Capítulo IV

Métodos de diferencias de la resolución de los problemas de contorno para ecuaciones diferenciales ordinarias

§ 1. Conceptos fundamentales de la teoría de esquemas de diferencias

Un método numérico universal para resolver ecuaciones diferenciales es el de diferencias finitas. Antes de pasar a su exposición hace falta introducir ciertos conceptos fundamentales referentes a la teoría de esquemas de diferencias, a saber, aproximación, estabilidad y convergencia.

1. Operadores de diferencias más simples. Con el fin de obtener una ecuación en diferencias en lugar de la ecuación diferencial, es necesario:

— sustituir el dominio de variación continua del argumento por un conjunto discreto de puntos (por una red);

— sustituir (aproximar en la red) la ecuación diferencial por una ecuación en diferencias.

El problema sobre la resolución numérica de una ecuación diferencial se reduce a la cuestión de resolver las ecuaciones en diferencias. En los capítulos antecedentes ya se han expuesto los ejemplos de las redes:

1) red uniforme en el segmento $0 \leqslant x \leqslant 1$ de paso h: un conjunto de nodos $\bar{\omega}_h = \{x_i = ih,\ i = 0,\ 1,\ 2,\ \ldots,\ N,\ h = 1/N\}$; $x_0 = 0$, $x_N = 1$ son los nodos de frontera; $\omega_h = \{x_i = ih,\ i = 1,\ 2,\ \ldots,\ N\text{-}1\}$ es el conjunto de nodos interiores;

2) red no uniforme: el segmento $0 \leqslant x \leqslant 1$ se divide en N partes mediante puntos arbitrarios $x_1 < x_2 < \ldots < x_{N-1}$; $h_i = x_i - x_{i-1}$ es el paso de la red;

$$\bar{\omega}_h = \{x_i,\ i = 0,\ 1,\ \ldots,\ N,\ x_0 = 0,\ x_N = 1\},$$

$$\sum_{i=1}^{N} h_i = 1 \qquad \omega_h = \{x_i,\ 0 < i < N\};$$

3) red en un segmento $0 \leqslant t \leqslant T$: $\bar{\omega} = \{t_n = n\tau, n = 0, 1, \ldots, n_0; n_0\tau = T\}$.

En lugar de la función de argumento continuo (por ejemplo, en el segmento $0 \leqslant x \leqslant 1$) se estudia la función $y(x_i) = y_i$ de argumento discreto x_i, donde x_i es un nodo de la red ω_h, o de argumento i que es el número del nodo. Esta función se denomina *reticular*. Cualquier función reticular puede ser representada en forma de un vector

$$Y = y_0, y_1, \ldots, y_{N-1}, y_N).$$

Por eso, el conjunto de funciones reticulares forma un espacio de dimensión finita H cuya dimensión, en el caso dado, es $(N + 1)$. Se analiza corrientemente una familia de redes $\{\omega_h\}$ que dependen del paso como de un parámetro, razón por la cual las funciones reticulares $y = y_h(x)$ también dependen del paso como de un parámetro (o de N), si la red ω_h es uniforme. Si la red no es uniforme, entonces por h se entiende un vector $h = (h_1, h_2, \ldots, h_N)$. Resulta natural por esta razón dotar el espacio de funciones reticulares con el índice h y escribir H_h. En el espacio H_h podemos introducir una norma $\|\cdot\|_h$. He qauí los tipos más simples de las normas:

$$\|y\|_C = \max_{x \in \bar{\omega}_h} y(x) \text{ o bien } \|y\|_C = \max_{0 \leqslant i \leqslant N} |y_i|;$$

$$\|y\| = \Big(\sum_{i=1}^{N-1} y_i^2 h\Big)^{1/2}.$$

El operador diferencial se sustituye por un operador de diferencias que actúa en el espacio de funciones reticulares.

Sea G un dominio del espacio euclídeo R^p $(p = 1, 2, 3)$ con la frontera Γ. Por ejemplo, G es el intervalo $0 < x < 1$, Γ expresa los puntos $x = 0$, $x = 1$; G es un rectángulo $0 < x_1 < l_1$, $0 < x_2 < l_2$, $x = (x_1, x_2) \in G$ $(p = 2)$, Γ consta de los segmentos de las rectas $x_2 = 0$, $x_2 = l_2$, $x_1 = 0$, $x_1 = l_1$, etc. Sea dado un operador diferencial lineal L que actúa contra una función $v(x)$, $x \in G$. Introduzcamos en $G = \bar{G} \cup \Gamma$ una red $\bar{\omega}_h$ y veamos una función reticular $v_h(x)$, $x \in \omega_h$. Sustituyamos Lv en el punto $x_i \in \omega_h$ por una combinación lineal consistente de los valores $v_h(x)$ de la función reticular en cierto conjunto de nodos de la red

el cual se llamará *molde*

$$(L_h v)_l = \sum_{x_j \in \sigma_l} a_{lj}^h v_h(x_j), \quad x_l \in \omega_h(G), \tag{1}$$

donde a_{lj}^h son los coeficientes, σ_l es el molde, $\sigma_l \in \overline{\omega}_h$.

Tal sustitución de Lv por $L_h v$ se denomina *aproximación en la red* de un operador diferencial L mediante el operador de diferencias L_h, o bien *aproximación de diferencias* del operador L. El estudio de las aproximaciones de diferencias L_h del operador L se realiza, corrientemente, de un modo local, es decir, en cualquier punto fijo de la red. La construcción de L_h se debe empezar con la elección de un molde $\sigma(x)$, es decir, de un conjunto de nodos, vecinos con el nodo $x \in \omega_h$, en los cuales los valores de la función reticular $v_h(x)$ pueden ser empleados al escribir la expresión para L_h.

Veamos algunos ejemplos de construcción de L_h.

EJEMPLO 1. Derivada primera: $Lv = \frac{dv}{dx} = v'(x)$. Tomemos tres nodos $(x-h,\ x,\ x+h)$. Podemos servirnos de cualquier de las expresiones

$$L_h^+ v = \frac{v(x+h)-v(x)}{h} = v_x \quad \text{(el molde } (x,\ x+h)\text{)};$$

$$L_h^- v = \frac{v(x)-v(x-h)}{h} = v_{\bar{x}} \quad \text{(el molde } (x-h,\ x)\text{)};$$

$$L_h^0 v = \frac{v(x+h)-v(x-h)}{2h} = v_{\mathring{x}} \text{ (el molde } (x-h,\ x+h)\text{)}.$$

Se emplean a menudo las siguientes denominaciones: $L_h^+ v = v_x$ es la derivada de diferencias *derecha*; $L_h^- v = v_{\bar{x}}$, la derivada de diferencias *izquierda* y $L_h^0 v = v_{\mathring{x}} = \frac{1}{2}(L_h^+ v + L_h^- v)$, *derivada de diferencias central*. Sobre el molde tripuntual $(x-h,\ x,\ x+h)$ podemos definir un operador de diferencias

$$L_h^{(\sigma)}(v) = \sigma v_x + (1-\sigma) v_{\bar{x}},$$

donde σ es un parámetro real. De este modo, existe una infinidad de aproximaciones de diferencias de la primera derivada sobre el molde tripuntual.

Se denomina *error de aproximación del operador* L mediante el operador L_h una diferencia

$$\psi = L_h v - Lv.$$

Dicen que L_h *tiene el m-ésimo orden de aproximación en el punto* x, si

$\psi(x) = L_h v(x) - Lv(x) = O(h^m)$, o bien $|\psi(x)| \leqslant Mh^m$,

donde $M = \text{const} > 0$ no depende de h, $m > 0$.

Haciendo uso de la fórmula de Taylor

$$v(x \pm h) = v(x) \pm hv'(x) + \\ + \frac{h^2}{2} v''(x) \pm \frac{h^3}{6} v'''(x) + \frac{h^4}{24} v^{IV}(x) + O(h^5),$$

no será difícil obtener las estimaciones

$$v_x - v' = O(h), \quad v_{\bar{x}} - v' = O(h), \quad v_{\overset{\circ}{x}} - v' = O(h^2),$$

$$\psi^{(\sigma)} = L_h^{(\sigma)} v - Lv = O\left(\left(\sigma - \frac{1}{2}\right) h + h^2\right).$$

EJEMPLO 2. Derivada segunda: $Lv = \dfrac{d^2v}{dx^2} = v''(x)$.

Tomemos el mismo molde tripuntual que figuraba en el ejemplo 1 y escribamos un operador de diferencias

$$L_h v(x) = \frac{v(x+h) - 2v(x) + v(x-h)}{h^2}.$$

Al notar que $v(x + h) = v(x) + hx_x$, $v(x - h) = v(x) - hv_{\bar{x}}$, transformemos $L_h v(x)$:

$$L_h v(x) = \frac{v_x(x) - v_{\bar{x}}(x)}{h} = \frac{v_{\bar{x}}(x+h) - v_{\bar{x}}(x)}{h} = v_{\bar{x}x}(x). \quad (2)$$

Aprovechando la fórmula de Taylor para $v(x \pm h)$, encontramos

$$\psi = L_h v - Lv = \frac{h^2}{12} v^{IV}(x) + O(h^4) = O(h^2),$$

es decir, L_h tiene el segundo orden de aproximación.

Habitualmente se requiere la estimación del error de aproximación sobre una red, es decir, en cierta norma reticular $\|\cdot\|_h$. Se dice que L_h tiene el m-ésimo *orden de aproxima-*

ción sobre una red, siempre que

$$\|L_h v_h - (Lv)_h\|_h = O(h^m).$$

2. Esquema de diferencias. La ecuación diferencial $Lu = f(x)$ se resuelve, como regla, con ciertas condiciones complementarias: iniciales (problemas de Cauchy), de contorno (problemas de contorno) o bien condiciones iniciales y las de contorno a la vez. Dichas condiciones complementarias, pasando a las ecuaciones en diferencias, se deben también aproximar.

Sea dado un dominio G con la frontera Γ y supongamos que se busca la solución $u = u(x)$, $x \in \bar{G}$, de una ecuación diferencial lineal

$$Lu = f(x), \qquad x \in G, \tag{3}$$

con la siguiente condición complementaria en la frontera:

$$u(x) = \mu(x), \qquad x \in \Gamma. \tag{4}$$

Introduzcamos en el dominio $\bar{G} = G + \Gamma$ una red $\bar{\omega}_h = \omega_h + \gamma_h$, $\omega_h \in G$, $\gamma_h \in \Gamma$, y al problema (3), (4) le pondremos en correspondencia un problema de diferencias con el operador lineal L_h del tipo (1):

$$L_h y_h = \varphi_h(x), \quad x \in \omega_h; \qquad y_h(x) = \nu_h(x), \quad x \in \gamma_h. \tag{5}$$

Las funciones $y_h(x)$, $\varphi_h(x)$, $\nu_h(x)$ dependen del paso h de la red. Al variar h, obtenemos las sucesiones $\{y_h\}$, $\{\varphi_h\}$, $\{\nu_h\}$. De este modo, se examina no uno de los problemas de diferencias, sino una familia de problemas que depende del parámetro h. Esta familia de problemas lleva el nombre de *esquema de diferencias*.

EJEMPLO 1. Problema de Cauchy:

$$Lu = \frac{du}{dt} + \lambda u = f(t), \quad t > 0, \quad u(0) = u_0.$$

El esquema de diferencias de Euler tiene por expresión:

$$L_\tau y = \frac{y_{n+1} - y_n}{\tau} + \lambda y_n = f_n,$$

$$y_n = y(t_n), \quad t_n = n\tau \in \omega_\tau, \quad n = 0, 1, \ldots, y_0 = u_0.$$

EJEMPLO 2. Primer problema de contorno:

$$Lu = u'' = -f(x), \quad 0 < x < 1, \quad u(0) = \mu_1, \quad u(1) = \mu_2. \tag{6}$$

Hagamos uso del operador de diferencias tripuntual (2): $L_h y_i = y_{\bar{x}x,i} = (y_{i+1} - 2y_i + y_{i-1})/h^2$ y obtendremos un problema de contorno en diferencias sobre la red $\bar{\omega}_h = \{x_i = ih,\ 0 \leqslant i \leqslant N,\ x_N = 1\}$:

$$L_h y_i = y_{\bar{x}x,i} = -f_i, \quad i = 1, 2, \ldots, N-1, \quad y_0 = \mu_1,\ y_N = \mu_2. \tag{6'}$$

3. Estabilidad. Nos resulta más conveniente pasar a la notación del esquema de diferencias (5) en la forma operacional. Con este fin escribamos al principio la ecuación (5) en la forma matricial

$$AY_h = \Phi_h,$$

donde Y_h es el vector buscado de N-ésima dimensión finita, la que es igual al número de nodos de la red, en los cuales no son conocidos los valores de la función reticular y_h (para el primer problema de contorno (6') la dimensión Y_h es igual a N-1, es decir, al número de nodos interiores de la red). Los valores de $y_h(x_i)$ en los nodos $x_i \in \omega_h$ son componentes del vector Y_h, mientras que $\varphi_h(x_i)$ representan los componentes del vector Φ_h, y A es la matriz cuadrada de dimensión $N \times N$.

Introduzcamos un espacio N-dimensional H_h de funciones reticulares y sea A_h un operador lineal correspondiente a la matriz A: $A_h : H_h \to H_h$. En lugar de (7) podemos escribir

$$A_h y_h = \varphi_h, \qquad \varphi_h \in H_h. \tag{8}$$

Sean $\|\cdot\|_{(1_h)}$y $\|\cdot\|_{(2_h)}$ ciertas normas en el espacio H_h.

Diremos que el esquema de diferencias (8) es estable, si existe una constante $M > 0$ (y dicha constante no depende de h ni del modo de elegir φ_h) tal que para la solución y_h de la ecuación (8) tiene lugar la estimación

$$\|y_h\|_{(1_h)} \leqslant M \|\varphi_h\|_{(2_h)} \tag{9}$$

con todo h suficientemente pequeño: $|h| \leqslant h_0$.

El esquema de diferencias (8) se denomina *correcto (correctamente planteado)*, si la solución de la ecuación (8) existe y es única, cualesquiera que sean los datos de entrada de $\varphi_h \in H_h$, y si el esquema de diferencias es estable, es decir, queda cumplida la desigualdad (9).

La estabilidad del esquema significa una dependencia continua de la solución y_h de los datos de entrada, con la particularidad de que dicha dependencia continua es uniforme respecto de h. Si $\tilde{y}_h$ es una solución de la ecuación $A_h \tilde{y}_h = \tilde{\varphi}_h$, entonces $A_h(\tilde{y}_h - y_h) = \tilde{\varphi}_h - \varphi_h$ en virtud de la linealidad de A_h; en este caso, de (9) proviene

$$\| \tilde{y}_h - y_h \|_{(1_h)} \leqslant M \| \tilde{\varphi}_h - \varphi_h \|_{(2_h)}. \qquad (10)$$

A la variación pequeña de los datos de entrada le corresponde la variación pequeña de la solución.

Si el esquema (8) es resoluble, existe un operador inverso A_h^{-1} y

$$y_h = A_h^{-1}\varphi_h, \quad \| y_h \|_{(1_h)} \leqslant \| A_h^{-1} \| \| \varphi_h \|_{(2_h)}, \qquad (11)$$

donde $\| A_h^{-1} \| = \| A_h^{-1} \|_{(2_h \to 1_h)}$ es la norma del operador A_h^{-1}.

La estabilidad es un testimonio de que el operador inverso está acotado uniformemente respecto de h

$$\| A_h^{-1} \| \leqslant M. \qquad (12)$$

El esquema es *inestable* si no existe tal constante M *(no dependiente de h) que supere* $\|A_h^{-1}\|$, *es decir*, $\|A_h^{-1}\|$ crece indefinidamente cuando $|h| \to 0$.

Puede suceder que en lugar de la condición de contorno de la primera especie $u = \mu$ para $x \in \Gamma$ viene prefijada la condición

$$lu = \mu(x), \qquad x \in \Gamma, \qquad (13)$$

donde l es cierto operador diferencial lineal, por ejemplo, $lu = u' - \sigma u$, $\sigma > 0$, o bien $lu = u'$ para $x = 0$ ó para $x = 1$. Entonces, en vez del problema (3), (4) tenemos el siguiente

$$Lu = f(x), \quad x \in G; \quad lu = \mu(x), \quad x \in \Gamma. \qquad (14)$$

El esquema de diferencias correspondiente tendrá

$$L_h y_h = \varphi_h \text{ para } x \in \omega_h, \quad l_h y_h = \bar{\mu}_h \text{ para } x \in \gamma_h, \quad (15)$$

donde l_h es un operador de diferencias lineal que aproxima el operador l. Puede ocurrir, además, que φ_h y $\bar{\mu}_h$ han de ser estimadas en las normas diferentes $\|\varphi_h\|_{(2_h)}$, $\|\bar{\mu}_h\|_{(3_h)}$.

El esquema (15) es *estable* si para su solución y_h queda válida la estimación

$$\|y_h\|_{(1_h)} \leqslant M_1 \|\varphi_h\|_{(2_h)} + M_2 \|\mu_h\|_{(3_h)}, \quad (16)$$

donde $M_1 > 0$, $M_2 > 0$ son unas constantes que no dependen ni de h ni del modo de elegir los datos de entrade φ_h y $\bar{\mu}_h$.

Se ha de notar que el esquema de diferencias (15) también puede ser escrito en la forma operacional $A_h y_h = \varphi_h$, sin embargo, en este caso, $\|\cdot\|_{(2_h)}$ en (9) y (16) pueden diferir, al igual que los propios miembros segundos (lo que ya está claro para el primer problema de contorno).

4. Ejemplo de un esquema estable. A título de ejemplo de un esquema estable analicemos el siguiente problema de contorno en diferencias

$$y_{\bar{x}x,\,i} = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2} = -\varphi_i, \quad i = 1, 2, \ldots, N-1,$$
$$y_0 = 0, \qquad y_N = 0, \qquad hN = 1. \quad (17)$$

Siguiendo las indicaciones del § 4, cap. I, definamos el operador A_h. Sea H_h un espacio de funciones reticulares definidas en los nodos interiores ($i = 1, 2, \ldots, N-1$) de la red. Tomemos $y \in H_h$ (el índice h de $y_h(x)$ queda por ahora omitido) y una función $\overset{\circ}{y}_0 = \overset{\circ}{y}_N = 0$. Entonces, el operador A_h se determina con ayuda de la identidad

$$(A_h y)_i = -\overset{\circ}{y}_{\bar{x}x,\,i}, \qquad i = 1, 2, \ldots, N-1,$$

y en lugar de (17) se obtiene una ecuación operacional

$$A_h y_h = \varphi_h. \quad (18)$$

En el espacio H_h introducimos el producto escalar

$$(y, v) = \sum_{i=1}^{N-1} y_i v_i h.$$

El operador A_h en H_h es autoconjugado y definido positivo y

$$\delta E \leqslant A_h \leqslant \Delta E, \text{ o bien } \delta \| y \|^2 \leqslant (A_h y, y) \leqslant \Delta \|y\|^2$$
$$\text{para todo } y \in H_h, \qquad (19)$$

donde δ y Δ son los valores propios mínimo y máximo, respectivamente, del operador A:

$$\delta = \frac{4}{h^2} \operatorname{sen}^2 \frac{\pi h}{2}, \qquad \Delta = \| A_h \| = \frac{4}{h^2} \cos^2 \frac{\pi h}{2}. \qquad (20)$$

El operador inverso A_h^{-1} es autoconjugado si $A_h = A_h^*$. En el § 4, cap I, se ha mostrado que las desigualdades (19) son equivalentes a las desigualdades operacionales

$$\frac{1}{\Delta} E \leqslant A_h^{-1} \leqslant \frac{1}{\delta} E, \quad \| A_h^{-1} \| = \frac{1}{\delta}. \qquad (21)$$

De aquí se deducen la acotación uniforme de la norma del operador inverso A_h^{-1} : $\|A_h^{-1}\| \leqslant 1/\delta < 1/8$ y la estimación apriorística

$$\| y_h \| \leqslant \frac{1}{\delta} \| \varphi_h \| \leqslant \frac{1}{8} \| \varphi_h \|, \qquad (22)$$

que es indicio de estabilidad del esquema (18). Esta estimación puede obtenerse por el método de desigualdades energéticas, sin recurrir a la estimación de los valores propios λ_k (A_h^{-1}). En efecto, multipliquemos la ecuación $A_h y_h = \varphi_h$ escalarmente por y_h: $(A_h y_h, y_h) = (\varphi_h, y_h)$ y aprovechemos las desigualdades $(\varphi_h, y_h) \leqslant \| \varphi_h \| \| y_h \|$, $\| y_h \|^2 \leqslant \frac{1}{\delta} (A_h y_h, y_h)$; entonces obtendremos la desigualdad $\delta \| y_h \|^2 \leqslant \| \varphi_h \| \| y_h \|$, de donde se deduce precisamente la estimación (22).

El esquema (17) es estable también en la norma $\| y \|_C$:

$$\| y_h \|_C \leqslant \frac{1}{2} \| \varphi_h \|_C, \quad \| y \|_C = \| y \|_{C_h} = \max_{0 < i < N} |y_i|. \qquad (23)$$

Esto proviene de la estimación de la solución del problema de controrno en diferencia tripuntual, obtenido en el p. 3 del § 5, cap. I. En el caso dado la estimación tiene por

expresión

$$\| y_h \|_C \leqslant \sum_{s=1}^{N-1} h \sum_{k=1}^{s} h |\varphi_h| \leqslant \| \varphi \|_C \sum_{s=1}^{N} x_s h < \frac{1}{2} \| \varphi_h \|_C,$$

puesto que

$$\sum_{s=1}^{N} x_s h = h^2 \sum_{s=1}^{N} s = \frac{N(N-1)}{2} h_2 = \frac{1-h}{2} < \frac{1}{2}.$$

5. Ejemplo de un esquema no correcto. Sea dado un esquema

$$A_h y_h = \varphi_h$$

y $\| A_h \| \to \infty$ cuando $| h | \to 0$. Veamos un problema inverso: determinar el segundo miembro φ_h por la solución conocida y_h:

$$B_h \varphi_h = y_h, \qquad B_h = A_h^{-1}.$$

El problema no es correctamente planteado, puesto que

$$\| B_h^{-1} \| = (A_h^{-1})^{-1} \| = \| A_h \| \to \infty \text{ cuando } | h | \to 0.$$

Esto significa que para cualquier constante M que no dependa de h puede indicarse tal h_* que $\| B_h^{-1} \| > M$ cuando $| h | \leqslant | h_* |$. Sea $\widetilde{\varphi}_h$ la solución de la ecuación $B_h \varphi_h = \widetilde{y}_h$, y sea φ_h la solución de la ecuación $B_h \varphi_h = y_h$, entonces

$$\| \widetilde{\varphi}_h - \varphi_h \| \leqslant \| B_h^{-1} \| \| \widetilde{y}_h - y_h \|.$$

Si, en cambio,

$$\| B_h^{-1} \| \leqslant M \text{ para } | h | \geqslant h_0,$$

de modo que se verifica la desigualdad

$$\| \widetilde{\varphi}_h - \varphi_h \| \leqslant M \| \widetilde{y}_h - y_h \|,$$

diremos que el esquema es *casi estable*. ¿Se podrá aplicar este esquema para determinar φ_h con la exactitud requerida ε, si y_h viene prefijada con cierta exactitud ε_0:

$$\| \widetilde{y}_h - y_h \| \leqslant \varepsilon?$$

De la desigualdad $\| \widetilde{\varphi}_h - \varphi_h \| \leqslant \| B_h^{-1} \| \| \widetilde{y}_h - y_h \|$ se desprende que la solución del problema $B_h \varphi_h = y_h$ se determina con la exactitud $\| B_h^{-1} \| \varepsilon_0$. Supongamos que se pide hallar φ_h con la exactitud $\varepsilon > 0$, de suerte que $\| \widetilde{\varphi}_h - \varphi_h \| \leqslant \varepsilon$; esto es posible bajo la condición

$$\| B_h^{-1} \| \cdot \varepsilon_0 \leqslant \varepsilon.$$

De aquí determinamos el paso admisible $h \geqslant h_0$, es decir, h_0.

Expliquemos esto con el problema concreto (17). Para dicho problema tenemos

$$\| B_h^{-1} \| = \| A_h \| = \Delta = \frac{4}{h^2} \cos^2 \frac{\pi h}{2} \leqslant \frac{4}{h^2},$$

y la condición $\| B_h^{-1} \| \varepsilon_0 = \Delta \varepsilon_0 \leqslant \varepsilon$ queda cumplida si $4\varepsilon_0/h^2 \leqslant \varepsilon$, o bien

$$h \geqslant h_0 = 2\sqrt{\varepsilon_0/\varepsilon}.$$

De aquí se ve que la precisión con la que se dan los datos de entrada ε_0 ha de ser *más alta que la exactitud* ε *con la* que se determina la solución.

Por ejemplo, sean prefijados el error del segundo miembro $\varepsilon_0 = 10^{-8}$ y la exactitud requerida $\varepsilon = 10^{-4}$. Entonces, $h_0 = 2 \cdot 10^{-2} = 1/50$, es decir, la exactitud $\varepsilon = 10^{-4}$ puede obtenerse sólo sobre una red de paso $h \geqslant 1/50$. *Si en cambio*, por ejemplo, $\varepsilon_0 = \frac{1}{4} \cdot 10^{-4}$, $\varepsilon = 10^{-4}$, entonces $h_0 = 1$ y la exactitud $\varepsilon = 10^{-4}$ no se conseguirá en ninguna red para tal precisión de prefijar los datos de entrada.

6. Aproximación y convergenica. Al resolver el problema (14) por el método de diferencias se debe saber con qué exactitud la resolución del problema de diferencias aproxima la solución del problema inicial. Con el fin de estimar el error obtenido al sustituir (14) por el esquema de difrencias (15), es necesario comparar las soluciones de estos problemas. La comparación se realizará en el espacio H_h de funciones reticulares. Denotemos con $u_h(x)$ los valores de las funciones $u(x)$ (soluciones exactas del problema (14)) sobre la red ω_h: $u_h \in H_h$. Veamos un error

$$z_h = y_h - u_h,$$

donde y_h es la solución del problema (15). Al sustitutir $y_h = z_h + u_h$ en (15) y al tomar $u = u(x)$ por la función prefijada, obtendremos para z_h un problema de diferencias

$$L_h z_h = \psi_h, \quad x \in \omega_h; \quad l_h z_h = \nu_h, \quad x \in \gamma_h, \qquad (24)$$

donde $\psi_h = \varphi_h - L_h u_h$ se denomina *error de aproximación para la ecuación* $L_h y_h = \varphi_h$ *en la solución* $u = u(x)$ *de la ecuación* $Lu = f(x)$ (*residuo para el esquema de diferencias en la solución*); $\nu_h = \mu_h - l_h u_h$ recibe el nombre de error de aproximación para la condición de contorno de diferencias $l_h y_h = \overline{\mu_h}$ en la solución del problema (14).

Diremos que:

el esquema de diferencias (15) *converge*, si

$$\| z_h \|_{(\Omega_h)} \to 0 \text{ para } | h | \to 0;$$

el esquema de diferencias (15) tiene *exactitud* de m-ésimo orden o *converge con la velocidad* $O(|h|^m)$, si

$$\| z_h \|_{(\Omega_h)} = \| y_h - u_h \|_{(\Omega_h)} \leqslant M | h |^m$$

o

$$\| z_h \|_{(\Omega_h)} = O(| h |^m), \qquad m > 0,$$

donde $M > 0$ es una constante no dependiente de h.

El esquema de diferencias (15) tiene el m-ésimo *orden de aproximación en la solución*, si

$$\| \psi_h \|_{(2h)} = O(| h |^m), \ \| \nu_h \|_{(3_h)} = O(| h |^m), \quad m > 0. \qquad (25)$$

La estimación de los residuos ψ_h y ν_h se realiza bajo el supuesto de que la solución del problema inicial existe y tiene tantas derivadas cuanto es necesario al obtener el m-ésimo orden de aproximación.

Demos a conocer dos ejemplos de estimar ψ_h.

EJEMPLOS. 1. Hay un problema

$$L_h y = -y_{\overline{x}x} = \varphi(x), \quad x = ih, \quad 1 \leqslant i \leqslant N - 1,$$
$$y_0 = y_N = 0, \qquad (26)$$
$$Lu = -u'' = f(x), \quad 0 < x < 1, \quad u(0) = u(1) = 0.$$

En este caso las condiciones de contorno se satisfacen con toda la exactutud, $v_h = 0$ (el índice h de $\varphi(x)$, $u(x)$ por ahora queda omitido) y

$$\psi_h = \varphi - L_h u = \varphi + u_{\bar{x}x} = \varphi + \left(u'' + \frac{1}{2}h^2 u^{IV} + O(h^4)\right) =$$

$$= (\varphi + u'') + \frac{h^2}{12}u^{IV} + O(h^4) = \varphi - f + O(h^2),$$

puesto que $u'' = -f(x)$. De aquí se ve que $\|\psi_h\|_C = O(h^2)$, si ponemos $\varphi = f$, o bien $\varphi = f + O(h^2)$.

En el p. 1 fue estimado el error $\psi = L_h v_h - (Lv)_h$ para una función arbitraria. En la estimación del error $z_h = y_h - u_h$ se une el residuo ψ_h que caracteriza el error de aproximación del operador $Lu - f$ mediante el operador $L_h u_h - \varphi_h$ en la solución $u = u(x)$ del problema inicial. Al tomar en consideración que $f - Lu = 0$, representemos $\psi_h = \varphi_h - L_h u_h$ en la forma

$$\psi_h = (\varphi_h - L_h u_h) - (f - Lu)_h =$$

$$= (\varphi_h - f_h) - (L_h u_h - (Lu)_h) = \psi_h^{(1)} + \psi_h^{(2)},$$

donde $\psi_h^{(1)} = -(L_h u_h - (Lu)_h)$, $\psi_h^{(2)} = \varphi_h - f_h$; $\psi_h^{(1)}$ es el error de aproximación de L por el operador L_h en la solución $u = u(x)$ del problema (6), $\psi_h^{(2)}$ es el error de aproximación del segundo miembro de la ecuación. La exigencia $\|\psi_h\|_{(2_h)} = O(|h|^m)$ se cumple, evidentemente, si $\|\psi_h^{(1)}\|_{(2_h)} = O(|h|^m)$, $\|\psi_h^{(2)}\|_{(2_h)} = O(|h|^m)$. No obstante, estas condiciones no son necesarias para la estimación de $\|\psi_h\|_{(2_h)} = O(|h|^m)$, lo que atestigua el siguiente ejemplo.

2. Primer problema de contorno (6). Calculemos

$$-\psi_h^{(1)} = u_{\bar{x}x} - u'' = \frac{1}{12}h^2 u^{IV} + O(h^4) = O(h^2).$$

Sea $\varphi = f + \frac{1}{12}h^2 f_{\bar{x}x}$, es decir, $\varphi - f = O(h^2)$. De aquí se ve que $\psi_h^{(1)} = O(h^2)$ y $\psi_h^{(2)} = O(h^2)$, sin embargo, el esquema tiene el cuarto orden de aproximación, puesto que

$$\psi_h = \psi_h^{(1)} + \psi_h^{(2)} = \varphi - f + \frac{h^2}{12}u^{IV} + O(h^4) =$$

$$= \frac{h^2}{12}(f_{\bar{x}x} + u^{IV}) + O(h^4) = \frac{h^2}{12}(f'' + u^{IV}) + O(h^4),$$

$\psi_h = O(h^4)$, dado que $u^{IV} + f''(x) = 0$ en virtud de la ecuación $u'' + f(x) = 0$.

7. Relación de la estabilidad y aproximación con la convergencia. Examinemos un esquema de diferencia lineal (15). Si el esquema es estable y aproxima el problema inicial, será convergente (se dice corrientemente: «de la estabilidad y aproximación proviene la convergencia del esquema»). En efecto, para el error $z_h = y_h - u_h$ obtenemos, en virtud de la linealidad de L_h y l_h, el problema (24) que es análogo al problema (15) para y_h. Por eso, si el esquema (15) es estable, es decir, si es justa la estimación (16), entonces para z_h será válida la estimación

$$\| z_h \|_{(1_h)} \leqslant M_1 \| \psi_h \|_{(2_h)} + M_2 \| \nu_h \|_{(3_h)}. \tag{27}$$

De aquí se deduce que

$$\| z_h \|_{(1_h)} = \| y_h - u_h \|_{(1_h)} = O(|h|^m),$$

siempre que

$$\| \psi_h \|_{(2_h)} = O(|h|^m), \qquad \| \nu_h \|_{(3_h)} = O(|h|^m).$$

De este modo, el estudio de la convergencia y del orden de exactitud de los esquemas de diferencias se reduce al estudio del error de aproximación y de la estabilidad, es decir, a la obtención de las estimaciones apriorísticas (16).

EJEMPLO. Para el esquema de diferencias (17) ($y_{\bar{x}x,i} = -\varphi_i$, $i = 1, 2, \ldots, N-1$, $y_0 = 0$, $y_N = 0$) se ha obtenido anteriormente la estimación (23). El error de aproximación es evidentemente, $\| \psi_h \|_{C_h} = O(h^2)$ para $\varphi_i = f_i$, $\| \psi_h \|_{C_h} \leqslant O(h^4)$ para $\varphi_i = f_i + \frac{h^2}{12} f_{\bar{x}x,i}$. Por cuanto $z_{\bar{x}x,i} = -\psi_{h,i}$ para $i = 1, 2, \ldots, N-1$, $z_0 = 0$, $z_N = 0$, entonces para z será también válida la estimación $\| z \|_C \leqslant \frac{1}{2} \| \psi \|_C$, de donde se desprende $\| y_h - u_h \|_C = O(h^m)$, donde $m = 2$ para $\varphi = f$, $m = 4$ para $\varphi = f + \frac{h^2}{12} f_{\bar{x}x}$.

Con ello se da por terminado el estudio del esquema (26) (el estudio del esquema (26) se ha ilustrado, de hecho, con los tres últimos ejemplos). Todo lo expuesto más arriba sirve de ejemplo típico de cómo se realiza el estudio de los esquemas de diferencias.

§ 2. Esquemas de diferencias homogéneos tripuntuales

1. Problema de partida. Consideremos el primer problema de contorno para una ecuación diferencial ordinaria de segundo orden:

$$Lu = \frac{d}{dx}\left(k(x)\frac{du}{dx}\right) - q(x)u = -f(x), \quad 0 < x < 1,$$
$$k(x) \geqslant q > 0, \quad q(x) \geqslant 0, \quad u(0) = \mu_1, \quad u(1) = \mu_2. \tag{1}$$

Una ecuación de este tipo describe la distribución estacionaria de temperatura, es decir, una distribución que no varía en tiempo (ecuación estacionaria de conductibilidad térmica), o bien la distribución de concentración (ecuación de difusión). Si $u = u(x)$ es la temperatura, entonces $W(x) = -k(x)\frac{du}{dx}$ es un flujo térmico ($k(x)$ es el coeficiente de conductibilidad térmica).

El problema (1) tiene una solución única, si $k(x)$, $q(x)$, $f(x)$ son funciones continuas a trozos. Si $k(x)$ tiene una discontinuidad de primera especie en el punto $x = \xi$, de modo que $[k] = k(\xi + 0) - k(\xi - 0) \neq 0$, en dicho punto deben ser continuos tanto la temperatura u como el flujo térmico $-(ku')$:

$$[u] = 0, \quad [ku'] = 0 \quad \text{para} \quad x = \xi.$$

Son posibles también otras condiciones de contorno para $x = 0$, $x = 1$: $ku' = \sigma_1 u - \mu_1$ para $x = 0$, $-ku' = \sigma_2 u - \mu_2$ para $x = 1$. Si $\sigma_1 > 0$, entonces la citada condición es de tercera especie; cuando $\sigma_1 = 0$, tenemos la condición de segunda especie ($ku' = -\mu_1$ para $x = 0$). Son posibles combinaciones de diferentes condiciones para $x = 0$ y $x = 1$.

2. Esquemas de diferencias tripuntuales. Introduzcamos en el segmento $0 \leqslant x \leqslant 1$ una red uniforme $\omega_h = \{x_i = ih,\ i = 0, 1, \ldots, N\}$ de paso $h = 1/N$ y elijamos un molde tripuntual (x_{i-1}, x_i, x_{i+1}), en el cual escribiremos el esquema de diferencias que aproxima el problema (1). Cualquier ecuación en diferencias en este molde tendrá por expresión

$$b_i y_{i+1} - c_i y_i + a_i y_{i-1} = -h^2 \varphi_i, \tag{2}$$

donde a_i, b_i, c_i son los coeficientes dependientes de $k(x)$, $q(x)$ y h. Estos coeficientes son, por ahora, desconocidos. Escribamos (2) de otra forma:

$$\frac{1}{h}\left(b_i\frac{y_{i+1}-y_i}{h}-a_i\frac{y_i-y_{i-1}}{h}\right)-d_iy_i=-\varphi_i, \qquad (3)$$

$$d_i=(c_i-a_i-b_i)/h^2.$$

Diremos que un esquema de diferencias es *homogéneo*, si sus coeficientes en todos los nodos de la red para cualesquiera coeficientes de la ecuación diferencial se calculan según unas mismas fórmulas. Así, por ejemplo, si introducimos las funcionales $A[\overline{k}(s)]$, $B[\overline{k}(s)]$, $D[\overline{k}(s)]$, $F(\overline{f}(s))$, definidas para cualesquiera funciones continuas a trozos sobre el segmento $-1\leqslant s\leqslant 1$, y calculamos los coeficientes del esquema (3) por las fórmulas

$$a_i=A[k(x_i+sh)], \qquad b_i=B[k(x_i+sh)],$$

$$d_i=D[k(x_i+sh)],\ \varphi_i=F[f(x_i+sh)],\ \overline{k}(s)=k(x_i+sh),$$

entonces tal esquema será homogéneo. He aquí las funcionales más simples

$$A[\overline{k}(s)]=\overline{k}(-0{,}5), \qquad a_i=k_{i-1/2}=k(x_i-0{,}5h),$$

$$F(\overline{f}(s))=f(0), \qquad \varphi_i=f_i=f(x_i), \text{ etc.}$$

Si un esquema es homogéneo, resulta más cómodo servirse del sistema de designaciones sin índices:

$$\Lambda y=\frac{1}{h}(by_x-ay_{\overline{x}})-dy=-\varphi, \qquad x\in\omega_h,$$

$$y(0)=\mu_1, \qquad y(1)=\mu_2, \qquad (4)$$

donde

$$a=a(x), \quad b=b(x), \quad y=y(x), \quad x=ih\in\omega_h,$$

$$y_x=(y(x+h)-y(x))/h, \quad y_{\overline{x}}=(y(x)-y(x-h))/h.$$

Para que el problema (4) sea resoluble, es suficiente que sea $a>0$, $b>0$, $d\geqslant 0$, y en este caso la solución puede ser determinada por el método de factorización (véase el cap. I, § 3).

3. Condiciones de aproximación. Calculemos el error de aproximación del esquema (4):

$$\psi = (\Lambda v + \varphi) - (Lv + f) = (\Lambda v - Lv) + (\varphi - f) =$$

$$= \left[\frac{1}{h}(bv_x - av_{\bar{x}}) - (kv')'\right] - (d - q)v + (\varphi - f),$$

donde $v(x)$ es una función arbitraria suficientemente suave; k, q, f cuentan con un número de derivadas necesarias en el transcurso de la exposición. Hegamos uso de la fórmula de Taylor:

$$v(x \pm h) = v(x) \pm hv'(x) + \frac{h^2}{2}v''(x) \pm \frac{h^3}{6}v'''(x) + O(h^4)$$

y hallemos

$$v_x = v' + \frac{h}{2}v'' + \frac{h^2}{6}v''' + O(h^3),$$

$$v_{\bar{x}} = v' - \frac{h}{2}v'' + \frac{h^2}{6}v''' + O(h^3).$$

Sustituyamos estas expresiones para v_x y $v_{\bar{x}}$ en la fórmula para ψ:

$$\psi = \left(\frac{1}{h}(b-a) - k'\right)v' + \left(\frac{b+a}{2} - k\right)v'' +$$

$$+ \frac{h(b-a)}{6}v''' - (d-q)v + (\varphi - f) + O(h^2).$$

De aquí se ve que el esquema tiene el segundo orden de aproximación si quedan cumplidas las condiciones

$$\frac{b-a}{2} = k'(x) + O(h^2), \quad \frac{b+a}{2} = k(x) + O(h^2),$$

$$d = q(x) + O(h^2), \quad \varphi = f(x) + O(h^2). \qquad (5)$$

En este caso $\psi = O(h^2)$.

