tag{5}$$

Ejemplo. Construir el polinomio interpolador de Hermite para la función $f = 1 - \cos(\pi x/2)$ sobre los nodos $x_0 = -1$, $x_1 = 0$, $x_2 = 1$ con las multiplicidades $\alpha_0 = 1$, $\alpha_1 = 2$, $\alpha_2 = 1$, respectivamente. Obtener una estimación uniforme del error sobre el segmento $[-1, 1]$.

△ Calculemos en los nodos dados los valores de la función y de su derivada: $f(x_0) = f(x_2) = 1$; $f(x_1) = f'(x_1) = 0$. Ahora vamos a construir el polinomio de Hermite teniendo en cuenta las multiplicidades de los nodos:

$$H_3(x) = 1 \cdot \frac{x^2(x-1)}{-2} + 0 \cdot \frac{(x-1)(x+1)}{-1} + 1 \cdot \frac{(x+1)x^2}{2} = x^2.$$

Nótese que en vez del polinomio de tercer grado hemos obtenido el de segundo grado lo que es consecuencia de la simetría de la información inicial (pero no de la función f).

Hallamos ahora la estimación del error. Utilizando la fórmula (4) y teniendo en cuenta que para la función dada $M_4 = (\pi/2)^4$, obtenemos

$$\Delta_1 \leqslant \left(\frac{\pi}{2}\right)^4 \cdot \frac{1}{4!} \cdot \max_{[-1,\,1]} |(x+1)\,x^2(x-1)|.$$

No es difícil mostrar que $\max\limits_{[-1,\,1]} |(x+1)\,x^2(x-1)| = 0{,}25$.

Por eso finalmente tenemos $\Delta_1 \leqslant 0{,}065$. ▲

§ 7.16. Aparato matemático de la interpolación trigonométrica

En los párrafos siguientes consideraremos el problema de aproximación de las funciones con ayuda del polinomio trigonométrico. Esto quiere decir que en calidad de función aproximante se toma la combinación lineal de las funciones trigonométricas sen nx y cos nx.

Para fundamentar formalmente la elección de las funciones trigonométricas en calidad de aproximantes necesitaremos algunas nociones del curso de análisis matemático que se refieren a las series de Fourier. A continuación citamos estas nociones omitiendo las demostraciones de las afirmaciones respectivas.

Sucesiones. Series. Consideremos cierta función $f(x)$. En calidad de dominio de definición de esta función tomemos el conjunto de los números naturales, o sea, el argumento x toma los valores $1, 2, \ldots, n$. Tal función se llama *sucesión*.

La sucesión se escribe en la forma

$$a_1, a_2, \ldots, a_{n-1}, a_n, a_{n+1}, \ldots, \text{ o bien } \{a_n\},$$

con la particularidad de que a_n se llama *término general*; a_{n-1}, término precedente a a_n; a_{n+1}, término siguiente a a_n.

Citemos algunos ejemplos de las sucesiones.

1) La sucesión 1, 2, 3, . . ., n, . . ., cuyo término general es $a_n = n$ se llama *serie de los números naturales.*

2) La sucesión a_1, a_2, . . ., a_{n-1}, a_n, . . . para la cual $a_n - a_{n-1} = d$, donde d es una magnitud constante, se denomina *progresión aritmética.* Para definir la progresión aritmética basta conocer su primer término a_1 y la diferencia de la progresión d. En efecto, el término general se expresa por la fórmula

$$a_n = a_1 + d(n - 1).$$

Puesto que con ayuda de la fórmula citada se puede hallar todo término de la sucesión, sustituyendo los valores $n = 1, 2, 3$, . . . la sucesión se considera definida.

3) La sucesión b_1, b_2, . . ., b_{n-1}, b_n, . . . para la cual $b_n = b_{n-1}q$, donde q es una magnitud constante, se llama *progresión geométrica.* Para definir una progresión geométrica basta conocer su primer término b_1 y la razón q. El término general se expresa por la fórmula

$$b_n = b_1 q^{n-1}.$$

4) La sucesión c_1, c_2, . . ., c_n, . . . para la cual $c_n = c$, donde c una magnitud constante, se denomina *sucesión constante.*

5) Consideremos un ejemplo más de la sucesión. Vamos a calcular el número e sucesivamente con una cifra, con dos cifras, con tres cifras, etc. Los resultados de cálculos pueden ser representados en la forma siguiente:

$$2;\ 2{,}7;\ 2{,}71;\ 2{,}718;\ \ldots;$$
$$(1);\ (2);\ (3);\ (4);\ \ldots$$

Al numerar los valores obtenidos por los números naturales de la serie, según se muestra entre paréntesis, obtenemos una sucesión.

Además de las sucesiones numéricas, consideraremos también las *sucesiones funcionales.*

Vamos a citar algunos ejemplos de sucesiones funcionales:

$$1)\ a_0,\ a_1x,\ a_2x^2,\ \ldots,\ a_{n-1}x^{n-1},\ \ldots;$$
$$2)\ \operatorname{sen} x,\ \operatorname{sen} 2x,\ \operatorname{sen} 3x,\ \ldots,\ \operatorname{sen} nx,\ \ldots$$

Recuérdese la definición del límite de una sucesión numérica. El número A se llama *límite de la sucesión* $\{a_n\}$ si para todo $\varepsilon > 0$ existe un tal número N que para todos los $n > N$ se cumpla la desigualdad $|a_n - A| < \varepsilon$. En este caso se escribe $\lim\limits_{n\to\infty} a_n = A$.

La sucesión que tiene el límite se llama *convergente* y en el caso contrario, *divergente.*

Ejemplo 1. Mostrar que la progresión geométrica b, bq, bq^2, ..., bq^{n-1}, . . . para $|q| < 1$ es una sucesión convergente y para $|q| \geqslant 1$, una sucesión divergente.

△ 1) Primero consideremos el caso cuando $|q| < 1$. Mostremos que

$$\lim_{n\to\infty} bq^{n-1} = 0,$$

o sea, según el dado $\varepsilon > 0$ hallemos un N tal que para $n > N$ se cumpla la desigualdad $|bq^{n-1} - 0| < \varepsilon$. Para esto resolvamos la desigualdad respecto a n. Escribámosla en la forma

$$|b|\,|q|^{n-1} < \varepsilon \text{ o bien } |q|^{n-1} \leqslant \varepsilon/|b|.$$

Logaritmemos la última desigualdad

$$(n-1)\ln|q| < \ln(\varepsilon/|b|).$$

Dividiendo ambos sus miembros por el número negativo $\ln|q|$, obtenemos

$$n-1 > \frac{\ln(\varepsilon/|b|)}{\ln|q|}, \text{ o sea, } n > 1 + \frac{\ln(\varepsilon/|b|)}{\ln|q|}.$$

Es evidente que en calidad de N es suficiente tomar

$$N = E\left(1 + \frac{\ln(\varepsilon/|b|)}{\ln|q|}\right),$$

donde $E(x)$ designa el número entero máximo que no supere x.

2) Examinemos ahora el caso cuando $q > 1$, $b > 0$. Mostremos que la sucesión es divergente. Para esto basta mostrar que para todo M grande como se quiera existe un N tal que para todos los $n > N$ se cumpla la desigualdad $bq^{n-1} > M$. Despejando en la última desigualdad n, obtenemos

$$n > 1 + \frac{\ln(M/b)}{\ln q}.$$

En calidad de N tomamos

$$N = E\left(1 + \frac{\ln(M/b)}{\ln q}\right).$$

Nótese que para $q = 1$ la sucesión es constante y $\lim\limits_{n\to\infty} bq^{n-1} = b$.

Se puede mostrar que en todos los demás casos la sucesión se diverge. ▲

Supongamos que hay una sucesión $a_1, a_2, \ldots, a_n, \ldots$. La expresión que tiene la forma

$$a_1 + a_2 + \ldots + a_n + \ldots = \sum_{n=1}^{\infty} a_n$$

se llama *serie*; aquí a_n es el n-ésimo término de la serie.

La suma de los primeros n términos de la serie se denomina *n-ésima suma parcial* de esta última:

$$S_n = \sum_{i=1}^{n} a_i.$$

Se llama *suma de la serie* el límite de la sucesión de sus sumas parciales

$$\lim_{n\to\infty} S_n = S.$$

Si la serie tiene la suma, se llama *convergente*, en el caso contrario se dice que la serie *diverge.*

Citemos algunos ejemplos de las series convergentes y divergentes.

Ejemplo 2. Consideremos una serie que se obtiene de la progresión aritmética y escribamos su n-ésima suma parcial

$$S_n = \frac{a_1 + a_n}{2} \cdot n = \frac{2a_1 + d(n-1)}{2} \cdot n = \frac{d}{2} n^2 + \frac{2a_1 - d}{2} n.$$

Es evidente que cuando n tiende hacia el infinito esta suma parcial crece indefinidamente en valor absoluto y, por consiguiente, la serie dada diverge.

Ejemplo 3. Consideremos una serie formada por la progresión geométrica para $|q| < 1$ y determinemos la suma de aquélla. Hagamos uso de la fórmula de la suma de n términos de la progresión geométrica:

$$S_n = \frac{b_1(1-q^n)}{1-q} = \frac{b_1}{1-q} - \frac{b_1}{1-q} q^n.$$

Antes hemos demostrado (véase el ejemplo 1) que $\lim\limits_{n\to\infty} q^n = 0$ $(|q| < 1)$; por lo tanto,

$$S = \lim_{n\to\infty} S_n = \frac{b_1}{1-q}.$$

Las series cuyos términos son funciones se llaman *series funcionales.* Tales son, por ejemplo, las series

$$a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1} + \ldots,$$
$$b_0 + b_1 \cos x + b_2 \cos 2x + \ldots + b_{n-1} \cos(n-1)x + \ldots.$$

La primera de ellas se denomina serie *de potencias* y la segunda, serie *trigonométrica.*

Consideremos la serie funcional

$$u_1(x) + u_2(x) + \ldots + u_n(x) + \ldots,$$

donde $u_n(x)$ son las funciones definides sobre el segmento $[a, b]$. Sea $x_0 \in [a, b]$; entonces la serie

$$u_1(x_0) + u_2(x_0) + \ldots + u_n(x_0) + \ldots$$

es una serie numérica y puede resultar convergente o divergente.

El conjunto de todos los valores de $x \in [a, b]$ para los cuales converge la serie numérica correspondiente se denomina *dominio de convergencia* de la serie funcional.

Es evidente que

$$S(x) = \lim_{n\to\infty} S_n(x),$$

donde $S_n(x) = \sum_{i=1}^{n} u_i(x)$ depende de la elección de la variable x, o sea, la suma $S(x)$ de la serie funcional es función del punto x.

Sea $\{S_n(x)\}$ la sucesión de las sumas parciales de la serie funcional definidas sobre el mismo intervalo cerrado $[a, b]$. La serie funcional se denomina *uniformemente convergente* hacia la función $S(x)$ definida sobre $[a, b]$ si a todo $\varepsilon > 0$ se puede hacer corresponder un tal número N, no dependiente de $x \in [a, b]$, que para todo $n > N$ se cumpla la desigualdad $|S_n(x) - S(x)| < \varepsilon$.

Examinemos un ejemplo que ilustra la diferencia existente entre los conceptos de convergencia de la serie y de convergencia uniforme de la misma.

Ejemplo 4. Para la serie $\sum_{n=1}^{\infty} \frac{x}{(1+x)^n}$, donde $0 \leqslant x \leqslant 1$, la suma parcial es

$$S_n(x) = \sum_{i=1}^{n} \frac{x}{(1+x)^i} = 1 - \frac{1}{(1+x)^n}.$$

Mostremos que en el segmento dado la serie converge.

En efecto, si $x = 0$, entonces $S_n(0) = S(0) = 0$. Sea ahora $x > 0$. Demostremos que en este caso $S(x) = 1$, o sea, según el dado $\varepsilon > 0$ determinemos un N tal que para $n > N$ se cumpla la desigualdad

$$\left| 1 - \frac{1}{(1+x)^n} - 1 \right| < \varepsilon.$$

Resolviendo esta desigualdad, obtenemos

$$N = E\left(-\frac{\ln \varepsilon}{\ln(1+x)} \right).$$

De la última igualdad se ve que, hablando en general, N depende no sólo de ε sino también de x, con la particularidad de que para un mismo ε, si x se aproxima a cero, N crece indefinidamente. Esto quiere decir que según el ε dado no podemos elegir el N *único para* todos los $x \in [0, 1]$, con otras palabras, la serie *no es uniformemente convergente sobre el segmento señalado.*

El ejemplo citado muestra que la sucesión de las sumas parciales, continuas sobre cierto segmento, puede converger hacia una función discontinua sobre este segmento. Una de las causas que originan la introducción del concepto de convergencia uniforme de una serie consiste en que una serie uniformemente convergente de las funciones continuas tiene como su suma también una función continua.

Desarrollo de las funciones en serie de Fourier. Muchos problemas de la ciencia y la técnica están vinculados con las funciones periódicas que reflejan procesos cíclicos.

La función $f(x)$ se llama *periódica* con el período $T > 0$ si satisface la igualdad

$$f(x) = f(x + T). \tag{1}$$

Partiendo de las consideraciones prácticas es cómodo, con un grado suficiente de precisión, representar tales funciones en forma de una serie trigonométrica o de su suma parcial.

La serie funcional que tiene la forma

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \operatorname{sen} nx) \tag{2}$$

se llama *trigonométrica*, con la particularidad de que a_n y b_n son los números reales que no dependen de x.

Supongamos que esta serie converge para todo x que está en el intervalo $[-\pi, \pi]$; entonces ella define la función periódica $f(x)$ de período $T = 2\pi$.

La serie que tiene la forma (2) se denomina *serie de Fourier* para la función $f(x)$, integrable sobre el segmento $[-\pi, \pi]$, si los coeficientes de esta serie se calculan con ayuda de las fórmulas siguientes:

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\,dx, \tag{3}$$

$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx\,dx \; (n = 1, 2, \ldots). \tag{4}$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \operatorname{sen} nx\,dx \; (n = 1, 2, \ldots). \tag{5}$$

Ahora bien, podemos formalmente considerar la serie de Fourier para la función dada $f(x)$. No obstante, en este caso surgen las preguntas siguientes: 1) ¿converge o no la serie de Fourier de la función $f(x)$? 2) si la serie converge ¿tendrá ésta como su suma $f(x)$? Es el teorema de Dirichlet que da las respuestas a estas preguntas. Antes de pasar al enunciado del mismo teorema, recordemos algunos conceptos.

La función $f(x)$ se considera *monótona* sobre un intervalo si para todos x_1 y x_2, pertenecientes a este intervalo y tales que $x_1 < x_2$, se cumple sólo una de las desigualdades $f(x_1) \leqslant f(x_2)$ o bien $f(x_1) \geqslant f(x_2)$.

La función $f(x)$ se dice *monótona a trozos* sobre un intervalo si éste puede ser dividido en un número finito de intervalos abiertos en cada uno de los cuales la función es monótona.

La función $f(x)$ se considera *continua a trozos* sobre un intervalo si tiene sobre éste un número finito de puntos de discontinuidad.

Designemos con $f(a+0)$ el límite de la función $f(x)$ si x tiende a a del lado derecho (límite por la derecha) y, respectivamente, designemos con $f(a-0)$ el límite por la izquierda.

Teorema de Dirichlet. *Si una función $f(x)$, dada sobre el segmento $[-\pi, \pi]$, es monótona a trozos y continua a trozos, la serie de Fourier de esta función converge sobre todo el segmento $[-\pi, \pi]$ y la suma de la serie es igual:*

1) a *$f(x)$ en todos los puntos de continuidad pertenecientes a $(-\pi, \pi)$;*

2) a $\frac{1}{2}[f(x-0)+f(x+0)]$ *en todos los puntos de discontinuidad pertenecientes a $(-\pi; \pi)$;*

3) a $\frac{1}{2}[f(-\pi+0)+f(\pi-0)]$ *en los extremos del segmento, es decir, en los puntos $x=-\pi$ y $x=\pi$.*

A continuación escribiremos que

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty}(a_n \cos nx+b_n \operatorname{sen} nx) \tag{6}$$

según el sentido del teorema de Dirichlet recién enunciado.

El teorema de Dirichlet no afirma que la serie de Fourier converge uniformemente hacia la función $f(x)$. No obstante, si se refuerzan las propiedades a las cuales debe satisfacer la función, es decir, exigir de ella que sea continua sobre todo el segmento $[-\pi, \pi]$ y sea monótona a trozos sobre éste, así como que se cumpla la igualdad $f(-\pi)=f(\pi)$, entonces para tal función la serie de Fourier convergirá uniformemente hacia la función $f(x)$ sobre todo el segmento $[-\pi, \pi]$.

Se puede mostrar que *para la función par todos los coeficientes b_n son iguales a cero y la serie de Fourier correspondiente no contiene senos:*

$$f(x)=\frac{a_0}{2}+\sum_{n=1}^{\infty} a_n \cos nx, \tag{7}$$

donde

$$a_n=\frac{2}{\pi}\int_0^{\pi} f(x)\cos nx\, dx \ (n=0,\ 1,\ 2,\ \ldots,). \tag{8}$$

Análogamente *para la función impar todos los coeficientes a_n son iguales a cero y la serie de Fourier correspondiente no contiene cosenos:*

$$f(x)=\sum_{n=1}^{\infty} b_n \operatorname{sen} nx, \tag{9}$$

donde

$$b_n=\frac{2}{\pi}\int_0^{\pi} f(x)\ \text{sen}\ nx\,\mathrm{d}x\ (n=1,\ 2,\ \ldots). \tag{10}$$

Ejemplo 1. Desarrollar en serie de Fourier la función

$$f(x)=\begin{cases} -x & \text{para}\quad -\pi\leqslant x\leqslant 0, \\ 2x & \text{para}\quad 0<x\leqslant\pi. \end{cases}$$

△ Determinemos los coeficientes de Fourier de la función $f(x)$. Con ayuda de las fórmulas (3) y (4) hallamos los coeficientes a_0 y a_n:

$$a_0=\frac{1}{\pi}\int_{-\pi}^{0}(-x)\,\mathrm{d}x+\frac{1}{\pi}\int_0^{\pi}2x\,\mathrm{d}x=\frac{1}{\pi}\cdot\frac{\pi^2}{2}+\frac{1}{\pi}\cdot\pi^2=\frac{3}{2}\pi;$$

$$a_n=\frac{1}{\pi}\int_{-\pi}^{0}(-x)\cos nx\,\mathrm{d}x+\frac{1}{\pi}\int_0^{\pi}2x\cos nx\,\mathrm{d}x.$$

Integrando por partes obtenemos

$$a_n=-\frac{1}{\pi}\left(\frac{1}{n}x\ \text{sen}\ nx\Big|_{-\pi}^{0}-\frac{1}{n}\int_{-\pi}^{0}\text{sen}\ nx\,\mathrm{d}x-\right.$$

$$\left.-\frac{2}{n}x\ \text{sen}\ nx\Big|_0^{\pi}+\frac{2}{n}\int_0^{\pi}\text{sen}\ nx\,\mathrm{d}x\right)=-$$

$$-\frac{1}{\pi n}\left(\frac{1}{n}\cos nx\Big|_{-\pi}^{0}-\frac{2}{n}\cos nx\Big|_0^{\pi}\right)=-\frac{3}{\pi n^2}[1-(-1)^n],$$

o sea,

$$a_{2k}=0,\ a_{2k-1}=-\frac{6}{\pi(2k-1)^2}\ (k=1,\ 2,\ 3,\ \ldots).$$

Obtenemos los coeficientes b_n con ayuda de la fórmula (5):

$$b_n=\frac{1}{\pi}\int_{-\pi}^{0}(-x)\,\text{sen}\ nx\,dx+\frac{1}{\pi}\int_0^{\pi}2x\ \text{sen}\ nx\,\mathrm{d}x=$$

$$=-\frac{1}{\pi}\left(-\frac{x}{n}\cos nx\Big|_{-\pi}^{0}+\frac{1}{n}\int_0^{\pi}\cos nx\,dx+\frac{2x}{n}\cos nx\Big|_0^{\pi}-\right.$$

$$\left.-\frac{2}{n}\int_0^{\pi}\cos nx\,\mathrm{d}x\right)=-\frac{1}{\pi}\left[-\frac{\pi}{n}(-1)^n+2\frac{\pi}{n}(-1)^n\right]=\frac{(-1)^{n+1}}{n},$$

o sea,

$$b_{2k}=-\frac{1}{2k},\ b_{2k-1}=\frac{1}{2k-1}\ (k=1,\ 2,\ 3,\ \ldots).$$

Ahora bien, la serie de Fourier de la función dada tiene la forma

$$f(x)=\frac{3}{4}\pi-\frac{6}{\pi}\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2}\cos(2k-1)x+$$

$$+\sum_{k=1}^{\infty}\frac{1}{2k-1}\operatorname{sen}(2k-1)x-\frac{1}{2}\sum_{k=1}^{\infty}\frac{1}{k}\operatorname{sen}2kx.$$

En el intervalo $(-\pi, \pi)$ la serie converge hacia la función $f(x)$ y en los puntos $x=\pm\pi$, hacia el número $\frac{1}{2}[f(-\pi+0)+$ $+f(\pi-0)]=\frac{3}{2}\pi$. ▲

Ejemplo 2. Desarrollar en serie de Fourier la función

$$f(x)=\begin{cases}-\operatorname{sen}x & \text{para } -\pi\leqslant x\leqslant 0,\\ \operatorname{sen}x & \text{para } \quad 0<x\leqslant\pi.\end{cases}$$

△ La función dada es par, por consiguiente, todos los coeficientes $b_n=0$ y a_n se determinan con ayuda de la fórmula (8):

$$a_n=\frac{2}{\pi}\int_0^{\pi}\operatorname{sen}x\cos nx\,dx\quad(n=0,\ 1,\ 2,\ \ldots).$$

De aquí

$$a_0=\frac{2}{\pi}\int_0^{\pi}\operatorname{sen}x\,dx=\frac{4}{\pi}.$$

Luego,

$$a_n=\frac{2}{\pi}\int_0^{\pi}\operatorname{sen}x\cos nx\,dx=\frac{1}{\pi}\int_0^{\pi}[\operatorname{sen}(n+1)x-\operatorname{sen}(n-1)x]\,dx=$$

$$=\begin{cases}0 & \text{para } n=2k-1,\\ -\dfrac{4}{\pi(n^2-1)} & \text{para } n=2k\ (k=1,\ 2,\ 3,\ \ldots).\end{cases}$$

Por lo tanto,

$$f(x)=|\operatorname{sen}x|=\frac{2}{\pi}-\frac{4}{\pi}\sum_{k=1}^{\infty}\frac{1}{4k^2-1}\cos 2kx.$$

Nótese que la serie obtenida converge hacia la función $|\operatorname{sen}x|$ sobre todo el segmento $[-\pi, \pi]$. ▲

Ejemplo 3. Desarrollar en serie de Fourier la función

$$f(x)=\begin{cases}-1 & \text{para } -\pi<x<0,\\ 0 & \text{para } \quad x=0,\\ 1 & \text{para } \quad 0<x<\pi.\end{cases}$$

△ La función dada es impar, por lo tanto, todos los coeficientes $a_n = 0$ y b_n se hallan con ayuda de la fórmula (10):

$$b_n = \frac{2}{\pi}\int_0^{\pi} 1\cdot \operatorname{sen} nx\, dx = -\frac{2}{\pi n}\cos nx \Big|_0^{\pi} = \frac{2}{\pi n}[1-(-1)^n].$$

Ahora bien, todos los coeficientes pares b_n son iguales a cero y los impares tienen la forma $b_{2k-1} = \frac{4}{\pi(2k-1)}$. Por consiguiente,

$$f(x) = \frac{4}{\pi}\sum_{k=1}^{\infty}\frac{1}{2k-1}\operatorname{sen}(2k-1)\,x.$$

Es evidente que la suma de la serie en los puntos $x = 0$ y $x = \pm\pi$ es igual a cero. ▲

Ejemplo 4. Desarrollar en serie de Fourier la función

$$f(x) = \begin{cases} x & \text{para } 0 < x \leqslant 1, \\ 1 & \text{para } 1 < x < 2. \end{cases}$$

△ Puesto que la función se da en un intervalo distinto de $(-\pi, \pi)$, sustituyamos la variable independiente con ayuda de la fórmula $x = (x' + \pi)/\pi$ o bien $x' = \pi(x-1)$. Ahora bien, obtenemos la función siguiente:

$$f(x') = \begin{cases} \frac{1}{\pi}(x'+\pi) & \text{para } -\pi < x' \leqslant 0, \\ 1 & \text{para } \quad 0 < x' < \pi. \end{cases}$$

Puesto que esta serie está definida sobre el intervalo $(-\pi, \pi)$, para ella se puede escribir la serie de Fourier. Calculemos los coeficientes de esta serie:

$$a_0 = \frac{1}{\pi}\int_{-\pi}^{0}\frac{x'+\pi}{\pi}\,dx' + \frac{1}{\pi}\int_0^{\pi} 1\cdot dx' = \frac{3}{2},$$

$$a_n = \frac{1}{\pi}\int_{-\pi}^{0}\frac{x'+\pi}{\pi}\cos nx'\,dx' + \frac{1}{\pi}\int_0^{\pi} 1\cdot\cos nx'\,dx' =$$

$$= \frac{1}{n\pi^2}\int_{-\pi}^{0}\operatorname{sen} nx'\,dx' = \frac{1}{\pi^2 n^2}[1-(-1)^n];$$

$$b_n = \frac{1}{\pi}\int_{-\pi}^{0}\frac{x'+\pi}{\pi}\operatorname{sen} nx'\,dx' + \frac{1}{\pi}\int_0^{\pi} 1\cdot\operatorname{sen} nx'\,dx' =$$

$$= \frac{1}{\pi^2}\int_{-\pi}^{0} x'\operatorname{sen} nx'\,dx' = -\frac{(-1)^n}{\pi n}.$$

Por lo tanto,

$$f(x)=\frac{3}{4}+\frac{2}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2}\cos[(2k-1)\pi(x-1)]-$$
$$-\frac{1}{\pi}\sum_{n=1}^{\infty}\frac{(-1)^n}{n}\operatorname{sen}[n\pi(x-1)].$$

Calculemos los valores de la suma de la serie en los extremos del intervalo:

$$\frac{1}{2}[f(0+0)+f(2-0)]=\frac{0+1}{2}=\frac{1}{2}.$$

El resultado obtenido ofrece la posibilidad de hallar la suma de la serie numérica

$$S=\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2}=1+\frac{1}{3^2}+\frac{1}{5^2}+\ldots+\frac{1}{(2k-1)^2}+\ldots$$

En efecto, en virtud del teorema de Dirichlet para $x=0$ o para $x=2$ tiene lugar la igualdad

$$\frac{1}{2}=\frac{3}{4}-\frac{2}{\pi^2}\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2},$$

de donde

$$\sum_{k=1}^{\infty}\frac{1}{(2k-1)^2}=\frac{\pi^2}{8}. \quad \blacktriangle$$

En conclusión nótese que la integral de la función $f(x)$ se obtiene integrando por términos la serie de Fourier que le corresponde y la derivada $f'(x)$, derivándola por términos. En caso de la derivación la condición $f(-\pi)=f(\pi)$ es necesaria.

§ 7.17. Interpolación trigonométrica

La operación de representar la función $f(x)$ por la serie de Fourier se llama *análisis armónico*. En los cálculos prácticos nos vemos obligados a considerar sólo algunos primeros términos de la serie de Fourier. Como resultado se obtiene únicamente una expresión analítica aproximada para la función $f(x)$ en la forma del polinomio trigonométrico de N-ésimo orden:

$$Q_N(x)=\frac{a_0}{2}+\sum_{n=1}^{N}(a_n\cos nx+b_n\operatorname{sen} nx)\;(-\pi\leqslant x\leqslant\pi). \qquad (1)$$

Además, las fórmulas (3) . . . (5) del § 7.16 para calcular los coeficientes de Fourier son útiles exclusivamente en caso de la representación analítica de la función. En la práctica la función $f(x)$ se representa, por regla general, en forma de una tabla o gráfico. Por eso surge el problema de determinar aproximadamente los coeficientes de Fourier valiéndose del número finito de los valores disponibles de la función.

Generalizando lo dicho, enunciemos el problema siguiente del análisis armónico numérico o bien, como con frecuencia se lo llama de otro modo, del análisis armónico práctico: aproximar sobre el intervalo $(0, T)$ por un polinomio trigonométrico de N-ésimo grado la función $y = f(x)$ para la cual se conocen m valores suyos $y_k = f(x_k)$ para $x_k = kT/m$ $(k = 0, 1, 2, \ldots, m-1)$.

El polinomio trigonométrico para la función definida sobre el intervalo $(0, T)$ tiene la forma

$$Q_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N} \left(a_n \cos n \frac{2\pi}{T} x + b_n \operatorname{sen} n \frac{2\pi}{T} x\right). \tag{2}$$

Los coeficientes a_n y b_n se definen por las relaciones siguientes:

$$a_n = \frac{2}{T} \int_0^T f(x) \cos n \frac{2\pi}{T} x \, dx, \tag{3}$$

$$(n = 0, 1, 2, \ldots, N).$$

$$b_n = \frac{2}{T} \int_0^T f(x) \operatorname{sen} n \frac{2\pi}{T} x \, dx. \tag{4}$$

Aplicando en las relaciones (3) y (4) la fórmula de los rectángulos para calcular los integrales según los valores de las expresiones subintegrales en los puntos $x_k = kT/m$ $(k = 0, 1, 2, \ldots, m-1)$, tenemos

$$a_n = \frac{2}{m} \sum_{k=0}^{m-1} y_k \cos n \frac{2\pi k}{m}, \tag{5}$$

$$(n = 0, 1, 2, \ldots, \ldots, N).$$

$$b_n = \frac{2}{m} \sum_{k=0}^{m-1} y_k \operatorname{sen} n \frac{2\pi k}{m}. \tag{6}$$

Ahora bien, el polinomio trigonométrico (2), cuyos coeficientes a_n y b_n se determinan con ayuda de las fórmulas (5) y (6), sirve de solución del problema planteado.

Se puede mostrar que para $m > 2N$ el polinomio (2) da la mejor aproximación a la función $f(x)$, en lo que se refiere al método de los cuadrados mínimos, si los coeficientes de aquél se calculan haciendo uso de las fórmulas (5) y (6). Con otras palabras, los coeficientes (5)

Tabla 7.13

Meses	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	$\frac{1}{6}\Sigma$
x_k	0	1	2	3	4	5	6	7	8	9	10	11	
Volumen de producción en unidades conv.	95	71	55	43	36	31	28	26	25	45	91	102	108
$\cos \frac{\pi}{6} x_k$	1	0,866	0,5	0	−0,5	−0,866	−1	−0,866	−0,5	0	0,5	0,866	
$\operatorname{sen} \frac{\pi}{6} x_k$	0	0,5	0,866	1	0,866	0,5	0	−0,5	−0,866	−1	−0,866	−0,5	
$\cos \frac{\pi}{3} x_k$	1	0,5	−0,5	−1	−0,5	0,5	1	0,5	−0,5	−1	−0,5	0,5	
$\operatorname{sen} \frac{\pi}{3} x_k$	0	0,866	0,866	0	−0,866	−0,866	0	0,866	0,866	0	−0,866	−0,866	
$\cos \frac{\pi}{2} x_k$	1	0	−1	0	1	0	−1	0	1	0	−1	0	
$\operatorname{sen} \frac{\pi}{2} x_k$	0	1	0	−1	0	1	0	−1	0	1	0	−1	
$\cos \frac{2\pi}{3} x_k$	1	−0,5	−0,5	1	−0,5	−0,5	1	−0,5	−0,5	1	−0,5	−0,5	
$\operatorname{sen} \frac{2\pi}{3} x_k$	0	0,866	−0,866	0	0,866	−0,866	0	0,866	−0,866	0	0,866	−0,866	

Meses	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	$\frac{1}{6}\Sigma$
x_k	0	1	2	3	4	5	6	7	8	9	10	11	
Volumen de producción en unidades conv.	95	71	55	43	36	31	28	26	25	45	91	102	108
$y_k \cos \frac{\pi}{6} y_k$	95	61,49	27,5	0	−18	−26,85	−28	−25,52	−12,5	0	45,5	83,33	34,99
$y_k \operatorname{sen} \frac{\pi}{6} y_k$	0	35,5	47,63	43	31,18	15,5	0	−13	21,65	−45	−78,81	−51	−6,11
$y_k \cos \frac{\pi}{3} y_k$	95	35,5	−27,5	−43	−18	15,5	28	13	−12,5	−45	−45,5	51	7,75
$y_k \operatorname{sen} \frac{\pi}{3} y_k$	0	61,49	47,63	0	31,18	−28,85	0	22,52	21,65	0	−78,81	−88,33	−11,98
$y_k \cos \frac{\pi}{2} y_k$	95	0	−55	0	36	0	−28	0	25	0	−91	0	−3
$y_k \operatorname{sen} \frac{\pi}{2} y_k$	0	71	0	−43	0	31	0	−26	0	45	0	−102	−4
$y_k \cos \frac{2\pi}{3} y_k$	95	−35,5	−27,5	43	−18	−15,5	28	−13	−12,5	45	−45,5	−51	−1,25
$y_k \operatorname{sen} \frac{2\pi}{3} y_k$	0	61.49	−47,63	0	31,18	−26,85	0	22,52	−21,65	0	78,81	−88,33	1,59

y (6) minimizan la suma de los cuadrados de desviaciones

$$\delta_N^2 = \sum_{k=0}^{m-1} [Q_N(x_k) - y_k]^2. \tag{7}$$

En el caso particular para $m = 2N$ los coeficientes a_n y b_n ($n = 0, 1, 2, \ldots, N-1$) se definen por las relaciones (5) y (6) y el coeficiente a_N es

$$a_N = \frac{1}{m} \sum_{k=0}^{m-1} (-1)^k y_k. \tag{8}$$

El mismo polinomio $Q_N(x)$ llega a ser un polinomio interpolador, ya que en este caso para todo b_N se cumplen las relaciones $Q_N(x_k) = y_k$ para todos los $x_k = kT/m$ ($k = 0, 1, 2, \ldots, m-1$).

Ejemplo. Investiguemos la dinámica de producción del azúcar de remolacha. Esta producción lleva un carácter periódico condicionado por la periodicidad del cultivo y por las condiciones de almacenamiento de las materias primas. Por eso en calidad de función que aproxima la dinámica de fabricación del azucar se puede tomar el polinomio trigonométrico (2) para $m = 12$ (esto corresponde al número de meses en el ciclo anual y permite revelar una particularidad específica, o sea, la temporalidad de la fabricación). Por consiguiente,

$$Q_N(x) = \frac{a_0}{2} + \sum_{n=1}^{N} \left(a_n \cos n\frac{\pi}{6}x + b_n \operatorname{sen} n\frac{\pi}{6}x\right) \quad (0 \leqslant x \leqslant 11).$$

En las investigaciones económicas para una buena aproximación de la serie dinámica periódica suelen elegirse cuatro armónicos, como máximo. Los coeficientes a_n y b_n se expresan en la forma siguiente:

$$a_n = \frac{1}{6} \sum_{k=0}^{11} y_k \cos n\frac{\pi}{6}x_k, \quad b_n = \frac{1}{6} \sum_{k=0}^{11} y_k \operatorname{sen} n\frac{\pi}{6}x_k.$$

Calculemos estos coeficientes para los primeros cuatro armónicos del polinomio $Q_N(x)$. Los cálculos necesarios se formalizan tabularmente (véase la tabla 7.13 en las págs 321 y 322).

De la tabla dada obtenemos que $a_0 = 108$; $a_1 = 34{,}99$; $a_2 = 7{,}75$; $a_3 = -3$; $a_4 = -1{,}25$; $b_1 = -6{,}11$; $b_2 = -11{,}98$; $b_3 = -4$; $b_4 = 1{,}59$. Ahora bien, tenemos los siguientes cuatro modelos matemáticos de temporalidad para la fabricación del azucar:

$$Q_1(x) = 54 + 34{,}99 \cos\frac{\pi}{6}x - 6{,}11 \operatorname{sen}\frac{\pi}{6}x;$$

$$Q_2(x) = 54 + 34{,}99 \cos\frac{\pi}{6}x - 6{,}11 \operatorname{sen}\frac{\pi}{6}x + 7{,}75 \cos\frac{\pi}{3}x - 11{,}98 \operatorname{sen}\frac{\pi}{3}x;$$

$$Q_3(x) = 54 + 34{,}99 \cos\frac{\pi}{6}x - 6{,}11 \operatorname{sen}\frac{\pi}{6}x + 7{,}75 \cos\frac{\pi}{3}x -$$

$$-11{,}98 \operatorname{sen} \frac{\pi}{3} x - 3 \cos \frac{\pi}{2} x - 4 \operatorname{sen} \frac{\pi}{2} x;$$

$$Q \qquad 4 + 34{,}99 \cos \frac{\pi}{6} x - 6{,}11 \operatorname{sen} \frac{\pi}{6} x + 7{,}75 \cos \frac{\pi}{3} x -$$

$$-11{,}98 \operatorname{sen} \frac{\pi}{3} x - 3 \cos \frac{\pi}{2} x - 4 \operatorname{sen} \frac{\pi}{2} x - 1{,}25 \cos \frac{2\pi}{3} x +$$

$$+ 1{,}59 \operatorname{sen} \frac{2\pi}{3} x.$$

La comparación de $Q_1(x_k)$ con los valores correspondientes de y_k muestra que ya el primer armónico da, en general, el modelo correcto de la dinámica de la fabricación del azúcar, reflejando su temporalidad.

Calculemos las desviaciones medias cuadráticas, o sea, las desviaciones estándar $\delta_1 = \sqrt{\sum_{k=0}^{11} [Q_i(x_k) - y_k]^2}$ para todos los $Q_i(x)$; encontramos $\delta_1 = 37{,}80$; $\delta_2 = 14{,}40$; $\delta_3 = 7{,}59$; $\delta_4 = 5{,}75$. Como era de esperar, los valores δ_i decrecen monótonamente con el crecimiento de i, con la particularidad de que δ_4 poco se distingue de δ_3. Además, los mismos valores δ_3 y δ_4 son bastante pequeños, así que el polinomio $Q_3(x)$ es una buena aproximación de la serie que caracteriza la dinámica anual de la fabricación del azúcar.

§ 7.18. Métodos numéricos de determinación de los coeficientes de Fourier

Supongamos que se da una serie de Fourier que converge hacia la función periódica $f(x)$:

$$f(x) = \frac{a_0}{2} + \sum_{m=1}^{\infty} (a_m \cos mx + b_m \operatorname{sen} mx), \tag{1}$$

donde

$$a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos mx \, dx \; (m = 0,\ 1,\ 2,\ \ldots). \tag{2}$$

$$b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \operatorname{sen} mx \, dx \; (m = 1,\ 2,\ 3,\ \ldots). \tag{3}$$

En el párrafo precedente hemos enunciado el problema de aproximación de la función $f(x)$ por el polinomio trigonométrico $Q_N(x)$. Allí mismo al calcular los coeficientes a_m y b_m con ayuda de los integrales hemos utilizado la fórmula de los rectángulos.

En el caso general el cálculo aproximado de los coeficientes a_m y b_m se funda en la sustitución de los integrales de los fórmulas (3) y (4) del § 7.17 por sus valores, obtenidos con ayuda de una de las fórmulas de integración aproximada. En este párrafo hacemos uso de la fórmula de los trapecios.

Vamos a suponer que la función $f(x)$ es peri ódica con le período de 2π. Nótese que en vez de los límites us uales de integración entre $-\pi$ y π, al determinar los coefici entes a_m y b_m se puede considerar todo segmento de integración de 2π de largo. Para comodidad de los cálculos tomemos un segmento de 0 a 2π, así que

$$a_m=\frac{1}{\pi}\int_0^{2\pi} f(x)\cos mx\,dx\ (m=0,\ 1,\ 2,\ \ldots), \tag{4}$$

$$b_m=\frac{1}{\pi}\int_0^{2\pi} f(x)\operatorname{sen} mx\,dx\,(m=1,\ 2,\ 3,\ \ldots). \tag{5}$$

Al dividir el segmento $[0, 2\pi]$ en N partes iguales, como resultado obtenemos los puntos de división $0,\ 1\cdot\frac{2\pi}{N},\ 2\cdot\frac{2\pi}{N},\ \ldots\ (N-1)\frac{2\pi}{N}$, 2π. En los p untos de división designemos los valores respectivos de N la N funci ón $f(x)$ con $y_0, y_1, y_2, \ldots, y_{N-1}, y_N = y_0$. Aplicando la fórmula d e los trapecios, obtenemos las siguientes fórmulas aproximadas para calcular los coeficientes a_m y b_m:

$$\frac{N}{2}a_0=\sum_{k=0}^{N-1} y_k=y_0+y_1+\ldots+y_{N-1};$$

$$\frac{N}{2}a_m=\sum_{k=0}^{N-1} y_k\cos k\frac{2m\pi}{N}=y_0+y_1\cos\frac{2m\pi}{N}+\ldots$$

$$\ldots\ y_{N-1}\cos(N-1)\frac{2m\pi}{N};$$

$$\frac{N}{2}b_m=\sum_{k=1}^{N-1} y_k\operatorname{sen} k\frac{2m\pi}{N}=y_1\operatorname{sen}\frac{2m\pi}{N}+\ldots$$

$$\ldots+y_{N-1}\operatorname{sen}(N-1)\frac{2m\pi}{N}.$$

Supongamos que $N = 12$, o sea, el segmento $[0, 2\pi]$ está dividido en 12 partes iguales, así que se utilizan los valores del argumento $0, \pi/6, \pi/3, \ldots, 11\pi/6$; a ellos les corresponden los valores de la función $y_0, y_1, y_2, \ldots, y_{11}$, y las magnitudes, por las cuales se multiplican estos valores, son tales: ± 1; $\pm\operatorname{sen}(\pi/6) = \pm 0{,}5$; $\pm\operatorname{sen}(\pi/3) = \pm 0{,}866$. De aquí, omitiendo los cálculos voluminosos, obtenemos

$$6a_0=y_0+y_1+y_2+y_3+y_4+y_5+y_6+y_7+y_8+y_9+y_{10}+y_{11},$$

$$6a_1=(y_2+y_{10}-y_4-y_8)\operatorname{sen}\frac{\pi}{6}+$$

$$+(y_1+y_{11}-y_5-y_7)\operatorname{sen}\frac{\pi}{3}+(y_0-y_6),$$

$$6a_2 = (y_1 + y_5 + y_7 + y_{11} - y_2 - y_4 - y_8 - y_{10}) \operatorname{sen} \frac{\pi}{6} +$$
$$+ (y_0 + y_6 - y_3 - y_9),$$
$$6a_3 = y_0 + y_4 + y_8 - y_2 - y_6 - y_{10},$$
$$6b_1 = (y_1 + y_5 - y_7 - y_{11}) \operatorname{sen} \frac{\pi}{6} +$$
$$+ (y_2 + y_4 - y_8 - y_{10}) \operatorname{sen} \frac{\pi}{3} + (y_3 - y_9),$$
$$6b_2 = (y_1 + y_2 + y_7 + y_8 - y_4 - y_5 - y_{10} - y_{11}) \operatorname{sen} \frac{\pi}{3},$$
$$6b_3 = y_1 + y_5 + y_9 - y_3 - y_7 - y_{11},$$

etc.

Para reducir al mínimo el número de operaciones aritméticas necesarias con el fin de obtener los valores a_m y b_m, se utiliza un esquema de cálculos especial llamado **esquema de Runge.**

Paso I. Se escriben los valores de la función $f(x)$ observando el orden siguiente:

$$\begin{array}{c} y_0 y_1 y_2 y_3 y_4 y_5 y_6 \\ y_{11} y_{10} y_9 y_8 y_7 \end{array}$$

Paso II. Se cuentan las sumas y diferencias de cada par de los valores que están uno debajo del otro. Las sumas y diferencias obtenidas se escriben del modo siguiente:

$$\begin{array}{lc} & y_0 y_1 y_2 y_3 y_4 y_5 y_6 \\ & \underline{y_{11} y_{10} y_9 y_8 y_7} \\ \text{Sumas} & u_0 u_1 u_2 u_3 u_4 u_5 u_6 \\ \text{Diferencias} & v_1 v_2 v_3 v_4 v_5 \end{array} \tag{6}$$

Paso III. Las operaciones análogas se realizan con las sumas y diferencias (6)

$$\begin{array}{lcc} & u_0 u_1 u_2 u_3 & v_1 v_2 v_3 \\ & \underline{u_6 u_5 u_4} & \underline{v_5 v_4} \\ \text{Sumas} & c_0 c_1 c_2 c_3 & g_1 g_2 g_3 \\ \text{Diferencias} & d_0 d_1 d_2 & h_1 h_2 \end{array} \tag{7}$$

Paso IV. Se calculan los valores a_m y b_m con ayuda de las fórmulas aproximadas

$$\left.\begin{array}{l} 6a_0 = c_0 \;\; c_1 + c_2 + c_3, \\ 6a_1 = d_0 + 0{,}866 d_1 + 0{,}5 d_2, \\ 6a_2 = (c_0 - c_3) + 0{,}5\,(c_1 - c_2), \\ 6a_3 = d_0 - d_2, \\ 6b_1 = 0{,}5 g_1 + 0{,}866 g_2 + g_3, \\ 6b_2 = 0{,}866\,(h_1 - h_2), \\ 6b_3 = g_1 - g_3, \end{array}\right\} \tag{8}$$

etc.

Para comparar de un modo evidente los coeficientes a_m y b_m obtenidos por las fórmulas aproximadas con los valores exactos de los mismos, citemos un ejemplo en el cual la función está representada analíticamente.