El esquema (4) con los coeficientes

$$b_i = k_{i+1/2}, \quad a_i = k_{i-1/2}, \quad d_i = q_i, \quad \varphi_i = f_i,$$

$$b_1 = \frac{k_1 + 2k_{i+1/2} + k_{i+1}}{4}, \quad a_i = k_{i-1/2}, \quad d_i - q_i \quad \varphi_i = f_i,$$

satisface las condiciones (5) del segundo orden de aproximación, mientras que el esquema con los coeficientes

$$b_i = k_{i+1}, \quad a_i = \frac{k_i + k_{i+1}}{2}$$

no satisface ni siquiera la condición del primer orden de aproximación, puesto que

$$\frac{1}{h}(b_i - a_i) - k_i' = O(1).$$

§ 3. Esquemas de diferencias conservativos

1. Esquemas conservativos homogéneos. En el § 4, cap. 1 fue establecido que la condición necesaria y suficiente para que un operador de diferencias Λy sea autoconjugado (la matriz sea simétrica) consiste en que $b_i = a_{i+1}$. En este caso el problema (2) del § 2 adquiere la forma

$$\Lambda y = \frac{1}{h}\left[a_{i+1}\frac{y_{i+1}-y_i}{h} - a_i\frac{y_i - y_{i-1}}{h}\right] - d_i y_i = -\varphi_i,$$
$$i = 1, 2, \ldots, N-1, \quad y_0 = \mu_1, \quad y_N = \mu_2. \qquad (1)$$

La ecuación

$$a_{i+1}\frac{y_{i+1}-y_i}{h} - a_i\frac{y_i - y_{i-1}}{h} - hd_i y_i = -h\varphi_i \qquad (2)$$

es un análogo reticular de la ecuación de balance del calor sobre el intervalo $(x_{-1,2}, x_{i+1,2})$:

$$w_{i+1/2} - w_{i-1/2} - \int_{x_{i-1/2}}^{x_{i+1/2}} qu\,dx = -\int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,dx, \quad w = ku',$$

(que se obtiene integrando la ecuación (1) del § 2 a lo largo del segmento $x_{i-1,2} \leqslant x \leqslant x_{i+1,2}$ y lleva el nombre de esquema *conservativo*, es decir, esquema para el cual se cumplen los análogos de diferencias de las leyes físicas de conservación.

El requisito $b_i = a_{i+1}$ para un esquema homogéneo significa que $B[k(x + sh)] = A[k(x + (s+1)/h]$, o bien, $B[\overline{k}(s)] = A[\overline{k}(s+1)]$ para cualesquiera funciones continuas a trozos $\overline{k}(s)$ en el segmento $[-1, 1]$. Esto es posi-

ble sólo cuando la funcional $A\,[\overline{k}\,(s)]$ no depende de los valores de $\overline{k}\,(s)$ para $0 \leqslant s \leqslant 1$, y $B\,[\overline{k}\,(s)]$ no depende de los valores de $\overline{k}\,(s)$ para $-1 \leqslant s \leqslant 0$, de modo que $a\,(x) = A[k(x+sh)]$ para $-1 \leqslant s \leqslant 0$. El coeficiente $a\,(x)$ del esquema conservativo depende sólo de los valores de $k\,(x)$ en el segmento $[x-h,\ x]$. Las condiciones del segundo orden de aproximación (5) del § 2 toman, para el esquema conservativo (2), la forma siguiente

$$\frac{a\,(x+h)-a\,(x)}{h} = k'\,(x) + O\,(h^2).$$
$$\frac{a\,(x+h)+a\,(x)}{2} = k\,(x) + O\,(h^2), \tag{3}$$

$$d\,(x) = q\,(x) + O\,(h^2), \quad \varphi\,(x) = f\,(x) + O\,(h^2). \tag{4}$$

De aquí, en particular, proviene que

$$a\,(x) = k\,(x) - {}^1/_2 hk'\,(x) + O\,(h^2) = k\,(x - {}^1/_2 h) + O\,(h^2).$$

Escribamos el esquema conservativo (2) utilizando las designaciones sin índices:

$$(ay_{\overline{x}})_x - d\,(x)\,y = -\varphi\,(x),$$
$$x = ih \in \omega_h, \quad y\,(0) = \mu_1, \quad y\,(1) = \mu_2. \tag{5}$$

Exigiremos que se cumplan también las condiciones

$$a \geqslant c_1 > 0, \qquad d \geqslant 0. \tag{6}$$

En la práctica se deben emplear las fórmulas sencillas para a, d y φ, por ejemplo, $a_i = k_{i-1,2}$, $d_i = q_i$, $\varphi_i = f_i$.

Si la discontinuidad de la función $k\,(x)$ se halla dentro del nodo $x = x_i$ de la red, calculemos los coeficientes del esquema homogéneo:

$$a_i = k_{i-1/2} \quad \text{o bien} \quad a_i = {}^1/_2\,(k\,(x_{i-1}+0) + k\,(x_i - 0)),$$
$$d_i = {}^1/_2\,(q\,(x_i - 0) + q\,(x_i + 0)), \quad \varphi_i = {}^1/_2\,(f\,(x_i - 0) + f\,(x_i + 0)).$$

En este caso las condiciones (3) se cumplen en todo punto, mientras que las condiciones (4) se sustituyen por las condiciones

$$d_i - {}^1/_2\,(q_{i-0} + q_{i+0}) = O\,(h^2), \qquad \varphi_i - {}^1/_2\,(f_{i-0} + f_{i+0}) = O\,(h^2).$$

Demos a conocer los ejemplos de un esquema cuyos coeficientes se calculan por integración en los intervalos de la red:

$$a_i = \left(\frac{1}{h}\int\limits_{x_{i-1}}^{x_i} \frac{dx}{k(x)}\right)^{-1} = \left(\int\limits_{-1}^{0} \frac{ds}{k(x_i+sh)}\right)^{-1};$$

$$\varphi_i = \frac{1}{h}\int\limits_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,dx = \int\limits_{-1/2}^{1/2} f(x_i+sh)\,ds,$$

$$d_i = \frac{1}{h}\int\limits_{x_{i-1/2}}^{x_{i+1/2}} g(x)\,dx = \int\limits_{-1/2}^{1/2} g(x_i+sh)\,ds.$$

Es evidente, pues, que las condiciones (3), (4) se cumplen.

2. Error de aproximación. Veamos un esquema conservativo de segundo orden de la aproximación. Sea $u = u(x)$ la solución exacta del problema

$$Lu = (ku')' - q(x)u = -f(x), \quad 0 < x < 1,$$
$$u(0) = \mu_1, \quad u(1) = \mu_2, \qquad (6)$$

y sea $y_i = y(x_i)$ la solución del problema de contorno en diferencias (5). Analicemos el error del esquema, es decir, una función reticular

$$z(x) = y(x) - u(x), \qquad x \in \bar{\omega}_h.$$

Al sustituir $y(x) = z(x) = u(x)$ en la ecuación (5) y al suponer que $u(x)$ es la función dada, obtendremos para el error $z(x)$ un problema

$$\Lambda z = (az_{\bar{x}})_x - dz = -\psi(x), \quad x \in \omega_h,$$
$$z(0) = 0, \quad z(1) = 0, \quad a \geqslant c_1 > 0, \quad d \geqslant 0, \qquad (6')$$

donde $\psi(x) = \Lambda u + \varphi(x) = (au_{\bar{x}})_x - du + \varphi$ es el residuo del esquema (5) en la solución $u = u(x)$ del problema diferencial de partida.

Teniendo presente que $Lu + f = 0$, escribamos

$$\psi = (\Lambda u + \varphi) - (Lu + f) = (\Lambda u - Lu) + (\varphi - f) =$$
$$= [(au_{\bar{x}})_x - (ku')'] - (d - q)\, u + (\varphi - f).$$

Por hipótesis, el esquema (5) satisface las condiciones del segundo orden de la aproximación. Esto significa que $\psi = O(h^2)$, si $k \in C^{(3)}$, $q, f \in C^{(2)}$, $u \in C^{(4)}$, y, por lo tanto,

$$\| \psi \|_C = O(h^2).$$

Con estas suposiciones el esquema tiene el segundo orden de exactitud.

No obstante, el mismo orden de exactitud tiene lugar también para las exigencias más débiles respecto a la suavidad:

$$k(x), \quad q(x), \quad f(x) \in C^{(2)}, \quad u \in C^{(3)}. \tag{7}$$

LEMA. *Si se cumplen las condiciones* (7), *queda lícita la fórmula*

$$\frac{(ku')_{i+1/2} - (ku')_{i-1/2}}{h} = (ku')'_i + O(h^2), \tag{8}$$

donde $u = u(x)$ *es la solución de la ecuación* (6).

DEMOSTRACIÓN. Hagamos uso de la fórmula de Taylor:

$$v_{i\pm 1/2} = v_i \pm \frac{1}{h} hv'_i + \frac{h^2}{8} v''_i + \frac{h^3}{48} v'''_i (x_i \pm \theta h),$$

$$0 \leqslant \theta \leqslant 1, \quad \frac{1}{h}(v_{i+1/2} - v_{i-1/2}) = v'_i + O(h_2).$$

Sustituyendo aquí $v = ku'$ y teniendo en cuenta que $(ku')'' = (qu - f)'$, $(ku')''' = (qu - f)''$, obtenemos la fórmula (8).

En virtud del lema, el error de aproximación ψ puede ser representado en la forma

$$\psi_i = \eta_{x,i} + \psi^*_i, \quad \eta_i = (au_{\bar{x}})_i - (ku')_{i-1/2}, \quad \psi^*_i = O(h^2)$$

bajo las condiciones (7).

Ahora, teniendo presente que

$$a_i = k_{i-1/2} + O(h^2) \quad \text{para} \quad k(x) \in C^{(2)},$$

$$u_{\bar{x},i} = \frac{u_i - u_{i-1}}{h} = (u')_{i-1/2} + O(h^2) \quad \text{para} \quad u \in C^{(3)},$$

obtenemos $\eta_i = O(h^2)$. Efectivamente, $u_i = u_{i-1/2} + {}^1/_2 h u'_{i-1/2} + {}^1/_8 h^2 u''_{i-1/2} + O(h^2)$.

$$u_{i-1} = u_{i-1/2} - \frac{1}{2} h u'_{i-1/2} + \frac{1}{8} h^2 u''_{i-1/2} + O(h^3),$$

$$u_{\bar{x},\, i} = u'_{i-1/2} + O(h^2),$$

$$a_i u_{\bar{x},\, i} = (k_{i-1/2} + O(h^2))\,(u'_{i-1/2} + O(h^2)) = (ku')_{i-1/2} + O(h^2),$$

$$\eta_i = O(h^2).$$

Más abajo se obtendrá la estimación apriorística $\| z \|_C$ directamente en términos de η y ψ^*.

3. Estimaciones apriorísticas. Pasemos a la estimación del error z en términos de ψ. Recordemos, ante todo, la estimación obtenida en el § 5 del cap. I con ayuda del método de factorización:

$$\| z \|_C \leqslant \frac{1}{c_1} \sum_{i=1}^{N-1} h \sum_{k=1}^{i} h |\psi_k|,$$

de donde se infiere

$$\| z \|_C \leqslant \frac{1}{2c_1} \| \psi \|_C.$$

Mostremos que para la solución del problema

$$(az_{\bar{x}})_x - dz = -\mu_x, \quad x \in \omega_h, \quad z(0) = z(1) = 0,$$
$$a \geqslant c_1 > 0, \qquad d \geqslant 0$$

es válida la estimación

$$\| z \|_C \leqslant \frac{2}{c_1} (1,\ |\mu|], \tag{9}$$

donde se designa $(y,\ v] = \sum_{i=1}^{N} y_i v_i h$.

Representemos z en forma de una suma $z = w + v$, donde w y v son las soluciones de los problemas

$$(aw_{\bar{x}})_x = -\mu_x, \quad x \in \omega_h, \quad w(0) = w(1) = 0;$$

$$\Lambda v = (av_{\bar{x}})_x - dv = -dw, \quad x \in \omega_h, \quad v(0) = v(1) = 0$$

La función w se hallará en la forma explícita, para estimar v se hará uso del principio del máximo. De la ecuación

$$(aw_{\bar{x}} + \mu)_x = 0, \quad (aw_{\bar{x}})_{i+1} = \mu_{i+1} = (aw_{\bar{x}})_i + \mu_i$$

se deduce que $aw_{\bar{x}} + \mu = \text{const} = c_0$. Realicemos las transformaciones evidentes:

$$w_i = w_{i-1} + \frac{(c_0 - \mu_i)\,h}{a_i} = c_0 \sum_{k=1}^{i} \frac{h}{a_k} - \sum_{k=1}^{i} \frac{\mu_k}{a_k} h + w_0,$$

$$0 = w_N = c_0 \sum_{k=1}^{N} \frac{h}{a_k} - \sum_{k=1}^{N} \frac{\mu_k}{a_k} h;$$

$$c_0 = \sum_{k=1}^{N} \frac{\mu_k}{a_k} h \Big/ \sum_{k=1}^{N} \frac{h}{a_k}.$$

Al introducir la designación

$$\alpha_i = \sum_{k=1}^{i} \frac{h}{a_k} \Big/ \sum_{k=1}^{N} \frac{h}{a_k}, \quad 0 < \alpha_i \leqslant 1,$$

encontramos

$$w_i = \alpha_i \sum_{k=1}^{N} \frac{h\mu_k}{a_k} - \sum_{k=1}^{i} \frac{h\mu_k}{a_k}.$$

De aquí proviene

$$|w_i| = \left| -(1-\alpha_i) \sum_{k=1}^{i} \frac{h\mu_k}{a_k} + \alpha_i \sum_{k=i+1}^{N} \frac{h\mu_k}{a_k} \right| \leqslant$$

$$\leqslant (1-\alpha_i) \sum_{k=1}^{i} \frac{h\,|\mu_k|}{a_k} + \alpha_i \sum_{k=i+1}^{N} \frac{h\,|\mu_k|}{a_k} \leqslant \sum_{k=1}^{N} \frac{h\,|\mu_k|}{a_k}.$$

Ahora nos queda tomar en consideración que $a_k \geqslant c_1 > 0$ y obtenemos

$$\| w \|_C \leqslant \frac{1}{c_1} \sum_{k=1}^{N} h\,|\mu_k| = \frac{1}{c_1}(1,\ |\mu|). \tag{10}$$

Con el fin de estimar v hagamos uso del teorema 4 del § 5 del cap. I:

$$\| v \|_C \leqslant \| w \|_C. \tag{11}$$

Al reunir las desigualdades (10 y (11), tenemos

$$\| z \|_C = \| w + v \|_C \leqslant 2 \| w \|_C \leqslant \frac{2}{c_1} (1, \ |\mu|),$$

es decir, queda demostrada la estimación (9).

Volvamos ahora al problema (6'), donde $\psi = \eta_x + \psi^*$. Representemos en la forma

$$\psi = \mu_x, \quad \text{donde } \mu_i = \eta_i + \sum_{k=1}^{i-1} h\psi_k^*, \tag{12}$$

y hagamos uso de la estimación (9). Entonces, para la solución del problema (6') obtendremos las siguientes estimaciones apriorísticas:

$$\begin{gathered} \| z \|_C \leqslant \frac{2}{c_1} \left\{ (1, \ |\eta|] + \sum_{k=1}^{N} h \left| \sum_{k=1}^{i-1} h\psi_k^* \right| \right\}, \\ \| z \|_C \leqslant \frac{2}{c_1} \{ (1, \ |\eta|] + (1, \ |\psi^*|] \}. \end{gathered} \tag{13}$$

Queda probar que tiene lugar la fórmula (12). En efecto, al designar $\rho_i = \sum_{k=1}^{i-1} h\psi_k^*$, vemos que $\rho_{i+1} - \rho_i = h\psi_i^*$, es decir, $\psi_i^* = \rho_{x,\,i}$ y $\psi = \eta_x + \rho_x = \mu_x$, donde $\mu_i = \eta_i + \rho_i$.

4. Convergencia y exactitud del esquema de diferencias. Pasemos a estimar la exactitud de un esquema de diferencias. Suponiendo que

$$k(x), \quad q(x), \quad f(x) \in C^{(2)}, \quad u(x) \in C^{(3)},$$

obtenemos $\eta(x) = O(h^2)$, $\psi^* = O(h^2)$. Ahora resta por utilizar la estimación apriorística (13), la que podría ser sustituida por una estimación más aproximada

$$\| z \|_C \leqslant \frac{2}{c_1} (\| \eta \|_C + \| \psi^* \|_C).$$

De aquí se desprende que el esquema (5) converge uniformemente con el segundo orden, es decir, $\| z \|_C = \| y - u \|_C \leqslant Mh^2$, si se cumplen las condiciones (7).

Resulta más difícil demostrar la convergencia del esquema en la clase de coeficientes discontinuos $k(x)$, $q(x)$, $f(x)$. Para simplificar, analicemos un caso en que $k(x)$ tiene la discontinuidad de primera especie en un punto, mientras que $q(x)$ y $f(x)$ son continuas y pertenecen ambas a la clase $C^{(2)}$.

Denotemos con $Q^{(k)}[a, b]$ un conjunto de funciones continuas a trozos que están definidas en el segmento $[a, b]$ y tienen en $[a, b]$ k derivadas continuas a trozos.

Así pues, sea $k(x) \in Q^{(2)}$, $q(x)$, $f(x) \in C^{(2)}$ y $k(x)$ tiene discontinuidad de primera especie en el punto ξ del segmento $[x_n, x_{n+1}]$, de modo que $\xi = x_n + \theta h$, $0 \leqslant \theta \leqslant 1$. Para $x = \xi$ se cumplen las condiciones de conjugación

$$u_- = u_+, \quad (ku')_- = (ku')_+ = w_0,$$

donde

$$v_+ = v(\xi + 0), \quad v_- = v(\xi - 0).$$

Entonces $\eta_i = O(h^2)$ para $i \neq n+1$, $\psi_i^* = O(h^2)$ para todo $i = 1, 2, \ldots, N-1$, $\eta_{n+1} = a_{n+1}u_{x,n} - (ku')_{n+1/2}$. Sustituyendo aquí

$$u_{n+1} = u(\xi) + (1-\theta)hu'_+ + O(h^2),$$
$$u_n = u(\xi) - \theta h u'_- + O(h^2),$$

$$u_{x,n} = (u_{n+1} - u_n)/h = \theta u'_- + (1-\theta)u'_+ + O(h) =$$
$$= \theta \frac{(ku')_-}{k_-} + (1-\theta)\frac{(ku')_+}{k_+} + O(h) =$$
$$= w_0\left(\frac{\theta}{k_-} + \frac{1-\theta}{k_+}\right) + O(h),$$

$(ku')_{n+1/2} = (ku')_- + O(h) = w_0 + O(h)$ para $\theta > 1/2$,

$(ku')_{n+1/2} = (ku')_+ + O(h) = w_0 + O(h)$ para $\theta < 1/2$,

obtenemos

$$\eta_{n+1} = w_0\left[a_{n+1}\left(\frac{\theta}{k_-} + \frac{1-\theta}{k_+}\right) - 1\right] + O(h),$$

es decir, $\eta_{n+1} = O(1)$ para cualquier esquema y sólo para un esquema con coeficiente

$$\mathring{a}_i = \left[\frac{1}{h}\int_{x_{i-1}}^{x_i} \frac{dx}{k(x)}\right]^{-1}$$

tenemos $\eta_{n+1} = O(h)$. En efecto,

$$\frac{1}{\overset{\circ}{a}_{n+1}} = \frac{1}{h}\int_{x_n}^{\xi} \frac{dx}{k(x)} + \frac{1}{h}\int_{\xi}^{x_{n+1}} \frac{dx}{k(x)} = \frac{\theta}{k_-} + \frac{1-\theta}{k_+} + O(h),$$

es decir, $\overset{\circ}{a}_{n+1}\left(\frac{\theta}{k_-} + \frac{1-\theta}{k_+}\right) = 1 + O(h)$, y, por consiguiente, $\eta_{n+1} = O(h)$. En el segundo miembro de la desigualdad (13) figura la magnitud

$$(1,\ |\eta|] = \sum_{i=1,\ i\neq n+1}^{N} h|\eta_i| + h|\eta_{n+1}|.$$

Con esto queda demostrado el teorema siguiente.

TEOREMA. En *la clase de coeficientes discontinuos* $k(x) \in Q^2$, $q(x)$, $f(x) \in C^{(2)}$ *cualquier esquema de diferencias homogéneo* (5) *de segundo orden de la aproximación tiene el primer orden de exactitud, mientras que el esquema con coeficiente* $a_i = \overset{\circ}{a}_i$ *tiene el segundo orden de exactitud.*

§ 4. Esquemas homogéneos sobre las redes no uniformes

1. Esquema conservativo en una red no uniforme. Elijamos en el segmento $0 \leqslant x \leqslant 1$ una red arbitraria no uniforme

$$\hat{\omega}_h = \{x_i,\ i = 0,\ 1,\ \ldots,\ N,\ x_0 = 0,\ x_N = 1\}.$$

Para obtener un esquema conservativo tripuntual en la red no uniforme, escribamos una ecuación de balance en el segmento $[x_{i-1/2},\ x_{i+1/2}]$:

$$w_{i+1/2} - w_{i-1/2} - \int_{x_{i-1/2}}^{x_{i+1/2}} qu\,dx = -\int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,dx,\quad w = ku'.$$

Dicha ecuación se anota igual tanto para una red uniforme como para una no uniforme. Nos queda aproximar las inte-

grales y derivadas que intervienen en la ecuación de balance:

$$w_{i-1/2} = (ku')_{i-1/2} \sim a_i \frac{u_i - u_{i-1}}{h_i}, \quad h_i = x_i - x_{i-1},$$

$$\int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,dx \sim \varphi_i \hbar_i, \quad \int_{x_{i-1/2}}^{x_{i+1/2}} gu\,dx \sim d_i u_i \hbar_i,$$

$$\hbar_i = \frac{1}{2}(h_i + h_{i+1}),$$

donde d_i y φ_i son unas funciones reticulares. Como resultado, se obtiene un esquema de diferencias

$$\frac{1}{\hbar_i}\left[a_{i+1}\frac{y_{i+1}-y_i}{h_{i+1}} - a_i \frac{y_i - y_{i-1}}{h_i}\right] - d_i y_i = -\varphi_i,$$
$$i = 1, 2, \ldots, N-1, \quad y_0 = \mu_1, \quad y_N = \mu_2. \quad (1)$$

Para determinar d_i y φ_i usaremos las fórmulas más sencillas $\varphi_i = f_i$, $d_i = q_i$, $i = 1, 2, \ldots, N-1$. El coeficiente a_i, se determina por los valores $k(x)$ en el intervalo (x_{i-1}, x_i), a consecuencia de lo cual puede tomarse igual al que figura sobre la red uniforme, de modo que $a_i = k_{i-1/2} + O(h_i^2)$ para $k(x) \in C^{(2)}$.

2. Error de aproximación. Introduzcamos las designaciones

$$y_{\overline{x},\,i} = \frac{y_i - y_{i-1}}{h_i}, \quad y_{x,\,i} = \frac{y_{i+1} - y_i}{h_{i+1}}, \quad y_{\hat{x},\,i} = \frac{y_{i+1} - y_i}{h_i}$$

y escribamos el esquema de diferencias en la forma

$$(ay_{\overline{x}})_{\hat{x}} - dy = -\varphi, \quad x = x_i \in \hat{\omega}_h, \quad y_0 = \mu_1, \quad y_N = \mu_2. \quad (1)$$

Al suponer $z = y - u$, obtendremos para z la ecuación

$$(az_{\overline{x}})_{\hat{x}} - dz = -\psi, \quad x \in \hat{\omega}_h, \quad z_0 = z_N = 0, \quad (2)$$

donde

$$\psi = \Lambda u + \varphi = (au_{\overline{x}})_{\hat{x}} - du + \varphi \quad (3)$$

es el residuo para el esquema (1) en la solución $u = u(x)$.

LEMA 1. *Si* $qu \in C^{(2)}$, $f \in C$ (2), *entonces para el error de aproximación* ψ *es válida la fórmula*

$$\psi = \eta_{\hat{x}} + \psi^*, \tag{4}$$

donde $\eta_i = (au_{\bar{x}})_i - (ku')_{i-1/2} - h_i^2 (qu - f)_i'/8$, $\psi_i^* = O(h^2)$ *para* $\varphi_i = f_i$, $d_i = q_i$.

Hagamos uso de la indentidad del p. 1, escribiéndola en la la forma

$$0 = w_{\hat{x},\, i} - \frac{1}{h_i} \int_{x_{i-1/2}}^{x_{i+1/2}} (qu - f)\, dx, \quad w_i = (ku')_{i-1/2}.$$

Sustrayeimos esta identidad de la igualdad (3):

$$\psi = [(au_{\bar{x}})_i - (ku')_{i-1/2}] - (du)_i + \varphi_i + \frac{1}{h_i} \int_{x_{i-1/2}}^{x_{i+1/2}} (du - f)\, dx. \tag{5}$$

La integral que figura en el segundo miembro se representará en forma de una suma de dos integrales: de $x_{i-1/2}$ a x_i y de x_i a $x_{i+1/2}$; al desarrollar después la función subintegral $\tilde{f} = qu - f$ en el entorno del nodo $x = x_i$, hallaremos

$$\frac{1}{h_i} \int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{f}(x)\, dx = \frac{1}{h_i} \left\{ \int_{x_{i-1/2}}^{x_i} [\tilde{f}_i + (x - x_i)\, \tilde{f}_i']\, dx + O(h_i^3) + \right.$$

$$\left. + \int_{x_i}^{x_{i+1/2}} [\tilde{f}_i + (x - x_i)\, \tilde{f}_i']\, dx + O(h_{i+1}^3) \right\} =$$

$$= \tilde{f}_i + \frac{1}{8h_i} (h_{i+1}^3 - h_i^3)\, \tilde{f}_i' + O(h_i^2),$$

puesto que $h_i^3 + h_{i+1}^3 < (2h_i)^3$. La sustitución $h_{i+1}^3 \tilde{f}_i' = h_{i+1}^3 \tilde{f}_{i+1}' + O(h_{i+1}^3)$ nos da

$$\frac{1}{h_i} \int_{x_{i-1/2}}^{x_{i+1/2}} \tilde{f}(x)\, dx = \tilde{f}_i + (h^2 \tilde{f}')_{\hat{x},\, i} + O(h_i^2).$$

Sustituyendo esta expresión con $\widetilde{f} = qu - f$ en (5), llegamos a la fórmula (4).

Para estimar η_i según el orden veamos la diferencia $(au_{\bar{x}})_i - (ku')_{i-1/2}$ a condición de que $k \in C^{(2)}$, $u \in C^{(3)}$. Empleando la suposición $a_i = k_{i-1/2} + O(h_i^2)$ y las fórmulas $u_i = u_{i-1/2} + h_i u'_{i-1/2}/2 + h_i^2 u''_{i-1/2}/8 + O(h_i^3)$, $u_{i-1} = u_{i-1/2} - h_i u'_{i-1/2}/2 + h_i^2 u''_{i-1/2}/8 + O(h_i^3)$, $u_{\bar{x},i} = (u_i - u_{i-1})/h_i = u_{i-1/2} + O(h_i^2)$, obtenemos

$$(au_{\bar{x}})_i - (ku')_{i-1/2} =$$

$$= (k_{i-1/2} + O(h_i^2))(u'_{i-1/2} + O(h_i^2)) - (ku')_{i-1/2} = O(h_i^2).$$

De este modo, es válida la estimación

$$\eta_i = O(h_i^2) \text{ para } (k(x),\ q(x),\ f(x) \in C^{(2)},\ u(x) \in C^{(3)}.$$

OBSERVACIÓN. Se suponía que d_i y φ_i se determinan según las fórmulas más sencillas: $d_i = q_i$, $\varphi_i = f_i$. Si, en cambio, se emplean las fórmulas más complejas, por ejemplo

$$\varphi_i = \frac{h_i f_{i-1/2} + h_{i+1} f_{i+1/2}}{2h_i}, \quad \varphi_i = \frac{1}{h_i} \int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,dx,$$

entonces la función reticular $\psi_i^* = O(h_i^2) - (d_i - q_i)u_i + (\varphi_i - f_i)$ puede ser representada en la forma $\psi_i^* = \rho_{\hat{x},i} + \psi_i^{**}$, donde $\psi_i^{**} = O(h_i^2)$, $\rho_i = O(h_i^2)$ y η_i, en la fórmula (4) se sustituye por la suma $\eta_i + \rho_i$:

$$\psi = (\rho_i + \eta_i)_{\hat{x}} + \psi^{**}, \qquad (4')$$

$$\rho_i = O(h_i^2) \quad \eta_i = O(h_i^2), \quad \psi_i^{**} = O(h_i^2)$$

para $k,\ q,\ f \in C^{(2)}$, $u \in C^{(3)}$.

3. Estimación de la velocidad de convergencia. Para el problema (2)-(4) es válida la siguiente estimación apriorística

$$\|z\|_C \leqslant \frac{1}{c_1}\{(1,\ |\eta|] + (1,\ |\psi^*|]\}, \qquad (6)$$

donde $(y,\ v)] = \sum_{i=1}^{N} y_i v_i h_i$. Si se cumplen las condiciones (7) del § 3, entonces $\eta_i = O(h_i^2)$, $\psi_i^* = O(h_i^2)$.

Al sustituir η_i y ψ_i^* en (6), nos convencemos de que es verídico el siguiente teorema.

TEOREMA. *En la clase de coeficientes suaves* k, q, $f \in C^{(3)}$ *todo esquema de la forma* (1) *mantiene el segundo orden de exactitud en una sucesión arbitraria de las redes no uniformes.*

Al tomar en consideración la observación del p. 2, podemos representar ψ_i^* en la forma $\psi_i^* = \rho \overset{*}{z}_{x,i} + \psi_i^{**}$, donde $\rho_i = O(h_i^2)$, $\psi_i^{**} = O(h_i^2)$. Entonces, en lugar de (6) queda válida la estimación

$$\|z\|_C \leqslant \frac{2}{c_1} \{(1, |\eta + \rho|] + (1, |\psi^{**}|]\};$$

el teorema sobre el segundo orden de exactitud sobre una red no uniforme queda en vigor.

Si el coeficiente $k(x)$ *tiene discontinuidades de primera* especie en un número finito de puntos, siempre podemos elegir tal red no uniforme $\hat{\omega}_h(k)$ que los puntos de discontinuidad sean los nodos de dicha red. En tal caso cualquier esquema tendrá el segundo orden de exactitud.

Así pues, cualquier esquema homogéneo de segundo orden de aproximación ($\psi = O(h^2)$) sobre una red no uniforme y en la clase de coeficientes suaves tiene el segundo orden de exactitud con la elección especial de las redes no uniformes $\hat{\omega}_h(k)$ en la clase de coeficientes discontinuos.

4. Esquema exacto. Para el problema (1) del § 2 podemos construir un esquema tripuntual exacto cuya solución en los nodos de una red arbitraria coincide con la solución exacta $u = u(x)$ del problema de contorno para una ecuación diferencial. Ilustremos la posibilidad de construir el esquema exacto con un caso particular del problema para $q(x) \equiv 0$:

$$(ku')' = -f(x), \quad 0 < x < 1, \quad u(0) = 0, \quad u(1) = 0. \quad (7)$$

Al haber integrado la ecuación desde x_i hasta x, obtendremos una ecuación

$$(ku') - (ku')_i + \int_{x_i}^{x} f(\xi)\, d\xi = 0.$$

Dividámosla por $k(x)$ y integremos respecto de x desde x_i hasta x_{i+1}:

$$u_{i+1} - u_i - (ku')_i \int_{x_i}^{x_{i+1}} \frac{dx}{k(x)} + \int_{x_i}^{x_{i+1}} \frac{dx'}{k(x')} \int_{x_i}^{x'} f(\xi)\, d\xi = 0, \quad (8)$$

y, luego, de x_{i-1} a x_i:

$$u_i - u_{i-1} - (ku')_i \int_{x_{i-1}}^{x_i} \frac{dx}{k(x)} + \int_{x_{i-1}}^{x_i} \frac{dx'}{k(x')} \int_{x_i}^{x'} f(\xi)\, d\xi = 0. \qquad (9)$$

Introduzcamos una designación

$$a_i^0 = \left[\frac{1}{h_i} \int_{x_{i-1}}^{x_i} \frac{dx}{k(x)}\right]^{-1}.$$

Multipliquemos (8) por a_{i+1}^0/h_{i+1}, (9) por a_i^0/h_i, y sustrayamos del primer resultado el segundo. Obtendremos una ecuación

$$\frac{1}{h_i}\left[a_{i+1}^0 \frac{u_{i+1}-u_i}{h_{i+1}} - a_i^0 \frac{u_i - u_{i-1}}{h_i}\right] + \varphi_i = 0,$$

o bien

$$(a^0 u_{\bar{x}})_{\hat{x},\,i} + \varphi_i = 0, \qquad (10)$$

donde

$$\varphi_i = \frac{a_i^0}{h_i h_i} \int_{x_{i-1}}^{x_i} \frac{dx'}{k(x')} \int_{x'}^{x_i} f(t)\, dt + \frac{a_{i+1}^0}{h_{i+1} h_i} \int_{x_i}^{x_{i+1}} \frac{dx'}{k(x')} \int_{x_i}^{x'} f(t)\, dt.$$

Si ponemos $x' = x_i + sh_i$ para $x_{i-1} \leqslant x' \leqslant x_i$, y $x' = x_i + sh_{i+1}$ para $x_i \leqslant x' \leqslant x_{i+1}$, entonces dicha fórmula puede reescribirse de la manera siguiente:

$$\varphi_i = \frac{h_i a_i^0}{h_i} \int_{-1}^{0} \frac{ds}{k(x_i + sh_i)} \int_{s}^{0} f(x_i + \lambda h_i)\, d\lambda + \\ + \frac{h_{i+1} a_{i+1}^0}{h_i} \int_{0}^{1} \frac{ds}{k(x_i + sh_{i+1})} \int_{0}^{s} f(x_i + \lambda h_{i+1})\, d\lambda.$$

De este modo, el esquema (10) es exacta sobre una red arbitraria no uniforme y para cualesquiera funciones continuas a trozos $k(x)$ y $f(x)$. Por supuesto, el empleo práctico de este esquema está obstaculizado por el hecho de que los coeficientes de dicho esquema se expresan a través de las in-

tegrales de $k(x)$ y $f(x)$, razón por la cual su cálculo requiere la aplicación de las fórmulas de integración numérica.

5. Aumento del orden de exactitud. De lo dicho se hace claro que para aumentar la exactitud de la solución aproximada se debe o bien disminuir el paso de la red h o bien aumentar el orden de exactitud del esquema. No obstante, es conveniente construir esquemas con orden de exactitud aumentado sólo para las ecuaciones de coeficientes constantes, puesto que la anotación de tales esquemas para las ecuaciones de coeficientes variables está relacionada con grandes dificultades técnicas y conduce, a menudo, a los algoritmos engarrosos. Ya hemos aducido un ejemplo del esquema $O(h^4)$ para la ecuación $u'' = -f(x)$.

Examinemos ahora una ecuación

$$u'' - qu = -f(x), \qquad q = \text{const} > 0.$$

Escribamos un esquema de diferencias sobre la red uniforme:

$$\Lambda y = y_{\bar{x}x} - dy = -\varphi(x)$$

y elijamos d y φ de un modo tal que tenga la aproximación $O(h^4)$. El error de la aproximación es

$$\psi = \Lambda u + \varphi = (\Lambda u - u'') - (d-q)u + \varphi - f = \\ = \frac{h^2}{12} u^{IV} - (d-q)u + \varphi - f + O(h^4).$$

Al sustituir aquí $u^{IV} = qu'' - f'' = q(qu - f) - f'' = q^2 u - qf - f''$, obtendremos

$$\varphi = -\left(d - q - \frac{h^2}{12} q^2\right) u + \varphi - \left(f + \frac{h^2}{12} qf + \frac{h^2}{12} f''\right) + O(h^4);$$

por consiguiente, $\psi = O(h^4)$, si se pone $d = q + \frac{h^2}{12} q^2$, $\varphi = f + \frac{h^2}{12}(qf + f'')$. El orden de exactitud queda intacto, si en la fórmula para φ sustituimos la derivada f'' por su aproximación de diferencias $f_{\bar{x}x}$, puesto que $h^2 f'' = h^2 f_{\bar{x}x} + O(h^4)$.

El aumento de exactitud del esquema disminuyendo h viene limitada también por el requisito de la economía del tiempo indispensable para la obtención de la solución con

una exactitud prefijada. Por ello, en la práctica se utiliza con frecuencia el cálculo según un mismo esquema sobre la sucesión de redes, el cual permite elevar la exactitud sin aumentar considerablemente el tiempo de cálculo (método de Runge), bajo el supuesto de que sea la solución lo suficientemente suave.

Supongamos que para resolver un problema de diferencias en cualquier red uniforme es válido un desarrollo asintótico

$$y_i^h = u_i + \alpha(x_i) h^{k_1} + O(h^{k_2}), \quad k_2 > k_1 > 0, \qquad (11)$$

donde $\alpha(x_i)$ no depende de h. Se pide hallar una función reticular $\tilde{y}_i$, para la cual

$$\tilde{y}_i = u_i + O(h^2) \qquad (12)$$

sobre cierto conjunto de nodos $\tilde{\omega}_h$.

Veamos dos redes ω_{h_1} y ω_{h_2} de pasos h_1 y h_2, respectivamente, que tienen nodos comunes; designemos con $\tilde{\omega}_h$ el conjunto de nodos comunes. Sean $y_i^{h_1}$ e $y_i^{h_2}$ las soluciones del problema de diferencias en las redes ω_{h_1} y ω_{h_2}, respectivamente. Formemos su combinación lineal $\tilde{y}_i = \sigma y_i^{h_1} + (1 - \sigma) y_i^{h_2}$ y sustituyamos aquí el desarrollo (11):

$$\tilde{y}_i = u_i + \alpha(x_i)(\sigma h_1^{k_1} + (1 - \sigma) h_2^{k_1} + O(h^{k_2}).$$

Igualando a cero el coeficiente de $\alpha(x_i)$, hallemos

$$\sigma = h_2^{k_1}/(h_2^{k_1} - h_1^{k_1}); \qquad (13)$$

con la particularidad de que en los nodos $x_i \in \tilde{\omega}_h$ se cumple el requisito (12).

De este modo, con el fin de aumentar la exactitud de la solución reticular en cierto conjunto de nodos $\tilde{\omega}_h$, se debe resolver el problema dos veces sobre las redes ω_{h_1} y ω_{h_2}, que se intersecan en dicho conjunto, y formar su combinación lineal con coeficientes σ y $(1 - \sigma)$, donde σ se determina de acuerdo con (13).

En particular, podemos tomar $h_2 = h_1/2$, $h_1 = h$; entonces $\tilde{\omega}_h = \omega_{h_1}$. Para el esquema de segundo orden de exactitud tenemos $k_1 = 2$, $k_2 = 4$, y $\sigma = -1/3$, $1 - \sigma = 4/3$.

La posibilidad de obtener el desarrollo

$$z_i = y_i - u_i = \alpha (x_i) h^2 + O(h^4)$$

proviene del desarrollo del residuo $\psi_i = \beta (x_i) h^2 + O(h^4)$, el cual constituye el segundo miembro del problema

$$\Lambda z = -\psi, \qquad z_0 = z_N = 0.$$

El empleo de las redes no uniformes concede grandes posibilidades de aumento empírico de la exactitud sin aumentar el número de nodos, siempre que se tiene una información preliminar sobre el comportamiento de la solución del problema de partida. Así, en la región de variación fuerte de los coeficientes y del segundo miembro de la ecuación resulta natural espesar la red. Cerca de la frontera de una discontinuidad de los coeficientes, la red se espesa corrientemente según la ley de una progresión geométrica. Para obtener la información preliminar, se pueden realizar los primeros cálculos en una red aproximada y a continuación, los cálculos definitivos en una red especial.