Ejemplo. Consideremos la función periódica con el período de 2π:

$$y=f(x)\begin{cases} x/\pi & \text{para } 0\leqslant x\leqslant \pi, \\ 1 & \text{para } \pi<x<2\pi, \\ 0 & \text{para } \qquad x=2\pi. \end{cases}$$

Hagamos la tabla:

x_k	0	$\frac{\pi}{6}$	$\frac{\pi}{3}$	$\frac{\pi}{2}$	$\frac{2\pi}{3}$	$\frac{5\pi}{6}$	π	$\frac{7\pi}{6}$	$\frac{4\pi}{3}$	$\frac{3\pi}{2}$	$\frac{5\pi}{3}$	$\frac{11\pi}{6}$	2π
y_k	0	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$	$\frac{2}{3}$	$\frac{5}{6}$	1	1	1	1	1	1	0

Escribamos según el esquema de Runge los valores y_k y realicemos las adiciones y sustracciones indicadas en éste (véanse las fórmulas (6)]

$$\begin{array}{lcccccccc}
 & 0 & \frac{1}{6} & \frac{1}{3} & \frac{1}{2} & \frac{2}{3} & \frac{5}{6} & 1 \\
 & & 1 & 1 & 1 & 1 & 1 & \\
\hline
\text{Sumas} & 0 & \frac{7}{6} & \frac{4}{3} & \frac{3}{2} & \frac{5}{3} & \frac{11}{6} & 1 \\
\text{Diferencias} & & -\frac{5}{6} & -\frac{2}{3} & -\frac{1}{2} & -\frac{1}{3} & -\frac{1}{6} &
\end{array}$$

Ahora realicemos las sustracciones y adiciones respecto a las sumas y diferencias halladas [véanse las fórmulas (7)]:

$$\begin{array}{lcccc}
 & \multicolumn{4}{c}{\text{sumas}} \\
 & 0 & \frac{7}{6} & \frac{4}{3} & \frac{3}{2} \\
 & 1 & \frac{11}{6} & \frac{5}{3} & \\
\hline
\text{sumas} & 1 & 3 & 3 & \frac{3}{2} \\
\text{diferencias} & -1 & -\frac{2}{3} & -\frac{1}{3} &
\end{array}
\qquad
\begin{array}{lccc}
 & \multicolumn{3}{c}{\text{diferencias}} \\
 & -\frac{5}{6} & -\frac{2}{3} & -\frac{1}{2} \\
 & -\frac{1}{6} & -\frac{1}{3} & \\
\hline
\text{sumas} & -1 & -1 & -\frac{1}{2} \\
\text{diferencias} & -\frac{2}{3} & -\frac{1}{3} &
\end{array}$$

Escribamos ahora las expresiones para a_m y b_m;

$6a_0=1+3+3+\frac{3}{2}$; $\qquad 6b_1=0{,}5\,(-1)+0{,}866\,(-1)-\frac{1}{2}$;

$6a_1=-1-\frac{2}{3}\cdot 0{,}866-0{,}5\cdot\frac{1}{3}$; $\qquad 6b_2=0{,}866\left(-\frac{2}{3}-\frac{1}{3}\right)$;

$6a_2=\left(1-\frac{3}{2}\right)+0{,}5\,(3-3)$; $\qquad 6b_3=-1-\left(-\frac{1}{2}\right)$;

$6a_3=-1-\left(-\frac{1}{3}\right)$;

De aquí

$$a_0=1{,}417;\ a_1=-0{,}291;\ a_2=-0{,}083;\ a_3=-0{,}111,$$
$$b_1=-0{,}311;\ b_2=-0{,}144;\ b_3=-0{,}083.$$

Para comparar damos los valores exactos de los coeficientes:

$$a_0=1{,}500;\ a_1=-0{,}203;\ a_2=0{,}000;\ a_3=-0{,}022;$$
$$b_1=-0{,}318;\ b_2=-0{,}159;\ b_3=-0{,}106.$$

Para obtener con una exactitud más alta los valores de los coeficientes haciendo uso de las fórmulas aproximadas, se pueden utilizar los esquemas provistos de un mayor número de ordenadas.

Cabe señalar que el análisis armónico práctico ofrece la posibilidad de obtener expresiones analíticas que aproximan las funciones dadas con un error estándar mínimo.

§ 7.19. Interpolación inversa

La interpolación de las funciones resulta un aparato útil para resolver varios problemas. El problema de determinación de la raíz de una ecuación o de la de una función es un ejemplo característico de utilización del polinomio interpolador.

Consideremos el siguiente problema de interpolación inversa. Sobre la red Λ_{2k}: $a\leqslant\ldots,\ x_{-k}<\ldots<x_0<\ \ldots<x_k<\ldots$ $\ldots\leqslant b$ está definida por sus valores f_i ($i=f_1,\ \pm1,\ \ldots,\ \pm k$) la función $f(x)$ que es continua sobre el segmento $[a,\ b]$. Se necesita determinar el valor del argumento $x^*\in(x_0,\ x_1)$ que corresponde al valor dado de la función f, $f^*=f_0+\theta\,(f_1-f_0)$; $\theta\in(0,\ 1)$. Se supone que el intervalo $(x_0,\ x_1)$ es tan pequeño que x^* es único.

En realidad, aquí se necesita determinar la raíz de la ecuación

$$f(x)=f^*. \tag{1}$$

Una de las vías posibles de resolver este problema consiste en lo siguiente. Aproximemos la función $f(x)$ por su polinomio interpolador $P_n(f,\ x)$ y reemplecemos la ecuación (1) por la siguiente:

$$P_n(f,\ x)=f^*. \tag{2}$$

Encontramos la raíz real $\overline{x^*}$ de la ecuación (2), perteneciente al intervalo $(x_0,\ x_1)$. Prácticamente obtenemos sólo una solución aproximada de la ecuación (2), o sea, el valor $\bar{\bar{x}}^*$. Ahora suponemos que $x^* = \bar{\bar{x}}^*$.

Vamos a estimar el error de tal solución. Supongamos que el error total de interpolación es igual a Δ, o sea,

$$| P_n (f, x) - f (x) | \leqslant \Delta \tag{3}$$

y el error de la solución de la ecuación (2) es igual a ε, o sea,

$$| \bar{x}^* - \bar{\bar{x}}^* | \leqslant \varepsilon \tag{4}$$

Entonces el incremento de la función f en el punto $\bar{x}^*$ puede ser representado en la forma

$$f^* - f(\bar{x}^*) = (x^* - \bar{x}^*) f'(\xi); \ \xi = \bar{x}^* + \theta (x^* - \bar{x}^*), \ \theta \in (0, 1).$$

De aquí, tomando en consideración que $f^* = P_n (f, \bar{x}^*)$, tenemos

$$P_n (f, \bar{x}^*) - f (x^*) = (x^* - \bar{x}^*) f' (\xi).$$

Suponiendo ahora que $\min\limits_{[x_0, x_1]} | f' (x) | = m_1 > 0$ y utilizando la estimación (3), obtenemos

$$| x^* - \bar{x}^* | \leqslant \Delta / m_1. \tag{5}$$

Luego,

$$| x^* - \bar{\bar{x}}^* | = | x^* - \bar{x}^* + \bar{x}^* - \bar{\bar{x}}^* | \leqslant | x^* - \bar{x}^* | + | \bar{x}^* - \bar{\bar{x}}^* |$$

y en virtud de las desigualdades (4) y (5) encontramos finalmente

$$|x^* - \bar{\bar{x}}^*| \leqslant \frac{\Delta}{m_1} + \varepsilon \tag{6}$$

Ahora bien, tanto la solución del problema planteado como el error (6) se determinan por dos procesos: por la construcción del polinomio interpolador y por la resolución de la ecuación (2), o sea, por la determinación de las raíces del polinomio interpolador.

Puede parecer que estos dos momentos de ningún modo están relacionados entre sí. Sin embargo, esto no es completamente así. Hay que tener presente que el aumento del grado del polinomio, por un lado, disminuye el error Δ, pero, por el otro lado, aumenta el volumen de trabajo necesario para resolver la ecuación (2).

Por eso el grado del polinomio interpolador debe ser mínimo a condición de que se alcance la precisión requerida.

Al resolver prácticamente el problema de interpolación inversa sobre una red uniforme en calidad de polinomios interpoladores, suelen utilizarse los polinomios de Stringer y de Bessel. En este caso la ecuación (2) escrita respecto a la variable $t = (x - x_0)/h$ se reduce a la forma $t = \varphi (t)$ y se resuelve por el método de iteraciones.

Al utilizar el polinomio de Stringer tenemos

$$t = \frac{1}{\mu f_0^1} \left(f^* - f_0 - \frac{f_0^2}{2!} t^2 - \frac{\mu f_0^3}{3!} t (t^2 - 1^2) - \dots \right). \tag{7}$$

La utilización del polinomio de Bessel da

$$t = \frac{1}{2} + \frac{1}{f'_{1/2}} \left(f^* - \mu f_{1/2} - \frac{\mu f''_{1/2}}{2!} t(t-1) - \right.$$

$$\left. - \frac{f'''_{1/2}}{3!} t \left(t - \frac{1}{2}\right)(t-1) - . \right). \tag{8}$$

En el primer caso como aproximación inicial t^0 se toma $\frac{1}{\mu f'_0}(f^* - f_0)$ y en el segundo, 0,5 ó bien $\frac{1}{2} + \frac{1}{f'_{1/2}}(f^* - \mu f_{1/2})$. Obtenida t^*, o sea la solución de la ecuación (7) ó (8), x^* se determina con ayuda de la fórmula $x^* = x_0 + t^*h$.

Análogamente, en caso de necesidad, se puede obtener la solución del problema planteado haciendo uso del primero o segundo polinomio interpolador de Newton.

Consideremos el segundo método de resolución del problema de interpolación inversa, método fundado en la existencia de la función $g(y)$, inversa a $f(x)$.

Supongamos que la función $g(y)$ es continua junto con una cantidad suficiente de sus derivadas sobre el intervalo mínimo, que contiene los valores $y_i = f_i$ ($i = 0, \pm 1, \ldots$) e $y^* = f^*$. En este caso la determinación de x es equivalente al cálculo de la función inversa $g(y)$, definida por sus valores x_i en los nodos y_i y en el punto $y = f^*$, ya que $x^* = g(f^*)$.

Ahora bien, el problema dado está reducido al de interpolación de la función inversa $g(y)$ al cálculo de $g(f^*)$.

El método mencionado de resolución del problema de interpolación inversa es más eficaz que el que contiene la resolución de una ecuación en calidad de una de las etapas. El método en cuestión es sobre todo cómodo si se necesita hallar la solución para una cantidad bastante grande de valores f^* o bien si se necesita obtener la expresión explícita para la raíz de la ecuación (1). El inconveniente del segundo método consiste en la exigencia de que exista una función inversa suave, lo que no siempre puede ser cumplido (por ejemplo, esta exigencia no se cumple para las funciones no monótonas).

En conclusión nótese que para calcular x^* con ayuda de la función inversa el más cómodo es el método de iteración-interpolación de Aitken expuesto en el § 7.13.

Ejemplo 1. Valiéndose de la tabla de valores de la función $f = 3^x$ (véase la tabla del ejemplo dado en el § 7.9,), determinar a qué valor del argumento x^* corresponde el valor de la función $f^* = 5$. Estimar el error.

△ En el ejemplo § 7.9 hemos determinado que el orden de exactitud de la tabla dada es igual a 3. Puesto que el valor dado f^* está situado al fin de la tabla, para calcular x^* es necesario utilizar el segundo polinomio interpolador de Newton que es de tercer grado. Suponiendo $x_0 = 1{,}50$ y $t = (x - x_0)/h$, así como utilizando la fórmula (10) del § 7.9, obtenemos la ecuación para determinar t^*:

$$5=5{,}196+\frac{1{,}248}{1!}t+\frac{0{,}300}{2!}t(t+1)+\frac{0{,}072}{3!}t(t+1)(t+2).$$

En este caso, teniendo en cuenta los resultados del § 7.9, calculemos el error del valor de $\bar{x}^*$ con ayuda de la fórmula (5). Puesto que $\Delta=\Delta_1+\Delta_2=0{,}004$ y $m_1=\min\limits_{[0,75;\,1,5]}|3^x\ln 3|\approx 2{,}5$, el error buscado constituye $|x^*-\bar{x}^*|\leqslant\Delta/m_1=0{,}0016$.

Vamos a transformar la ecuación para determinar t^* en forma cómoda para la aplicación del método de iteraciones

$$t=\frac{5-5{,}196}{1{,}248}-\frac{1}{1{,}248}\left[\frac{0{,}300}{2!}t(t+1)+\frac{0{,}072}{3!}t(t+1)(t+2)\right].$$

Resolvamos esta ecuación, tomando por aproximación inicial $t_0=(5-5{,}196):1{,}248\approx-0{,}16$. Entonces

$$t_1=-0{,}16-\frac{1}{1{,}248}\left[\frac{0{,}300}{2!}(-0{,}16)(0{,}84)+\right.$$

$$\left.+\frac{0{,}072}{3!}(-0{,}16)\cdot(0{,}84)(1{,}84)\right]=-0{,}141;$$

$$t_2=-0{,}16-\frac{1}{1{,}248}\left[\frac{0{,}300}{2!}(-0{,}141)(0{,}859)+\right.$$

$$\left.+\frac{0{,}072}{3!}(-0{,}141)(0{,}859)(1{,}859)\right]=-0{,}143.$$

Ahora bien, en calidad de solución aproximada de la ecuación se puede tomar el valor $t^*=-0{,}14\pm0{,}003$. De aquí

$$\bar{\bar{x}}^*=x_0+t^*h=1{,}465$$

y el error de la solución de la ecuación

$$|\bar{x}^*-\bar{\bar{x}}^*|=\varepsilon<0{,}0008,$$

Ahora bien, la solución final $x^*=1{,}465\pm0{,}003$. ▲

Ejemplo 2. Haciendo uso de la tabla de valores de la función $f=\ln x$ (véase el ejemplo 1 del § 7.13), calcular e^2 con exactitud de hasta 0,01.

△ La función dada f tiene la función inversa $g(y)=e^y$ que es continua junto con sus derivadas sobre el intervalo $(-\infty,+\infty)$. Por eso el cálculo de e^y puede ser reducido al cálculo en el punto $y=y^*=2$ de la función e^y representada en forma de la tabla:

y	0,00	0,69	1,39	1,61	2,08	2,30
g	1	2	4	5	8	10

Vamos a resolver el problema de interpolación obtenido con ayuda del método de Aitken. Numeremos los nodos y_1 en el orden siguiente: $y_0=2{,}08$; $y_1=2{,}30$; $y_2=1{,}61$; $y_3=1{,}39$; $y_4=0{,}69$; $y_5=0{,}00$. Utilizando ahora la fórmula (1) del § 7.13 y reemplazando x por y y x_m por y_m, calculamos los valores de los polinomios interpoladores P_n (y^*):

$$P_1^{01}(2)=\frac{1}{0,22}\cdot\begin{vmatrix}-0,08 & 8\\ -0,30 & 10\end{vmatrix}=7,27;$$

$$P_1^{02}(2)=\frac{1}{0,47}\cdot\begin{vmatrix}0,39 & 5\\ -0,08 & 8\end{vmatrix}=7,49;$$

$$P_2^{012}(2)=\frac{1}{0,69}\cdot\begin{vmatrix}0,39 & 7,49\\ -0,30 & 7,27\end{vmatrix}=7,37;$$

$$P_1^{23}(2)=\frac{1}{0,22}\cdot\begin{vmatrix}0,61 & 4\\ 0,39 & 5\end{vmatrix}=6,77;$$

$$P_2^{023}(2)=\frac{1}{0,69}\cdot\begin{vmatrix}0,61 & 6,77\\ 0,08 & 7,49\end{vmatrix}=7,41;$$

$$P_3^{0123}(x)=\frac{1}{0,91}\cdot\begin{vmatrix}0,61 & 7,41\\ -0,30 & 7,37\end{vmatrix}=7,38.$$

Puesto que $|P_3^{0123}(2)-P_2^{012}(2)|=0,01$ y la exactitud requerida está alcanzada, cesamos los cálculos y suponemos $e^2=7,38\pm$ $\pm\,0,01$. ▲

Ejercicios

1. La función $y=f(x)$ se da tabularmente:

x	1,522	1,523	1,524
y	20,477	20,906	21,354

Determinar su valor en el punto $x=1,5228$ con ayuda de la primera fórmula de interpolación de Newton.

2. La función $y=f(x)$ se da tabularmente:

x	1,529	1,530	1,531
y	23,911	24,498	25,115

Determinar su valor en el punto $x=1,5303$, haciendo uso de la segunda fórmula de interpolación de Newton.

3. Construir el polinomio interpolador de Lagrange para la función representada tabularmente:

x	-2	-1	2	3
y	-12	-8	3	5

4. Construir el polinomio interpolador de Lagrange para la función $f(x)=e^{-x}$ si de nodos de interpolación sirven los puntos $x_0=1$, $x_1=2$, $x_2=3$. Estimar el error para $x=1,5$.

5. Hacer la tabla de diferencias finitas para la función representada tabularmente:

x	-2	-1	0	1	2	3
y	10	5	1	-15	-20	-100

6. Hacer la tabla de diferencias divididas para la función representada tabularmente:

x	-3	1	0	2	3
y	-15	-7	1	25	47

7. Para la función $y = f(x)$ representada en la forma tabular:

x	1,03	1,08	1,016	1,23	1,26	1,33	1,39
y	2,80107	2,94468	3,18993	3,42123	3,52542	3,78104	4,01485

calcular el valor en el punto $x = 1{,}21555$ con exactitud de hasta 10^{-5}, ha iendo uso de la fórmula de Aitken.

8. C lcular el valor de la función en el punto $x = 1{,}34627$ utilizando la fórmula de Stirling si la función $y = f(x)$ está representada en la forma tabular:

x	1,355	1,340	1,345	1,350	1,355	1,360
y	4,16206	4,25562	4,35325	4,45522	4,56184	4,67433

9. Para la función representada en la forma tabular:

x	1,435	1,440	1,445
y	0,892687	0,893698	0,894700

determinar el valor del argumento, correspondiente al de la función 0,892914.

10. Haciendo uso del método de interpolación inversa, hallar con la exactitud de hasta 10^{-5} la raíz de la ecuación $\sqrt{x+1} - \frac{1}{x} = 0$, que está sobre el segmento $[0{,}7;\ 0{,}8]$.

11. Construir un polinomio interpolador que refleje la variación de la exportación realizada por la industria de construcción durante el período de 1981 a 1983, utilizando los datos siguientes:

Años	1981	1982	1983
Exportación, en millones de rublos convertibles	384,6	507,9	477,9

12. Construir un polinomio interpolador que refleje la variación de la importación realizada por la industria de materiales de construcción durante el período de 1981 a 1983, utilizando los datos

Años	1981	1982	1983
Importación, en millones de rublos convertibles	349,1	416,9	430,6

13. Construir un polinomio interpolador que refleje la variación de la importación realizada por la industria de construcción durante el período de 1981 a 1983, utilizando los datos siguientes:

Años	1981	1982	1983
Importación, en millones de rublos convertibles	657,8	980,6	949,6

14. Construir un polinomio interpolador que refleje la variación de la exportación realizada por la energética eléctrica durante el período de 1980 a 1983, utilizando los datos siguientes:

Años	1980	1981	1982	1983
Exportaciones, en millones de rublos convertibles	421,1	469,4	544,6	619,2

15. La importación de los productos de las industrias forestal, papelera y maderera constituyó según los año

Años	1980	1981	1982	1983
Importación, en millones de rublos convertibles	889,2	938,7	883,6	796,8

Construir un polinomio interpolador que refleje la variación de la importación durante el período de 1980 a 1983.

16. Construir un polinomio interpolador que refleje la variación de la exportación realizada por las industrias forestal, papelera y maderera en el período de 1979 a 1981, utilizando los datos siguientes:

Años	1979	1980	1981
Exportación, en millones de rublos convertibles	1742,7	2008,5	1893,4

17. Costruir un polinomio interpolador que refleje la variación de la importación realizada por la industria ligera en el período de 1980 a 1983, utilizando los datos siguientes:

Años	1980	1981	1982	1983
Importación, en millones de rublos convertibles	7085,3	8454,4	8821,1	8760,8

18. El nivel de instrucción de la población urbana ocupada está representado en la tabla:

Años	1939	1959	1979	1983
Tienen instrucción superior por 1000 personas de la población urbana ocupada	32	59	130	142

¿Cuál fue el nivel de instrucción de la población urbana ocupada en 1982?

19. El nivel de instrucción de la población rural ocupada está representado en la tabla:

Años	1939	1959	1979	1983
Tienen instrucción superior por 1000 personas de la población rural ocupada	3	11	42	50

¿Cuál fue el nivel de instrucción en los lugares rurales en 1982?

20. La cantidad de egresados de los centros de enseñanza superior, especialidad de geología, está representada en la tabla:

Años	1965	1970	1975	1980
Cantidad de especialistas, en miles de personas	3,2	5,1	5,9	6,2

Determinar la cantidad de egresados, especialidad de geología, en 1979.

21. La renta nacional utilizada para la acumulación y otros gastos constituyó

Años	1965	1970	1975	1980
Renta nacional, en miles de millones de rublos	50,2	84,2	96,6	107,2

Determinar la renta nacional en 1979.

22. El consumo de energía eléctrica en la agricultura constituyó:

Años	1965	1970	1975	1980
Consumo de energía eléctrica, en miles de millones de kWh	21,1	38,6	73,8	111

¿Cuál fue el consumo de energía eléctrica en 1979?

23. La fabricación de los motores eléctricos de corriente alterna constituyó:

Años	1970	1975	1980	1982
Fabricación de motores eléctricos con consumo de energía, en millones de kWh	36,3	45,9	51,8	53,3

¿Cuál fue la fabricación de los motores eléctricos (según el consumo de energía) en 1981?

24. Calcular el valor de la función $f(x)$ en el punto $x = x_1$ con ayuda del polinomio interpolador correspondiente, utilizando las tablas de cuatro cifras de funciones trigonométricas con el paso igual a 1°. Determinar el error absoluto del resultado:

a) $f(x) = \text{sen}\, x$, $x_1 = 37{,}7°$; b) $f(x) = \cos x$, $x_1 = 19°48'$;
c) $f(x) = \text{son}\, x$, $x = 53°12'$; d) $f(x) = \cos x$, $x_1 = 36°48'$;
e) $f(x) = \cos x$, $x_1 = 71{,}6°$; f) $f(x) = \text{tg}\, x$, $x_1 = 67°48'$.

25. Para la función $y = f(x)$ definida sobre el intervalo $(0; \pi)$ y representada en forma de la tabla de valores $y_k = f(x_k)$:

k	0	1	2	3
x_k	0	$\pi/4$	$\pi/2$	$3\pi/4$
y_k	1	2	2,4	2,6

formar el polinomio interpolador trigonométrico.

26. Para la función $y = f(x)$ definida sobre el intervalo $(0, 1)$ y representada en forma de la tabla de valores $y_k = f(x_k)$:

k	0	1	2	3	4	5
x_k	0	1/6	1/3	1/2	2/3	5/6
y_k	1	0	−2	−3	0	2

formar un polinomio trigonométrico cuyo orden no sea inferior al segundo.

CAPÍTULO VIII

Derivación e integración numéricas

§ 8.1. Planteamiento del problema y fórmulas elementales de la derivación numérica

Al resolver muchos problemas prácticos surge la necesidad de obtener los valores de las derivadas de diferentes órdenes de la función f, representada en forma de una tabla o de una expresión analítica compleja. En estos casos es imposible o difícil aplicar directamente los métodos de cálculo diferencial. Entonces se utilizan los métodos aproximados de derivación numérica.

Las expresiones elementales para las derivadas se obtienen como resultado de derivación de las fórmulas de interpolación.

Ahora bien, examinemos el problema siguiente. Sobre la red $a \leqslant x_0 < x_1 < \ldots < x_n \leqslant b$ en los nodos x_i se dan los valores f_i de la función f continuamente derivable $n + 1 + m$ veces. Se necesita calcular la derivada $f^{(m)}(x^*)$, $x^* \in [a, b]$ y estimar el error.

Uno de los métodos posibles de resolver este problema consiste en lo siguiento. Vamos a construir para la función f sobre los nodos x_i $(i = 0, 1, \ldots, n)$ el polinomio interpolador con el término residual R_n, así que

$$f(x) = P_n(x) + R_n(x). \tag{1}$$

Derivemos los miembros primero y segundo de la relación (1) m veces y pongamos $x = x^*$:

$$f^{(m)}(x^*) = P_n^{(m)}(x^*) + R_n^{(m)}(x^*). \tag{2}$$

Para funciones bastante suaves, o sea, para funciones con derivadas restringidas, suficiente cantidad de nodos y suficiente exactitud de los cálculos la magnitud $R_n^{(m)}(x^*)$ es pequeña y $P_n^{(m)}(x^*)$ es una buena aproximación para $f^{(m)}(x^*)$, así que se puede poner

$$f^{(m)}(x^*) \approx P_n^{(m)}(x^*). \tag{3}$$

En los cálculos prácticos la derivación numérica resulta muy sensible a los erroresen la información inicial, a la supresión de los términos de la serie y a otras operaciones semejantes. Además, una alta exactitud de interpolación [pequeñez de $R_n(x)$] no garantiza, en absoluto, una alta exactitud de la fórmula de interpolación para las derivadas [pequeñez de $R_n^{(m)}(x)$]. Por eso la derivación numérica ha de aplicarse con cuidado y, por lo general, para pequeños m.

Teniendo en cuenta lo dicho, así como el hecho de que el cálculo de derivadas superiores puede ser reducido al cálculo sucesivo de las derivadas inferiores, consideremos con más detalles la obtención de las fórmulas de cálculo para f' y f'' en los nodos de una red uniforme. Para obtener las derivadas en los puntos nodales es conveniente utilizar el polinomio interpolador de Stirling y su término residual [véanse las fórmulas (5) y (6) del § 7 8]. Así, derivando el polinomio de Stirling y su término residual respecto a x y suponiendo $x^* = x_0$ $(t^* = 0)$, obtenemos las siguientes expresiones para la derivada:

$$f'(x_0) = \frac{1}{h}\mu f_0^1 \pm \frac{M_3}{6}h^2 \quad (k=1), \tag{4}$$

$$f'(x_0) = \frac{1}{h}\left(\mu f_0^1 - \frac{\mu f_0^3}{6}\right) \pm \frac{M_5}{30}h^4 \quad (k=2). \tag{5}$$

Derivando el polinomio de Stirling dos veces respecto a x y calculando el valor de la segunda derivada en el punto $x^* = x_0$, tenemos

$$f''(x_0) = \frac{1}{h^2} f_0^2 \pm \frac{M_4}{12}h^2 \quad (k=1), \tag{6}$$

$$f''(x_0) = \frac{1}{h^2}\left(f_0^2 - \frac{1}{12}f_0^4\right) \pm \frac{M_6}{90}h^4 \quad (k=2). \tag{7}$$

Para calcular la derivada exactamente en el centro entre los nodos $x^* = x_0 + \frac{h}{2}$ se emplea el polinomio de Bessel. En este caso las fórmulas correspondientes para la derivada tienen la forma

$$f'\left(x_0 + \frac{h}{2}\right) = \frac{1}{h} f_{1/2}^1 \pm \frac{M_3}{24}h^2 \quad (k=1), \tag{8}$$

$$f'\left(x_0 + \frac{h}{2}\right) = \frac{1}{h}\left(f_{1/2}^1 - \frac{1}{24}f_{1/2}^3\right) \pm \frac{3M_5}{640}h^4 \quad (k=2). \tag{9}$$

Representan interés práctico también las así llamadas fórmulas de derivación unilateral que permiten calcular $f'(x_0)$ por los nodos $x_i = x_0 + ih$ $(i = 0, 1, \ldots, k, \ldots$ o bien $i = 0, -1, \ldots, -k, \ldots)$. Es cómodo construir estas fórmulas con ayuda de los polinomios interpoladores primero y segundo de Newton.

Derivando el primer polinomio de Newton respecto a x y calculando el valor de la derivada en el punto $x = x_0$ $(t = 0)$ para $k = 1$ y $k = 2$, obtenemos, respectivamente, las siguientes fórmulas

$$f'(x_0) = \frac{1}{h}\Delta f_0 \pm \frac{1}{2}M_2 h, \tag{10}$$

$$f'(x_0) = \frac{1}{h}\left(\Delta f_0 - \frac{1}{2}\Delta^2 f_0\right) \pm \frac{1}{3}M_3 h^2. \tag{11}$$

Análogamente, derivando el segundo polinomio de Newton, para $k = -1$ y $k = -2$ tenemos, respectivamente,

$$f'(x_0) = \frac{1}{h}\nabla f_0 \pm \frac{1}{2}M_2 h, \tag{12}$$

$$f'(x_0) = \frac{1}{h}\left(\nabla f_0 + \frac{1}{2}\nabla^2 f_0\right) \pm \frac{1}{3}M_3 h^2. \tag{13}$$

§ 8.2. Particularidades de la derivación numérica

Escribamos una vez más todas las fórmulas de segundo orden, expresando directamente por los valores de la función f_i las diferencias finitas que forman parte de ellas. De las relaciones (4), (6) y (8) del § 8.1 tenemos

$$f'(x_0)=\frac{f_1-f_{-1}}{2h}\pm\frac{M_3}{6}h^2, \tag{1}$$

$$f''(x_0)=\frac{f_1-2f_0+f_{-1}}{h^2}\pm\frac{M_4}{12}h^2, \tag{2}$$

$$f'\left(x_0+\frac{h}{2}\right)=\frac{f_1-f_0}{h}\pm\frac{M_3}{24}h^2. \tag{3}$$

Las relaciones (11) y (13) del § 8,1 dan, respectivamente,

$$f'(x_0)=\frac{1}{2h}(-3f_0+4f_1-f_2)\pm\frac{M_3}{3}h^2, \tag{4}$$

$$f'(x_0)=\frac{1}{2h}(3f_0-4f_{-1}+f_{-2})\pm\frac{M_3}{3}h^2. \tag{5}$$

De las fórmulas anteriormente citadas se ve que con la disminución del paso de la red disminuye también el error del método. No obstante, si los valores de la función f_i se dan aproximadamente, por ejemplo con igual error absoluto ε, entonces al utilizar las fórmulas de derivación numérica el error total contendrá un sumando adicional, inversamente proporcional a h^m (m es el orden de la derivada). Por eso es racional disminuir h sólo en determinados límites.

Ilustrando lo dicho, consideremos el segundo miembro de la fórmula (3). Su error total constituye

$$\Delta=\frac{M_3}{24}h^2+\frac{2\varepsilon}{h}. \tag{6}$$

Igualando $\Delta'(h)$ a cero, obtenemos el punto del extremo de la función Δ h):

$$h_0=2\sqrt[3]{\frac{3\varepsilon}{M_3}}\approx 2{,}9\sqrt[3]{\frac{\varepsilon}{M_3}}. \tag{7}$$

Puesto que $\Delta''(h)>0$, entonces h_0 es el punto del mínimo $\Delta(h)$, con la particularidad de que

$$\Delta(h_0)=\frac{3}{2}\sqrt[3]{\frac{1}{3}M_3\varepsilon^2}\approx\sqrt[3]{M_3\varepsilon^2}. \tag{8}$$

Esta relación significa, en particular, que para ningún h se puede garantizar que el error del resultado sea la magnitud $o(\varepsilon^{2/3})$.

Análogamente, de la fórmula (2) para el paso óptimo obtenemos la expresión

$$h_0=2\sqrt[4]{\frac{3\varepsilon}{M_4}}\approx 2{,}6\sqrt[4]{\frac{\varepsilon}{M_4}} \tag{9}$$

y de las fórmulas (4) y (5), la expresión

$$h_0=\sqrt[3]{\frac{6\varepsilon}{M_3}}\approx 1{,}8\sqrt[3]{\frac{\varepsilon}{M_3}}. \tag{10}$$

Ahora bien, al calcular las derivadas es necesario determinar previamente el paso óptimo de la tabla inicial de valores f_i.

Ejemplo 1. Calcular f' (1,6) y f'' (1,4) para la función $f=\ln x$ representada en forma de la tabla siguiente:

x	1,2	1,3	1,4	1,5	1,6
f	0,1823	0,2626	0,3364	0,4054	0,4700

que contiene los valores f_i con todas las cifras justas en un amplio sentido. Estimar el error del resultado.

△ Para calcular las derivadas requeridas apliquemos las fórmulas (5) y (2), respectivamente. Entonces, utilizando las igualdades (10) y (9), así como los datos iniciales, obtenemos los valores siguientes para el paso óptimo:

$$h_{01}=\sqrt[3]{\frac{6\cdot 10^{-4}}{0{,}73}}\approx 0{,}1 \text{ al calcular } f'\ (1{,}6)$$

$$h_{02}=\sqrt[4]{\frac{48\cdot 10^{-4}}{2{,}1}}\approx 0{,}22 \text{ al calcular } f''\,(1{,}4).$$

Puesto que los datos tabulares no permiten elegir en calidad de paso 0,22, tomamos por h_2 el número más próximo posible 0,2. Por consiguiente

$$f'\,(1{,}6)=\frac{1}{0{,}2}\,(3\cdot 0{,}4700-4\cdot 0{,}4054+0{,}3364)=0{,}624,$$

con la particularidad de que el error total no supera

$$\Delta=\frac{0{,}73}{3}\cdot 0{,}1^2+\frac{4\cdot 10^{-4}}{0{,}1}=0{,}007$$

y

$$f''\,(1{,}4)=\frac{1}{0{,}2^2}\,(0{,}4700-2\cdot 0{,}3364+0{,}1823)=-0{,}512,$$

con la particularidad de que el error total no supera

$$\Delta=\frac{2{,}9}{12}\cdot 0{,}2^2+\frac{4\cdot 10^{-4}}{0{,}2^2}=0{,}02.$$

Las estimaciones mencionadas del error, por ser, como regla, demasiado aumentadas, videncian, a pesar de lo dicho, que la operación de determinación de la segunda derivada es menos fiable que la de determinación de la primera. ▲

En algunos casos prácticos para determinar la derivada se representa sólo la tabla de valores de la función. Entonces no es posible, evidentemente, estimar el error. Los valores aproximados de las derivadas se calculan directamente con ayuda de una de las fórmulas (4)...(13) del § 8.1 sin tener en cuenta el error.

Ejemplo 2. Calcular $f'(1,3)$ y $f''(1,4)$ para la función $f(x)$ representada en forma de la tabla siguiente:

x	1,2	1,3	1,4	1,5	1,6
$y=f(x)$	0,18	0,26	0,34	0,41	0,47

En virtud de las fórmulas (4) y (6) del § 8.1 obtenemos, respectivamente:

$$f'(1,3)=\frac{1}{0,1}\cdot\frac{1}{2}(0,34-0,26+0,26-0,18)=0,8;$$

$$f''(1,4)=\frac{1}{0,1^2}(0,41-0,34-0,34+0,26)=-1. \quad \blacktriangle$$

§ 8.3. Planteamiento del problemas de integración numérica

Supongamos que se necesita calcular la integral

$$I=\int_a^b f(x)\,dx. \tag{1}$$

Del curso de análisis matemático se sabe que para una función f, continua sobre el segmento $[a, b]$, la integral (1) existe y es igual a la diferencia entre los valores de la primitiva F para la función f en los puntos b y a:

$$I=\int_a^b f(x)\,dx=F(b)-F(a). \tag{2}$$

Sin embargo, en la mayoría aplastante de los problemas prácticos no se logra expresar la primitiva por las funciones elementales. Además, la función f se representa con frecuencia en forma de la tabla de sus valores para determinados valores del argumento. Todo esto da lugar a que se apliquen los métodos aproximados de cálculo de la integral (1), los cuales pueden ser subdivididos, convencionalmente, en análiticos y numéricos. Los primeros consisten, de hecho, en la construcción aproximada de la primitiva y utilización posterior de la fórmula (2). En cambio, los segundos permiten hallar inmediatamente el valor numérico de la integral, basándose en los valores conocidos de la función subintegral (y, a veces, de sus derivadas) en los puntos dados que se llaman *nodos*. En este capítulo examinaremos sólo los métodos numéricos de integración de las funciones. El mismo proceso de determinación numérica de la integral se denomina *cuadratura* y las fórmulas correspondientes, *fórmulas de integración numérica* (fórmulas de cuadratura).

Según sea el método de representación de la función subintegral examinaremos dos casos de integración numérica, diferentes en el sentido de su realización.

Problema I. Sobre el segmento $[a, b]$ en los nodos x_i se dan los valores f_i de cierta función f perteneciente a determinada clase de F. Se necesita calcular aproximadamente la integral (1) y estimar el error del valor obtenido.

Así suele plantearse el problema de integración numérica en el caso en que la función subintegral esté representada en forma tabular.

Problema II. Sobre el segmento $[a, b]$ la función $f(x)$ se da en forma de una expresión analítica. Se necesita calcular la integral (1) con el error máximo admisible prefijado ε.

Uno de los métodos posibles de resolver los problemas enunciados se funda en la utilización de diferentes fórmulas de integración numérica que tienen la forma

$$I \equiv \int_a^b f(x)\,dx \approx (b-a) \sum_{i=1}^{n} A_i f(x_i) \equiv I_n \tag{3}$$

con el término residual conocido $R_n[f] = I - I_n$ o con su estimación.

En el caso general, tanto los puntos nodales x_i como los factores de ponderación (pesos) A_i no se conocen de antemano y se someten a la determinación al deducir cada fórmula concreta de integración numérica (3), basándose en los requisitos que ésta debe reunir.

En realidad, el problema de integración numérica es equivalente a la estimación del valor medio de la función. Efectivamente, el valor medio de la función sobre el segmento $[a, b]$ se determina del modo siguiente:

$$\bar{f} = \frac{1}{b-a} \int_a^b f(x)\,dx.$$

Por eso

$$\int_a^b f(x)\,dx = (b-a)\,\bar{f}.$$

A su vez, la determinación del valor medio de la función es un problema estadístico que comprende problemas de muestreo sucesivo y de planificación del experimento. Debido a la dificultad que presenta tal planteamiento de la cuestión, en este capítulo nos limitamos a examinar sólo los métodos clásicos de integración numérica basados en la previa determinación tanto de los puntos nodales, en los cuales debe darse la información sobre la función a integra, como de la misma información mencionada.

Vamos a pasar ahora a los algoritmos de resolución de los problemas anteriormente enunciados.

Algoritmo de resolución del problema I.

1°. Se elige la fórmula concreta de integración numérica (3) y se calcula I_n. Si los valores de la función f_i se dan aproximadamente, se calcula de hecho sólo el valor aproximado de $\bar{I}_n$ para I_n exacto.

2°. Se toma aproximadamente que $I \approx \bar{I}_n$.

3°. Haciendo uso de la expresión concreta para el término residual o de su estimación para la fórmula elegida de integración numérica, se calcula el error del método

$$\Delta_1 = | I - I_n | = | R_n |.$$

4°. Se determina el error de cálculo de I_n:

$$\Delta_2 = | I_n - \bar{I}_n |$$

utilizando los errores de los valores aproximados f_i.

5°. Se halla el error absoluto completo del valor aproximado $\bar{I}_n$:

$$\Delta = | I - \bar{I}_n | \leqslant \Delta_1 + \Delta_2.$$

6°. Se obtiene la solución del problema en la forma

$$I = \bar{I}_n \pm \Delta.$$

Para las funciones bastante suaves, o sea, para las funciones con una variación restringida de las derivadas el error de las fórmulas de integración numérica (3) para grandes bastante n es, por lo general, pequeño. Por eso para una exactitud suficiente de los valores iniciales f_i y para una exactitud suficiente de cálculo de $\bar{I}_n$ se puede esperar que $\bar{I}_n$ será una buena aproximación de I. En estas consideraciones se basa precisamente el algoritmo siguiente.

Algoritmo de resolución del problema II.

1°. Se representa ε en forma de la suma de tres sumandos no negativos:

$$\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3,$$

donde ε_1 es el error máximo admisible del método; ε_2, el error máximo admisible de cálculo de $\bar{I}_n$; ε_3, el error máximo admisible de redondeo del resultado.

2°. Se elige n en la fórmula de integración numérica de un modo tal que se cumpla la desigualdad

$$\Delta_1 = | I - I_n | = | R_n | \leqslant \varepsilon_1.$$

3 . Se calcula f_i con una exactitud tal que al calcular I_n con ayuda de la fórmula (3) se asegure el cumplimiento de la desigualdad

$$\Delta_2 = | I_n - \bar{I}_n | \leqslant \varepsilon_2.$$

Es evidente que para eso basta con calcular todos los f_i con un error absoluto igual a $\dfrac{\varepsilon}{(b-a)\sum\limits_{i=1}^{n} |A_i|}$.

4°. La magnitud $\overline{I}_n$ hallada en el subp. 3° se redondea (si $\varepsilon_3 \neq 0$) con el error máximo admisible ε_3 hasta la magnitud $\overline{\overline{I}}_n$.

5°. Se obtiene la solución del problema en la forma

$$I = \overline{\overline{I}}_n \pm \varepsilon.$$

Según ya hemos dicho, las fórmulas de integración numérica, utilizadas en los algoritmos de ambos problemas, se construyen sobre la base de unos u otros criterios que determinan la posición de los puntos nodales y las magnitudes de los factores de ponderación. De tales criterios pueden servir: la representación de la integral en forma de la suma integral; la aproximación de la función subintegral (por ejemplo, por medio de un polinomio) y la integración sucesiva de la función aproximativa la exigencia de que la fórmula (3) sea absolutamente exacta para una clase determinada de las funciones, etc.

§ 8.4. Fórmulas elementales de integración numérica

Fórmulas de los rectángulos. Como se sabe, la integral definida es, en virtud de su construcción, el límite de las sumas integrales

$$\int_a^b f(x)\,dx = \lim_{\max h_i \to 0} \sum_{i=1}^{n} h_i f(\xi_i), \tag{1}$$

cada una de las cuales corresponde a cierta partición D_n: $x_0 = x_0 < < x_1 < \ldots < x_n = b$ del segmento $[a, b]$ y a la colección arbitraria de los puntos $\xi_i \in [x_{i-1}, x_i]$ para cada partición; $h_i = x_i - x_{i-1}$.

Limitándose por un número finito de sumandos en el segundo miembro de la igualdad (1) y tomando en calidad de colección de ξ_i unos u otros valores del argumento contenidos en los segmentos $[x_{i-1}, x_i]$ se puede obtener diferentes fórmulas de integración aproximada. Así, tomando por colección de ξ_i los valores de los extremos izquierdos o derechos de los segmentos $[x_{i-1}, x_i]$, obtenemos, respectivamente, las *fórmulas de los rectángulos izquierdos* o *derechos* ($h_i = 1/n = \text{const}$):

$$I \equiv \int_a^b f(x)\,dx \approx (b-a) \sum_{i=0}^{n-1} \frac{1}{n} f_i \equiv I_{izq}, \tag{2}$$

$$I \equiv \int_a^b f(x)\,dx \approx (b-a) \sum_{i=1}^{n} \frac{1}{n} f_i \equiv I_d. \tag{3}$$

Las denominaciones de estas fórmulas proceden de su interpretación geométrica. Si en el plano xOy se construye una curva $y = = f(x)$ y se divide el segmento $[a, b]$ en n partes por los puntos x_i de la red D_n, la fórmula de los rectángulos izquierdos dará como valor aproximado de la integral el área total de los rectángulos rayados en la fig. 8.1, mientras que la fórmula de los rectángulos derechos dará el area total de los rectángulos rayados en la fig. 8.2

Ejemplo 1. Con ayuda de las fórmulas de los rectángulos izquierdos y derechos calcular $\int_1^9 \frac{dx}{x+2}$, suponiendo $n = 4$.

△ Conociendo los límites de integración $a = 1$ y $b = 9$, encontramos el paso $h = (b - a)/n = 2$; entonces de puntos de partición

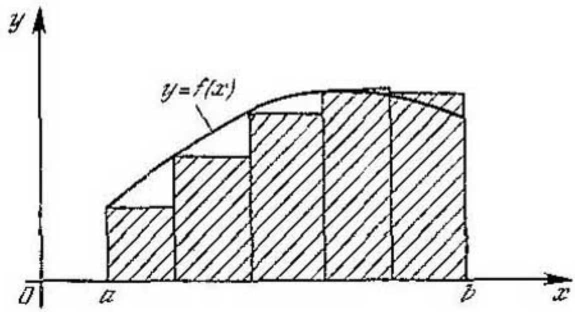


Fig. 8.1

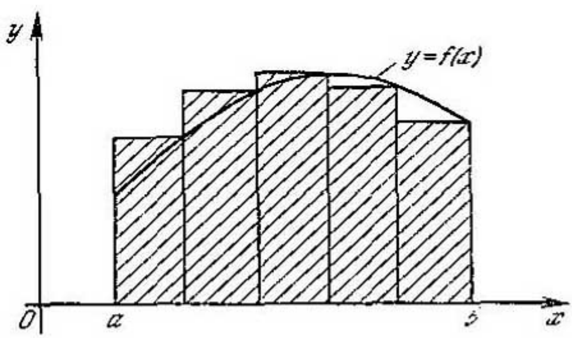


Fig. 8.2

sirven $x_0 = 1$, $x_1 = 3$, $x_2 = 5$, $x_3 = 7$, $x_4 = 9$ y los valores de la función subintegral $f(x) = 1/(x + 2)$ en estos puntos son tales:

$$y_0 = f(x_0) = 1/3; \quad y_1 = f(x_1) = 1/5;$$

$$y_2 = f(x_2) = 1/7; \quad y_3 = f(x_3) = 1/9; \quad y_4 = f(x_4) = 1/11.$$

Luego vamos a determinar el valor numérico de la integral, haciendo uso de la fórmula (2):

$$I_{izq} = \frac{b-a}{n}(y_0 + y_1 + y_2 + y_3) = 2\left(\frac{1}{3} + \frac{1}{5} + \frac{1}{7} + \frac{1}{9}\right) \approx 1,6024.$$

En cambio, si el cálculo de la integral definida se realiza con ayuda de la fórmula (3), obtendremos

$$I_d = \frac{b-a}{n}(y_1+y_2+y_3+y_4) = 2\left(\frac{1}{5}+\frac{1}{7}+\frac{1}{9}+\frac{1}{11}\right) \approx 1{,}1053. \ \blacktriangle$$

La fórmula de uso más frecuente basada en la idea de representación de la integral definida en forma de la suma integral es la de los *rectángulos* donde por ξ_i se toman los centros de los segmentos $[x_{i-1}, x_i]$. Para una red uniforme ($h_i = h$) esta fórmula tiene el aspecto siguiente:

$$I \equiv \int_a^b f(x)\,dx \approx \frac{b-a}{n}\sum_{i=1}^{n} f_{i-1/2} \equiv I_n, \tag{4}$$

donde $f_{i-1/2} = f\left(x_i - \frac{h}{2}\right)$; $\quad x_0 = a,\ x_n = b.$

Determinemos la expresión para el término residual de la fórmula aproximada (4). Con este fin representemos la integral que entra en el primer miembro de la relación (4) en forma de la suma:

$$\int_a^b f(x)\,dx = \sum_{i=1}^{n}\int_{x_{i-1}}^{x_i} f(x)\,dx. \tag{5}$$

Suponiendo que la función $f(x)$ es dos veces derivable, o sea $f \in C^2[a, b]$, escribamos para la funcion $f(x)$ sobre cada uno de los segmentos $[x_{i-1}, x_i]$ la fórmula de Taylor con el término residual en forma de Lagrange:

$$f(x) = f_{i-1/2} + \left(x - x_i + \frac{h}{2}\right) f'_{i-1/2} + \frac{\left(x - x_i + \frac{h}{2}\right)^2}{2} f''(\eta_i);$$
$$\eta_i \in (x_{i-1}, x_i). \tag{6}$$

Sustituyamos en el segundo miembro de la relación (5) en vez de la función f su representación (6) y ejecutemos la integración, utilizando el segundo teorema del valor medio de la función:

$$\int_a^b f(x)\,dx = (b-a)\frac{1}{n}\sum_{i=1}^{n} f_{i-1/2} + \frac{h^3}{24}\sum_{i=1}^{n} f''(\bar{\eta}_i);$$
$$\bar{\eta}_i \in (x_{i-1}, x_i).$$

En virtud de la continuidad de la segunda derivada existe tal punto $\eta \in (a, b)$ que

$$\sum_{i=1}^{n} f''(\bar{\eta}_i) = nf''(\eta) = \frac{b-a}{h} f''(\eta).$$

Utilizando esta relación, tenemos definitivamente

$$\int_a^b f(x)\,dx = (b-a)\sum_{i=1}^{n} \frac{1}{n}\cdot f_{i-1/2} + \frac{b-a}{h} h^2 f''(\eta). \tag{7}$$

Comparando las fórmulas (4) y (7), óbtenemos la expresión para el término residual de la fórmula de integración numérica (4):

$$R_n[f] = I - I_n = \frac{b-a}{24} h^2 f''(\eta). \tag{8}$$

Así pues, la estimación del error de la fórmula de integración numérica (4) puede ser representada en la forma siguiente:

$$\Delta_1 = \left| \int_a^b f(x)\,dx - (b-a)\sum_{i=1}^{n} \frac{1}{n} f_{i-1/2} \right| \leqslant \frac{b-a}{24} h^2 M_2, \tag{9}$$

$$\text{donde } M_2 = \max_{[a,\, b]} |f''(x)|.$$

Las expresiones obtenidas para el término residual (8) y para el error (9) muestran que la fórmula (4) es exacta para toda función lineal, puesto que la segunda derivada de tal función es igual a cero y, por consiguiente, el término residual y el error también son iguales a cero.

Mostremos que la estimación obtenida no puede ser mejorada, o sea, que existe una función para la cual el error de cálculo de la integral con ayuda de la fórmula (4) es exactamente igual al segundo miembro de (9). Para esto en calidad de función integrable cónsideremos $f = x^2$ y apliquemos a ésta la fórmula (4):

$$I_n = (b-a)\frac{1}{n}\sum_{i=1}^{n} \left(a + \frac{2i-1}{2} h\right)^2.$$

Abriendo paréntesis bajo el signo de la suma y realizando las sumaciones necesarias, obtenemos

$$I_n = \frac{b^3-a^3}{3} - (b-a)\frac{h^2}{12}.$$

Por otro lado, la integración inmediata de la función x^2 da

$$I = \int_a^b x^2\,dx = \frac{b^3-a^3}{3}.$$

Planteando la diferencia entre el valor exacto de la integral y su valor aproximado, para el término residual obtenemos la expresión

$$I - I_n = (b-a)\frac{h^2}{12}.$$

Volviendo a la estimación del error (9) y notando que para la función x^2 la segunda derivada (y, por consiguiente, también M_2)

es igual a 2, obtenemos para el error el mismo valor

$$\Delta_1 = (b-a)\frac{h^2}{12},$$

o sea, la estimación del error (9) se alcanza sobre la parábola $y = x^2$. Este resultado puede ser extendido también a una parábola arbitraria en virtud de la linealidad de la operación de integración y en virtud del hecho de que para las funciónes lineales la fórmula (4) es exacta.

Es evidente que la estimación (9) no tiene en cuenta los errores relaciónados con el cálculo de I_n. El error Δ_1 refleja la diferencia entre la fórmula exacta de Newton — Leibniz y la fórmula aproximada (4), o sea, es el error del método.

Pasemos ahora a examinar la estimación del error del valor aproximado $\overline{I}_n$. Si los valores de la función, utilizables en la fórmula de integración numérica, han sido obtenidos aproximadamente o si por cualesquiera causas los cálculos no pueden ser realizados con exactitud absoluta, esto da lugar a que aparezca un error de cálculos y errores de redondeo. Supongamos, por ejemplo, que los valores $f_{i-1/2}$ en la fórmula (4) han sido calculados con igual error absoluto ε; entonces el error total de cálculo $\overline{I}_n$ valdrá

$$\Delta_2 = (b-a)\sum_{i=1}^{n} \frac{1}{n}\varepsilon = (b-a)\varepsilon. \tag{10}$$

Nótese una particularidad característica de este error: no depende del número de particiones del segmento de integración y es proporcional sólo a su longitud.

Ejemplo 2. Con ayuda de la fórmula de los rectángulos calcular la integral $\int_0^1 \frac{dx}{1+x}$, suponiendo $n=4$. Estimar el error del valor aproximado obtenido.

△ Según los límites de integración y el número de particiones n determinemos el paso: $h = (1-0)/4 = 0{,}25$. Luego, en virtud de la fórmula (4), tenemos

$$I_4 = 0{,}25\left[f\left(\frac{1}{8}\right) + f\left(\frac{3}{8}\right) + f\left(\frac{5}{8}\right) + f\left(\frac{7}{8}\right)\right].$$

Calculando los valores necesarios de la funcion con tres cifras justas en sentido estrecho ($\varepsilon = 0{,}0005$), obtenemos

$$\int_0^1 \frac{dx}{1+x} = 0{,}25\,(0{,}889 + 0{,}727 + 0{,}615 + 0{,}533) = 0{,}691.$$

Estimemos el error del método con ayuda de la fórmula (9) para lo cual encontramos previamente el máximo del valor absoluto de la segunda derivada de la función subintegral:

$$M_2 = \max_{[0,\,1]} \left| \left(\frac{1}{1+x} \right)'' \right| = \max_{[0,\,1]} \frac{2}{(1+x)^3} = 2.$$

Así pues, el error del método vale

$$\Delta_1 \leqslant (1/24) \cdot 0{,}25^2 \cdot 2 \approx 0{,}0053.$$

Haciendo uso de la fórmula (10), hallamos el error de cálculo

$$\Delta_2 \leqslant 1 \cdot 0{,}0005 = 0{,}0005.$$

Por lo tanto, como error total del valor aproximado de la integral se puede tomar $\Delta = \Delta_1 + \Delta_2 = 0{,}006$ y escribir la respuesta definitiva en la forma $\int\limits_0^1 \frac{dx}{1+x} = 0{,}691 \pm 0{,}006.$

Para comparar citemos algunas cifras del valor exacto de la integral calculada: $\ln 2 = 0{,}693147 \ldots$ ▲

Ejemplo 3. Con ayuda de la fórmula de los rectángulos calcular la integral $\int\limits_0^1 \frac{dx}{1+x}$ con exactitud de hasta 0,001.

△ Empleando el algoritmo de resolución del problema II dado en el § 8.3, representemos el error total en forma de la suma de tres sumandos: $0{,}001 = 0{,}0009 + 0{,}00005 + 0{,}00005$. Luego elegimos n a partir de la condición

$$\Delta_1 = \frac{b-a}{24} h^2 M_2 = \frac{b-a}{24} \left(\frac{b-a}{n} \right)^2 M_2 \leqslant 0{,}0009.$$

Resolviendo esta desigualdad respecto a n, para $b - a = 1$ y $M_2 = 2$ obtenemos $n \geqslant 10$.

Hagamos la tabla de valores de la función $1/(x + 1)$ con cuatro cifras justas en sentido estrecho:

0,05	0,15	0,25	0,35	0,45	0,55	0,65	0,75	0,85	0,95
0,9254	0,8696	0,8000	0,7407	0,6897	0,6452	0,6061	0,5714	0,5405	0,5128

Utilizando la fórmula de los rectángulos (4), obtenemos

$$\bar{I}_{10} = 0{,}1\,(0{,}9524 + 0{,}8696 + 0{,}8000 + 0{,}7407 + 0{,}6897 + 0{,}6452 + 0{,}6061 + 0{,}5714 + 0{,}5405 + 0{,}5128) = 0{,}69284.$$

Redondeando el resultado obtenido, tenemos $I = 0{,}6928 \pm 0{,}001$. ▲

Fórmula de los trapecios. Vamos ahora a examinar otro método de construcción de las fórmulas de integración numérica, relacionado con la aproximación de la función subintegral por medio del polinomio interpolador. Consideremos el caso elemental de aproximación por el polinomio de primer grado con los nodos presentes en los puntos a y b:

$$f(x)=f(a)+\frac{x-a}{b-a}[f(b)-f(a)]+(x-a)(x-b)\frac{f''(\bar{\eta})}{2};$$

$$\bar{\eta}\in(a,\ b).$$

Integrando los miembros segundo y primero de esta igualdad y utilizando el segundo teorema del valor medio de la función al integrar el último sumando del segundo miembro, obtenemos

$$\int_a^b f(x)\,\mathrm{d}x=\frac{b-a}{2}[f(a)+f(b)]-\frac{(b-a)^3}{12}f''(\eta);\quad \eta\in(a,\ b).$$

Ahora bien, suponiendo que el segmento de integración es pequeño, obtenemos la fórmula de integración numérica llamada *fórmula de los trapecios*:

$$I\equiv\int_a^b f(x)\,\mathrm{d}x\ \approx\frac{b-a}{2}[f(a)+f(b)]\equiv I_2 \tag{11}$$

con el término residual

$$R_2[f]=I-I_2=-\frac{(b-a)^3}{12}f''(\eta);\quad \eta\in(a,\ b). \tag{12}$$

Utilizando la expresión (12) para el término residual, la estimación del error de la fórmula de integración numérica (11) puede ser representada en la forma

$$\Delta_1=\left|\int_a^b f(x)\,\mathrm{d}x-\frac{b-a}{2}(f(a)+f(b))\right|\leqslant\frac{(b-a)^3}{12}M_2, \tag{13}$$

$$\text{donde } M_2=\max_{[a,\ b]}|f''(x)|.$$

Las expresiones obtenidas para el término residual (12) y para el error (13) muestran que la fórmula de integración numérica (11) es exacta para todas las funciones lineales, puesto que la segunda derivada de tales funciones es igual a cero y, por consiguiente, son iguales a cero el término residual y el error.

De un modo análogo a como se ha hecho para la estimación (9) podemos mostrar que la estimación (13) también es inmejorable, puesto que se alcanza sobre una parábola arbitraria.

Cuando el cálculo se hace con ayuda de la fórmula (11), la estimación del error de cálculo para el caso en que los valores de la función están determinados con igual exactitud ε tiene la forma

$$\Delta_2\leqslant\frac{b-a}{2}(\varepsilon+\varepsilon)=(b-a)\,\varepsilon. \tag{14}$$

Nótese que los errores de cálculo de las fórmulas de integración numérica (11) y (4) son iguales.

Ejemplo 4. Con ayuda de la fórmula de los trapecios calcular la integral $\int\limits_0^1 \frac{dx}{1+x}$. Estimar el error del valor aproximado obtenido.

△ En virtud de la fórmula (11) tenemos

$$I_2 = 0{,}5\,[f(0) + f(1)].$$

Calculando los valores necesarios de la función, obtenemos

$$\int\limits_0^1 \frac{dx}{1+x} \approx 0{,}5\,(1+0{,}5) = 0{,}75.$$

Estimemos el error del método con ayuda de la fórmula (13), utilizando el valor M = 2, obtenido en el ejemplo 2

$$\Delta_1 \leqslant \frac{1^3}{12}\cdot 2 \approx 0{,}17.$$

Es evidente que el error de cálculo es igual a cero, ya que los valores de la función e I_2 están determinados con exactitud absoluta.

Así pues, finalmente tenemos $\int\limits_0^1 \frac{dx}{1+x} = 0{,}75 \pm 0{,}17$. ▲

Nótese que en el ejemplo 4 hemos obtenido una solución menos exacta que en el ejemplo 2. Sin embargo, no se debe sacar conclusiones apresuradas, ya que la utilización de la fórmula de los trapecios en el ejemplo 4 tiene también algunas ventajas. En primer lugar, si la función subintegral está representada en forma de una tabla de los valores que ella tiene en los nodos x_i, para utilizar la fórmula de los rectángulos es necesario determinar los valores que esta función tiene, además, en los puntos $x_i \pm h/2$, lo que origina dificultades adicionales e introduce un error adicional. En segundo lugar, en el ejemplo 4 los valores de la función subintegral han sido calculados sólo en dos puntos, mientras que en el ejemplo 2 este cálculo ha sido realizado en cuatro puntos, lo que, desde luego, ha exigido un mayor tiempo.

Los razonamientos citados muestran que la validez de la fórmula de integración numérica se determina no sólo por la forma de su término residual (o del error) sino también por otros factores, por ejemplo por el tiempo de cálculo.

Otros tipos de fórmulas de integración numérica. Consideremos un método más de construir las fórmulas de integración numérica, por medio del cual la integral se representa como combinación lineal de los valores que la función subintegral y sus derivadas tienen en algunos nodos x_i, con la determinación sucesiva de los coeficientes desconocidos (de los factores de ponderación).

Supongamos, por ejemplo, que construimos la fórmula de integración numérica que tiene el aspecto siguiente:

$$\int_a^b f(x)\,dx \approx (b-a)\,[A_1 f(a) + A_2 f(b) + A_3 f'(a) + A_4 f'(b)]. \qquad (15)$$

Determinemos los factores de ponderación A_i ($i = 1, 2, 3, 4$) de manera que la fórmula (15) sea exacta para polinomios arbitrarios de grado nulo, primero, segundo y tercero. Puesto que las operaciones de integración y derivación son lineales, esta condición será cumplida siempre que ella se cumpla para los polinomios 1, x, x^2, x^3.

Sustituyendo estos polinomios en vez de $f(x)$ en la relación (15), a condición de que ésta se cumpla exactamente obtenemos el siguiente sistema de ecuaciones lineales respecto a A_i:

$$\begin{cases} A_1 + A_2 = 1, \\ aA_1 + bA_2 + A_3 + A_4 = \frac{1}{2}(a+b), \\ a^2A_1 + b^2A_2 + 2aA_3 + 2bA_4 = \frac{1}{3}(a^2 + ab + b^2), \\ a^3A_1 + b^3A_2 + 3a^2A_3 + 3b^2A_4 = \frac{1}{4}(a^3 + a^2b + ab^2 + b^3). \end{cases}$$

Resolviendo este sistema, encontramos

$$A_1 = A_2 = 1/2; \quad A_3 = -A_4 = (b-a)/12.$$

Ahora bien, la fórmula buscada de integración numérica tiene el aspecto

$$I \equiv \int_a^b f(x)\,dx \approx (b-a)\left[\frac{f(a)+f(b)}{2} + (b-a)\,\frac{f'(a)-f'(b)}{12}\right] \equiv I_4. \quad (16)$$

Hallemos la expresión para el término residual de esta fórmula. Representemos la función integrable en forma de la suma del polinomio interpolador de Hermite de tercer grado con dos nodos dobles a y b y del término residual, y luego integremos los miembros segundo y primero de esta representación sobre el segmento $[a, b]$:

$$\int_a^b f(x)\,dx = \int_a^b H_3(x)\,dx + \int_a^b \frac{(x-a)^2(x-b)^2}{4!} f^{IV}(\bar{\eta})\,dx; \quad \bar{\eta} \in (a, b).$$

El primer sumando del segundo miembro da el segundo miembro de la fórmula de integración numérica (16) en virtud del hecho de que esta fórmula es exacta para todos los polinomios de tercer grado y, por consiguiente, también para el polinomio de Hermite $H_3(x)$. El segundo sumando del segundo miembro da la expresión para el término residual de la fórmula (16). Utilizando el segundo teorema del valor medio de la función y cumpliendo la integración, obtenemos

$$R_4[f] = I - I_4 = \frac{(b-a)^5}{720} f^{\mathrm{IV}}(\eta); \quad \eta \in (a,\ b). \tag{17}$$

La expresión obtenida para el término residual permite escribir la estimación del error de la fórmula de integración numérica (16) así:

$$\Delta_1 = |\, I - I_4 \,| \leqslant \frac{(b-a)^5}{720} M_4, \tag{18}$$

$$\text{donde } M_4 = \max_{[a,\ b]} |\, f^{\mathrm{IV}}(x) \,|.$$

La estimación (18) es inmejorable, ya que se alcanza con el polinomio arbitrario de cuarto grado lo que no es difícil demostrar de un modo análogo a como se ha hecho para la estimación (9).

Para estimar el error de cálculo del resultado obtenido con ayuda de la fórmula (16) supongamos que los valores de la función se dan con exactitud ε_1 y los valores de las derivadas, con exactitud ε_2. Entonces el error de cálculo vale

$$\Delta_2 \leqslant (b-a)\,\varepsilon_1 + \frac{(b-a)^2}{6}\,\varepsilon_1. \tag{19}$$

Ejemplo 5. Con ayuda de la fórmula de integración numérica (16) calcular la integral $\int\limits_0^1 \frac{dx}{1+x}$. Estimar el error del valor aproximado obtenido.

△ Calculando los valores necesarios de la función subintegral y de sus derivadas con ayuda de la fórmula (16), encontramos

$$I_4 = 1 \cdot \left(\frac{1+0{,}5}{2} \quad 1 \cdot \frac{-1+0{,}25}{12} \right) = 0{,}6875.$$

Haciendo uso de la fórmula (18) estimemos el error del método, para lo cual determinemos previamente el máximo del valor absoluto de la cuarta derivada de la función subintegral $M_4 = 24$:

$$\Delta_1 \leqslant \frac{1^5}{720} \cdot 24 \approx 0{,}034$$

El error de cálculo es, evidentemente, igual a cero, ya que los valores de la función y de las derivadas están calculados con exactitud absoluta.

Ahora bien, redondeando los valores aproximados de la integral y del error, finalmente obtenemos

$$\int\limits_0^1 \frac{dx}{1+x} = 0{,}69 + 0{,}004 \quad ▲$$

Hasta ahora en todas las fórmula de integración numérica examinadas los nodos de cuadratura estaban fijados de antemano. Vamos a considerar ahora el caso cuando la posición de todos los nodos, así como todos los factores de ponderación sa suponen parámetros libres. Para que los cálculos sucesivos no sean demasiado complicados, pero al mismo tiempo no completamente triviales, buscaremos el valor de la integral en la forma

$$\int_a^b f(x)\,dx = (b-a)\,[A_1 f(x_1) + A_2 f(x_2)]. \tag{20}$$

Con el fin de determinar los cuatro parámetros libres A_1, A_2, x_1 y x_2 exijamos que la fórmula (20) sea absolutamente exacta para todos los polinomios de grados nulo, primero, segundo y tercero. Puesto que la operación de integración y el segundo miembro de la relación (20) son lineales, con el fin de que la fórmula de integración numérica (20) son lineales, con el fin de que la fórmula de integración numérica (20) sea exacta para todos los polinomios de tercer grado es necesario y suficiente que sea exacta para las funciones 1, x, x^2, x^3. Por consiguiente, han de cumplirse las relaciones

$$\begin{cases} A_1 + A_2 = 1, \\ x_1A_1 + x_2A_2 = \frac{1}{2}(a+b), \\ x_1^2A_1 + x_2^2A_2 = \frac{1}{3}(a^2+ab+b^2), \\ x_1^3A_1 + x_2^3A_2 = \frac{1}{4}(a^3+a^2b+ab^2+b^3), \end{cases}$$

que son un sistema no lineal de ecuaciones respecto a los parámetros a determinar x_1, x_2, A_1, A_2.

La solución de este sistema es así:

$$A_1 = A_2 = \frac{1}{2},$$

$$x_1 = \frac{b+a}{2} - \frac{b-a}{2}\cdot\frac{1}{\sqrt{3}}, \quad x_2 = \frac{b+a}{2} + \frac{b-a}{2}\cdot\frac{1}{\sqrt{3}}. \tag{21}$$

Así pues, la fórmula de integración numérica (20) toma el aspecto siguiente:

$$I \equiv \int_a^b f(x)\,dx \approx \frac{b-a}{2}\left[f\left(\frac{b+a}{2} - \frac{b-a}{2}\frac{1}{\sqrt{3}}\right) + f\left(\frac{b+a}{2} + \frac{b-a}{2}\frac{1}{\sqrt{3}}\right)\right] \equiv I_2. \tag{22}$$

Las fórmulas de tal tipo, cuando no sólo los factores de ponderación sino también los nodos no se fijan de antemano, se llaman *fórmlas gaussianas*.

Determinemos la expresión para el término residual de la fórmula (22). Para esto representemos la función integrable en forma de la suma del polinomio interpolador de Hermite de tercer grado con dos nodos dobles x_1 y x_2, que se determinan por las relaciones (21), y del término residual. Integrando los miembros segundo y primero de esta representación sobre el segmento $[a, b]$, obtenemos

$$\int_a^b f(x)\,dx = \int_a^b H_3(x)\,dx + \int_a^b \frac{(x-x_1)^2\,(x-x^2)^2}{4!} f(\bar{\eta})\,dx; \quad \bar{\eta}\in(a, b)$$

El primer sumando del segundo miembro dará el segundo miembro de la fórmula de integración numérica (22), puesto que esta fórmula es exacta para todos los polinomios de tercer grado y, por consiguiente, también para $H_3(x)$. El segundo sumando del segundo miembro dará el término residual de la fórmula (22). Utilizando el segundo teorema del valor medio de la función y ejecutando la integración, tenemos

$$R_2[f] = I - I_2 = \frac{(b-a)^5}{4320} f^{\mathrm{IV}}(\eta); \quad \eta\in(a, b). \tag{23}$$

Por lo tanto, la estimación del error se expresa por la relación

$$\Delta_1 = |I - I_2| \leqslant \frac{(b-a)^5}{4320} M_4, \tag{24}$$

$$\text{donde } M_4 = \max_{[a,\, b]} |f^{\mathrm{IV}}(x)|.$$

La estimación obtenida es inmejorable, ya que se alcanza sobre un polinomio arbitrario de segundo grado, lo que no es difícil mostrar por los cálculos directos como lo hemos hecho para la estimación (9).

Si en la fórmula (22) los valores de los nodos están determinados con exactitud práctica y los valores de la función, con el error absoluto ε, entonces para el error de cálculo al utilizar la fórmula (2) obtendremos la misma expresión que para el error de cálculo al utilizar las fórmulas (4) y (11).

$$\Delta_2 \leqslant (b-a)\,\varepsilon. \tag{25}$$

Ejemplo 6. Con ayuda de la fórmula de integración numérica (22) calcular la integral $\int_0^1 \frac{dx}{1+x}$. Estimar el error del valor aproximado obtenido.

△ Ante todo determinemos los nodos de la fórmula de integración numérica:

$$x_1 = \frac{1}{2}\left(1 - \frac{1}{\sqrt{3}}\right) = 1{,}2113249\ldots;$$

$$x_2 = \frac{1}{2}\left(1 + \frac{1}{\sqrt{3}}\right) = 0{,}7886751\ldots$$

Ahora, calculados los valores necesarios de la función integrable con exactitud hasta tres cifras justas en sentido estrecho, hagamos uso de la fórmula (22):

$$\int_0^1 \frac{dx}{1+x} \approx \frac{1}{2}(0,826+0,559) = 0,6925.$$

Vamos a estimar el error del método con ayuda de la fórmula (24), para lo cual utilizamos el valor máximo absoluto de la derivada cuarta $M_4 = 24$, valor hallado en el ejemplo 3:

$$\Delta_1 \leqslant \frac{1^5}{4320} \cdot 24 \approx 0,0056.$$

El error de cálculo puede ser hallado con ayuda de la fórmula (24), teniendo en cuenta que la exactitud de cálculo de los valores de la función integrable es igual a 0,0005.

Ahora bien, el error total es $\Delta = \Delta_1 + \Delta_2 = 0,0061$.

Por último, redondeando el valor aproximado de la integral, finalmente tenemos $\int_0^1 \frac{dx}{1+x} = 0,692 \pm 0,007$. ▲

El objetivo principal del presente párrafo consiste en mostrar, citando ejemplos sencillos, cómo se deducen las diferentes fórmulas de integración numérica. Claro está que hemos considerado no todos los métodos de construcción de las fórmulas. Al mismo tiempo los ejemplos citados son característicos, así que, utilizándolos, se puede construir una fórmula concreta de integración numérica que sea la que más corresponde al problema práctico dado.

§ 8.5. Fórmulas de integración numérica de Newton—Cotes

En este párrafo vamos a considerar las fórmulas de integración numérica que tienen una estructura más complicada. Los métodos de integración numérica examinados hasta ahora eran los de interpolación, incluyendo, en cierto sentido, también las fórmulas de los rectángulos. Esto quiere decir que la función subintegral se aproximaba por medio del polinomio interpolador. Si la función integrable es bastante suave y el segmento de integración es finito, se pueden obtener resultados buenos. Por otro lado, es difícil suponer que se alcance una buena aproximación de la función integrable si la misma función o sus derivadas de bajos órdenes tienen particularidades. En tales casos es conveniente representar la función subintegral como producto de dos factores: $\rho(x) f(x)$ que deben poseer las tres propiedades siguientes. En primer lugar, el factor ponderal $\rho(x)$ debe reflejar todas las peculiaridades de la función integrable. En segundo lugar, los momentos

$$m_k = \int_a^b \rho(x) x^k \, dx \quad (k = 0, 1, \ldots), \tag{1}$$

donde $[a, b]$ es el segmento de integración, deben calcularse analíticamente. En tercer lugar, el error de aproximación de la función $f(x)$ por medio del polinomio debe ser pequeño.

Pasemos ahora a construir las propias fórmulas de integración numérica. Vamos a construirlas en la misma forma que antes:

$$I \equiv \int_a^b \rho(x) f(x)\,\mathrm{d}x \approx (b-a) \sum_{i=1}^{n} A_i f(x_i) \equiv I_n. \tag{2}$$

Como ya hemos mencionado, en el caso general la fórmula (2) tiene $2n$ parámetros libres, que son los nodos de la cuadratura x_i y los factores ponderales A_i. Supondremos fijo el número n. La elección de los parámetros libres se determina por los requisitos que debe reunir la fórmula de integración numérica según las condiciones del problema práctico. Tales requisitos pueden ser, por ejemplo, la exactitud máxima posible, de error de cálculo mínimo, la fijación de algunos (y posiblemente de todos) factores ponderales o de los nodos de cuadratura.

Comencemos con un caso relativamente simple cuando los nodos están determinados de antemano y se puede variar sólo con la elección de los factores ponderales A_i. La idea de las cuadraturas de interpolación consiste en lo siguiente. Aproximemos la función f por medio del polinomio interpolador en la forma de Lagrange de grado $n-1$ respecto de n nodos diferentes x_i:

$$f(x) = L_{n-1}(x) + R_{n-1}(x).$$

Integremos los miembros segundo y primero de esta igualdad sobre el segmento $[a, b]$, multiplicándolos previamente por la función ponderante $\rho(x)$:

$$\int_a^b \rho(x) f(x)\,\mathrm{d}x = \int_a^b \rho(x) L_{n-1}(x)\,\mathrm{d}x + \int_a^b \rho(x) R_{n-1}(x)\,\mathrm{d}x. \tag{3}$$

Transformemos el primer sumando del segundo miembro de esta relación. Para esto sustituyamos en vez de L_{n-1} su expresión explícita y cambiemos de lugar las operaciones de integración y de sumación:

$$\int_a^b \rho(x) f(x)\,\mathrm{d}x = \sum_{i=1}^{n} \left(\int_a^b \rho(x) \frac{\omega_{n-1}(x)}{(x-x_i)\,\omega'_{n-1}(x_i)}\,\mathrm{d}x \right) f_i. \tag{4}$$

Aquí $\omega_{n-1}(x) = \prod_{k=1}^{n} (x - x_k)$. El primero de los factores que entran bajo el signo de suma es el coeficiente numérico que es proporcional a la longitud del segmento de integración y depende sólo de la disposición de los nodos y de las propiedades de la función $\rho(x)$ (pero no de la función f).

Suponiendo ahora que el segundo sumando del segundo miembro de la relación (3) es pequeño, obtenemos la fórmula aproximada de integración numérica (2) con los nodos x_i dados y con los coeficientes A_i que se determinan del modo siguiente:

$$A_i = \frac{1}{b-a} \int_a^b \rho(x) \frac{\omega_{n-1}(x)}{(x-x_i)\,\omega'_{n-1}(x_i)}\,dx \quad (i=1, 2, \ldots n). \tag{5}$$

La fórmula de integración numérica (2) construida de este modo se lama *fórmula de interpolación.*

Pasemos a estimar el error de la fórmula (2) con los coeficientes (5). Para esto vamos a integrar el término residual de la fórmula de interpolación $R_{n-1}(x) = \frac{\omega_{n-1}(x)}{n!} f^{(n)}(\bar{\eta})$. Sustituyendo la expresión dada para $R_{n-1}(x)$ en el segundo sumando de la igualdad (3), obtenemos

$$R_{n-1}[f] = I - I_n = \frac{1}{n!} \int_a^b \rho(x)\,\omega_{n-1}(x)\,f^{(n)}(\bar{\eta})\,dx; \quad \bar{\eta} \in (a, b).$$

Si la función f tiene la derivada continua de orden n sobre el segmento de integración y el producto $\rho(x)\,\omega_n(x)$ conserva sobre el mismo segmento el propio signo, para el término residual se puede obtener la expresión siguiente:

$$R_{n-1}[f] = \frac{f^{(n)}(\eta)}{n!} \int_a^b \rho(x)\,\omega_{n-1}(x)\,dx; \quad \eta \in (a, b). \tag{6}$$

Por consiguiente, la estimación del error de la cuadratura toma la forma:

$$\Delta_1 \leqslant \frac{M_n}{n!} \left| \int_a^b \rho(x)\,\omega_{n-1}(x)\,dx \right|, \tag{7}$$

donde $M_n = \max\limits_{[a,\, b]} |f^{(n)}(x)|$. Para las condiciones anteriormente indicadas la estimación obtenida es inmejorable.

En cambio, si el producto $\rho(x)\,\omega_{n-1}(x)$ no conserva el signo sobre el segmento de integración, en este caso se obtiene sólo una estimación muy aproximada del error

$$\Delta_1 \leqslant \frac{M_n}{n!} \int_a^b |\rho(x)\,\omega_{n-1}(x)|\,dx$$

la cual puede estar lejos de la óptima. Por eso en los casos semejantes se utilizan otras consideraciones al construir una expresión explícita para el término residual y para el error. Examinemos más abajo uno de tales procedimientos al estudiar la fórmula de Simpson.

Ejemplo 1. Construir la fórmula de integración numérica (2) para el segmento $[-1, 1]$ con los nodos $x_1 = -1$, $x_2 = 0$, $x_3 = 1$ y con la función ponderante $\rho(x) = (1 - x^2)^{-1/2}$.

△ En realidad, es necesario determinar los coeficientes A_i ($i = 1, 2, 3$) de la fórmula (2). Utilizando la expresión (5) para los coeficientes buscados, tenemos

$$2A_1 = \int_{-1}^{1} \frac{(x-0)(x-1)}{(-1-0)(-1-1)} \cdot \frac{dx}{\sqrt{1-x^2}} = \frac{\pi}{4};$$

$$2A_2 = \int_{-1}^{1} \frac{(x+1)(x-1)}{(0+1)(0-1)} \cdot \frac{dx}{\sqrt{1-x^2}} = \frac{\pi}{2};$$

$$2A_3 = \int_{-1}^{1} \frac{(x+1)(x-0)}{(1+1)(1-0)} \cdot \frac{dx}{\sqrt{1-x^2}} = \frac{\pi}{4}.$$

Ahora bien, la fórmula buscada tiene el aspecto

$$\int_{-1}^{1} f(x) \frac{dx}{\sqrt{1-x^2}} \approx \frac{\pi}{4} [f(-1) + 2f(0) + f(1)]. \ ▲$$

Ejemplo 2. Construir la fórmula de integración numérica (2) para el segmento $[0, 1]$ con los nodos $x_1 = 0$, $x_2 = 0{,}5$, $x_3 = 1$ y con la función ponderante $\rho(x) = \ln x$.

△ Al igual que en el ejemplo precedente, haciendo uso de la fórmula (5), determinemos los coeficientes:

$$A_1 = \int_0^1 \ln x \frac{(x-0{,}5)(x-1)}{(0-0{,}5)(0-1)} dx = -\frac{17}{36};$$

$$A_2 = \int_0^1 \ln x \frac{(x-0)(x-1)}{(0{,}5-0)(0{,}5-1)} dx = -\frac{20}{36};$$

$$A_3 = \int_0^1 \ln x \frac{(x-0)(x-0{,}5)}{(1-0)(1-0{,}5)} dx = \frac{1}{36}.$$

Así pues,

$$\int_0^1 \ln x\, f(x)\, dx \approx -\frac{1}{36} [17f(0) + 20f(0{,}5) - f(1)]. \ ▲$$

Nótese las particularidades características de los ejemplos considerados. En el ejemplo 1 los nodos dispuestos simétricamente y la paridad de la función ponderante respecto al centro del segmento han conducido a la simetría de los coeficientes de integración numérica. En el ejemplo 2, a pesar de la simetría de los nodos, la simetría de los coeficientes está alterada lo que es consecuencia de la falta de simetría (de paridad) en la función ponderante.

- En los cálculos prácticos representa especial interés el caso cuando los nodos de la fórmula de integración numérica se dan en forma de puntos equidistantes del segmento $[a, b]$: $x_i = a + (i-1)h$ $(i = 1, 2, \ldots, n)$ y la función ponderante $\rho(x)$ es idénticamente igual a la unidad. Con tales suposiciones la fórmula (2) puede transformarse así:

$$\int_a^b f(x)\,dx \approx (b-a) \sum_{i=1}^{n} H_i f_i. \qquad (8)$$

Para diferentes n obtenemos diferentes *fórmulas de integración numérica de Newton-Cotes*. Los coeficientes H_i, llamados *coeficientes de Cotes*, se determinan por la relación (5):

$$H_i = A_i = \frac{(-1)^{n-i}}{(n-1)(i-1)!(n-1)!} \int_1^n \frac{(t-1)\ldots(t-n)}{t-i}\,dt; \qquad (9)$$

$$n > 1; \quad i = 1, 2, \ldots, n; \ 0! = 1.$$

Estos coeficientes poseen las siguientes propiedades, útiles para calcularlos.

1°. *Los coeficientes simétricos (primero y* n-esimo, *segundo y* $n-1, \ldots$) *son iguales entre sí*:

$$H_i = H_{n+1-i}.$$

□ Reemplacemos en la expresión (9) i por $n+1-i$:

$$H_{n+1-i} = \frac{(-1)^{i-1}}{(n-1)(n-i)!(i-1)!} \int_1^n \frac{(t-1)\cdots(t-n)}{i-n-1+i}\,dt.$$

Pasando bajo el signo integral a una nueva variable $q = n + 1 - t$ y realizando transformaciones poco complicadas, obtenemos

$$H_{n+1-i} = \frac{(-1)^{n-i}}{(n-1)(n-i)!(i-1)!} \int_1^n \frac{(q-n)\ldots(q-1)}{q-i}\,dq$$

lo que coincide con la expresión (9) para los coeficientes H_i. ■

2°. La suma de todos los coeficientes es igual a la unidad:

$$\sum_{i=1}^{n} H_i = 1.$$

□ La validez de esta propiedad se deduce inmediatamente de la fórmula (8), si se pone $f(x) = 1$, puesto que el término residual de esta fórmula, determinado por la expresión (6), es igual a cero cuando $f(x) = 1$. ■

En la tabla 8.1 se dan los valores de los coeficientes de Cotes para $n = 2, 3, 4, 5, 6$.

Tabla 8.1

$n \setminus i$	1	2	3	4	5	6
2	1/2	1/2				
3	1/6	4/6	1/6			
4	1/8	3/8	3/8	1/8		
5	7/90	32/90	12/90	32/90	7/90	
6	19/288	75/288	50/288	50/288	75/288	19/288

Examinemos con más detalles un importante caso particular de la fórmula de integración numérica (8), el cual se obtiene para $n = 3$. Con el fin de construir esta fórmula se podría utilizar los datos de la tabla 8.1. No obstante, en calidad de ejemplo realicemos los cálculos correspondientes. Aplicando la fórmula (9), obtenemos

$$H_1 = \frac{(-1)^2}{2 \cdot 1 \cdot 2} \cdot \int_1^3 (t-2)(t-3)\,\mathrm{d}t = \frac{1}{6}.$$

Luego, utilizando las igualdades (10), encontramos

$$H_3 = H_1 = 1/6,\ H_2 = 1 - (H_1 + H_3) = 2/3.$$

Ahora bien, la fórmula buscada de integración numérica, llamada *fórmula de Simpson*, tiene el aspecto

$$I \equiv \int_a^b f(x)\,\mathrm{d}x \approx \frac{b-a}{6}\left[f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right] \equiv I_3. \quad (11)$$

Esta fórmula en virtud de su construcción (en virtud de la aproximación de la función subintegral (por medio del polinomio de segundo grado) es exacta para todos los polinomios de grados nulo, primero y segundo. Se podría probar a obtener la expresión para el término residual utilizando inmediatamente la relación (6). Sin embargo, la fórmula de Simpson posee la así llamada **propiedad de exactitud elevada** que consiste en que es exacta no sólo para los polinomios de segundo grado sino también para los de tercer grado. En virtud de la linealidad de la operación de integración para demostrar esta afirmación es suficiente establecer la igualdad exacta de los miembros primero y segundo de la fórmula (11) para el polinomio elemental de tercer grado x^3. En efecto, calculando el primer miembro de la fórmula (11) para $f = x^3$, tenemos

$$\int_a^b x^3\,dx = \frac{b^4-a^4}{4}.$$

Por otro lado, calculando el segundo miembro de la fórmula (11), obtenemos

$$\frac{b-a}{6}\left[a^3+4\left(\frac{b+a}{2}\right)^3+b^3\right]=\frac{b^4-a^4}{4},$$

lo que se trataba de demostrar.

Hagamos uso de la afirmación demostrada para construir el término residual de la fórmula de Simpson. Representemos la función f como suma del polinomio interpolador de Hermite, con nodos simples a y b y con un nodo doble $(a+b)/2$, y del término residual:

$$f(x)=H_3(x)+(x-a)\left(x-\frac{a+b}{2}\right)^2(x-b)\frac{f^{IV}(\bar{\eta})}{4!}\ ;\quad \bar{\eta}\in(a,\ b).$$

Vamos a integrar los miembros segundo y primero de esta igualdad sobre el segmento $[a,\ b]$. En virtud de lo recién demostrado la integral del polinomio de Hermite dará el segundo miembro de la fórmula (11) y la integral del segundo sumando, el término residual de esta fórmula

$$R_3[f]=I-I_3=-\frac{(b-a)^5}{2880}f^{IV}(\eta);\quad \eta\in(a,\ b). \tag{12}$$

Por consiguiente, la estimación del error puede ser escrita en la forma

$$\Delta_1\leqslant|\ I-I_3\ |\leqslant\frac{(b-a)^5}{2880}\cdot M_4, \tag{13}$$

donde $M_4 = \max\limits_{[a,\ b]} |\ f^{IV}\ (x)\ |$. Esta estimación es inmejorable, puesto que se alcanza, por ejemplo, en la función $f = x^4$.

Ejemplo 3. Con ayuda de la fórmula de Simpson calcular la integral $\int\limits_0^1 \frac{dx}{1+x}$. Estimar el error del valor aproximado obtenido.

$\triangle$ Calculando los valores necesarios de la función subintegral en los puntos $x_1 = 0$, $x_2 = 0{,}5$ y $x_3 = 1$, los sustituimos en la fórmula (11):

$$I_3=\frac{1}{6}(1+4\cdot 0{,}667+0{,}5)=0{,}6947.$$

Teniendo en cuenta que $M_4=\max\limits_{[0,\ 1]}\left|\left(\frac{1}{1+x}\right)^{IV}\right|=24$ y utilizando la fórmula (13), para el error del método obtenemos $\Delta_1\leqslant 0{,}0084$.

Hallamos el error de cálculo

$$\Delta_2 \leqslant \frac{1}{6} \cdot (0+4 \cdot 0{,}0005+0)=0{,}00034.$$

Sumando los errores y redondeando el resultado, finalmente obtenemos $\int\limits_0^1 \frac{dx}{1+x}=0{,}695 \pm 0{,}01$. ▲

La propiedad de exactitud elevada, descrita anteriormente, la poseen todas las fórmulas de integración numérica del tipo (8) construidas con un número impar de nodos. Esta propiedad es consecuencia directa de la «simetría» de la cuadratura, o sea, de la igualdad de los coeficientes simétricos $H_i = H_{n+1-i}$ y consecuencia directa de la linealidad de las operaciones de cálculo de los valores de la función y de las de integración. Para obtener una estimación precisada del término residual de tal fórmula es necesario aproximar la función subintegral por el polinomio interpolador de Hermite con doble nodo central.

Retornando a la fórmula de integración numérica (2), se puede afirmar que para los coeficientes simétricos ($A_i = A_{n+1-i}$) y para la función ponderante par esta cuadratura construida sobre $2k+1$ nodos es exacta para todos los polinomios de grado $2k+1$, puesto que es exacta para toda función f que sea impar respecto al centro del segmento de integración. En efecto, por un lado, para tales funciones $\int\limits_a^b \rho(x) f(x)\, dx = 0$ y por el otro, $\sum\limits_{i=1}^{2k+1} A_i f_i = 0$ en virtud de la simetría de A_i y de la imparidad de f.

Ejemplo 4. Utilizando la fórmula de integración numérica construida en el ejemplo 1, calcular la integral $\int\limits_{-1}^{1} \frac{\cos x\, dx}{\sqrt{1-x^2}}$ y estimar el error.

△ Calculamos los valores necesarios de la función $f = \cos x$: $\cos(-1) = 0{,}540$; $\cos 0 = 1$; $\cos 1 = 0{,}540$. Sustituyendo estos valores en la fórmula de integración numérica, tenemos

$$\int\limits_{-1}^{1} \frac{\cos x\, dx}{\sqrt{1-x^2}} \approx \frac{\pi}{4}(0{,}540+2 \cdot 1+0{,}540)=2{,}419.$$

Ahora encontramos la expresión para el término residual de la fórmula de integración numérica escrita en el ejemplo 1. Puesto que la función ponderante es par y los nodos son de disposición simétrica y su número es impar, esta fórmula posee la propiedad de exactitud elevada, o sea, es exacta para todos los polinomios de tercer grado. Por eso de un modo análogo a cómo lo hemos hecho para la fórmula de Simpson, utilizando el polinomio interpolador de Hermite con el nodo central doble $x_2 = 0$, obtenemos la siguiente expresión para el término residual:

$$R_3[f]=\frac{f^{\mathrm{IV}}(\eta)}{4!}\int\limits_{-1}^{1}\frac{(x+1)\,x^2\,(x-1)}{\sqrt{1-x^2}}\,\mathrm{d}x=-\frac{\pi}{192}f^{\mathrm{IV}}(\eta);\quad \eta\in(-1,\ 1).$$

Por lo tanto, el error del método vale

$$\Delta_1\leqslant\frac{\pi}{192}\max_{[-1,\,1]}|\cos x|=0{,}017.$$

Luego, puesto que los valores primero y tercero del coseno están calculados con un error de 0,0005 y el segundo, con una exactitud absoluta, para el error de cálculo obtenemos

$$\Delta_2\leqslant\frac{\pi}{4}\,(0{,}0005+2\cdot 0+0{,}0005)=0{,}0008.$$

Sumando los errorres y realizando los redondeos necesarios, finalmente tenemos

$$\int\limits_{-1}^{1}\frac{\cos x\,\mathrm{d}x}{\sqrt{1-x^2}}=2{,}42\pm 0{,}02.\ \blacktriangle$$

Para concluir examinemos un procedimiento más de determinar los coeficientes A_i de la fórmula de integración numérica (2) con los nodos dados. Exijamos que la fórmula (2) sea exacta para las funciones $1,\ x,\ x^2,\ \ldots$ hasta un grado lo más alto posible. En este caso para hallar los coeficientes A_i obtenemos el siguiente sistema de ecuaciones:

$$m_k=(b-a)\sum_{i=1}^{n}x_i^kA_i\quad (k=0,\ 1,\ \ldots,\ N).\tag{14}$$

Si los nodos x_i son no coincidentes y $N=n-1$, el determinante del sistema (14) es el determinante de Vandermond y la solución de este sistema (la colección de los coeficientes A_i) existe y es única.

Mostremos que tal método de determinar los coeficientes de integración numérica es equivalente al descrito anteriormente, o sea, que las fórmulas (5) dan los mismos valores para A_i que el sistema (14). Con el fin de demostrar esto sustituyamos las expresiones (5) para los coeficientes A_i en los segundos miemros de las ecuaciones (14) y cambiemos de lugar las operaciones de sumación y de integración:

$$m_k=\int\limits_{a}^{b}\rho(x)\left[\sum_{i=1}^{n}\frac{\omega_{n-1}(x)}{(x-x_i)\,\omega_{n-1}'(x_i)}\,x_i^k\right]\mathrm{d}x.$$

La expresión puesta entre paréntesis es un polinomio interpolador de Lagrange de grado no superior a $n-1$ para la función x^k $(k<n)$. El término residual de tal polinomio es, evidentemente,

igual a cero $\left(\frac{d^n}{dx^n}(x^k)=0 \text{ para } k<n\right)$; por eso

$$m_k=\int_a^b \rho(x)\,x^k\,dx \qquad (k=0,\ 1,\ \ldots,\ n-1).$$

De las identidades obtenidas y de la unicidad de la solución del sistema (14) se desprende la equivalencia de ambos métodos de construcción de las fórmulas de integración numérica (2).

A título de ejemplo vamos a construir la fórmula que tiene el aspecto

$$\int_a^b f(x)\,\mathrm{d}x \approx (b-a)\left[A_1 f'(a)+A_2 f\left(\frac{a+b}{2}\right)+A_3 f'(b)\right].$$

Con el fin de construir la fórmula buscada de integración numérica exijamos que ésta sea exacta para todos los polinomios de grados nulo, primero y segundo. Para esto, en virtud de la linealidad de las operaciones de integración, del cálculo de los valores de la función y de la derivación, es suficiente exigir que la fórmula de integración numérica sea exacta para 1, x, x^2. Así pues, sea $f(x)=1$; entonces $(b-a)=(b-a)(A_1\cdot 0+A_2\cdot 1+A_3\cdot 0)$, es decir, $A_2=1$. Sea ahora $f(x)=x$; entonces

$$\frac{b+a}{2}=A_1\cdot 1+1\cdot\frac{b+a}{2}+A_3\cdot 1,\ \text{ es decir, } A_1=-A_3.$$

Por último, suponiendo $f(x)=x^2$, obtenemos la ecuación

$$\frac{a^2+ab+b^2}{3}=A_1\cdot 2a+1\cdot\left(\frac{a+b}{2}\right)^2-A_1\cdot 2b;$$

al resolver esta ecuación, encontramos $A_1=-\frac{b-a}{24}$.

Ahora bien, la fórmula buscada se escribe así:

$$I\equiv\int_a^b f(x)\,\mathrm{d}x \approx \frac{b-a}{24}\Big[-(b-a)f'(a)+24f\left(\frac{a+b}{2}\right)+$$
$$+(b-a)f(b)\Big]\equiv I_3.$$

Nótese que esta fórmula también tiene la propiedad de exactitud elevada, como, por ejemplo, la tiene la fórmula de Simpson. Citemos, sin deducir, la expresión para el término residual de la fórmula obtenida:

$$R_3[f]=I-I_3=-\frac{7}{5760}(b-a)^5 f^{\mathrm{IV}}(\eta);\quad \eta\in(a,\ b).$$

Proponemos que en calidad de ejercicio el mismo lector demuestre la última relación.

§ 8.6. Fórmulas de integración numérica de grado algebraico superior de precisión

Volvamos a considerar la fórmula de integración numérica

$$I \equiv \int_a^b \rho(x) f(x)\, dx \approx (b-a) \sum_{i=1}^{n} A_i f(x_i) \equiv I_n. \qquad (1)$$

Según ya hemos mencionado, en el caso general no sólo los coeficientes A_i sino también los nodos x_i son parámetros arbitrarios que deben determinarse en dependencia de los requisitos que ha de reunir la relación (1). Puesto que el número total de parámetros libres es igual a $2n$, se puede esperar que superponiendo $2n$ condiciones a la relación (1), obtendremos un sistema de ecuaciones para determinar estos parámetros. Desde luego, en el caso general es imposible decir si tiene solución tal sistema y si la tiene, si es o no única. La respuesta a esta pregunta puede obtenerse sólo si se consideran los requisitos concretos a los que debe satisfacer la fórmula de integración numérica (1). Por lo visto, conviene relacionar estos requisitos con la magnitud del error de la fórmula (1), procurar elegir los nodos y coeficientes de un modo tal que se miniminice, en cierto sentido, el valor absoluto del término residual. En el párrafo precedente hemos señalado que las estimaciones obtenidas del error de las fórmulas de integración numérica se alcanzan con los polinomios cuyo grado es en una unidad mayor que el grado máximo del polinomio, para el cual la cuadratura correspondiente es exacta. Por eso es natural que se pruebe a aumentar el grado del polinomio para el cual la fórmula (2) sería absolutamente exacta. Tal planteamiento de la cuestión origina el siguiente problema de optimización.