§ 5. Métodos de construcción de los esquemas de diferencias

De lo expuesto más arriba está claro que los esquemas de diferencias para una ecuación diferencial concreta han de reflejar correctamente, en el espacio de funciones reticulares, las propiedades principales del problema de partida (autoconjugación, definición de signo y otras). Para el problema de contorno analizado por nosotros anteriormente, el requisito principal resultó ser una propiedad de conservación que es equivalente a la propiedad de autoconjugación del operador de diferencias. El problema de importancia consiste en obtener los esquemas de diferencias con una calidad prefijada. Para construir tales esquemas se emplean actualmente toda una serie de métodos, de los cuales se trata en este párrafo.

1. Método integral de interpolación. Una ecuación diferencial expresa habitualmente cierta ley física de conservación. Dicha ley puede ser escrita en la forma integral para un intervalo (célula) de una red (ecuación de balance). La ecuación diferencial se obtiene de la ecuación de balance,

cuando el paso de la red tiende a cero bajo el supuesto de que existen derivadas continuas que figuran en la ecuación. Las derivadas e integrales que intervienen en la ecuación de balance sobre la red se deben sustituir por las expresiones aproximadas en la red. De resultas se obtendrá un esquema homogéneo. Este método se denomina *integral de interpolación* o bien *método de balance*. Ilustrémoslo con un ejemplo de un problema

$$(ku')' - qu = -f(x), \qquad 0 < x < 1, \qquad (ku') - \sigma_1 u = \\ = -\mu_1 \text{ para } x = 0, \quad u(1) = \mu_2. \qquad (1)$$

Escribamos la ecuación de balance del calor en el segmento $0 \leqslant x \leqslant 1$.:

$$w_{i+1/2} - w_{i-1/2} + \int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\,dx = \\ = \int_{x_{i-1/2}}^{x_{i+1/2}} q(x)\,u(x)\,dx, \quad w = ku', \qquad (2)$$

donde $(-w(x))$ es el flujo térmico, $q(x)\,u(x)$ es la potencia de las corrientes (de las fuentes, cuando $q < 0$) de calor, la cual es proporcional a la temperatura, y $f(x)$, la densidad de distribución de las fuentes exteriores (de las corrientes) de calor. En el primer miembro de esta ecuación figura la cantidad de calor que queda a cuenta de los flujos térmicos en el segmento $[x_{i-1/2}, x_{i+1/2}]$ y a cuenta de las fuentes exteriores; en el segundo miembro se indica la cantidad de calor que se disipa al ambiente exterior a cuenta del intercambio térmico en la superficie lateral.

Con el objeto de obtener de (2) una ecuación en diferencias tripuntual, sustituyamos $w_{i-1\,2}$, $w_{i+1\,2}$ y las integrales en la ecuación (2), por la combinación lineal de valores de las funciones subintegrales en los nodos de la red (x_{i-1}, x_i, x_{i+1}), por ejemplo,

$$\frac{1}{h}\int_{x_{i-1/2}}^{x_{i+1/2}} q(x)\,u(x)\,dx \approx d_i u_i, \quad d_i = \frac{1}{h}\int_{x_{i-1/2}}^{x_{i+1/2}} q(x)\,dx.$$

Integremos la igualdad $u' = w/k$ respecto de x entre x_{i-1} y x_i:

$$u_i - u_{i-1} = \int_{x_{i-1}}^{x_i} \frac{w}{k(x)}\, dx \approx h w_{i-1/2} \frac{1}{a_i},$$

$$a_i = \left[\frac{1}{h}\int_{x_{i-1}}^{x_i} \frac{dx}{k(x)}\right]^{-1}.$$

Como resultado, obtenemos de (2) un esquema

$$\frac{1}{h}\left[a_{i+1}\frac{y_{i+1}-y_i}{h} - a_i\frac{y_i - y_{i-1}}{h}\right] - d_i y_i = -\varphi_i,$$

$$\varphi_i = \frac{1}{h}\int_{x_{i-1/2}}^{x_{i+1/2}} f(x)\, dx.$$

Deduciendo esta expresión se suponía, de hecho, que $u = \text{const}$ para $x_{i-1/2} \leqslant x \leqslant x_{i+1/2}$, $w = \text{const}$ para $x_{i-1} \leqslant x \leqslant x_i$.

En lugar de las expresiones para a_i, d_i, φ_i conviene tomar las fórmulas más sencillas, como lo hicimos en los párrafos anteriores. Escribamos una aproximación de diferencias para la condición de contorno de tercera especie cuando $x = 0$. Con este fin hagamos uso de la ecuación de balance para $0 \leqslant x \leqslant x_{1/2} = h/2$

$$w_{1/2} - w_0 - \int_0^{x_{1/2}} qu\, dx = -\int_0^{x_{1/2}} f(x)\, dx.$$

Sustituyendo aquí

$$w_{1/2} = a_1 u_{\overline{x},\,1}, \qquad w_0 = (ku')_0 = \sigma_1 u_0 - \mu_1,$$

$$\int_0^{x_{1/2}} qu\, dx \sim q_0 u_0 \frac{1}{2} h, \qquad \int_0^{x_{1/2}} f(x)\, dx \sim f_0 \frac{1}{2} h$$

y cambiando en todos los casos u por y, obtendremos la condición de contorno en diferencias

$$a_1 y_{\overline{x},\,1} - \sigma_1 y_0 + \mu_1 - h q_0 y_0/2 = -h f_0/2.$$

la cual puede escribirse en la forma

$$a_1 y_{\bar{x},1} = \bar{\sigma}_1 y_0 - \bar{\mu}_1, \text{ donde } \bar{\sigma}_1 = \sigma_1 + hq_0/2, \ \mu_1 = \mu_1 + hf_0/2. \quad (3)$$

Estimemos en la solución $u = u(x)$ de la ecuación (1) el valor del residuo

$$\nu = a_2 u_{\bar{x},1} - \bar{\sigma} u_0 + \bar{\mu}_1.$$

Al sustituir aquí $a_1 = k_{1/2} + O(h^2) = k_0 + {}^1/_2 hk_0' + O(h^2)$, $u_1 = u_0 + hu_0' + h^2 u_0''/2 + O(h^3)$, $u_{\bar{x},1} = (u_1 - u_0)/h = u_0' + hu_0''/2 + O(h^2)$, obtenemos

$$\nu = (ku')_0 + {}^1/_2 h (ku')_0' - \bar{\sigma}_1 u_0 + \mu_1 + O(h^2) =$$
$$= [(ku')_0 - \sigma_1 u_0 + \mu_1) + {}^1/_2 h [(ku')' - qu + f]_0 +$$
$$+ O(h^2) = O(h^2),$$

es decir, la condición de contorno en diferencias de tercera especie (3) aproxima la condición $ku' = \sigma_1 u - \mu_1$ para $x = 0$ con un error de segundo orden $\nu = O(h^2)$.

Para el empleo práctico la condición de contorno (3) ha de escribirse en la forma

$$y_0 = \varkappa_1 y_1 - \tilde{\mu}_1, \quad \varkappa_1 = \frac{a_1}{a_1 + h\bar{\sigma}_1}, \quad \tilde{\mu}_1 = \frac{h\bar{\mu}_1}{a_1 + h\bar{\sigma}_1}.$$

Con el fin de aumentar la exactitud del esquema al calcular las integrales se debe utilizar la interpolación de orden más elevado.

2. Método de aproximación de una funcional cuadrática. Un problema de contorno

$$Lu = (ku')' - qu = -f(x), \quad 0 < x < 1, \quad u(0) = 0, \quad u(1) = 0$$

es equivalente al problema de buscar el elemento minimizador de la funcional cuadrática

$$J[u] = \int_0^1 (k(u')^2 + qu^2\, dx - 2\int_0^1 fu\, dx.$$

Introduzcamos en el segmento $0 \leqslant x \leqslant 1$ una red $\bar{\omega}_h = \{x_i = ih, i = 0, 1, \ldots, N\}$ y aproximemos la funcional.

Con este objeto representémosla, al principio, como una suma de integrales en los intervalos de la red:

$$J[u]=\sum_{i=1}^{N} J_i[u], \quad J_i[u]=\int_{x_{i-1}}^{x_i}(k(u')^2+qu^2-2fu)\,dx,$$

después de lo cual aproximemos J_i, por ejemplo, así:

$$\int_{x_{i-1}}^{x_i} k(u')^2\,dx \approx a_i (u_{\overline{x},i})^2 h,$$

$$\int_{x_{i-1}}^{x_i}(qu^2-2fu)\,dx \approx \frac{h}{2}|(qu^2-2fu)_i+(qu^2-2fu)_{i-1}|,$$

donde a_i es un coeficiente, por ejemplo,

$$a_i=\frac{1}{h}\int_{x_{i-1}}^{x_i} k(x)\,dx.$$

Obtenemos, como resultado, una funcional

$$J_h[y]=\sum_{k=1}^{N} ha_k (y_{\overline{x},k})^2+\sum_{k=1}^{N-1}(q_k y_k^2-2f_k y_k)\,h,$$

donde $y_i = y(i)$ es una función reticular arbitraria que se reduce a cero cuando $i = 0, N$.

La ecuación

$$Ay=\varphi \text{ ó } \sum_{j=1}^{N} a_{ij}y_j=\varphi_i, \qquad A=A^*>0,$$

tiene una solución que minimiza la funcional

$$I_A[y]=(Ay,\ y)-2(\varphi,\ y)=\sum_{i,\ j=1}^{N} a_{ij}y_j y_i-2\sum_{i=1}^{N}\varphi_i y_i.$$

De esto podemos convencernos al igualar a cero la derivada

$$\frac{\partial I_A[y]}{\partial y_{i_0}}=2\sum_{j=1}^{N} a_{i_0 j}y_j-2\varphi_{i_0}=0, \qquad \frac{\partial^2 I_A}{\partial y_{i_0}^2}>0,$$

puesto que $a_{ii} > 0$ para cualesquiera $i = 1, 2, \ldots, N$, en virtud de que A es positivo $(A > 0)$.

Al calcular las derivadas

$$\frac{\partial J_h}{\partial y_i} = 2a_i y_{\bar{x},\, i} - 2a_{i+1} y_{\bar{x},\, i+1} + (2q_i y_i - 2f_i)\, h,$$

$$\frac{\partial J_h}{\partial y_i^2} = \frac{2a_i}{h} + \frac{2a_{i+1}}{h} + 2q_i > 0,$$

nos cercioramos de que el elemento $y = y(x) \in H_h$, que minimiza la funcional cuadrática, es la solución del problema

$$(a y_{\bar{x}})_{x,\, i} - q_i y_i = -f_i, \quad i = 1, 2, \ldots, N-1, \quad y_0 = 0 = y_N = 0$$

3. Método de aproximación de una identidad integral (método de identidades sumadoras). Sea

$$(ku')' - qu + f(x) = 0, \quad 0 < x < 1, \quad u(0) = u(1) = 0. \quad (4)$$

Multiplicando la ecuación (4) por una función diferenciable arbitraria $v(x)$, que se anula para $x = 0$, $x = 1$, e integrando respecto de k entre 0 y 1, obtenemos una identidad

$$I(u, v) = \int_0^1 (ku'v' + quv - fv)\, dx = 0.$$

Cambiando, por analogía con el p. 2, la integral y las derivadas u', v', escribamos una identidad sumadora

$$I_h[y, v] = \sum_{i=1}^{N} a_i y_{\bar{x},\, i} v_{\bar{x},\, i} h + \sum_{i=1}^{N-1} (q_i y_i - f_i)\, v_i h = 0.$$

Luego, suponiendo, por ejemplo, que $v_i = \delta_{i,\, i_0}$, $0 < i_0 < N$, y teniendo presente que $v_{\bar{x},\, i} = 0$ para $i < i_0$, y $i > i_0 + 1$, $v_{\bar{x},\, i_0+1} = -1/h$, $v_{\bar{x},\, i_0} = 1/h$, obtendremos

$$h\left(\frac{1}{h} a_{i+1} y_{x,\, i} - \frac{1}{h} a_i y_{\bar{x},\, i}\right) + (q_i - f_i)\, h = 0 \quad \text{cuando} \quad i = i_0,$$

es decir, $(a_{\bar{x}} y)_x - qy = -f$.

4. Métodos de Ritz y de Bubnov—Galerkin (métodos variacionales de diferencias). El problema sobre el mínimo

de una funcional

$$I[u] = (Au, u) - 2(u, f),$$

donde A es un operador lineal autoconjugado y definido positivo en el espacio de Hilbert H con el producto escalar (x, y), es equivalente al problema sobre la resolución de una ecuación

$$Au = f.$$

Se introduce una sucesión de espacios de dimensión finita V_n con base $\{\varphi_i^{(n)}\}$, $i = 1, 2, \ldots, n$.

El *método de Ritz* consiste en que se busca un elemento $u_n \in V_n$ que minimiza la funcional $I(u)$ en V_n. La solución aproximada u_n se busca en forma de la suma

$$u_n = \sum_{j=1}^{n} y_j \varphi_j, \tag{5}$$

donde $y_1, \ldots, y_n$ son unos coeficientes desconocidos. Los cálculos nos dan

$$I[u_n] = \sum_{i,\, j=1}^{n} \alpha_{ij} y_i y_j - 2 \sum_{i=1}^{n} \beta_i y_i,$$

$$\alpha_{ij} = \alpha_{ji} = (A\varphi_i, \varphi_j), \quad \beta_i = (f, \varphi_i);$$

$I[u_n] = \Phi(y_1, y_2, \ldots, y_n)$ es una función de n coeficientes y_i. Igualando a cero las derivadas $\partial I[u_n]/\partial y_i$, obtendremos un sistema de n ecuaciones

$$\sum_{j=1}^{n} \alpha_{ij} y_j - \beta_i = 0, \quad i = 1, 2, \ldots, n,$$

para determinar $y_1, y_2, \ldots, y_n$.

Ilustremos el método de Ritz con un ejemplo del problema (4). A título de la función $\varphi_i(x)$ tomemos

$$\varphi_i(x) = \eta\left(\frac{x - x^i}{h}\right) = \eta_i(x), \qquad \eta(s) \begin{cases} 0, & s < -1, \ s > 1, \\ 1 + s, & -1 < s < 0, \\ 1 - s, & 0 < s < 1. \end{cases}$$

Al sustituir en la fórmula para $\alpha_{ij} A\varphi_i = -(k\varphi_i')' + q\varphi_i$, tenemos

$$\alpha_{ij} = (A\varphi_i,\ \varphi_j) = \int_0^1 \left(k \frac{d\eta_i}{dx} \frac{d\eta_j}{dx} + q\eta_i\eta_j \right) dx,$$

$$\beta_i = \int_0^1 f(x)\,\eta_i(x)\,dx. \tag{6}$$

Los cálculos nos dan

$$\frac{d\eta_i}{dx} = 0 \text{ para } x < x_{i-1},\ x > x_{i+1}$$

$$\frac{d\eta_i}{dx} \begin{cases} 1/h \text{ para } x_{i-1} < x < x_i \\ -1/h \text{ para } x_i < x < x_{i+1} \end{cases}$$

De aquí y de (6) se ve que la matriz $[a_{ij}]$ es tridiagonal, puesto que son diferentes de cero sólo aquellos α_{ij}, para los cuales $j = i - 1,\ i,\ i + 1$. Por esto, para y_i se obtiene un sistema

$$\alpha_{i,\,i-1} y_{i-1} + \alpha_{i,i} y_i + \alpha_{i,i+1} y_{i+1} - \beta_i = 0.$$

Introduciendo las designaciones

$$a_i = -h\alpha_{i,i-1} + h^2\alpha_i = h\alpha_{i,i} + h\,(\alpha_{i,\,i-1} + \alpha_{i,i+1}),\quad \beta_i = -h^2\varphi_i$$

y observando que $\alpha_{i+1,\,i} = \alpha_{i,\,i+1}$, obtenemos un esquema

$$a_i y_{i-1} - (a_i + a_{i+1} + h^2 d_i)\, y_i + a_{i+1} y_{i+1} + h^2\varphi_i = 0,$$

o bien

$$(a y_{\bar{x}})_x - dy + \varphi = 0, \tag{7}$$

donde

$$a_i = \int_{-1}^{0} k(x_i + sh)\,ds + h^2 \int_{-1}^{0} q(x_i + sh)\,s(1+s)\,ds,$$

$$d_i = \int_{-1}^{0} q(x_i + sh)(1+s)\,ds + \int_{0}^{1} q(x_i + \mathrm{sh})(1-s)\,ds,$$

$$\varphi_i = \int_{-1}^{0} f(x_i + sh)(1+s)\,ds + \int_{0}^{1} f(x_i + sh)(1-s)\,ds.$$

Este es el esquema de segundo orden de aproximación.

En el método de Bubnov — Galerkin la solución u_n se busca también en la forma (6), más los coeficientes y_i se hallan de la condición de ortogonalidad del residuo $Au_n - f$ respecto de las funciones básicas $\varphi_i(x)$:

$$(Au_n - f, \varphi_i) = 0, \qquad i = 1, 2, \ldots, n, \tag{8}$$

con la particularidad de que no se requiere que el operador A sea autoconjugado. Para el problema (4) eligimos de nuevo las mismas funciones básicas. Al sustituir (6) en (8), obtendremos un sistema de ecuaciones para y_i. Calculando α_{ij} y β_i, llegamos al mismo esquema (7) que se ha obtenido por el método de Ritz.

Con la elección indicada de las funciones coordenadas $\varphi_i(x) = \eta\left(\frac{x - x_i}{h}\right)$ los métodos de Ritz y de Bubnov — — Galerkin coinciden con el método de elementos finitos.

Capítulo V

Problema de Cauchy para las ecuaciones diferenciales ordinarias

En este capítulo examinaremos los esquemas de diferencias destinados para resolver ecuaciones diferenciales ordinarias (no lineales, en el caso general) de primer orden con datos iniciales (problema de Cauchy). La resolución de las ecuaciones mencionadas representa un dominio clásico de aplicación de los métodos numéricos. Existen varios métodos de diferencias, una parte de los cuales se ha elaborado en la época precedente a la invención de los ordenadores y, no obstante, resultó ser aplicable también para las máquinas electrónicas modernas. Nos limitaremos a una exposición breve de los esquemas de diferencias principales que son de amplio uso en la práctica y para los cuales se tienen los programas estándar correspondientes.

§ 1. Métodos de Runge—Kutta

1. Problema de Cauchy para una ecuación de primer orden. Supongamos que se pide hallar una función $u = u(t)$, continua para $0 \leqslant t \leqslant T$, que satisfaga la ecuación diferencial para $t > 0$ y la condición inicial para $t = 0$:

$$\frac{du}{dt} = f(t, u(t)), \quad 0 < t \leqslant T, \quad u(0) = u_0, \tag{1}$$

donde $f(t, u)$ es la función continua prefijada de dos argumentos.

Si la función $f(t, u)$ está definida en un rectángulo $D = \{0 \leqslant t \leqslant T, \ |u - u_0| \leqslant U\}$ y satisface en el dominio D según la variable u la condición de Lipschitz:

$$|f(t, u_1) - f(t, u_2)| \leqslant K |u_1 - u_2|$$

para cualesquiera $(t, u_1), (t, u_2) \in D$, (2)

donde $K = \text{const} > 0$, entonces el problema (1) tiene una solución única.

Para demostrar esta afirmación la ecuación (1) se integra de 0 a t:

$$u(t) = u_0 + \int_0^t f(s,\ u(s))\, ds, \tag{3}$$

y la ecuación integral obtenida se resuelve por el método de aproximaciones sucesivas (método de Picard):

$$u_{n+1}(t) = u_0 + \int_0^t f(s,\ u_n(s))\, ds, \tag{4}$$

donde n es el número de la aproximación (iteración). El método de Picard converge y determina la única solución de la ecuación (3) o del problema de Cauchy (1).

Este método permite hallar la solución aproximada del problema (1), si en (4) sustituimos la integral por una fórmula de cuadratura cualquiera. Sin embargo, el volumen de los cálculos para el algoritmo obtenido es bastante grande, puesto que para cada iteración (con t fijo) debe calcularse una integral.

Para la resolución aproximada del problema (1) se emplea a veces, un método analítico basado en la idea de desarrollo de la solución del problema de Cauchy (1) en una serie de Taylor. La solución aproximada $u_n(t)$ se busca en la forma

$$u_n(t) = \sum_{k=1}^{n} \frac{t^k}{k!} u^{(k)}(0) + u_0, \quad 0 \leqslant t \leqslant T, \tag{5}$$

donde $u^{(1)}(0) = \dfrac{du}{dt}(0) = f(0,\ u_0)$, y los valores de las derivadas $u^{(k)}(0)$ $(k \geqslant 2)$ se hallan mediante la diferenciación sucesiva de la ecuación (1)

$$u^{(2)}(0) = u''(0) = \frac{d}{dt} f(t,\ u)\Big|_{t=0} = f_t(0,\ u_0) + f(0,\ u_0) f_u(0,\ u_0).$$

$$u^{(3)}(0) = u'''(0) = \frac{d^2}{dt^2} f(t, u)\Big|_{t=0} =$$

$$= f_{tt}(0, u_0) + 2f_{ut}(0, u_0) f(0, u_0) +$$

$$+ f_{uu}(0, u_0) u''(0), \ \ldots,$$

$$f_t = \frac{\partial f}{\partial t}, \quad f_u = \frac{\partial f}{\partial u}, \quad f_{ut} = \frac{\partial^2 f}{\partial u\, dt}, \quad \text{etc.}$$

Para t pequeños el método de series (5) puede asegurar buena aproximación hacia la solución exacta $u(t)$ si n son no muy grandes. Aquí el volumen de los cálculos depende no sólo de la exactitud $\varepsilon > 0$ ($|u(t) - u_n(t)| < \varepsilon$) y de $n = n(\varepsilon)$, sino también del tipo de la función $f(t, u)$, puesto que la determinación de las derivadas $u^{(k)}(t)$ puede resultar muy engorrosa.

En lo sucesivo se supondrá siempre que la función $f(t, u)$ es bastante suave, es decir, tiene tantas derivadas (respecto de t y de u) cuantas sean necesarias en el transcurso de la exposición.

Antes de pasar a la exposición de los esquemas de diferencias para el problema (1), detendrémonos en la cuestión de estabilidad de la solución del problema (1). ¿Cómo variará la solución del problema (1) si cambian las condiciones iniciales? Sea $\tilde{u}(t)$ la solución de la ecuación (1) con las condiciones iniciales $u(0) = \tilde{u}_0$. Para el error $z(t) = \tilde{u}(t) - u(t)$ obtenemos una ecuación

$$\frac{dz}{dt} + \alpha(t) z, \quad 0 < t \leqslant T, \quad z(0) = z_0 = \tilde{u}_0 - u_0, \qquad (6)$$

donde $\alpha(t) = |f(t, \tilde{u}) - f(t, u) - f(t, u)|/z = f_u(t, u + \theta z)$, $0 \leqslant \theta \leqslant 1$.

Como solución de (6) interviene la función

$$z(t) = z(0) \exp\left\{\int_0^t \alpha(s)\, ds\right\}.$$

Si $f_u \leqslant 0$ para cualesquiera t, u, entonces

$|z(t)| \leqslant |z(0)|$, o bien $|\tilde{u}(t) - u(t)| \leqslant |\tilde{u}_0 - u_0|$

para todo $t \in [0, T]$,

es decir, la solución del problema (1) es *estable* respecto de los datos iniciales (el error en los datos iniciales no crece). El problema (1) es estable también respecto del segundo miembro:

$$| \tilde{u}(t) - u(t) | \leqslant | \tilde{u}_0 - u_0 | + \varepsilon T \text{ para } 0 \leqslant t \leqslant T, \quad \text{si } f_u \leqslant 0,$$

donde $\tilde{u}(t)$ es la solución del problema (1) con el segundo miembro

$$\tilde{f} = f(t, \tilde{u}) + \delta f, \quad | \delta f | \leqslant \varepsilon, \quad \varepsilon = \text{const} > 0.$$

La solución del problema (6) para $t \to \infty$ se comporta igual que la solución de una ecuación lineal

$$\frac{dz}{dt} + \lambda z = 0, \quad 0 < t \leqslant T, \quad z(0) = z_0,$$

que puede considerarse en el estudio de la estabilidad como la ecuación modelo.

2. Esquema de diferencias de Euler. Introduzcamos en el segmento de integración $0 \leqslant t \leqslant T$ una red $\omega_\tau = \{t_n = n\tau, \quad n = 0, 1, \ldots\}$. Denotaremos con $y_n = y(t_n)$ una función reticular. El método numérico más simple para resolver la ecuación (1) está representado por *el esquema de diferencias de Euler*:

$$\frac{y_{n+1} - y_n}{\tau} = f(t_n, y_n), \quad n = 0, 1, \ldots, y_0 = u_0. \tag{7}$$

Los valores de $y_n = y(t_n)$ se determinan sucesivamente a partir de $y_0 = u_0$ según una fórmula explícita

$$y_{n+1} = y_n + \tau f(t_n, y_n), \quad n = 0, 1, \ldots, y_0 = u_0.$$

En lugar de $u = u(t)$ encontramos una función reticular $y_n = y(t_n)$ que es la solución aproximada del problema (1).

Una función reticular

$$z_n = y_n - u(t_n)$$

es el *error* del esquema de diferencias. Escribamos la ecuación para z_n. Con este fin sustituyamos $y_n = z_n + u_n$ en

(7) y tomemos en consideración que

$$y_{n+1} - y_n = (z_{n+1} - z_n) + (u_{n+1} - u_n),$$

$$f(t_n, y_n) = f(t_n, u_n) + [f(t_n, u_n + z_n) - f(t_n, u_n)] = \\ = f(t_n, u_n) = \alpha_n z_n,$$

donde

$$\alpha_n = f_u(t_n, u_n + \theta z_n), \quad 0 \leqslant \theta \leqslant 1.$$

Como resultado, obtenemos para z_n un problema

$$\frac{z_{n+1} - z_n}{\tau} = \alpha_n z_n + \psi_n, \quad n = 0, 1, \ldots, z_0 = 0, \tag{8}$$

donde ψ_n es el residuo o error de la aproximación del esquema (7) en la solución $u = u(t)$ del problema (1), que es igual a

$$\psi_n = f(t_n, u_n) - \frac{u_{n+1} - u_n}{\tau}. \tag{9}$$

Estimemos ψ_n para $\tau \to 0$. Para ello sustituyamos

$$u_{n+1} = u_n + \tau \dot{u}_n + \frac{\tau^2}{2} \ddot{u}_n + \ldots \left(\dot{u} = \frac{du}{dt}\right)$$

en (9), y, teniendo en cuenta que, de acuerdo con (1), $\dot{u}_n = f(t_n, u_n)$, obtendremos: $\psi_n = O(\tau)$, o bien $\| \psi \|_C = \max\limits_{0 \leqslant t_n \leqslant T} |\psi_n| = O(\tau)$. Esto es testimonio de que el esquema de Euler tiene *primer orden de aproximación*.

Mostremos que el esquema de Euler converge, es decir, $\| z_n \|_C = \| y_n - u_n \|_C \to 0$ para $\tau \to 0$, y tiene *primer orden de exactitud*, es decir,

$$\| z \|_C = \max_{0 \leqslant t_n \leqslant T} |z_n| = O(\tau).$$

La demostración se aduce bajo el supuesto de que

$$-K \leqslant f_u(t, u) \leqslant 0, \qquad \tau \leqslant 2/K. \tag{10}$$

De (8) determinamos

$$z_{n+1} = (1 + \tau\alpha_n) z_n + \tau\psi_n,$$

$$|z_{n+1}| \leqslant |1 + \alpha\tau_n| |z_n| + \tau |\psi_n| \leqslant |z_n| + \tau |\psi_n|,$$

puesto que $|1 + \tau\alpha_n| \leqslant 1$ conforme a (10). De aquí se deduce que

$$|z_{n+1}| \leqslant |z_0| + \sum_{s=0}^{n} \tau |\psi_s| = \sum_{s=0}^{n} \tau |\psi_s|, \qquad (11)$$

es decir, $\| z \|_C = O(\tau)$.

Si la condición (10) no se cumple, pero $|f_u| \leqslant K$, entonces en lugar de (11) obtenemos $|z_{n+1}| \leqslant Te^{KT} \| \psi \|_C$, y la afirmación $|z|_C = O(\tau)$ queda en vigor.

3. Aumento del orden de exactitud. El método de Euler es muy sencillo, mas no es de elevada exactitud. Se puede aumentar el orden de exactitud de la solución numérica respecto de τ sin complicar el algoritmo. Existe el *método de Runge* para elevar la exactitud cuya idea consiste en lo siguiente. Supongamos que la solución $u = u(t)$ es suficientemente suave y tiene lugar el desarrollo siguiente del error $z_n = y_n - u_n$ en potencias de τ:

$$y_n = u_n + \alpha(t)\tau + \beta(t)\tau^2 + \ldots, \qquad (12)$$

donde $\alpha(t)$ y $\beta(t)$ son unas funciones que no dependen de τ.

Elijamos dos redes de pasos τ_1 y τ_2 que tienen nodos comunes (por ejemplo, $\tau_1 = \tau$, $\tau_2 = \tau/2$), resolvamos en cada red el problema (7) y encontremos $y^{(1)}(t_{n_1})$ e $y^{(2)}(t_{n_2})$, respectivamente. Tomemos un nodo, común para las dos redes, $t_n^* = t_{n1} = t_{n2}$, y escribamos (12) para $n = n^*$:

$$y^{(1)}(t_{n^*}) = u(t_{n^*}) + \alpha(t_{n^*})\tau_1 + O(\tau_1^2),$$
$$y^{(2)}(t_{n^*}) = u(t_{n^*}) + \alpha(t_{n^*})\tau_2 + O(\tau_2^2).$$

Formemos una combinación lineal con el parámetro σ:

$$\tilde{y}(t_{n^*}) = \sigma y^{(1)}(t_{n^*}) + (1 - \sigma) y^{(2)}(t_{n^*}) =$$
$$= u(t_{n^*}) + [\sigma\tau_1 + (1 - \sigma)\tau_2]\alpha(t_{n^*}) + O(\tau_1^2 + \tau_2^2).$$

Eligiendo σ de la condición $\sigma\tau_1 + (1 - \sigma)\tau_2 = 0$, es decir suponiendo $\sigma = \tau_2/(\tau_2 - \tau_1)$, obtenemos

$$\tilde{y}(t_{n^*}) = u(t_{n^*}) + O(\tau^2), \quad \tau = \max(\tau_1, \tau_2).$$

La función reticular y aproxima la solución $u = u(t)$ con el segundo orden de exactitud respecto de τ. De este modo, he-

mos elevado la exactitud del método de Euler realizando dos cálculos en las redes de pasos τ_1 y τ_2. Este procedimiento puede ser continuado teniendo presente (12). Al realizar los cálculos según el esquema (7) en tres redes de pasos τ_1, τ_2, τ_3, hallaremos la solución del problema (1) con el tercer orden de exactitud en los nodos comunes para las tres redes elegidas.

4. Esquemas de Runge-Kutta. El orden de exactitud puede ser aumentado complicando el esquema de diferencias. Son de amplio uso en la práctica los *esquemas de Runge-Kutta* del segundo y cuarto órdenes de exactitud.

El cálculo por el esquema de Runge—Kutta del segundo orden de exactitud se realiza en dos etapas. En la primera etapa se halla el valor intermedio de $\bar{y}_n$ según el esquema de Euler de paso $\alpha\tau$:

$$\bar{y}_n = y_n + \alpha\tau f(t_n, y_n);$$

en la segunda etapa se determina el valor de y_{n+1} por la fórmula

$$y_{n+1} = y_n + \tau(1-\sigma) f(t_n, y_n) + \sigma\tau f(t_n + \alpha\tau, \bar{y}_n),$$

donde $\alpha > 0$, $\sigma > 0$ son los parámetros. Al eliminar $\bar{y}_n$, obtendremos para y_{n+1} un esquema

$$\frac{y_{n+1} - y_n}{\tau} = (1-\sigma) f(t_n, y_n) + \\ + \sigma f(t_n + \alpha\tau, y_n + \alpha\tau f(t_n, y_n)). \quad (13)$$

El orden de exactitud del esquema depende de los parámetros α, τ.

Hallemos una expresión para el residuo o error de aproximación del esquema (13). Con este fin, por analogía con el p. 2, trasladamos $(y_{n+1} - y_n)/\tau$ en el segundo miembro y sustituyamos u_n, u_{n+1} en lugar de y_n, y_{n+1}. De resultas, obtendremos la siguiente expresión para el residuo:

$$\psi_n = (1-\sigma) f(t_n, u_n) + \sigma f(t_n + \alpha\tau, u_n + \\ + \alpha\tau f(t_n, u_n)) - (u_{n+1} - u_n)/\tau. \quad (13')$$

Recurriendo al desarrollo por la fórmula de Taylor, obtenemos

$$\psi_n = \tau(\sigma\alpha - 1/2) u_n'' + O(\tau^2).$$

De aquí se ve que el esquema (13) tiene segundo orden de aproximación $\psi_n = O(\tau^2)$, si se cumple la condición

$$\sigma\alpha = 1/2. \tag{14}$$

De este modo, existe una familia (de un solo parámetro) de esquemas (13), (14) de segundo orden de aproximación.

Veamos los casos particulares:

1) $\sigma = 1$, $\alpha = 1/2$:

$$\frac{\bar{y}_n - y_n}{\tau/2} = f(t_n y_n), \quad \frac{y_{n+1} - y_n}{\tau} = f\left(t_n + \frac{\tau}{2}, \bar{y}_n\right). \tag{15}$$

Este es el conocido esquema *predictor—corrector*, o bien *cálculo—recálculo*. Puede ser escrito de otra forma:

$$\bar{y}_n = y_n + \frac{\tau}{2} f(t_n, y_n), \quad y_{n+1} = y_n + \tau f\left(t_n + \frac{\tau}{2}, \bar{y}_n\right),$$

o, al eliminar $\bar{y}_n$, en la forma

$$(y_{n+1} - y_n)/\tau = f\left[t_n + \frac{\tau}{2}, y_n + \frac{\tau}{2} f(t_n, y_n)\right]. \tag{15'}$$

2) $\sigma = 1/2$, $\alpha = 1$:

$$\frac{y_{n+1} - y_n}{\tau} = \frac{1}{2}\{f(t_n, y_n) + f(t_{n+1}, y_n + \tau f(t_n, y_n))\}. \tag{16}$$

Este esquema también puede considerarse como un esquema predictor—corrector: al principio, el esquema de Euler de paso τ (*predictor*):

$$\bar{y}_n = y_n + \tau f(t_n y_n);$$

después, el esquema con una semisuma (*corrector*):

$$(y_{n+1} - y_n)/\tau - {}^1/_2 [f(t_n, y_n) + f(t_{n+1}, \bar{y}_n)].$$

La idea del método predictor—corrector se usa con frecuencia al escribir esquemas de diferencias para las ecuaciones de la física matemática con derivadas parciales.

He aquí las fórmulas para el esquema de Runge—Kutta del cuarto orden de exactitud:

$$\frac{y_{n+1} - y_n}{\tau} = \frac{1}{6}[k_1(y_n) + 2k_2(y_n) + 2k_3(y_n) + k_4(y_n)],$$

$$n = 0, 1, \ldots, y_0 = u_0, \tag{17}$$

donde k_1, k_2, k_3, k_4 son las correcciones que se calculan según las fórmulas

$$k_1 = f(t_n, y_n),\ k_2 = f(t_n + \tau/2,\ y_n + \tau k_1/2),$$
$$k_3 = f(t_n + \tau/2,\ y_n + \tau k_2/2),\ k_4 = f(t_n + \tau,\ y_n + \tau k_3). \tag{18}$$

Determinando y_{n+1} según el y_n dado se debe cuatro veces calcular el segundo miembro.

Demos a conocer el método de los cálculos según este esquema. Para $n = 0$ sabemos $y_0 = u_0$. Podemos calcular sucesivamente k_1, k_2, k_3, k_4, y hallar

$$y_1 = y_0 + \frac{1}{6}\tau(k_1(y_0) + 2k_2(y_0) + 2k_3(y_0) + k_4(y_0)),$$

después de lo cual los cálculos se realizan para $n = 1, 2, \ldots$ Para el residuo obtenemos una expresión

$$\psi_n = \frac{1}{6}[k_1(u_n) + 2k_2(u_n) + 2k_3(u_n) + k_4(u_n)] - \frac{u_{n+1} - u_n}{\tau}, \tag{19}$$

donde $k_i(u_n)$ $(i = 1, 2, 3, 4)$ se determinan según las fórmulas (18), en las cuales y_n se ha sustituido por u_n.

Al desarrollar u_{n+1}, $k_2(u_n)$, $k_3(u_n)$, $k_4(u_n)$ en el entorno de $t = t_0$, nos convencemos de que $\psi_n = O(\tau^4)$, es decir, el esquema (7), (18) tiene el cuarto orden de aproximación, si $u = u(t)$ tiene cuatro derivadas continuas.

Todos los métodos de Runge—Kutta son *explícitos* (para determinar y_{n+1} se debe realizar los cálculos según las fórmulas explícitas) y de *un paso* (para determinar y_{n+1} se debe hacer un paso en la red desde t_n hasta t_{n+1}).

5. Estabilidad de los esquemas de diferencias. En el p. 1 se ha considerado una propiedad importante de la ecuación diferencial (1), a saber, la estabilidad (respecto de los datos iniciales y del segundo miembro). Para estudiar la estabilidad respecto de los datos iniciales de la ecuación no lineal (1) analizaremos una ecuación modelo

$$\frac{du}{dt} + \lambda u = 0,\quad \lambda = \text{const} > 0,\quad t > 0,\quad u(0) = u_0. \tag{20}$$

Su solución $u(t) = u_0 e^{-\lambda t}$ decrece para $\lambda > 0$, y

$$|u(t)| \leqslant |u_0| \text{ cuando } \lambda \geqslant 0 \text{ para todo } t \geqslant 0, \quad (21)$$

es decir, la ecuación (20) es *estable para* $\lambda \geqslant 0$, lo que corresponde a la condición $f_u \leqslant 0$.

Se introduce una exigencia natural: para los esquemas de diferencias que aproximan las ecuaciones modelo ha de cumplirse un análogo de la desigualdad (21):

$$|y_n| \leqslant |y_0| \text{ para cualesquiera } n = 1, 2, \ldots \quad (22)$$

Veremos más abajo que esto no siempre se cumple.

Veamos una serie de ejemplos.

1) ESQUEMA EXPLÍCITO DE EULER:

$$\frac{y_{n+1} - y_n}{\tau} + \lambda y_n = 0, \qquad y_{n+1} = (1 - \tau\lambda) y_n. \quad (23)$$

De aquí se ve que la condición

$$|y_{n+1}| \leqslant |y_n| \leqslant \ldots \leqslant |y_0| \quad (24)$$

queda cumplida para $|1 - \tau\lambda| \leqslant 1$, o bien $-1 \leqslant 1 - \tau\lambda \leqslant 1$, es decir, para

$$\tau\lambda \leqslant 2. \quad (25)$$

Si, por ejemplo, $\tau\lambda \leqslant 3$, entonces

$$|y_{n+1}| = |\tau\lambda - 1|\,|y_n| \geqslant 2|y_n| \geqslant \ldots \geqslant 2^{n+1}|y_0|,$$
$$|y_n| \geqslant 2^n |y_0| \to \infty \text{ cuando } n \to \infty.$$

El esquema es inestable, la condición (24) no se cumple. De este modo, el esquema de Euler (23) es *convencionalmente estable* para $\tau \leqslant 2/\lambda$, $\lambda > 0$.

2) Esquema implícito de Euler:

$$\frac{y_{n+1} - y_n}{\tau} + \lambda y_{n+1} = 0 \qquad y_{n+1} = \frac{1}{1 + \tau\lambda} y_n. \quad (26)$$

Por cuanto $1/(1 + \tau\lambda) \leqslant 1$ para cualesquiera $\tau\lambda \geqslant 0$, entonces el esquema es *absolutamente estable*:

$$|y_n| \leqslant |y_0| \text{ para cualesquiera } \tau \text{ y } \lambda \geqslant 0, \; n = 0, 1, 2, \ldots \quad (27)$$

3) ESQUEMA CON PESOS

$$\frac{y_{n+1}-y_n}{\tau}+\lambda\left(\sigma y_{n+1}+(1-\sigma)\,y_n\right)=0,\quad y_{n+1}=qy_n. \tag{28}$$

El esquema es estable para

$$|\,q\,|\leqslant 1, \qquad q=\frac{1-(1-\sigma)\,\tau\lambda}{1+\sigma\tau\lambda}.$$

Vemos que $|\,q\,|\leqslant 1$, si $-1-\sigma\tau\lambda\leqslant 1-(1-\sigma)\,\tau\lambda\leqslant 1+\sigma\tau\lambda$, o bien $1+\tau\,(\sigma-1/2)\,\lambda\geqslant 0$, de modo que $1+\sigma\tau\lambda\geqslant\tau\lambda/2>0$. De este modo el *esquema con pesos es absolutamente* (*para todo* τ) *estable* para $\sigma>1/2$, y *condicionalmente estable* para $\sigma<1/2$, siempre que $\tau\leqslant 1/((1/2-\sigma)\,\lambda)$.

4. ESQUEMA DE RUNGE-KUTTA DE SEGUNDO ORDEN. Al sustituir en la fórmula (13) $f=-\lambda y$, obtenemos

$$y_{n+1}=qy_n, \qquad q=1-\tau\lambda+\frac{1}{2}\,\tau^2\lambda^2. \tag{29}$$

El esquema es estable, $|\,y_n\,|\leqslant|\,y_0\,|$, si $|\,q\,|=1-\tau\lambda+\frac{1}{2}\,\tau^2\lambda^2\leqslant 1$, lo que tiene lugar cuando

$$\tau\lambda\leqslant 2. \tag{25}$$

El esquema de Runge—Kutta de segundo orden es estable bajo la misma condición que el esquema explícito de Euler.

5) ESQUEMA DE RUNGE-KUTTA DE CUARTO ORDEN. Sustituyendo $f=-\lambda y$ en (17), (18), obtenemos

$$y_{n+1}=qy_n,$$

$$q=1-\tau\lambda+\frac{1}{2}\,\tau^2\lambda^2-\frac{1}{6}\,\tau^3\lambda^3+\frac{1}{24}\,\tau^4\lambda^4. \tag{30}$$

La desigualdad $|\,q\,|\leqslant 1$ se cumple para $\tau\lambda\leqslant 2{,}78$, es decir, la condición de estabilidad del esquema de cuarto orden es *un poco más débil que la condición* (25) para el esquema de segundo orden.

Estos ejemplos muestran que los esquemas explícitos de *un paso son condicionalmente* estables, y entre los esquemas implícitos se tienen absolutamente estables (por ejemplo (28) cuando $\sigma\geqslant 1/2$). Si $\lambda>0$ es grande, el paso τ, en virtud de (25), debe elegirse para los esquemas explícitos lo suficientemente pequeño.

6. Sobre la convergencia y la exactitud. El esquema de Runge—Kutta para una ecuación no homogénea

$$\frac{du}{dt} + \lambda u = f(t), \quad t > 0, \quad u(0) = u_0 \tag{31}$$

tiene por expresión

$$y_{n+1} = qy_n + \tau\varphi_n, \quad q = q(\tau\lambda), \tag{32}$$

donde las expresiones para q y φ_n dependen del orden del esquema. Así, para el esquema de segundo orden tenemos

$$q = 1 - \tau\lambda + \frac{1}{2}\tau^2\lambda^2,$$

$$\varphi_n = (1-\sigma) f(t_n) + \sigma f(t_n + \alpha\tau), \quad \alpha\sigma = \frac{1}{2}.$$

Para el error $z_n = y_n - u_n$ obtenemos

$$\frac{z_{n+1} - z_n}{\tau} + \left(\lambda - \frac{\lambda^2\tau}{2}\right) z_n = \psi_n$$

o bien

$$z_{n+1} = qz_n + \tau\psi_n, \; n = 0, 1, 2, \ldots, z_0 = 0,$$

donde ψ_n es el residuo igual a

$$\psi_n = \varphi_n - (u_{n+1} - u_n)/\tau = O(\tau^2).$$

En virtud de la condición de estabilidad (25), $|q| \leqslant 1$ y

$$|z_{n+1}| \leqslant |z_n| + \tau |\psi_n| \leqslant \sum_{k=0}^{n} \tau |\psi_k|, \tag{33}$$

de donde precisamente proviene que el esquema (32) converge y tiene el segundo orden de exactitud (converge con la velocidad $O(\tau^2)$, o converge con el segundo orden):

$$\| z \|_C = O(\tau^2).$$

De este modo, si un esquema es estable y aproxima la ecuación (1), es convergente. Esta afirmación demostrada para el problema modelo tiene un carácter general y es verídica para cualquiera de los esquemas de segundo orden.

De modo análogo se demuestra la convergencia con la velocidad $O(\tau^2)$ del esquema de Runge—Kutta (13) a con-

dición de que $f_u \leqslant 0$. En este caso, para $z_n = y_n - u_n$ con $\sigma\alpha = {}^1/_2$ obtenemos un problema

$$\frac{z_{n+1}-z_n}{\tau} = \beta_n \left(1 + \frac{1}{2}\,\tau\gamma_n\right) z_n + \tau\psi_n, \qquad (34)$$

donde $\beta_n = f_u\,(t_n,\ u_n + \theta_1 z_n)$, $\gamma_n = f_u\,(t_n\,(+\,\tau/2,\ u_n + \theta_2 z_n)$ $(0 \leqslant \theta_i \leqslant 1,\ i = 1, 2)$, y ψ se determina por la fórmula (13'). Reescribamos (34) en la forma

$$z_n = q_n z_n + \tau\psi_n, \quad q_n = 1 + \tau\beta_n\,(1 + \tau\gamma_n/2).$$

La condición de estabilidad $|\,q_n\,| \leqslant 1$ ó $-1 \leqslant q_n \leqslant 1$ se cumple, si $2 - \tau\,|\,\beta_n\,| + {}^1/_2\tau^2\,|\,\beta_n\,|\,|\,\gamma_n\,| \geqslant 0$, ${}^1/_2\tau\,|\,\beta_n\,|\,|\,\gamma_n\,| \leqslant |\,\beta_n\,|$, o bien $\tau\,|\,\gamma_n\,| \leqslant 2$. La primera desigualdad queda cumplida también para $\tau\,|\,\beta_n\,| \leqslant 2$, y, por consiguiente, es suficiente que sea

$$\tau K \leqslant 2, \qquad (35)$$

siempre que $f_u \leqslant 0$, $|\,f_u\,| \leqslant K$, $(t,\ u) \in D$. La condición (35) es análoga a (25) y asegura el cumplimiento de la estimación (33), de la cual se deduce precisamente la convergencia del esquema (13) con el segundo orden $\|\,z\,\|_C = O\,(\tau^2)$.

§ 2. Esquemas de varios pasos. Métodos de Adams

1. Esquemas de varios pasos. En el § 1 fueron considerados los métodos de Runge—Kutta para la resolución numérica del problema de Cauchy

$$\frac{du}{dt} = f(t,\ u), \quad 0 < t \leqslant T, \quad u\,(0) = u_0. \qquad (1)$$

Estos son los *métodos de un paso*: al determinar el nuevo valor de y_{n+1} se usa sólo el valor de y_n. Para determinar el valor aproximado de y_n pueden analizarse, en el caso general, los *esquemas de diferencias de m pasos* $(m \geqslant 1)$, es decir, las ecuaciones del tipo

$$\sum_{k=0}^{m} \frac{a_k}{\tau}\, y_{n-k} = \sum_{k=0}^{m} b_k f_{n-k}, \quad n = m,\ m+1,\ \ldots, \qquad (2)$$

donde a_k, b_k son ciertos coeficientes numéricos,

$$f_{n-k} = f(t_{n-k}, y_{n-k}), \quad a_0 \neq 0, \quad b_m \neq 0.$$

En particular, para $m = 1$, $b_0 = 0$, $b_1 = -a_0$, $a_1 = -a_0$, llegamos al esquema de Euler.

El esquema (2) se denomina explícito (de extrapolación), si $b_0 = 0$ y los valores de y_n se determinan en términos de los valores antecedentes de $y_{n-1}, y_{n-2}, \ldots, y_{n-m}$ según una fórmula explícita

$$y_n = \frac{1}{a_0} \sum_{k=1}^{m} (b_k \tau f_{n-k} - a_k y_{n-k}) = \frac{1}{a_0} F(y_{n-1}, y_{n-2}, \ldots, y_{n-m}).$$

Los cálculos empiezan con $n = m$. Para hallar y_m, se deben prefijar m valores iniciales $y_0, y_1, \ldots, y_{m-1}$; éstos pueden determinarse, por ejemplo, mediante el método de Runge—Kutta en el que se utiliza solamente el valor inicial de $y_0 = u_0$.

Si $b_0 \neq 0$, el esquema (2) se denomina implícito (de interpolación): al hallar y_n, se debe resolver, para cada n, una ecuación no lineal

$$a_0 y_n - b_0 f(t_n, y_n) = F(y_{n-1}, y_{n-2}, \ldots, y_{n-m}). \tag{3}$$

Dicha ecuación no lineal puede resolverse, por ejemplo, mediante el método de Newton.

El error de aproximación del esquema (2) en la solución $u = u(t)$ de la ecuación (1) o el residuo se determina por la fórmula

$$\psi_n = \sum_{k=0}^{m} b_k f(t_{n-k}, u_{n-k}) - \frac{1}{\tau} \sum_{k=0}^{m} a_k u_{n-k}. \tag{4}$$

Suele decirse que el esquema (2) tiene el s-ésimo *orden de aproximación* (o simplemente que el esquema (2) tiene el s-ésimo *orden*), si

$$\| \psi \|_C = O(\tau^s), \quad \text{o bien} \quad \| \psi \|_C \leqslant M\tau^s, \quad s > 0, \tag{5}$$

donde $M = \text{const} > 0$ no depende de τ.

Los coeficientes a_k, b_k se eligen a partir de los requisitos de aproximación y estabilidad. Sin perturbar la generalidad

de razonamientos podemos considerar que

$$\sum_{k=0}^{m} b_k = 1, \tag{6}$$

puesto que los coeficientes de la ecuación (2) están determinados con una exactitud de hasta un factor. Desarrollando ψ_n en potencias de τ y exigiendo que el residuo tenga un orden prefijado, obtenemos las condiciones para determinar a_k, b_k. Por cuanto $u = 1$ es la solución de la ecuación $u_t = f(t, u)$ para $f = 0$, de (2) se desprende que

$$\sum_{k=0}^{m} a_k = 0. \tag{7}$$

Con el objeto de construir los esquemas (2) se aplican, corrientemente, otros procedimientos en los cuales se emplean fórmulas de cuadratura y de interpolación. Así por ejemplo, integrando la ecuación diferencial (1) respecto de t dentro de los límites de t_{n-n_0} a t_n, obtenemos

$$u_n - u_{n-n_0} = \int_{t_{n-n_0}}^{t_n} f(t, u(t))\, dt. \tag{8}$$

Para obtener de aquí un esquema de diferencias se puede usar para la integral una fórmula de cuadratura cualquiera.

2. Método de Adams. Toda fórmula de cuadratura engendra el método correspondiente de resolución numérica de la ecuación diferencial ordinaria (1). En una identidad

$$u_n - u_{n-1} = \int_{t_{n-1}}^{t_n} f(t, u(t))\, dt, \tag{9}$$

correspondiente a la identidad (8) cuando $n_0 = 1$, sustituyamos la integral por una fórmula de cuadratura

$$\int_{t_{n-1}}^{t_n} f(t, u(t))\, dt \approx \tau \sum_{k=0}^{m} b_k f(t_{n-k}, u_{n-k}). \tag{10}$$

Teniendo presente (9) y (10), podemos escribir el *esquema de diferencias de Adams*:

$$\frac{y_n - y_{n-1}}{\tau} = \sum_{k=0}^{m} b_k f(t_{n-k}, y_{n-k}). \qquad (11)$$

Dicho esquema puede ser obtenido de (2), si ponemos $a_k = 0$ para $k = 2, 3, \ldots, m$, y $a_0 = 1$, $a_1 = -1$.

La fórmula de cuadratura (10), en cuya base está construido el esquema de Adams, contiene los nodos de las redes que no pertenecen al intervalo de integración $t_{n-1} \leqslant t \leqslant t_n$. Habitualmente se utiliza la exigencia de que la fórmula de cuadratura sea exacta para un polinomio de grado m. Con ello se elige un polinomio de interpolación con los nodos t_n, $t_{n-1}, \ldots, t_{n-m}$.

Con tal construcción del esquema su error de aproximación coincide con el error de la fórmula de cuadratura. Efectivamente, el residuo para el esquema (11) es

$$\psi_n = \sum_{k=0}^{m} b_k f(t_{n-k}, u_{n-k}) - \frac{u_n - u_{n-1}}{\tau}.$$

Sustituyendo aquí de (9) la expresión

$$\frac{u_n - u_{n-1}}{\tau} = \frac{1}{\tau} \int_{t_{n-1}}^{t_n} f(t, u(t))\, dt,$$

obtenemos la fórmula para el residuo:

$$\psi_n = \sum_{k=0}^{m} b_k f(t_{n-k}, u_{n-k}) - \frac{1}{\tau} \int_{t_{n-1}}^{t_n} f(t, u(t))\, dt. \qquad (12)$$

3. **Esquemas explícitos e implícitos.** Si $b_0 = 0$, el esquema (11) será explícito y

$$y_n = y_{n-1} + \tau \sum_{k=1}^{m} b_k f_{n-k}. \qquad (13)$$

De ejemplo más simple del esquema explícito de Adams sirve el de Euler

$$y_n - y_{n-1} = \tau f_{n-1} \quad \text{para } m = 1,\ b_0 = 0,\ b_1 = 1. \qquad (14)$$

Al poner en (11) $m = 1$, $b_0 = 1$, $b_1 = 0$, obtendremos el esquema de Adams implícito

$$\frac{y_n - y_{n-1}}{\tau} = f_n, \quad \text{o bien} \quad y_n - \tau f(t_n, y_n) = y_{n-1}. \tag{15}$$

El esquema implícito *simétrico* de un paso ($m = 1$)

$$\frac{y_n - y_{n-1}}{\tau} = \frac{1}{2}[f(t_n, y_n) + f(t_{n-1}, y_{n-1})] \tag{16}$$

corresponde a los valores $m = 1$, $b_0 = b_1 = 1/2$ y tiene el segundo orden de aproximación: $\psi_n = O(\tau^2)$. Para determinar y_n se debe resolver (con cada n) una ecuación no lineal $y_n = {}^1/_2\tau f(t_n, y_n) = F_{n-1}$, donde $F_{n-1} = y_{n-1} + {}^1/_2\tau f(t_{n-1}, y_{n-1})$.

Veamos ahora los esquemas de Adams de dos pasos que corresponden a $m = 2$. El esquema explícito de dos pasos ($m = 2$) tiene por expresión

$$\frac{y_n - y_{n-1}}{\tau} = \frac{3}{2} f_{n-1} - \frac{1}{2} f_{n-2},$$

$$m = 2, \quad b_0 = 0, \quad b_1 = \frac{3}{2}, \quad b_i = -\frac{1}{2}. \tag{17}$$

El esquema es de segundo orden de aproximación:

$$\psi_n = \frac{3}{2} f(t_{n-1}, u_{n-1}) - \frac{1}{2} f(t_{n-2}, u_{n-2}) - \frac{u_n - u_{n-1}}{\tau} = O(\tau^2).$$

Investiguemos la estabilidad del esquema modelo correspondiente

$$\frac{y_n - y_{n-1}}{\tau} + \lambda\left(\frac{3}{2} y_{n-1} - \frac{1}{2} y_{n-2}\right) = 0. \tag{18}$$

Al sustituir aquí $y_n = q^n$, obtendremos

$$q^2 - \left(1 - \frac{3}{2}\mu\right) q - \frac{1}{2}\mu = 0, \quad \mu = \lambda\tau. \tag{19}$$

Por cuanto $D = 1 - \mu + \frac{9}{4}\mu^2 > 0$ para μ cualquiera, las raíces q_1 y q_2 serán reales y distintas. La estabilidad significa que $|q_1| \leqslant 1$ y $|q_2| \leqslant 1$. Hagamos uso de la siguien-

te propiedad que se comprueba inmediatamente: las raíces de una ecuación cuadrática $q^2 + bq + c = 0$ no superan en módulo la unidad:

$$| q_{1,2} | \leqslant 1, \text{ si } | b | \leqslant 1 + c, \ c \leqslant 1. \tag{20}$$

Para la ecuación (19) tenemos $b = 3\mu/2 - 1$, $c = -\mu/2$, y la condición $| 3\mu/2 - 1 | \leqslant 1 - \mu/2$ queda cumplida para $\mu \leqslant 1$, ó

$$\tau\lambda \leqslant 1,$$

es decir, el esquema (18) es condicionalmente estable (el paso τ debe ser dos veces menor que el paso admisible en el esquema de Euler).

Escribamos un esquema implícito de Adams de dos pasos ($m = 2$). Exigiendo que la fórmula de cuadratura (10) sea exacta para los polinomios de grados 0, 1, 2, es decir, que $F(t) = f(t, u(t)) = \{1, t, t^2\}$, encontramos los coeficientes $b_0 = 5/12$, $b_1 = 8/12$, $b_2 = -1/12$. El esquema es de la forma

$$\frac{y_n - y_{n-1}}{\tau} = \frac{1}{12}(5f_n + 8f_{n-1} - f_{n-2}). \tag{21}$$

Investiguemos la estabilidad del problema modelo

$$\frac{y_n - y_{n-1}}{\tau} + \frac{\lambda}{12}(5y_n + 8y_{n-1} - y_{n-2}) = 0. \tag{22}$$

Suponiendo $y_n = q^n$, obtendremos una ecuación característica

$$aq^2 + bq + c = 0, \quad a = 1 + \frac{5}{12}\tau\lambda, \qquad b = \frac{8}{12}\tau\lambda - 1,$$

$$c = -\frac{1}{12}\tau\lambda.$$

Las condiciones (20), para las cuales $| q_{1,2} | \leqslant 1$, adquieren la forma $| b | \leqslant a + c$, $c \leqslant a$. De aquí se infiere que el esquema (22) es estable cuando $\tau\lambda \leqslant 6$.

4. Problema de Cauchy para una ecuación de segundo orden. Analicemos un problema de Cauchy:

$$\frac{d^2u}{dt^2} = f(t, u(t)), \ t > 0, \ u(0) = u_0,$$
$$\frac{du}{dt}(0) = u_1. \tag{23}$$

Los más usados son los *métodos de Störmer*

$$\frac{y_{n+1}-2y_n+y_{n-1}}{\tau^2}=\sum_{k=-1}^{m} b_k f(t_{n-k},\ y_{n-k}),$$

$$m \geqslant 0,\ n=1,\ 2,\ \ldots, \qquad (24)$$

$$y_0=u_0,\quad y_1=\bar{u}_1 \text{ o bien } \frac{y_1-y_0}{\tau}=\tilde{u}_1.$$

El valor de $\bar{u}_1$ (o de $\tilde{u}_1$) se elige de una manera tal que el error de aproximación $\nu=\frac{1}{\tau}[u(\tau)-u(0)]-\dot{u}(0)-\tilde{u}_1$ tenga cierto orden, por ejemplo, $\nu=O(\tau^p)$, donde p es el orden de aproximación del esquema (24). Por ejemplo, para $p=2$ hallamos

$$u(\tau)=u(0)+\tau\dot{u}(0)+\frac{1}{2}\tau^2\ddot{u}(0)+O(\tau^3),$$

$$\nu=u_1+\frac{\tau}{2}\ddot{u}(0)-\tilde{u}_1+O(\tau^2)=\frac{\tau}{2}f(0,\ u(0))+$$

$$+0(\tau^2)-\tilde{u}_1+u_1=O(\tau^2),$$

si ponemos

$$\tilde{u}_1=u_1+{}^1/_2\tau f(0,\ u_0),\ \bar{u}_1=u_0+\tau\tilde{u}_1.$$

Si $b_{-1}=0$, el esquema (24) será explícito, puesto que en el segundo miembro figuran sólo valores conocidos de y_n, $y_{n-1}, \ldots, y_{n-m}$. Si $b_{-1}=0$, el esquema (24) es implícito y para determinar y_{n+1} se debe resolver la ecuación

$$y_{n+1}-b_{-1}f(t_{n+1},\ y_{n+1})=F(y_n,\ y_{n-1},\ \ldots,\ y_{n-m},\ t_n).$$

Para obtener el esquema de diferencias (24) calculemos una integral

$$\int_{t_{n-1}}^{t_{n+1}} u^n v\,dt=\int_{t_{n-1}}^{t_n} u^n v\,dt+\int_{t_n}^{t_{n+1}} u^n v\,dt=$$

$$=(u'v-uv')\Big|_{t_{n-1}}^{t_n}+(u'v-uv')\Big|_{t_{n-1}}^{t_n}+\int_{t_{n-1}}^{t_n} uv^n\,dt, \qquad (25)$$

donde $v(t)$ es una función continua a trozos

$$v(t)=\begin{cases}(t-t_{n-1})/\tau \text{ para } t_{n-1}\leqslant t\leqslant t_n,\\ (t_{n+1}-t)/\tau \text{ para } t_n\leqslant t\leqslant t_{n+1}.\end{cases} \tag{26}$$

Sustituyamos (26) en (25) teniendo presente que $v''(t)=0$:

$$\int_{t_{n-1}}^{t_{n+1}} u''v\,dt=\frac{1}{\tau}(u_{n-1}-2u_n+u_{n+1}). \tag{27}$$

Luego, multiplicando la ecuación (23) por $v(t)$ y tomando en consideración (27), obtendremos una identidad

$$\frac{u_{n+1}-2u_n+u_{n-1}}{\tau}=\frac{1}{\tau}\int_{t_{n-1}}^{t_{n+1}} f(t,\ u(t))\ v(t)\,dt. \tag{28}$$

El error de aproximación del esquema (24) en la solución $u=u(t)$, o el residuo para el esquema (24) se determina mediante la fórmula

$$\psi_n=\sum_{k=-1}^{m} b_k f(t_{n-k},\ u_{n-k})-\frac{u_{n+1}-2u_n+u_{n-1}}{\tau^2},$$

la cual puede ser escrita, en virtud de la identidad (28), en la forma

$$\psi_n=\sum_{k=-1}^{m} b_k f(t_{n-k},\ u_{n-k})-\frac{1}{\tau}\int_{t_{n-1}}^{t_{n+1}} f(t,\ u(t))\ v(t)\,dt. \tag{29}$$

Al introducir una nueva variable $s=(t-t_n)/\tau$, escribamos la integral en la forma más cómoda:

$$\frac{1}{\tau}\int_{t_{n-1}}^{t_{n+1}} F(t)\,v(t)\,dt=\int_{-1}^{1} F(t_n+s\tau)\,\bar{v}(s)\,ds,$$

$$F=f(t,\ u(t)),\quad \bar{v}(s)=\begin{cases}1+s,\ s<0,\\ 1-s,\ s>0.\end{cases}$$

De (29) se ve que el primer sumando es una fórmula de cuadratura para la integral de la función $F(t)=f(t,\ u(t))$

con el peso $v(t) \geqslant 0$. El error de aproximación del esquema se determina completamente por el de la fórmula de cuadratura. Los métodos construidos en esta base se denominan, además *métodos de Adams—Störmer*.

La fórmula más simple de un rectángulo da un esquema

$$\frac{y_{n+1}-2y_n+y_{n-1}}{\tau^2} = f(t_n, y_n),$$

puesto que $\frac{1}{\tau}\int\limits_{t_{n-1}}^{t_{n+1}} v(t)\,dt = 1$.

Para el problema modelo

$$\frac{d^2u}{dt^2}+\lambda u = 0, \quad t>0, \quad u(0)=0, \quad \frac{du}{dt}(0)=u_1$$

tenemos

$$\frac{y_{n+1}-2y_n+y_{n-1}}{\tau^2}+\lambda y_n = 0.$$

Sustituyendo aquí $y_n = q^n$, encontramos $q^2 - 2(1 - \tau^2\lambda/2)\,q + 1 = 0$; $D < 0$ para $\lambda\tau^2 \leqslant 4$, $\tau \leqslant 2/\sqrt{\lambda}$; además, $|q_1| = |q_2|$ y el esquema es estable a condición de que $\tau \leqslant 2\sqrt{\lambda}$ ó $\tau\sqrt{\lambda} \leqslant 2$.

5. Sistemas de ecuaciones. Muchos métodos se extienden sin cambios algunos al problema de Cauchy para el sistema de ecuaciones

$$\frac{du}{dt} = f(t, u), \quad t>0, \quad u(0)=u_0, \tag{30}$$

donde $u = (u^1(t), u^2(t), \ldots, u^N(t))$ es el vector buscado, y $f = (f^1, f^2, \ldots, f^N)$, el vector prefijado. Escribamos (30) en componentes

$$\frac{du^i}{dt} = f^i(t, u), \quad t>0, \quad u^i(0)=u_0^i, \quad i=1, 2, \ldots, N. \tag{31}$$

Sean u, v dos soluciones del problema (30) con los datos iniciales $u(0) = u_0$, $v(0) = v_0$. Para su diferencia $z =$

$= u - v$ obtendremos un sistema de ecuaciones lineales

$$\frac{dz^i}{dt} = \sum_{j=1}^{n} \alpha_{ij}(t)\, z^j,$$

donde α_{ij} es el valor de la derivada $\partial f^i/\partial u^j$ en cierto punto medio $(t, \bar{u}_j)$, $\bar{u}_j = (v^1, v^2, \ldots, v^{j-1}, u^j + \theta, z^j, u^{j+1}, \ldots, u^N)$ $(0 \leqslant \theta_j \leqslant 1, j = 1, 2, \ldots, N)$. Por eso el modelo lineal del sistema de ecuaciones no lineales (30) será representado por un sistema lineal

$$\frac{du^i}{dt} + \sum_{j=1}^{N} a_{ij} u^j = f^i(t). \tag{32}$$

o, en la forma vectorial,

$$\frac{du}{dt} + Au = f(t), \qquad A = (a_{ij}). \tag{33}$$

Para que dicha ecuación sea estable respecto de los datos iniciales, es suficiente que la matriz A sea no negativa. En el párrafo que sigue se indicarán las condiciones necesarias y suficientes de estabilidad de los esquemas para los sistemas de ecuaciones lineales (33).

En la práctica nos encontramos a menudo con los sistemas de ecuaciones que se llaman rígidos y cuya resolución por medios corrientes representa grandes dificultades. Supongamos que $\{\lambda_k\}$ son los números propios de la matriz A (si A no es simétrica, los números λ_k pueden ser complejos). El sistema de ecuaciones (33) se denominará *rígido*, si $\operatorname{Re} \lambda_k > 0$ $(k = 1, 2, \ldots, N)$ y si la razón $\xi = \max\limits_k \operatorname{Re} \lambda_k / \min\limits_k \operatorname{Re} \lambda_k$ es grande.

Si la matriz A es simétrica, entonces todos los números propios son reales y la rigidez del sistema (33) es testimonio de que la matriz A es positiva y que el sistema (33) está mal condicionado, es decir,

$$\xi = \frac{\max\limits_k \lambda_k}{\min\limits_k \lambda_k} \gg 1.$$

Son rígidas, en particular, las ecuaciones que se obtienen al reducir las ecuaciones con derivadas parciales a los siste-

mas de ecuaciones diferenciales ordinarias por medio de la aproximación de diferencias de un operador que contiene derivadas respecto de las variables espaciales (por ejemplo, el operador de Laplace en el caso de la ecuación de conductibilidad térmica).

Los métodos explícitos resultaron ser inútiles para la resolución numérica de los esquemas rígidos, puesto que conducen a grandes restricciones referentes al paso a causa de las exigencias de estabilidad en perjuicio de las de precisión. Aclaremos esto con un ejemplo del sistema de dos ecuaciones

$$\frac{du_1}{dt}+a_1u_1=0,\quad \frac{du_2}{dt}+a_2u_2=0,\quad A=\begin{pmatrix}a_1 & 0\\ 0 & a_2\end{pmatrix},\quad t>0$$
$$a_2>0,\quad a_1>0,\quad a_2\gg a_1. \tag{34}$$

La solución de este sistema es un vector

$$u(t)=(u_1(t),\ u_2(t)),$$
$$u_1(t)=u_1(0)\,e^{-a_1t},\ u_2(t)=u_2(0)\,e^{-a_2t};$$

los componentes de este vector decrecen cuando crece t, con la particularidad de que $|u_2(t)| \ll |u_1(t)|$ para t suficientemente grande.

Tomemos un esquema explícito

$$\frac{y_1^{n+1}-y_1^n}{\tau}+a_1y_1^n=0,\quad \frac{y_2^{n+1}-y_2^n}{\tau}+a_2y_2^n=0,$$
$$n=0,\ 1,\ \ldots,\ y_i^n=y_i(t_n),\quad i=1,\ 2. \tag{35}$$

El sistema se descompone en dos ecuaciones, cada una de las cuales puede resolverse separadamente, no obstante dichas ecuaciones están ligadas entre sí por la elección del paso común τ. El esquema es estable, si se cumplen simultáneamente dos condiciones $a_1\tau \leqslant 2$ y $a_2\tau \leqslant 2$. Por cuanto $a_2 \gg a_1$, ambas condiciones quedan cumplidas, si $\tau \leqslant 2/a_2$. El paso admisible τ se determina, de hecho, por aquel componente $u_2(t)$ de la solución que decrece con mayor rapidez.

Para la resolución del sistema (34) es aplicable un esquema implícito

$$\frac{y_1^{n+1}-y_1^n}{\tau}+a_1y_1^{n+1}=0,\quad \frac{y_2^{n+1}-y_2^n}{\tau}+a_2y_2^{n+1}=0,$$

que es estable para cualesquiera τ y $a_1 \geqslant 0$, $a_2 \geqslant 0$.

Ultimamente ha aparecido toda una serie de esquemas implícitos, algoritmos para éstos y programas nuevos, útiles para resolver sistemas rígidos de ecuaciones diferenciales lineales y no lineales.

6. **Observaciones generales.** 1. Al escoger tal o cual método numérico se toman en consideración varias circunstancias, a saber, el volumen de los cálculos, el volumen requerido de memoria de acceso rápido del ordenador, el orden de exactitud, la estabilidad respecto de los errores de redondeo y otras. Hemos considerado en todo caso los métodos de paso constante $\tau = t_{n+1} - t_n$. La introducción del paso variable $\tau_{n+1} = t_{n+1} - t_n$ lleva un carácter formal y, cuando se trata de los esquemas de un paso, no conduce a nuevas cuestiones de principio. Para los esquemas de varios pasos ($m \geqslant 2$) las fórmulas se alteran.

En el caso general la solución puede ser una función no monótona fuertemente variable. Es natural de emplear una red no uniforme y disminuir el paso (espesar la red) en el dominio de variación rápida de la función $u(t)$ con el fin de asegurar una aproximación más exacta de $u(t)$ por la solución reticular. Sin embargo, no sabemos de antemano el comportamiento de la solución $u = u(t)$. Por eso en la práctica se procede de la manera siguiente: al principio se realizan los cálculos en la red uniforme; si se pone claro que la solución $u = u(t)$ varía fuertemente en cierto intervalo $t_* < t < t^*$, entonces la red se hace espesar en $[t_*, t^*]$ y el problema se resuelve en la red no uniforme de esta índole. Se recomienda, en general, realizar los cálculos en varias redes que se hacen espesar. Si, al espesar la red, la solución varía poco, la exactitud requerida se considera lograda. Con el fin de elevar el orden de exactitud resulta aplicable el método de Runge en el que se usan cálculos sobre diferentes redes (siempre que la solución $u = u(t)$ posee una suavidad suficiente). En el transcurso de los cálculos puede resultar necesario emplear los esquemas de diferentes órdenes de exactitud en distintos dominios de variación del argumento.

2. Nos encontramos a menudo con la necesidad de resolver las ecuaciones cuyos coeficientes cambian furtemente, por ejemplo

$$\frac{du}{dt} = \alpha(t)\, u, \quad t > 0, \quad u(0) = u_0. \tag{36}$$

Una ecuación de este género se encuentra en la descripción de los problemas de la cinética química. A título de su solución interviene una función

$$u(t) = u_0 \exp\left\{\int_0^t \alpha(s)\, ds\right\}.$$

Si $\alpha(t) \geqslant 0$, puede utilizarse el esquema de Euler para cualquiera:

$$y_{n+1} = y_n + \tau\alpha_n y_n = (1 + \tau\alpha_n)\, y_n. \qquad (37)$$

Si, en cambio, $\alpha(t) < 0$, puede suceder que $1 + \tau\alpha_{n_1} < 0$ para cierto $n = n_1$, e $y_{n_1+1} < 0$, es decir, la solución pierde sentido. En este caso puede emplearse un esquema implícito

$$y_{n+1} = y_n + \tau\alpha_n y_{n+1},$$
$$y_{n+1} = y_n/(1 - \tau\alpha_n), \qquad 1 - \tau\alpha_n > 1, \qquad (38)$$

que es estable para cualquier τ. Si $\alpha(t)$ cambia de signo para ciertos valores de t, entonces en aquellos nodos, donde $\alpha(t) > 0$, se debe emplear el esquema explícito (37), y en los nodos en que $\alpha(t) < 0$, el esquema implícito (38).

Los métodos de Adams son menos laboriosos en comparación con los de Runge—Kutta. La deficiencia de los métodos de Adams radica en lo que el comienzo de los cálculos no es usual; para determinar $y_1, y_2, \ldots, y_{m-1}$ se emplea corrientemente el método de Runge—Kutta. Si se emplean los esquemas de Adams de dos pasos (y, con mayor razón, de varios pasos) el cambio del paso τ requiere cierta complicación de las fórmulas, lo que no ocurre al emplear el método de Runge—Kutta. En la práctica se emplea la combinación de los métodos de Runge—Kutta y de Adams con un programa de elección automática del paso para la obtención de la exactitud prefijada.

§ 3. Aproximación del problema de Cauchy para un sistema de ecuaciones diferenciales lineales ordinarias de primer orden

1. Problema de Cauchy. En este párrafo se estudiarán los esquemas de diferencias lineales (de un paso o de dos capas) que surgen al aproximar el problema de Cauchy para un

sistema de ecuaciones diferenciales lineales ordinarias de primer orden, como también al aproximar las ecuaciones diferenciales con derivadas parciales (método de las rectas).

Tomemos un problema de Cauchy

$$\frac{du^i}{dt}+\sum_{j=1}^{N} a_{ij}u^j=f^i(t), \quad t\geqslant 0, \quad u^i(0)=u_0^i, \quad i=1, 2, \ldots, N. \quad (1)$$

Al designar por $A=(a_{ij})$ una matriz cuadrada de dimensión $N\times N$ con elementos a_{ij} que no dependen de t, por $u(t)=(u^1(t), u^2(t), \ldots, u^N(t))$ y $f(t)=(f^1(t), f^2(t), \ldots, f^N(t))$ los vectores buscado y prefijado de dimensión N, respectivamente, escribamos el sistema en la forma

$$\frac{du}{dt}+Au=f(t), \quad t\geqslant 0, \quad u(0)=u_0. \quad (2)$$

La misma designación A se empleará también para un operador correspondiente que actúa en el espacio H^N de dimensión N (A: $H^N\to H^N$). Introduzcamos en el espacio H^N un producto escalar (u, v) y una norma $\|u\|=\sqrt{(u, u)}$. Supondremos que el operador A es positivo

$$A>0, \text{ ó } (Ax, x)>0 \text{ para todos los } x\in H^N, \ x\neq 0.$$

El problema de Cauchy (1) bajo las condiciones (2) tiene la solución única. En efecto, supongamos que existen dos soluciones $\overline{u}(t)$ y $\overline{\overline{u}}(t)$ del problema (2). En este caso su diferencia satisface las condiciones homogéneas

$$\frac{dz}{dt}+Az=0, \quad t>0, \quad z(0)=0, \quad z(t)=\overline{u}(t)-\overline{\overline{u}}(t). \quad (3)$$

Multiplicando (3) escalarmente por z, y tomando en consideración que $\left(z, \frac{dz}{dt}\right)=\frac{1}{2}\frac{d}{dt}(z, z)$, obtenemos

$$\frac{1}{2}\frac{d}{dt}\|z\|^2+(Az, z)=0,$$

$$\|z(t)^2\| \int_0^t (Az(t'), z(t'))\,dt'=\|z(0)\|^2.$$

Puesto que $A > 0$, $z(0) = 0$, de aquí se deduce que

$$\| z(t) \|^2 = 0, \quad z(t) \equiv 0, \quad \bar{u}(t) \equiv \bar{\bar{u}}(t).$$

Indiquemos una propiedad de importancia de la solución del problema (2) cuando $f(t) \equiv 0$:

$$\| u(t) \| \leqslant e^{-\lambda_1 t} \| u(0) \|, \text{ si } A = A^* > 0, \qquad (4)$$

donde λ_1 es el valor propio mínimo del operador A:

$A\xi_k = \lambda_k \xi_k$, $k = 1, 2, \ldots, N$, $0 < \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N$.

Con el fin de demostrar (4) buscaremos la solución $u(t)$ del problema (2) en la forma

$$u(t) = \sum_{k=1}^{N} \alpha_k(t)\, \xi_k, \qquad \|u(t)\|^2 = \sum_{k=1}^{N} \alpha_k^2(t).$$

Al sustituir esta expresión en la ecuación (2) con $f(t) \equiv 0$, hallamos

$$\sum_{k=1}^{N} \left(\frac{d\alpha_k}{dt} + \lambda_k \alpha_k \right) \xi_k = 0,$$

y, por lo tanto, $\frac{d\alpha_k}{dt} + \lambda_k \alpha_k = 0$, $\alpha_k(t) = \alpha_k(0)\, e^{-\alpha_k t}$, de suerte que

$$\|u(t)\|^2 = \sum_{k=1}^{N} \alpha_k^2(0)\, e^{-2\lambda_k t} \leqslant e^{-2\lambda_1 t} \sum_{k=1}^{N} \alpha_k^2(0) =$$

$$= e^{-2\lambda_1 t} \|u(0)\|^2.$$

2. Esquemas de diferencias. Introduzcamos una red de paso τ según la variable t: $\omega_\tau = \{t_n = n\tau, n = 0, 1, 2, \ldots\}$ y designemos con $y_n = y(t_n)$ una función reticular del argumento $t_n = n\tau$ (o bien n) con valores en H^N. Escribamos un esquema explícito

$$\frac{y_{n+1} - y_n}{\tau} + A y_n = f_n, \quad n = 0, 1, 2, \ldots, \quad y_0 = u_0, \qquad (5)$$

de modo que y_{n+1} se halla por la fórmula explícita

$$y_{n+1} = y_n - \tau (A y_n - f_n), \; n = 0, 1, 2, \ldots, \; y_0 = u_0. \quad (5')$$

La solución y_n del problema (5) depende no sólo de τ, sino también de N o del parámetro $h = 1/N$: $y_n = y_{n,\tau,h}$. En realidad analizamos no solo un problema (5), sino un conjunto de problemas $\{5_{\tau,h}\}$ para cualesquiera τ y h. Esto es precisamente un esquema de diferencias. A título de su solución interviene una familia de funciones $\{y_{n,\tau,h}\}$. Para no complicar las notaciones, omitiremos los índices τ y h en los casos cuando esto no lleve a equivocaciones algunas. El esquema (5) es *esquema de diferencias de un paso* (o *de dos capas*).

En general, por *esquema de dos capas* se entiende una ecuación que liga los valores del vector $y(t)$ para dos valores del argumento $t = t_n$ y $t = t_{n+1}$ (para dos capas):

$$By_{n+1} = Cy_n + F_n, \qquad n = 0, 1, \ldots,$$

donde B, C son las matrices cuadradas $N \times N$ (operadores lineales B, C: $H^N \to H^N$); y_n, F_n son los vectores de dimensión N. Dicha ecuación puede ser siempre escrita en la siguiente forma canónica:

$$B\frac{y_{n+1}-y_n}{\tau} + Ay_n = \varphi_n, \quad n = 0, 1, 2, \ldots, \quad y_0 = u_0. \tag{6}$$

Para determinar y_{n+1} se debe resolver la ecuación

$$BY_{n+1} = \Phi_n, \quad \Phi_n = By_n - \tau(Ay_n - \varphi_n).$$

Supondremos siempre que existe el operador inverso B^{-1}.