Construir una fórmula de integración numérica tipo (1), con n fijo, la cual sea exacta para un polinomio arbitrario de grado más alto posible r.

Mostremos que el problema enunciado es equivalente al de construir una fórmula de integración numérica (1) que sea exacta para todas las funciones x^k $(k = 0, 1, \ldots, r)$. En efecto, supongamos que se da un polinomio arbitrario $P_r(x) = \sum_{k=0}^{r} a_k x^k$. Entonces

$$R_n[P_r] = I[P_r] - I_n[P_r] =$$

$$= \sum_{k=0}^{r} a_k \left(\int_a^b \rho(x) x^k\, dx - \sum_{i=1}^{n} A_i x_i^k \right) = \sum_{k=0}^{r} a_k R_n[x^k]. \qquad (2)$$

De aquí en virtud de la arbitrariedad de a_k (el polinomio P_r es arbitrario) obtenemos que para $R_n[P_r] = 0$ es necesario $R_n[x^k] = 0$ $(k = 0, 1, \ldots, r)$. La suficiencia es consecuencia directa de la linealidad del término residual de la fórmula (1) respecto a la función f y es evidente en virtud de la misma relación (2).

Así pues, llegamos al sistema de $r+1$ ecuaciones respecto a $2n$ parámetros desconocidos A_i y x_i:

$$m_k = \int_a^b \rho(x)\, x^k \, dx = (b-a) \sum_{i=1}^{n} A_i x_i^k \quad (k=0, 1, \ldots, r). \tag{3}$$

Es natural probar a resolver este sistema, o sea, construir la fórmula de integración numérica para $r+1=2n$ (el número de ecuaciones es igual al de incógnitas.) La racionalidad de tal prueba es confirmada por los teoremas que presentamos a continuación.

Primero enunciemos, sin demostrar, el teorema auxiliar que caracteriza la propiedad de los polinomios ortogonales.

Teorema 1. *Sea* 1°) $\rho(x) > 0$ *casi por doquier sobre* $[a, b]$; 2°) $P_{n-1}(x)$, *un polinomio arbitrario de grado no superior a* $n-1$. *Entonces existe un polinomio y además único*

$$\Psi_n(x) = (x-x_1)(x-x_2)\ldots(x-x_n), \tag{4}$$

ortogonal con el peso $\rho(x)$ a $P_{n-1}(x)$, o sea, tal que

$$\int_a^b \rho(x)\, \Psi(x)\, P_{n-1}(x)\, dx = 0, \tag{5}$$

con la particularidad de que todas las raíces $x_1, x_2, \ldots, x_n$ *del polinomio* $\Psi_n(x)$ *son distintas y están dentro del segmento* $[a, b]$.

Pasemos ahora a los teoremas que permiten determinar inmediatamente los nodos y coeficientes de la fórmula (1), exacta para todos los polinomios de grado $r = 2n-1$.

Teorema 2. *Para que la fórmula* (1) *sea exacta para los polinomios de grado* $2n-1$ *es necesario y suficiente que*: 1°) *los nodos* x_i *sean raíces del polinomio* Ψ_n *determinado por la relación* (5); 2°) *los factores de ponderación* A_i *se determinen por las relaciones* (5) *del* § 8.5.

□ Comencemos por examinar la necesidad de la condición 2°. Si la fórmula (1) es exacta para todos los polinomios de grado $2n-1$, es asimismo exacta para los polinomios de todo grado inferior, incluyendo el grado $n-1$; entonces esta fórmula es de interpolación y obtendremos la unica colección de los coeficientes A_i definida por la fórmula (5) del § 8.5.

Consideremos ahora el polinomio $Q_{2n-1} = \omega_{n-1}(x) P_{n-1}(x)$. Utilizando el hecho de que para el polinomio Q_{2n-1} la fórmula (1) es exacta, obtenemos la condición 1°:

$$\int_a^b \rho(x)\, \omega_{n-1}(x)\, P_{n-1}(x)\, dx = (b-a) \sum_{i=1}^{n} A_i\, (Q_{2n-1}(x_i) = 0;$$

la última relación se deduce del hecho de que $\omega_{n-1}(x_i) = 0$ para todos los $i = 1, 2, \ldots, n$.

Vamos a comprobar la suficiencia de las condiciones 1° y 2°. Representemos un polinomio arbitrario $Q_{2n-1}(x)$ de grado $2n-1$

en la forma

$$Q_{2n-1} = \Psi_n P_{n-1}(x) + S_{n-1}(x),$$

donde P_{n-1} y S_{n-1} son, respectivamente, el cociente y el resto de la división del polinomio Q_{2n-1} por el polinomio Ψ_n, con la particularidad de que P_{n-1} es el polinomio de grado $n-1$, y S_{n-1} es el polinomio de grado no superior a $n-1$. Luego tenemos

$$\int_a^b \rho(x) Q_{2n-1}\,dx = \int_a^b \rho(x)\Psi_n P_{n-1}\,dx + \int_a^b \rho(x) S_{n-1}\,dx.$$

El primer sumando del segundo miembro es igual a cero en virtud de la ortogonalidad de Ψ_n y P_{n-1} (condición 1°) y para el segundo sumando la fórmula (1) es exacta en virtud de la condición 2°. Por eso, teniendo en cuenta que $Q_{2n-1}(x_i) = S_{n-1}(x_i)$, ya que $\Psi_n(x_i) = 0$, finalmente tenemos

$$\int_a^b \rho(x) Q_{2n-1}\,dx - (b-a)\sum_{i=1}^{n} A_i Q_{2n-1}(x_i). \ \blacksquare$$

Pongamos ahora la pregunta: ¿se puede construir una fórmula de integración numérica (1) que sea exacta para todos los polinomios de grado superior a $2n-1$? El siguiente teorema da la respuesta a esta pregunta.

Teorema 3. *Sea* $\rho(x) > 0$ *casi por doquier sobre* $[a, b]$. *Entonces no existe una fórmla de integración numérica* (1) *que sea exacta para todos los polinomios de grado* $2n$.

$\blacksquare$ Consideremos el polinomio $Q_{2n}(x) = (x-x_1)^2 \ldots (x-x_n)^2 = \omega_{n-1}^2(x)$. Entonces el primer miembro de la igualdad (1) es $\int_a^b \rho(x)\omega_{n-1}^2(x)\,dx > 0$, mientras que su segundo miembro $\sum_{i=1}^{n} A_i \omega_{n-1}^2(x_i) = 0$ lo que se demuestra precisamente por el teorema $\blacksquare$.

Así pues, la fórmula de integración numérica (1) con los nodos x_i, definidos por las relaciones (5), y con los coeficientes A_i, definidos por las relaciones (5) del § 8.5, es exacta para todo polinomio de grado no superior a $2n-1$. Ella se llama *fórmula de integración numérica de grado algebraico superior de precisión* (o bien *fórmula de integración numérica de Gauss*).

La fórmula de Gauss posee una propiedad útil desde el punto de vista del error de cálculo: todos los coeficientes A_i $(i = 1, 2, \ldots, n)$ para todo n son positivos.

Para demostrar esta afirmación consideremos la función $\left[\frac{\Psi_n(x)}{x-x_p}\right]^2$ que es un polinomio de grado $2n-2$ y se anula en todos los nodos $x_i \neq x_p$. Para esta función la fórmula de integración numérica de Gauss es exacta y por eso

$$\int_a^b \rho(x) \frac{\Psi_n^2(x)}{(x-x_p)^2} \, dx = (b-a) A_p [\Psi'(x_p)]^2.$$

De aquí se deduce inmediatamente que $A_p > 0$. Es evidente que en virtud de la arbitrariedad de ρ todos los coeficientes de integración numérica $A_i > 0$.

Pasemos ahora a la estimación del error de la fórmula de integración numérica de Gauss.

Teorema 4. *Sea:* 1°) $\rho(x) > 0$ *casi por doquier sobre* $[a, b]$; 2°) $f(x) \in C^{2n}[a, b]$; 3°) x_i y A_i $(i = 1, 2, \ldots, n)$ *son nodos y coeficientes de la fórmula de integración numérica de Gauss, respectivamente. Entonces existe un tal punto* $\xi \in (a, b)$ *que*

$$R_n[f] = I - I_n = \frac{f^{(2n)}(\xi)}{(2n)!} \int_a^b \rho(x) \omega_{n-1}^2(x) \, dx. \tag{6}$$

□ Representemos la función a integrar en la forma

$$f(x) = H_{2n-1}(x) + \frac{f^{(2n)}(\bar{\eta})}{(2n)!} \omega_{n-1}^2(x); \quad \bar{\eta} \in (a, b),$$

donde $H_{2n-1}(x)$ es el polinomio interpolador de Hermite cuyo grado no supera a $2n-1$ con los nodos dobles x_i $(i = 1, 2, \ldots, n)$. Es obvio que para $H_{2n-1}(x)$ la fórmula de Gauss es exacta. Por eso

$$I = I_n + \frac{1}{(2n)!} \int_a^b \rho(x) f^{(2n)}(\bar{\eta}) \omega_{n-1}^2(x) \, dx.$$

Aplicando al segundo sumando del segundo miembro el segundo teorema del valor medio de la función y pasando I_n al primer miembro, obtenemos la relación requerida (6). ■

Prácticamente la cuadratura de Gauss se obtiene del modo siguiente. Primero se encuentran las raíces del polinomio $\Psi_n(x)$ ortogonal a todos los polinomios de grado inferior a n. Luego para los coeficientes A_i se construye un sistema de ecuaciones lineales, suponiendo que la fórmula de integración numérica es justa para las funciones $1, x, \ldots x^{n-1}$:

$$m_k \equiv \int_a^b \rho(x) x^k \, dx = \sum_{i=1}^{n} A_i x_i^k.$$

Resolviendo este sistema, se determina A_i.

De ilustración elemental de la cuadratura de Gauss sirve la fórmula (22) construida en el § 8.4.

§ 8.7. Fórmulas compuestas de integración numérica

Las fórmulas con un gran número de nodos equidistantes se emplean, relativamente, muy rara vez. Esto se explica por diferentes causas. En primer lugar, para las funciones que tienen la singularidad incluso no en el mismo segmento de integración sino en la proximidad de este último, el término residual de la fórmula de integración numérica de alto orden es, por lo general, grande. Además, hasta para ciertas funciones analíticas, al situarse uniformemente los puntos nodales, es posible que $R_n[f] = I - I_n \to \infty$ cuando $n \to \infty$. En segundo lugar, para $n \geqslant 10$ entre los coeficientes H_i los hay negativos, lo que aumenta considerablemente el error de cálculo esperado. En efecto, supongamos que todos los valores de la función f_i de la fórmula (8) dada en el § 8.5 se conocen con la misma exactitud ε. Entonces el error de cálculo total se puede estimar por la magnitud $\Delta_2 = (b - a)\,\varepsilon \sum_{i=1}^{n} |H_i|$. Puesto que la suma de todos los coeficientes de Cotes es igual a la unidad, el hecho de que entre ellos existen coeficientes negativos origina el aumento de Δ_2. De la tabla 8.2 se ve cuán rápido es el crecimiento de Δ_2.

Tabla 8.2

n	10	15	20
$\sum_{i=1}^{n} \|H_i\|$	≈ 3	≈ 8	≈ 560

Las fórmulas de tipo gaussiano poseen ciertas ventajas en comparación con las de Newton-Cotes, puesto que carecen de los inconvenientes anterormente descritos. Sin embargo, las fórmulas de Gauss, siendo fórmulas de grado algebraico de precisión superior, para grandes n tienen un término residual proporcional a la derivada de la función integrable de orden superior. Esto es una desventaja esencial al integrar funciones que no posean derivadas continuas de órdenes superiores o bien al integrar dependencias funcionales construidas empíricamente. Además, incluso para los procesos de cuadratura convergentes, construidos con ayuda de las fórmulas de Gauss, de antemano se desconoce generalmente con cuán grande n la convergencia empieza a realizarse prácticamente, o sea, de antemano se desconoce cuál debe ser n para garantizar la exactitud deseada.

Las consideraciones recién expuestas determinan precisamente el hecho de que en la práctica suele darse preferencia a las así llamadas *fórmulas compuestas* (*o generalizadas*). El sentido de estas fórmulas consiste en que el segmento de integración $[a, b]$ se divide en varias partes, se aplica una u otra fórmula de integración numérica con pequeño n a cada parte separada y luego se suman los resultados.

Lo racional de tal enfoque se funda en los razonamientos siguientes. Para muchas fórmulas de integración numérica (incluyendo todas las que se consideran en este capítulo) el término residual es proporcional a cierto grado de longitud del segmento de integración. Supongamos, por ejemplo, que la aplicación de la fórmula de integración numérica elegida al segmento $[a, b]$ da para el término residual la expresión siguiente:

$$R[f] = (b - a)^k \varphi(a, b), \tag{1}$$

donde $\varphi(a, b)$ es la función lentamente variable del segmento de integración.

Dividiendo el segmento inicial en m partes iguales y aplicando a cada una de ellas la misma fórmula de integración numérica, obtenemos que en cada parte separada el término residual es, aproximadamente, m^k veces menor que (1). Sumando luego los resultados de integración y los términos residuales, obtenemos que el error de la integral inicial es, aproximadamente, m^{k-1} veces menor que al aplicar la fórmula de integración numérica elegida a todo el segmento inicial.

Este método es bastante general y puede ser realizado para toda fórmula de integración numérica.

Como ejemplo elemental de la fórmula compuesta de integración numérica puede servir la de los rectángulos (4) con el término residual (8), examinada en el § 8.5.

En el presente párrafo vamos a considerar las fórmulas compuestas más frecuentes construidas sobre la base de las cuadraturas elementales antes citadas.

Fórmula compuesta de los trapecios. Supongamos que se necesita calcular la integral de la función f sobre el segmento $[a, b]$. Dividamos el segmento $[a, b]$ en m partes iguales con los puntos de frontera $a = x_0, x_1, \ldots, x_m = b$ de un modo tal que la longitud de cada parte sea igual a $h = (b - a)/m$. Representemos la integral en forma de la suma:

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_1} f(x)\,dx + \int_{x_1}^{x_2} f(x)\,dx + \ldots + \int_{x_{m-1}}^{x_m} f(x)\,dx.$$

Apliquemos a cada sumando del segundo miembro de esta igualdad la fórmula de los trapecios con el término residual [véanse las relaciones (11) y (12) del § 8.5]. Reduciendo los términos semejantes, obtenemos

$$\int_a^b f(x)\,dx = \frac{b-a}{m}\left(\frac{f_0+f_m}{2} + \sum_{i=1}^{m-1} f_i\right) -$$

$$-\frac{(b-a)^3}{12m^3}\sum_{i=1}^{m} f''(\eta_i); \quad \eta_i \in (x_{i-1}, x_i). \tag{2}$$

Suponiendo que la derivada segunda de la función integrable es continua sobre todo el segmento $[a, b]$, en virtud del teorema de Weierstrass tenemos

$$\sum_{i=1}^{m} f''(\eta_i) = mf''(\eta); \quad \eta \in (a, b).$$

Sustituyendo esta expresión para la suma en la relación (2), obtenemos la *fórmula compuesta de los trapecios*

$$I \equiv \int_a^b f(x)\,dx \approx \frac{b-a}{m}\left(\frac{f_0+f_m}{2} + \sum_{i=1}^{m-1} f_i\right) \equiv I_2^m \tag{3}$$

con el término residual

$$R_2^m[f] = -\frac{(b-a)^3}{12m^2} f''(\eta). \tag{4}$$

Por consiguiente, la estimación del error de la cuadratura (3) se puede representar en la forma

$$\Delta_1 = |I - I_2^m| \leqslant \frac{(b-a)^3}{12m^2} M_2, \tag{5}$$

donde $M_2 = \max\limits_{[a,\, b]} |f''(x)|$. La estimación obtenida es inmejorable, puesto que se alcanza, por ejemplo, sobre la parábola $f = x^2$, lo que es fácil mostrar mediante una verificación inmediata. Nótese que los cálculos análogos han sido realizados al analizar la fórmula de cuadratura (4) del § 8.4. Mostremos cómo se hace esto para la fórmula compuesta (3).

Calculemos el miembro izquierdo de la fórmula de integración numérica (3) para $f = x^2$:

$$I = \int_a^b x^2\,dx = \frac{b^3-a^3}{3}.$$

Calculemos la suma

$$I_2^m = \frac{b-a}{m}\left[\frac{a^2+b^2}{2} + \sum_{i=1}^{m-1}\left(a + i\,\frac{b-a}{m}\right)^2\right].$$

Transformando la expresión puesta entre paréntesis y utilizando en este caso las siguientes fórmulas de sumación:

$$\sum_{i=1}^{m-1} i = \frac{(m-1)\,m}{2}; \quad \sum_{i=1}^{m-1} i^2 = \frac{(m-1)\,m\,(2m-1)}{6},$$

obtenemos la magnitud buscada

$$I_2^m = \frac{b^3-a^3}{3} + \frac{(b-a)^3}{6m^2}.$$

Así pues, el error de la fórmula de integración numérica (3) para la función $f = x^2$ es igual a

$$\Delta_1 = | I - I_2^m | = \frac{(b-a)^3}{6m^2}$$

lo que coincide exactamente con el segundo miembro de la estimación (5), puesto que $M_2 = 2$.

En virtud de la linealidad de las operaciones $I[f]$ e $I_2^m[f]$ y en virtud de que la fórmula (3) es exacta para toda función lineal, se puede afirmar que la estimación (5) se alcanza sobre una parábola arbitraria de segundo grado.

La expresión (5) para el error del método Δ_1 de la fórmula de integración numérica (3) muestra que $\lim\limits_{m\to\infty} \Delta_1 = 0$, o sea, aumentando m, podemos garantizar toda exactitud prefijada de antemano I_2^m desde el punto de vista del error del método (al faltar un error de cálculo). En tales casos se dice que el método es *convergente*. Nos queda aclarar la influencia que el error de los datos iniciales (de los valores de la función f_i) ejerce en el error del resultado. En virtud de la linealidad del segundo miembro de la fórmula de integración numérica (3) respecto a los valores de la función integrable, el error de cálculo total es proporcional al del cálculo de cada valor de la función (en caso de la igualdad de estos últimos). Por eso se puede esperar que en caso de una exactitud suficientemente alta de los datos iniciales el error de cálculo total será bastante pequeño.

Vamos a deducir las fórmulas concretas para el número de particiones m y para el error admisible de cada valor de la función f_i, que garantizan la exactitud requerida ε al utilizar la fórmula de integración numérica (3) con el error (5).

Ante todo, utilizando el algoritmo de resolución del problema II, citado en el § 8.3, representemos ε en forma de la suma: $\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3$ (si, por ejemplo, $\varepsilon = 10^{-k}$ entonces se supone de ordinario $\varepsilon_1 = 0{,}3 \cdot 10^{-k}$, $\varepsilon_2 = 0{,}2 \cdot 10^{-k}$, $\varepsilon_3 = 0{,}5 \cdot 10^{-k}$).

Escojamos luego el número de particiones m de un modo tal que se asegure el cumplimiento de la desigualdad $\Delta_1 \leqslant \varepsilon_1$. Para esto, en virtud de (5), es suficiente exigir que

$$\frac{(b-a)^3}{12m^2} M_2 \leqslant \varepsilon_1.$$

Transformando esta desigualdad, obtenemos para el número de particiones la fórmula

$$m \geqslant (b-a) \sqrt{\frac{(b-a) M_2}{12\varepsilon_1}}. \qquad (6)$$

Así pues, el error del método Δ_1 no superará ε_1 si el número de particiones m satisface la desigualdad (6).

Determinemos ahora cómo debe ser el error de los valores de la función integrable para que el error de cálculo total de I_2^m por la fórmula (3) no supere ε_2. Supongamos que el error buscado es $\Delta[f_i]$;

entonces, utilizando la fórmula (3), tenemos

$$\Delta [I_2^m] \leqslant \frac{b-a}{m} (\Delta [f_i] + (m-1) \Delta [f_i]) = (b-a) \Delta [f_i].$$

Por lo tanto, para cumplir la desigualdad $\Delta [I_2^m] \leqslant \varepsilon_2$ basta exigir que

$$\Delta [f_i] \leqslant \varepsilon_2/(b-a). \tag{7}$$

Ejemplo 1. Utilizando la fórmula compuesta de los trapecios, calcular la integral $\int_0^1 \frac{dx}{1+x}$ con exactitud hasta 0,01.

$\triangle$ Sea $\varepsilon_1 = 0{,}004$; $\varepsilon_2 = 0{,}001$; $\varepsilon_3 = 0{,}005$. Encontramos $M_2 = \max\limits_{[0,\,1]} \left| \left(\frac{1}{1+x} \right)'' \right| = 2$. Con ayuda de la desigualdad (6) determinemos el número necesario de particiones:

$$m \geqslant (1-0) \cdot \sqrt{\frac{(1-0) \cdot 2}{12 \cdot 0{,}004}} = 6{,}4.$$

Tomamos $m = 7$. Para el error admisible de los valores de la función subintegral la desigualdad (7) da el valor $\Delta [f_i] = 0{,}001$. Hacemos la tabla de los valores necesarios de la función integrable con tres cifras justas:

x	0	1/7	2/7	3/7	4/7	5/7	6/7	1
$(1+x)^{-1}$	1	0,875	0,778	0,700	0,636	0,583	0,538	0,500

Con ayuda de la fórmula (3) hallamos

$$I_2^7 = \frac{1-0}{7} \left(\frac{1+0{,}5}{2} + 0{,}875 + 0{,}778 + 0{,}700 + 0{,}636 + \right.$$

$$\left. + 0{,}583 + 0{,}538 \right) = 0{,}694.$$

Redondeando el resultado obtenido, finalmente tenemos

$$\int_0^1 \frac{dx}{1+x} = 0{,}69 \pm 0{,}01.$$

Compárese la solución obtenida con la del ejemplo 6 dado en el § 8.4 y con la del ejemplo 3 dado en el § 8.5. $\blacktriangle$

Fórmula compuesta de Simpson. En este caso dividamos el segmento de integración en un número par $2m$ de partes iguales con los puntos de frontera $a = x_0, x_1, \ldots, x_{2m} = b$, así que la longitud de cada parte vale $h = (b-a)/(2m)$. Representemos la integral en forma de la suma:

$$\int_a^b f(x)\,dx = \int_{x_0}^{x_2} f(x)\,dx + \int_{x_2}^{x_4} f(x)\,dx + \ldots + \int_{x_{2m-2}}^{x_{2m}} f(x)\,dx.$$

Apliquemos la fórmula de Simpson con el término residual [véase las relaciones (11) y (12) del § 8.5] a cada sumando del segundo miembro de la igualdad dada. Reduciendo los términos semejantes, obtenemos

$$\int_a^b f(x)\,dx = \frac{b-a}{6m}(f_0 + f_{2m} + 4\sigma_1 + 2\sigma_2) - \frac{(b-a)^5}{2880m^5}\sum_{i=1}^{m} f^{\mathrm{IV}}(\eta_i); \tag{8}$$

$$\eta_i \in (x_{2i-2},\ x_{2i});$$

$$\sigma_1 = f_1 + f_3 + \ldots + f_{2m-1}; \quad \sigma_2 = f_2 + f_4 + \ldots + f_{2m-2}.$$

Suponiendo que la cuarta derivada de la función integrable es continua sobre todo el segmento $[a, b]$, en virtud del teorema de Weierstrass tenemos

$$\sum_{i=1}^{m} f^{\mathrm{IV}}(\eta_i) = mf^{\mathrm{IV}}(\eta); \quad \eta \in (a, b).$$

Sustituyendo la expresión obtenida para la suma de las derivadas en la relación (8), llegamos a la *fórmula compuesta de Simpson*

$$I \equiv \int_a^b f(x)\,dx \approx \frac{b-a}{6m}(f_0 + f_{2m} + 4\sigma_1 + 2\sigma_2) \equiv I_3^m; \tag{9}$$

$$\sigma_1 = f_1 + f_3 + \ldots f_{2m-1}; \quad \sigma_2 = f_2 + f_4 + \ldots + f_{2m-2}$$

con el término residual

$$R_3^m[f] = -\frac{(b-a)^5}{2880m^4} f^{\mathrm{IV}}(\eta); \quad \eta \in (a, b). \tag{10}$$

Por consiguiente, la estimación del error de la fórmula de integración numérica (9) puede ser representada en la forma

$$\Delta_1 = |I - I_3^m| \leqslant \frac{(b-a)^5}{2880m^4} M_4, \tag{11}$$

donde $M_4 = \max\limits_{[a,\, b]} |f^{\mathrm{IV}}(x)|$. La estimación obtenida es inmejorable, puesto que se alcanza, por ejemplo, en la función $f = x^4$.

Al igual que para la fórmula de los trapecios, obtenemos las fórmulas concretas para el número de particiones m y para el error admisible de cada valor f_i, que aseguran la exactitud deseada ε al utilizar la fórmula de integración numérica (9) con el error (11).

Sea $\varepsilon = \varepsilon_1 + \varepsilon_2 + \varepsilon_3$. Teniendo en cuenta la relación (11), para el cumplimiento de la desigualdad $\Delta_1 \leqslant \varepsilon_1$ exijamos que

$$\frac{(b-a)^5}{2880m^4} M_4 \leqslant \varepsilon_1.$$

De aquí hallamos la condición para el número de particiones $2m$ con el cual el error del método no superará ε_1:

$$2m \geqslant (b-a)\sqrt[4]{\frac{(b-a)M_4}{180\varepsilon_1}}. \tag{12}$$

Determinemos ahora cómo debe ser el error de los valores de la función subintegral para que el error total de cálculo de I_3^m por la fórmula (9) no supere ε_2. Supongamos que el error buscado es $\Delta[f_i]$; entonces, utilizando la fórmula (9), tenemos

$$\Delta[I_3^m] \leqslant \frac{b-a}{6m}(2\Delta[f_i]+4m\Delta[f_i]+2(m-1)\Delta[f_i]=(b-a)\Delta[f_i].$$

Por lo tanto, para el cumplimiento de la desigualdad $\Delta[I_3^m] \leqslant$ $\leqslant \varepsilon_2$ basta exigir que

$$\Delta[f_i] \leqslant \varepsilon_2/(b-a). \tag{13}$$

Como era de esperar, hemos obtenido el mismo resultado que para la fórmula de los trapecios. Esto es consecuencia de que (3) y (9) son fórmulas del mismo tipo (3) del § 8.3, con la particularidad de que todos los coeficientes A_i son positivos y su suma es igual a la unidad.

Ejemplo 2. Utilizando la fórmula compuesta de Simpson, calcular la integral $\int_0^1 \frac{dx}{1+x}$ con exactitud de hasta 0,001.

△ Sea $\varepsilon_1 = 0{,}00045$; $\varepsilon_2 = 0{,}00005$, $\varepsilon_3 = 0{,}0005$. Encontramos $M_4 = \max\limits_{[0,\ 1]} \left|\left(\frac{1}{1+x}\right)^{\mathrm{IV}}\right| = 24$. Con ayuda de las desigualdades (12) determinemos el número necesario de particiones:

$$2m \geqslant (1-0)\cdot\sqrt[4]{\frac{(1-0)\cdot 24}{180\cdot 0{,}00045}} = 4{,}1\ \ldots$$

Puesto que $2m$ es un número par, tomamos $2m = 6$. Para el error admisible de los valores de la función subintegral la desigualdad (13) da el valor $\Delta[f_i] = 0{,}00005$. Hacemos la tabla de los valores necesarios de la función integrable con cuatro cifras justas:

x	0	1/6	2/6	3/6	4/6	5/6	1
$(1+x)^{-1}$	1	0,8571	0,75	0,6667	0,6	0,5455	0,5

Por la fórmula (9) encontramos

$$I_3^3 = \frac{1-0}{18}[1+0{,}5+4\cdot(0{,}8571+0{,}6667+0{,}5455)+$$
$$+2\cdot(0{,}75+0{,}6)] = 0{,}69317\ \ldots$$

Redondeando el resultado obtenido, finalmente tenemos

$$\int_0^1 \frac{dx}{1+x} = 0{,}693 \pm 0{,}001. \ \blacktriangle$$

Analicemos el resultado obtenido en el ejemplo 2. Comparando el valor aproximado de $I_3^3 = 0{,}69317\ldots$ con la magnitud exacta de la integral $I = 0{,}69314\ldots$, notamos que la diferencia real entre los valores exacto y aproximado (cerca de 0,00003) es alrededor de 17 veces menor que el error admisible dado ($\varepsilon_1 + \varepsilon_2 = 0{,}0005$). Es natural poner la pregunta ¿en qué consisten las causas de tal diferencia entre el error teóricamente predicho y el obtenido prácticamente?

Respondiendo a la pregunta puesta, nótese, ante todo, que tal divergencia entre el error verdadero y la estimación requerida se observa no siempre, ni mucho menos, y en los casos semejantes al examinado en el ejemplo 2 puede ser predicha de antemano y, más aún, ser reducida al mínimo al utilizar ciertas consideraciones adicionales y al construir más racionalmente el proceso de cálculo.

En efecto, en el ejemplo 2 debido a la desigualdad (12) el error del método, igual a 0,0004, corresponde al número de particiones $2m \geqslant 4{,}1\ldots$. El valor límite $4{,}1\ldots$ no podía ser elegido en virtud del hecho de que $2m$ debe ser entero y par, y al valor mínimo admisible $2m = 6$, que hemos elegido, corresponde, desde luego, también un error menor del método que es igual, aproximadamente, a 0,0001. Esta es precisamente la primera causa de disminución del error práctico.

Luego, por ser inmejorables todas las estimaciones consideradas del error, ellas se obtenían con los polinomios en los cuales la derivada, que como multiplicador forma parte de la expresión para el término residual, era una magnitud constante. En cambio, si la derivada mencionada cambia considerablemente sobre el segmento de integración (lo que es natural cuando este segmento es grande), la estimación respectiva resulta, por lo general, distar mucho de la óptima. En este caso es conveniente determinar el número de particiones partiendo de la forma del término residual, representada en las relaciones (2) y (8). Por ejemplo, para la fórmula de Simpson se obtendría la siguiente desigualdad destinada a determinar $2m$:

$$\frac{(b-a)^5}{180\,(2m)^4}\cdot\frac{1}{m}\cdot\sum_{i=1}^{m} M_4^i \leqslant \varepsilon_1, \tag{14}$$

donde $M_4^i = \max\limits_{[x_{2i-2}x_{2i}]} |f^{\mathrm{IV}}(x)|$. Es evidente que a grandes cambios de la cuarta derivada sobre el segmento de integración, la desigualdad (14) es más ventajosa que la (12) desde el punto de vista del número de particiones en $\sqrt[4]{\dfrac{M_4}{(1/m)\sum\limits_{i=1}^{m} M_4^i}}$ veces. En realidad, debajo del signo de la raíz está la relación entre el valor absoluto máximo de la cuarta derivada sobre todo el segmento de integración y el valor medio calculado respecto de m intervalos $[x_{2i-2}, x_{2i}]$; $i = 1, 2, \ldots, m$. Así, la utilización de la desigualdad (14) en el ejemplo 2 para el número de particiones da el valor $2m = 4$ y para $2m = 6$ la estimación del error del método disminuye hasta 0,000045, que es sólo una vez y media más que el error verdadero. Esto es ya un buen resultado.

Ejercicios

1. Calcular la integral definida $\int_2^8 \sqrt{x+2}\,dx$, utilizando la fórmula de los rectángulos izquierdos para $n=6$.

2. Haciendo uso de la fórmula de los rectángulos derechos para $n=8$, calcular $\int_1^9 \frac{dx}{1+x}$.

3. Haciendo uso de la fórmula de los trapecios para $n=8$, calcular $\int_0^8 \frac{dx}{1+x}$.

4. Con ayuda de la fórmula de los trapecios calcular $\int_0^5 \frac{dx}{\sqrt{x+4}}$, suponiendo $n=5$.

5. Con ayuda de la fórmula de Simpson calcular $\int_0^1 \frac{dx}{x^2+9}$, suponiendo $2m=10$.

6. Calcular $\int_0^1 \frac{\cos x}{1+x}\,dx$ con ayuda de la fórmula de Simpson; poner $2m=10$.

7. Calcular la integral $\int_{-1}^1 \frac{dx}{x+3}$, haciendo uso de la fórmula de Gauss para $2m=6$.

8. La función se da en forma de una tabla

x	0,525	0,526	0,527	0,528
f	0,50121	0,50208	0,50294	0,50381

Haciendo uso del método de integración numérica, hallar la derivada primera en el punto $x^* = 0,525$.

9. Encontrar la derivada de primer orden en el punto $x^* = 50$ para la función dada en la forma tabular

x	50	55	60	65
f	1,6990	1,7404	1,7782	1,8129

10. Calcular la integral $\int_0^1 f(x)\,dx$ con exactitud de hasta ε.

a) $f(x) = x^2 \cos x$, $\varepsilon = 0{,}001$; b) $f(x) = x^2 \operatorname{sen} x$, $\varepsilon = 0{,}001$; c) $f(x) = xe^x$, $\varepsilon = 0{,}0001$; d) $f(x) = x\sqrt{x}e^x$, $\varepsilon = 0{,}0001$; e) $f(x) = e^{x^2}$, $\varepsilon = 0{,}0001$; f) $f(x) = e^{x\sqrt{x}}$, $\varepsilon = 0{,}0001$; g) $f(x) = x^2 + 1$, $\varepsilon = 0{,}0001$; h) $f(x) = \operatorname{sen} x$, $\varepsilon = 0{,}001$; i) $f(x) = e^x$, $\varepsilon = 0{,}001$.

11. Calcular los valores de la derivada de la función $f(x)$ en el punto $x = x_1$, utilizando las tablas matemáticas de cuatro cifras decimales de las funciones trigonométricas con el paso igual a 1°. Determinar el valor absoluto del resultado:

a) $f(x) = \operatorname{sen} x$; $x_1 = 41°$; b) $f(x) = \cos x$, $x_1 = 27{,}5°$; c) $f(x) = \operatorname{tg} x$, $x_1 = 50°$; d) $f(x) = \operatorname{sen} x$, $x_1 = 17°30'$; e) $f(x) = \cos x$, $x_1 = 63°$, e) $f(x) = \operatorname{tg} x$; $x_1 = 33{,}5°$.

CAPÍTULO IX

Resolución aproximada de las ecuaciones diferenciales ordinarias

§ 9.1. Concepto de ecuación diferencial

La ecuación en la cual la función incógnita entra bajo el signo de derivada o de diferencial se llama *ecuación diferencial*.

Por ejemplo,

$$\frac{dy}{dx} = 2(y-3); \quad \frac{d^2y}{dt^2} = t+1; \quad \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0;$$

$$y' = x^2; \quad x\,dy = y^3\,dx.$$

Si la función incógnita que forma parte de la ecuación diferencial depende sólo de una variable independiente, la ecuación diferencial se llama *ordinaria*.

Tales son, por ejemplo, las ecuaciones diferenciales

$$x^2 \cdot \frac{d^2y}{dx^2} = 2; \quad 2s\,dt = t\,ds.$$

En cambio, si la función incógnita que forma parte de la ecuación diferencial es función de dos o más variables independientes, la ecuación diferencial se denomina *ecuación en derivadas parciales*.

Por ejemplo, la ecuación diferencial

$$\frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 0$$

es ecuación en derivadas parciales.

Se llama *orden* de una ecuación diferencial al orden superior de la derivada (o de la diferencial) que forma parte de la ecuación.

Así, por ejemplo, las ecuaciones

$$\frac{d^2s}{dt^2} = t-1; \quad \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = 1$$

son ecuaciones de segundo orden y las ecuaciones

$$\frac{ds}{dt} \cdot \cos t + \operatorname{sen} t = 1; \quad (x^2 - y^2)\,dx + (x^2 + y^2)\,dy = 0,$$

ecuaciones de primer orden.

En este capítulo vamos a considerar sólo las ecuaciones diferenciales ordinarias.

En el caso más general una ecuación diferencial ordinaria de n-ésimo orden contiene una variable independiente, una función incógnita y sus derivadas o diferenciales hasta el n-ésimo orden, inclusivamente, y tiene la forma

$$F(x, y, y', y'', \ldots, y^{(n)}) = 0. \tag{1}$$

En esta ecuación x es la variable independiente; y, la función incógnita; $y', y'', \ldots, y^{(n)}$ son las derivadas de esta función.

Si el primer miembro de la ecuación diferencial (1) es el polinomio respecto a la derivada de la función incógnita, el grado de este polinomio se llama *grado* de la ecuación diferencial.

Por ejemplo, la ecuación

$$(y'')^4 + (y')^2 - y^6 + x^7 = 0$$

es la ecuación de cuarto grado y de segundo orden, mientras que la ecuación

$$(y')^2 + x^4y^5 - y^8 + x^{10} = 0$$

es la ecuación de segundo grado y de primer orden.

La ecuación diferencial de n-ésimo orden resuelta respecto a la derivada superior puede ser escrita en la forma

$$y^{(n)} = f(x, y, y', y'', \ldots, y^{(n-1)}). \tag{2}$$

Se llama *solución* (*o integral*) de la ecuación (2) a toda función derivable $y = \varphi(x)$ que satisfaga esta ecuación, o sea, a tal función que al ser sustituida en la ecuación (2), convierte a ésta en identidad.

El gráfico de solución de una ecuación diferencial se denomina *curva integral* de esta ecuación.

La solución de la ecuación diferencial que contiene una cantidad de constantes (parámetros) arbitrarias independientes igual a su orden se llama *solución general* (o *integral general*) de esta ecuación.

Geométricamente la solución general de la ecuación diferencial es una familia de curvas integrales de esta ecuación.

Se llama *solución particular* de una ecuación diferencial a toda solución que pueda ser obtenida de la general para ciertos valores numéricos de las constantes arbitrarias que forman parte de la solución general.

Las constantes arbitrarias que entran en la solución general se determinan de las llamadas condiciones iniciales.

El problema con condiciones iniciales se plantea así: hallar la solución $y = \varphi(x)$ de la ecuación $y^{(n)} = f(x, y, y', y'', \ldots, y^{(n-1)})$, que satisface las condiciones adicionales consistentes en que la solución $y = \varphi(x)$ debe tomar junto con sus derivadas hasta el orden $(n-1)$ los valores numéricos asignados $y_0, y_0', y_0'', \ldots, y_0^{(n-1)}$ para el valor numérico asignado $x = x_0$ de la variable independiente x:

$$y = y_0,\ y' = y_0',\ y'' = y_0'',\ \ldots,\ y^{(n-1)} = y_0^{(n-1)} \quad \text{para} \quad x = x_0. \tag{3}$$

Las condiciones (3) se denominan *condiciones iniciales*; los números $x_0, y_0, y_0', y_0'', \ldots, y_0^{(n-1)}$ se llaman *datos iniciales* de la solución y el problema de determinación de la solución $y = \varphi(x)$ de la ecuación diferencial (2), que satisface las condiciones iniciales (3), *problema con condiciones iniciales* (o *problema de Cauchy*).

En caso de una ecuación de primer orden, o sea, para $n = 1$ obtenemos el problema de Cauchy para la ecuación $y' = f(x, y)$ con la condición inicial $x = x_0$, $y = y_0$.

Geométricamente el problema de Cauchy (para la ecuación de primer orden) consiste en que de todo el conjunto de las curvas integrales que representan la solución general se separa la curva integral que pasa por el punto M_0 con las coordenadas $x = x_0$, $y = y_0$.

Ejemplo. Para la ecuación diferencial $\frac{dy}{dx} = 2x$ con condición inicial $y_0 = 2$ cuando $x_0 = 1$ la solución general tiene la forma $y = x^2 + c$. Es una familia de parábolas. Si ahora en la solución general sustituimos los datos iniciales, obtenemos $2 = 1 + C$, o sea, $C = 1$. Por consiguiente, la solución particular que satisface la condición inicial es $y = x^2 + 1$. Geométricamente esto quiere decir que de todo el conjunto de las parábolas que representan la solución general de la ecuación diferencial se elige una que pasa por el punto M_0 (1; 2) (fig. 9.1).

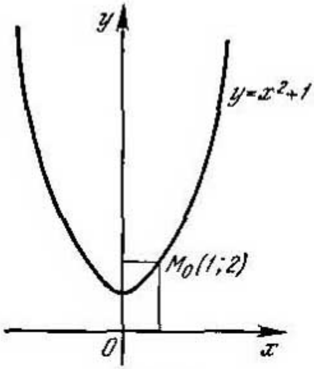


Fig. 9.1

El problema de Cauchy tiene una única solución que satisface la condición $y(x_0) = y_0$ si la función $f(x, y)$ es continua en cierto dominio $R_{[a,\, b]} = \{\, |x - x_0| < a,\ |y - y_0| < b\}$ y satisface en este dominio la *condición de Lipschitz*

$$|f(x, \bar{y}) - f(x, y)| \leqslant N\, |\bar{y} - y|,$$

donde N es la constante de Lipschitz que depende de a y b (a y b son las fronteras del dominio).

Los métodos de integración exacta de las ecuaciones diferenciales son útiles sólo para una parte relativamente pequeña de ecuaciones que se encuentran en la práctica.

Por eso adquieren gran importancia los métodos de resolución aproximada de las ecuaciones diferenciales, los cuales según la forma de representar la solución pueden ser divididos en dos grupos:

1) **métodos analíticos** que ofrecen una solución aproximada de la ecuación diferencial en forma de una expresión analítica;

2) **métodos numéricos** que ofrecen una solución aproximada en forma tabular.

En este capítulo para el primer grupo de métodos consideraremos el método de aproximaciones sucesivas (método de Picard) y el de integración de las ecuaciones diferenciales con ayuda de las series de potencias; para el segundo grupo examinaremos el método de Euler y sus modificaciones, los métodos de Runge — Kutta y de Adams.

§ 9.2. Método de aproximaciones sucesivas (método de Picard)

Este método apareció debido a la demostración del teorema de existencia y de unicidad de la solución de las ecuaciones diferenciales. Lleva el nombre de **método de Picard.**

Supongamos que se da la ecuación

$$y' = f(x, y), \tag{1}$$

cuyo segundo miembro en el rectángulo $\{ | x - x_0 | \leqslant a, \ | y - y_0 | \leqslant b\}$ es continuo y tiene una derivada parcial continua respecto a y. Es necesario hallar la solución de la ecuación (1), que satisfaga la condición inicial

$$x = x_0, \ y(x_0) = y_0. \tag{2}$$

Integrando ambos miembros de la ecuación entre x_0 y x, obtenemos

$$\int_{y_0}^{y} dy = \int_{x_0}^{x} f(x, y)\, dx$$

o bien

$$y(x) = y_0 + \int_{x_0}^{x} f(x, y)\, dx. \tag{3}$$

La ecuación (1) se reemplaza por la ecuación integral (3) en la cual la función incógnita y está bajo el signo integral. La ecuación integral (3) satisface la ecuación diferencial (1) y la condición inicial (2). En efecto,

$$y(x_0) = y_0 + \int_{x_0}^{x} f(x, y)\, dx = y_0.$$

Reemplazando en la igualdad (3) la función y por el valor y_0, obtenemos la primera aproximación

$$y_1(x) = y_0 + \int_{x_0}^{x} f(x, y_0)\, dx.$$

Luego en la ecuación (3) sustituimos y por el valor hallado y_1 y obtenemos la segunda aproximación

$$y_2(x) = y_0 + \int_{x_0}^{x} f(x, y_1)\, dx.$$

Continuando luego el proceso, encontramos sucesivamente

$$y_3(x)=y_0+\int_{x_0}^{x} f(x,\ y_2)\,dx.$$

. . . .

$$y_n(x)=y_0+\int_{x_0}^{x} f(x,\ y_{n-1})\,dx.$$

Ahora bien, formamos la sucesión de las funciones

$$y_1(x),\ y_2(x),\ y_3(x),\ \ldots,\ y_n(x).$$

Es válido el siguiente teorema que citaremos sin demostración.

Teorema. *Supongamos que en el entorno del punto* $(x_0;\ y_0)$ *la función* $f(x,\ y)$ *es continua y tiene una derivada parcial acotada* $f'_y(x,\ y.)$ *Entonces en cierto intervalo, que contiene el punto* x_0, *la sucesión* $\{y_i(x)\}$ *converge hacia la función* $y(x)$ *que sirve de solución de la ecuación diferencial* $y = f(x,\ y)$ *y que satisface la condición* $y(x_0) = y_0$.

La estimación del error del método de Picard se determina por la fórmula

$$|y-y_n| \leqslant N^n M\,\frac{h^{n+1}}{(n+1)!}\,, \tag{4}$$

donde $M = \text{máx}\ |f(x,\ y)|$ para $(x,\ y) \in R_{[a,\ b]}$ y N es la constante de Lipschitz para el dominio $R_{[a,\ b]}$, igual a $N = \text{máx}\ |f'_y(x,\ y)|$. La magnitud h para determinar el entorno $[x_0 - h \leqslant x \leqslant x_0 + h]$ se calcula con ayuda de la fórmula

$$h = \text{mín}\ (a,\ b/M); \tag{5}$$

a y b son las fronteras del dominio R.

Ejemplo. Haciendo uso del método de Picard, resolver la ecuación diferencial $y' = x^2 + y^2$ que satisface la condición inicial $x_0 = 0,\ y(x_0) = y_0 = 0$.