Si $B = E$ es un operador unidad, obtendremos el *esquema explícito* (5). Cuando $B \neq E$, el esquema (6) se denomina *implícito*. Se encuentran con frecuencia los esquemas

$$\frac{y_{n+1}-y_n}{\tau} + Ay_{n+1} = \varphi_n \text{ (esquema explícito puro)}, \tag{7}$$

$$\frac{y_{n+1}-y_n}{\tau} + \frac{1}{2}A(y_n + y_{n+1}) = \varphi_n \text{ (esquema simétrico)}. \tag{8}$$

Representan ambos los casos particulares (para $\sigma = 1$ y $\sigma = 1/2$) del *esquema con pesos*

$$\frac{y_{n+1}-y_n}{\tau} + A(\sigma y_{n+1} + (1-\sigma)y_n) = \varphi_n, \quad u = 0, 1, \ldots, \tag{9}$$

el cual puede escribirse en la forma canónica (6) con

$$B = E + \sigma\tau A, \tag{10}$$

si tenemos en cuenta que $\sigma y_{n+1} + (1 - \sigma)\, y_n = y_n + \sigma\tau\,(y_{n+1} - y_n)/\tau$.

3. Error de aproximación. Sea $u = u(t)$ la solución del problema (2) y sea $y_n = y\,(t_n)$ la solución del problema (6); al sustituir en (6) $y_n = u_n + z_n$, obtenemos para el error $z_n = y_n - u_n$, $u_n = u\,(t_n)$:

$$B\frac{z_{n+1} - z_n}{\tau} + Az_n = \psi_n, \quad n = 0, 1, 2, \ldots, \quad z_0 = 0, \quad (11)$$

donde

$$\psi_n = \varphi_n - Au_n - B\frac{u_{n+1} - u_n}{\tau} \quad (12)$$

es el residuo o error de aproximación para el esquema (6) en la solución $u = u\,(t)$ del problema de partida (2).

Sean $\| u \|_1$, $\| v \|_{(2)}$ ciertas normas en $H^N = H_h$. El esquema (6) converge, si $\| z_n \|_{(1)} \to 0$ para $\tau \to 0$, cualquiera que sea $n = 1, 2, \ldots$. El esquema (6) es de *m-ésimo orden de exactitud*, o converge con la velocidad $O\,(\tau^m)$, siempre que

$$\| z_n \|_{(1)} = O\,(\tau^m), \text{ es decir } \| z_n \|_{(1)} \leqslant M\tau^m, \quad (13)$$

donde $M = \text{const}$ no depende de τ.

Recordemos que el esquema (6) es de *m-ésimo orden de aproximación* en la solución de la ecuación (1), si para el residuo ψ_n se cumple la estimación

$$\| \psi_n \|_{(2)} = O_{\_}(\tau^m). \quad (14)$$

Aclaremos las condiciones de aproximación del esquema (6) con $m = 1, 2$. Suponiendo que $u = u\,(t)$ tiene tantas derivadas cuanto sea necesario en el transcurso de exposición, encontramos

$$u_{n+1} = \left(u + \frac{\tau}{2}\dot{u} + \frac{\tau^2}{8}\ddot{u}\right)_{n+1/2} + O\,(\tau^3),$$

$$\dot{u}_n = \left(\frac{du}{dt}\right)_n, \quad \ddot{u}_n = \left(\frac{d^2u}{dt^2}\right)_n,$$

$$u_n = \left(u - \frac{\tau}{2}\dot{u} + \frac{\tau^2}{8}\ddot{u}\right)_{n+1/2} + O\,(\tau^2),$$

$$\frac{1}{\tau}(u_{n+1} - u_n) = \dot{u}_{n+1/2} + O\,(\tau^2),$$

$$\psi_n = \varphi_n - (Au + B\dot{u})_{n+1/2} + \frac{\tau}{2} A\dot{u}_{n+1/2} + O(\tau^2) =$$

$$= \varphi_n - f_{n+1/2} + (f - Au - \dot{u})_{n+1/2} +$$

$$+ \left(E - B + \frac{\tau}{2} A\right) \dot{u}_{n+1/2} + O(\tau^2) =$$

$$= \varphi_n - f_{n+1/2} + \left(E - B + \frac{\tau}{2} A\right) \dot{u}_{n+1/2} + O(\tau^2).$$

De aquí se ve que la condición (14) se cumplirá, si

$$\|\varphi_n - f_{n+1/2}\|_{(2)} = O(\tau^m),$$

$$\left\| \left(E - B + \frac{\tau}{2} A\right) \dot{u} \right\|_{(2)} = O(\tau^m), \quad m = 1, 2. \tag{15}$$

En particular, para el esquema explícito (en el caso $B = E$) tenemos

$$\left\| \frac{\tau}{2} A\dot{u} \right\|_{(2)} = O(\tau),$$

y $\|\psi_n\|_{(2)} = O(\tau)$ cuando $\|\varphi_n - f_{n+1/2}\| = O(\tau)$, por ejemplo, cuando $\varphi_n = f_n$.

Si, en el caso del esquema simétrico ($\sigma = 1/2$) $B = E + \tau A/2$, $\|\varphi_n - f_{n+1/2}\|_{(2)} = O(\tau^2)$, entonces $\|\psi_n\|_{(2)} = O(\tau^2)$, puesto que $\|(E - B + \tau A/2)\dot{u}\|_{(2)} = 0$; en tal caso podemos tomar, por ejemplo, $\varphi_n = f_{n+1/2}$.

El esquema de adelantamiento ($\sigma = 1$) es de primer orden de aproximación, puesto que $\|(E - B + \tau A/2)\dot{u}\|_{(2)} = \tau \|A\dot{u}\|_{(2)}/2 = O(\tau)$.

4. Estabilidad y convergencia. Según se ha observado más arriba, el esquema (6) es *estable* (respecto de los datos iniciales y del segundo miembro), si su solución depende continuamente de los datos de entrada (de y_0 y de φ_n), con la particularidad de que dicha dependencia es continua según τ y N, o según h. Para estimar la solución del problema se usará la norma $\|u\|_{(1)}$, y para estimar el segundo miembro, la norma $\|v\|_2$. *Hagamos uso de la definición de estabilidad.*

El esquema (6) será *estable*, si para cualesquiera y_0, φ_n existen tales constantes $M_1 > 0$ y $M_2 > 0$, independientes

tanto de τ, como de N, y_0, φ_n, que para la solución del problema (6) se cumple la desigualdad

$$\| y_n \|_{(1)} \leqslant M_1 \| y_0 \|_{(1)} + M_2 \max_{0 \leqslant k < n} \| \varphi_k \|_{(2)}. \quad (16)$$

Si el esquema (6) es estable y posee una aproximación $\| \psi_n \|_{(2)} \to 0$ cuando $\tau \to 0$, es convergente:

$$\| y_n - u_n \|_{(1)} \to 0 \text{ cuando } \tau \to 0, \; n = 1, 2, \ldots \quad (17)$$

(de la aproximación y de la estabilidad se desprende la convergencia del esquema). En efecto, si el esquema (6) es estable, entonces para la solución $z_n = y_n - u_n$ del problema (11) se cumple, de acuerdo con (16), la estimación

$$\|z_n\|_{(1)} \leqslant M_1 \max_{0 \leqslant k < n} \|\psi_k\|_{(2)}. \quad (18)$$

De aquí precisamente proviene que $\| z_n \|_{(1)} \to 0$, si $\| \psi_n \|_{(2)} \to 0$ cuando $\tau \to 0$.

El estudio de la convergencia y del orden de exactitud se reduce al estudio del error de aproximación y de estabilidad del esquema de diferencias (6).

§ 4. Estabilidad del esquema de dos capas

1. Estabilidad respecto de los datos iniciales. Examinaremos un esquema de dos capas en la forma canónica

$$B \frac{y_{n+1} - y_n}{\tau} + A y_n = \varphi_n, \quad n = 0, 1, \ldots,$$

se ha prefijado el valor inicial $y_0 \in H$, (1)

donde A, B: $H \to H$ ($H = H^N$). La solución del problema (1) puede ser representada como una suma $y = y^{(1)} + y^{(2)}$ de las soluciones de dos problemas

$$B \frac{y_{n+1} - y_n}{\tau} + A y_n = 0, \quad n = 0, 1, \ldots, \; y_0 = u_0, \quad (2)$$

$$B \frac{y_{n+1} - y_n}{\tau} + A y_n = \varphi_n, \quad n = 0, 1, \ldots, \; y_0 = 0, \quad (3)$$

($y^{(1)}$ es la solución del problema (2), $y^{(2)}$ es la solución del problema (3)).

El esquema (1) es estable respecto de los datos iniciales, si para la solución del problema (2) es justa la estimación

$$\| y_n \|_{(1)} \leqslant M_1 \| y_0 \|_{(1)}. \qquad (4)$$

El esquema (1) es estable respecto del segundo miembro, si para la solución del problema (3) es justa la estimación

$$\|y_n\|_{(1)} \leqslant M_2 \max_{0 \leqslant k < n} \|\varphi_k\|_{(2)}. \qquad (5)$$

Aquí M_1, M_2 no dependen de N, τ, n.

Utilizaremos una condición más simple para la estabilidad respecto de los datos iniciales:

$$\| y_n \|_{(1)} \leqslant \| y_{n-1} \|_{(1)}, \ \ldots, \ \| y_1 \|_{(1)} \leqslant \leqslant \| y_0 \|_{(1)} \ (M_1 = 1), \qquad (6)$$

y, además, la condición de ρ-estabilidad:

$$\| y_n \|_{(1)} \leqslant \rho \| y_{n-1} \|_{(1)} \leqslant \ldots \leqslant \rho^n \| y_0 \|_{(1)}, \quad \rho > 0. \qquad (7)$$

Es evidente que el esquema es estable en el sentido de la definición (4), si $\rho = e^{c_0 \tau}$, donde $c_0 = \text{const}$ no depende de n, τ, N. En este caso $\rho^n = e^{c_0 t_n} \leqslant e^{c_0 T} = M_1$ para $0 \leqslant t_n \leqslant T$, $c_0 > 0$, o bien $\rho^n \leqslant 1$ para $c_0 \leqslant 0$.

Introduzcamos en el espacio H un producto escalar $(,)$ y una norma $\| x \| = \sqrt{(x, x)}$. Sea $D = D^* > 0$ un operador *positivo autoconjugado*. A título de norma $\| y \|_{(1)}$ elijamos una norma energética

$$\|y\|_{(1)} = \|y\|_D = \sqrt{(Dy, y)}, \qquad (8)$$

En particular, $D = A$, $D = E$ o $D = B$ (para $B = B^* > 0$). De (2) proviene que

$$y_{n+1} = S y_n, \quad S = E - \tau B^{-1} A, \qquad (9)$$

donde S es el operador de transición de una capa a la otra.

El esquema (2) es estable en H_D, si es justa la estimación

$$\| y_{n+1} \|_D \leqslant \| y_n \|_D. \qquad (10)$$

De la estimación $\| y_{n+1} \|_D = \| S y_n \|_D \leqslant \| S \|_D \| y_n \|_D$ se deduce que la desigualdad (10) es equivalente a la condición

$$\| S \|_D \leqslant 1. \qquad (11)$$

Esta última condición es equivalente, a su vez, a la siguiente

$$J_D = \| y \|_D^2 - \| Sy \|_D^2 = (Dy, y) - (DSy, Sy) \geqslant 0$$
para todo $y \in H$. (12)

De este modo, (10), (11) y (12) son equivalentes, es decir, el cumplimiento de cualquiera de ellas provoca el cumplimiento de dos otras.

2. Condición necesaria y suficiente de estabilidad. Teorema fundamental.

TEOREMA 1. *Si $A = A^*$ es un operador positivo autoconjugado y existe el operador B^{-1}, entonces para que el esquema* (2) *sea estable en H_A:*

$$\| y_{n+1} \|_A \leqslant \| y_n \|_A \tag{13}$$

es necesario y suficiente que se verifique la desigualdad

$$(By, y) - \frac{\tau}{2} (Ay, y) \geqslant 0 \text{ para todo } y \in H, \text{ o bien}$$
$$B \geqslant \frac{\tau}{2} A. \tag{14}$$

DEMOSTRACION. Es suficiente convencerse de equivalencia entre (14) y la desigualdad $J_A \geqslant 0$, donde

$$J_A = (Ay, y) - (ASy, Sy) =$$
$$= (Ay, y) - (Ay - \tau AB^{-1}Ay, y - \tau B^{-1}Ay) =$$
$$= 2\tau (AB^{-1}Ay, y) - \tau^2 (AB^{-1}Ay, B^{-1}Ay).$$

Al designar $B^{-1}Ay = x$, $Ay = Bx$, obtendremos

$$J_A = 2\tau \left((Bx, x) - \frac{\tau}{2} (Ax, x) \right) \geqslant 0 \text{ para todo } x \in H, \tag{15}$$

es decir, las desigualdades (14), (15) y, por lo tanto, (13), (14) son equivalentes. Esto quiere decir que de (14) provienen (11), (12) para $D = A$ y (13) (la condición (14) es suficiente para la estabilidad). Por cuanto el esquema es estable, es decir, se verifica (13) o bien $\| S \|_A \leqslant 1$, entonces $J_A \geqslant 0$, y, por consiguiente, $B \geqslant \tau A/2$ (necesidad de la condición (14)).

OBSERVACION. La condición (14) puede aclararse con un ejemplo del esquema de diferencias

$$b \frac{y_{n+1} - y_n}{\tau} + a y_n = 0, \quad n = 0, 1, 2, \ldots, \quad a > 0, \quad b > 0$$

con coeficientes numéricos a, b. Este esquema corresponde al problema de Cauchy

$$bu'(t) + au(t) = 0 \quad t > 0, \; u(0) = u_0.$$

De la fórmula $y_{n+1} = (1 - \tau a/b)\, y_n$ se ve que el esquema es estable, es decir, $|y_{n+1}| \leqslant |y_n| \leqslant \ldots \leqslant |y_0|$, si $|1 - \tau a/b| \leqslant 1$, $-1 \leqslant 1 - \tau p/b \leqslant 1$, es decir, $b \geqslant \tau a/2$. La analogía con la desigualdad operacional $B \geqslant \tau A/2$ es evidente.

3. Ejemplos de aplicación del teorema fundamental.

EJEMPLO 1. Un esquema explícito: $B = E$, $A = A^* > 0$. De la desigualdad de Cauchy—Buniakovski $(Ax, x) \leqslant \leqslant \|Ax\| \|x\| \leqslant \|A\| \|x\|^2$ se deduce $A \leqslant \|A\| E$, o bien

$$E \geqslant \frac{1}{A} A \tag{16}$$

Veamos ahora la diferencia $B - \frac{1}{2}\tau A = E - \frac{1}{2}\tau A \geqslant \frac{1}{\|A\|} A - \frac{1}{2}\tau A = \left(\frac{1}{\|A\|} - \frac{\tau}{2}\right) A$. Como que $A > 0$, entonces la condición $B - \frac{1}{2}\tau A \geqslant 0$ se cumplirá para $\frac{1}{\|A\|} - \frac{\tau}{2} \geqslant 0$, es decir, cuando

$$\tau \leqslant 2/\|A\|. \tag{17}$$

Esta es una condición necesaria y suficiente de estabilidad del esquema explícito en H_A ($\|y_n\|_A \leqslant \|y_0\|_A$).

EJEMPLO 2. Esquema (9) del § 3 con pesos, $A = A^* > 0$. Para dicho esquema $B = E + \sigma\tau A$ y $B - \frac{1}{2}\tau A = E + \left(\sigma - \frac{1}{2}\right)\tau A \geqslant \frac{1}{\|A\|} + \left(\sigma - \frac{1}{2}\right)\tau\Big) A \geqslant 0$, siempre que

$$1 + \left(\sigma - \frac{1}{2}\right)\tau\|A\| \geqslant 0. \tag{18}$$

De aquí se ve que el esquema con pesos es estable en H_A para todo $\tau > 0$ (incondicionalmente estable), si $\sigma \geqslant 1/2$, y es condicionalmente estable para $\tau \leqslant 1/(1/2 - \sigma)\|A\|$, si $\sigma < 1/2$.

EJEMPLO 3. Estabilidad en H (para $D = E$) del esquema con pesos (9) del § 3:

$$(E + \sigma\tau A)\frac{y_{n+1} - y_n}{\tau} + Ay_n = 0, \quad n = 0, 1, 2, \ldots,$$

$$B = E + \sigma\tau A. \tag{19}$$

Aplicando el operador A^{-1} a los dos miembros de la ecuación (19), obtenemos

$$\widetilde{B}\frac{y_{n+1} - y_n}{\tau} + \widetilde{A}y_n = 0, \quad n = 0, 1, 2, \ldots,$$

$$\widetilde{B} = A^{-1} + \sigma\tau E, \quad \widetilde{A} = E, \tag{20}$$

Este esquema es estable, en virtud del teorema 1, en $H_{\widetilde{A}} = H$ $(\widetilde{A}^* = \widetilde{A} = E > 0)$ para $\widetilde{B} - \frac{1}{2}\tau\widetilde{A} = A^{-1} + \left(\sigma - \frac{1}{2}\right)\tau E \geqslant \left(\frac{1}{\|A\|} + \left(\sigma - \frac{1}{2}\right)\tau\right) E \geqslant 0$, es decir, si se cumple (18) (en este caso se ha tomado en consideración la estimación $A^{-1} \geqslant \frac{1}{\|A\|}E$, la cual se deduce de (16)). De este modo, de (18) se infiere que para (19) es justa la estimación (10) cuando $D = \widetilde{A}$, es decir,

$$\|y_n\| \leqslant \|y_0\|. \tag{21}$$

El esquema (19) puede ser escrito en la forma

$$y_{n+1} = Sy_n, \quad S = (E + \sigma\tau A)^{-1}(E - (1 - \sigma)\tau A),$$

$$A = A^* > 0. \tag{22}$$

Por eso, si se cumple la condición (18), para dicho esquema es justa la estimación (21), lo que significa

$$\|(E + \sigma\tau A)^{-1}(E - (1 - \sigma)\tau A)\| \leqslant 1,$$

$$\text{siempre que } 1 + \left(\sigma - \frac{1}{2}\right)\tau\|A\| \geqslant 0. \tag{23}$$

Esta estimación nos hará falta en lo que sigue.

4. Estabilidad en H .

TEOREMA 2. *Si $A = A^* > 0$, $B = B^* > 0$, entonces para que el esquema* (2) *sea estable en H_B:*

$$\|y_{n+1}\|_B \leqslant \|y_n\|_B, \tag{24}$$

es necesario y suficiente que se cumpla la condición (14).

DEMOSTRACIÓN. Escribamos el esquema (2) en la forma (9) y mostremos que la condición

$$\| S \|_B \leqslant 1 \tag{25}$$

es equivalente a la desigualdad (14), es decir, de (14) proviene (25), y, viceversa, de (25) proviene (14).

Sea y un vector arbitrario de H; representémoslo en la forma

$$y = \sum_{k=1}^{N} \alpha_k \xi_k,$$

donde $\{\xi_k\}$ son los vectores propios del problema

$$A\xi_k = \lambda_k B\xi_k, \qquad \lambda_k > 0,$$
$$(B\xi_k, \xi_m) = \delta_{km} = \begin{cases} 1, & k = m, \ k, \ m = 1, \ 2, \ \ldots, \ N \\ 0, & k \neq m. \end{cases} \tag{26}$$

Teniendo presente que $S\xi_k = \xi_k - \tau B^{-1} A\xi_k = (1 - \tau\lambda_k)\,\xi_k$, $BS\xi_k = (1 - \tau\lambda_k)\,B\xi_k$, encontramos

$$(By, \ y) = \sum_{k=1}^{N} \alpha_k^2, \qquad (Ay, \ y) = \sum_{k=1}^{N} \lambda_k \alpha_k^2,$$
$$(BSy, \ Sy) = \sum_{k=1}^{N} \alpha_k^2 (1 - \tau\lambda_k)^2 \leqslant \|S\|_B^2 \sum_{k=1}^{N} \alpha_k^2 = \tag{27}$$
$$= \|S\|_B^2 \, (By, \ y),$$

donde

$$\|S\|_B^2 = \max_{1 \leqslant k \leqslant N} (1 - \tau\lambda_k)^2, \tag{28}$$

La desigualdad (25) es equivalente a la condición

$$\tau\lambda_k \leqslant 2, \qquad k = 1, \ 2, \ \ldots, \ N, \tag{29}$$

la cual, a su vez, es equivalente a la desigualdad (14), puesto que

$$(By, \ y) - \frac{\tau}{2} (Ay, \ y) = \sum_{k=1}^{N} \alpha_k^2 \left(1 - \frac{\tau\lambda_k}{2}\right).$$

Con ello queda demostrada la equivalencia de (24) y (14).

5. ρ-estabilidad.

TEOREMA 3. *Si* $A = A^* > 0$, $B = B^* > 0$, *entonces la condición necesaria y suficiente de la ρ-estabilidad del esquema* (2) *con* $\rho > 0$ *cualquiera*:

$$\| y_{n+1} \|_D \leqslant \rho \| y_n \|_D, \quad D = A, B, \tag{30}$$

será representada por las desigualdades operacionales

$$\frac{1-\rho}{\tau} B \leqslant A \leqslant \frac{1+\rho}{\tau} B. \tag{31}$$

DEMOSTRACION. Las desigualdades (31) son equivalentes a las condiciones (véase el cap. I, § 4, p. 4):

$$\frac{1-\rho}{\tau} \leqslant \lambda_k \leqslant \frac{1+\rho}{\tau}, \quad k = 1, 2, \ldots, N, \tag{32}$$

donde λ_k son los números propios del problema (26).

Admitamos que $D = B$ y son justas (31) ó (32). De (32) se deduce $-\rho \leqslant \tau\lambda_k - 1 \leqslant \rho$, $|1 - \tau\lambda_k| \leqslant \rho$, y, en virtud de (27), $\| S \|_B \leqslant \rho$ (puesto que $\| S \|_B^2$ es una constante mínima, para la cual se verifica la desigualdad $(BSy, Sy) \leqslant M(By, y)$, es decir, es justa la estimación (30) suficiencia). Si es justa la estimación (30), entonces $|1 - \tau\lambda_k| \leqslant \rho$, y, por lo tanto, quedan cumplidas (32) y (31) (necesidad).

Análogamente se demuestra el teorema para $D = A$, si se toma en consideración que

$$(ASy, Sy) = \sum_{k=1}^{N} \alpha_k^2 \lambda_k (1 - \tau\lambda_k)^2 \leqslant \max_{1 \leqslant k \leqslant N} (1 - \tau\lambda_k)^2 (Ay, y).$$

De (30) se infiere

$$\| y_n \|_D \leqslant \rho^n \| y_0 \|_D.$$

Surge una pregunta: ¿en qué condiciones tiene lugar la estimación apriorística (30) con $\rho < 1$? La respuesta se ofrece por el siguiente

TEOREMA 4. Supongamos cumplidas las condiciones

$$A = A^* > 0, \quad B = B^* > 0, \quad \gamma_1 B \leqslant A \leqslant \gamma_2 B, \quad \gamma_1 > 0. \tag{33}$$

En este caso para la solución del problema (2) *es justa la estimación*

$$\| y_{n+1} \|_D \leqslant \rho \| y_n \|_D, \quad \rho = 1 - \tau\gamma_1, \quad D = A, B, \tag{34}$$

siempre que

$$\tau \leqslant \tau_0, \qquad \tau_0 = \frac{2}{\gamma_1 + \gamma_2}. \tag{35}$$

Para la demostración se debe calcular la norma $\| S \|_B = \| S \|_A = \max\limits_{1 \leqslant k \leqslant N} | 1 - \tau\lambda_k |$ bajo la condición de que $\gamma_1 \leqslant \lambda_k \leqslant \gamma_2$, $0 < \gamma_1 = \lambda_1 \leqslant \lambda_2 \leqslant \ldots \leqslant \lambda_N = \lambda_2$. Veamos una diferencia $\varphi_k = (1 - \tau\lambda_1)^2 - (1 - \tau\lambda_k)^2 = 2\tau\,(\lambda_k - \lambda_1) \times \left(1 - \frac{\tau}{2}\,(\lambda_k - \lambda_1)\right)$. De aquí se ve que $\varphi_k \geqslant 0$ para $1 - \frac{\tau}{2}\,(\lambda_k + \lambda_1) \geqslant 1 - \frac{\tau}{2}\,(\gamma_2 + \gamma_1) \geqslant 1 - \frac{\tau_0}{2}\,(\gamma_1 - \gamma_2) = 0$, es decir, $\max\limits_{1 \leqslant k \leqslant N} | 1 - \tau\lambda_k | = 1 - \tau\gamma_1$, si $\tau \leqslant \tau_0$. El teorema está demostrado.

6. Estabilidad respecto del segundo miembro. Método de las desigualdades energéticas. Analicemos el problema y reescribámoslo en la forma

$$y_{n+1} + Sy_n + \tau B^{-1}\varphi_n, \quad n = 0, 1, \ldots,$$
$$S = E - \tau B^{-1}A, \quad y_0 = 0. \tag{36}$$

Hagamos uso de la desigualdad triangular

$$\| y_{n+1} \|_D \leqslant \| Sy_n \|_D + \tau \| B^{-1}{}_n \|_D \leqslant$$
$$\leqslant \| S \|_D \| y_n \|_D + \tau \| B^{-1}\varphi_n \|_D. \tag{37}$$

Si se cumplen las condiciones del teorema 2, entonces $B = B^* > 0$, $D = B$, y $\| S \|_D = \| S \|_B \leqslant 1$ para $B \geqslant \frac{\tau}{2}A$, $\| B^{-1}\varphi_n \|_B^2 = (B\,(B^{-1}\varphi_n),\ B^{-1}\varphi_n) =$
$= (B^{-1}\varphi_n,\ \varphi_n) = \| \varphi_n \|_{B^{-1}}^2$, y de (37) proviene

$$\| y_{n+1} \|_B \leqslant \| y_n \|_B + \tau \| \varphi_n \|_{B^{-1}}.$$

Al sumar según $n = 0, 1, 2 \ldots$ y teniendo presente que $y_0 = 0$, obtendremos

$$\| y_n \|_B \leqslant \sum_{k=0}^{n-1} \tau \| \varphi_k \|_{B^{-1}}. \tag{38}$$

Esta estimación apriorística expresa la estabilidad del esquema (1) respecto del segundo miembro bajo la misma condición (14).

Pueden obtenerse también otras estimaciones. Con este fin aprovechemos un método, bastante general, de desigualdades energéticas. Sustituyamos $y_n = \frac{1}{2}(y_n + y_{n+1}) - \frac{\tau}{2} \times \frac{y_{n+1} - y_n}{\tau}$ en (1):

$$\left(B - \frac{\tau}{2} A\right) \frac{y_{n+1} - y_n}{\tau} + \frac{1}{2} A (y_{n+1} + y_n) = \varphi_n.$$

Multipliquemos esta ecuación escalarmente por $2(y_{n+1} - y_n)$ y tengamos en cuenta que $(A(y_{n+1} + y_n), y_{n+1} - y_n) = (Ay_{n+1}, y_{n+1}) + (Ay_n, y_{n+1}) + (Ay_{n+1}, y_n) - (Ay_n, y_n) = (Ay_{n+1}, y_{n+1}) - (Ay_n, y_n)$, puesto que $(Ay_n, y_{n+1}) = (Ay_{n+1}, y_n)$ en virtud de que A es autoconjugado. De resultas se obtiene una «identidad energética»

$$2\tau \left(\left(B - \frac{\tau}{2} A\right) \frac{y_{n+1} - y_n}{\tau}, \frac{y_{n+1} - y_n}{\tau}\right) + (Ay_{n+1}, y_{n+1}) = (Ay_n, y_n) + 2(\varphi_n, y_{n+1} - y_n). \quad (39)$$

De aquí se ve que para $\varphi_n = 0$ y $B \geqslant \frac{\tau}{2} A$ será justa la estimación (13).

Transformemos $2(\varphi_n, y_{n+1} - y_n) = 2\tau \left(\varphi_n, \frac{y_{n+1} - y_n}{\tau}\right)$. Con este fin hagamos uso de la desigualdad

$$|ab| = (\sqrt{2\varepsilon a}) \left(\sqrt{\frac{1}{2\varepsilon}} b\right) \leqslant \varepsilon a^2 + \frac{1}{4\varepsilon} b^2,$$

donde a, b, $\varepsilon > 0$ son unos números cualesquiera. En nuestro caso

$$2(\varphi_n, y_{n+1} - y_n) \leqslant 2\tau \|\varphi_n\| \left\| \frac{y_{n+1} - y_n}{\tau} \right\| \leqslant 2\tau\varepsilon \left\| \frac{y_{n+1} - y_n}{\tau} \right\|^2 + \frac{\tau}{2\varepsilon} \|\varphi_n\|^2.$$

Al sustituir esta estimación en la identidad (39), obtendremos

$$2\tau \left(\left(B - \varepsilon E - \frac{\tau}{2} A\right) \frac{y_{n+1} - y_n}{\tau}, \frac{y_{n+1} - y_n}{\tau}\right) + \|y_{n+1}\|_A^2 \leqslant \|y_n\|_A^2 + \frac{\tau}{2\varepsilon} \|\varphi_n\|^2. \quad (40)$$

Si se cumple la desigualdad

$$B \geqslant \varepsilon E + \frac{\tau}{2} A, \qquad \varepsilon > 0, \tag{41}$$

entonces de (40) proviene (sustituyendo n por k)

$$\| y_{k+1} \|_A^2 \leqslant \| y_k \|_A^2 + \frac{\tau}{2\varepsilon} \| \varphi_k \|^2.$$

Al sumar según $k = 0, 1, 2, \ldots, n-1$, obtenemos una estimación

$$\| y_n \|_A^2 \leqslant \| y_0 \|_A^2 + \frac{1}{2\varepsilon} \sum_{k=0}^{n-1} \tau \| \varphi_k \|^2, \tag{42}$$

que expresa la estabilidad del esquema (1) respecto del segundo miembro y de los datos iniciales en H_A.

EJEMPLO. Esquema con pesos (1): $B = E + \sigma\tau A$. Para dicho esquema la condición (41) significa que

$$(1 - \varepsilon) E + \left(\sigma - \frac{1}{2}\right) \tau A \geqslant 0.$$

En particular, la estimación (42) es justa para $\varepsilon = 1$ y $\sigma \geqslant 1/2$.

7. Estabilidad asintótica. Para el problema de Cauchy

$$\frac{du}{dt} + Au = 0, \qquad t > 0 \qquad u(0) = u_0$$

se ha obtenido en el § 3, p. 1, la estimación

$$\| u(t) \| \leqslant e^{-\gamma_1 t} \| u(0) \|,$$

donde $\lambda_1 = \min\limits_k \lambda_k (A)$.

Buscaremos las condiciones, bajo las cuales la estimación análoga tiene lugar para el esquema (2). Usaremos el teorema 4. Supongamos cumplidas las condiciones (33). Entonces, debido a (34), (35)

$$\| y_n \|_A \leqslant \rho^n \| y_0 \|_A, \quad \rho = 1 - \tau\gamma_1, \quad \tau \leqslant \tau_0 = \frac{2}{\gamma_1 + \gamma_2}. \tag{43}$$

De aquí se desprende una estimación que expresa la propiedad de la estabilidad asintótica

$$\| y_n \|_A \leqslant e^{-\gamma_1 t n} \| y_0 \|_A \tag{44}$$

(aquí se ha tomado en consideración que $\rho = 1 - \tau\gamma_1 < e^{-\tau\gamma_1}$).

Analicemos el esquema con pesos y supongamos que

$$\delta E \leqslant A \leqslant \Delta E, \quad \delta = \lambda_1 > 0, \quad \Delta = \lambda_N > 0. \quad (45)$$

Calculemos γ_1 y γ_2. Teniendo presente (45), tenemos

$$B = E + \sigma\tau A \geqslant \left(\frac{1}{\Delta} + \sigma\tau\right) A = \frac{1}{\gamma_2} A;$$

$$B \leqslant \left(\frac{1}{\delta} + \sigma\tau\right) A = \frac{1}{\gamma_1} A, \quad (46)$$

$$\gamma_1 = \frac{\delta}{1+\sigma\tau\delta}, \qquad \gamma_2 = \frac{\Delta}{1+\sigma\tau\Delta}.$$

Para un esquema explícito $\gamma_1 = \delta$, $\gamma_2 = \Delta$ la condición de estabilidad asintótica

$$\tau \leqslant 2/(\delta + \Delta) \quad (47)$$

es próxima a la condición de estabilidad usual con $\rho = 1$. Cuando $\sigma \neq 0$, la condición $\tau \leqslant 2/(\gamma_1 + \gamma_2)$ conduce a una desigualdad

$$2 + 2(\sigma - 1/2)\tau(\delta + \Delta) - 2\sigma(1 - \sigma)\tau^2\delta\Delta \geqslant 0.$$

Cuando $\sigma = 1$, se cumple para cualquier τ, es decir, un esquema implícito puro con $\sigma = 1$ es incondicionalmente estable asintóticamente. El esquema simétrica

$$\frac{y_{n+1} - y_n}{\tau} + \frac{1}{2} A(y_{n+1} + y_n) = 0, \quad \sigma = \frac{1}{2}, \quad (48)$$

es asintóticamente estable a condición de que

$$\tau \leqslant \tau^*, \quad \tau^* = 2/\sqrt{\delta\Delta}, \quad (49)$$

y incondicionalmente estable en el sentido habitual. En este caso

$$\rho = e^{-\lambda_1\tau + O(\tau^2)} < e^{-\lambda_1\tau}$$

y resulta lícita la estimación

$$\|y_n\| \leqslant e^{-\lambda_1 t_n} \|y_0\| \quad \text{para} \quad \tau \leqslant \tau^*, \quad \sigma = 1/2. \quad (50)$$

¿Qué sucederá, si la condición $\tau \leqslant \tau_0$ no se cumple, es decir, si $\tau > \tau_0$? Entonces, máx $|1 - \tau\lambda_k|$ se logra no para $k = 1$, sino para $k = N$ y $\rho = \tau\gamma_2 - 1$. La asintótica de la solución de un problema de diferencias (con t_n grandes) nada tiene en común con la solución asintótica del problema de partida. De este modo, la perturbación de la estabilidad asintótica lleva a la pérdida de la exactitud del esquema para t grandes.

Capítulo VI

Métodos de diferencias para las ecuaciones elípticas

En este capítulo se examinarán los esquemas de diferencias y los métodos de resolución de las ecuaciones en diferencias para la ecuación de Poisson y ecuaciones elípticas de coeficientes variables.

§ 1. Esquemas de diferencias para la ecuación de Poisson

1. Problema de partida. Examinemos la ecuación de Poisson

$$\Delta u = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} = -f(x_1, x_2). \tag{1}$$

Buscaremos su solución que sea continua en un rectángulo

$$\overline{G} = G \cup \Gamma = \{x = (x_1, x_2): 0 \leqslant x_\alpha \leqslant l_\alpha, \quad \alpha = 1, 2\}$$

y que tome en la frontera Γ los valores prefijados:

$$u|\Gamma = \mu(x) \tag{2}$$

Un *problema* definido por la ecuación (1) y la condición (2). recibe el nombre *de Dirichlet* (*primer problema de contorno*).

2. Esquema de diferencias «cruz». Para la resolución numérica del problema (1), (2), introduzcamos en $\overline{G}$ una red $\overline{\omega}_h = \omega_h \cup \gamma_h = \{x_i = (i_1 h_1, i_2 h_2), i_\alpha = 0, 1, \ldots, N\alpha, h_\alpha = l_\alpha/N_\alpha, \alpha = 1, 2\}$ y designemos con $y_i = y_{i_1 i_2} = y(i_1, i_2) = y(x_i)$ una función reticular definida sobre $\overline{\omega}_h$; h_1 y h_2 son los pasos de la red según las coordenadas de x_1 y x_2.

Con el fin de escribir un esquema de diferencias para (1), (2) aproximemos cada una de las derivadas $\partial^2 u/\partial x_\alpha^2$ en un

molde tripuntual, suponiendo

$$\frac{\partial^2 u}{\partial x_1^2} \sim \frac{u(x_1-h_1, x_2)-2u(x_1+x_2)+u(x_1+h_1, x_2)}{h_1^2} = u_{\bar{x}_1 x_1},$$

$$\frac{\partial^2 u}{\partial x_2^2} \sim \frac{u(x_1, x_2-h_2)-2u(x_1, x_2)+u(x_1, x_2+h_2)}{h_2^2} = u_{\bar{x}_2 x_2},$$

el signo $\sim$ significa la aproximación. Haciendo uso de estas expresiones, sustituyamos (1) por la ecuación en diferencias

$$\frac{y(i_1-1, i_2)-2y(i_1, i_2)+y(i_1+1, i_2)}{h_1^2} + \\ + \frac{y(i_1, i_2-1)-2y(i_1, i_2)+y(i_1, i_2+1)}{h_2^2} = -f(i_1, i_2), \quad (3)$$

o bien, en la forma abreviada,

$$y_{\bar{x}_1 x_1}(i_1, i_2) + y_{\bar{x}_2 x_2}(i_1, i_2) = -f(i_1, i_2).$$

En las designaciones sin índices tenemos

$$y_{\bar{x}_1 x_1}(x) + y_{\bar{x}_2 x_2}(x) = -f(x), \quad x = (i_1 h_1, i_2 h_2) \in \omega_h(G). \quad (4)$$

A esta ecuación se le debe agregar las condiciones de contorno

$$y = \mu(x), \quad x = (i_1 h_1, i_2 h_2) \in \gamma_h. \quad (5)$$

La frontera γ_h de la red está constituida por todos los nodos $(0, i_2)$, (N_1, i_2), $(i_1, 0)$, (i_1, N_2), a excepción de los vértices del rectángulo $(0, 0)$, $(0, N_2)$, $(N_1, 0)$, (N_1, N_2) que no se emplean. La ecuación en diferencias (3) está escrita en un molde pentapuntual

$$(i_1 - 1, i_2), (i_1 + 1, i_2), (i_1, i_2), (i_1, i_2 - 1), (i_1, i_2 + 1).$$

El esquema (4) se denomina a menudo esquema *cruz*. Si $h_1 = h_2 = h$, es decir, si las redes según x_1 y x_2 coinciden, la red ω_h se llamará *cuadrada*. En esta red el esquema de diferencias (4) puede ser escrito en la forma

$$y(i_1, i_2) = \\ = \frac{y(i_1-1, i_2)+(y(i_1+1, i_2)+y(i_1, i_2-1)+y(i_1, i_2+1)+h^2 f(i_1, i_2)}{4}.$$

Para la ecuación homogénea ($f = 0$) obtenemos

$$y(i_1, i_2) = \frac{1}{4}[y(i_1-1, i_2) + y(i_1+1, i_2) + \\ + y(i_1, i_2-1) + y(i_1, i_2+1)],$$

es decir, el valor en el centro del molde se determina como media aritmética de los valores en los nodos restantes del molde.

3. Error de aproximación. Sea u = u(x) la solución del problema de Dirichlet (1), (2), y sea $y = y(i_1, i_2)$ la solución del problema en diferencias (4), (5). Veamos un error

$$z(x) = y(x) - u(x), \qquad x = (i_1 h_1, i_2 h_2) \in \omega_h.$$

Al sustituir $y = z + u$ en (4), (5), obtenemos para el error $z = z(x)$ una ecuación no homogénea

$$\Lambda z = z_{\bar{x}_1 x_1} + z_{\bar{x}_2 x_2} = -\psi(x), \quad x \in \omega_h(G), \tag{6}$$

con la condición de contorno homogénea

$$z = 0 \quad \text{cuando} \quad x \in \gamma_h. \tag{7}$$

Aquí

$$\psi(x) = \Lambda u + f(x) = u_{\bar{x}_1 x_1} + u_{\bar{x}_2 x_2} + f(x) \tag{8}$$

es el residuo o error de aproximación para el esquema (4) en la solución $u = u(x)$ de la ecuación (1).

Mostremos que

$$|\psi| \leqslant M_4 \frac{h_1^2 + h_2^2}{24}, \tag{9}$$

donde

$$M_4 = \max_{x \in G} \left(\left| \frac{\partial^4 u}{\partial x_1^4} \right|, \left| \frac{\partial^4 u}{\partial x_2^4} \right| \right).$$

En efecto, tomando en consideración las fórmulas

$$u(x_1 \pm h_1, x_2) = u(x_1, x_2) \pm h_1 \frac{\partial u}{\partial x_1}(y_1, x_2) + \\ + \frac{h_1^2}{2} \frac{\partial^2 u}{\partial x_1^2}(x_1, x_2) \pm \frac{h_1^3}{6} \frac{\partial^3 u}{\partial x_1^3}(x_1, x_2) + \\ + \frac{h_1^4}{24} \frac{\partial^4 u}{\partial x_1^4}(\bar{x}_1, x_2), \quad \bar{x}_1 = x_1 + \theta_1 h_1 \quad 0 \leqslant \theta_1 \leqslant 1,$$

$$u(x_1, x_2 \pm h_2) = u(x_1, x_2) \pm h_2 \frac{\partial u}{\partial x_2}(x_1, x_2) + \\ + \frac{h_2^2}{2} \frac{\partial^2 u}{\partial x_2^2}(x_1, x_2) \pm \frac{h_2^3}{6} \frac{\partial^3 u}{\partial x_2^3}(x_1, x_2) + \\ + \frac{h_2^4}{24} \frac{\partial^4 u}{\partial x_2^4}(x_1, \bar{x}_2), \quad \bar{x}_2 = x_2 + \theta_2 h_2 \quad 0 \leqslant \theta_2 \leqslant 1,$$

encontramos

$$\psi = \left(\frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2} - f(x) \right) + \frac{h_1^2}{24} \frac{\partial^4 u}{\partial x_1^4} (\bar{x}_1, x_2) + \frac{h_2^2}{24} \frac{\partial^4 u}{\partial x_2^4} (x_1, \bar{x}_2).$$

De aquí y de (1) proviene (9).

Así pues, el esquema (4) es de segundo orden de aproximación.

4. **Esquema de orden aumentado de exactitud.** Haciendo uso de un molde nonapuntual (x_1, x_2), $(x_1 \pm h_1, x_2)$, (x_1), $x_2 \pm h_2)$, $(x_1 \pm h_1, x_2 \pm h_2)$, podemos construir un esquema que tenga el cuarto orden de aproximación (y de exactitud), si suponemos que la solución del problema (1) — (2) $u = u(x) \in C^{(6)}(\bar{G})$. Dicho esquema tiene por expresión

$$\begin{gathered} \Lambda' y = \left(\Lambda_1 + \Lambda_2 + \frac{h_1^2 + h_2^2}{12} \Lambda_1 \Lambda_2 \right) y = -\varphi(x), \quad x \in \omega_h, \\ y(x) = \mu(x), \quad x \in \gamma_h, \\ \Lambda_1 y = y_{\bar{x}_1 x_1}, \quad \Lambda_2 y = y_{\bar{x}_2 x_2}, \\ \varphi = f + \frac{h_1^2}{12} \Lambda_1 f + \frac{h_2^2}{12} \Lambda_2 f. \end{gathered} \tag{10}$$

La comprobación inmediata muestra que el residuo es

$$\psi = \Lambda' u + \varphi = O(|h|^4). \tag{11}$$

Para el error $z = y - u$, donde y es la solución del problema (10), obtenemos

$$\Lambda' z = -\psi(x), \quad x \in \omega_h; \quad z = 0, \quad x \in \gamma_h. \tag{12}$$

5. **Propiedades del operador de diferencias.** Sea $\overset{\circ}{y}(x)$ una función reticular definide sobre la red $\bar{\omega}_h = \omega_h (\bar{G})$ e igual a cero en la frontera γ_h de la red, y sea $\overset{\circ}{\Omega}$ un conjunto de funciones reticulares y.

El operador A se definirá del modo siguiente:

$$Ay = -\Lambda \overset{\circ}{y} = -\overset{\circ}{y}_{\bar{x}_1 x_1} - \overset{\circ}{y}_{\bar{x}_2 x_2} \text{ para cualquier } y \in \Omega, \tag{13}$$

donde Ω es un espacio de funciones reticulares que están definidas en los nodos interiores de la red ω_h y coinciden en

dichos nodos con $\overset{\circ}{y}$, $y(x) = \overset{\circ}{y}(x)$ para $x \in \omega_h$. Al designar

$$\varphi = f + \frac{\mu(l_1, x_2)}{h_1^2} \text{ para } x_1 = l_1 - h_1, \quad 0 < x_2 < l_2,$$

$$\varphi = f + \frac{\mu(0, x_2)}{h_1^2}, \quad x_1 = h_1 \quad 0 < x_2 < l_2,$$

$$\varphi = f + \frac{\mu(x_1, l_2)}{h_2^2}, \quad 0 < x_1 < l_1, \quad x_2 = l_2 - h_2,$$

$$\varphi = f + \frac{\mu(x_1, 0)}{h_2^2}, \quad 0 < x_1 < l_1, \quad x_2 = h_2,$$

$\varphi(x) = f(x)$ en los demás puntos $x \in \omega_h$, escribamos el esquema de diferencias (4), (5) en una forma operacional

$$A\varphi = \varphi, \qquad y, \ \varphi \in H, \tag{14}$$

donde $H = \Omega$.

Introduzcamos en H un producto escalar

$$(y, v) = \sum_{i_1=1}^{N_1-1} \sum_{i_2=1}^{N_2-1} \overset{\circ}{y}(i_1, i_2) \overset{\circ}{v}(i_1, i_2) h_1 h_2$$

y probemos que el operador A es autoconjugado. Representemos A en forma de una suma $A = A_1 + A_2$, donde $A_1 y = -\overset{\circ}{y}_{\bar{x}_1 x_1}$, $A_2 y = -\overset{\circ}{y}_{\bar{x}_2 x_2}$, y mostremos que cada uno de los operadores «unidimensionales» A_1 y A_2 es autoconjugado. Será suficiente probarlo para el operador A_1. Veamos un producto escalar

$$(A_1 y, v) =$$

$$= -\sum_{i_2=1}^{N_2-1} h_2 \Big(\sum_{i_1=1}^{N_1-1} \overset{\circ}{y}_{\bar{x}_1 x_1}(i_1, i_2) \overset{\circ}{v}(i_1, i_2) h_1 \Big). \tag{15}$$

Aprovechemos la fórmula unidimensional de Green (cap. I, § 4):

$$\sum_{i_1=1}^{N_1-1} \overset{\circ}{y}_{\bar{x}_1 x_1}(i_1, i_2) \overset{\circ}{v}(i_1, i_2) h_1 =$$

$$= \sum_{i_1=1}^{N_1-1} \overset{\circ}{y}(i_1, i_2) \overset{\circ}{v}_{\bar{x}_1 x_1}(i_1, i_2) h_1.$$

Sustituyendo esta expresión en (15), obtenemos

$$(A_1 y,\ v) = -\sum_{i_2=1}^{N_2-1} h_2 \Big(\sum_{i_1=1}^{N_1-1} \overset{\circ}{y}(i_1,\ i_2)\, \overset{\circ}{v}_{\bar{x}_1 x_1}(i_1,\ i_2)\, h_1 \Big) = (y,\ A_1 v).$$

De un modo análogo nos convencemos de que $A_2^* = A_2$, y, por consiguiente,

$$(Ay,\ v) = ((A_1 + A_2)\, y,\ v) = (A_1 y,\ v) + (A_2 v,\ y) = \\ = (y,\ A_1 v) + (y,\ A_2 v) = (y,\ Av),$$

es decir, $A^* = A$.

Si hacemos uso de la primera fórmula de diferencias de Green

$$\sum_{i_1=1}^{N_1-1} \overset{\circ}{y}_{\bar{x}_1 x_1}(i_1,\ i_2)\, \overset{\circ}{y}(i_1,\ i_2)\, h_1 = -\sum_{i_1=1}^{N_1} (\overset{\circ}{y}_{\bar{x}_1}(i_1,\ i_2))^2 h_1,$$

obtendremos

$$(A_1 y,\ y) = \sum_{i_2=1}^{N_2-1} h_2 \sum_{i_1=1}^{N_1} (\overset{\circ}{y}_{\bar{x}_1}(i_1,\ i_2))^2\, h_1 > 0,$$

y, análogamente, $(A_2 y,\ y) > 0$, de suerte que $A > 0$, es decir, A es un operador definido positivo y autoconjugado.

No es difícil encontrar las fronteras δ y Δ del operador A, es decir, los números, para los cuales se verifican las desigualdades $\delta E \leqslant A \leqslant \Delta E$, donde E es un operador unidad. En efecto, se ha mostrado en el § 4, cap. I que

$$\delta_1 \sum_{i_1=1}^{N_1-1} (\overset{\circ}{y}(i_1,\ i_2))^2\, h_1 \leqslant \sum_{i_1=1}^{N_1} (\overset{\circ}{y}_{\bar{x}_1}(i_1,\ i_2))^2 h_1 \leqslant \\ \leqslant \Delta_1 \sum_{i_1=1}^{N_1-1} (\overset{\circ}{y}(i_1,\ i_2))^2 h_1,$$

donde

$$\delta_1 = \frac{4}{h_1^2} \operatorname{sen}^2 \frac{\pi h_1}{2l_1}, \qquad \Delta_1 = \frac{4}{h_1^2} \cos^2 \frac{\pi h_1}{2l_1}.$$

Al sumar estas desigualdades según $i_2 = 1, 2, \ldots, N_2 - 1$, obtendremos $\delta_1 (y,\ y) \leqslant (A_1 y,\ y) \leqslant \Delta_1 (y,\ y)$.

Del modo análogo encontramos $\delta_2 (y, y) \leqslant (A_2 y, y) \leqslant \Delta_2 (y, y)$, donde

$$\delta_2 = \frac{4}{h_2^2} \operatorname{sen}^2 \frac{\pi h_2}{2l_2}, \quad \Delta_2 = \frac{4}{h_2^2} \cos^2 \frac{\pi h_2}{2l_2}.$$

De aquí se desprende

$$\delta \| y \|^2 \leqslant (Ay, y) \leqslant \Delta \| y \|^2, \tag{16}$$

donde

$$\begin{aligned} \delta &= \delta_1 + \delta_2 = \frac{4}{h_1^2} \operatorname{sen}^2 \frac{\pi h_1}{2l_1} + \frac{4}{h_2^2} \operatorname{sen}^2 \frac{\pi h_2}{2l_2}, \\ \Delta &= \Delta_1 + \Delta_2 = \frac{4}{h_1^2} \cos^2 \frac{\pi h_1}{2l_1} + \frac{4}{h_2^2} \cos^2 \frac{\pi h_2}{2l_2}, \end{aligned} \tag{17}$$

En el cuadrado ($l_1 = l_2 = 1$) en la red cuadrada ($h_1 = h_2 = h$) tenemos

$$\delta = \frac{8}{h^2} \operatorname{sen}^2 \frac{\pi h}{2}, \quad \Delta = \frac{8}{h^2} \cos^2 \frac{\pi h}{2}, \quad \delta + \Delta = \frac{8}{h^2}. \tag{18}$$

6. Problema de diferencias en valores propios. Planteemos un problema: hallar tales valores del parámetro λ (valores propios), para los cuales el problema homogéneo

$$y_{\bar{x}_1 x_1} + y_{\bar{x}_2 x_2} + \lambda y = 0, \quad x \in \omega_h, \quad y = 0, \quad x \in \gamma_h \tag{19}$$

tenga soluciones no triviales (funciones propias). Recurriremos al método de separación de variables y buscaremos la solución del problema (19) en forma de un producto

$$y (x_1, x_2) = v (x_1) w (x_2) \not\equiv 0 \tag{20}$$

de función $v (x_1)$, dependiente sólo de x_1, y de función $w (x_2)$ que depende sólo de x_2. Al sustituir (20) en (19) y al dividir por $y = vw$, obtendremos

$$\frac{v_{\bar{x}_1 x_1}}{v} = -\frac{w_{\bar{x}_2 x_2}}{w} - \lambda, \qquad (x, x_2) \in \omega_h. \tag{21}$$

El primer miembro depende sólo de x_1, mientras que el segundo, sólo de x_2; la igualdad (21) es posible sólo bajo la condición de que

$$\frac{v_{\bar{x}_1 x_1}}{v} = \lambda^{(1)}, \quad -\frac{w_{\bar{x}_2 x_2}}{w} - \lambda = \lambda^{(1)},$$

donde $\lambda^{(1)} = \text{const}$. De aquí resultan dos problemas unidimensionales en valores propios para los segmentos $0 \leqslant x_1 \leqslant l_1$ y $0 \leqslant x_2 \leqslant l_2$, respectivamente:

$$v_{x_1x_1} + \lambda^{(1)} v = 0, \quad 0 < x_1 = i_1 h_1 < l_1,$$

$$v = 0, \quad i_1 = 0, \quad N_1, \tag{22}$$

$$w_{x_2x_2} + \lambda^{(2)} w = 0, \quad 0 < x_2 = i_2 h_2 < l_2,$$

$$w = 0, \quad i_2 = 0, \quad N_2, \tag{23}$$

donde $\lambda^{(2)} = \lambda - \lambda^{(1)}$, o bien $\lambda = \lambda^{(1)} + \lambda^{(2)}$.

Recurriendo al p. 8 del § 4, cap. I, escribamos la solución de los problemas (22), (23) en la forma

$$\lambda_{k_1}^{(1)} = \frac{4}{h_1^2} \operatorname{sen}^2 \frac{\pi k_1 h_1}{2l_1},$$

$$v_{k_1}^{(1)}(x_1) = \sqrt{\frac{2}{l_1}} \operatorname{sen} \frac{\pi k_1 x_1}{l_1}, \quad k_1 = 1, 2, \ldots, N_1 - 1,$$

$$\lambda_{k_2}^{(2)} = \frac{4}{h_2^2} \operatorname{sen}^2 \frac{\pi k_2 h_2}{2l_2},$$

$$w_{k_2}^{(2)}(x_2) = \sqrt{\frac{2}{l_2}} \operatorname{sen} \frac{\pi k_2 x_2}{l_2}, \quad k_2 = 1, 2, \ldots, N_2 - 1,$$

donde $x_\alpha = i_\alpha h_\alpha$, $i_\alpha = 0, 1, \ldots, N_\alpha$, $\alpha = 1, 2$.

De aquí se deduce que el problema (19) tiene valores propios

$$\lambda_{k_1k_2} = \frac{4}{h_1^2} \operatorname{sen}^2 \frac{\pi k_1 h_1}{2l_1} + \frac{4}{h_2^2} \operatorname{sen}^2 \frac{\pi k_2 h_2}{2l_2},$$

$$k_\alpha = 1, 2, \ldots, N_\alpha - 1, \ \alpha = 1, 2, \tag{24}$$

y funciones propias correspondientes $y_k = v_{k_1}^{(1)}(x_1) w_{k_2}^{(2)}(x_2)$:

$$y_k = y_{k_1k_2}(x_1, x_2) = \sqrt{\frac{4}{l_1 l_2}} \operatorname{sen} \frac{\pi k_1 x_1}{l_1} \operatorname{sen} \frac{\pi k_2 x_2}{l_2},$$

$$x_\alpha = i_\alpha h_\alpha, \quad i_\alpha = 0, 1, \ldots, N_\alpha, \quad k_\alpha = 1, 2, \ldots, N_\alpha - 1, \quad \alpha = 1, 2. \tag{25}$$

Estas funciones propias están ortonormalizadas:

$$(y_{k_1k_2}, y_{m_1m_2}) = \delta_{k_1m_1} \delta_{k_2m_2}.$$

De (17) y (25) se ve que

$$\delta = \min \lambda_{k_1k_2} = \lambda_{1,1} \qquad \Delta = \max \lambda_{k_1k_2} = \lambda_{N_1-1, N_2-1},$$

donde δ y Δ se determinan según las fórmulas (17). Para δ y Δ son justas las estimaciones

$$\delta \geqslant 8\left(\frac{1}{l_1^2}+\frac{1}{l_2^2}\right), \quad \Delta < \frac{4}{h_1^2}+\frac{4}{h_2^2}. \tag{26}$$

7. Estimación de la velocidad de convergencia del esquema «cruz». Principio del máximo. Para el error $z = y - u$ del esquema en el p. 3 se ha obtenido el problema (6), (7), donde

$$\psi(x) = O(|h|^2), \qquad |h|^2 = h_1^2 + h_2^2 \tag{27}$$

bajo el supuesto de que la solución $u = u(x) \in C^{(4)}(\overline{G})$ del problema de partida (1), (2) sea suficientemente suave. Demostremos que el esquema (4) converge con la velocidad $O(|h|^2)$ (es de segundo orden de exactitud) en la norma reticular C, es decir, que $\|z\|_C = O(|h|^2)$, donde $\|z\|_C = \max\limits_{x \in \omega_h} |z(x)|$. Para esto nos hará falta la estimación de la solución del problema (6), (7) a través del segundo miembro de ψ. El problema de contorno de Dirichlet es un caso particular del problema

$$\mathcal{L}[y] = a_{i_1 i_2} y_{i_1 i_2} - b_{i_1-1,\, i_2} y_{i_1-1,\, i_2} - b_{i_1+1,\, i_2} y_{i_1+1,\, i_2} - \\ - b_{i_1,\, i_2-1} y_{i_1,\, i_2-1} - b_{i_1,\, i_2+1} y_{i_1,\, i_2+1} = \varphi_{i_1 i_2}, \\ x = (i_1 h_1,\ i_2 h_2) \in \omega_h; \quad y = \mu, \quad x \in \gamma_h, \tag{28}$$

donde $a = a_{i_1 i_2}$, $b = b_{i_1,\, i_2}$ son los coeficientes.
En el caso (4) tenemos

$$a_{i_1 i_2} = 2\left(\frac{1}{h_1^2}+\frac{1}{h_2^2}\right), \\ b_{i_1 \pm 1,\, i_2} = \frac{1}{h_1^2}\, b_{i_1,\, i_2 \pm 1} = \frac{1}{h_2^2}, \tag{29} \\ b_{i_1,\, i_2} = 0.$$

El operador $\mathcal{L}[y]$ puede anotarse de otra forma:

$$\mathcal{L}[y] = d_{i_1 i_2} y_{i_1 i_2} + b_{i_1-1,\, i_2}(y_{i_1,\, i_2} - y_{i_1-1,\, i_2}) + \\ + b_{i_1+1,\, i_2}(y_{i_1,\, i_2} - y_{i_1+1,\, i_2}) + b_{i_1,\, i_2-1}(y_{i_1,\, i_2} - y_{i_1,\, i_2-1}) + \\ + b_{i_1 i_2+1}(y_{i_1 i_2} - y_{i_1,\, i_2+1}), \tag{30}$$

donde $d_{i_1 i_2} = a_{i_1 i_2} - b_{i_1-1,\, i_2} - b_{i_1+1,\, i_2} - b_{i_1,\, i_2-1} - b_{i_1,\, i_2+1}$.

Supondremos cumplidas las condiciones

$$d = d_{i_1 i_2} \geqslant 0, \quad b_{i_1 \pm 1, i_2} > 0, \quad b_{i_1, i_2 \pm 1} \geqslant 0. \tag{31}$$

Para el problema (4) tenemos $d \equiv 0$.

TEOREMA 1. *Supongamos cumplidas las condiciones* (31) *y que* $\varphi(x) \geqslant 0$, $y|_\gamma \geqslant 0$. *Entonces, la solución de la ecuación* (28) *es no negativa, es decir,* $y(x) \geqslant 0$ *en todos los nodos de la red* $\bar{\omega}_h = \omega_h \cup \gamma_h$.

DEMOSTRACIÓN. Supongamos que la afirmación del teorema es falsa y existe por lo menos un nodo $x_{i_*} = (i_1^0 h_1, i_2^0 h_2)$ en el cual $y(x_{i_*}) < 0$. Entonces la función $y(x)$ ha de tomar en cierto nodo interior de la red un valor negativo mínimo $\min\limits_{x \in \omega_h} y(x) = y(x_*)$. En este nodo se verifica la ecuación (28). Si $d(x_*) = 0$ y $\varphi(x_*) = 0$, la ecuación (28) se verificará sólo bajo la condición de que $y(x) = y(x_*)$ en todos los nodos del molde. Sin embargo, por cuanto $\varphi(x) \not\equiv 0$, existe un nodo $x_{i_{**}}$ en el que $y(x_{i_{**}}) = y(x_{i_*}) = \min y(x) = c_0 < 0$, y, al menos en un nodo, por ejemplo, para $x = x_{i_1+1}$ tenemos $y_{i_1+1} > c_0$, y, por consiguiente, $\mathcal{L}[y]|_{x=x_{i_{**}}} < 0$, que contradice la condición $\mathcal{L}[y] = \varphi(x) \geqslant 0$. La contradicción obtenida demuestra el teorema.

TEOREMA 2 (TEOREMA DE COMPARACIÓN). *Sea* $\bar{y}(x)$ *la solución del problema*

$$\mathcal{L}[\bar{y}] = \bar{\varphi}, \quad x \in \omega_h, \quad \bar{y} = \bar{\mu}, \quad x \in \gamma_h \tag{32}$$

y supongamos cumplidas las condiciones (31). *Si*

$$|\varphi(x)| \leqslant \bar{\varphi}(x), \quad x \in \omega_h, \quad |\mu(x)| \leqslant \bar{\mu}(x), \quad x \in \gamma_h, \tag{33}$$

entonces, para la ecuación del problema (28) *es justa la estimación*

$$|y(x)| \leqslant \bar{y}(x) \text{ para todos los } x \in \bar{\omega}_h.$$

Basta por convencerse de que para las funciones $u = \bar{y}(x) + y(x)$, $v = \bar{y}(x) - y(x)$ quedan cumplidas las condiciones del teorema 1, y, por lo tanto, $u(x) \geqslant 0$, $v(x) \geqslant 0$, o bien $y(x) \geqslant -\bar{y}(x)$, $y(x) \leqslant \bar{y}(x)$, es decir, $|y| \leqslant \bar{y}$.

Así pues, la función $\overline{y}(x)$ es una *mayorante*. Si la mayorante $\overline{y}(x)$ queda hallada, la solución del problema (28) viene estimada de acuerdo con el teorema 2. Para el problema (4) elijamos, a título de mayorante, una función

$$\overline{y}(x) = C[L^2 - (x_1^2 + x_2^2)], \qquad L^2 = l_1^2 + l_2^2. \tag{34}$$

Calculemos al principio $\overline{\varphi} = \mathscr{L}[\overline{y}] = -\Lambda\overline{y} = C\Lambda(x_1^2 + x_2^2) = C(\Lambda_1 x_1^2 + \Lambda_2 x_2^2) = 4C$, puesto que $(x_1^2)_{\overline{x}_1 x_1} = \frac{1}{h_1^2}((x_1 + h_1)^2 - 2x_1^2 + (x_1 - h_1)^2) = 2$. De la fórmula (34) se ve que $\overline{\mu} = y(x) > 0$ en la frontera γ_h. Volvamos ahora al problema (6), (7) para el error $z = y - u$ del esquema (4). Eligiendo $4C = |\psi|_C$, y teniendo presente que $z|_{\gamma_h} = 0$, obtenemos $|z(x)| < \overline{y}(x) < CL^2$, de suerte que

$$\|z\|_C \leqslant \frac{L^2}{4}\|\psi\|_C. \tag{35}$$

De aquí y de (9) se deduce la convergencia uniforme del esquema (4) con el segundo orden de exactitud.

OBSERVACION. La ecuación (28) puede ser sustituida por una ecuación de la forma más general

$$\mathscr{L}[y] = a(x)y(x) - \sum_{\substack{\xi \in \sigma(\kappa) \\ \xi \neq x}} b(x, \xi)y(\xi) = \varphi(x), \tag{36}$$

donde $a(x) > 0$, $b(x, \xi) > 0$, $\sigma(x)$ es un conjunto de nodos $\xi \neq x$ del molde con centro en el nodo x, con la particularidad de que

$$d(x) = a(x) = \sum_{\xi \in \sigma(x)} b(x, \xi) \geqslant 0.$$

Para la ecuación (36) son verídicos los teoremas 1 y 2. Cuando se trata de un esquema con el orden aumentado de exactitud, el molde consta de nueve nodos, el conjunto $\sigma(x)$ de ocho nodos, y en este caso $a = \frac{5}{3}(h_1^{-2} + h_2^{-2})$, mientras que en el segundo miembro se tienen coeficientes $\frac{1}{6}(5h_1^{-2} - h_2^{-2})$,

$\frac{1}{6}(5h^{-2} - h^{-2})$, los cuales son positivos sólo a condición de que

$$1/\sqrt{5} \leqslant h_1/h_2 \leqslant \sqrt{5},$$

y, por consiguiente, la estimación (35) se obtendrá bajo dicha condición.

§ 2. Resolución de las ecuaciones en diferencias

1. Métodos directos. Método de separación de variables. El sistema de ecuaciones en diferencias para el problema de Dirichlet del § 1,

$$\Lambda y = y_{\bar{x}_1 x_1} + y_{\bar{x}_2 x_2} = -f(x), \quad x \in \omega_h \quad y = \mu, \quad x \in \gamma_h \quad (1)$$

cuenta con una matriz de alto orden $(N_1 - 1)(N_2 - 1)$. Se toman habitualmente N_1, $N_2 \sim 50 - 100$, de modo que el número de ecuaciones en el sistema (1) es igual a 10^3—10^4. La resolución de un sistema de orden tan alto, por el método de Gauss, exigiría aproximadamente $(N_1 - 1)^3 (N_2 - 1)^3$, esto es, 10^9—10^{12} operaciones, si el sistema (1) no tuviera una calidad muy buena: la matriz del sistema está débilmente llenada y sólo tiene $\sim 5N_1N_2$ elementos distintos de cero. Por esta razón, para la resolución de un sistema de ecuaciones en diferencias se logra construir los métodos que requieren $O(N \ln N)$ e incluso $O(N)$ operaciones, donde $N = (N_1 - 1)(N_2 - 1)$. Describamos uno de los métodos directos de resolución del problema en diferencias de Dirichlet de la ecuación de Poisson en un rectángulo.

Escribamos el problema (1) en una forma

$$\Lambda \overset{*}{y} = \overset{*}{y}_{\bar{x}_1 x_1} + \overset{*}{y}_{\bar{x}_2 x_2} = -\varphi(x), \quad x \in \omega_h, \quad \overset{*}{y}|_{\gamma_h} = 0, \quad (2)$$

donde $\overset{*}{y}(x) \equiv y(x)$ para $x \in \omega_h$, y $\varphi(x)$ se determina según las fórmulas (14) del § 1.

Su solución puede encontrarse por el método de separación de variables. Sean $\{v_{k_2}^{(2)}(x_2), \lambda_{k_2}^{(2)}\}$ $(k = 1, 2, \ldots, \ldots, N_2 - 1)$ funciones propias y valores propios del problema

$$\Lambda_2 v + \lambda v = 0, \quad x \in \omega_h; \quad v(0) = v(l_2) = 0. \quad (3)$$

Las expresiones para $\lambda_{k_2}^{(2)}$ y $v_{k_2}^{(2)}(x_2)$ se han aducido en el p. 6, § 1.

Desarrollemos la solución $\overset{\circ}{y}(x_1, x_2)$ y el segundo miembro $\varphi(x_1, x_2)$ según las funciones propias $\{v_{k_2}^{(2)}\}$:

$$\overset{\circ}{y}(x_1, x_2) = \sum_{k_2=1}^{N_2-1} c_{k_2}(x_1)\, v_{k_2}(x_2), \tag{4}$$

$$\varphi(x_1, x_2) = \sum_{k_2=1}^{N_2-1} \varphi_{k_2}(x_1)\, v_{k_2}(x_2), \tag{5}$$

donde $x_\alpha = i_\alpha h_\alpha$, $i_\alpha = 1, 2, \ldots, N_\alpha - 1$, $\alpha = 1, 2$, $c_{k_2}(x_1)$ y $\varphi_{k_2}(x_1)$ son los coeficientes de Fourier, por ejemplo,

$$\varphi_{k_2}(x_1) = \sum_{i_2=1}^{N_2-1} h_2 \varphi(x_1, i_2 h_2)\, v_{k_2}(i_2 h_2).$$

Apliquemos un operador $\Lambda = \Lambda_1 + \Lambda_2$ al producto $c_{k_2} v_{k_2}$:

$$\begin{aligned}
&\Lambda c_{k_2}(x_1)\, v_{k_2}(x_2) = \\
&\quad = v_{k_2}(x_2)\, \Lambda_1 c_{k_2}(x_1) + c_{k_2}(x_1)\, \Lambda_2 v_{k_2}(x_2) = \\
&\quad = v_{k_2}(x_2)\, \Lambda_1 c_{k_2}(x_1) - \lambda_{k_2}^{(2)} c_{k_2}(x_1)\, v_{k_2}(x_2) = \\
&\qquad\qquad = [\Lambda_1 c_{k_2}(x_1) - \lambda_{k_2}^{(2)} c_{k_2}(x_1)]\, v_{k_2}(x_2).
\end{aligned}$$

Ahora, al sustituir esta expresión en (2) y al tomar en consideración (5), obtendremos

$$\sum_{k_1=1}^{N_1-1} \{\Lambda_1 c_{k_2}(x_1) - \lambda_{k_2}^{(2)} c_{k_2}(x_1) + \varphi_{k_2}(x_1)\}\, v_{k_2}(x_2) = 0. \tag{6}$$

En virtud de que la función $\{v_{k_2}(x_2)\}$ es ortogonal, esta identidad se verifica sólo cuando es igual a cero la expresión encerrada dentro de las llaves

$$\begin{aligned}
&\Lambda_1 c_{k_2}(x_1) - \lambda_{k_2}^{(2)} c_{k_2}(x_1) = -\varphi_{k_2}(x_1), \\
&\qquad\qquad k_2 = 1, 2, \ldots, N_2 - 1, \\
&= i_1 h_1,\ 0 < i_1 < N_1,\ c_{k_2}(i_1 h_1) = 0,\ i_1 = 0,\ N_1. \qquad (7)
\end{aligned}$$

Efectivamente, multiplicando (6) escalarmente por $v_{k_2}(x_2)$, tenemos

$$0 = \sum_{k=1}^{N_1-1} \{\cdot\}_k (v_k, v_{k_2}) = \sum_{k=1}^{N_1-1} \{\cdot\}_k \delta_{kk_2} = \{\cdot\}_{k_2} = 0,$$

donde $\{\cdot\}_{k_2}$ es el contenido de las llaves (6).

Los problemas (7) se resuelven por el método de factorización; se necesita emplear $N_2 - 1$ veces en total el alogritmo de factorización para $k_2 = 1, 2, \ldots, N_2 - 1$. Conociendo $c_{k_2}(x_1)$, hallaremos la solución del problema (2) según la fórmula (4). Con este fin se requiere, primero, calcular los coeficientes de Fourier $\varphi_{k_2}(x_1)$ ($k_2 = 1, 2, \ldots, N_2 - 1$). De las fórmulas (4) y (5) se ve que $y(x_1, x_2)$ y $\varphi_{k_2}(x_1)$ se calculan según las fórmulas de una misma forma:

$$w_l = \sum_{k=1}^{N-1} \alpha_k \operatorname{sen} \frac{k\pi l}{N}, \quad l = 1, 2, \ldots, N-1. \qquad (8)$$

Se ha elaborado un algoritmo especial de transformación rápida de Fourier para calcular sumas, el cual permite obtener la suma (8) en el transcurso de $5N \log_2 N$ operaciones aritméticas (cuando $N = 2^n$, siendo n un número entero) en lugar de $O(N^2)$ si se usa el modo de sumar habitual. Este algoritmo permite hallar la solución del problema de partida (2) en el transcurso de $O(N_1 N_2 \log_2 N_2)$ operaciones. El método de separación de variables puede combinarse con el de reducción o descomposición que representa una modificación del método de Gauss. De resultas obtenemos un algoritmo con el número de operaciones $Q \approx 5N_1 N_2 \log_2 N_2$ que es dos veces menor que el número correspondiente para el algoritmo de separación aducido más arriba.

2. Métodos iterativos. Los métodos directos son más económicos cuando se resuelve el problema de Dirichlet en diferencias para la ecuación de Poisson en un rectángulo. Actualmente existen programas estándar en el lenguaje algorítmico FORTRAN y ALGOL para resolver las ecuaciones de Poisson en un rectángulo con las condiciones de contorno de tres tipos y también con las condiciones de contorno mixtas. No obstante, cuando el dominio no es un rectángulo o cuando se analizan las ecuaciones de coeficientes variables,

se aplican los métodos iterativos. En realidad los métodos directos son económicos sólo en el caso en que las variables se separen.

En el cap. III se ha estudiado la teoría de los métodos iterativos para una ecuación.

$$Ay = \varphi,$$

donde $A = A^* > 0$. La comparación de diferentes métodos se realizaba para el problema unidimensional modelo en el segmento $0 \leqslant x \leqslant 1$:

$$y_{\bar{x}x} = -f(x), \quad x = ih, \quad 0 < i < N, \quad y_0 = y_N = 0.$$

Para el problema citado el operador A tiene la forma $Ay = \overset{\circ}{=} -y_{\bar{x}x}$. Las fronteras del operador se determinan por las constantes

$$\delta = \frac{4}{h^2}\operatorname{sen}^2\frac{\pi h}{2}, \qquad \Delta = \frac{4}{h^2}\cos^2\frac{\pi h}{2}.$$

El número de iteraciones para los métodos, estudiados en el cap. III depende de la razón

$$\eta = \frac{\delta}{\Delta} = \operatorname{tg}^2\frac{\pi h}{2} \approx \frac{\pi^2 h^2}{4}. \tag{9}$$

Veamos ahora, a título del problema modelo, un problema de Dirichlet bidimensional en un cuadrado unitario ($l_1 = l_2 = 1$) sobre la red cuadrada de paso $h = h_1 = h_2$:

$$Ay = -\overset{\circ}{y}_{\bar{x}_1 x_1} - \overset{\circ}{y}_{\bar{x}_2 x_2} = \varphi, \quad \varphi,\ y \in H. \tag{10}$$

El número de intervalos por cada una de las direcciones es N, de modo que $h = 1/N$.

Las fronteras δ y Δ del operador A se han determinado en el § 1 (véase (18) del § 1), la razón $\eta = \delta/\Delta$ coincide con (9). De aquí proviene que el número de iteraciones no depende del número de mediciones (si $h_1 \neq h_2$, $l_1 \neq l_2$, depende poco). Por esta razón, las estimaciones del número de iteraciones de diferentes métodos iterativos, obtenidas para el problema modelo unidimensional, siguen siendo en rigor para el caso bidimensional.

En el caso de una red no cuadrada, el número de iteraciones para el problema bidimensional puede diferir un poco

del número de iteraciones para el problema unidimensional.

Aquí se examinará sólo el método iterativo alternado triangular para resolver el problema de Dirichlet en diferencias (10).

3. Método alternado triangular. Para la resolución de una ecuación operacional

$$Au = f, \quad A = A^* > 0, \quad A: H \to H, \tag{11}$$

hemos considerado en el cap. III los métodos iterativos de un paso (de dos capas), los cuales se anotaban en la siguiente forma canónica:

$$B\frac{y_{k+1}-y_k}{\tau_{k+1}} + Ay_k = f, \quad k = 0, 1, \ldots, n,$$

para todo $y_0 \in H$, (12)

donde B: $H \to H$, $B = B^* > 0$. Para A y B se cumplen las condiciones

$$\gamma_1 B \leqslant A \leqslant \gamma_2 B, \qquad \gamma_1 > 0, \tag{13}$$

donde γ_1 y γ_2 son unas constantes.

El número mínimo de iteraciones $\min_{\tau_k} n(\varepsilon)$ con γ_1 y γ_2 prefijadas se consigue al elegir los parámetros de Chébishev

$$\tau_k = \frac{\tau_0}{1+\rho_0\sigma_k}, \quad \tau_0 = \frac{2}{\gamma_1+\gamma_2} \quad \rho_0 = \frac{1-\xi}{1+\xi},$$

$$\xi = \frac{\gamma_1}{\gamma_2}, \qquad k = 1, 2, \ldots, n, \tag{14}$$

donde σ_k pertenece a cierto conjunto, especialmente ordenado, de ceros del polinomio de Chébishev; con tal ordenación el método (12) es estable desde el punto de vista de los cálculos.

Para determinar la $(k+1)$-ésima iteración tenemos una ecuación

$$By_{k+1} = F_k, \qquad F_k = By_k - \tau_{k+1}(Ay_k - f).$$

El número de operaciones al calcular y_{k+1} depende de B. Al elegir

$$B = (D + \omega A_1) D^{-1} (D + \omega A_2), \tag{15}$$

donde A_1 y A_2 son los operadores con matrices triangulares $A_1^* = A_2$, $A_1 + A_2 = A$ y $D = D^* > 0$ es un operador

arbitrario, obtenemos el **método alternado triangular.** Corrientemente, $D = (d_{i_1 i_1})$ es una matriz diagonal. En el cap. III fue elaborada la teoría de este método y se han encontrado las constantes γ_1, γ_2 y ω para las condiciones prefijadas

$$A \geqslant \delta D, \quad A_1 D^{-1} A_2 \leqslant \frac{\Delta}{4} A, \quad \delta > 0, \quad \Delta \geqslant \delta > 0, \qquad (16)$$

las cuales pueden ser escritas en la forma equivalente:

$$(Ay, y) \geqslant \delta (Dy, y), \quad (D^{-1} A_2 y, A_2 y) \leqslant \frac{\Delta}{4} (Ay, y).$$

En este caso tenemos

$$\omega = \frac{2}{\sqrt{\delta \Delta}}, \qquad \xi = \frac{2\sqrt{\eta}}{1+\sqrt{\eta}}, \qquad \eta = \frac{\delta}{\Delta}, \qquad (17)$$

y para el número de iteraciones es justa la estimación

$$n(\varepsilon) \approx n_0(\varepsilon) = \frac{1}{2\sqrt{2}\sqrt[4]{\eta}} \ln \frac{2}{\varepsilon}. \qquad (18)$$

4. Método alternado triangular para el problema de Dirichlet en diferencias. Volvamos al problema (10). Representemos el operador A en forma de una suma $A = A_1 + A_2$, donde

$$A_1 y = \frac{y_{\bar{x}_1}}{h_1} + \frac{y_{\bar{x}_2}}{h_2}, \qquad A_2 y = -\frac{y_{x_1}}{h_1} - \frac{y_{x_2}}{h_2},$$

y pongamos $D = E$. El carácter conjugado de A_1 y A_2: $A_2 = A_1^*$ se establece por comparación de sus matrices o bien con ayuda de la primera fórmula de Green en diferencias: $(A_1 y, v) = (y, A_1^* v) = (y, A_2 v)$.

Para determinar y_{k+1} obtenemos una ecuación

$$B y_{k+1} = (E + \omega A_1)(E + \omega A_2) y_{k+1} = F_k,$$

$$F_k = B\overset{\circ}{y}_k + \tau_{k+1}(\Lambda y_k + \varphi) \quad (y_k = \mu, \ \overset{\circ}{y}_k = 0 \text{ para } x \in \gamma_h).$$

Los valores de y_{k+1} se hallan sucesivamente de la ecuación

$$(E + \omega A_1)\overset{\circ}{y}{}^{(1)}_k = F_k, \qquad (E + \omega A_2)\overset{\circ}{y}_{k+1} = \overset{\circ}{y}{}^{(1)}_k.$$

De aquí obtenemos las fórmulas

$$\overset{\cdot}{y}{}_k^{(1)}(i_1, i_2) = \left[\frac{\varkappa_1 \overset{\cdot}{y}{}_k^{(1)}(i_1-1, i_2) + \varkappa_2 \overset{\cdot}{y}{}_k^{(1)}(i_1, i_2-1) + F_k(i_1, i_2)}{(1+\varkappa_1+\varkappa_2)}\right],$$

$$\varkappa_1 = \frac{\omega}{h_1^2}, \qquad \varkappa_2 = \frac{\omega}{h_2^2},$$

$$\overset{\cdot}{y}{}_{k+1}(i_1, i_2) = \left[\frac{\varkappa_1 \overset{\cdot}{y}{}_{k+1}(i_1+1, i_2) + \varkappa_2 \overset{\cdot}{y}{}_{k+1}(i_1, i_2+1) + \overset{\cdot}{y}{}_k^{(1)}(i_1, i_2)}{(1+\varkappa_1+\varkappa_2)}\right]. \tag{19}$$

Para determinar $\overset{\cdot}{y}{}_k^{(1)}(i_1, i_2)$ elijamos un nodo $i_1 = 1$, $i_2 = 1$ en la esquina izquierda del rectángulo; entonces, los dos nodos restantes $(i_1 - 1, i_2)$ y $(i_1, i_2 - 1)$ del molde $\{(i_1, i_2), (i_1 - 1, i_2), (i_1, i_2-1)\}$ se disponen en la frontera y, por lo tanto, $\overset{\cdot}{y}{}^{(1)}(i_1 - 1, i_2) = \overset{\cdot}{y}{}^{(1)}(i_1, i_2 - 1) = 0$ son conocidos. Sabiendo $\overset{\cdot}{y}{}_k^{(1)}$ para $i_1 = 1$, $i_2 = 1$, hallamos sucesivamente $\overset{\cdot}{y}{}_k^{(1)}$ para $i_1 = 2, 3, \ldots, N_1 - 1$ y $i_2 = 1$ (en la primera fila). Luego suponemos $i_2 = 2$ y encontramos sucesivamente $\overset{\cdot}{y}{}_k^{(1)}$ en la segunda fila para $i_1 = 1, 2, \ldots$ $\ldots, N - 1$. Con el fin de hallar $\overset{\cdot}{y}{}_{k+1}$ realizamos los cálculos en el molde $\{(i_1, i_2), (i_1 + 1, i_2), (i_1, i_2 + 1)\}$ según las columnas de arriba abajo: fijamos $i_1 = N_1 - 1$, $N_1 - 2, \ldots, 2, 1$, y para cada i_1 cambiamos $i_2 = N_2 - 1$, $N_2 - 2, \ldots, 2, 1$. Empezamos la cuenta de $\overset{\cdot}{y}{}_{k+1}$ con el nodo $(i_1 = N_1 - 1, i_2 = N_2 - 1)$ en la esquina derecha superior. Se debe observar que la cuenta de $\overset{\cdot}{y}{}_{k+1}$ puede realizarse también según las filas de derecha a izquierda: fijamos $i_2 = N_2 - 1, N_2 - 2, \ldots, 2, 1$ y para cada i_2 cambiamos $i_1 = N_1 - 1, N_1 - 2, \ldots, 2, 1$. Es más, el cálculo de $\overset{\cdot}{y}{}_k^{(1)}$ podemos llevarlo a cabo no por las filas, sino por las columnas de abajo arriba. Esto lo muestran las mismas fórmulas.