△ Pasemos a la ecuación integral

$$y=y_0+\int_{x_0}^{x}(x^2+y^2)\,dx$$

o bien, teniendo en cuenta las condiciones iniciales,

$$y=\int_{0}^{x}(x^2+y^2)\,dx.$$

Obtenemos las aproximaciones sucesivas:

$$y_1=\int_{0}^{x}(x^2+y_0^2)\,dx=\int_{0}^{x}(x^2+0)\,dx=\frac{x^3}{3}\,;$$

$$y_2=\int_0^x (x^2+y_1^2)\,dx=\int_0^x \left(x^2+\frac{x^6}{9}\right)dx=\frac{x^3}{3}+\frac{x^7}{63};$$

$$y_3=\int_0^x (x^2+y_2^2)\,dx=\int_0^x \left(x^2+\frac{x^6}{9}+\frac{2x^{10}}{189}+\frac{x^{14}}{3969}\right)dx=$$

$$=\frac{x^3}{3}+\frac{x^7}{63}+\frac{2x^{11}}{2079}+\frac{x^{15}}{59\,535}.$$

Vamos a estimar el error de la tercera aproximación con ayuda de la fórmula (4):

$$|y-y_n|\leqslant N^n M\,\frac{h^{n+1}}{(n+1)!}.$$

Puesto que la función $y'=x^2+y^2$ está definida y es continua en todo el plano, en calidad de a y b se pueden tomar todos números. Con el fin de precisar elijamos el rectángulo

$$R\{\ |x-x_0|\leqslant 0{,}5,\ \ |y-y_0|\leqslant 1\},$$

o sea, $R\{-0{,}5\leqslant x\leqslant 0{,}5,\ -1\leqslant y\leqslant 1\}$
Entonces

$$M=\text{máx}\,|f(x,y)|=\text{máx}\,(x^2+y^2)=1{,}25,$$
$$N=\text{máx}\,|f'_y(x,y)|=\text{máx}\,|2y|=2.$$

Puesto que $a=0{,}5$ y $b/M=0{,}8$, según la fórmula (5) tenemos

$$h=\text{mín}\,(a,b/M)=0{,}5.$$

La solución y se dará para $-0{,}5\leqslant x\leqslant 0{,}5$. Si $n=3$, encontramos

$$|y-y_3|\leqslant 1{,}25\cdot 2^3\cdot 0{,}5^4/4!=5/192.$$

La estimación obtenida del error es muy aproximada, de hecho el error es mucho más pequeño. ▲

§ 9.3. Integración de las ecuaciones diferenciales con ayuda de las series de potencias

Método de derivación sucesiva. Supongamos que se da la ecuación diferencial de n-ésimo orden:

$$y^{(n)}=f(x,y,y',\ldots,y^{(n-1)}) \tag{1}$$

con las condiciones iniciales

$$x=x_0,\ y(x_0)=y_0,\ y'(x_0)=y'_0,\ \ldots,\ y^{(n-1)}(x_0)=y_0^{(n-1)}. \tag{2}$$

El segundo miembro de esta ecuación es una función analítica en el punto inicial $M_0(x_0;\ y_0;\ y'_0,\ \ldots,\ y_0^{(n-1)})$. Representemos la solución $y=y(x)$ de la ecuación (1) en el entorno del punto x_0 como serie de

Taylor

$$y = y_0 + y_0'(x - x_0) + \frac{y_0''}{2!}(x - x_0)^2 + \ldots + \frac{y_0^{(n)}}{n!}(x - x_0)^n + \ldots, \tag{3}$$

donde $|x - x_0| < h$ y h es una magnitud bastante pequeña. Para encontrar los coeficientes de la serie (3) la ecuación (1) se deriva respecto a x un número necesario de veces, utilizando las condiciones (2).

En la práctica la magnitud $|x - x_0|$ se toma tan pequeña que en caso de un grado requerido de precisión el resto de la serie puede ser despreciado.

Si $x_0 = 0$, se obtiene la serie de Taylor respecto a las potencias de x

$$y = y_0 + y_0'x + \frac{y_0''}{2!}x^2 + \ldots + \frac{y_0^{(n)}}{n!}x^{(n)} + \ldots \tag{4}$$

Ejemplo 1. Halla la solución de la ecuación diferencial $y = y - 4x + 3$ que satisface la condición inicial $x_0 = 0$, $y_0 = 3$.

△ Sustituyendo en el desarrollo (4) $y_0 = 3$, obtenemos

$$y = 3 + \frac{y_0'}{1}x + \frac{y_0''}{2!}x^2 + \frac{y'''}{3!}x^3 + \ldots + \frac{y_0^{(n)}}{n!}x^{(n)} + \ldots \tag{*}$$

Derivando sucesivamente la ecuación dada, tenemos

$$y'' = y' - 4 = y - 4x - 1,\ y''' = y'' = y - 4x - 1,$$
$$y^{IV} = y^{III}, \text{ etc.}$$

Utilizando la condición inicial, encontramos

$$y_0' = y_0 - 4x_0 + 3 = 3 + 3 = 6;\ y'' = y_0 - 4x_0 - 1 = 3 - 1 = 2;$$
$$y_0''' = 2;\ y_0^{IV} = 2;\ \ldots;\ y_0^{(n)} = 2.$$

Sustituyamos y_0', y_0'', y_0''' en el segundo miembro de la relación (*); entonces obtenemos

$$y = 3 + 6x + x^2 + \frac{x^3}{3} + \frac{x^4}{12} + \cdots$$

La solución exacta de la ecuación dada es una función

$$y = 2e^x + 4x + 1.$$

Si se pone $h = 0{,}1$, se puede hacer la tabla de los valores de la solución de la ecuación diferencial dada. ▲

Ejemplo 2. Hallar la solución de la ecuación diferencial $y'' - x^2y = 0$, que satisface la condición inicial $x_0 = 0$, $y_0 = 1$, $y_0' = 0$.

△ Al igual que en el ejemplo 1, buscaremos la solución de la ecuación dada en forma de la serie (4) respecto a las potencias de x.

Tabla 9.1

x_i	0	0,1	0,2	0,3
Valores obtenidos de la resolución analítica	3	3,6021	4,2428	4,8996
Solución aproximada obtenida con ayuda de la serie de potencias	3	3,6103	4,2427	4,8999

Puesto que $y_0 = 1$, $y'_0 = 0$, la serie (4) tiene la forma

$$y = 1 + \frac{y''_0}{2!} x^2 + \frac{y'''_0}{3!} x^3 + \ldots + \frac{y_0^{(n)}}{n!} x^n + \ldots \qquad (*)$$

Escribamos la ecuación diferencial dada en la forma $y'' = x^2 y$. Derivando sucesivamente esta igualdad, tenemos

$y''' = 2xy + x^2 y'$,

$y^{IV} = 2y + 2xy' + 2xy' + x^2 y'' = 2y + 4xy' + x^2 y''$,

$y^{V} = 2y' + 4y' + 4xy'' + 2xy'' + x^2 y''' = 6y' + 6xy'' + x^2 y'''$,

$y^{VI} = 12y'' + 8xy''' + x^2 y^{IV}$,

$y^{VII} = 20y''' + 10xy^{IV} + x^2 y^{V}$,

$y^{VIII} = 30y^{IV} + 12xy^{V} + x^2 y^{VI}$.

Sustituyendo sucesivamente en cada una de las desigualdades obtenidas las condiciones iniciales $x_0 = 0$; $y_0 = 1$, $y'_0 = 0$, encontramos $y''(0) = 0$; $y'''(0) = 0$; $y^{IV}(0) = 2$; $y^{V} = y^{VI} = y^{VII} = 0$; $y^{VIII} = 30 \cdot 2 = 60$.

Por último, sustituyendo y'', y'''_0, y^{IV} ... en el segundo miembro de la relación (*), obtenemos la solución buscada

$$y = 1 + \frac{1}{12} x^4 + \frac{1}{672} x^8 + \ldots \quad \blacktriangle$$

El método de desarrollo en serie de la solución se puede utilizar también para resolver los sistemas de ecuaciones diferenciales ordinarias.

Ejemplo 3. Hallar la solución del sistema

$$\begin{cases} \dfrac{dx}{dt} = x \cos t - y \operatorname{sen} t, \\ \dfrac{dy}{dt} = x \operatorname{sen} t + y \cos t, \end{cases}$$

que satisface las condiciones iniciales $x(0) = 1$, $y(0) = 0$.

$\triangle$ Pongamos

$$x = x(0) + x'(0)\, t + x''(0) \frac{t^2}{2} + \frac{x'''(0)}{3!} t^3 + \ldots;$$

$$y = y(0) + y'(0)\, t + y''(0) \frac{t^2}{2} + \frac{y'''(0)}{3!} t^3 + \ldots$$

Derivando las ecuaciones del sistema, tenemos

$$x'' = x' \cos t - x \operatorname{sen} t - y' \operatorname{sen} t - y \cos t,$$
$$y'' = x' \operatorname{sen} t + x \cos t + y' \cos t - y \operatorname{sen} t, \text{ etc.}$$

Sustituyendo sucesivamente en cada una de las igualdades obtenidas las condiciones obtenidas, hallamos $x'(0) = \cos t\,|_{t=0} = 1$, $y'(0) = 0$, $x''(0) = 1$, $y'(0) = 1$, $x'''(0) = 0$, $y'''(0) = 3$, etc.

La solución buscada tiene la forma

$$x = 1 + t + \frac{t^2}{2} + \cdots, \quad y = \frac{t^2}{2} + \frac{t^3}{2} + \cdots \quad \blacktriangle$$

Método de coeficientes indeterminados. Supongamos que se da la ecuación diferencial

$$y' = f(x, y) \tag{5}$$

con la condición inicial $y(x_0) = y_0$. El método de coeficientes indeterminados consiste en que la solución de la ecuación (5) se busca en forma de una serie con coeficientes desconocidos

$$y = a_0 + a_1(x - x_0) + a_2(x - x_0)^2 + a_3(x - x_0)^3 + \ldots, \tag{6}$$

que se determinan con ayuda de la sustitución de la serie (6) en la ecuación (5), luego se igualan los coeficientes de x de iguales grados y se utiliza la condición inicial. Los valores hallados de los coeficientes $a_0, a_1, a_2, a_3, \ldots$ se sustituyen en la serie (6).

Ejemplo 4. Con ayuda del método de coeficientes indeterminados hallar la solución de la ecuación diferencial $y' = x^2 + y^2$, que satisface la condición inicial $x_0 = 0$, $y(x_0) = 1$.

$\triangle$ Puesto que $x_0 = 0$, la serie (6) tomará la forma

$$y = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4 + \ldots \tag{*}$$

Sustituyendo en la relación (*) $x = x_0$ e $y(x_0) = 1$, obtenemos $a_0 = 1$.

En adelante es cómodo desarrollar el segundo miembro de la ecuación $y' = x^2 + y^2$ respecto a las potencias de $(y - 1)$:

$$y' = x^2 + [(y - 1) + 1]^2 = x^2 + 1 + 2(y - 1) + (y - 1)^2. \tag{**}$$

Derivando la serie (*), tenemos

$$y' = a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \ldots \tag{***}$$

Vamos a sustituir ahora en la igualdad (**) la expresión para y' de la igualdad (***), para y de la igualdad (*) y $a_0 = 1$. Entonces obtenemos

$$a_1 + 2a_2 x + 3a_3 x^2 + 4a_4 x^3 + \ldots = 1 + x^2 + 2(a_1 x +$$
$$+ a_2 x^2 + a_3 x^3 + a_4 x^4 + \ldots) + (a_1 x + a_2 x^2 + a_3 x^2 + \ldots)^2.$$

Abramos paréntesis en el segundo miembro de la última igualdad, reduzcamos los términos semejantes e igualemos los coeficientes de

x de iguales potencias. Como resultado obtenemos

$$a_1 = 1,\ 2a_2 = 2a_1, 3a_3 = 1 + 2a_2 + a_1^2, 4a_4 = 2a_3 + 2a_1a_2,$$

de donde $a_1 = 1$, $a_2 = 1$, $a_3 = 4/3$, $a_4 = 7/6$.

Sustituyendo los valores hallados de los coeficientes en la serie (*), finalmente tenemos

$$y = 1 + x + x^2 + \frac{4}{3} x^3 + \frac{7}{6} x^4 + \ldots \quad \blacktriangle$$

Ejemplo 5. Hallar la solución de la ecuación diferencial $y'' - x^2y = 0$, que satisface las condiciones iniciales $x_0 = 0$, $y(x_0) = 1$, $y'(x_0) = 0$.

$\triangle$ Puesto que $x_0 = 0$, buscaremos la solución en forma de la serie

$$y = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots \tag{*}$$

Derivando dos veces la igualdad (*), tenemos

$$y' = a_1 + 2a_2x + 2a_3x^2 + 4a_4x^3 + \ldots + na_nx^{n-1}, \tag{**}$$

$$y'' = 2a_2 + 6a_3x + 12a_4x^2 + \ldots + a_nn(n-1)x^{n-2}. \tag{***}$$

Utilizando las condiciones iniciales, de las igualdades (*) y (**) obtenemos $a_0 = 1$, $a_1 = 0$.

Sustituyendo los valores hallados de los coeficientes en la serie (*), tenemos

$$y = 1 + a_2x^2 + a_3x^3 + a_4x^4 + \ldots + a_nx^n. \tag{****}$$

Con el fin de determinar los demás coeficientes de la serie (****), sustituyamos en la ecuación diferencial dada la expresión para y'' de la igualdad (***) y para y de la igualdad (****)

$$2a_2 + 6a_3x + 12a_4x^2 + 20a_5x^3 + 30a_6x^4 + \ldots + n(n-1) \times \\ \times a_nx^{n-2} - x^2(1 + a_2x^2 + a_3x^3 + a_4x^4 + \ldots + a_nx^n + \ldots) = \\ = 0$$

Agrupemos los términos de potencias iguales:

$$2a_2 + 6a_3x + (12a_4 - 1)x^2 + (20a_5)x^3 + (30a_6 - a_2)x^4 + \\ + (42a_7 - a_3)x^5 + \ldots = 0.$$

La igualdad obtenida es una identidad. Puesto que buscamos la solución para $x \neq 0$, nos queda poner iguales a cero todos los coeficientes de x de distintas potencias, o sea,

$a_2 = 0$, $a_3 = 0$, $12a_4 - 1 = 0$, de donde $a_4 = 1/12$,

$a_5 = 0$, $30a_6 - a_2 = 0$, de donde $a_6 = 0$, etc.

La solución buscada tiene la forma

$$y = 1 + \frac{1}{12} x^4 + \frac{1}{672} x^8 + \ldots \quad \blacktriangle$$

§ 9.4. Integración numérica de las ecuaciones diferenciales. Método de Euler

Resolver la ecuación diferencial $y' = f(x, y)$ por el método numérico, quiere decir que para la sucesión dada de los argumentos x_0, $x_1, \ldots, x_n$ y el número y_0, sin determinar la función $y = F(x)$, han de encontrarse tales valores $y_1, y_2, \ldots, y_n$ que $y_i = F(x_i)$ ($i = 1, 2, \ldots, n$) y $F(x_0) = y_0$.

Ahora bien, los métodos numéricos permiten, en vez de hallar la función $y = F(x)$, obtener una tabla de valores de esta función para la sucesión dada de argumentos. La magnitud $h = x_k - x_{k-1}$ se llama *paso de integración*.

Consideremos algunos de los métodos numéricos.

Método de Euler. Este método es comparativamente muy aproximado y se emplea, en lo fundamental, para los cálculos preliminares. Sin embargo, las ideas en que se basa el método de Euler son de partida para varios otros métodos.

Supongamos que se da la ecuación diferencial de primer orden

$$y' = f(x, y) \tag{1}$$

con la condición inicial

$$x = x_0, \; y(x_0) = y_0. \tag{2}$$

Se necesita hallar la solución de la ecuación (1) en el segmento $[a, b]$.

Dividamos el segmento $[a, b]$ en n partes iguales; en este caso obtenemos la sucesión $x_0, x_1, x_2, \ldots, x_n$, donde $x_i = x_0 + ih$ ($i = 0, 1, 2, \ldots, n$) y $h = (b - a)/n$ es el paso de integración.

Elijamos el k-ésimo tramo $[x_k, x_{k+1}]$ e integremos la ecuación (1):

$$\int_{x_k}^{x_{k+1}} f(x, y)\,\mathrm{d}x = \int_{x_k}^{x_{k+1}} y'\,\mathrm{d}x = y(x)\Big|_{x_k}^{x_{k+1}} = y(x_{k+1}) - y(x_k) = y_{k+1} - y_k,$$

o sea,

$$y_{k+1} = y_k + \int_{x_k}^{x_{k+1}} f(x, y)\,\mathrm{d}x. \tag{3}$$

Si en la última integral a función subintegral se toma constante sobre el tramo $[x_k, x_{k+1}]$ e igual a su valor inicial en el punto $x = x_k$, obtenemos

$$\int_{x_k}^{x_{k+1}} f(x, y)\,\mathrm{d}x = f(x_k, y_k)\cdot x\Big|_{x_k}^{x_{k+1}} = f(x_k, y_k)(x_{k+1} - x_k) = y'_k h.$$

Entonces la fórmula (3) se escribe así:

$$y_{k+1} = y_k + y'_k h. \tag{3'}$$

Designando $y_{k+1} - y_k = \Delta y_k$, o sea, $y'_k h = \Delta y_k$, obtenemos

$$y_{k+1} = y_k + \Delta y_k. \tag{4}$$

Continuando este proceso y tomando cada vez como constante la función subintegral en el trozo correspondiente e igual a su valor al comienzo del trozo, obtenemos la tabla de soluciones de la ecuación diferencial sobre el segmento dado $[a, b]$.

El sentido geométrico del método de Euler consiste en lo siguiente. En el intervalo (x_0, x_1) la curva integral se reemplaza por el seg-

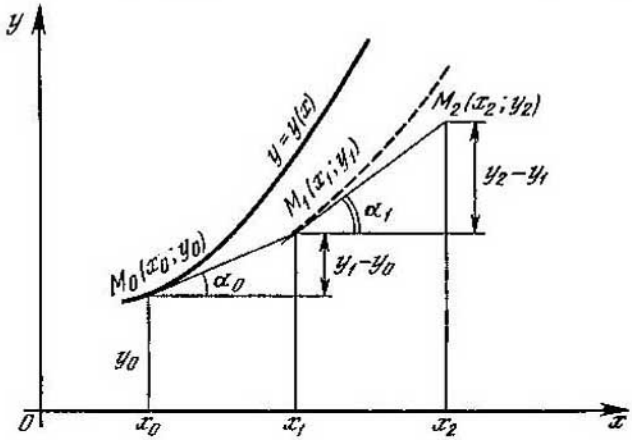


Fig. 9.2

mento de la tangente a ella, que pasa por el punto M_0 $(x_0; y_0)$. Como se ve de la fig. 9.2, el coeficiente angular de esta tangente vale

$$\operatorname{tg} \alpha_0 = (y_1 - y_0)/(x_1 - x_0) = f(x_0, y_0) = y'(x_0).$$

Luego del punto M_1 $(x_1; y_1)$ se traza un nuevo segmento de la tangente a la curva integral, que pasa por este punto. El coeficiente angular de tal tangente es

$$\operatorname{tg} \alpha_1 = (y_2 - y_1)/(x_2 - x_1) = f(x_1, y_1).$$

Continuando la construcción de tales segmentos, se obtiene la *quebrada de Euler*. La quebrada de Euler pasa por el punto dado M_0 (x_0, y_0) y aproxima la curva integral buscada.

Si la función $f(x, y)$ en cierto rectángulo

$$R\{\,|x - x_0| \leqslant a, \ |y - y_0| \leqslant b\}$$

satisface la condición

$$|f(x_1, y_1) - f(x_1, y_2)| \leqslant N\,|y_1 - y_2| \ (N = \text{const}) \tag{5}$$

y, además,

$$\left|\frac{df}{dx}\right| = \left|\frac{\partial f}{\partial x} + f\frac{\partial f}{\partial y}\right| \leqslant M \ (M = \text{const}), \tag{6}$$

tiene lugar la siguiente estimación del error:

$$|y(x_n)-y_n| \leqslant \frac{hM}{2N}[(1+hN)^n-1], \tag{7}$$

donde $y(x_n)$ es el valor de la solución exacta de la ecuación (1) para $x = x_n$ e y_n el valor aproximado, obtenido en el n-ésimo paso.

La fórmula (7) tiene, en lo principal, un empleo teórico. En la práctica se aplica, por regla general, un «cálculo doble». Primero el cálculo se realiza con paso h, luego el paso se divide y el cálculo repetido se ejecuta con paso $h/2$. El error de un valor más exacto y_n^* se estima por la fórmula

$$|y_n^* - y(x_n)| \approx |y_n^* - y_n|. \tag{8}$$

Ejemplo 1. Haciendo uso del método de Euler, integrar en el segmento [0; 1,5] la ecuación diferencial $y' = y - x$ que satisface la condición inicial $x_0 = 0$, $y_0 = 1{,}5$; el paso $h = 0{,}25$. Llevar a cabo los cálculos con cuatro cifras después de la coma.

△ Para la comodidad de los cálculos hagamos la tabla siguiente.

Tabla 9.2

i	x_i	y_i	$y_i' = y_i - x_i$	$\Delta y_i = hy_i'$
(1)	(2)	(3)	(4)	(5)
0	0	1,5000	1,5000	0,3750
1	0,25	1,8750	1,6250	0,4062
2	0,50	2,2812	1,7812	0,4453
3	0,75	2,7265	1,9765	0,4941
4	1,00	3,2206	2,2206	0,5552
5	1,25	3,7758	2,5258	0,6314
6	1,50	4,4702		

Paso I. Haciendo uso de los datos iniciales, llenamos la primera fila en las columnas (2) y (3).

Paso II. De la ecuación $y_i' = y_i - x_i$ calculamos y_i' ($i = 0, 1, \ldots, 5$) en la columna (4).

Paso III. Multiplicamos el contenido de la columna (4) por h (calculamos $\Delta y_i' = hy_i'$; $i = 0, 1, \ldots, 5$) y escribimos el resultado en la columna (5) de esta misma fila.

Paso IV. Al contenido de la columna 3 adicionamos el contenido de la columna (5) de esta misma fila (calculamos $y_{i+1} = y_i + \Delta y_i$; $i = 0, 1, \ldots, 5$) y escribimos el resultado en la columna (3) de la siguiente fila. Determinamos $x_{i+1} = x_i + h$ y luego repetimos los pasos II, III, IV hasta que quede recorrido todo el segmento [0; 1,5]. ▲

El método de Euler puede ser aplicado a la resolución de los sistemas de ecuaciones diferenciales y de las ecuaciones diferenciales de órdenes superiores. Sin embargo, en el último caso las ecuaciones diferenciales han de ser reducidas a un sistema de ecuaciones diferenciales de primer orden.

Supongamos que se da el sistema de dos ecuaciones de primer orden

$$\begin{cases} y' = f_1(x, y, z), \\ z' = f_2(x, y, z) \end{cases} \tag{9}$$

con las condiciones iniciales

$$y(x_0) = y_0, \; z(x_0) = z_0. \tag{10}$$

Los valores aproximados de $y(x_i) \approx y_i$ y de $z(x_i) \approx z_i$ se determinan por las fórmulas

$$y_{i+1} = y_i + \Delta y_i, \; z_{i+1} = z_i + \Delta z_i, \tag{11}$$

donde

$$\Delta y_i = hf_1(x_i, y_i, z_i), \; \Delta z_i = hf_2(x_i, y_i, z_i)$$
$$(i = 0, 1, 2, \ldots). \tag{12}$$

Ejemplo 2. Aplicando el método de Euler, resolver numéricamente el sistema de ecuaciones diferenciales $\begin{cases} y' = (z - y)x \\ z' = (z + y)x \end{cases}$ para las condiciones iniciales $y(0) = 1,0000$, $z(0) = 1,0000$ en el segmento $[0; 0,6]$; el paso $h = 0,1$. Realizar el cálculo con una cifra de reserva.

Tabla 9.3

i	x_i	y_i	$y_i' = (z_i - y_i)x_i$	$\Delta y_i = y_i'h$	z_i	$z_i' = (z_i + y_i)x_i$	$\Delta z_i = z_i'h$
0	0	1,0000	0	0	1,0000	0	0
1	0,1	1,0000	0	0	1,0000	0,2000	0,0200
2	0,2	1,0000	0,0040	0,0004	1,0200	0,4040	0,0404
3	0,3	1,0004	0,0180	0,0018	1,0604	0,6182	0,0618
4	0,4	1,0022	0,0480	0,0048	1,1222	0,8498	0,0850
5	0,5	1,0070	0,1001	0,0100	1,2072	1,1071	0,1107
6	0,6	1,0170			1,3179		

Para realizar los cálculos utilicemos la tabla 9.3. La sucesión de operaciones se ve claramente de la tabla. ▲

Ejemplo 3. Aplicando el método de Euler, construir en el segmento $[1; 1,5]$ la tabla de soluciones de la ecuación diferencial $y'' + \frac{y'}{x} + y = 0$ para las condiciones iniciales $y(1) = 0,77$, $y'(1) = -0,44$; el paso $h = 0,1$.

△ Con ayuda de la sustitución $y' = z$, $y'' = z'$ reemplacemos la ecuación dada por el sistema de ecuaciones

$$\begin{cases} y' = z, \\ z' = -\frac{z}{x} - y \end{cases}$$

i	x_i	y_i	$y'_i = z_i$	$\Delta y_i = hy'_i$	z_i	$z'_i = -\frac{z_i}{x_i} - y_i$	$\Delta z_i = z'_i h$
0	1,0	0,77	−0,44	−0,044	−0,44	−0,33	−0,033
1	1,1	0,726	−0,473	−0,047	−0,473	−0,296	−0,030
2	1,2	0,679	−0,503	−0,050	−0,503	−0,260	−0,026
3	1,3	0,629	−0,529	−0,053	−0,529	−0,222	−0,022
4	1,4	0,576	−0,551	−0,055	−0,551	−0,183	−0,018
5	1,5	0,521			−0,569		

para las condiciones iniciales $y\ (1) = 0{,}77$, $z\ (1) = -\ 0{,}44$. Ejecutemos los cálculos con una cifra de reserva. Para realizar los cálculos utilizamos la tabla 9.4. ▲

§ 9.5. Modificaciones del método de Euler

Método perfeccionado de Euler. Consideremos la ecuación diferencial

$$y' = f\ (x,\ y) \tag{1}$$

con la condición inicial

$$y\ (x_0) = y_0. \tag{2}$$

Se necesita hallar la solución de la ecuación (1) en el segmento $[a,\ b]$.

Dividamos el segmento $[a,\ b]$ en n partes iguales por medio de los puntos $x_i = x_0 + ih$ $(i = 0, 1, 2, \ldots, n)$, donde $h = (b - a)/n$ es el paso de integración. La esencia del método perfeccionado de Euler consiste en lo siguiente: primero se calculan los valores auxiliares de la función buscada $y_{i+1/2}$ en los puntos $x_{i+1/2} = x_i + \frac{h}{2}$ con ayuda de la fórmula

$$y_{i+1/2} = y_i + \frac{h}{2}\, y'_i, \tag{3}$$

luego se halla el valor el segundo miembro de la ecuación (1) en el punto medio $y'_{i+1/2} = f\ (x_{i+1/2},\ y_{i+1/2})$ y se determina

$$y'_{i+1} = y_i + hy'_{i+1/2}. \tag{4}$$

Observación. La estimación del error en el punto x_i puede ser obtenida con ayuda del «cálculo doble»; el cálculo se repite con paso $h/2$ y el error de un valor más exacto y_i^* (para el paso $h/2$) se estima aproximadamente del modo siguiente:

$$|y_i^* - y\,(x_i)| \approx \frac{1}{3}\,|y_i^* - y_i|, \tag{5}$$

donde $y(x)$ es la solución exacta de la ecuación diferencial. El método perfeccionado de Euler es más exacto en comparación con el examinado en el § 9.4.

Ejemplo 1. Integrar por el método perfeccionado de Euler la ecuación diferencial $y' = y - x$ para las condiciones iniciales $x_0 = 0$, $y_0 = 1{,}5$ en el segmento $[0, 1]$, tomando $h = 0{,}25$. Realizar los cálculos con cuatro cifras después de la coma.

△ Los resultados de cálculos se citan en la tabla 9.5. Esta tabla se llena del modo siguiente.

Tabla 9.5

i	x_i	y_i	$y'_i = f(x_i, y_i)$	$\frac{h}{2} \cdot y'_i$	$x_{i+1/2} = x_i + \frac{h}{2}$	$y_{i+1/2} = y_i + \frac{h}{2} y'_i$	$y'_{i+1/2} = f(x_{i+1/2}, y_{i+1/2})$	$hy'_{i+1/2}$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	0	1,5000	1,5000	0,1875	0,125	1,6875	1,5625	0,3906
1	0,25	1,8906	1,6406	0,2051	0,3750	2,0957	1,7207	0,4302
2	0,50	2,3208	1,8208	0,2276	0,6750	2,5484	1,8734	0,4684
3	0,75	2,7892	2,0392	0,2549	0,8750	3,0441	2,1691	0,5423
4	1,00	3,3315	2,3315	0,2914	1,1250	3,6229	2,4974	0,6243
5	1,25	3,9558	2,7058	0,3382	1,3750	4,2940	2,9190	0,7298
6	1,50	4,6856						

Paso I. Según los datos iniciales llenamos la primera fila en las columnas (2) y (3).

Paso II. De la ecuación $y'_i = f(x_i, y_i) = y_i - x_i$ calculamos y'_i para la columna (4) ($i = 0, 1, \ldots, 5$).

Paso III. Multiplicamos el contenido de la columna (4) por $h/2$ y de este modo determinamos $(h/2) \cdot y'_i$; escribimos el resultado en la columna (5).

Paso IV. Obtenemos el contenido de la columna (6) mediante la adición del valor corriente x_i y $h/2$.

Paso V. Al contenido de la columna (3) adicionamos el de la columna (5) y escribimos el resultado en la columna (7).

Paso VI. Sustituimos respectivamente los valores hallados $x_{i+1/2}$, $y_{i+1/2}$ [columnas (6) y (7)] en el segundo miembro de la ecuación diferencial dada, determinamos $y'_{i+1/2}$ y lo escribimos en la columna (8).

Paso VII. Multiplicamos el contenido de la columna (8) por el paso de integración h determinamos $hy'_{i+1/2}$ [columna (9)].

Paso VIII. Adicionamos el contenido de la columna (3) al de la columna (9) y escribimos el resultado obtenido $y_{i+1}=y_i+hy'_{i+1/2}$ ($i=0, 1, \ldots, 5$) en la columna (3) de la fila siguiente.

Luego repetimos todo el proceso de los cálculos, comenzando por el paso II. ▲

Método perfeccionado de Euler — Cauchy. La esencia del método de Euler — Cauchy consiste en lo siguiente. Primero se determina una magnitud auxiliar

$$\tilde{y}_{i+1}=y_i+hy'_i, \tag{6}$$

luego se calcula $\tilde{y}'_{i+1}=f(x_{i+1}, \tilde{y}_{i+1})$ y por la fórmula

$$y_{i+1}=y_i+h\cdot\frac{y'_i+\tilde{y}'_{i+1}}{2} \tag{7}$$

se halla la solución respectiva.

La estimación del error puede ser realizada por la fórmula (5) después de llevar a cabo el cálculo repetido con paso $h/2$.

Ejemplo 2. Valiéndose del método perfeccionado de Euler—Cauchy, integrar la ecuación diferencial dada en el ejemplo 1.

△ Los resultados de los cálculos se dan en la tabla 9.6.

Tabla 9.6

i	x_i	y_i	$y'_i=y_i-x_i$	hy'_i	x_{i+1}	$\tilde{y}_{i+1}=y_i+hy'_i$	$\tilde{y}'_{i+1}=\tilde{y}_{i+1}-x_{i+1}$	$h\tilde{y}'_{i+1}$	$\Delta y_i=\frac{hy'_i+h\tilde{y}'_{i+1}}{2}$
(1)	(2)	(2)	(3)	(4)	(5)	(7)	(8)	(9)	(10)
0	0	1,5000	1,5000	0,3750	0,25	1,8750	1,625	0,4062	0,3906
1	0,25	1,8906	1,6406	0,4102	0,50	2,3008	1,8008	0,4502	0,4302
2	0,50	2,3208	1,8208	0,4552	0,75	2,7760	2,0260	0,5065	0,4808
3	0,75	2,8016	2,0516	0,5129	1,00	3,3145	2,3145	0,5786	0,5458
4	1,00	3,3474	2,3474	0,5868	1,25	3,9342	2,6842	0,6710	0,6289
5	1,25	3,9763	2,7263	0,6816	1,50	4,6579	3,1579	0,7895	0,7355
6	1,50	4,7118							

La tabla se llena del modo siguiente.

Paso I. Según los datos iniciales llenamos la primera fila en las columnas (2) y (3).

Paso II. Determinamos el valor $y'_i=f(x_i, y_i)=y_i-x_i$ ($i=0, 1, \ldots, 5$) para la columna (4).

Paso III. El valor hallado y_i de la columna (4) lo multiplicamos por el paso de integración h y escribimos el resultado en la columna (5).

Paso IV. Determinamos $x_{i+1} = x_i + h$ $(i = 0, 1, \ldots, 5)$ para la columna (6).

Paso V. Adicionamos al contenido de la columna (3) el de la columna (5) y apuntamos el resultado en la columna (7), o sea, determinamos $\tilde{y}_{i+1} = y_i + hy'_i$.

Paso VI. Sustituimos los valores hallados x_{i+1} e y_{i+1} en el segundo miembro de la ecuación diferencial dada y determinamos $\tilde{y}'_{i+1}$ para la columna (8).

Paso VII. Multiplicamos el resultado de la columna (8) por el paso de integración h y determinamos $h\tilde{y}'_{i+1}$ [columna (9)].

Paso VIII. Encontramos Δy_i [columna (10)] para lo cual determinamos la semisuma de las magnitudes escritas en las columnas (5) y (9).

Paso IX. Adicionamos al contenido de la columna (3) el contenido de la columna (10) y apuntamos el resultado en la columna (3) de la fila siguiente, o sea, determinamos $y_{i+1} = y_i + \Delta y_i$.

Luego repetimos todo el proceso de los cálculos, comenzando por el paso II. ▲

Método perfeccionado de Euler—Cauchy con tratamiento iterativo sucesivo. El método de Euler—Cauchy con tratamiento iterativo es más exacto que el de Euler— Cauchy anteriormente considerado. Su esencia consiste en que se realiza el tratamiento iterativo de cada valor hallado y_i. Primero se elige la aproximación no exacta

$$y_{i+1}^{(0)} = y_i + hf(x_i,\ y_i), \tag{8}$$

luego se construye el proceso iterativo:

$$y_{i+1}^{(k)} = y_i + \frac{h}{2}\,[f(x_i,\ y_i) + f(x_{i+1},\ y_{i+1}^{(k-1)})]. \tag{9}$$

La iteración se continúa hasta que dos aproximaciones sucesivas $y_{i+1}^{(k)}$ y $y_{i+1}^{(k+1)}$ no coincidan en los signos que interesan al calculador. Luego se toma $y_{i+1} \approx y_{i+1}^{k+1}$. Si, realizadas tres o cuatro iteraciones con el valor elegido de h, no llega la coincidencia de los signos necesarios, hace falta disminuir el paso de cálculo h.

Ejemplo 3. Aplicando el método de tratamiento iterativo, hallar con una exactitud de hasta cuatro cifras decimales coincidentes la solución de la ecuación $y' = y - x$ para las condiciones iniciales $y(0) = 1$. La solución ha de obtenerse sobre el segmento [0, 1,5], al elegir $h = 0{,}25$.

△ Según la fórmula (8) encontramos

$$y_1^{(0)} = y_0 + h\,(y_0 - x_0) = 1{,}5000 + 0{,}375 = 1{,}8750$$

Luego, aplicando el proceso iterativo (9), determinamos sucesivamente

$$y_1^{(1)} = y_0 + \frac{h}{2}\,[(y_0 - x_0) - (y_1^{(0)} - x_1)] =$$
$$= 1{,}5000 + 0{,}125\,(1{,}5000 + 1{,}875 - 0{,}25) = 1{,}89062;$$
$$y_1^{(2)} = y_0 + \frac{h}{2}\,[(y_0 - x_0) + (y_1^{(1)} - x_1)] =$$
$$= 1{,}5000 + 0{,}125\,(1{,}5000 + 1{,}89062 - 0{,}25) = 1{,}89258;$$
$$y_1^{(3)} = y_0 + \frac{h}{2}\,[(y_0 - x_0) + (y_1^{(2)} - x_1)] =$$
$$= 1{,}5000 + 0{,}125\,(1{,}5000 + 1{,}89258 - 0{,}25) = 1{,}89282;$$
$$y_1^{(4)} = y_0 + \frac{h}{2}\,[(y_0 - x_0) + (y_1^{(3)} - x_1)] =$$
$$= 1{,}5000 + 0{,}125\,(1{,}5000 + 1{,}89282 - 0{,}25) = 1{,}89285.$$

En las dos últimas aproximaciones coinciden cuatro cifras. Por eso después del redondeo se puede tomar $y_1 \approx 1{,}8929$.

Volviendo a hacer uso de la fórmula (8) para $i = 1$ encontramos

$$y_2^{(0)} = y_1 + hf\,(x_1, y_1) = y_1 + h\,(y_1 - x_1) =$$
$$= 1{,}8929 + 0{,}25\,(1{,}8929 - 0{,}25) = 2{,}3036.$$

Por la fórmula (9) determinamos las aproximaciones sucesivas:

$$y_2^{(1)} = 1{,}8929 + 0{,}125\,[1{,}6429 + (2{,}3036 - 0{,}50)] = 2{,}3237;$$
$$y_2^{(2)} = 1{,}8929 + 0{,}125\,[1{,}6429 + (2{,}3237 - 0{,}50)] = 2{,}32622;$$
$$y_2^{(3)} = 1{,}8929 + 0{,}125\,[1{,}6429 + (2{,}32622 - 0{,}50)] = 2{,}32654;$$
$$y_2^{(4)} = 1{,}8929 + 0{,}125\,[1{,}6429 + (2{,}32654 - 0{,}50)] = 2{,}32658.$$

Se pueden cesar las iteraciones y tomar $y_2 \approx 2{,}3266$. Aplicando luego las fórmulas (8) y (9), obtenemos la solución de la ecuación dada. Los resultados de los cálculos se dan en la tabla 9.7.▲

Tabla 9.7

i	x_i	y_i	$y_{i+1}^{(0)}$	$y_{i+1}^{(1)}$	$y_{i+1}^{(2)}$	$y_{i+1}^{(3)}$	$y_{i+1}^{(4)}$	y_{i+1}
0	0	1,5000	1,875	1,89062	1,89258	1,89282	1,89285	1,8929
1	0,25	1,8929	2,3036	2,3237	2,32622	2,32654	2,32658	2,3266
2	0,50	2,3266	2,78325	2,80908	2,81231	2,81271	2,81276	2,8128
3	0,75	2,8128	3,3285	3,36171	3,36586	3,3664	3,36645	3,3664
4	1,00	3,3664	3,9580	4,0007	4,00603	4,0067	4,00679	4,0068
5	1,25	4,0068	4,6960	4,7509	4,75776	4,75870	4,75872	4,7587
6	1,50	4,7587						

§ 9.6. Método de Runge—Kutta

El método de Runge—Kutta es uno de los métodos de precisión elevada. Tiene mucho de común con el de Euler.

Supongamos que sobre el segmento $[a, b]$ se necesita hallar la solución numérica de la ecuación

$$y' = f(x, y) \tag{1}$$

con la condición inicial

$$y(x_0) = y_0. \tag{2}$$

Dividamos el segmento $[a, b]$ en n partes iguales por los puntos $x_i = x_0 + ih$ $(i = 0, 1, \ldots, n)$, donde $h = (b - a)/n$ es el paso de integración. Al igual que en el método de Euler, en el de Runge-Kutta los valores sucesivos y_i de la función buscada y se determinan con ayuda de la fórmula

$$y_{i+1} = y_1 + \Delta y_i. \tag{3}$$

Si desarollamos la función y en serie de Taylor y nos limitamos a los términos hasta h^4 inclusivamente, el incremento de la función Δy puede ser representado en la forma

$$\Delta y = y(x+h) - y(x) = hy'(x) + \frac{h^2}{2} y''(x) + \\ + \frac{h^3}{6} y'''(x) + \frac{h^4}{24} y^{\mathrm{IV}}(x), \tag{4}$$

donde las derivadas $y''(x)$, $y'''(x)$, $y^{\mathrm{IV}}(x)$ se obtienen de la ecuación (1) mediante la derivación sucesiva.

En vez de los cálculos inmediatos con ayuda de la fórmula (4) en el método de Runge-Kutta se determinan cuatro números:

$$\begin{aligned} k_1 &= hf(x, y), \\ k_2 &= hf\left(x + \frac{h}{2}, y + \frac{k_1}{2}\right), \\ k_3 &= hf\left(x + \frac{h}{2}, y + \frac{k_2}{2}\right), \\ k_4 &= hf(x + h, y + k_3). \end{aligned} \tag{5}$$

Se puede mostrar que si a los números k_1, k_2, k_3, y k_4 se les asigna el peso de 1/6; 1/3; 1/3; 1/6, respectivamente, la media ponderada de estos números, o sea,

$$\frac{1}{6} k_1 + \frac{1}{3} k_2 + \frac{1}{3} k_3 + \frac{1}{6} k_4 \tag{6}$$

con exactitud de hasta los cuartos grados, es igual al valor de Δy calculado por medio de la fórmula (4):

$$\Delta y = \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4). \tag{7}$$

Ahora bien, para cada par de valores corrientes x_i e y_i con ayuda de las fórmulas (5) se determinan los valores

$$
\begin{aligned}
k_1^{(i)} &= hf(x_i,\ y_i),\\
k_2^{(i)} &= hf\left(x_i+\frac{h}{2},\ y_i+\frac{k_1^{(i)}}{2}\right),\\
k_3^{(i)} &= hf\left(x_i+\frac{h}{2},\ y_i+\frac{k_2^{(i)}}{2}\right),\\
k_4^{(i)} &= hf(x_i+h,\ y_i+k_3^{(i)})
\end{aligned}
\tag{8}
$$

y por la fórmula (7) se halla

$$\Delta y_i = \frac{1}{6}(k_1^{(i)}+2k_2^{(i)}+2k_3^{(i)}+k_4^{(i)})$$

y luego

$$y_{i+1} = y_i + \Delta y_i.$$

Los números k_1, k_2, k_3 y k_4 tienen un sentido geométrico simple. Supongamos que la curva M_0CM_1 (fig. 9.3) es la solución de la ecua-

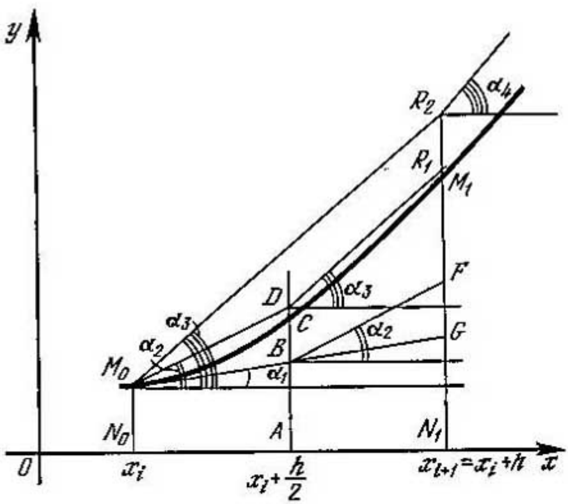


Fig. 9.3

ción diferencial (1) con la condición inicial (2). El punto C de esta curva se halla en la recta que es paralela al eje Oy y biseca el segmento $[x_i, x_{i+1}]$; B y G son los puntos de intersección de la tangente trazada a la curva en el punto M_0 con las ordenadas AC y N_1M_1. Entonces el número k_1, con exactitud de hasta el multiplicador h (donde $h = x_{i+1} - x_i$), es el coeficiente angular de la tangente en el punto M_0 a la curva integral M_0CM_1, o sea, $k_1 = hy'_i = hf(x_i, y_i)$.

El punto B tiene las coordenadas $x = x_i + \frac{h}{2}$, $y = y_i + \frac{k_1}{2}$. Por consiguiente, el número k_2, con exactitud de hasta el multiplicador h, es el coeficiente angular de la tangente trazada a la curva integral en el punto B (BF es el segmento de esta tangente).

Tracemos por el punto M_0 la recta paralela al segmento BF. Entonces la curva D tiene las coordenadas $x = x_i + \frac{h}{2}$, $y = y_i + \frac{k_2}{2}$ y k_3, con exactitud de hasta el multiplicador h, que es el coeficiente angular de la tangente trazada a la curva integral en el punto D (DR_1 es el segmento de esta tangente). Por último, por el punto M_0 tracemos la recta, paralela a DR_1, que cortará la prolongación M_1N_1 en el punto R_2 ($x_i + h$; $y_i + k_3$). Entonces k_4, con la exactitud de hasta el multiplicador h, es el coeficiente angular de la tangente trazada a la curva integral en el punto R_2.

Los cálculos según el método de Runge-Kutta es cómodo disponerlos de acuerdo con el esquema indicado en la tabla 9.8. Esta tabla se llena del modo siguiente.

Paso I. En las columnas (2) y (3) de la fila correspondiente se apuntan los valores necesarios de x e y. (Si la fila es la primera, se apuntan los datos iniciales x_0 e y_0).

Paso II. Los valores de x e y de la fila correspondiente se sustituyen en el segundo miembro de la ecuación diferencial (1), se determina $f(x, y)$ y se escribe en la columna (4) de esta misma fila.

Paso III. El valor obtenido de $f(x, y)$ en la columna (4) se multiplica por el paso de integración h, se calcula $k = hf(x, y)$ y se apunta en la columna (5) de esta misma fila.

Paso IV. Los valores hallados de k se multiplican por el coeficiente correspondiente (por 1 si éste es k_1 o k_4, o bien por 2 si éste es k o k_3), el resultado se escribe en la columna (6) de la fila respectiva.

Los pasos I, II, III, IV se repiten para determinar cada k en la i-ésima solución.

Los resultados de la sexta fila se suman, se dividen por 6 y se determina $\Delta y_i = \frac{1}{6}\Sigma$ e $y_{i+1} = y_i + \Delta y_i$.

Luego se repiten todos los cálculos, comenzando por el paso I, hasta que no se recorra todo el segmento $[a, b]$.

El método de Runge—Kutta tiene el orden de precisión igual a h^4 sobre todo el segmento $[a, b]$. Es muy difícil estimar la exactitud de este método. Una estimación aproximada del error puede ser obtenida con ayuda del «cálculo doble», haciendo uso de la fórmula

$$|y_i^* - y(x_i)| \approx \frac{y_i^* - y_i}{15}, \tag{9}$$

donde $y(x_i)$ es el valor de la solución exacta de la ecuación (1) en el punto x_i, mientras que y_i^* e y_i son los valores aproximados, obtenidos con el paso $h/2$ y h.

Tabla 9.8

i	x	y	$y' = f(x, y)$	$k = hf \times$ $\times (x, y)$	Δy
(1)	(2)	(3)	(4)	(5)	(6)
0	x_0	y_0	$f(x_0, y_0)$	$k_1^{(0)}$	$k_1^{(0)}$
	$x_0+\frac{h}{2}$	$y_0+\frac{k_1^{(0)}}{2}$	$f\left(x_0+\frac{h}{2}, y_0+\frac{k_1^{(0)}}{2}\right)$	$k_2^{(0)}$	$2k_2^{(0)}$
	$x_0+\frac{h}{2}$	$y_0+\frac{k_2^{(0)}}{2}$	$f\left(x_0+\frac{h}{2}, y_0+\frac{k_2^{(0)}}{2}\right)$	$k_3^{(0)}$	$2k_3^{(0)}$
	x_0+h	$y_0+k_3^{(0)}$	$f(x_0+h, y_0+k_3^{(0)})$	$k_4^{(0)}$	$k_4^{(0)}$
					$\frac{1}{6}\Sigma = \Delta y_0$
1	x_1	$y_1 = y_0 + \Delta y_0$	$f(x_1, y_1)$	$k_1^{(1)}$	$k_1^{(1)}$
	$x_1+\frac{h}{2}$	$y_1+\frac{k_1^{(1)}}{2}$	$f\left(x_1+\frac{h}{2}, y_1+\frac{k_1^{(1)}}{2}\right)$	$k_2^{(1)}$	$2k_2^{(1)}$
	$x_1+\frac{h}{2}$	$y_1+\frac{k_2^{(1)}}{2}$	$f\left(x_1+\frac{h}{2}, y_1+\frac{k_2^{(1)}}{2}\right)$	$k_3^{(1)}$	$2k_3^{(1)}$
	x_1+h	$y_1+k_3^{(1)}$	$f(x_1+h, y_1+k_3^{(1)})$	$k_4^{(1)}$	$k_4^{(1)}$
					$\frac{1}{6}\Sigma = \Delta y_1$
2	x_2	$y_2 = y_1 + \Delta y_1$			

Si ε es la exactitud prefijada de la solución, el número n (número de divisiones) para determinar el paso de integración $h = (b - a)/n$ se elige de modo que

$$h^4 < \varepsilon \tag{10}$$

No obstante, el paso de cálculo se puede cambiar al pasar de un punto a otro.