Los cálculos se realizan por las fórmulas recurrentes (19); la cuenta es, evidentemente, estable. El algoritmo del tipo semejante se llama (como ya se ha indicado) *algoritmo de cómputo móvil*.

Calculemos el número de operaciones aritméticas que corresponden a un nodo de la red: el cálculo de F_k requiere

10 operaciones de adición y 10 operaciones de multiplicación; el cálculo de y_{k+1} con F_k prefijada exige 4 operaciones de adición y 6 operaciones de multiplicación.

En total se exigen 14 operaciones de adición y 16 operaciones de multiplicación para determinar y_{k+1} en un solo nodo. El número de operaciones puede ser disminuido conservando en la memoria de acceso rápido no una, sino dos sucesiones, $\{y_k\}$ y $\{w_{k+1}\}$, y empleando para la determinación de y_{k+1} un algoritmo

$$(E + \omega A_1)\,\mathring{w}_{k+1/2} = \Lambda y_k + f,\ (E + \omega A_2)\mathring{w}_{k+1} = \mathring{w}_{k+1/2},$$
$$y_{k+1} = y_k + \tau_{k+1}\mathring{w}_{k+1}.$$

En este caso para pasar de y_k a y_{k+1} son suficientes 10 operaciones de adición y 10 operaciones de multiplicación por un nodo.

5. Elección de los parámetros del método alternado triangular para el problema de Dirichlet en diferencias. Para poder aprovechar la teoría general expuesta en el cap. III (véase el § 5, cap. III), es necesario hallar las constantes δ y Δ que intervienen en la condición (16). En nuestro caso $A = A_1 + A_2 \geqslant \delta E$, donde δ es el valor propio mínimo del operador A igual a

$$\delta = 4\left(\frac{1}{h_1^2}\operatorname{sen}^2\frac{\pi h_1}{2l_1} + \frac{1}{h_2^2}\operatorname{sen}^2\frac{\pi h_2}{2l_2}\right). \qquad (20)$$

Examinemos el operador $A_1D^{-1}A_2 = A_1A_2$. Teniendo presente que

$$A_1^* = A_2,\ (a_1b_1 + a_2b_2)^2 \leqslant (a_1^2 + a_2^2)\ (b_1^2 + b_2^2),$$

encontramos

$$(A_1A_2y,\ y) = (A_2y,\ A_2y) =$$
$$= \left(\left(\frac{1}{h_2}y_{x_1} + \frac{1}{h_2}y_{x_2}\right)^2,\ 1\right) \leqslant \left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right)((y_{x_1})^2 + (y_{x_2})^2,\ 1)) =$$
$$= \left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right)\sum_{i_1=1}^{N_1-1}\sum_{i_2=1}^{N_2-1}[(y_{x_1})^2 + (y_{x_2})^2]_{i_1i_2}h_1h_2 \leqslant$$
$$\leqslant \left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right)(Ay,\ y),$$

puesto que (véase el § 1, cap. V)

$$(Ay, y) = \sum_{i_1=1}^{N_1-1} h_2 \sum_{i_2=0}^{N_2-1} (y_{x_1})^2_{i_1 i_2} h_1 + \sum_{i_1=1}^{N_1-1} h_1 \sum_{i_2=0}^{N_2-1} (y_{x_2})^2_{i_1 i_2} h_2.$$

Al comparar las desigualdades

$$(A_1 A_2 y, y) \leqslant \left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right)(Ay, y) \quad \text{y} \quad A_1 A_2 \leqslant \frac{\Delta}{4} A,$$

concluimos que

$$\Delta = 4\left(\frac{1}{h_1^2} + \frac{1}{h_2^2}\right). \tag{21}$$

Conociendo δ y Δ, encontramos $\eta = \delta/\Delta$, y, según las fórmulas del § 5 del cap. V, determinamos los parámetros γ_1, γ_2, ξ, después de lo cual estimamos el número de iteraciones por la fórmula

$$n(\varepsilon) \approx \ln \frac{\varepsilon}{2} \Big/ \ln \frac{1}{\rho_1}, \quad \rho_1 = \frac{1-\sqrt{\xi}}{1+\sqrt{\xi}}.$$

Haciendo uso de $n(\varepsilon)$, eligimos una totalidad estable de los parámetros de Chébishev σ_k, τ_{k+1} y $\omega = 2/\sqrt{\delta\Delta}$.

Comparamos los siguientes métodos de resolución referente al número de iteraciones $n_0(\varepsilon)$: el método de iteración simple ($n_0^{(1)}(\varepsilon)$), el esquema explícito con una totalidad de Chébishev ($n_0^{(2)}(\varepsilon)$) y el método alternado triangular ($n_0^{(3)}(\varepsilon)$) para el problema bidimensional modelo (10), empleando las fórmulas aproximadas $n_0^{(1)}(\varepsilon) \approx 2/h^2$, $n_0^{(2)}(\varepsilon) \approx$ $\approx 3{,}2/h$, $n_0^{(3)}(\varepsilon) \approx 2{,}9\sqrt{h}$ para $\varepsilon = 10^{-4}$ (tabla 2).

TABLA 2

h	$n_0^{(1)}(\varepsilon)$	$n_0^{(2)}(\varepsilon)$	$n_0^{(3)}(\varepsilon)$
1/10	200	32	9
1/50	5 000	160	21
1/100	20 000	320	29

6. Ecuaciones en diferencias con coeficientes variables. Supongamos que se pide resolver en un rectángulo $\bar{G} = \{(x_1, x_2): 0 \leqslant x_\alpha \leqslant l_\alpha, \alpha = 1, 2\}$ un problema de Dirichlet

para la ecuación elíptica de coeficientes variables:

$$Lu = L_1 u + L_2 u = -f(x),$$
$$x = (x_1, x_2) \in G, \quad u = \mu(x), \quad x \in \Gamma, \quad (22)$$
$$L_\alpha u = \frac{\partial}{\partial x_\alpha}\left(k_\alpha(x)\frac{\partial u}{\partial x_\alpha}\right), \quad 0 < c_1 \leqslant k_\alpha(x) \leqslant c_2, \quad \alpha = 1, 2,$$

donde c_1 y c_2 son unas constantes. Para $k_1 \equiv k_2 \equiv 1$ obtenemos la ecuación de Poisson $\Delta u = -f$.

El esquema de diferencias se construye sobre una red $\omega_h = \{x_i = (i_1 h_1, i_2 h_2) \mid i_\alpha = 0, 1, \ldots, N_\alpha, h_\alpha = l_\alpha / N_\alpha, \alpha = 1, 2\}$. Todo operador L_α se sustituye en el molde tripuntual $(x_\alpha - h_\alpha, x_\alpha + h_\alpha)$ por un operador de diferencias

$$\Lambda_\alpha u = (a_\alpha u_{\bar{x}_\alpha})_{x_\alpha} = \frac{1}{h_\alpha}\left[\frac{a_\alpha^{(+1_\alpha)}(u_\alpha^{(+1_\alpha)} - u)}{h_\alpha} - \frac{a_\alpha(u - u^{(-1_\alpha)})}{h_\alpha}\right],$$

donde $u^{(\pm 1_s)} = u((i_1 \pm 1)h_1, i_2 h_2)$, $u^{(\pm 1_s)} = u(i_1 h_1, (i_2 \pm \pm 1)h_2)$. Para a_1 y a_2 pueden elegirse las expresiones más simples

$$a_1(x_1, x_2) = k_1(x_1 - 1/2h_1, x_2) = k^{(-1/2_1)},$$
$$a_2(x_1, x_2) = k_2(x_1, x_2 - 1/2h_2) = k^{(-1/2_2)},$$

que aseguran el segundo orden de aproximación:

$$\Lambda_\alpha u - L_\alpha u = O(h_\alpha^2).$$

De resultas, al operador Lu se le pone en correspondencia un operador de diferencias sobre el molde pentapuntual:

$$\Lambda u = \Lambda_1 u + \Lambda_2 u = (a_1 u_{\bar{x}_1})_{x_1} + (a_2 u_{\bar{x}_2})_{x_2}.$$

Escribamos un esquema de diferencias

$$\Lambda y = -f(x), \quad x \in \omega_h, \quad y = \mu(x), \quad x \in \gamma_h, \quad (23)$$
$$0 < c_1 \leqslant a_\alpha \leqslant c_2, \quad \alpha = 1, 2,$$

correspondiente al problema (22).

Introduzcamos en un espacio de funciones reticulares $H = \Omega_N$ el operador

$$Ay = -\Lambda \overset{\circ}{y}, \qquad A = A_1 + A_2,$$
$$A_1 y = -\Lambda_1 \overset{\circ}{y}, \qquad A_2 y = -\Lambda_2 \overset{\circ}{y}$$

y escribamos (23) en la forma operacional

$$Ay = \varphi, \qquad y,\ \varphi \in H,$$

donde φ difiere de f sólo en 4 nodos de frontera $(i_1 = 1,\ N_1 - 1,\ 0 < i_2 < N_2)$ y $(0 < i_1 < N_1,\ i_2 = 1,\ N_2 - 1)$.

El operador A es, evidentemente, autoconjugado: $(Ay,\ v) = (y,\ Av)$.

De la fórmula

$$-\sum_{i_1=1}^{N_1-1} (a_1 \overset{\circ}{y}_{\bar{x}_1})_{x_1,\, i_1} \overset{\circ}{y}_{i_1} h_1 = \sum_{i_1=1}^{N_1} (a_1 (\overset{\circ}{y}_{\bar{x}_1})^2)_{i_1} h_1$$

y de la desigualdad $0 < c_1 \leqslant a \leqslant c_2$ proviene que $c_1 (Ry,\ y) \leqslant (Ay,\ y) \leqslant c_2 (Ry,\ y)$, o bien $c_1 R \leqslant A \leqslant c_2 R$, (24)

donde R es el operador de Laplace estudiado más arriba

$$Ry = -\overset{\circ}{y}_{\bar{x}_1 x_1} - \overset{\circ}{y}_{\bar{x}_2 x_2}. \tag{25}$$

De aquí concluimos que

$$c_1 \overset{\circ}{\delta} E \leqslant A \leqslant c_2 \overset{\circ}{\Delta} E,$$

donde $\overset{\circ}{\delta}$ y $\overset{\circ}{\Delta}$ se definen por las fórmulas (20), (21).

Para resolver el problema (23) podemos aprovechar el método alternado triangular con un operador

$$B = (E + \omega R_1)(E + \omega R_2), \quad R_1 + R_2 = R, \quad R_1^* = R_2 \text{ para } D = E.$$

En este caso tenemos $\gamma_1 B \leqslant A \leqslant \gamma_2 B$, donde $\gamma_1 = c_1 \overset{\circ}{\gamma}_1$, $\gamma_2 = c_2 \overset{\circ}{\gamma}_2$, mientras que $\overset{\circ}{\gamma}_1$ y $\overset{\circ}{\gamma}_2$ se han encontrado para el operador (25). Para el número de iteraciones tenemos la siguiente estimación

$$n_0(\varepsilon) \approx \sqrt{\frac{c_2}{c_1}}\, \overline{n}_0(\varepsilon), \qquad \overline{n}_0(\varepsilon) = \frac{2}{2\sqrt[4]{2}\sqrt[4]{\eta}} \ln \frac{2}{\varepsilon}.$$

Para una ecuación de coeficientes variables se requieren $\sqrt{c_2/c_1}$ veces más iteraciones que para la ecuación de Poisson.

Sin embargo, podemos omitir la introducción del operador R, correspondiente al operador de Laplace, representando inmediatamente el operador de coeficientes variables en la forma

$$A = A_1 + A_2,$$

$$A_1 y = \frac{1}{h_1}\left(a_1 y_{\bar{x}_1} + \frac{1}{2} y a_{x_1}\right) + \frac{1}{h_2}\left(a_2 y_{\bar{x}_2} + \frac{1}{2} y a_{x_2}\right),$$

$$A_2 y = -\frac{1}{h_1}\left(a_1^{(+1_1)} y_{x_1} + \frac{1}{2} y a_{x_1}\right) - \frac{1}{h_2}\left(a_2^{(+1_2)} y_{x_2} + \frac{1}{2} y a_{x_2}\right).$$

El operador B se elige en la forma

$$B = (D + \omega A_1) D^{-1} (D + \omega A_2), \tag{26}$$

donde $D = d(x) E$ es una matriz diagonal. Para poder aplicar la teoría general se deben hallar las constantes δ y Δ que figuran en las condiciones $A \geqslant \delta D$, $A_1 D^{-1} A_2 \leqslant \frac{\Delta}{4} A$. El coeficiente $d(x)$ se escoge a partir de la condición de mínimo de la razón $\eta = \delta/\Delta$, y, por consiguiente, de máximo de $\xi = \gamma_1/\gamma_2$. De resultas se obtiene un algoritmo en el que el número de iteraciones $n_0(\varepsilon)$ depende poco de la razón c_2/c_1. De esto precisamente nos dice la tabla 3.

TABLA 3

$\frac{c_2}{c_1}$	$h = 1/32$		$h = 1/128$	
	$D = E$	$D = d(x) E$	$D = E$	$D = d(x) E$
2	23	20	45	39
8	46	23	90	47
32	92	25	180	53
128	184	26	360	57
512	367	26	720	59

Capítulo VII

Métodos de diferencias para resolver la ecuación de conductibilidad térmica

En el presente capítulo se examinan los esquemas de diferencias para resolver la ecuación de conductibilidad térmica. Se ha investigado detalladamente una ecuación unidimensional con coeficientes constantes. Se aducen esquemas de diferencias para una ecuación multidimensional de conductibilidad térmica con coeficientes variables.

§ 1. Ecuación de conductibilidad térmica con coeficientes constantes

1. Problema de partida. El proceso de difusión del calor en un vástago unidimensional $0 < x < l$ se describe por la ecuación de conductibilidad térmica

$$c\rho \frac{\partial u}{\partial t} = \frac{\partial}{\partial x}\left(k \frac{\partial u}{\partial x}\right) + f_0(x, t), \qquad (1)$$

donde $u = u(x, t)$ es la temperatura en el punto x del vástago en el momento t; c es la capacidad calorífica de la unidad de masa, ρ es la densidad de la masa, $c\rho$ es la capacidad calorífica de la unidad de longitud, k es el coeficiente de conductibilidad térmica y f_0, la densidad de las fuentes térmicas. En el caso general k, c, ρ, f_0 pueden depender no sólo de x y t, sino también de la temperatura $u = u(x, t)$ (ecuación casi lineal de conductibilidad térmica) e incluso de $\partial u/\partial x$ (ecuación no lineal). Si k, c, ρ son constantes, entonces (1) puede anotarse en la forma

$$\frac{\partial u}{\partial t} = a^2 \frac{\partial^2 u}{\partial x^2} + f, \qquad f = \frac{f_0}{c\rho}, \qquad (2)$$

donde $a^2 = k/(c\rho)$ es el coeficiente de conducción de temperatura. Sin perjudicar la generalidad de nuestros razona-

mientos podemos considerar $a = 1$, $l = 1$. En efecto, introduciendo las variables $x_1 = \frac{x}{l}$, $t_1 = \frac{a^2 t}{l^2}$, $f_1 = \frac{u^2}{a^2} f$, obtenemos

$$\frac{\partial u}{\partial t_1} = \frac{\partial^2 u}{\partial x_1^2} + f_1, \qquad 0 < x_1 < 1.$$

Se examinará aquí el primer problema de contorno (a veces suele decirse: problema inicial de contorno) en el dominio $\overline{D} = \{0 \leqslant x \leqslant 1,\ 0 \leqslant t \leqslant T\}$. Se pide hallar la solución $u(x, t)$, continua en $\overline{D}$, del problema

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2} + f(x, t), \quad 0 < x < 1, \quad 0 < t \leqslant T,$$
$$u(x, 0) = u_0(x),\ 0 \leqslant x \leqslant 1,\ u(0, t) = u_1(t), \quad (3)$$
$$u(1, t) = u_2(t), \qquad 0 \leqslant t \leqslant T.$$

2. Algunas propiedades de las soluciones de la ecuación de conductibilidad térmica. En virtud del principio del máximo, para la solución del problema (3) tiene lugar una estimación

$$\max_{0 \leqslant x \leqslant 1,\ 0 \leqslant t \leqslant T} |u(x, t)| \leqslant$$
$$\leqslant \max\Big(\max_{0 \leqslant x \leqslant 1} |u_0(x)|, \max_{0 \leqslant t \leqslant T} |u_1(t)|, \max_{0 \leqslant t \leqslant T} |u_2(t)|\Big) +$$
$$+ \int_0^T \max_{0 \leqslant x \leqslant 1} |f(x, t)|\, dt. \quad (4)$$

Tomemos una ecuación homogénea con las condiciones de contorno homogéneas:

$$\frac{\partial u}{\partial t} = \frac{\partial^2 u}{\partial x^2}, \qquad 0 < x < 1, \qquad 0 < t < T,$$
$$u(0, t) = u(1, t) = 0, \qquad 0 \leqslant t \leqslant T, \quad (5)$$
$$u(x, 0) = u_0(x), \qquad 0 \leqslant x \leqslant 1.$$

La solución de este problema se halla por el método de separación de variables en la forma

$$u(x, t) = \sum_{k=1}^{\infty} c_k e^{-\lambda_k t} X_k(x), \quad (6)$$

donde λ_k y $X_k(x)$ son los valores propios y las funciones propias ortonormalizadas del problema

$$X'' + \lambda X = 0, \qquad 0 < x < 1, \qquad X(0) = X(1) = 0,$$

iguales a

$$\lambda_k = k^2\pi^2, \qquad X_k(x) = \sqrt{2}\,\operatorname{sen} k\pi x, \tag{7}$$

con la particularidad de que

$$(X_k, X_m) = \int_0^1 X_k(x)\, X_m(x)\, dx = \delta_{km},$$

$$\delta_{km} = \begin{cases} 1, & k = m, \\ 0, & k \neq m. \end{cases}$$

Efectivamente, todas las soluciones particulares (armónicas) $u_k(x, t) = c_k e^{-\lambda_k t} X_k(x)$ satisfacen la ecuación y las condiciones de contorno (5). De la condición inicial

$$u(x, 0) = u_0(x) = \sum_{k=1}^{\infty} c_k X_k(x) \tag{8}$$

se hallan los coeficientes $c_k = (u_0, X_k)$.

De (6) y (8) se infiere

$$\| u(t) \|^2 = (u(x, t), u(x, t)) =$$

$$= \sum_{k=1}^{\infty} c_k^2 e^{-2\lambda_k t} \| X_k \|^2 \leqslant e^{-2\lambda_1 t} \sum_{k=1}^{\infty} c_k^2 = e^{-2\lambda_1 t} \| u_0 \|^2,$$

puesto que

$$\| u_0 \|^2 = \sum_{k=1}^{\infty} c_k^2, \qquad \lambda_k > \lambda_{k-1} > \ldots > \lambda_1 = \pi^2.$$

De este modo, para la solución del problema (5) resulta justa la estimación

$$\| u(t) \| \leqslant e^{-\lambda_1 t} \| u_0 \|, \qquad \lambda_1 = \pi^2, \tag{9}$$

que exprese la propiedad de estabilidad asintótica (para $t \to \infty$) del problema (5) respecto de los datos iniciales (§ 4, p. 7, cap. V). Debido a que $\lambda_k = k^2\pi^2$ crece con el crecimiento de k, a partir de cierto momento t, en la suma (6)

se hará preponderante el primer sumando (primera armónica), es decir, tendrá lugar una igualdad aproximada

$$u(x, t) \approx c_1 e^{-\lambda_1 t} X_1(x).$$

Esta etapa del proceso lleva el nombre de régimen regular.

3. Esquemas de diferencias. En el dominio D introduzcamos una red

$$\bar{\omega}_{h\tau} = \{(x_i t_j)\colon x_i = ih, \quad t_j = j\tau, \quad i = 0, 1, \ldots$$
$$\ldots, N, h = 1/N, \quad j = 0, 1, \ldots, L, \quad \tau = T/L\}$$

con los pasos: h según x y τ según t. Al cambiar la derivada respecto de x por la expresión de diferencias

$$\left(\frac{\partial^2 u}{\partial x^2}\right)_i \sim \frac{u_{i+1} - 2u_i + u_{i-1}}{h^2} = u_{\bar{x}x,\, i} = \Lambda u_i,$$

obtendremos en lugar de (3) un sistema de ecuaciones diferenciales en diferencias (*método de las rectas*)

$$\frac{dv_i}{dt} = \Lambda v_i + f_i, \qquad i = 1, 2, \ldots,$$

con las condiciones de contorno e iniciales

$$v_0(t) = u_1(t), \quad v_N(t) = u_2(t), \quad v_i(0) = u_0(x_i).$$

Para la resolución numérica de este problema sustituyamos, por analogía con el cap. V, la derivada respecto de t por una razón de diferencias

$$\frac{dv_i}{dt} \sim \frac{v_i(t_{j+1}) - v_i(t_j)}{\tau} = \frac{v_i^{j+1} - v_i^j}{\tau} = (v_t)_i^j,$$

y tomemos el segundo miembro en forma de una combinación lineal de valores para $t = t_j$ (en la j-ésima capa) y $t = t_{j+1}$ (en la $(j+1)$-ésima capa):

$$\frac{y_i^{j+1} - y_i^j}{\tau} = \sigma \Lambda y_i^{j+1} + (1 - \sigma)\Lambda y_i^j + \varphi_i^j, \tag{10}$$

donde σ es el parámetro, mientras que φ_i^j es cierto segundo miembro, por ejemplo, $\varphi_i^j = f_i^j$, $\varphi_i^j = f_i^{j+1/2}$, etc. Se deben agregar aquí las condiciones complementarias

$$y_0^j = u_1(t_j), \quad y_N^j = u_2(t_j), \quad y_i^0 = u_0(x_i), \tag{11}$$
$$j = 0, 1, 2, \ldots, \qquad 0 \leqslant i \leqslant N.$$

El esquema (10) está definido en un molde 6-puntual

$$\begin{array}{ccc} (x_{i-1},\ t_{j+1}) & (x_i,\ t_{j+1}) & (x_{i+1},\ t_{j+1}) \\ \times & \times & \times \\ \times & \times & \times \\ (x_{i-1},\ t_j) & (x_i,\ t_j) & (x_{i+1},\ t_j) \end{array}$$

Examinemos un esquema explícito ($\sigma = 0$) en el molde 4-puntual:

$$\frac{y_i^{j+1}-y_i^j}{\tau} = \frac{y_{i-1}^j - 2y_i^j + y_{i+1}^j}{h^2} + \varphi_i^j. \qquad (12)$$

Los valores en la $(j+1)$-ésima capa se hallan por la fórmula explícita

$$y_i^{j+1} = \left(1 - \frac{2\tau}{h^2}\right) y_i^j + \frac{\tau}{h^2}\,(y_{i-1}^j + y_{i+1}^j) + \tau\varphi_i^j.$$

En el caso de $\sigma = 1$ obtenemos un esquema completamente implícito, esto es, un esquema con adelantamiento en el molde$^{***}_{*}$:

$$\frac{y_i^{j+1}-y_i^j}{\tau} = \frac{y_{i-1}^{j+1} - 2x_i^{j+1} + y_{i+1}^{j+1}}{h^2} + \varphi_i^j. \qquad (13)$$

Para determinar y_i^{j+1} de (13) obtenemos el problema de contorno

$$\frac{\tau}{h^2}\, y_{i-1}^{j+1} - \left(1 + \frac{2\tau}{h^2}\right) y_i^{j+1} + \frac{\tau}{h^2}\, y_{i-1}^{j+1} = -F_i^j, \quad 0 < i < N,$$

$$F_i^j = y_i^j + \tau\varphi_i^j, \quad y_0^{j+1} = u_1(t_{j+1}), \quad y_N^{j+1} = u_2(t_{j+1}),$$

el cual se resuelve por el método de factorización.

Se usa frecuentemente un esquema implícito simétrico (a veces se denomina *esquema de Crank — Nickolson*) con $\sigma = 1/2$ y un molde$^{***}_{***}$:

$$\frac{y_i^{j+1}-y_i^j}{\tau} = \frac{1}{2}\left(\frac{y_{i-1}^{j+1} - 2y_i^{j+1} + y_{i+1}^{j+1}}{h^2} + \frac{y_{i-1}^j - 2y_i^j + y_{i+1}^j}{h^2}\right) + \varphi_i^j. \qquad (14)$$

Los valores de y^{+1} en la nueva capa se determinan en este caso también por el método de factorización para el problema

de contorno:

$$\frac{\tau}{2h^2}y_{i-1}^{j+1}-\left(1+\frac{\tau}{h^2}\right)y_i^{j+1}+\frac{\tau}{2h^2}y_{i-1}^{j+1}=-F_i^j, \quad 0<i<N,$$

$$y_0^{j+1}=u_1(t_{j+1}), \qquad y_N^{j+1}=u_2(t_{j+1}), \tag{15}$$

$$F_i^j=\left(1-\frac{\tau}{h^2}\right)y_i^j+\frac{\tau}{2h^2}(y_{i+1}^j+y_{i+1}^j)+\tau\varphi_i^j.$$

En el caso general (para σ cualquiera) el esquema (10) se denomina *esquema con pesos*. Cuando $\sigma = 0$, el esquema es implícito e y_i^{j+1} se determina por el método de factorización como una solución del problema

$$\sigma\tau\Lambda y_i^{j+1}-y_i^{j+1}=-F_i^j, \qquad 0<i<N,$$

$$y_0^{j+1}=u_1(t_{j+1}), \quad y_N^{j+1}=u_2(t_{j+1}), \qquad j=0,\ 1,\ \ldots \tag{16}$$

Procedamos ahora al estudio de las propiedades del esquema (10) con σ cualquiera.

4. Estimación del error de aproximación. Con el fin de estimar el orden de exactitud del esquema con pesos (10), se debe estimar primero el error de aproximación (el residuo) y hallar las estimaciones apriorísticas que expresan la estabilidad del esquema respecto del segundo miembro. El esquema de diferencias (10), (11) toma en consideración los datos iniciales y de frontera exactos. Escribamos el esquema (10) en la forma sin índices. Al introducir las designaciones

$$y=y_i^j, \quad \hat{y}=y_i^{j+1}, \quad \Lambda y=y_{\bar{x}x}=\frac{y_{i+1}-2y_i+y_{i-1}}{h^2},$$

$$y_t=\frac{y_i^{j+1}-y_i^j}{\tau}, \quad y^{(\sigma)}=\sigma y_i^{j+1}+(1-\sigma)\,y_i^j,$$

obtenemos

$$y_t=\Delta y^{(\sigma)}+\varphi, \quad (x_i t_j)\in\omega_{h\tau}, \quad y(x,\,0)=u_0(x),$$

$$y_0=\mu_1(t), \quad y_N=u_2(t) \quad (t=t_j=j\tau, \quad j=0,\ 1,\ \ldots). \tag{17}$$

Sea $u = u(x_i,\ t_j)$ la solución exacta del problema de partida (3) y sea y la solución del problema de diferencias (17).

Sustituyendo en (17) $y = z + u$, obtendremos para el error $z = y - u$ las siguientes condiciones:

$$z_t = \Lambda z^{(\sigma)} + \psi, \qquad (x, t) \in \omega_{h\tau},$$
$$z(x, 0) = 0, \qquad z(0, t) = z(1, t) = 0, \quad (18)$$

donde

$$\psi = \Lambda u^{(\sigma)} + \varphi - u_t \quad (19)$$

es el error de aproximación del esquema (17) en la solución $u = u(x, t)$ del problema (3) (el residuo del esquema).

Encontraremos el desarrollo de ψ según las potencias de h y τ en el entorno del punto $(x_i, \bar{t} = t_j + {}^1/_2\tau)$. Al tomar en consideración que

$$u^{(\sigma)} = \sigma\hat{u} + (1-\sigma)u = \frac{u+\hat{u}}{2} + \left(\sigma - \frac{1}{2}\right)\tau u_t,$$

$$\hat{v} = \bar{v} + \frac{1}{2}\tau\frac{\partial\bar{v}}{\partial t} + \frac{\tau^2}{8}\frac{\partial^2\bar{v}}{\partial t^2} + \frac{\tau^3}{48}\frac{\partial^3\bar{v}}{\partial t^3} + O(\tau^4),$$

$$\bar{v} = v\left(x, t_j + \frac{1}{2}\tau\right),$$

$$v = \bar{v} - \frac{1}{2}\tau\frac{\partial\bar{v}}{\partial t} + \frac{\tau^2}{8}\frac{\partial^2\bar{v}}{\partial t^2} - \frac{\tau^3}{48}\frac{\partial^3\bar{v}}{\partial t^3} + O(\tau^4),$$

$$\Lambda u = u_{\bar{x}x} = Lu + \frac{h^2}{12}u^{IV} + O(h^4), \qquad Lu = \frac{\partial^2 u}{\partial x^2},$$

obtenemos

$$\psi = L\bar{u} + \bar{f} - \frac{\partial\bar{u}}{\partial t} + \varphi - \bar{f} + \left(\sigma - \frac{1}{2}\right)\tau L\frac{\partial u}{\partial t} + \\ + \frac{h^2}{12}L^2u + O(\tau^2 + h^4).$$

Por cuanto, en virtud de la ecuación (3),

$$L\bar{u} + \bar{f} - \frac{\partial\bar{u}}{\partial t} = 0,$$

entonces

$$L\frac{\partial u}{\partial t} = L^2u + Lf$$

y

$$\psi + \left(\varphi - \overline{f} + \left(\sigma - \frac{1}{2}\right)\tau L\overline{f}\right) + \\ + \left(\frac{h^2}{12} + \left(\sigma - \frac{1}{2}\right)\tau\right) L^2\overline{u} + O(\tau^2 + h^4).$$

De aquí se ve que

$$\psi = O(\tau + h^2) \text{ para } \varphi = f \text{ y } \sigma \neq \frac{1}{2},$$

$$\psi = O(\tau^2 + h^2) \text{ para } \varphi = \overline{f} \text{ y } \sigma = \frac{1}{2}.$$

Si elegimos σ de un modo tal que el coeficiente de $L^2\bar{u}$ sea nulo:

$$\sigma = \sigma_* = \frac{1}{2} - \frac{h^2}{12\tau}, \tag{20}$$

y ponemos φ igual a

$$\varphi = \overline{f} + \frac{h^2}{12} L\overline{f}, \text{ o bien } \varphi = \overline{f} + \frac{h^2}{12} \Lambda\overline{f} \tag{21}$$

(ambas expresiones se diferencian en una magnitud $O(h^4)$, puesto que $\Lambda f - Lf = O(h^2)$), obtendremos un esquema con el orden de aproximación aumentado (respecto de x): $\psi = O(h^4 + \tau^2)$ para $\sigma = \sigma_*$. Este esquema es también implícito, por lo cual y_i^{j+1} se halla de la ecuación $\sigma_*\tau\Lambda\hat{y} - \hat{y} = -F$ por el método de factorización.

5. Estabilidad del esquema. Volvamos al estudio de la estabilidad y de la convergencia del esquema (17). Veamos, primero, el esquema explícito ($\sigma = 0$) y el esquema implícito puro ($\sigma = 1$). La ecuación (17) para el esquema explícito se escribirá en la forma

$$y_i^{j+1} = \left(1 - \frac{2\tau}{h^2}\right) y_i^j + \frac{\tau}{h^2}(y_{i-1}^j + y_{i+1}^j) + \tau\varphi_i^j, \quad 0 < i < N, \tag{22}$$

$$y_0^{j+1} = 0, \quad y_N^{j+1} = 0, \quad y_i^0 = u_0(x_i), \quad 0 \leqslant i \leqslant N.$$

Si el coeficiente de y_i^j es no negativo, es decir, si

$$\tau \leqslant h^2/2, \tag{23}$$

entonces de (22) proviene que

$$\| y^{j+1} \|_C \leqslant \| y^j \|_C + \tau \| \varphi^j \|_C, \qquad (24)$$

donde $\| y \|_C = \max\limits_{0 \leqslant i \leqslant N} | y_i |$. La sumación según k de 0 a $j-1$ nos da

$$\| y^j \|_C \leqslant \| y^0 \|_C + \sum_{k=0}^{j-1} \tau \| \varphi^k \|_C. \qquad (25)$$

Esta desigualdad expresa precisamente la estabilidad en una norma reticular C del esquema explícito respecto de los datos iniciales y del segundo miembro bajo la condición (23) (el esquema explícito es condicionalmente estable).

El esquema implícito (17) con $\sigma = 1$ se escribirá en la forma

$$\tau \Lambda y_i^{j+1} - y_i^{j+1} = -F_i, \qquad F_i = y_i^j + \tau \varphi_i^j$$

o bien

$$\frac{\tau}{h^2} y_{i-1}^{j+1} - \left(1 + \frac{2\tau}{h^2}\right) y_i^{j+1} + \frac{\tau}{h^2} y_{i+1}^{j+1} = -F_i^j, \quad 0 < i < N,$$

$$y_0^{j+1} = y_N^{j+1} = 0.$$

Ahora, hagamos uso del teorema 3 del § 5, cap. I: para la solución del problema

$$A_i y_{i-1} - C_i y_i + A_{i+1} y_{i+1} = -F_i,$$

$$C_i = A_i + A_{i+1} + D_i, \quad 0 < i < N, \quad y_0 = y_N = 0$$

es justa la estimación

$$\| y \|_C \leqslant \left\| \frac{F}{D} \right\|_C.$$

En nuestro caso $A_i = A_{i+1} = \tau/h^2$, $D_i = 1$,

$$\| y^{k+1} \|_C \leqslant \| F^k \|_C \leqslant \| y^k \|_C + \tau \| \varphi^k \|_C. \qquad (26)$$

De aquí, sumando según $k = 0, 1, \ldots, j-1$, obtenemos la estimación (25). De este modo, un esquema implícito puro es incondicionalmente estable, es decir, es estable para cualesquiera τ y h. Siendo σ arbitrario, la ecuación en di-

ferencias tiene por expresión

$$\frac{\sigma\tau}{h^2}\,y_{i-1}^{j+1}-\left(1+\frac{2\sigma\tau}{h^2}\right)y_i^{j+1}+\frac{\sigma\tau}{h^2}\,y_{i+1}^{j+1}=-F_i^j$$

$$0<i<N, \qquad y_0^{j+1}=y_N^{j+1}=0,$$

$$F_i^j=\left(1-\frac{2(1-\sigma)\tau}{h^2}\right)y_i^j+\frac{(1-\sigma)}{h^2}\tau\,(y_{i-1}^j+y_{i+1}^j)+\tau\varphi_i^j.$$

De aquí se ve que el coeficiente de y_i^j es no negativo, si

$$\tau\leqslant\frac{h^2}{2(1-\sigma)} \quad \text{o bien} \quad \sigma\geqslant 1-\frac{h^2}{2\tau}. \tag{27}$$

Bajo esta condición $\|F\|_C\leqslant\|y\|_C+\tau\|\varphi\|_C$; aprovechando, luego, el teorema 3 del § 5, cap. I, obtendremos la estimación (25) para la condición (27). En particular, para un esquema simétrico la estabilidad en C tiene lugar cuando $\tau\leqslant h^2$. En realidad el esquema (17), con $\sigma\geqslant 1/2$ es incondicionalmente estable en C respecto de los datos iniciales, de suerte que

$$\|y^j\|_C\leqslant M_0\,\|\mathring{y}\|_C,$$

donde $M_0=\text{const}>1$. No obstante, esta desigualdad se demuestra de un modo bastante complejo.

Más abajo se probará que en otra norma la condición de estabilidad de un esquema con pesos tiene por expresión

$$\sigma\geqslant\sigma_0=\frac{1}{2}-\frac{h^2}{4\tau}, \tag{28}$$

de modo que el esquema con $\sigma\geqslant 1/2$ es incondicionalmente estable, mientras que para $\sigma<1/2$ la condición de estabilidad, en lugar de (27), se expresa así

$$\tau\leqslant\frac{h^2}{4(1/2-\sigma)}. \tag{29}$$

El resultado obtenido (29) se obtiene a base de la teoría general de estabilidad.

Por analogía con el § 4, cap. I, introduzcamos un operador A:

$$Ay=-\Lambda\mathring{y}, \qquad y\in\Omega, \qquad \mathring{y}\in\mathring{\Omega},$$

donde $\mathring{\Omega}$ es un conjunto de funciones y definidas sobre la red $\mathring{\omega}_h=\{x_i:\ x_i=ih,\ i=0,\ 1,\ \ldots,\ N,\ h=1/N\}$ e igua-

les a cero en la frontera para $i = 0, N$; o y es un conjunto de funciones definidas en los nodos interiores de la red $x \in \in \omega_h = \{x_i : x_i = ih,\ i = 1, 2, \ldots, N-1,\ h = 1/N\}$.

Escribamos el esquema con pesos en la forma canónica:

$$Bz_t + Az = \psi(t),\ t \in \bar{\omega}_\tau,\quad Z(0) = 0,\ B = E + \sigma\tau A. \tag{30}$$

Con este fin resulta suficiente sustituir $z^{(\sigma)} = \sigma\hat{z} + (1 - \sigma)z = z + \sigma(\hat{z} - z) = z + \sigma\tau z_t$ en (18).

El operador A es, de acuerdo con lo mostrado en el cap. I, autoconjugado y positivo: $A = A^* > 0$, si el producto escalar en H lo definimos según la fórmula

$$(y, v) = \sum_{i=1}^{N-1} y_i v_i h.$$

La estabilidad del esquema (30) fue investigada en el cap. V, donde probamos que el esquema (30) es estable en H_A cuando

$$\sigma \geqslant \sigma_0 = \frac{1}{2} - \frac{1}{\tau \| A \|}. \tag{31}$$

En nuestro caso $\| A \| = \frac{4}{h^2}\cos^2\frac{\pi h}{2}$. De aquí proviene que el esquema (17) es estable para cualesquiera τ y h, siempre que $\sigma \geqslant 1/2$. Si $\sigma < 1/2$, el esquema será estable para

$$\tau \leqslant \frac{1}{(1/2 - \sigma)\| A \|}.$$

Al sustituir aquí $\| A \| \approx 4/h^2$, obtenemos

$$\tau \leqslant \frac{h^2}{4(1/2 - \sigma)} \quad \text{y} \quad 4(1/2 - \sigma)\tau \leqslant h^2.$$

En particular, cuando $\sigma = \sigma_*$, tenemos $4(1/2 - \sigma_*)\tau = h^2/3 < h^2$, es decir, el esquema con un orden de aproximación aumentado es incondicionalmente estable.