Para estimar lo correcto de la elección del paso h se utiliza la igualdad

$$q = \left| \frac{k_2^{(i)} - k_3^{(i)}}{k_1^{(i)} - k_2^{(i)}} \right|, \tag{11}$$

donde q debe ser igual a algunas centésimas, en el caso contrario el paso h ha de disminuirse.

Ejemplo 1. Se da la ecuación diferencial $y' = y - x$ que satisface la condición inicial $y(0) = 1,5$. Calcular con exactitud de hasta 0,01 la solución de esta ecuación para $x = 1,5$. Realizar el cálculo, haciendo uso del método de Runge—Kutta, con dos cifras de reserva.

△ Elegimos el paso inicial de los cálculos de h partiendo de la condición $h^4 < 0,001$. Entonces $h < 0,3$. Para comodidad de los cálculos tomemos $h = 0,25$. Todo el segmento de integración $[0, 1,5]$ se dividirá en seis partes iguales por los puntos $x_0 = 0$; $x_1 = 0,25$; $x_2 = 0,50$, $x_3 = 0,75$; $x_4 = 1,00$; $x_5 = 1,25$; $x_6 = 1,50$. De las condiciones iniciales tenemos $x_0 = 0$; $y_0 = 1,5$. Determinemos la primera aproximación $y_1 + \Delta y_0$, donde

$$\Delta y_0 = \frac{1}{6}(k_1^{(0)} + 2k_2^{(0)} + 2k_3^{(0)} + k_4^{(0)}).$$

Utilizando las fórmulas (8), obtenemos

$$k_1^{(0)} = (y_0 - x_0)\,h = 1,5000 \cdot 0,25 = 0,3750.$$

$$k_2^{(0)} = \left[\left(y_0 + \frac{k_1^{(0)}}{2}\right) - \left(x_0 + \frac{h}{2}\right)\right] h = [(1,5000 + 0,1875) - \\ - 0,125] \cdot 0,25 = 0,3906,$$

$$k_3^{(0)} = \left[\left(y_0 + \frac{k_2^{(0)}}{2}\right) - \left(x_0 + \frac{h}{2}\right)\right] h = \\ = [(1,5000 + 0,1953) - 0,125] \cdot 0,25 = 0,3926.$$

$$k_4^{(0)} = [(y_0 + k_3^{(0)}) - (x_0 + h)]\,h = [(1,5000 + 0,3926) - \\ - 0,125]\,0,25 = 0,4106.$$

Por lo tanto,

$$\Delta y_0 = \frac{1}{6}(0,3750 + 2 \cdot 0,3906 + 2 \cdot 0,3926 + 0,4106) = 0,3920;$$

$$y_1 = 1,5000 + 0,3920 = 1,8920.$$

La ulterior resolución de la ecuación se representa en la tabla 9.9 (véanse las págs. 404). Ahora bien, finalmente tenemos $y(1,5) = 4,74$. ▲

El método de Runge—Kutta puede ser aplicado también para la resolución de los sistemas de ecuaciones diferenciales.

Supongamos que se da el sistema de ecuaciones diferenciales de primer orden.

$$\begin{cases} y' = f(x, y, z), \\ z' = g(x, y, z) \end{cases} \tag{12}$$

con las condiciones iniciales

$$x = x_0,\ y(x_0) = y_0,\ z(x_0) = z_0. \tag{13}$$

Tabla 9.9

i	x	y	$y' = f(x, y)$	$k = hf(x, y)$	Δy
(1)	(2)	(3)	(4)	(5)	(6)
0	0 0,125 0,125 0,25	1,5000 1,6875 1,6953 1,8926	1,5000 1,5625 1,5703 1,6426	0,3750 0,3906 0,3926 0,4106	0,3750 0,7812 0,7852 0,4106
					0,3920
1	0,25 0,375 0,375 0,50	1,8920 2,0973 2,1073 2,3251	1,6420 1,7223 1,7323 1,8251	0,4105 0,4306 0,4331 0,4562	0,4105 0,8612 0,8662 0,4562
					0,4323
2	0,50 0,625 0,625 0,75	2,2343 2,5523 2,5652 2,8093	1,8243 1,9273 1,9402 2,0593	0,4561 0,4818 0,4850 0,5148	0,4561 0,9636 0,9700 0,5148
					0,4841
3	0,75 0,875 0,875 1,00	2,8084 3,0657 3,0823 3,3602	2,0584 2,1907 2,2073 2,3602	0,5146 0,5477 0,5518 0,5900	0,5146 1,0954 1,1036 0,5900
					0,5506
4	1,00 1,125 1,125 1,25	3,3590 3,6539 3,6751 3,9965	2,3590 2,5289 2,5501 2,7465	0,5898 0,6322 0,6375 0,6686	0,5898 1,2641 1,2750 0,6866
					0,6360
5	1,25 1,375 1,375 1,50	3,9950 4,3381 4,3654 4,7426	2,7450 2,9631 2,9904 3,2426	0,6862 0,7408 0,7476 0,8106	0,6862 1,4816 1,4952 0,8106
					9,7456
6	1,50	4,7406			

En este caso paralelamente se determinan los números Δy_i y Δz_i:

$$\Delta y_i = \frac{1}{6}\left(k_1^{(i)} + 2k_2^{(i)} + 2_3^{(i)} + k_4^{(i)}\right),$$
$$\Delta z_i = \frac{1}{6}\left(l_1^{(i)} + 2l_2^{(i)} + 2l_3^{(i)} + l_4^{(i)}\right), \tag{14}$$

donde

$$k_1^{(i)} = hf(x_i,\ y_i,\ z_i);$$
$$l_1^{(i)} = hg(x_i,\ y_i,\ z_i);$$
$$k_2^{(i)} = hf\left(x_i + \frac{h}{2},\ y_i + \frac{k_1^{(i)}}{2},\ z_i + \frac{l_1^{(i)}}{2}\right);$$
$$l_2^{(i)} = hg\left(x_i + \frac{h}{2},\ y_i + \frac{k_1^{(i)}}{2},\ z_i + \frac{l_1^{(i)}}{2}\right);$$
$$k_3^{(i)} = hf\left(x_i + \frac{h}{2},\ y_i + \frac{k_2^{(i)}}{2},\ z_i + \frac{l_2^{(i)}}{2}\right);$$
$$l_3^{(i)} = hg\left(x_i + \frac{h}{2},\ y_1 + \frac{k_2^{(i)}}{2},\ z_i + \frac{l_2^{(i)}}{2}\right);$$
$$k_4^{(i)} = hf(x_i + h,\ y_i + k_3^{(i)},\ z_i + l_3^{(i)});$$
$$l_4^{(i)} = hg(x_i + h,\ y_i + k_3^{(i)},\ z_i + l_3^{(i)}).$$

Entonces obtenemos la solución del sistema

$$y_{i+1} = y_i + \Delta y_i,\ z_{i+1} = z_i + \Delta z_i$$

Ejemplo 2. Se da el sistema de ecuaciones diferenciales

$$\begin{cases} y = \dfrac{2y - x}{z} \\ z = \dfrac{2y}{z + x} \end{cases}$$

para las condiciones iniciales $x_0 = 0{,}5$, $y_0 = 1$, $z_0 = 1$. Hallar la solución del sistema para $x = 0{,}6$. Los cálculos se realizarán con cinco cifras después de la coma.

△ Elegimos el paso $h = 0{,}1$ y hallamos los números k_1, l_1, k_2, l_2, k_3, l_3, k_4, l_4.

$$k_1 = h \cdot \frac{2y_0 - x_0}{z_0} = 0{,}1 \cdot \frac{2 - 0{,}5}{1} = 0{,}15000;$$

$$l_1 = h \cdot \frac{2y_0}{z_0 + x_0} = 0{,}1 \cdot \frac{2}{1{,}5} = 0{,}13333;$$

$$k_2 = h\left[\frac{2\left(y_0 + \frac{k_1}{2}\right) - \left(x_0 + \frac{h}{2}\right)}{z_0 + \frac{l_1}{2}}\right] = 0{,}1\,\frac{2 \cdot 1{,}075 - 0{,}55}{1{,}06667} = 0{,}14100;$$

$$l_2=h\left[\frac{2\left(y_0+\frac{k_1}{2}\right)}{\left(z_0+\frac{l_1}{2}\right)+\left(x_0+\frac{h}{2}\right)}\right]=0,1\,\frac{2\cdot 1,075}{1,06667+0,55}=0,13299;$$

$$k_3=h\left[\frac{2\left(y_0+\frac{k_2}{2}\right)-\left(x_0+\frac{h}{2}\right)}{z_0+\frac{l_2}{2}}\right]=0,1\cdot\frac{2\cdot 1,07050-0,55}{1,06650}=0,14918;$$

$$l_3=h\left[\frac{2\left(y_0+\frac{k_2}{2}\right)}{\left(z_0+\frac{l_2}{2}\right)+\left(x_0+\frac{h}{2}\right)}\right]=0,1\cdot\frac{2\cdot 1,07050}{1,06650+0,55}=0,13245;$$

$$k_4=h\left[\frac{2\,(y_0+k_3)-(x_0+h)}{z_0+l_3}\right]=0,1\cdot\frac{2\cdot 1,14918-0,16}{1,13245}=0,14998;$$

$$l_4=h\left[\frac{2\,(y_0+k_3)}{(z_0+l_3)+(z_0+h)}\right]=0,1\cdot\frac{2\cdot 1,14918}{1,13245+0,6}=0,13266;$$

Por consiguiente,

$$\Delta y_0=\frac{1}{6}\,(0,15+2\cdot 0,14100+2\cdot 0,14918+0,14998)=0,14672;$$

$$\Delta z_0=\frac{1}{6}\,(0,13333+2\cdot 0,13299+2\cdot 0,13245+0,13266)=0,13281$$

y finalmente obtenemos el valor de las funciones buscadas en el punto $x=0,6$:

$$y_1=1+0,14672=1,14672;\quad z_1=1+0,13281=1,13281.\ \blacktriangle$$

§ 9.7. Método de extrapolación de Adams

Al resolver una ecuación diferencial por el método de Runge-Kutta es necesario realizar muchos cálculos para hallar cada y_i. En el caso cuando el segundo miembro de la ecuación tiene una expresión analítica complicada, la resolución de la misma con ayuda del método de Runge-Kutta es acompañada de grandes dificultades. Por eso en la práctica se aplica el *método de Adams* que no requiere hacer un cálculo múltiple del segundo miembro de la ecuación.

Supongamos que se da la ecuación diferencial

$$y'=f\,(x,\,y)\tag{1}$$

con la condición inicial

$$x=x_0,\ y\,(x_0)=y_0.\tag{2}$$

Se necesita hallar la solución de esta ecuación en el segmento $[a,\,b]$.

Dividamos el segmento $[a,\,b]$ en n partes iguales por los puntos $x_i=x_0+ih$ $(i=0,\,1,\,2,\,\ldots,\,n)$. Elijamos el trozo $[x_i,\ x_{i+1}]$ e integremos la ecuación diferencial (1); entonces obtenemos

$$y_{i+1}=y_i+\int_{x_i}^{x_{i+1}} y'\,dx,$$

o bien

$$\Delta y_i=\int_{x_i}^{x_{i+1}} y'\,dx. \tag{3}$$

Para hallar la derivada hagamos uso de la segunda fórmula de interpolación de Newton (limitándonos en este caso por las diferencias de tercer orden);

$$y'=y'_i+t\Delta y'_{i-1}+\frac{t(t+1)}{2!}\Delta^2 y'_{i-2}+$$
$$+\frac{t(t+1)(t+2)}{3!}\Delta^3 y'_{i-3}, \tag{4}$$

donde $t=(x-x_i)/h$ o bien

$$y'=y'_i+t\Delta y'_{i-1}+\frac{t^2+t}{2}\Delta^2 y'_{i-2}+\frac{t^3+3t^2+2}{6}\Delta^3 y'_{i-3}. \tag{4'}$$

Sustituyendo la expresión para y' de la fórmula (4′) en la relación (3) y teniendo en cuenta que $dx = h\,dt$, tenemos

$$\Delta y_i=h\int_0^1\left(y'_i+t\Delta y'_{i-1}+\frac{t^2+t}{2}\Delta^2 y'_{i-2}+\frac{t^3+3t^2+2t}{6}\Delta^3 y'_{i-3}\right)dt=$$
$$=hy'_i+\frac{1}{2}\Delta(hy'_{i-1})+\frac{5}{12}\Delta^2(hy'_{i-2})+\frac{3}{8}\Delta^3(hy'_{i-3}). \tag{5}$$

Designemos a continuación

$q_i = y'_ih = f(x_i, y_i)\cdot h \ (i = 0, 1, 2, \ldots, n)$.

Entonces para toda diferencia tenemos $\Delta^m q_i = \Delta^m(y'_ih)$ y

$$\Delta y_i=q_i+\frac{1}{2}\Delta q_{i-1}+\frac{5}{12}\Delta^2 q_{i-2}+\frac{3}{8}\Delta^3 q_{i-3}. \tag{6}$$

Con ayuda de la fórmula $y_{i+1} = y_i+\Delta y_i$ obtenemos la solución de la ecuación. La fórmula (6) lleva el nombre de *fórmula de extrapolación de Adams.*

Para el comienzo del proceso se necesitan cuatro valores iniciales y_0, y_1, y_2 e y_3, o sea, el llamado *segmento inicial* que puede ser hallado partiendo de la condición inicial (2) con la utilización de uno de los métodos conocidos. Por lo general, el segmento inicial de la solución se encuentra con ayuda del método de Runge—Kutta. Conociendo y_0, y_1, y_2, y_3, se puede determinar

$$q_0 = hy'_0 = hf(x_0, y_0);\ q_1 = hy'_1 = hf(x_1, y_1);$$
$$q_2 = hy'_2 = hf(x_2, y_2);\ q_3 = hy'_3 = hf(x_3, y_3). \tag{7}$$

Luego se hace la tabla de diferencias de la magnitud q (tabla 9.10).

Tabla 9.10

i	x_i	y_i	Δy_i	$y'_i = f(x_i, y_i)$	$q = hy'_i$	Δq_i	$\Delta^2 q_i$	$\Delta^3 q_i$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	x_0	y_0		$f(x_0, y_0)$	q_0	Δq_0	$\Delta^2 q_0$	$\Delta^3 q_0$
1	x_1	y_1		$f(x_1, y_1)$	q_1	Δq_1	$\Delta^2 q_1$	
2	x_2	y_2		$f(x_2, y_2)$	q_2	Δq_2		
3	x_3	y_3	Δy_3	$f(x_3, y_3)$	q_3			
4	x_4	y_4						
5	x_5							
6	x_6							

El método de Adams consiste en la prolongación de la tabla diagonal de diferencias con ayuda de la fórmula (6). Utilizando los números q_3, Δq_2, $\Delta^2 q_1$, $\Delta^3 q_0$ que se disponen en la tabla en diagonal, con ayuda de la fórmula (6), suponiendo en ésta $n = 3$ (el último valor conocido de y es y_3), se obtiene:

$$\Delta y_3 = q_3 + \frac{1}{2}\Delta q_2 + \frac{5}{12}\Delta^2 q_1 + \frac{3}{8}\Delta^3 q_0.$$

El valor obtenido Δy_3 se apunta en la tabla y se halla $y_4 = y_3 + \Delta y_3$. Luego, utilizando x_4 y el valor hallado y_4, se encuentra $f(x_4, y_4)$, q_4, Δq_3, $\Delta^2 q_2$, $\Delta^3 q_1$, o sea, se obtiene una nueva diagonal. Basándose en estos datos, se halla

$$\Delta y_4 = q_3 + \frac{1}{2}\Delta q_3 + \frac{5}{12}\Delta^2 q_2 + \frac{3}{8}\Delta^3 q_1;\ y_5 = y_4 + \Delta y_4.$$

Ahora bien, se prolonga la tabla de solución, calculando el segundo miembro de la ecuación diferencial (1) en cada etapa sólo una vez.

Para la estimación aproximada del error se aplica el **principio de Runge** que consiste en lo siguiente:

1) se encuentra la solución de la ecuación diferencial para el paso h;

2) el valor del paso h se duplica y se halla la solución con el paso $H = 2h$;

3) se calcula el error del método, haciendo uso de la fórmula

$$\varepsilon = \frac{|\tilde{y}_n - \tilde{y}_{2n}|}{2^m - 1}, \tag{8}$$

donde $\tilde{y}_n$ es el valor del cálculo aproximado con el paso doble $H = 2h$, e $\tilde{y}_{2n}$, el valor del cálculo aproximado con el paso h.

Observación. Al hacer el cálculo con el paso h se supone que en cada paso se comete un error proporcional a h^{m+1} y con el paso $2h$, un error proporcional a $(2h)^{m+1}$ si el orden de precisión del método está determinado y es igual a h^m.

Nótese que en la fórmula de extrapolación de Adams (6) las diferencias finitas terceras Δ_q^2 se consideran constantes. Por eso la magnitud h del paso inicial se puede obtener de la desigualdad $h^4 < \varepsilon$, donde ε es la exactitud prefijada de la solución.

En la práctica se controla el proceso de diferencias finitas terceras, eligiendo h tal que las diferencias vecinas $\Delta^3 q_i$ y $\Delta^3 q_{i+1}$ se distingan entre sí no más que en una-dos unidades del orden prefijado (sin contar las cifras de reserva).

Ejemplo 1. Calcular para $x = 1{,}5$ con exactitud de hasta 0,01, valiéndose del método de Adams, el valor de la solución de la ecuación diferencial $y' = y - x$ si $x_0 = 0$ e $y_0 = 1{,}5$. Es cómodo hacer todos los cálculos con dos cifras de reserva.

△ Al igual que antes elegimos h de la relación $h^4 < 0{,}01$, o sea, $h = 0{,}25$. Tomemos el segmento inicial y_0, y_1, y_2, y_3 de la solución del ejemplo 1 dado en el § 9.6. Para resolver esta ecuación hacemos dos tablas: la principal (tabla 9.11) y la auxiliar (tabla 9.12). Su finalidad es clara de las mismas tablas.

Tabla 9.11

i	x_i	y_i	Δy_i	$y'_i = f(x_i y_i)$	$q_i = hy'_i$	Δq_i	$\Delta^2 q_i$	Δq^3_i
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
0	0	1,5000		1,5000	0,3750	0,0355	0,0101	0,0028
1	0,25	1,8920		1,6420	0,4105	0,0456	0,0129	0,0037
2	0,50	2,3243		1,8243	0,4561	0,0585	0,0166	0,0047
3	0,75	2,8084	0,5504	2,0584	0,5146	0,0751	0,0213	
4	1,00	3,3588	0,6356	2,3588	0,5897	0,0964		
5	1,25	3,9944	0,7450	2,7444	0,6861			
6	1,50	4,7394						

Definitivamente tenemos $y(1{,}5) = 4{,}74$. ▲

El método de Adams se emplea también para resolver los sistemas de ecuaciones diferenciales y las ecuaciones diferenciales de n-ésimo orden.

Tabla 9.12

i	q_i	$\frac{1}{2}\Delta q_{i-1}$	$\frac{5}{12}\Delta^2 q_{i-2}$	$\frac{3}{8}\Delta^3 q_{i-3}$	Δy_i
3	0,5146	0,0293	0,0054	0,0011	0,5504
4	0,5897	0,0376	0,0069	0,0014	0,6356
5	0,6861	0,0482	0,0089	0,0018	0,7450

Tabla 9.13

i	x_i	y_i	Δy_i	p_i	Δp_i	$\Delta^2 p_i$	$\Delta^3 p_i$
0	0	1,0000		0,0000	0,0000	0,0004	0,0006
1	0,1	1,0000		0,0000	0,0004	0,0010	0,0010
2	0,2	1,0000		0,0004	0,0014	0,0020	0,0004
3	0,3	1,0004	0,0032	0,0018	0,0034	0,0024	
4	0,4	1,0036	0,0081	0,0052	0,0058		
5	0,5	1,0117	0,0150	0,0110			
6	0,6	1,0267					

i	x_i	z_i	Δz_i	g_i	Δg_i	$\Delta^2 g_i$	$\Delta^3 g_i$
0	0	1,0000		0,0000	0,0200	0,0004	0,0006
1	0,1	1,0000		0,0200	0,0204	0,0010	0,0013
2	0,2	1,0200		0,0404	0,0214	0,0023	0,0007
2	0,3	1,0604	0,0732	0,0618	0,0237	0,0030	
4	0,4	1,1336	0,0984	0,0855	0,0267		
5	0,5	1,2323	0,1271	0,1122			
6	0,6	1,3594					

Supongamos que se da el sistema de dos ecuaciones

$$\begin{cases} y' = f_1(x, y, z), \\ z' = f_2(x, y, z). \end{cases} \tag{9}$$

Entonces para este sistema las fórmulas de extrapolación de Adams se escriben así:

$$\begin{aligned} \Delta y_i &= p_i + \frac{1}{2}\Delta p_{i-1} + \frac{5}{12}\Delta^2 p_{i-2} + \frac{3}{8}\Delta^3 p_{i-3}, \\ \Delta z_i &= g_i + \frac{1}{2}\Delta g_{i-1} + \frac{5}{12}\Delta^2 g_{i-2} + \frac{3}{8}\Delta^3 g_{i-3}, \end{aligned} \tag{10}$$

donde

$$p_i = hy_i' = hf_1(x_i, y_i, z_i), \quad g_i = hz_i' = hf_2(x_i, y_i, z_i)$$

y

$$y_{i+1} = y_i + \Delta y_i, \quad z_{i+1} = z_i + \Delta z_i.$$

Ejemplo 2. Aplicando el método de Adams, resolver numéricamente el sistema de ecuaciones diferenciales

$$\begin{cases} y' = (z - y)\,x, \\ z' = (z + y)\,x \end{cases}$$

para las condiciones iniciales $y\,(0) = 1{,}000$; $z\,(0) = 1{,}000$ en el segmento $[0;\ 0{,}6]$; el paso $h = 0{,}1$.

△ Tomemos el segmento inicial de la solución de la tabla 9.3 (antes hemos resuelto este sistema con ayuda del método de Euler). Buscaremos los valores de las funciones $y\,(x)$ y $z\,(x)$ para $x_4 = 0{,}4$; $x_5 = 0{,}5$ y $x_6 = 0{,}6$, haciendo uso de las fórmulas (10) y designando $f_1\,(x,\ y,\ z) = y' = (x - y)\,x$ y $f_2\,(x,\ y,\ z) = z' = (z + y)\,x$. Los cálculos se dan en las tablas 9.13, 9.14 y 9.15 (las tablas 9.14 y 9.15 son auxiliares).

La tabla 9.14 se destina a determinar los segundos miembros del sistema dado y hallar p_i y g_i; la tabla 9.15 sirve para determinar Δy_i y Δz_i con ayuda de las diferencias de las magnitudes p y g, obtenidas en la tabla 9.13. ▲

Tabla 9.14

i	x_i	y_i	z_i	y'_i	p_i	z'_i	g_i
0	0,0	1,0000	1,0000	0,0000	0,0000	0,0000	0,0000
1	0,1	1,0000	1,0000	0,0000	0,0000	0,2000	0,0200
2	0,2	1,0000	1,0200	0,0040	0,0004	0,4040	0,0404
3	0,3	1,0004	1,0604	0,0180	0,0018	0,6182	0,0618
4	0,4	1,0036	1,1336	0,0520	0,0052	0,8549	0,0855
5	0,5	1,0117	1,2323	0,1103	0,0110	1,1220	0,1122

Tabla 9.15

i	p_i	$\frac{1}{2}\Delta p_{i-1}$	$\frac{5}{12}\Delta^2 p_{i-2}$	$\frac{3}{8}\Delta^3 p_{i-3}$	Δy_i
3	0,0018	0,0007	0,000425	0,000225	0,0032
4	0,0052	0,0017	0,000850	0,000375	0,0081
5	0,0110	0,0029	0,0010	0,00015	0,0150

i	g_i	$\frac{1}{2}\Delta g_{i-1}$	$\frac{5}{12}\Delta^2 g_{i-2}$	$\frac{3}{8}\Delta^3{}_{i-3}$	Δz_i
3	0,0618	0,0107	0,000425	0,000225	0,0732
4	0,0855	0,0118	0,000958	0,000488	0,0987
5	0,1122	0,0134	0,00125	0,00026	1,1271

§ 9.8. Método de Milne

Al igual que el método de Runge—Kutta, el de **Milne** es un método de exactitud elevada.

Supongamos que en el segmento $[a, b]$ se necesita hallar la solución numérica de la ecuación diferencial

$$y' = f(x, y) \tag{1}$$

con la condición inicial

$$y(x_0) = y_0. \tag{2}$$

Dividamos el segmento $[a, b]$ en n partes iguales por los puntos $x_i = x_0 + ih$ $(i = 0, 1, \ldots, n)$, donde $h = (b - a)/n$ es el paso de integración.

Utilizando los datos iniciales, encontramos por un método cualquiera los valores sucesivos $y_1 = y(x_1)$, $y_2 = y(x_2)$, $y_3 = y(x_3)$ de la función buscada $y(x)$. Así pues, llegan a ser conocidas y'_i $(i = 0, 1, 2, 3)$.

Las aproximaciones $\overline{y}_1$ e $\overline{\overline{y}}_j$ para los siguientes valores y_1 $(i = 4, 5, \ldots, n)$ se determinan sucesivamente con ayuda de las *fórmulas de Milne*:

$$\overline{y}_i = y_{i-4} + \frac{4h}{3}(2y'_{i-3} - y'_{i-2} + 2y'_{i-1}), \tag{3}$$

$$\overline{\overline{y}}_i = y_{i-2} + \frac{h}{3}(\overline{y}'_i + 4y'_{i-1} + y'_{i-2}), \tag{4}$$

$$\text{donde } \overline{y}'_i = f(x_i, \overline{y}_i).$$

Se puede mostrar que el error absoluto del valor $\overline{\overline{y}}_i$ es, aproximadamente, igual a

$$\varepsilon_i = \frac{1}{29}\left|\overline{\overline{y}}_i - \overline{y}_i\right|. \tag{5}$$

Por eso, si $\varepsilon_i \leqslant \varepsilon$, donde ε es el error límite prefijado de la solución, se puede poner $y_i \approx \overline{\overline{y}}_i$, e $y'_i = f(x_i, \overline{\overline{y}}_i)$. Esto tiene lugar si $\overline{y}_i$ e $\overline{\overline{y}}_i$ coinciden en las cifras decimales que nos interesan. Si la condición (5) se cumple, pasamos a calcular el siguiente valor y_{i+1}, repitiendo el proceso. En el caso contrario el paso h, comenzando con el lugar conocido, se disminuye y el segmento inicial respectivo se recalcula. Al igual que en el método de Runge—Kutta, la magnitud del paso inicial se obtiene de la desigualdad $h^4 < \varepsilon$.

Con el fin de deducir las fórmulas de Milne (3) y (4) utilicemos la primera fórmula de interpolación de Newton para la derivada y' en cierto punto escogido x_k. En este caso nos limitamos por las diferencias de tercer orden; esto es equivalente al hecho de que la integral $y = y(x)$ de la ecuación diferencial (1) se aproxima por un polinomio de cuarto grado. Tenemos

$$y' = y'_k + q\Delta y'_k + \frac{q(q-1)}{2!}\Delta^2 y'_k + \frac{q(q-1)(q-2)}{3!}\Delta^3 y'_k. \qquad (6)$$

Abriendo paréntesis, encontramos

$$y' = y'_k + q\Delta y'_k + \frac{1}{2}(q^2-q)\Delta^2 y'_k + \frac{1}{6}(q^3-3q^2+2q)\Delta^3 y'_k, \qquad (7)$$

donde $q = (x - x_k)/h$.

Suponiendo en la relación (7) $k = i-4$, obtenemos

$$y' = y'_{i-4} + q\Delta y'_{i-4} + \frac{1}{2}(q^2-q)\Delta^2 y'_{i-4} + \\ + \frac{1}{6}(q^3-3q^2+2q)\Delta^3 y'_{i-4}. \qquad (8)$$

Vamos a integrar la igualdad (8) respecto a x entre x_{i-4} y x_i:

$$\int\limits_{x_{i-4}}^{x_i} y'\,dx = \int\limits_{x_{i-4}}^{x_i} \Big[y'_{i-4} + q\Delta y'_{i-4} + \frac{q^2-q}{2}\Delta^2 y'_{i-4} + \\ + \frac{q^3-3q^2+2q}{6}\Delta^3 y'_{i-4}\Big]\,dx.$$

Teniendo en cuenta que $q = (x - x_{i-4})/h$ y $dx = h\,dq$, tenemos

$$y_i - y_{i-4} = h\,\Big\{y'_{i-4}\int\limits_0^4 dq + \Delta y'_{i-4}\int\limits_0^4 q\,dq + \\ + \Delta^2 y'_{i-4}\int\limits_0^4 \frac{q^2-q}{2}\,dq + \Delta^3 y'_{i-4}\int\limits_0^4 \frac{q^3-3q^2+2q}{6}\,dq\Big\} = \\ = h\left(4y'_{i-4} + 8\Delta y'_{i-4} + \frac{20}{3}\Delta^2 y'_{i+4} + \frac{8}{3}\Delta^3 y'_{i-4}\right). \qquad (9)$$

Puesto que

$$\Delta y'_{i-4} = y'_{i-3} - y'_{i-4},$$
$$\Delta^2 y''_{i-4} = y'_{i-2} - 2y'_{i-3} + y'_{i-4},$$
$$\Delta^3 y'_{i-4} = y'_{i-1} - 3y'_{i-2} + 3y'_{i-3} - y'_{i-4},$$

entonces, sustituyendo estas expresiones en la igualdad (9), obtenemos la primera fórmula de Milne (3):

$$\overline{y}_i = y_{i-4} + \frac{4h}{3}(2y'_{i-3} - y'_{i-2} + 2y'_{i-1}).$$

Para deducir la segunda fórmula de Milne (4) pongamos en la relación (7) $k = i - 2$:

$$y' = y'_{i-2} + q\Delta y'_{i-2} + \frac{1}{2}(q^2 - q)\,\Delta^2 y'_{i-2} +$$
$$+ \frac{1}{6}(q^3 - 3q^2 + 2q)\,\Delta^3 y'_{i-2}, \qquad (10)$$

donde $q = (x - x_{i-2})/h$ y $\mathrm{d}x = h\,\mathrm{d}q$.

Vamos a integrar la igualdad (10) respecto a x entre x_{i-2} y x_i:

$$\int_{x_{i-2}}^{x_i} y'\,\mathrm{d}x = h\int_0^2 \Big[y'_{i-2} + q\Delta y'_{i-2} + \frac{1}{2}(q^2 - q)\,\Delta^2 y'_{i-2} +$$
$$+ \frac{1}{6}(q^3 - 3q^2 + 2q)\,\Delta^3 y'_{i-2}\Big]\,\mathrm{d}q.$$

De aquí tenemos

$$y_i - y_{i-2} = h\left(2y'_{i-2} + 2\Delta y'_{i-2} + \frac{1}{3}\Delta^2 y'_{i-2}\right). \qquad (11)$$

Tomando en consideración que $\Delta y'_{i-2} = y'_{i-1} - y'_{i-2}$ y $\Delta^2 y_{i-2} = y'_i - 2y'_{i-1} + y'_{i-2}$, obtenemos la segunda fórmula de Milne (4):

$$\bar{\bar{y}}_i = y_{i-2} + \frac{h}{3}(y'_i \pm 4y'_{i-1} + y'_{i-2}).$$

Ejemplo 1. Se da la ecuación diferencial $y' = y - x$ que satisface la condición inicial $x_0 = 0$, $y(x_0) = 1{,}5$. Calcular con exactitud de hasta 0,01 el valor de la solución de esta ecuación para $x = 1{,}5$. El cálculo ha de realizarse por el método combinado de Runge—Kutta y de Milne con dos cifras de reserva.

△ Vamos a elegir el paso inicial del cálculo. De la condición de $h^4 < 0{,}01$ obtenemos $h = 0{,}25$. Entonces todo el segmento de integración se divide en seis partes iguales por los puntos $x_0 = 0$; $x_1 = 0{,}25$; $x_2 = 0{,}50$; $x_3 = 0{,}75$; $x_4 = 1{,}00$; $x_5 = 1{,}25$; $x_6 = 1{,}50$. Tomemos el segmento inicial y_0, y_1, y_2, y_3 de la solución del ejemplo 1 dado en el § 9.6. Para resolver esta ecuación hagamos la tabla 9.16.

Así pues, obtenemos la respuesta: $y = (1{,}5) = 4{,}74$. ▲

Al resolver el sistema de ecuaciones diferenciales

$$\begin{cases} y' = f(x,\ y,\ z) \\ z' = \varphi(x,\ y,\ z) \end{cases}$$

con las condiciones iniciales $y(x_0) = y_0$; $z(x_0) = z_0$ las fórmulas de Milne se escriben por separado para las funciones $y(x)$ y $z(x)$. El orden de los cálculos queda el de antes.

Ejemplo 2. Se da el sistema de ecuaciones

$$\begin{cases} y' = \cos(y + 1{,}1z) + 1, \\ z' = \dfrac{1}{x + 2{,}1y^2} + x + 1 \end{cases}$$

Tabla 9.16

i	x_i	y_i	$y_i' = f(x_i, y_i) = y_i - x_i$	$\bar{y}_i$	$\bar{y}' = f(x_i, \bar{y}_i) = \bar{y}_i - x_i$	$\bar{\bar{y}}_i$	ε_i	y_i	$y_i' = f(x_i, y_i) = y_i - x_i$	Revisión del paso de cálculo según los resultados de la fórmula (5)
0	0	1,5000	1,5000							
1	0,25	1,8920	1,6420							
2	0,50	2,3243	1,8243							
3	0,75	2,8084	2,0584							
4	1,00			3,3588	2,3588	3,3590	$7\cdot10^{-5}$	3,3590	2,3590	No se nesecita
5	1,25			3,9947	2,7447	3,9950	10^{-5}	3,9950	2,7450	»
9	1,50			4,7402	3,2402	4,7406	$1,4\cdot10^{-5}$	4,7406		»

para las condiciones iniciales $y(0) = 3{,}14159$, $z(0) = 0$. Haciendo uso del método de Milne, con exactitud de hasta $\varepsilon = 0{,}0001$ hallar la solución de este sistema en el segmento $[0,\ 0{,}5]$, tomando el paso $h = 0{,}1$ y considerando conocidos los valores de las funciones $y(x)$ y $z(x)$ para $x_1 = 0{,}1$; $x_2 = 0{,}2$; $x_3 = 0{,}3$. Estos valores son tales: $y(0,1) = 3{,}14184$; $y(0,2) = 3{,}14364$; $y(0,3) = 3{,}14903$; $z(0,1) = 0{,}10981$; $z(0,2) = 0{,}22960$ $z(0,3) = 0{,}35934$.

△ Buscaremos los valores de las funciones $y(0{,}4)$, $y(0{,}5)$ y $z(0,4)$, $z(0{,}5)$ con ayuda de las fórmulas (3) a (5). Realicemos los cálculos utilizando dos tablas: una principal (tabla 9.17) y otra auxiliar (tabla 9.18).

Tabla 9.17

(1)	(2)	(3)	(4)	(5)	(6)
i	x_i	$\bar{Y}_i = (y_i, z_i)$		$\bar{Y}_i'$	
		$\bar{y}_i$	$\bar{z}_i$	$\bar{y}_i'$	$\bar{z}_i'$
0	0,0				
1	0,1				
2	0,2				
3	0,3				
4	0,4	3,16057	0,49906	0,15698	1,44678
5	0,5	3,18164	0,64870	0,27079	1,54596

(7)	(8)	(9)	(10)	(11)	(12)
$\overline{\overline{Y}}_i = Y_i$		$\overline{\overline{Y}}'_i = Y'_i$		ε_i	
$\overline{\overline{y}}_i$	$\overline{\overline{z}}_i$	$\overline{\overline{y}}'_i = y'_i$	$\overline{\overline{z}}'_i = z'_i$	$\varepsilon_{i,1}$	$\varepsilon_{i,2}$
3,14159	0,00000	0,00000	1,04824		
3,14184	0,10981	0,00732	1,14801		
3,14364	0,22960	0,03224	1,24773		
3,14903	0,35934	0,08001	1,34734		
3,16062	0,49905	0,15701	1,44678	$1,7 \cdot 10^{-6}$	0
3,18166	0,64869			$7 \cdot 10^{-7}$	$4 \cdot 10^{-7}$

Tabla 9.18

	4		5	
	y_i	z_i	y_i	z_i
$2Y'_{i-3}$	0,01464	2,29602	0,06448	2,49546
$-Y'_{i-2}$	−0,30224	−1,24773	−0,08001	−1,34734
$2Y'_{i-1}$	0,16002	2,69468	0,31402	2,89356
$\Sigma_i^{(1)}$	0,14242	3,74297	0,29849	4,04168
$\frac{4}{3} h\Sigma_i^{(1)}$	0,01898	0,49906	0,03980	0,53889
Y_{i-4}	3,14159	0,00000	0,14184	0,10981
$\overline{Y}_i$	3,16057	0,49906	3,18164	0,64870
Y'_{i-2}	0,03224	1,24773	0,08001	1,34734
$4Y'_{i-1}$	0,32004	5,38936	0,62804	5,7871
Y'_i	0,15698	1,44678	0,27079	1,54596
$\Sigma_i^{(2)}$	0,50926	8,08387	0,97884	8,68042
$\frac{h}{3} \Sigma_i^{(2)}$	0,01698	0,26945	0,03263	0,28935
Y_{i-1}	3,14364	0,22960	3,14903	0,35934
$\overline{\overline{Y}}_i$	3,16062	0,49905	3,18166	0,64869

Apuntamos el segmento inicial de la solución en las columnas (7) y (8) de la tabla 9.17 y encontramos los valores y_i' y z_i' [columnas (9) y (10)]. Luego, utilizando los datos obtenidos y la fórmula (3), hallamos $\bar{y}_4$ y $\bar{z}_4$ (tabla 9.18).

Transponemos los valores $\bar{y}_4$ y $\bar{z}_4$ a la tabla principal [columnas (3) y (4)]. Determinamos $\bar{y}_4'$ y $\bar{z}_4'$ [columnas (5) y (6) y la tabla 9.17] y luego, volviendo a utilizar la tabla auxiliar, encontramos ya $\bar{\bar{y}}_4$ y $\bar{\bar{z}}_4$. Sus valores los pasamos a las columnas (7) y (8) de la tabla 9.17. Hallamos

$$\varepsilon_{4,1}=\frac{1}{29}(\bar{\bar{y}}_4-\bar{y}_4), \quad \varepsilon_{4,2}=\frac{1}{29}(\bar{\bar{z}}_4-\bar{z}_4)$$

[las columnas (11) y (12) de la tabla 9.17]. Para el error admisible prefijado $\varepsilon = 0{,}0001$ vemos que no se necesita volver a calcular el segmento inicial y por y_4 se puede tomar $\bar{\bar{y}}_4$. Para encontrar $y\,(0{,}5)$ y $z\,(0{,}5)$ repetimos todo el proceso de cálculos.

Finalmente obtenemos: $y\,(0{,}4) = 3{,}16062$, $z\,(0{,}4) = 0{,}49905$, $y\,(0{,}5) = 3{,}18166$, $z\,(0.5) = 0{,}64869$. ▲

§ 9.9. Concepto de problemas de contorno para ecuaciones diferenciales ordinarias

Consideremos la resolución de un problema de contorno para ecuaciones diferenciales ordinarias, citando como ejemplo una ecuación de segundo orden

$$F\,(x, y, y', y'') = 0. \tag{1}$$

Un problema de contorno elemental de dos puntos para la ecuación (1) se plantea del modo siguiente: se necesita hallar una función $y = y\,(x)$ que dentro del segmento $[a, b]$ satisfaga la ecuación (1) y en los extremos del segmento, las condiciones de contorno

$$\begin{aligned} \varphi_1\,[y\,(a), y'\,(a)] &= 0, \\ \varphi_2\,[y\,(b), y'\,(b)] &= 0. \end{aligned} \tag{2}$$

Examinemos algunos tipos de un problema de contorno de dos puntos para la ecuación (1).

Supongamos, por ejemplo, que se da la ecuación diferencial de segundo orden

$$y'' = f\,(x, y, y') \tag{3}$$

con las condiciones de contorno $y\,(a) = A$, $y\,(b) = B$ $(a < b)$ o sea, se conocen los valores de la función buscada $y = y\,(x)$ en los puntos de frontera $x = a$ y $x = b$. Entonces la solución de la ecuación (3) representa geométricamente la curva integral $y = y\,(x)$ que pasa por los puntos dados $M\,(a;\, A)$ y $N\,(b;\, B)$ (fig. 9.4).

Supongamos ahora que para la ecuación (3) se conocen los valores de las derivadas de la función buscada a en los puntos de frontera, o sea, $y'(a) = A_1$, $y'(b) = B_1$. Entonces la solución de la ecuación (3) significa geométricamente que es necesario hallar tal curva integral $y = y(x)$ de esta ecuación que corte a las rectas $x = a$ y $x = b$ bajo los ángulos $\alpha = \operatorname{arctg} A_1$ y $\beta = \operatorname{arctg} B_1$, respectivamente, (fig. 9.5).

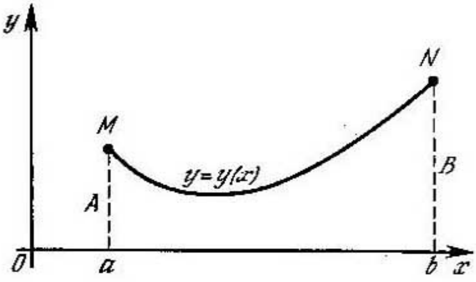


Fig. 9.4

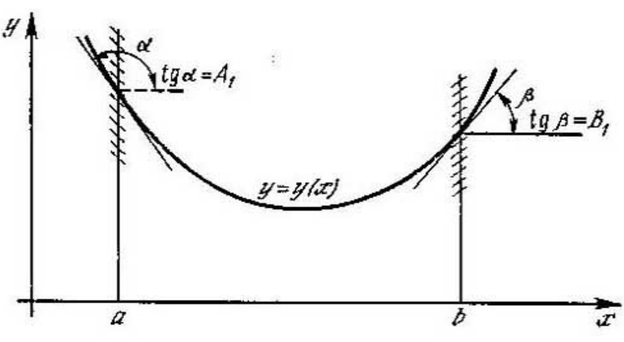


Fig. 9.5

Supongamos, por último, que para la ecuación (3) en un punto de frontera se conoce el valor de la función buscada $y(a) = A$ y en otro, el valor de la derivada de esta función $y'(b) = B_1$. Tal problema de contorno se llama *mixto*. La solución de la ecuación (3) significa geométricamente que se necesita hallar una curva integral $y = y(x)$ de esta ecuación que pase por el punto $M(a; A)$ y corte la recta $x = b$ bajo el ángulo $\beta = \operatorname{arc} B_1$ (fig. 9.6).

Si la ecuación diferencial y las condiciones de contorno son lineales, tal problema se dice *lineal*. En este caso la ecuación diferencial (1) y las condiciones de contorno (2) se escriben así:

$$y'' + p(x)\,y' + q(x)\,y = f(x); \tag{4}$$

$$\left\{\begin{aligned} \alpha_0 y(a) + \alpha_1 y'(a) &= \gamma_1, \\ \beta_0 y(b) + \beta_1 y'(b) &= \gamma_2, \end{aligned}\right. \tag{5}$$

donde $p(x)$, $q(x)$ y $f(x)$ son las funciones continuas conocidas sobre el segmento $[a, b]$; α_0, α_1, β_0, β_1, γ_1, γ_2, las constantes dadas con la particularidad de que $|\alpha_0| + |\alpha_1| \neq 0$, $|\beta_0| + |\beta_1| \neq 0$.

Si $f(x) = 0$ para $a \leqslant x \leqslant a$, la ecuación se denomina *homogénea* y en el caso contrario, *heterogénea*.

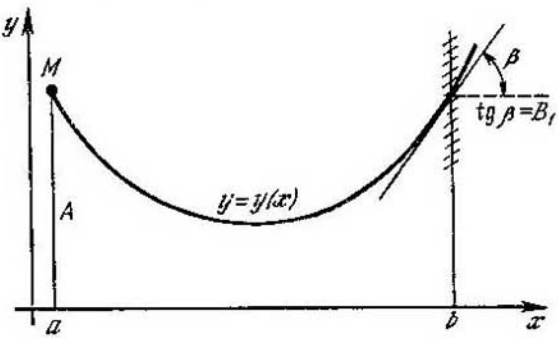


Fig. 9.6

Si $\gamma_1 = 0$ y $\gamma_2 = 0$, la condición de contorno respectiva se llama *homogénea*. Si la ecuación diferencial y las condiciones de contorno son homogéneas, el problema de contorno se dice *homogéneo*.

§ 9.10. Método de diferencias finitas para las ecuaciones diferenciales lineales de segundo orden

Supongamos que se da la ecuación diferencial lineal de segundo orden

$$y'' + p(x)\,y' + q(x)\,y = f(x) \tag{1}$$

con las condiciones de contorno lineales de dos puntos

$$\begin{cases} \alpha_0 y(a) + \alpha_1 y'(a) = A, \\ \beta_0 y(b) + \beta_1 y'(b) = B \end{cases} \tag{2}$$

$$(|\alpha_0| + |\alpha_1| \neq 0; \ |\beta_0| + |\beta_1| \neq 0)$$

y $p(x)$, $q(x)$ y $f(x)$ son continuas sobre el segmento $[a, b]$. Luego, supongamos que $x_0 = a$, $x_n = b$, $x_i = x_0 + ih$ $(i = 1, 2, \ldots$ $\ldots, n-1)$ son los sistemas de nodos equidistantes con cierto paso $h = (b-a)/h$ y $p_i = p(x_i)$ $q_i = q(x_i)$, $f_i = f(x_i)$.

Designemos los valores aproximados de la función buscada $y(x)$ y de sus derivadas $y'(x)$, $y''(x)$ en los puntos x_i, obtenidos como resultado del cálculo, por y_i, y_i' e y_i'', respectivamente. En cada nodo interior señalemos, aproximadamente, las derivadas $y'(x_i)$ e $y''(x_i)$ por las relaciones de diferencia finita

$$y_i' = \frac{y_{i+1} - y_i}{h}, \quad y_i'' = \frac{y_{i+2} - 2y_{i+1} + y_i}{h^2} \tag{3}$$

y para los puntos extremos $x_0=a$ y $x_n=b$ pongamos

$$y'_0=\frac{y_1-y_0}{h},\quad y_n=\frac{y_n-y_{n-1}}{h}. \tag{4}$$

Utilizando las fórmulas (3) y (4), reemplacemos aproximadamente la ecuación (1) las condiciones de contorno (2) por el sistema de ecuaciones

$$\begin{cases} \dfrac{y_{i+2}-2y_{i+1}+y_i}{h^2}+p_i\,\dfrac{y_{i+1}-y_i}{h}+q_iy_i=f_i, \\ \alpha_0y_0+\alpha_1\,\dfrac{y_1-y_0}{h}=A;\quad \beta_0y_n+\beta_1\,\dfrac{y_n-y_{n-1}}{h}=B. \end{cases} \tag{5}$$

Así pues, llegamos al sistema algebraico de $n+1$ ecuaciones con $n+1$ incógnitas. Como resultado de la resolución de tal sistema obtenemos una tabla de valores aproximados de la función buscada.

Si se reemplazan $y'(x_i)$ e $y''(x_i)$ por las relaciones de diferencia central:

$$y'_i=\frac{y_{i+1}-y_{i-1}}{2h},\quad y''_i=\frac{y_{i+1}-2y_i+y_{i-1}}{h^2}, \tag{6}$$

se puede obtener fórmulas más exactas. Sin embargo, para las derivadas en los puntos extremos pueden ser utilizadas las fórmulas (5). Entonces obtenemos el sistema

$$\begin{cases} \dfrac{y_{i+1}-2y_i+y_{i-1}}{h^2}+p_i\,\dfrac{y_{i+1}-y_{i-1}}{2h}+q_iy_i=f_i, \\ \alpha_0y_0+\alpha_1\,\dfrac{y_1-y_0}{h}=A;\quad \beta_0y_n+\beta_1\,\dfrac{y_n-y_{n-1}}{h}=B. \end{cases} \tag{7}$$

Ejemplo. Haciendo uso del método de diferencias finitas, con exactitud de hasta 0,001 hallar la solución del problema de contorno

$$\begin{cases} y''-xy'+2y=x+1, \\ y(0{,}9)-0{,}5y'(0{,}9)=2, \\ y(1{,}2)=1. \end{cases}$$

△ Dividamos el segmento $[0{,}9;\ 1{,}2]$ en partes con un paso de $h=0{,}1$. Entonces obtenemos cuatro puntos nodales con las abscisas $x_0=0{,}9$; $x_1=1{,}0$; $x_2=1{,}1$; $x_3=1{,}2$. En los puntos interiores $x_1=1{,}0$; $x_2=1{,}1$ reemplacemos la ecuación dada por la de diferencias finitas

$$\frac{y_{i+1}-2y_i+y_{i-1}}{h^2}-x_i\,\frac{y_{i+1}-y_{i-1}}{2h}+2y_i=x_{i+1}\quad (i=1,\ 2). \tag{*}$$

Utilizando las condiciones de contorno, escribamos las ecuaciones de diferencias finitas en los puntos finales

$$\begin{cases} y_0-0{,}5\,\dfrac{y_1-y_0}{h}=2, \\ y_3=1. \end{cases} \tag{**}$$

Reduciendo los términos semejantes y teniendo en cuenta que $h = 0,1$, rescribimos las ecuaciones (*) y (**), respectivamente, en la forma

$$y_{i-1}(2 + 0,1x_i) - 4y_i(1 - 0,01) + y_{i+1}(2 - 0,1x_i) = \\ = 0,02(x_i + 1),$$
$$1,2y_0 - y_1 = 0,4; \; y_3 = 1.$$

El problema dado se reduce a la resolución del sistema de ecuaciones

$$\begin{cases} 1,2y_0 - y_1 = 0,4, \\ 2,1y_0 - 3,96y_1 + 1,9y_2 = 0,04, \\ 1,11y_1 - 3,96y_2 + 1,89y_3 = 0,042. \end{cases}$$

Resolviendo este sistema, obtenemos $y_0 = 1,406$; $y_1 = 1,287$; $y_2 = 1,149$; $y_3 = 1,000$. ▲

Ejercicios

1. Utilizando el método de Picard, hallar tres aproximaciones sucesivas de la solución de la ecuación diferencial:
a) $y' = 4y(1 + x)$; la condición inicial $y(0) = 1$,
b) $y' = x - y$; la condición inicial $y(0) = 1$.

2. Encontrar los primeros siete términos del desarrollo en serie de potencias de la solución $y = y(x)$ de la ecuación $y'' + 0,1(y')^2 + (1 + 0,1x)y = 0$ para las condiciones iniciales $y(0) = 1$; $y'(0) = 2$.

3. Hallar la solución de la ecuación diferencial $y' = x^2 + y^2$, que satisface la condición inicial $x_0 = 0$, $y(x_0) = 0$. Limitarse a los términos del desarrollo en serie de potencias que contienen x^7.

4. Suponiendo $h = 0,1$, con ayuda del método de Euler resolver las ecuaciones diferenciales para las condiciones iniciales dadas en los segmentos indicados:
a) $y' = y + 3x$; $y(0) = -1$; $x \in [0, 0,5]$;
b) $y' = x - 2y$; $y(0) = 0$; $x \in [0, 1]$.

5. Aplicando el método perfeccionado de Euler, encontrar en el segmento $[0, 1]$ la tabla de solución de la ecuación diferencial $y' = y - \frac{2x}{y}$ para la condición inicial $y(0) = 1$; tomar $h = 0,2$.

6. Aplicando el método perfeccionado de Euler-Cauchy, resolver la ecuación diferencial dada en el ejercicio 5.

7. Con ayuda del método de Runge-Kutta, tomando $h = 0,1$, encontrar las soluciones de las ecuaciones diferenciales para las condiciones iniciales dadas en los segmentos indicados:
a) $y' = x + y^2$; $y(1) = 0$; $x \in [1,2]$;
b) $y' = x^2 - y$; $y(0) = 2$; $x \in [0, 1]$.

8. Aplicando el método de extrapolación de Adams, resolver la ecuación diferencial $y' = 2x - y$ para la condición inicial $y(0) = 1$ en el segmento $[0, 1]$. El segmento inicial de solución está dado: $y_0 = 1$, $y_1 = 0,9145$, $y_2 = 0,8562$, $y_3 = 0,8225$ (tomar $h = 0,1$).

CAPÍTULO X

Métodos aproximados de solución de las ecuaciones diferenciales en derivadas parciales

§ 10.1. Clasificación de las ecuaciones diferenciales de segundo orden

Son muchos los problemas, prácticamente importantes, de hidrodinámica, transferencia de masa y de calor, conducción del calor, difusión, teoría de elasticidad y de otros campos del saber, que se describen por las ecuaciones diferenciales lineales en derivadas parciales de segundo orden. Entre tales ecuaciones admiten una interpretación física más simple y clara las *ecuaciones con dos variables independientes* *):

$$a_{11} u_{xx} + 2a_{12}u_{xy} + a_{22}u_{yy} + b_1u_x + b_2u_y + cu = F, \qquad (1)$$

donde $u(x, y)$ es la función incógnita a determinar; a_{11}, a_{12}, a_{22}, b_1, b_2, c son las funciones dadas de las variables independientes x e y que se llaman *coeficientes* de la ecuación y F, la función dada de x e y, llamada *segundo miembro de la ecuación*. Si los coeficientes de la ecuación (1) son constantes, ésta se denomina *ecuación lineal de coeficientes constantes*. La ecuación (1) se dice homogénea si $F = 0$.

Se llama *solución* (o *integral*) de la ecuación (1) a toda función que, siendo sustituida en vez de u en la ecuación, la convierte en identidad.

Las condiciones necesarias para la existencia y unicidad de la solución de la ecuación (1) dependen considerablemente de los coeficientes a_{11}, a_{12}, a_{22}. Vamos a suponer que al menos uno de estos coeficientes es distinto del cero idéntico (en el caso contrario tendríamos la ecuación de primer orden). Resulta que las propiedades de la solución de la ecuación (1) se determinan en gran medida por la magnitud (más exactamente, por el signo) del *discriminante* $\Delta = a_{12}^2 - a_{11}a_{22}$.

Teniendo en cuenta la diferencia de las propiedades de la solución, y, por consiguiente, también la diferencia de los métodos de resolución ha sido adoptada la siguiente clasificación de las ecuaciones. Supongamos que en cierto dominio D el discriminante conserva el signo o por doquier es igual a cero.

---

*) Aquí y a continuación se utilizan las siguientes designaciones para las derivadas: $u_x = \dfrac{\partial u}{\partial x}$, $u_{xx} = \dfrac{\partial^2 u}{\partial x^2}$, $u_{xy} = \dfrac{\partial^2 u}{\partial x\, \partial y}$, etc.

La ecuación (1) se llama ecuación de *tipo elíptico* si $\Delta < 0$, ecuación de *tipo parabólico*, si $\Delta = 0$ y ecuación de *tipo hiperbólico* si $\Delta > 0$.

Si el signo del discriminante cambia al pasar de un punto del dominio D a otro, la ecuación se denomina *ecuación del tipo mixto*.

En su esencia la clasificación citada se debe al hecho de que las ecuaciones de los tipos elíptico, parabólico e hiperbólico describen problemas muy distintos por su sentido físico. Estos problemas se determinan por fenómenos físicos que son diferentes por su naturaleza. Así, las ecuaciones de conducción del calor y de difusión (de tipo parabólico) expresan las leyes de conservación de la energía y la materia. Estas ecuaciones se construyen basándose en las leyes de Fourier y de Nernst que son iguales desde el punto de vista de la formulación matemática.

Por otro lado, la ecuación de vibración de una cuerda (de tipo hiperbólico) representa la ley de conservación del impulso y se funda en la segunda ley de Newton.

Por último, la ecuación del tipo elíptico define la función de un género completamente distinto, o sea, la función que representa los procesos estacionarios que no varían en el tiempo.

Del curso de física matemática se sabe que para ciertas condiciones que se sobreponen a los coeficientes a_{11}, a_{12}, a_{22} (por ejemplo, si son derivables continuamente dos veces), existe la tranformación de las variables

$$\xi = \varphi(x, y), \ \eta = \psi(x, y) \tag{2}$$

que reducen la ecuación (1) a una de las formas canónicas siguientes:

$$u_{\xi\xi} + u_{\eta\eta} = f \text{ (ecuación de tipo elíptico);} \tag{3}$$

$$u_{\xi\xi} = f \text{ (ecuación de tipo parabólico);} \tag{4}$$

$$u_{\xi\eta} = f \text{ o bien } u_{\xi\xi} - u_{\eta\eta} = f \text{ (ecuación de tipo hiperbólico).} \tag{5}$$

Aquí $f = f(\xi, \eta, u, u_\xi, u_\eta)$ es la función de las variables independientes, de la función incógnita y de sus primeras derivadas. Nótese que la ecuación de tipo hiperbólico tiene dos formas canónicas equivalentes.

Para la ecuación (1) de coeficientes constantes la transformación de las variables (2) es lineal y tiene una forma simple. Citemos estas transformaciones para cada tipo de ecuaciones.

Para las ecuaciones de tipo elíptico::

$$\xi = y - \frac{a_{12}}{a_{11}} x; \quad \eta = \frac{\sqrt{a_{11}a_{22} - a_{12}^2}}{a_{11}} x. \tag{6}$$

Para la ecuaciones de tipo parabólico:

$$\xi = y - \frac{a_{12}}{a_{11}} x; \quad \eta = x. \tag{7}$$

Para las ecuaciones de tipo hiperbólico:

$$\xi = y - \frac{a_{12}}{a_{11}} x; \quad \eta = -\frac{\sqrt{a_{12}^2 - a_{11}a_{22}}}{a_{11}} x. \tag{8}$$

Nótese que para las ecuaciones de tipo parabólico η, hablando en general, puede ser función arbitraria que no depende de ξ.

Ejemplo. Transformar la ecuación $u_{xx} + u_{xy} = F$, reduciéndola a la forma canónica.

$\triangle$ Aquí $a_{11} = 1$; $a_{12} = 0{,}5$; $a_{22} = 0$; por eso el discriminante $\Delta = 0{,}5^2 - 1 \cdot 0 = 0{,}25 > 0$. Por consiguiente, esta ecuación es de tipo hiperbóbico. Utilizando la transformación correspondiente de las variables (8), encontramos

$$u_{xx} = \frac{1}{4}(u_{\xi\xi} + 2u_{\xi\eta} + u_{\eta\eta}), \quad u_{xy} = \frac{1}{2}(u_{\xi\xi} + u_{\eta\eta}).$$

Ahora bien, la forma canónica de la ecuación inicial se escribe así:

$$u_{\xi\xi} - u_{\eta\eta} = -4\overline{F}(\xi, \eta),$$

donde

$$\overline{F}(\xi, \eta) = F(-2\eta, \xi - \eta).$$

Hemos obtenido la segunda forma canónica.

Se puede mostrar que la transformación de las variables $\alpha = \xi + \eta$, $\beta = \xi - \eta$ lleva a la primera forma canónica; $u_{\alpha\beta} = -F(\beta - \alpha, \beta)$. ▲

A continuación consideraremos precisamente las formas canónicas (3) ... (5) de la ecuación (1).

§ 10.2. Clasificación de los problemas de contorno

En este párrafo consideraremos las ecuaciones canónicas elementales (3)...(5) del § 10.1. Según ya hemos mencionado en el § 10.1 distintos tipos de ecuaciones describen diferentes procesos físicos. Así, la ecuación *)

$$u_{tt} - u_{xx} = f(x, t), \tag{1}$$

llamada *ecuación de vibración de una cuerda* describe los procesos relacionados con las oscilaciones mecánicas, eléctricas, acústicas y otras.

La ecuación

$$u_t = u_{xx} = f(x, t), \tag{2}$$

que suele llamarse *ecuación de conducción del calor*, describe la propagación del calor, la difusión y otros procesos de transferencia.

La ecuación

$$u_{xx} + u_{yy} = f(x, y), \tag{3}$$

---

*) Aquí y a continuación para una mayor claridad física las variables x e y corresponden a las coordenadas espaciales y t, a la coordenada temporal.

denominada *ecuación de Poisson*, describe el campo calorífico estacionario, flujo potencial de un líquido y otros fenómenos físicos vinculados con la puesta en régimen estacionario.

Para describir por completo (unívocamente) uno u otro proceso físico es necesario, además de la misma ecuación de este proceso, asignar las condiciones iniciales (estado inicial del proceso) y el régimen de variación de la función a determinar en la frontera del dominio en que se desarrolla el proceso. Matemáticamente esto está relacionado con la no unicidad de la solución de la ecuación diferencial. Por eso para determinar unívocamente la solución conviene, además de la ecuación, asignar condiciones adicionales, llamadas *condiciones de contorno* que se subdividen en condiciones iniciales y condiciones de frontera. En consonancia con la clasificación citada se distinguen tres tipos de problemas de contorno.

Se llaman *condiciones iniciales* las condiciones que prefijan en cierto instante de tiempo, denominado generalmente inicial, el valor de la solución buscada (y, a veces, sus derivadas temporales) para todos los puntos del dominio en cuestión.

El primer tipo del problema de contorno es el *problema de Cauchy*. Así se llama el problema de resolución de la ecuación (1) ó (2) para el cual en calidad de condiciones adicionales se dan sólo las condiciones iniciales (estado inicial del proceso).

En el problema de Cauchy faltan las condiciones de frontera. Este problema se enuncia para las ecuaciones hiperbólicas y parabólicas. La ausencia de las condiciones de frontera se determina por el hecho de que se considera un dominio indefinido o un pequeño intervalo inicial de tiempo, cuando la influencia de las fronteras es despreciablemente pequeña.

El segundo tipo del problema de contorno es el *problema sin condiciones iniciales* en el cual se asignan sólo las condiciones de frontera. A su vez, las condiciones de frontera suelen subdividirse en tres géneros.

Se llama *condición de frontera de primer género* la condición con la cual en la frontera del dominio en cuestión la función buscada toma los valores asignados.

Se llama *condición de frontera de segundo género* la condición con la cual en la frontera del dominio en cuestión la derivada normal de la función buscada debe tomar los valores asignados.

Se llama *condición de frontera de tercer género* la condición con la cual en la frontera del dominio en cuestión se asigna la combinación lineal de la función buscada y de su derivada normal.

Desde el punto de vista matemático las condiciones de frontera de los géneros I y II son casos particulares de la condisión del género, III. Sin embargo, aquéllas se separan en forma indepedieiente no sólo en virtud de las causas históricas sino también merced a su diferente interpretación física y cierta diferencia en los métodos de resolución de problemas de contorno correspondientes.

La ausencia de las condiciones iniciales en los problemas reales puede determinarse por la consideración de los instantes de tiempo que están bastante alejados del inicial, cuando disminuye la influencia de las condiciones iniciales. Tales problemas se denominan, con frecuencia, *problemas de régimen estacionario*. Los problemas de este tipo pueden ser formulados para todos los tipos de ecuaciones (1) (3).

El tercer tipo de problemas de contorno es el *problema mixto*, en el cual se asignan las condiciones iniciales y de frontera. En cierto sentido este problema es la generalización de los problemas de los dos primeros tipos. El problema de Cauchy y el de contorno sin condiciones iniciales son dos casos límite contrarios del problema mixto. El primero se da para un lapso de tiempo bastante pequeño y el segundo, para un lapso de tiempo bastante grande. El problema mixto se formula para ecuaciones hiperbólicas y parabólicas.

§ 10.3. Planteamiento de los problemas de contorno más elementales

En este párrafo expondremos diferentes planteamientos de los problemas de contorno suponiendo que sus soluciones son suficientemente suaves. Aquí y a continuación por suavidad suficiente de una función entendremos la continuidad de la función y de número necesario de sus derivadas. En el planteamiento de los problemas de contorno la suavidad suficiente significa de ordinario la continuidad de las funciones y de las derivadas que forman parte de la ecuación diferencial y de las condiciones de contorno.

Se llama *solución clásica de un problema de contorno* a toda función que satisfaga la ecuación diferencial en cada punto dentro del dominio de representación de esta ecuación y que sea continua en el dominio en cuestión, incluyendo la frontera.

El planteamiento respectivo del problema de contorno se denomina *clásico*. Así pues, el planteamiento clásico impone automáticamente ciertas limitaciones para los datos de entrada del problema de contorno. Así, por ejemplo, se necesita la continuidad del segundo miembro de las ecuaciones (1)... (3) dadas en el § 10.2 y la suficiente suavidad de las funciones de frontera asignadas. Nótese que en los problemas prácticos más interesantes los segundos miembros, por ejemplo, tienen particularidades sustanciales, por eso el planteamiento clásico resulta ya insuficiente. En tales casos se introduce el concepto de solución generalizada que dejamos sin examinar limitándonos a la consideración de la solución clásica.

Problema de Cauchy para un dominio infinito. Vamos a formular este problema para la ecuación de vibración de la cuerda y para la ecuación de conducción del calor.

Consideremos el proceso de vibración de una cuerda delgada infinita (muy larga) sometida a la fuerza exterior continuamente distribuida con densidad l. Supongamos que la fuerza actúa en un plano (fig. 10.1) que es el plano de vibración de la cuerda (x, u), y ésta no es sino un hilo elástico flexible. Supongamos que la tensión producida en la cuerda debido a su flexión se subordina a la ley de Hooke y que las mismas vibraciones son bastante pequeñas. Entonces la magnitud del desplazamiento $u(x, t)$ satisface la ecuación de vibración de la cuerda

$$u_{tt} - u_{xx} = f(x, t) \quad (t > 0, \ -\infty < x < \infty). \tag{1}$$

Para la univocidad del proceso es necesario asignar, además, el desplazamiento inicial y la distribución inicial de las velocidades. Matemáticamente, esto corresponde a la representación de las condiciones iniciales:

$$u(x, 0) = u_0(x); \ u_1(x, 0) = u_1(x). \tag{2}$$

Se necesita hallar la solución clásica de la ecuación (1), que satisface las condiciones iniciales (2).

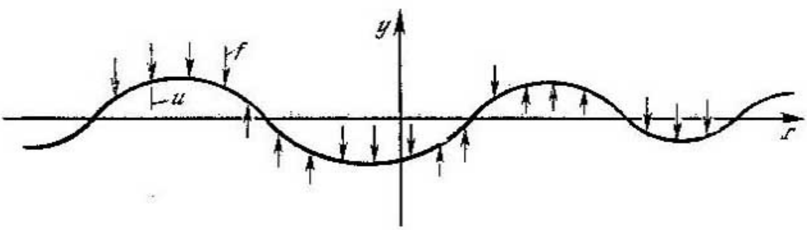


Fig. 10.1

El problema (1), (2) formulado de esta manera se llama *problema de Cauchy para la ecuación hiperbólica.*

Vamos a investigar ahora el proceso de distribución de la temperatura en una barra delgada infinita (muy pequeña). Se supone que el flujo calorífico se subordina a la ley de Fourier y que la variación de la temperatura del cuerpo es proporcional a la cantidad de calor comunicada a este último. Supongamos que dentro de la barra puede desprenderse y absorberse el calor que se caracteriza por la densidad de las fuentes térmicas f. Entonces la distribución de la temperatura en la barra se describe por la ecuación de conducción del calor:

$$u_t - u_{xx} = f(x, t) \quad (t > 0, \ -\infty < x < \infty). \tag{3}$$

Para la representación unívoca del proceso es necesario señalar la distribución inicial de la temperatura. Esto corresponde a la representación de la condición inicial

$$u(x, 0) = u_0(x). \tag{4}$$

Se necesita hallar la solución clásica de la ecuación (3), que satisface la condición inicial (4).

El problema formulado (3), (4) se llama *problema de Cauchy para la ecuación parabólica.*

Al plantear los problemas de contorno y, sobre todo, al resolverlos numéricamente es necesario responder a las tres siguientes cuestiones fundamentales correspondientes a los requisitos físicos naturales:

1. ¿Existe o no la solución del problema planteado? ¿No está ésta predeterminada por las condiciones de contorno?

2. Si la solución existe ¿es ésta única?

3. ¿Depende o no la solución continuamente de los datos iniciales del problema de contorno (f, u_0, u_1, etc,) o sea, cambia o no la solución al variar continuamente el segundo miembro de la ecuación y las condiciones de contorno? Esta propiedad se llama *estabilidad de la solución* respecto a los datos de entrada.

El problema de contorno se dice *correcto* si su solución existe, es única y estable.

El problema clásico de Cauchy, citado anteriormente, para la ecuación de vibración de la cuerda es correcta si las funciones f, u_0 y u_1 son suficientemente suaves.

Para que sea correcto el problema de Cauchy concerniente a la ecuación de conducción del calor se necesita, además de la suavidad de f y u_0, el carácter limitado de la solución.

Problema estacionario (problema sin los datos iniciales). Consideremos el régimen estacionario de distribución de la temperatura en una placa delgada limitada de forma arbitraria con una frontera suave. Supongamos que la función $u(x, y)$ expresa la temperatura de cada punto de la placa. Siempre que se trate de leyes corrientes de propagación del calor, descritas anteriormente al formular el problema de Cauchy para la ecuación de conducción del calor, la función $u(x, y)$ satisface la ecuación de Poisson

$$u_{xx} + u_{yy} = f(x, y); \; (x, y) \in D, \tag{5}$$

donde la función f asigna la densidad de las fuentes térmicas de la placa. En caso de no existir las fuentes ($f = 0$) la ecuación (5) se llama *ecuación de Laplace*:

$$u_{xx} + u_{yy} = 0. \tag{6}$$

Para describir unívocamente el proceso es necesario asignar el régimen térmico en la frontera de la placa. Esto puede hacerse asignando la distribución de la temperatura en la frontera o la distribución del flujo calorífico.

Es posible también el régimen de equilibrio térmico del cuerpo radiante con el medio ambiente. Según el régimen térmico en la frontera se obtienen tres condiciones de frontera para la función $u(x, y)$. Sea Γ la frontera del campo dado D de definición de la ecuación (6). La formulación matemática de las condiciones de frontera puede ser representada del modo siguiente:

condición de frontera de género I:

$$u|_{\Gamma} = \varphi_0(x, y); \; (x, y) \in \Gamma; \tag{7}$$

condición de frontera de género II:

$$\frac{\partial u}{\partial n}\bigg|_{\Gamma} = \varphi_1(x, y); \quad (x, y) \in \Gamma; \tag{8}$$

condición de frontera de género III:

$$\frac{\partial u}{\partial n}\bigg|_{\Gamma} + \lambda u\,|_{\Gamma} = \varphi_2(x, y); \quad (x; y) \in \Gamma. \tag{9}$$

La derivada se toma sobre la normal exterior a la curva Γ; $\lambda > 0$ es el coeficiente de conducción del calor; φ_0, φ_1, φ_2 son los asignados sobre la función Γ, con la particularidad de que φ_2 es el producto del coeficiente de conductibilidad térmica por la temperatura del medio ambiente que tiene contacto con el cuerpo.
De este modo, el problema de contorno consiste en encontrar la solución clásica de la ecuación (5) ó (6), que satisfaga una de las condiciones de frontera (7) ...(9).

De acuerdo con el tipo de condiciones de frontera se distinguen: el primer problema de contorno (5), (7), o sea, el *problema de Dirichlet;* el segundo problema de contorno (5), (8), o sea, el *problema de Neumann* y el tercer problema de contorno (5), (9).

Para los datos de entrada suficientemente suaves los problemas de contorno primero y tercero concernientes a las ecuaciones de Poisson y de Laplace son correctos.

Para el problema de Neumann el teorema de unicidad consiste en que al ser iguales las condiciones de frontera dos soluciones suyas pueden distinguirse en una magnitud constante.

En los problemas de física matemática ocupan importante lugar las funciones armónicas.

La función $u(x, y)$ se llama *armónica* en cierto dominio D si es continua junto con sus derivadas de segundo orden y satisface la ecuación (6) en este dominio.

Citemos algunas propiedades de las funciones armónicas.

1° (principio del máximo). *Si la función $u(x, y)$ está definida y es continua en $\bar{D} = D \cup \Gamma$, y satisface la ecuación* (6) *en D, los valores máximo y mínimo se alcanzan en la frontera Γ*, o sea,

$$\max_{(x, y)\in D} u(x, y) \leqslant \max_{(x, y)\in \Gamma} u(x, y), \quad \min_{(x, y)\in D} u(x, y) \geqslant \min_{(x, y)\in \Gamma} u(x, y).$$

2° (corolario de la propiedad 1°). *Si las funciones $u(x, y)$ y $v(x, y)$ son continuas en $\bar{D}$ y armónicas en $\bar{D}$, $u \leqslant v$ para $(x, y) \in \Gamma$, entonces $u \leqslant v$ para $(x, y) \in D$.*

3° (corolario de la propiedad 1°). *Si las funciones u y v son continuas en $\bar{D}$ y armónicas en D, con la particularidad de que $|u| \leqslant x$ para $(x, y) \in \Gamma$, entonces $|u| \leqslant v$ para $(x, y) \in$ D.*

Problema de contorno mixto. Consideremos el problema de propagación del calor en una barra delgada de longitud unitaria. Coloquemos uno de sus extremos en el punto $x = 0$ y otro en el punto $x = 1$. En tal barra la distribución de la temperatura durante cierto intervalo de tiempo $0 < t < T$ se describe por la ecuación

$$u_t - u_{xx} = f \ (0 < x < 1, 0 < t \leqslant T) \tag{10}$$

con la condición inicial

$$u(x, 0) = u_0(x) \ (0 \leqslant x \leqslant 1) \tag{11}$$

y, además, en este caso para la unicidad de la solución es necesario asignar el régimen de temperatura en los extremos de la barra. Esto se puede hacer con ayuda de las condiciones de frontera análogas a las que hemos formulado para las ecuaciones de Poisson y de Laplace.

Condición de frontera de género I (en el extremo de la barra $x = 0$ está asignada la temperatura):

$$u(0, t) = \varphi_0(t) \quad (0 < t \leqslant T). \tag{12}$$

Condición de frontera de género II (en el extremo de la barra $x = 0$ está asignado el flujo calorífico):

$$u_x(0, t) = \varphi_1(t) \ (0 < t \leqslant T) \tag{13}$$

Condición de frontera de género III:

$$-u_x(0, t) + \lambda u(0, t) = \varphi_2(t) \quad (0 < t \leqslant T). \tag{14}$$

Para el otro extremo de la barra $x = 1$ los segundos miembros de las condiciones de frontera (12 ... (14) se reemplazan por $\psi_0(t)$, $\psi_1(t)$ y $\psi_2(t)$, respectivamente. Nótese que las condiciones inicial y de frontera deben satisfacer las así llamadas *condiciones de conjugación*, o sea para la condición (12) $u_0(0) = \varphi_0(0)$, para la condición (13) $u_{0x}(0) = \varphi_1(0)$ y para la condición (14) $-u_{0x}(0) + \lambda u_0(0) = \varphi_2(0)$. Las condiciones análogas de conjugación han de cumplirse también en el otro extremo de la barra $z = 1$.

Así, para el primer problema de contorno las condiciones de conjugación significan que

$$u_0(0) = \varphi_0(0); \quad u_0(1) = \psi_0(0). \tag{15}$$

En el caso general en diferentes extremos de la barra pueden existir diferentes condiciones, así que el número total de todas las combinaciones posibles es igual a 6.

Enunciemos uno de los posibles problemas de contorno. Hallar la solución clásica de la ecuación (10), que satisface la condición inicial (11) y las siguientes condiciones de frontera

$$u(0, t) = \varphi_0(t); \ u(1, t) = \psi_0(t) \ (0 < t \leqslant T). \tag{16}$$

Este problema (10), (11), (16) suele llamarse primer problema de contorno para la ecuación de conducción del calor. El problema de contorno (10), (11) con las condiciones de frontera (13) ó (14) en ambos extremos de la barra se llama segundo o tercer problema de contorno, respectivamente.

De un modo análogo se plantean también otros problemas de contorno con diferentes combinaciones de las condiciones de frontera (12) ... (14) en ambos extremos de la barra.

Los problemas de contorno de los géneros primero, segundo tercero son correctos si se cumplen las condiciones correspondientes de suavidad y de conjugación para los datos de entrada.

Las soluciones de la ecuación de conducción del calor poseen la siguiente propiedad importante, análoga a la que hemos citado para la solución de la ecuación de Laplace.

Principio del máximo. *Si la función* $u(x, t)$ *es continua* en el dominio $\overline{D}_T$ $\{0 \leqslant t \leqslant T;\ 0 \leqslant x \leqslant 1$ *y satisface la ecuación* (10), *para* $f = 0$ *los valores máximo y mínimo de la función* $u(x, t)$ *se alcanzan en el instante inicial o en los puntos de la frontera* $x = 0$ *ó* $x = 1$.

Consideremos ahora las vibraciones de una cuerda delgada de longitud unitaria. La magnitud del desplazamiento $u(x, t)$ se describe por la ecuación de tipo hiperbólico

$$u_{tt} - u_{xx} = f \quad (0 < x < 1,\ t > 0). \tag{17}$$

En este caso las condiciones iniciales tienen la forma

$$u(x, 0) = u_0(x);\ u_t(x, 0) = u_1(x)\ (0 \leqslant x \leqslant 1). \tag{18}$$

Consideraremos las condiciones de frontera en la misma forma que para la ecuación de conducción del calor, o sea (12) ... (14). El problema (17) y (18) con iguales condiciones de frontera en ambos extremos de la forma (12) ... (14) se llama, respectivamente, problema de contorno primero, segundo y tercero para la ecuación hiperbólica. Todos estos problemas son correctos si se cumplen las condiciones correspondientes de suavidad y de conjugación para los datos de entrada.

A título de ejemplo enunciemos el primer problema de contorno para la ecuación de vibración de la cuerda.

Hallar la función $u(x, t)$ que satisface la ecuación (17), las condiciones iniciales (18) y las siguientes condiciones de contorno:

$$u(0, t) = \varphi_0(t);\ u(1, t) = \psi_0(t)\ (t > 0). \tag{19}$$

El sentido físico de las condiciones de frontera de género I consisse en que ambos extremos de la cuerda vibran en los regímenes dados según las leyes de las funciones dadas φ_0 y ψ_0. Las condiciones de frontera de género II corresponden al hecho de que en el extremo está asignada la ley de acción de la fuerza. La condición de frontera III corresponde a la sujeción elástica del extremo de la cuerda.

§ 10.4. Método de diferencias finitas. Conceptos fundamentales

Para los problemas de contorno elementales, enunciados en el § 10.3, en diferentes cursos de física matemática se dan algunas soluciones exactas. Al mismo tiempo incluso las ecuaciones lineales que describen los procesos reales en su mayoría son tales que no permiten construir una solución exacta con ayuda de funciones elementales. En semejantes casos se recurre a los métodos aproximados. De ordinario se consideran dos tipos de soluciones aproximadas: analíticas y numéricas. Examinemos los métodos numéricos fundados en la aproximación de diferencias de las derivadas. Tal enfoque se denomina **método de diferencias, método de diferencias finitas o método de rejillas.**

Para reducir los cálculos vamos a ilustrar este método citando las ecuaciones elementales (para las cuales, tal vez, se obtuvo también la solución exacta), teniendo presente que los principios fundamentales de construcción de los esquemas de diferencias se extienden a ecuaciones más generales.

Supongamos que se da una ecuación diferencial lineal escrita en la siguiente forma simbólica:

$$Lu\,(x,\,y) = f\,(x,\,y);\ (x,\,y) \in D. \tag{1}$$

Aquí u es la solución buscada de la ecuación; L, cierto operador diferencial que designa abreviadamente la operación diferencial correspondiente; f, el segundo miembro de la ecuación (función asignada).

Se sabe que para la unicidad de la solución de la ecuación (1) es necesario adjuntarle, además, también las condiciones de contorno (iniciales y de frontera). Escribamos estas condiciones también en la forma de una igualdad simbólica

$$lu\,(x,\,y) = \varphi\,(x,\,y),\quad (x,\,y) \in \Gamma, \tag{2}$$

donde l es cierto operador que designa abreviadamente el primer miembro de la condición de contorno; φ, el segundo miembro de la condición de contorno (función asignada); Γ, la frontera del dominio D.

Vamos a ilustrar los conceptos fundamentales del método de diferencias finitas, citando a título de ejemplo la resolución del problema de Dirichlet para la ecuación de Laplace en el cuadrado $D^0\,\{0 < x < 1,\ 0 < y < 1\}$ con la frontera $\Gamma^0\,\{x = 0,\ x = 1,\ 0 \leqslant y \leqslant 1;\ y = 0,\ y = 1,\ 0 \leqslant x \leqslant 1\}$:

$$Lu \equiv u_{xx} + u_{yy} = 0;\ (x,\,y) \in D^0; \tag{3}$$

$$u\,(x,\,y) = \frac{1}{4}\,xy\,(x+1)\,(y+1);\quad (x,\,y) \in \Gamma^0. \tag{4}$$

Ahora bien, para el problema (3), (4) el operador L transforma la función u en expresión diferencial $u_{xx} + u_{yy}$. En semejantes casos se escribe que

$$L \equiv \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2};$$

el segundo miembro de la ecuación $f = 0$; el operador de condiciones de frontera es operador idéntico, o sea, transforma la función u en u : $lu \equiv u$; el segundo miembro de la condición de contorno tiene la forma

$$\varphi(x, y)=\begin{cases} 0; & x=0,\ 0\leqslant y\leqslant 1; \\ 0; & 0\leqslant x\leqslant 1,\ y=0; \\ \frac{1}{2}y(y+1); & x=1,\ 0\leqslant y\leqslant 1; \\ \frac{1}{2}x(x+1); & 0\leqslant x\leqslant 1,\ y=1. \end{cases}$$

El método de diferencias de resolución del problema (1), (2) puede ser representado en forma de dos etapas:

1) la construcción del esquema de diferencias que aproxima el problema continuo dado;

2) la obtención de la solución del problema de diferencias y la estimación del error de esta solución.

Examinemos con más detalles estas cuestiones.

Al construir el esquema de diferencias el primer paso consiste en reemplazar el dominio $\overline{D}$ de variación continua de los argumentos por el dominio de su variación discreta, o sea, por el dominio reticular $\overline{\omega}_h$ (o simplemente por la *red*), es decir, por el conjunto de los puntos (x_n, y_m) llamados *nodos de la red*. Para el cuadrado $\overline{D}^0$ el dominio reticular puede ser construido del modo siguiente. Tracemos las rectas

$$x_n = nh,\ y_m = mh\ (h = 1/N;\ n, m = 0, 1, \ldots, N). \tag{5}$$

El conjunto de los puntos de intersección (x_n, y_m) de estas rectas es precisamente el dominio reticular $\overline{\omega}_h$ y los mismos puntos forman los nodos de la red. Toda función $v(x, y)$ definida sobre la red $\overline{\omega}_h$ se denomina *función de red* y designa con frecuencia v_h.

El segundo paso en la construcción del esquema de diferencias consiste en aproximar la expresión diferencial Lu por cierta expresión de diferencias, y la función del argumento continuo f, por la función de red, o sea, en construir cierto análogo de diferencias para la ecuación (1). Lo mismo se refiere también a las condiciones de contorno (2).

Tal aproximación conduce al sistema de ecuaciones algebraicas respecto a los valores de cierta función de red v_h. Este sistema de ecuaciones puede ser escrito en la forma siguiente:

$$L_h v_h = f_h; \tag{6}$$

$$l_h v_h = \varphi_h; \tag{7}$$

donde L_h y φ_h son los operadores de diferencias *) que aproximan L y l, respectivamente; v_h, la función de red buscada que aproxima la solución u; f_h, φ_h son las funciones de red dadas que aproximan f y φ, respectivamente.

*) Nótese que al concepto «operador» le damos un único sentido, que es la forma simbólica abreviada de notación de las expresiones diferenciales y de diferencias.

El conjunto de las ecuaciones de diferencias (6), (7) que aproximan el problema inicial (1), (2) se llama *esquema de diferencias*. Nótese que el problema inicial puede ser aproximado, hablando en general, mediante distintos esquemas de diferencias, y un mismo esquema de diferencias puede aproximar distintos problemas continuos.

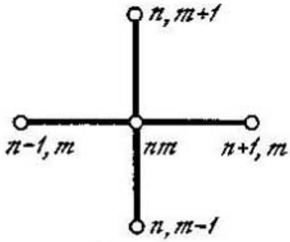


Fig. 10.2

A título de ejemplo vamos a construir el esquema de diferencias para el problema (3), (4).

Para resolver $u(x, y)$ construyamos la función de red u_h definida de la manera siguiente: $u_h(x_n, y_m) = u(x_n, y_m)$. A continuación por simplificar la forma de notación, allí donde esto no dará lugar a equivocaciones, omitiremos el índice h en las funciones de red y, sobre todo, en sus valores. De este modo, $u_{nm} = u_h \times \times (x_n, y_m)$. Utilizando esta designación, aproximemos cada una de las derivadas (3) mediante la relación de diferencias:

$$u_{xx}(y_n, y_m) \approx \frac{1}{h^2}(u_{n-1,m} - 2u_{nm} + u_{n+1,m}),$$

$$u_{yy}(x_n, y_m) \approx \frac{1}{h^2}(u_{n,m-1} - 2u_{nm} + u_{n,m+1}). \tag{8}$$

Entonces la ecuación diferencial (3) se puede aproximar mediante las ecuaciones de diferencias:

$$\frac{1}{h^2}(v_{n-1,m} + v_{n,m-1} + v_{n+1,m} + v_{n,m+1} - 4v_{nm}) = 0 \tag{9}$$
$$(n, m = 1, 2, \ldots, N-1).$$

De aquí obtenemos

$$v_{nm} = \frac{1}{4}(v_{n-1,m} + v_{n+1,m} + v_{n,m-1} + v_{n,m+1}) \tag{10}$$
$$(n, m = 1, 2, \ldots, N-1).$$

La condición de frontera (4) se aproxima del modo siguiente:

$$v_{0m} = 0, \quad v_{n0} = 0;$$

$$v_{Nm} = \frac{1}{2}\frac{m(m+N)}{N^2}, \quad v_{nN} = \frac{1}{2}\frac{n(n+N)}{N^2} \quad (n, m = 0, 1, \ldots, N). \tag{11}$$

El sistema de ecuaciones (10), (11) se resuelve de ordinario por el método de iteración simple o por el método de Seydel.

Comparando las relaciones (6), (7) y (9), (11) es fácil comprender el sentido de los operadores de diferencias y de las funciones de red del sistema (6) y (7) y su relación con los operadores y funciones respectivas del problema (1), (2).

La expresión de diferencias $L_h v_h$ no es sino la combinación lineal de los valores de la función de red en ciertos nodos llamado *molde*. En particular, la expresión de diferencias (9) contiene cinco nodos (molde de cinco puntos). Este molde lleva el nombre de «cruz» (fig. 10.2).

Ahora bien, para $(N-1)^2$ valores desconocidos v_{nm} $(n, m = 1, 2, \ldots, N-1)$ de la función de red v_h obtenemos el sistema de $(N-1)^2$ ecuaciones (9) ó (10) en el cual las magnitudes v_{0m}, v_{n0}, v_{Nm}, v_{nN} están definidas por las condiciones de frontera (11). Las relaciones (9), (11) pueden considerarse juntas como sistema único de $(N+1)^2-4$ ecuaciones con $(N+1)^2-4$ incógnitas v_{nm} $(n, m = 0, 1, \ldots, N)$, a excepción de v_{00}, v_{0N}, v_{NN}, v_{N0}.

Intuitivamente está claro que cuanto más exactas son las aproximaciones de tipo (8) tanto más proxima es la función de v_h *a* u_h. Por eso en la etapa dada introduzcamos el concepto riguroso de aproximación.

La solución v_h del problema (6), (7) es una función de red y, por consiguiente, depende del parámetro h, o sea, del paso de la red. Surge una pregunta natural sobre la posibilidad de principio de aproximar la solución v_h a la solución $u(x, y)$ del problema (1), (2) mediante un número finito de operaciones con ayuda de la elección correspondiente del paso h.

Vamos a efectuar la comparación para dos funciones de red v_h y $u_h = u(x, y)$, $(x, y) \in \overline{\omega}_h$. Para determinar la proximidad de dos funciones de red asignemos el concepto de norma sobre el conjunto de las funciones de red así:

$$\| v_h \| = \max_{(x, y) \in \overline{\omega}_h} | v(x, y) | . \tag{12}$$

En la definición (12) el máximo se toma en el dominio de definición de la función que está bajo el signo de norma.

Consideremos el error del esquema de diferencias (6), (7): $z_h = v_h - u_h$. Sustituyendo $v_h = z_h + u_h$ en (6), (7) obtenemos para z_h un problema análogo al problema para v_h:

$$L_h z_h = f_h - L_h u_h, \tag{13}$$

$$l_h z_h = \varphi_h - l_h u_h. \tag{14}$$

Los segundos miembros de las ecuaciones (13) y (14) se llaman *error de aproximación de la ecuación* (1) mediante la ecuación de diferencias (6) y *error de aproximación de la condición de contorno* (2) mediante la condición de diferencia (7) en la solución del problema inicial (1), (2).

Diremos que el esquema de diferencias (6), (7) *aproxima el problema* (1), (2) *con el orden* $k > 0$ *respecto a* h en la solución $u(x, y)$ si

$$\| f_h - L_h u_h \| \leqslant c_1 h^k; \quad \| \varphi_h - l_h u_h \| ^k c_2 h^k, \tag{15}$$

donde c_1 y c_2 son las constantes que no dependen de h.

Determinemos el orden de aproximación para el esquema (9), (11) o bien, que es lo mismo, para (10), (11). Puesto que las condiciones de frontera están dadas en la red exactamente, el primer miembro de la segunda de las igualdades (15) es igual a cero y el orden de aproximación se determinará por la primera de las desigualdades (15).

El segundo miembro de la ecuación (3) es igual a cero, por eso se estima sólo la norma $L_h u_h$.

Hagamos uso de la fórmula de Taylor:

$$u_{n\pm 1,\, m} = u_{nm} \pm h\left(\frac{\partial u}{\partial x}\right)_{nm} + \frac{h^2}{2}\left(\frac{\partial^2 u}{\partial x^2}\right)_{nm} \pm$$
$$\pm \frac{h^3}{6}\left(\frac{\partial^3 u}{\partial x^3}\right)_{nm} + \frac{h^4}{24}\left(\frac{\partial^4 u}{\partial x^4}\right)^{\pm}_{xy},$$
$$\left(\frac{\partial^4 u}{\partial x^4}\right)^{\pm}_{xy} = \frac{\partial^4}{\partial x^4} u(x_n \pm \theta_1 h,\, y_m), \quad 0 < \theta_1 < 1;$$
$$u_{n,\, m\pm 1} = u_{nm} \pm h\left(\frac{\partial u}{\partial y}\right)_{nm} + \frac{h^2}{2}\left(\frac{\partial^2 u}{\partial y^2}\right)_{nm} \pm$$
$$\pm \frac{h^3}{6}\left(\frac{\partial^3 u}{\partial y^3}\right)_{nm} + \frac{h^4}{24}\left(\frac{\partial^4 u}{\partial y^4}\right)^{\pm}_{xy},$$
$$\left(\frac{\partial^4 u}{\partial y^4}\right)^{\pm}_{xy} = \frac{\partial^4}{\partial y^4} u(x_n,\, y_m \pm \theta_2 h), \quad 0 < \theta_2 < 1.$$

Sumando todas estas cuatro relaciones, encontramos

$$u_{n+1,\, m} + u_{n-1,\, m} + u_{n,\, m+1} + u_{n,\, m-1} = 4u_{nm} + h^2\left[(u_{xx})_{nm} + (u_{yy})_{nm}\right] +$$
$$+ \frac{h^4}{24}\left[\left(\frac{\partial^4 u}{\partial x^4}\right)^{+}_{xy} + \left(\frac{\partial^4 u}{\partial x^4}\right)^{-}_{xy} + \left(\frac{\partial^4 u}{\partial y^4}\right)^{+}_{xy} + \left(\frac{\partial^4 u}{\partial y^4}\right)^{-}_{xy}\right].$$

Utilizando las ecuaciones de diferencias (6), obtenemos

$$\| L_h u_h \| \leqslant \frac{h^2}{6} M_4, \tag{16}$$

donde $M_4 = \max\limits_{(x,\, y)\in \overline{D}} \left(\left|\frac{\partial^4 u}{\partial x^4}\right|, \left|\frac{\partial^4 u}{\partial y^4}\right|\right)$. Aquí se supone la existencia de las derivadas correspondientes. Ahora bien, el esquema (9), (11) tiene el segundo orden de aproximación.

En esto terminamos el examen de la primera etapa, o sea, la construcción del esquema de diferencias y pasamos a la segunda etapa, o sea, a la resolución del problema de diferencias y a la estimación del error.

En esta etapa las cuestiones más importantes son la cuestión de resolubilidad del problema de diferencias, la de unicidad de su solución y la de dependencia continua de la solución de los datos de entrada. Por *datos de entrada* se entienden los segundos miembros de las ecuaciones de diferencias, o sea, las condiciones iniciales y de frontera del problema de diferencias. Análogamente al hecho de cómo se plantea la cuestión acerca de lo correcto de los problemas de la física matemática, se puede formular el concepto de carácter correcto del esquema de diferencias. Sea v_h la solución y f_h, φ_h los datos de entrada de cierto problema de diferencias (6), (7). Es evidente que la solución y los datos de entrada dependen de h.

Diremos que el *problema (esquema) de diferencias es correcto* si para todos los $h < h_0$ $(h_0 > 0)$ se cumplen las condiciones:

1°) la solución del problema de diferencias existe y es única;

2°) la solución del problema de diferencias depende continuamente de los datos de entrada.

Para el problema de diferencias (6), (7) la condición 2° puede ser escrita de la manera siguiente:

$$\| \bar{v}_h - v_h \| \leqslant M_1 \| \bar{f}_h - f_h \| + M_2 \| \bar{\varphi}_h - \varphi_h \| , \qquad (17)$$

donde los símbolos sin raya corresponden a un problema y con raya, a otro.

La condición señalada se llama *estabilidad del problema (esquema) de diferencias referente a los datos de entrada* o simplemente *estabilidad.*

En la teoría de esquemas de diferencias se demuestra que el esquema «cruz» construido para el problema de Dirichlet y para la ecuación de Laplace (hablando en general, también para la ecuación de Poisson) es correcto.

Una vez enunciado el concepto de aproximación y de estabilidad para los esquemas de diferencias, hemos llegado a la cuestión más importante, es decir, a la cuestión sobre la convergencia de la solución del problema de diferencias (6), (7), a la solución del problema continuo (1), (2).

Diremos que el esquema de diferencias (6), (7) *converge con la velocidad de orden* $s > 0$ *respecto a* h, si se cumple la condición

$$\| v_h - u_h \| \leqslant ch^s,$$

donde c es la constante que no depende de h.

Entre los conceptos de aproximación, de carácter correcto y de estabilidad existe una relación íntima que se determina por el teorema siguiente.

Teorema. *Supongamos que el problema de diferencias* (6), (7) *aproxima el problema* (1), (2) *en la solución* $u(x, y)$ *con orden* $k > 0$ *respecto a* h *y es correcto. Entonces este esquema converge con un orden igual al orden de aproximación* k, o sea, *se cumple la estimación*

$$\| v_h - u_h \| \leqslant ch^k. \qquad (18)$$

□ Por definición de la aproximación tenemos

$$\| f_h - \mathcal{L}_h u_h \| \leqslant c_1 h^k; \ \| \varphi_h - l_h u_h \| \leqslant c_2 h^k.$$

Utilizando las igualdades (13) y (14), obtenemos

$$\| L_h z_h \| \leqslant c_1 h^k; \ \| l_h z_h \| \leqslant c_2 h^k.$$

Luego, en virtud de la suposición de que el esquema de diferencias es estable [relación (17)], tenemos

$$\| z_h \| \leqslant M_1 \| L_h z_h \| + M_2 \| l_h z_h \| ,$$

de donde, utilizando las estimaciones que acabamos de obtener, encontramos

$$\| z_h \| = \| v_h - u_h \| \leqslant M_1 c_1 h^k + M_2 c_2 h^k = ch^k. \quad \blacksquare$$

Ejemplo 1. Hallar la solución de los problemas (3) y (4):

$$u_{xx} + u_{yy} = 0; \ (x, y) \in D^0,$$

$$u(x, y) = \frac{1}{4} xy(x+1)(y+1); \ (x, y) \in \Gamma^0.$$

Aquí D^0 es el cuadrado $\{0 < x < 1, \ 0 < y < 1\}$ con la frontera $\Gamma^0 \{x = 0, x = 1, \ 0 \leqslant y \leqslant 1; \ y = 0, y = 1, \ 0 \leqslant x \leqslant 1\}$.

△ Para este problema el sistema de ecuaciones de diferencias finitas ya está escrito en el caso general y tiene la forma:

$$v_{nm} = \frac{1}{2}(v_{n-1, m} + v_{n+1, m} + v_{n, m-1} + v_{n, m+1})$$
$$(n; m = 1, 2, \ldots, N-1);$$
$$v_{0m} = v_{n0} = 0;$$
$$v_{Nm} = \frac{1}{2} \cdot \frac{m(m+N)}{N^2}, \quad v_{nN} = \frac{1}{2} \cdot \frac{n(n+N)}{N^2} \quad (n; m = 0, 1, \ldots, N).$$

Tomando en calidad de paso de la red $h = 1/3$ $(N = 3)$, construyamos la tabla inicial de condiciones de frontera y de valores desconocidos:

Tabla 10.1

	0,222	0,556	
0	v_{12}	v_{22}	0,556
0	v_{11}	v_{21}	0,222
	0	0	

El sistema inicial de ecuaciones para los valores desconocidos tomará la forma siguiente:

$$v_{11} = \frac{1}{4}(0 + v_{21} + 0 + v_{12}),$$
$$v_{12} = \frac{1}{4}(0 + v_{22} + v_{11} + 0{,}222),$$
$$v_{21} = \frac{1}{4}(v_{11} + 0{,}222 + 0 + v_{22}),$$
$$v_{22} = \frac{1}{4}(v_{12} + 0{,}556 + v_{21} + 0{,}556).$$

Resolvamos este sistema por el método de iteración simple. Para esto es necesario obtener los valores iniciales para las incógnitas

Vamos a obtenerlos con ayuda de la interpolación lineal respecto a los valores de frontera: primero en filas y luego en columnas.

La interpolación lineal en filas se realiza por la fórmula

$$\bar{v}_{nm} = v_{0m} + (v_{Nm} - v_{0m}) \frac{m}{N}$$

lo que da (n; m = 1, 2)

$\overline{v_{12}} = 0{,}185$; $\overline{v_{22}} = 0{,}371$; $\overline{v_{11}} = 0{,}074$; $\overline{v_{21}} = 0{,}148$.

La interpolación lineal en columnas se realiza por la fórmula

$$\bar{\bar{v}}_{nm} = v_{n0} + (v_{nN} - v_{n0}) \frac{n}{N}$$

lo que da (n; $m = 1{,}2$).

$\overline{\overline{v_{12}}} = 0{,}148$; $\overline{\overline{v_{11}}} = 0{,}074$; $\overline{\overline{v_{22}}} = 0{,}371$; $\overline{\overline{v_{21}}} = 0{,}185$.

Por valores iniciales tomamos la semisuma de las magnitudes obtenidas:

$$v_{nm}^0 = \frac{1}{2} (\bar{v}_{nm} + \bar{\bar{v}}_{nm}),$$

o sea, $v_{12}^0 = v_{21}^0 = 0{,}166$; $v_{22}^0 = 0{,}371$; $v_{11}^0 = 0{,}074$.

Ahora se puede realizar el proceso iterativo:

$$v_{11}^{k+1} = \frac{1}{2} v_{12}^k; \quad v_{12}^{k+1} = \frac{1}{4} (v_{11}^k + v_{22}^k + 0{,}222);$$

$$v_{22}^{k+1} = \frac{1}{2} (v_{12}^k + 0{,}556).$$

Hemos utilizado la simetría de los datos iniciales ($v_{12}^0 = v_{21}^0$) y del sistema de ecuaciones. Vamos a resolver este sistema hasta que dos iteraciones sucesivas coincidan con una exactitud de hasta 0,001. Representemos los resultados de los cálculos en forma de la tabla 10.2.

Tabla 10.2

	Número de iteración		
	0	1	2
v_{11}	0,074	0,083	0,084
$v_{12} = v_{21}$	0,166	0,167	0,166
v_{22}	0,371	0,361	0,362

Para obtener la solución resultó suficiente dos iteraciones. Esto es consecuencia de la sencillez del problema de diferencias debido a una red gruesa. ▲

Ejemplo 2. Hallar la solución de la ecuación de Laplace (3) en el cuadrado unitario para las siguientes condiciones de frontera:

$$u(x, y) = \begin{cases} 0; & 0 \leqslant x \leqslant 1, \ y = 0; \\ \dfrac{8}{3} y (64y^2 - 60y + 29); & x = 0, \ 0 \leqslant y \leqslant 1; \\ \dfrac{8}{3} (1 - x)(64x^2 - 68x + 33); & 0 \leqslant x \leqslant 1, \ y = 1; \\ 0; & x = 1, \ 0 \leqslant y \leqslant 1. \end{cases}$$

△ Pongamos $h = 0{,}25$ y construyamos el sistema (10), teniendo en cuenta los valores de frontera:

$$v_{11}^{k+1} = \frac{1}{4}(12 + v_{21}^k + 0 + v_{12}^k);$$

$$v_{21}^{k+1} = \frac{1}{4}(v_{11}^k + v_{31}^k + 0 + v_{22}^k);$$

$$v_{31}^{k+1} = \frac{1}{4}(v_{21}^k + 0 + 0 + v_{32}^k) = \frac{1}{2} v_{21}^k;$$

$$v_{12}^{k+1} = \frac{1}{4}(20 + v_{22}^k + v_{11}^k + v_{13}^k);$$

$$v_{22}^{k+1} = \frac{1}{4}(v_{12}^k + v_{32}^k + v_{23}^k + v_{21}^k) = \frac{1}{2}(v_{12}^k + v_{21}^k);$$

$$v_{13}^{k+1} = \frac{1}{4}(40 + v_{23}^k + v_{12}^k + 40) = 20 + \frac{1}{2} v_{12}^k.$$

Las condiciones de frontera y los valores desconocidos se dan en la tabla 10.3.

Tabla 10.3

	40	20	12	
40	v_{13}	v_{23}	v_{33}	0
20	v_{12}	v_{23}	v_{32}	0
12	v_{11}	v_{21}	v_{31}	0
	0	0	0	

Al construir este sistema hemos utilizado la propiedad de simetría: $v_{nm} = v_{N-m, N-n}$.

El cálculo de la aproximación inicial se realiza con ayuda de la interpolación lineal respecto a los valores de frontera en los nodos interiores. Utilizando para el cálculo de v_{n1}^0 la fórmula

$$v_{n1}^0 = 12\left(1 - \frac{n}{4}\right),$$

obtenemos $v^0_{11}=9$, $v^0_{21}=6$, $v^0_{31}=3$. En virtud de la simetría suponemos $v^0_{32}=v^0_{21}=6$, $v^0_{33}=v^0_{11}=9$.

Utilizando para calcular v^0_{12} y v^0_{22} la fórmula

$$v^0_{n2}=20\left(1-\frac{7}{30}n\right),$$

tenemos $v^0_{12}=15,\ 33$, $v^0_{22}=10,\ 66$. En virtud de la simetría suponemos que $v^0_{23}=v^0_{12}=15,\ 33$.

El último valor v^0_{13} se calcula con ayuda de la fórmula

$$v^0_{13}=+40-\frac{40-15,33}{2}\cdot 1=27,67.$$

Vamos a realizar la solución del sistema por dos métodos: por el de iteración simple (tabla 10.4) y por el de Seydel (tabla 10.5). Ejecutaremos el cálculo hasta que coincidan dos soluciones sucesivas para cada variable con exactitud hasta 0,1. El cálculo con ayuda del método de iteración simple necesitó cuatro iteraciones y con ayuda del método de Seydel, tres.

Tabla 10.4

	40	20	12	
40	28,5	17,0	8,6	0
20	17,0	11,3	5,6	0
12	8,6	5,6	2,8	0
	0	0	0	

Tabla 10.5

	40	20	12	
40	28,6	17,0	8,6	0
20	17,0	11,4	5,7	0
12	8,6	5,7	2,8	0
	0	0	0	

Las soluciones finales se dan en las tablas 10.4 y 10.5. ▲

§ 10.5. Esquemas de diferencias para resolver la ecuación de conducción del calor

Consideremos el primer problema de contorno para la ecuación de conducción del calor en el rectángulo $\overline{D}\{0\leqslant x\leqslant 1,\ 0\leqslant t\leqslant T\}$.

Se necesita hallar la solución del problema, continua en $\overline{D}$:

$$Lu \equiv u_t - u_{xx} = f \quad (0 < x < 1, \quad 0 < t \leqslant T); \tag{1}$$

$$u(x,\ 0) = u_0(x) \quad (0 \leqslant x \leqslant 1); \tag{2}$$

$$u(0,\ t) = \varphi_0(t); \quad u(1,\ t) = \psi_0(t) \ (0 \leqslant t \leqslant T). \tag{3}$$

De un modo análogo al utilizado en el § 10.4 para la ecuación de Poisson, construyamos con ayuda del método de diferencias la solución del problema (1) ... (3).

En el dominio $\overline{D}$ introduzcamos una red uniforme rectangular $\overline{\omega}_{h\tau}\{x_n,\ t_k\}$ con un paso de $h = 1/N$ en la coordenada x y con un paso de $\tau = T/M$ en la coordenada t:

$$x_n = nh \ (n = 0,\ 1,\ \ldots,\ N), \quad t_k = k\tau \ (k = 0,\ 1,\ \ldots,\ (M).$$

Vamos a aproximar las derivadas del primer miembro de la ecuación (1) mediante las siguientes expresiones de diferencias:

$$(u_t)_n^k \approx \frac{u_n^{k+1} - u_n^k}{\tau} \ \text{ o bien } \ (u_t)_n^k \approx \frac{u_n^k - u_n^{k-1}}{\tau}$$

$$(u_{xx})_n^k \approx \frac{u_{n-1}^k - 2u_n^k + u_{n+1}^k}{h^2}. \tag{4}$$

De acuerdo con la aproximación (4) construyamos dos análogos de diferencias (11) con la función de red desconocida $v_{h\tau}$:

$$L_h v_{h\tau} = \frac{v_n^{k+1} - v_n^k}{\tau} - \frac{v_{n-1}^k - 2v_n^k + v_{n+1}^k}{h^2} = f_n^k, \tag{5}$$

$$L_h v_{h\tau} = \frac{v_n^k - v_n^{k-1}}{\tau} - \frac{v_{n-1}^k - 2v_n^k + v_{n+1}^k}{h^2} = f_n^k. \tag{6}$$

Aquí f_n^k son los valores de cierta función de red $f_{h\tau}$ correspondiente al segundo miembro de la ecuación (1), por ejemplo $f_n^k = f(x_n,\ t_k)$. Para el esquema (5) suele tomarse $f_n^k = f\left(x_n,\ t_k + \frac{\tau}{2}\right)$ y para el esquema (6), $f_n^k = f\left(x_n,\ t_k - \frac{\tau}{2}\right)$. Para el primer problema de contorno las condiciones inicial y de frontera se aproximan exactamente:

$$\begin{gathered} v_n^0 = u_0(nh) \ (n = 0,\ 1,\ \ldots,\ N): \\ v_0^k = \varphi_0(k\tau),\ v_N^k = \psi_0(k\tau) \ (k = 0,\ 1,\ \ldots,\ M). \end{gathered} \tag{7}$$

En caso de los problemas de contorno segundo y tercero las condiciones de frontera se aproximan basándose en las fórmulas análogas a las relaciones (4).

Los esquemas (5) y (6) se ilustran en moldes de cuatro puntos representados en la fig. 10.3.

El esquema (5) se dice *explícito* y el esquema (6), *implícito*.

Tal definición se debe al hecho de que el esquema (5) determina en forma explícita los valores de la función de red desconocida, sucesivos en el tiempo, en dependencia de los precedentes. En efecto, de la igualdad (5), suponiendo $r = \tau/h^2$, es fácil obtener que

$$v_n^{k+1} = r\,(v_{n-1}^k + v_{n+1}^k) + (1 - 2r)\,v_n^k + \tau f_n^k. \tag{8}$$

Ahora bien, utilizando las condiciones (7) y la forma explícita (8), se puede obtener sucesivamente todo valor de v_n^k. Por consiguiente, la solución del sistema (7), (8) existe y ésta es única. El asunto es otro con el esquema (6). Escribámoslo en la forma siguiente:

$$rv_{n-1}^k - (1 + 2r)\,v_n^k + rv_{n+1}^k = -\,(v_n^{k-1} + \tau f_n^k). \tag{9}$$

Este esquema da los valores de la función de red buscada en forma implícita, es decir, en forma de un sistema de ecuaciones. Se puede mos-

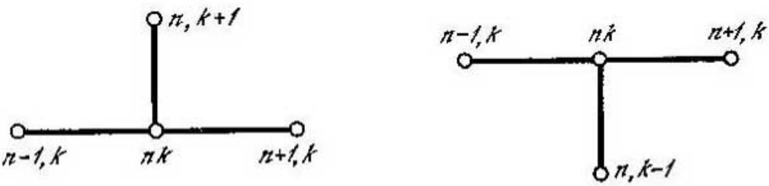


Fig. 10.3

trar que la solución del sistema (7), (9) existe y es única. Por lo general, se determina por el método de pasadas que aquí dejamos sin considerar.

El orden de aproximación para los esquemas (5) y (6) se determina basándose en las fórmulas respectivas de Taylor de un modo análogo al empleado para la ecuación de Poisson en el § 10.4. Como resultado obtenemos que los esquemas de diferencias (5), (7) y (6), (7) aproximan el problema (1) ... (3) con un error de $O\,(\tau + h^2)$, o sea,

$$\| L_h u_{h\tau} - f_{h\tau} \| \leqslant M\,(\tau + h^2). \tag{10}$$

En la teoría de esquemas de diferencias se demuestra la validez de las propiedades siguientes.

1°. *El esquema explícito* (5), (7) *tiene para* $r \leqslant 1/2$ *la solución única y es estable* y *para* $r > 1/2$ *es inestable.*

2°. *El esquema implícito* (6), (7) *tiene la solución única y es estable para todo* r.

Así pues, basándose en el teorema dado en el § 10.4, la relación (10) y las propiedades que acabamos de formular, se puede afirmar la convergencia del esquema explícito para $r \leqslant 1/2$ y del esquema implícito para todos h y τ con error $O\,(\tau + h^2)$.

Ejemplo. Resolver el problema (1) ... (3) para $f = 0$; $u_0 = x\,(1 - x)$; $\varphi_0 = \psi_0 = 0$; $T = 0{,}1$.

△ En este caso la ecuación (1) y las condiciones (2), (3) toman la forma

$$u_t - u_{xx} = 0 \ (0 < x < 1, \ 0 < t \leqslant 0{,}1);$$
$$u\,(x, 0) = x\,(1 - x) \ (0 \leqslant x \leqslant 1);$$
$$u\,(0, t) = u\,(1, t) = 0 \ (0 \leqslant t \leqslant 0{,}1).$$

En calidad de esquema de cálculo tomemos el esquema explícito (8). Pongamos $h = 0{,}25$; entonces $\tau \leqslant 0{,}03$. Puesto que $T = 0{,}1$, escojamos $\tau = 0.025$ para que M sea un número entero ($M = 4$). Calculemos $r = \tau/h^2 = 0{,}4$. La fórmula de cálculo se escribe así:

$$v_n^{k+1} = 0{,}4\,(v_{n-1}^k + v_{n+1}^k) + 0{,}2v_n^k \quad (n = 1, 2, 3; \ k = 0, 1, 2, 3);$$
$$v_n^0 = \frac{n\,(4-n)}{16}; \quad v_0^k = v_4^k = 0 \quad (n = 1, 2, 3; \ k = 1, 2, 3, 4).$$

Así pues, obtenemos las condiciones iniciales $v_1^0 = 0{,}1875$; $v_2^0 = 0{,}2500$; $v_3^0 = 0{,}1875$ y las condiciones de frontera $v_0^k = v_4^k = 0$. En el primer paso tenemos

$$v_1^1 = 0{,}4\,(v_0^0 + v_2^0) + 0{,}2v_1^0 = 0{,}1375;$$
$$v_2^1 = 0{,}4\,(v_1^0 + v_3^0) + 0{,}2v_2^0 = 0{,}2000.$$

En virtud de la simetría $v_3^1 = v_1^1 = 0{,}1375$. De un modo análogo se realizan los cálculos también en los pasos sucesivos.

Todos los cálculos se dan en la tabla 10.6. ▲

Tabla 10.

k	0	1	2	3	4
v_0^k	0	0	0	0	
v_1^k	0,1875	0,1375	0,1075	0,0815	0,0627
v_2^k	0,2500	0,2000	0,1500	0,1160	0,0884
v_3^k	0,1875	0,1375	0,1075	0,0815	0,0627
v_4^k	0	0	0	0	

§ 10.6. Esquemas de diferencias para resolver la ecuación de vibración de la cuerda

Consideremos el primer problema de contorno para la ecuación de la vibración de una cuerda en el rectángulo $\bar{D}\,\{0 \leqslant x \leqslant 1; \ 0 \leqslant t \leqslant T\}$. Se necesita hallar la solución del problema, continua en $\bar{D}$:

$$Lu = u_{tt} - u_{xx} = f \ (0 < x < 1, \ 0 < t \leqslant T); \qquad (1)$$
$$u\,(x, 0) = u_0\,(x); \ u_t\,(x, 0) = u_1\,(x) \quad (0 \leqslant x \leqslant 1); \qquad (2)$$
$$u\,(0, t) = \varphi_0\,(t); \ u\,(1, t) = \psi_0\,(t) \ (0 \leqslant t \leqslant T). \qquad (3)$$

La aplicación del método de diferencias finitas a la resolución del problema (1) ... (3) poco se distingue, en realidad, de su aplicación a la ecuación de conducción del calor *). El dominio $\overline{D}$ se cubre por la red $\overline{\omega_{h\tau}}$. La distinción consiste en aproximar la derivada segunda respecto a la variable t:

$$(u_{tt})_n^k \approx \frac{u_n^{k-1} - 2u_n^k + u_n^{k+1}}{\tau^2}. \tag{4}$$

La aproximación de diferencias para la ecuación toma la forma

$$L_h v_{h\tau} \equiv \frac{v_n^{k-1} - 2v_n^k + v_n^{k+1}}{\tau^2} - \frac{v_{n-1}^k - 2v_n^k + v_{n+1}^k}{h^2} = f_n^k, \tag{5}$$

Las condiciones iniciales se aproximan de la manera siguiente:

$$v_n^0 = u_0(nh); \quad \frac{v_n^1 - v_n^{-1}}{\tau} = u_1(nh), \tag{6}$$

$$(n = 0, 1, \ldots, N).$$

Las condiciones de frontera se aproximan exactamente igual que para la ecuación de conducción del calor:

$$v_0^k = \varphi_0(k\tau); \; v_N^k = \psi_0(k\tau) \; (k = 0, 1, \ldots, M). \tag{7}$$

El esquema (5) está determinado en el molde «cruz» de cinco puntos (véase la fig. 10.2).

El valor v_n^{-1} es una incógnita ficticia que se puede determinar de la relación (6) y sustituir en la ecuación (5). En este caso obtenemos un esquema explícito simple ($\gamma = \tau/h$):

$$v_n^{k+1} = -v_n^{k-1} + \gamma^2 (v_{n-1}^k + v_{n+1}^k) + 2(1-\gamma^2) v_n^k + \tau^2 f_n^k. \tag{8}$$

El orden de aproximación del esquema de diferencias (6) ...(8) se determina análogamente al indicado en el § 10.4 para la ecuación de Laplace. El análisis muestra que el error de aproximación del esquema (6) ... 8) es $O(\tau^2 + h^2)$ y, además, este esquema es estable cuando $\gamma^2 = (\tau/h)^2 \leqslant 1/(1+\varepsilon)$; $\varepsilon > 0$. Ahora bien, éste converge con un error de orden $O(\tau^2 + h^2)$ para la condición mencionada.

Ejemplo. Resolver el problema (1) . . . (3) para $f = 0$; $u_0 = x(1-x)$, $u_1 = \varphi_0 = \psi_0 = 0$; $T = 0{,}6$.

△ Pongamos $h = 0{,}25$, entonces $\tau < 0{,}25$. Puesto que $T = 0{,}6$, escojamos $\tau = 0{,}2$ para que M sea un número entero ($M = 3$). Calculemos $\gamma^2 = (\tau/h)^2 = 0{,}64$. La fórmula de cálculo se escribe así:

$$v_n^{k+1} = -v_n^{k-1} + 0{,}64(v_{n-1}^k + v_{n+1}^k) + 0{,}72 v_n^k \quad (n = 1, 2, 3),$$

$$v_n^0 = \frac{n(4-n)}{16} \cdot \frac{v_n^1 - v_n^{-1}}{0{,}2} = 0 \quad (n = 0, 1, 2, 3, 4).$$

$$v_0^k = v_4^k = 0 \quad (k = 0, 1, 2, 3, 4).$$

Ahora bien, obtenemos las condiciones iniciales $v_1^0 = 0{,}188$; $v_2^0 = 0{,}250$; $v_3^0 = 0{,}188$; $v_n^{-1} = v_1^1$ y las condiciones de frontera $v_0^k = v_4^k = 0$ $(k = 0, 1, 2, 3, 4)$.

*) En este párrafo hemos utilizado las designaciones y conceptos considerados en el § 10.4 y 10.5.

Para calcular v_1^1 al primer paso se necesitará el valor ficticio v_1^{-1}. Lo obtenemos de la condición inicial $v_1^{-1} = v_1^1$. Así pues, al primer paso se tiene

$$v_n^1 = -v_n^1 + 0{,}64\,(v_{n-1}^0 + v_{n+1}^0) + 0{,}72 v_n^0.$$

De aquí

$$v_1 = 0{,}32\,(v_0^0 + v_2^0) + 0{,}36 v_1^0 = 0{,}148;$$
$$v_2^1 = 0{,}32\,(v_1^0 + v_3^0) + 0{,}36 v_2^0 = 0{,}210.$$

En virtud de la simetría del problema $v_3^1 = v_1^1 = 0{,}148$. Al segundo paso obtenemos:

$$v_1^2 = -v_1^0 + 0{,}64\,(v_0^1 + v_2^1) + 0{,}72 v_1^1 = 0{,}053;$$
$$v_2^2 = -v_2^0 + 0{,}64\,(v_1^1 + v_3^1) + 0{,}72 v_2^1 = 0{,}091.$$

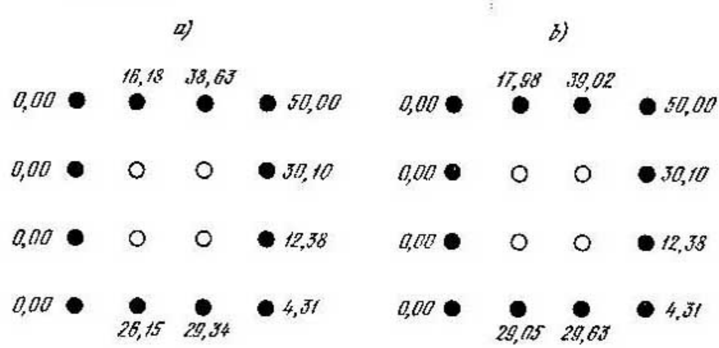
En virtud de la simetría del problema $v_3^2 = v_1^2 = 0{,}053$. De un modo análogo se realizan los cálclous también a los pasos sucesivos. Todos los cálculos se dan en la tabla 10.7.▲

Tabla 10.7

k	0	1	2	3
v_0^k	0	0	0	0
v_1^k	0,188	0,148	0,053	−0,052
v_2^k	0,250	0,210	0,091	−0,077
v_3^k	0,188	0,148	0,053	−0,052
v_4^k	0	0	0	0

Ejercicios

1. Hallar la solución aproximada de la ecuación de Laplace $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0$ para el cuadrado al ser válidas las condiciones de contorno indicadas



2. Hallar la solución aproximada de la ecuación de Laplace $\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}=0$ con un paso de $h=1/6$ para el cuadrado siendo válidas las condiciones de contorno indicadas

		9,81	19,78	29,12	40,16	42,31		
0,00	×	×	×	×	×	×	×	50,00
0,00	×	○	○	○	○	○	×	40,16
0,00	×	○	○	○	○	○	×	33,11
0,00	×	○	○	○	○	○	×	19,14
0,00	×	○	○	○	○	○	×	13,00
0,00	×	○	○	○	○	○	×	6,98
0,00	×	×	×	×	×	×	×	4,31
		17,28	31,96	40,00	30,50	17,28		

3. Hallar la solución aproximada de la ecuación $\frac{\partial^2 u}{\partial x^2}=\frac{\partial u}{\partial t}$, que satisface las condiciones de contorno $u(x, 0) = g_0(x)$, $u(0, t) = f_0(t)$, $u(1, t) = f_1(t)$ para los valores $0 \leqslant t \leqslant T$, tomando en el argumento x el paso $h = 0{,}1$. Considerar dos variantes de las condiciones de contorno:

a) $g_0(x) = (1{,}1x^2 + 1{,}1)$ sen πx, $f_0(t) = f_1(t) = 0$; $T = 0{,}02$; $r = 1/2$;

b) $g_0(x) = (1{,}1x^2 + 2{,}3)\, e^{-x}$; $f_0(t) = 2{,}3$; $f_1(t) = 3{,}4\, e^{-1}$; $T = 0{,}01$; $r = 1/6$.

Respuestas a los ejercicios

Capítulo I

1. a) 2,7546; 2,755; 2,75; 2,8; 3; b) 3,1416; 3,142; 3,14; 3,1; 3; c) 0,5645; 0,565; 0,56; 0,6; 1; d) 4,194; 4,19; 4,2; 4; e) 0,6065; 0,607; 0,61; 0,6; 1. **2.** a) 1,14; $\Delta_a = 0{,}026$ $\delta_a = 0{,}23\%$; b) 0,0102; $\Delta_a = 0{,}00005$; $\delta_a = 0{,}5\%$; c) 0,124; $\Delta_a = 0{,}0005$; $\delta_a = 0{,}41\%$; d) 922; $\Delta_a = 0{,}45$; $\delta_a = 0{,}049\%$; e) 0,00246; $\Delta_a = 0{,}000002$; $\delta_a = 0{,}082\%$; **3.** a) 0,018; b) 0,099; c) 0,047; d) 2,0; e) 0,00035. **4.** a 3; b) 4; c) 3; d) 2; e) 3. **5.** a) segunda; b) segunda; c) primora; d) segunda; e) segunda. **6.** $a = 47{,}5$. **7.** 46,39. **8.** 3,29. **9.** a) $\Delta_a = 0{,}0005$, $\delta_a = 0{,}0075\%$; b) $\Delta_a = 0{,}0005$; $\delta_a = 0{,}003\%$. **10.** a) $y = 0{,}085$; $\Delta_y = 0{,}0012$; $\delta_y = 1{,}4\%$; b) $y = 1{,}20$; $\Delta_y = 0{,}056$; $\delta_y = 4{,}7\%$; c) $y = 0{,}0552$; $\Delta_y = 0{,}00043$; $\delta_y = 0{,}77\%$; d) $y = 2{,}747$; $\Delta_y = 0{,}0090$; $\delta_y = 0{,}33$. **11.** a) $s = 0{,}594$; b) $s = 0{,}687$.

Capítulo II

1. a) $\begin{bmatrix} 6 & & 4 \\ 9 & 9 & 6 \\ 8 & 9 & 1 \end{bmatrix}$; b) $\begin{bmatrix} 1 & 5 & -5 \\ 3 & 10 & 0 \\ 2 & 9 & -7 \end{bmatrix}$. **2.** a) $\begin{bmatrix} 8 & -56 & 54 \\ -30 & -100 & 146 \\ 118 & -82 & 28 \end{bmatrix}$

b) $\begin{bmatrix} -72 & -72 & 78 \\ 36 & 54 & -6 \\ 66 & 240 & 88 \end{bmatrix}$. **3.** a) $\begin{bmatrix} 5 & 10 & -10 & 15 \\ 7 & 14 & -14 & 21 \\ -3 & -6 & 6 & -9 \\ 2 & 4 & -4 & 6 \end{bmatrix}$; b) $\begin{bmatrix} 4 & 5 \\ 8 & 10 \\ 12 & 15 \end{bmatrix}$; c) 409.

4. a) $\begin{bmatrix} -9 \\ 20 \\ -18 \end{bmatrix}$; b) $\begin{bmatrix} 8 \\ 2 \\ 2 \end{bmatrix}$. 5. a) 22; b) -26; c) 4279,1. 6. a) $A^{-1} = \begin{bmatrix} -12 & -1 & 8 \\ 9 & 1 & 6 \\ -7 & 1 & 5 \end{bmatrix}$;

b) $A^{-1} = \frac{1}{48} \begin{bmatrix} -4 & 10 & -46 & 24 \\ 10 & -1 & 67 & -36 \\ -14 & -13 & 7 & 12 \\ 0 & 0 & -96 & 48 \end{bmatrix}$; c) $A^{-1} = \frac{1}{24} \begin{bmatrix} 24 & 0 & 0 & 0 \\ 0 & 12 & 0 & 0 \\ -12 & 9 & 6 & 0 \\ -28 & -5 & 2 & 8 \end{bmatrix}$.

7. $AB = \begin{bmatrix} 24 & 57 & 15 & 31 \\ -4 & 16 & 6 & 11 \\ 16 & 27 & 7 & 14 \\ 10 & 53 & 13 & 29 \end{bmatrix}$. 8. a) $A^{-1} = \frac{1}{18} \begin{bmatrix} 1 & 2 & 3 & 2 \\ 2 & -1 & 2 & -3 \\ 3 & -2 & -1 & 2 \\ -2 & -3 & 2 & 1 \end{bmatrix}$; b) $A^{-1} =$

$= \frac{1}{18} \begin{bmatrix} 6 & -4 & 0 & 8 \\ -12 & 5 & 9 & 1 \\ -12 & 11 & 9 & -13 \\ -6 & -5 & 9 & 1 \end{bmatrix}$. **9.** a) $A = T_1 T_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 5 & 0 & 0 \\ 3 & 4 & 18/5 & 0 \\ 2 & -7 & 36/5 & 18 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 & -2 \\ 0 & 1 & 8/5 & -1/5 \\ 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 1 \end{bmatrix}$

$A^{-1} = T_2^{-1} T_1^{-1} = \begin{bmatrix} 1 & -2 & 1/5 & 2 \\ 0 & 1 & 8/5 & -3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2/5 & -1/5 & 0 & 0 \\ 7/18 & 4/18 & -5/18 & 0 \\ -2/18 & -3/18 & 2/18 & 1/18 \end{bmatrix} = \frac{1}{18} \times$

$$\times\begin{bmatrix} 1 & 2 & 3 & 2 \\ 2 & -1 & 2 & -3 \\ 3 & -2 & -1 & 2 \\ -2 & -3 & 2 & 1 \end{bmatrix};\ \text{b) } A=R_1R_2=\begin{bmatrix} 3 & 0 & 0 & 0 \\ 2 & 7/3 & 0 & 0 \\ 3 & 1 & 1/7 & 0 \\ 1 & 8/3 & -9/7 & 18 \end{bmatrix}\begin{bmatrix} 1 & -2/3 & 2/3 & 0 \\ 0 & 1 & -1/7 & -6/7 \\ 0 & 0 & 1 & 13 \\ 0 & 0 & 0 & 1 \end{bmatrix};$$

$$A^{-1}=R_2^{-1}R_1^{-1}=\begin{bmatrix} 1 & 2/3 & -4/7 & 8 \\ 0 & 1 & 1/7 & -1 \\ 0 & 0 & 1 & -13 \\ 0 & 0 & 0 & 1 \end{bmatrix}\begin{bmatrix} 1/3 & 0 & 0 & 0 \\ -2/7 & 3/7 & 0 & 0 \\ -5 & -3 & 7 & 0 \\ -1/3 & -5/18 & 9/18 & 1/18 \end{bmatrix}=\frac{1}{18}\times$$

$$\times\begin{bmatrix} 6 & -4 & 0 & 8 \\ -12 & 5 & 9 & -1 \\ -12 & 11 & 9 & -13 \\ -6 & -5 & 9 & 1 \end{bmatrix}.\ \textbf{10.}\ \text{a) } X=\begin{bmatrix} 3 & 2 & -1 \\ 0 & -1 & 2 \\ 5 & 7 & 1 \end{bmatrix};\ \text{b) } X=\begin{bmatrix} 0 & -1 & 2 \\ 1 & 0 & -2 \\ 3 & 1 & 2 \end{bmatrix}.$$

11. a) $r=2$; b) $r=2$. **12.** a) Linealmente dependiente; b) linealmente independiente. **13.** Forman la base, por ejemplo, los vectores $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_4$; $\mathbf{x}_3=\mathbf{x}_1-\mathbf{x}_2$. **14.** $\mathbf{y}=(5/4,\ 1/4;\ -1/4,\ -1/4)$.

Capítulo III

1. a) $x_1=-(11/7)x_3$, $x_2=-(1/7)x_3$; b) $x_1=(3x_3-13x_4)/17$, $x_2=(19x_3-20x_4)/17$. 2. a) Solución general: $x_1=(x_3-9x_4-2)/11$, $x_2=(-5x_3+x_4+10)/11$; solución particular: $x_1=-1$, $x_2=1$, $x_3=0$, $x_4=1$; b) solución general: $x_3=22x_1-33x_2-11$, $x_4=-16x_1+24x_2+8$; solución particular: $x_1=1$, $x_2=0$, $x_3=11$, $x_4=-8$. 3. a) $x_1=0$, $x_2=-1$, $x_3=2$; b) $x=2$, $y=-2$, $z=1$. 4. a) $x_1=1$, $x_2=-1$, $x_3=1$, $x_4=-1$; b) $x_1=-1$, $x_2=2$, $x_3=0$, $x_4=3$. 5. a) $x_1=1{,}120$; $x_2=-0{,}341$; $x_3=-0{,}008$; b) $x=0{,}008$; $y=-0{,}231$; $z=0{,}042$. 6. a) $d=88$; b) $d=2111{,}97$. 7. a) $A^{-1}=$

$$=\begin{bmatrix} 10/3 & -7/6 & 1/2 & -1/6 \\ -5/3 & 5/6 & -1/2 & -7/6 \\ 1 & -1/2 & 1/2 & 3/2 \\ 1 & -1/2 & 1/2 & 1/2 \end{bmatrix};\ \text{b) } A^{-1}=\begin{bmatrix} 1{,}19 & -0{,}31 & -0{,}82 & -0{,}12 \\ -0{,}17 & 1{,}57 & 1{,}23 & 0{,}70 \\ -1{,}75 & 0{,}11 & 0{,}30 & 0{,}87 \\ -0{,}12 & -2{,}92 & -1{,}09 & 0{,}17 \end{bmatrix}.$$

8. a) $x_1=-0{,}72$; $x_2=1{,}88$; $x_3=-0{,}92$; $x_4=-1{,}94$; b) $x=1{,}22$; $y=-0{,}67$; $z=0{,}35$. 10. a) $\|A\|_1=1{,}9$, $\|A\|_2=1{,}9$ $\|A\|_3=2{,}55$; b) $\|A\|_1=1{,}45$, $\|A\|_2=1{,}07$, $\|A\|_3=1{,}20$. 11. a) $x_1=1$, $x_2=-1$, $x_3=2$, $x_4=0$; b) $x_1=1/2$, $x_2=3/2$, $x_3=-1/2$, $x_4=-2$. 12. a) $x_1=-2$, $x_2=2$, $x_3=-3$, $x_4=3$; b) $x_1=1$, $x_2=1$, $x_3=-1$.

Capítulo IV

1. Resto $r=P_5(3)=430$; cociente $P_4^{(1)}=x^4+6x^3+16x^2+48x+143$. 2. Sí. 3. a) 0,423; b) 0,940; c) 1,386; d) 1,221; e) 0,809; f) 0,309. 4. a) 3,464; b) 7,483; c) 6,481.

Capítulo V

1. a) $-1{,}325$; b) 1,180; c) $-1{,}876$; 0,578; d) 0,781, 2,401; e) 0,0; 0,399; 6,352; f) 0,310; 4,0. 2. a) 1,213; b) 0,706; c) 0,841; d) $-0{,}438$; 0,438; e) 0,0; 0,787; f) 1,897. 3. a) $-4{,}071$; 0,468; 0,993; b) $-0{,}695$; $-3{,}067$; 3,757; c) $-3{,}523$; $-1{,}567$; 1,086; d) 0,398; 4,862; e) 0,0; 2,753; f) 0,739. 4. a) 0,760; b) $-2{,}258$; c) $-0{,}465$; d) $-0{,}567$; $-0{,}335$; 0,0; e) 3,473; f) 1,422. 5. a) $-0{,}532$; 0,653; 2,879; b) $-0{,}475$; 1,395; c) $-1{,}582$; 0,402; 1,373; d) $-1{,}453$; 1,164. 6. a) 0,187; b) 0,755; c) 0,739; d) 0,607; e) 0,672.

Capítulo VI

1. a) $\lambda^3-2\lambda^2-66\lambda+1$, b) $\lambda^4-4\lambda^3-6\lambda-41$, c) $\lambda^4-4\lambda^3+16\lambda^2-16$. 2. a) $\lambda_1=7$; $x_1=c\begin{bmatrix}1\\1\end{bmatrix}$; $\lambda_2=-2$, $x_2=c\begin{bmatrix}4\\-5\end{bmatrix}$; b) $\lambda_1=-2$, $x_1=c_1\begin{bmatrix}-2\\1\\0\end{bmatrix}+$ $+c_2\begin{bmatrix}1\\0\\1\end{bmatrix}$. 3. a) $\lambda^3+\lambda^2-3\lambda-7$; b) $\lambda^4+8\lambda^3+6\lambda^2+6\lambda-54$; c) $\lambda^4+\lambda^3-6x^2-$ -18λ. 4. a) $\lambda^4-7\lambda^2+15\lambda^2-2\lambda-34$; b) $\lambda^4+\lambda^3+7\lambda^2-20\lambda-54$. 6. $D(\lambda)=$ $=\lambda^3-3\lambda^2+3\lambda-1$, $\lambda_1=\lambda_2=\lambda_3=1$, $x^{(1)}=x^{(3)}=x^{(3)}=\begin{bmatrix}1\\1\\-1\end{bmatrix}$, si $e=\begin{bmatrix}1\\0\\0\end{bmatrix}$. 7. $\lambda_1\approx 4{,}46$; $\lambda_2=1{,}59$.

Capítulo VII

1. 20.819. 2. 24,680. 3. $L_3(x)=-(1/15)\,x^3-(3/20)\,x^2+(241/60)\,x-3{,}9$. 4. $L_2(x)=0{,}0735x^2-0{,}4530x+0{,}7474$; $\Delta_1=0{,}23\cdot 10^{-1}$. 7. 3,37215. 8. 4,379 9. 1,43612. 10. 0,75487. 24. a) $0{,}6115\pm 0{,}00013$; b) $0{,}9409\pm 0{,}0007$, c) $0{,}9456\pm$ $\pm 0{,}002$; d) $0{,}8007\pm 0{,}0009$; e) $0{,}3156\pm 0{,}00012$; f) $2{,}4505\pm 0{,}0018$. 25. $Q_2(x)=$ $=2-0{,}7\cos 2x-0{,}3\operatorname{sen} 2x-0{,}3\cos 4x+b_2\operatorname{sen} 4x$. 26. $Q_2(x)=-1+$ $+(7/3)\cos 2\pi x-(2/\sqrt{3})\operatorname{sen} 2\pi x-\cos 4\pi x$.

Capítulo VIII

1. 15,160. 2. 1,429. 3. 2,28. 4. 2,0[illegible]2. 5. 0,107250. 6. 0,67363. 7. 0,6931472; 8. 0,007. 9. 0,0087. 10. a) 0,239; b) 0,223; c) 1,0000; d) 0,8349; e) 1,4627; f) 1,5625; g) 1,333; h) 0,460; i) 1,718. 11. a) $0{,}754\pm 0{,}002$ (para $h_0=3°$); b) $-0{,}471\pm 0{,}002$ (para $h_0=7°$); c) $2{,}421\pm 0{,}07$ (para $h_0=1°$); d) $0{,}953\pm$ $\pm 0{,}002$ (para $h_0=6°$); e) $-0{,}892\pm 0{,}0015$ (para $h_0=3°$); f) $1{,}438\pm 0{,}003$ (para $h_0=3°$)

Capítulo IX

1. a) $y_3=1+8x^2+(56/3)\,x^3+18x^4+8x^5+(4/3)\,x^6$; b) $y_3=1-x+x^2-$ $-(1/3)\,x^3+(1/24)\,x^4$. 2. $y(x)=1+2x-0{,}7x^2-0{,}2567x^3+0{,}051x^4+0{,}00147x^5-$ $-0{,}00101x^6$. 3. $y=x^3/3+x^7/63+\ldots$ 4. a) $y_1=-1{,}1$; $y_2=-1{,}18$; $y_3=-1{,}238$; $y_4=-1{,}2718$; $y_5=-1{,}2790$; b) $y_1=0$, $y_2=0{,}01$, $y_3=0{,}0278$, $y_4=0{,}0524$, $y_5=0{,}08192$, $y_6=0{,}115536$, $y_7=0{,}152429$, $y_8=0{,}191943$, $y_9=0{,}233554$, $y_{10}=$ $=0{,}276844$. 5. $y_0=1$, $y_1=1{,}1836$, $y_2=1{,}3426$, $y_3=1{,}4850$, $y_4=1{,}6152$, $y_5=$ $=1{,}7362$. 6. $y_0=1$, $y_1=1{,}1867$, $y_2=1{,}3484$, $y_3=1{,}4938$, $y_4=1{,}6272$, $y_5=1{,}7542$. 7. a) $y_0=0$, $y_1=0{,}10536$, $y_2=0{,}223136$, $y_3=0{,}356601$, $y_4=0{,}510424$, $y_5=$ $=0{,}691497$, $y_6=0{,}910454$, $y_7=1{,}184648$, $y_8=1{,}54449$, $y_9=2{,}048721$, $y_{10}=$ $=2{,}827617$; b) $y_0=2{,}00$, $y_1=1{,}81$, $y_2=1{,}64$, $y_3=1{,}49$, $y_4=1{,}36$, $y_5=1{,}25$, $y_6=1{,}16$, $y_7=1{,}09$, $y_8=1{,}04$, $y_9=1{,}01$. 8. $y_4=0{,}8110$, $y_5=0{,}8196$, $y_6=0{,}8464$, $y_7=0{,}8898$, $y_8=0{,}9480$, $y_9=1{,}0197$, $y_{10}=1{,}1037$.

Capítulo X

1. a)

0,00	16,18	38,63	50,00
0,00	14,12	26,09	30,10
0,00	15,20	20,53	12,38
0,00	20,15	29,34	4,31

b)

0,00	17,88	39,92	50,00
0,00	15,18	36,39	30,10
0,00	16,37	21,26	12,38
0,00	29,05	29,63	4,31

2.

	9,81	19,78	29,12	40,16	42,31	
0,00	8,97	17,58	25,36	32,18	36,11	40,16
0,00	8,68	16,00	22,29	26,86	29,69	33,11
0,00	8,36	15,59	20,71	23,05	22,62	19,11
0,00	9,43	17,22	21,71	21,85	18,55	13,00
0,00	12,20	22,09	26,96	24,01	16,70	6,98
	17,28	31,96	40,00	30,50	17,28	

3.

	t	1	2	3	4	5	6	7	8	9
a)	0,000	0,343	0,672	0,970	1,213	1,375	1,423	1,327	1,062	0,618
	0,005	0,336	0,656	0,943	1,172	1,318	1,351	1,243	0,973	0,531
	0,010	0,328	0,639	0,914	1,131	1,262	1,281	1,162	0,887	0,486
	0,015	0,320	0,621	0,885	1,088	1,206	1,212	1,084	0,824	0,443
	0,020	0,311	0,602	0,855	1,045	1,150	1,145	1,018	0,764	0,412
b)	0,000	2,091	1,919	1,777	1,660	1,562	1,480	1,410	1,350	1,297
	0,017	2,097	1,924	1,781	1,663	1,564	8,482	1,411	1,351	1,298
	0,033	2,102	1,929	1,785	1,666	1,567	1,484	1,413	1,352	1,299
	0,050	2,106	1,934	1,789	1,670	1,570	1,486	1,415	1,354	1,300
	0,067	2,110	1,939	1,794	1,673	1,572	1,488	1,416	1,355	1,301

Índice alfabético de materias

A NUESTROS LECTORES:

Mir edita libros soviéticos traducidos al español, inglés, francés, árabe y otros idiomas extranjeros. Entre ellos figuran las mejores obras de las distintas ramas de la ciencia y la técnica, manuales para los centros de enseñanza superior y escuelas tecnológicas, literatura sobre ciencias naturales y médicas. También se incluyen monografías, libros de divulgación científica y ciencia ficción.

Dirijan sus opiniones a la Editorial Mir, 1 Rizhski per., 2, 129820, Moscú, I-110, GSP, URSS.