6. Convergencia del esquema. Con el fin de demostrar la convergencia del esquema (17) se debe obtener la estimación apriorística para el problema (30). Hagamos uso de la desigualdad para z, obtenida al investigar la convergencia de los esquemas en el cap. V, en virtud de la cual para (30)

y (18) es justa la estimación del error

$$\| z^j \|_A \leqslant \sum_{k=0}^{j-1} \tau \| \psi^k \| \text{ para } \sigma \geqslant 0,\ \sigma \geqslant \sigma_0. \tag{32}$$

Al sustituir aquí $Az = \overset{\circ}{z}_{\bar{x}x}$, hallaremos

$$\| z \|_A^2 = (Az, z) = -(\overset{\circ}{z}_{\bar{x}, x}, \overset{\circ}{z}) = (\overset{\circ}{z}_{\bar{x}}, \overset{\circ}{z}_{\bar{x}}] = \sum_{i=1}^{N} h (z_{\bar{x}, i})^2$$

y aprovecharemos la estimación

$$\| z \|_C = \max_{x \in \omega_h} | z | \leqslant \frac{1}{2} \left(\sum_{i=1}^{N} h (z_{\bar{x}, i})^2 \right)^{1/2} = \frac{1}{2} \| z \|_A.$$

De resultas obtenemos

$$\| z^j \|_C \leqslant \frac{1}{2} \sum_{k=0}^{j-1} \tau \| \psi^k \|, \tag{33}$$

es decir, el esquema (17) converge en la norma reticular C con la velocidad $\| y^j - u^j \|_C = \| z^j \|_C = O(h^2 + \tau)$ para $\sigma \neq 1/2$, $\sigma \geqslant \sigma_0$, $\| z^j \|_C = O(h^2 + \tau^2)$, $\sigma = 1/2$. Si $\sigma_* \geqslant 0$, es decir, si $\tau \geqslant h^2/\sigma$, entonces para el esquema $\sigma = \sigma_*$ es también justa la estimación (33) y

$$\| z^j \|_C = O(h^4 + \tau^2) \text{ para } \sigma = \sigma_*.$$

7. Estabilidad asintótica. La propiedad de estabilidad *asintótica (para $t \to \infty$) del problema (5) respecto de los* datos iniciales se expresa por la estimación (9). Para t grandes la solución del problema (5) se determina por la primera armónica

$$u(x, t) \approx c_1 e^{-\lambda_1 t} X_1(x)$$

(régimen regular). Es natural exigir que la solución del problema de diferencias

$$\begin{aligned} y &= \sigma \Lambda \hat{y} + (1 - \sigma) \Lambda y; \quad x = ih, \quad t = j\tau, \\ & i = 1, 2, \ldots, N - 1, \quad j = 0, 1, \ldots, \\ & y(0, t) = 0,\ y(1, t) = 0, \quad y(x, 0) = u_0(x) \end{aligned} \tag{34}$$

posea las propiedades analíticas.

En el cap. V para el esquema operacional de diferencias con pesos

$$By_t + Ay = 0, \quad t \in \omega_\tau, \quad y(0) = y_0, \quad B = E + \sigma\tau A,$$
$$\delta E \leqslant A \leqslant \Delta E, \quad \delta > 0, \quad A = A^* > 0$$

se ha establecido la estabilidad asintótica del esquema con pesos

$$\|y^j\| \leqslant e^{-\delta t_j} \|\overset{\circ}{y}\|$$

con la condición complementaria

$$\tau \leqslant \tau_0 \ (\sigma)$$

donde $\tau_0 = 2/(\delta + \Delta)$ para un esquema explícito ($\sigma = 0$), $\tau_0 = \infty$ (τ es cualquiera) para un esquema implícito ($\sigma = 1$) y $\tau_0 = 2/\sqrt{\delta\Delta}$ para un esquema simétrico ($\sigma = 1/2$). Para el esquema (34) tendremos

$$\delta = \frac{4}{h^2} \operatorname{sen}^2 \frac{\pi h}{2}, \quad \Delta = \frac{4}{h^2} \cos^2 \frac{\pi h}{2}, \quad \delta + \Delta = \frac{4}{h^2}.$$

Para un esquema explícito ($\sigma = 0$) $\tau_0 = h^2/2$ y la condición de estabilidad asintótica coincide con la de estabilidad corriente; el esquema implícito ($\sigma = 1$) es, como antes, incondicionalmente estable. Sin embargo, el esquema simétrico ($\sigma = 1/2$) será incondicionalmente estable en el sentido habitual y asintóticamente estable bajo la condición

$$\tau \leqslant \tau_0, \quad \tau_0 = \frac{h^2}{\operatorname{tg} \pi h} \approx \frac{h}{\pi}.$$

En este caso la solución del problema de diferencias (34) con $\sigma = 1/2$, para t grandes, se determina por la primera armónica:

$$y_i^j \approx c_1 \rho^j \operatorname{sen} \pi x_i \approx c_1 e^{-\lambda_1 t_j} \operatorname{sen} \pi x.$$

Aquí $\rho = (1 - 1/2\tau\delta)/(1 + 1/2\tau\delta) = e^{-\lambda_1 t} (1 + 0 (\tau^2))$.

Si la condición $\tau \leqslant \tau_0$ está perturbada, es decir, si $\tau > \tau_0$, entonces para t grandes predomina no la primera, sino la última armónica:

$$y_i^j \approx c_1 \rho^j \operatorname{sen} \pi (N - 1) x_i \approx c_1 \rho^j (-1)^i \operatorname{sen} \pi x_i,$$

donde $\rho = \dfrac{1/2\tau\Delta - 1}{1/2\tau\Delta + 1} < e^{-\lambda_1 \tau}$, lo que, por supuesto, no tiene nada en común con la solución de una ecuación diferencial.

La exigencia de la estabilidad asintótica está estrechamente ligada con la exactitud del esquema y de hecho significa también la exigencia de exactitud asintótica. Esto se pone de manifiesto con mayor claridad en los cálculos sobre las redes reales para t grandes. Hemos de observar que la condición $\tau \approx h/\pi$ para un esquema simétrico no parece abrumadora. Se demuestra que un esquema implícito puro ($\sigma = 1$) puede asegurar una exactitud admisible para grandes valores de t sólo cuando el paso τ sea comparable con el de un esquema explícito, con lo que en los cálculos para t grandes el esquema implícito puro queda privado de su ventaja principal, a saber, la estabilidad para τ y h cualesquiera.

§ 2. Problemas multidimensionales de la conductibilidad térmica

1. Esquema de diferencias con pesos. Examinemos en un plano $x = (x_1, x_2)$ un dominio G con la frontera Γ. Buscaremos la solución del problema de conductibilidad térmica en el dominio $\overline{G} = G + \Gamma$ para todo $0 \leqslant t \leqslant T$. Se pide hallar una función $u(x, t)$ que esté definida en el cilindro $\overline{Q_T} = \overline{G} \times [0, T] = \{(x, t): x \in G, 0 \leqslant t \leqslant T\}$ y que satisfaga en $Q_T = G \times (0, T] = \{(x, t): x \in G, 0 < t \leqslant T\}$ la ecuación de conductibilidad térmica

$$\frac{\partial u}{\partial t} = Lu + f(x, t), \quad Lu = \frac{\partial^2 u}{\partial x_1^2} + \frac{\partial^2 u}{\partial x_2^2}, \tag{1}$$

las condiciones de contorno de primera especie en la frontera Γ del dominio G

$$u = \mu(x, t), \quad x \in \Gamma, \quad 0 \leqslant t \leqslant T, \tag{2}$$

y la condición inicial para $t = 0$:

$$u(x, 0) = u_0(x), \quad x \in \overline{G}. \tag{3}$$

Supongamos que $\overline{G}$ es un rectángulo:

$$\overline{G} = \{x = (x_1, x_2): 0 \leqslant x_1 \leqslant l_1, \quad 0 \leqslant x_2 \leqslant l_2\}.$$

Introduzcamos en $\overline{G}$ una red rectangular

$$\overline{\omega}_h = \{x_i = (x_1^{(i_1)}, x_2^{(i_2)}): x_\alpha^{(i_\alpha)} = i_\alpha h_\alpha,$$
$$i_\alpha = 0, 1, \ldots, N_\alpha, \ h_\alpha = l_\alpha/N_\alpha, \ \alpha = 1, 2\}$$

con la frontera

$$\gamma_h = \{x_i = (i_1h_1,\ i_2h_2):\ i_1 = 0,\ N_1,\ 0 < i_2 < N_2:$$
$$i_2 = 0,\ N_2,\ 0 < i_1 < N_1\}.$$

Aproximemos el operador de Laplace $Lu = \Delta u$ mediante un operador de diferencias en el molde pentapuntual (véase el cap. VI, § 1)

$$Lu \sim \Lambda u = u_{\bar{x}_1x_1} + u_{\bar{x}_2x_2}.$$

Sustituyamos el problema (1) — (3) por un problema diferencial en diferencias (método de las rectas):

$$\frac{dv_i(t)}{dt} = \Lambda v_i(t) + f_i(t),\ i = (i_1, i_2),\ v_i(0) = u_0(x_i),$$
$$x_i \in \omega_h,\quad v_i(t)|_{\gamma_h} = \mu_i(t),\quad 0 \leqslant t \leqslant. \qquad (4)$$

Introduzcamos en el segmento $0 \leqslant t \leqslant T$ una red $\bar{\omega}_\tau = \{t_j = j\tau:\ 0 \leqslant t_j \leqslant T\}$ de paso τ. Escribamos un esquema con pesos

$$\frac{y^{j+1} - y^j}{\tau} = \Lambda\left(\sigma y^{j+1}\right) + (1-\sigma)\, y^j) + \varphi^j,\quad j = 0,\ 1,\ \ldots, \qquad (5)$$

donde $y^j = y(x_i, t_j) = y(i_1h_1,\ i_2h_2;\ t_j)$, $x = (i_1h_1,\ i_2h_2) \in \in \omega_h$. Agregamos a las ecuaciones (5)

$$y(x, 0) = u_0(x),\quad x = (i_1h_1,\ i_2h_2) \in \bar{\omega}_h,$$
$$y(x_1, t) = \mu_i(t),\quad x \in \gamma_h,\ t = j\tau \in \omega_h. \qquad (5')$$

De aquí se ve que para determinar $\hat{y} = y^{j+1}$ en una capa nueva $t = t_{j+1}$ se debe resolver una ecuación en diferencias

$$\hat{y} - \sigma\tau\Lambda\hat{y} = F,\ F = y + (1-\sigma)\,\tau\Lambda y + \tau\varphi,\ x \in \omega_h,$$
$$\hat{y} = \mu,\quad x \in \gamma_h. \qquad (6)$$

La resolubilidad de este problema se desprende de lo que el operador $(E - \sigma\tau\Lambda)$ es definido positivo para $\sigma > > -1/(\tau \| A \|)$, puesto que $(E - \sigma\tau\Lambda)\mathring{y} = (E + \sigma\tau A)\, y$ en el espacio de funciones reticulares $\mathring{y}$ que vienen dadas en la red $\bar{\omega}_h$ y que se reducen a cero en la frontera γ_h (compárese con el cap. VI). Mostrémoslo.

Introduciendo un producto escalar

$$(y, v) = \sum_{x_j \in \omega_h} y(x_j)\, v(x_j)\, h_1 h_2 =$$

$$= \sum_{i_1=1}^{N_1-1} h_1 \sum_{i_2=1}^{N_2-1} h_2 y(i_1 h_1, i_2 h_2)\, v(i_1 h_1, i_2 h_2) \qquad (7)$$

y teniendo presente que $(Ay, y) \leqslant \| Ay \| \| y \| \leqslant \| A \| \cdot \| y \|^2$, encontramos

$$((E - \sigma\tau\Lambda)\, \mathring{y}, \mathring{y}) = ((E + \sigma\tau A)\, y, y) =$$

$$= \|y\|^2 + \sigma\tau (Ay, y) \geqslant \left(\frac{1}{\|A\|} + \sigma\tau \right) (Ay, y) > 0,$$

puesto que $(Ay, y) \geqslant \delta \| y \|^2 > 0$ (véase el cáp. VI, § 1, p. 5).

Escribamos detalladamente en la forma de índices una ecuación en diferencias

$$\sigma\gamma_1 (\hat{y}_{i_1-1,\, i_2} + \hat{y}_{i_1+1,\, i_2}) - (1 + 2\sigma(\gamma_1 + \gamma_2)) \times$$

$$\times \hat{y}_{i_1 i_2} + \sigma\gamma_2 (\hat{y}_{i_1 i_2 - 1} + \hat{y}_{i_1 i_2 + 1}) = -F_{i_1 i_2}, \qquad (8)$$

donde

$$y_{i_1 i_2} = y(i_1 h_1, i_2 h_2), \quad \gamma_1 = \tau/h_1^2, \quad \gamma_2 = \tau/h_2^2,$$

$$F_{i_1 i_2} = (1 - 2(1 - \tau)(\gamma_1 + \gamma_2))\, y_{i_1 i_2} + (1 - \sigma) \times$$

$$\times \gamma_1 (y_{i_1-1,\, i_2} + y_{i_1+1,\, i_2}) + (1 - \sigma)\, \gamma_2 \times$$

$$\times (y_{i_1,\, i_2-1} + y_{i_1 i_2 + 1}) + \varphi_{i_1 i_2},$$

$$\hat{y}_{i_1 i_2} = \hat{\mu}_{i_1 i_2}, x_i = (i_1 h_1, i_2 h_2) \in \gamma_h.$$

Dicho problema de contorno en diferencias se resuelve respecto de $\hat{y}$ por los mismos métodos que se usan en la resolución del problema de Dirichlet en diferencias para la ecuación de Poisson (véase el cap. VI, § 2). Aquí los coeficientes de la ecuación son constantes, el dominio G es un rectángulo, razón por la cual los métodos directos de resolución de las ecuaciones en diferencias (8) resultan los más económicos. Los métodos iterativos son menos económicos.

2. Estabilidad y convergencia. Haciendo uso del operador A, definido anteriormente en el cap. VI:

$$Ay = -\Lambda \overset{\circ}{y} = -\overset{\circ}{y}_{\bar{x}_1 x_1} - \overset{\circ}{y}_{\bar{x}_2 x_2}, \quad \overset{\circ}{y} \in \overset{\circ}{\Omega}, \quad y \in \Omega = H,$$

escribamos el esquema (5) en la forma canónica:

$$B\frac{y^{j+1}-y^j}{\tau} + Ay^j = \varphi^j, \quad j = 0, 1, \ldots, \quad \overset{\circ}{y} = u_0, \quad y \in H,$$

$$B = E + \sigma\tau A. \tag{9}$$

El operador A fue estudiado en el cap. VI. Es autoconjugado y definido positivo en el espacio $H = \Omega$ de dimensión

$$(N_1 - 1)(N_2 - 1), \quad A = A^*, \quad \delta_0 E \leqslant A \leqslant \Delta_0 E,$$

donde

$$\delta_0 = \frac{4}{h_1^2}\operatorname{sen}^2\frac{\pi h_1}{2l_1} + \frac{4}{h_2^2}\operatorname{sen}^2\frac{\pi h_2}{2l_2};$$

$$\Delta_0 = \frac{4}{h_1^2}\cos^2\frac{\pi h_1}{2l_1} + \frac{4}{h_2^2}\cos^2\frac{\pi h_2}{2l_2}, \quad \Delta_0 = \|A\|. \tag{10}$$

En virtud de la teoría general (véase el cap. V), el esquema (9) es estable en H_A cuando

$$\delta \geqslant \delta_0, \qquad \delta_0 = \frac{1}{2} - \frac{1}{\tau\|A\|}. \tag{11}$$

En particular, para un esquema explícito tenemos la condición

$$\tau \leqslant \frac{2}{\Delta_0}, \quad \text{o bien} \quad \tau < \left(\frac{2}{h_1^2} + \frac{2}{h_2^2}\right)^{-1}. \tag{12}$$

En la red cuadrada ($h_1 = h_2 = h$) la condición de estabilidad del esquema explícito tiene por expresión

$$\tau < h^2/4$$

(compárese con las condiciones $\tau < h^2/2$ para el problema unidimensional). De (11) se ve que los esquemas con

$$\sigma \geqslant 1/2,$$

incluidos el esquema implícito ($\sigma = 1$) y el simétrico ($\sigma = 1/2$), son incondicionalmente estables. El esquema ex-

plícito ($\sigma = 0$) puede escribirse en la forma

$$y^{j+1}_{i_1 i_2} = (1 - 2(\gamma_1 + \gamma_2)) y^j_{i_1 i_2} + \gamma_1 (y^j_{i_1 - 1,\, i_2} + y^j_{i_1 + 1,\, i_2}) + \\ + \gamma_2 (y^j_{i_1,\, i_2 - 1} + y^j_{i_1,\, i_2 + 1}) + \tau \varphi^j_{i_1 i_2}. \quad (13)$$

La suma de los coeficientes de y en el segundo miembro de (12) es igual a uno. Si todos los coeficientes son no negativos, es decir, si se cumple la condición $\gamma_1 + \gamma_2 \leqslant 1/2$, $\gamma_1 = \tau/h_1^2$, $\gamma_2 = \tau/h_2^2$, equivalente a la condición de estabilidad (12), entonces de (13) proviene una desigualdad

$$\| y^{j+1} \|_C \leqslant \| y^j \|_C + \tau \| \varphi^j \|_C.$$

Al sumar según $k = 0, 1, \ldots, j - 1$, obtenemos la estimación (*compárese con el* § 1)

$$\| y^j \|_C \leqslant \| y^0 \|_C + \sum_{k=0}^{j-1} \tau \| \varphi^k \|_C, \quad (14)$$

que queda en vigor para cualesquiera pasos de la red si el esquema es implícito puro ($\sigma = 1$). En todos los demás casos la estimación (14) tiene lugar para $\sigma \geqslant 1 - 1/\tau\Delta_0$. Para demostrar la convergencia se debe, como siempre, investigar el residuo

$$\psi = \Lambda (\sigma \hat{u} + (1 - \sigma) u) + \varphi - u_t.$$

Teniendo presente que $\Lambda u = Lu + O(|h|^2)$, $|h|^2 = h_1^2 + h_2^2$, encontramos, por analogía con el caso unidimensional,

$$\psi = O(|h|^2 + \tau^2) + \left(\tau - \frac{1}{2}\right) O(\tau).$$

Para el error $z = y - u$ tenemos un problema

$$B = \frac{z^{j+1} - z^j}{\tau} + Az^j = \psi^j, \quad j = 0, 1, \ldots, \quad z^0 = z(0) = 0.$$

De aquí y de las estimaciones apriorísticas proviene la convergencia en C del esquema (5) con la velocidad $O(\tau + |h|^2)$ para $\sigma \neq 1/2$, y $O(\tau^2 + |h|^2)$ para $\sigma = 1/2$ (analogía completa con el caso unidimensional), siempre que $\sigma \geqslant 1 - \frac{1}{\tau\Delta_0}$.

Para la solución del problema se cumple, en virtud de la estimación obtenida en el cap. V, una desigualdad

$$\|z^{j+1}\|_A \leqslant \sum_{k=0}^{j} \tau \|\psi^k\| \text{ para } \sigma \geqslant \sigma_0 = \frac{1}{2} - \frac{1}{\tau \Delta_0}, \quad \sigma \geqslant 0,$$

donde

$$\|z\|_A^2 = \|z\|_{A_1}^2 + \|z\|_{A_2}^2, \quad A_1 y = -y_{\bar{x}_1 x_1}, \quad A_2 y = -y_{\bar{x}_2 x_2},$$

$$\|z\|_A^2 = \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2-1} h_2 \, (z_{\bar{x}_1}(i_1, i_2))^2 + \sum_{i_1=1}^{N_1-1} \times$$

$$\times \sum_{i_2=1}^{N_2} h_2 \, (z_{\bar{x}_2}(i_1, i_2))^2.$$

De aquí proviene la estabilidad incondicional de la convergencia del esquema (5) en H_A con la velocidad $O(\tau + |h|^2)$ para $\sigma \neq 1/2$, $\sigma \geqslant 1/2$, y $O(\tau^2 + |h|^2)$ para $\sigma = 1/2$.

La investigación realizada más arriba debe ser complementada con las condiciones de estabilidad asintótica. Por cuanto las condiciones citadas $\tau \leqslant \tau_0$ se han obtenido para un esquema operacional en diferencias con pesos y operador arbitrario

$$A = A^* > 0, \qquad \delta_0 E \leqslant A \leqslant \Delta_0 E,$$

pueden ser aplicadas, pues, también para nuestro esquema (5). Haciendo uso de las expresiones (10) para δ_0 y Δ_0, obtenemos las condiciones de estabilidad asintótica $\tau \leqslant \tau_0^{(1)}$, $\tau_0^{(1)} = 2\left(\frac{4}{h_1^2} + \frac{4}{h_2^2}\right)^{-1}$ para un esquema explícito ($\sigma = 0$) $\tau \leqslant \tau_0^{(2)}$, $\tau_0^{(2)} = \frac{2}{\sqrt{\delta_0 \Delta_0}}$, δ_0, Δ_0 de (10), para un esquema simétrico ($\sigma = 1/2$).

En particular, para $h_1 = h_2 = h$, $l_1 = l_2 = l$ tenemos

$$\delta_0 = \frac{8}{h^2} \operatorname{sen}^2 \frac{\pi h}{2l}, \quad \Delta_0 = \frac{8}{h^2} \cos^2 \frac{\pi h}{2l}, \quad \tau_0^{(1)} = \frac{h^2}{4},$$

$$\tau_0^{(2)} = \frac{h^2}{2} \left(\operatorname{sen} \frac{\pi h}{l}\right)^{-1} \approx \frac{hl}{2\pi}.$$

El valor límite de $\tau_0^{(2)}$ es dos veces menor que para el esquema unidimensional (5) del § 1.

Un esquema implícito puro ($\sigma = 1$) es absolutamente estable asintóticamente.

3. Coeficientes variables. Examinemos el problema (1) suponiendo que L es un operador elíptico de segundo orden, con coeficientes variables, privado de derivadas mixtas:

$$Lu = L_1 u + L_2 u, \quad L_1 u = \frac{\partial}{\partial x_1}\left(k_1(x, t)\frac{\partial u}{\partial x_1}\right),$$

$$L_2 u = \frac{\partial}{\partial x_2}\left(k_2(x, t)\frac{\partial u}{\partial x_2}\right),$$

$$c_1 \leqslant k_\alpha(x, t) \leqslant c_2, \quad (x, t) \in \bar{Q}_T = G \times (0, T).$$

Aproximemos cada uno de los operadores L_1 y L_2 mediante un operador de diferencias tripuntual:

$$L_1 \sim \Lambda_1, \qquad L_2 \sim \Lambda_2,$$

$$\Lambda_1 v = (a_1 v_{\bar{x}_1})_{x_1}, \qquad \Lambda_2 v = (a_2 v_{\bar{x}_2})_{x_2},$$

donde $a_1 = a_1(i_1 h_1, i_2 h_2, t)$, $a_2 = a_2(i_1 h_1, i_2 h_2, t)$ son ciertas funcionales de los valores k_1 y k_2, respectivamente; en un caso más simple $a_1 = k_1((i_1 - 1/2)h_1, i_2 h_2, t)$, $a_2 = k_2(i_1 h_1; (i_2 - 1/2)h_2, t)$, lo que asegura el segundo orden de aproximación: $\Lambda_\alpha u - L_\alpha u = O(h_\alpha^2)$, $\alpha = 1, 2$. Al operador L se le pone en correspondencia el operador de diferencias Λ:

$$\Lambda v = \Lambda_1 v + \Lambda_2 v = (a_1 v_{\bar{x}_1})_{x_1} + (a_2 v_{\bar{x}_2})_{x_2}. \tag{15}$$

Escribamos $\Lambda_1 v$ y $\Lambda_2 v$ en forma de índices

$$\Lambda_1 v = \frac{1}{h_1}\left[a_1((i_1+1)h_1, i_2 h_2; t)\frac{v_{i_1+1, i_2} - v_{i_1 i_2}}{h_1} - a_1(i_1 h_1, i_2 h_2; t)\frac{v_{i_1, i_2} - v_{i_1-1, i_2}}{h_1}\right],$$

$$\Lambda_2 v = \frac{1}{h_2}\left[a_2(i_1 h_1, (i_2+1)h_2; t)\frac{v_{i_2+1} - v_{i_1 i_2}}{h_2} - a_2(i_1 h_1, i_2 h_2; t)\frac{v_{i_1 i_2} - v_{i_1, i_2-1}}{h_2}\right].$$

El esquema de diferencias con pesos tiene la misma expresión (5) que en el p. 1. Se escoge el mismo espacio reticular $H = \Omega$ con producto escalar (7) y se introduce el operador A:

$$Ay = -\Lambda \overset{\circ}{y} = -(a_1 \overset{\circ}{y}_{\bar{x}_1})_{x_1} - (a_2 \overset{\circ}{y}_{\bar{x}_2})_{x_2}.$$

Teniendo en cuenta que para el caso unidimensional del operador A: $Ay = -(a\overset{\circ}{y}_{\bar{x}})_x$

$$c_1 (\mathring{A}y, y) \leqslant (Ay, y) \leqslant c_2 (\mathring{A}y, y), \quad \mathring{A}y = -\overset{\circ}{y}_{\bar{x}x},$$
$$0 < c_1 \leqslant a \leqslant c_2,$$

no es difícil convencerse de que las desigualdades semejantes se verifican también para el operador bidimensional (15):

$$c_1 \mathring{A} \leqslant A \leqslant c_2 \mathring{A}, \quad \mathring{A}y = -\overset{\circ}{y}_{\bar{x}_1 x_1} - \overset{\circ}{y}_{\bar{x}_2 x_2}.$$

De aquí se ve que $\delta E \leqslant A \leqslant \Delta E$, $\delta = c_1\delta_0$, $\Delta = c_2\Delta_0$, donde δ_0 y Δ_0 se determinan según las fórmulas (10). Para hallar $\hat{y} = y^{j+1}$ sobre la capa nueva obtenemos el problema (6), en el que Λ se determina de (15). En el caso de un esquema explícito $\hat{y}$ se determina en cada nodo $x \in \omega_h$ por la fórmula

$$\hat{y} = y + (1 - \sigma)\,\tau\Lambda y + \tau\varphi.$$

Para los esquemas implícitos ($\sigma \neq 0$) se debe resolver una ecuación en diferencias pentapuntual con coeficientes variables. Aquí se emplean los métodos iterativos, de los cuales el método alternado triangular resulta ser más económico (véase el cap. V, § 5); el número de iteraciones para dicho método se expresa por la magnitud $O\left(\frac{1}{\sqrt{h}} \ln \frac{1}{\varepsilon}\right)$, siempre que ($\tau \approx O(h)$). La descripción del método alternado triangular para las ecuaciones en diferencias con coeficientes variables se ha dado en el cap. VI; si se aplica a la ecuación (6) con el operador Λ del tipo (15), debe ser un tanto modificado.

§ 3. Esquemas económicos

1. Método de direcciones variables. Al comparar los esquemas explícitos e implícitos (5), detendrémonos en dos características: el volumen de los cálculos para hallar y^{j+1} y las restricciones que se imponen sobre el paso τ.

ESQUEMA EXPLÍCITO: para determinar y^{j+1} sobre la red ω_h hay que realizar un número de operaciones proporcional al número de nodos, es decir, el número de operaciones correspondientes a un nodo no depende de la red ω_h. Sin embargo, el paso τ está rígidamente limitado superiormente por la condición $\tau \leqslant \tau_0(h)$: $\tau \leqslant h^2/4$ con $h_1 = h_2 = h$ para el esquema (13).

ESQUEMA IMPLÍCITO ($\sigma \geqslant 1/2$); para determinar y^{j+1} se debe resolver un sistema de $(N_1 - 1)(N_2 - 1)$ ecuaciones en diferencias pentapuntuales; con este objeto, por lo menos en el caso de coeficientes variables, se requiere un número de operaciones (correspondientes a un nodo de la red ω_h) que crece cuando $|h| \to 0$.

Surge un problema: construir los esquemas en que se combinen las mejores cualidades de esquemas explícitos e implícitos, es decir, los esquemas a construir deben ser incondicionalmente estables con un número de operaciones en cada capa proporcional al número de nodos de la red ω_h. Los esquemas de este género suelen llamarse *económicos*. Por supuesto, hemos de hacer una especificación de que los esquemas incondicionalmente estables en el sentido habitual han de ser asintóticamente estables, lo que conduce a una restricción del paso mucho más débil (por ejemplo, $\tau \leqslant lh/(2\pi)$ para $\sigma = 1/2$, $h_1 = h_2 = h$, $l_1 = l_2 = l$) que la condición de estabilidad ($\tau \leqslant h^2/4$) para el esquema explícito. Además, la condición $\tau = O(h)$ es natural para el esquema $O(\tau^2 + |h|^2)$.

Los primeros esquemas económicos han aparecido en los años 1955—1956 y se han denominado *métodos de direcciones variables*. La idea algorítmica principal (en la que se radica la economía) consiste en que para pasar de una capa t_j a otra capa t_{j+1} se deben resolver, por el método de factorización, las ecuaciones en diferencias tripuntuales, primero a lo largo de las filas y, a continuación, a lo largo de las columnas de la red ω_h.

Demos a conocer las fórmulas del método de direcciones variables (*esquema* longitudinal transversal *de Pismann-Recford*) para el problema (1) con un operador L: $Lu = L_1u + L_2u$, donde L_α es uno de los operadores:

$$L_\alpha u = \frac{\partial^2 u}{\partial x_\alpha^2} \text{ o bien } L_\alpha u = \frac{\partial}{\partial x_\alpha}\left(k_\alpha(x, t)\frac{\partial u}{\partial x_\alpha}\right), \quad \alpha = 1, 2.$$

Sean Λ_1, Λ_2 los operadores tripuntuales correspondientes y supongamos que $\Lambda = \Lambda_1 + \Lambda_2$. Introduciendo un valor intermedio de $\bar{y} = y^{j+1/2}$, enunciamos el esquema de diferencias de direcciones variables:

$$\frac{y^{j+1/2} - y^j}{\tau/2} = \Lambda_1 y^{j+1/2} + \Lambda_2 y^j + \varphi^j, \quad x \in \omega_h,$$

$$y^{j+1/2} = \bar{\mu} \text{ para } i_1 = 0, N_1, \qquad (1)$$

$$\frac{y^{j+1} - y^{j+1/2}}{\tau/2} = \Lambda_1 y^{j+1/2} + \Lambda_2 y^{j+1} + \varphi^j, \quad x \in \omega_h,$$

$$y^{j+1} = \mu^{j+1} \text{ para } i_2 = 0, N_2, \; y^0 = u_0(x), \; x \in \bar{\omega}_h, \qquad (2)$$

donde $\bar{\mu}$ es el valor intermedio de la función $\mu(x, t)$ igual a

$$\bar{\mu} = \frac{\mu^j + \mu^{j+1}}{2} - \frac{\tau}{4}\Lambda_2(\mu^{j+1} - \mu^j).$$

Para hallar $y^{j+1/2}$ e y^{j+1} tenemos problemas de contorno en diferencias

$$\begin{gathered}
{}^1/_2\tau\Lambda_1 y^{j+1/2} - y^{j+1/2} = -F^j, \\
F^j = y^j + 1/2\tau(\Lambda_2 y^j + \varphi^j), \quad x \in \omega_h, \\
y^{j+1/2} = \bar{\mu}, \quad i_1 = 0, N_1, \\
{}^1/_2\tau\Lambda_2 y^{j+1} - y^{j+1} = -F^{j+1/2}, \\
F^{j+1/2} = y^{j+1/2} + {}^1/_2\tau(\Lambda_1 y^{j+1/2} + \varphi^j), \quad x \in \omega_h, \\
y^{j+1} = \mu^{j+1}, \quad i_2 = 0, N_2.
\end{gathered} \qquad (3)$$

El primer problema se resuelve mediante la factorización por las filas ($i_2 = 1, 2, \ldots, N_2 - 1$); el segundo, mediante la factorización por las columnas ($i_1 = 1, 2, \ldots, \ldots, N_1 - 1$). El número de operaciones correspondientes a un nodo es finito y no depende de la red.

El esquema (3) es estable tanto respecto de los datos iniciales, como respecto del segundo miembro para cualesquiera τ y $|h|$ y tiene la exactitud $O(\tau^2 + |h|^2)$. De esto podemos convencernos eliminando $y^{j+1/2}$ y reduciendo el esquema (1), (2) a un esquema equivalente de dos capas con el operador factorizado B_j:

$$B\frac{y^{j+1}-y^j}{\tau} + Ay^j = \Phi^j, \quad j=0,\ 1,\ \ldots,\ y^0 = u_0 \in H, \qquad (4)$$

$$B = \left(E + \frac{\tau}{2} A_1\right)\left(E + \frac{\tau}{2} A_2\right), \quad A_\alpha y = -\Lambda_\alpha \overset{\circ}{y} = -\overset{\circ}{y}_{\overline{x}_\alpha x_\alpha},$$

$$\alpha = 1,\ 2,$$

donde $H = \Omega$ es el espacio de funciones reticulares definidas en los nodos interiores de la red ω_h.

Es evidente que $A_\alpha = A_\alpha^* > 0$, $\alpha = 1, 2$, $A_1 A_2 = A_2 A_1$. Por eso $B = E + \tau A/2 + \tau^2 A_1 A_2/2 \geqslant E + \tau A/2 > \tau A/2$, y el esquema es estable.

2. Esquemas factorizados. El operador B, representado como un producto de varios operadores $B = B_1 B_2 \ldots B_p$ se llamará *factorizado*, y el esquema correspondiente

$$B\frac{y^{j+1}-y^j}{\tau} + Ay^j = \varphi^j, \quad j=0,\ 1,\ \ldots,\ y^0 = y(0), \qquad (5)$$

esquema factorizado.

Si para la resolución del problema

$$B_\alpha v = F_\alpha, \qquad \alpha = 1,\ 2,$$

con el segundo miembro prefijado F_α se requiere $O(N_1 N_2)$ número de operaciones, entonces para determinar y^{j+1}, partiendo de y^j conocido, también son necesarias $O(N_1 N_2)$ operaciones (el operador B es «económico)». Por cuanto

$$By^{j+1} = B_1 B_2 y^{j+1} = F^j,$$

el algoritmo se reducirá a la resolución sucesiva de las ecuaciones

$$B_1 y^{j+1/2} = F^j, \qquad B_2 y^{j+1} = y^{j+1/2}.$$

Apoyándose en la teoría de estabilidad de los esquemas de dos capas, no es difícil, partiendo del esquema con pesos, construir un esquema factorizado económico (por el método de regularización).

Así pues, supongamos que

$$A = A_1 + A_2, \quad B = E + \sigma\tau A = E + \sigma\tau (A_1 + A_2),$$
$$A_1 = A_1^*, \qquad A_2 = A_2^*.$$

En este caso el esquema (9) del § 2 será estable para $\sigma \geqslant \sigma_0 = \frac{1}{2} - \frac{1}{\tau \|A\|}$. Sustituyamos en (9) el operador B por un operador factorizado

$$\widetilde{B} = (E + \sigma\tau A_1)\ (E + \sigma\tau A_2),$$

que se diferencia de B en el término $\sigma^2\tau^2 A_1 A_2$,

$$\widetilde{B} = B + \sigma^2\tau^2 A_1 A_2.$$

Como resultado obtenemos un esquema factorizado

$$\widetilde{B}\,\frac{y^{j+1} - y^j}{\tau} + Ay^j = \Phi^j, \quad j = 0,\ 1,\ \ldots,\ y^0 = u_0 \in H, \qquad (6)$$

del mismo orden de aproximación $O\,((\sigma - 1/2)\,\tau + \tau^2)$ que tiene el esquema de partida con pesos. Por cuanto el esquema de partida con pesos es estable ($\sigma \geqslant \sigma_0$), el esquema factorizado (6) será estable en virtud de la condición

$$\widetilde{B} > B > \tau A/2,$$

que se verifica, siempre que A_1 y A_2 son permutables y $A_\alpha^* = A_\alpha > 0$, $\alpha = 1, 2$.

Para hallar y^{j+1} obtenemos una ecuación $\widetilde{B}y^{j+1} = F^j$, o bien

$$(E + \sigma\tau A_1)\ (E + \sigma\tau A_2)\ y^{j+1} = F^j,$$

$$F^j = \widetilde{B}y^j + \tau\,(\Phi^j - Ay^j),$$

la cual se resuelve sucesivamente:

$$(E + \sigma\tau A_1)\ \bar{y} = F^j, \qquad (E + \sigma\tau A_2)\ y^{j+1} = \bar{y}$$

(con las condiciones de contorno correspondientes). El algoritmo que sigue es más económico (a cuenta del cálculo del segundo miembro F^j):

$$(E + \sigma\tau A_1)\ w^{j+1/2} = F^j = \Phi^j - Ay^j,$$
$$(E + \sigma\tau A_2)\ w^{j+1} = w^{j+1/2}, \quad y^{j+1} = y^j = \tau w^{j+1}. \qquad (7)$$

Sin embargo, en este caso deben guardarse no uno, sino dos vectores ($w^{j+1/2}$ ó w^{j+1} e y^j). Cuando $\sigma = 1$, de (7) se deduce el segundo esquema de direcciones variables (*esquema de Duglass—Recford*)

$$\frac{y^{j+1/2}-y^j}{\tau}+A_1y^{j+1/2}+A_2y^j=\Phi^j,$$

$$(E+\tau A_2)\frac{y^{j+1}-y^{j+1/2}}{\tau}=\frac{y^{j+1/2}-y^j}{\tau}.$$

3. Método de aproximación sumaria. Con el fin de obtener esquemas económicos para la amplia clase de problemas (ecuaciones con coeficientes variables, dominios de forma compleja, etc.) hemos de cambiar el concepto de esquema de diferencias.

Dejamos a parte el concepto habitual de aproximación que se ha examinado anteriormente y lo cambiamos por un concepto más débil de *aproximación sumaria*. Aclaremos esto. Supongamos que el paso de una capa j a la otra $j+1$ se efectúa en varias etapas, en cada una de las cuales se utiliza el esquema corriente de dos capas que no aproxima la ecuación de partida y, no obstante, la suma de residuos para todo esquema intermedio

$$\psi=\sum_{\alpha=1}^{p}\psi\alpha \tag{8}$$

tiende a cero, cuando tiende a cero el paso τ según la variable t.

La idea del método de aproximación sumaria puede explicarse con un ejemplo del problema de Cauchy para la ecuación diferencial ordinaria

$$\frac{du}{dt}+au=f(t), \qquad t>0, \qquad u(0)=u_0, \tag{9}$$

donde $a>0$ es un número. Supongamos que

$$a=a_1+a_2, \quad a_1>0, \quad a_2>0, \quad f(t)=f_1(t)+f_2(t) \tag{10}$$

Es evidente que una representación de tal índole es siempre posible.

Introduzcamos una red $\omega_\tau = \{t_j = j\tau,\ j = 0, 1, \ldots\}$ y en cada paso (t_j, t_{j+1}) resolvemos sucesivamente (en lugar de (9)) dos ecuaciones

$$\frac{1}{2}\frac{dv_{(1)}}{dt} + a_1 v_{(1)} = f_1(t), \quad t_j \leqslant t \leqslant t_{j+1/2} = t_j + \frac{\tau}{2},$$

$$\frac{1}{2}\frac{dv_{(2)}}{dt} + a_2 v_{(2)} = f_2(t), \quad t_{j+1/2} \leqslant t \leqslant t_{j+1} \qquad (11)$$

con los siguientes datos iniciales

$$v_{(1)}(t_j) = v(t_j), \quad v_{(2)}(t_{j+1/2}) = v_{(1)}(t_{j+1/2}),$$
$$j = 0, 1, \ldots, \quad v_{(1)}(0) = u_0. \qquad (12)$$

Como solución del problema (11)— 12) interviene la función

$$v(t) = v_{(2)}(t). \qquad (13)$$

Cada una de las ecuaciones (11) se aproximará mediante un esquema de diferencias de dos capas con paso $\tau/2$. Por ejemplo, tomemos un esquema implícito

$$\frac{y^{j+1/2} - y^j}{\tau} + a_1 y^{j+1/2} = f_1^j, \qquad (14)$$

$$\frac{y^{j+1} - y^{j+1/2}}{\tau} + a_2 y^{j+1} = f_2^j$$

Calculemos los residuos ψ_1 y ψ_2 para los esquemas (11). Sutituyamos en (11)

$$y^j = z^j + u^j, \quad y^{j+1/2} = z^{j+1/2} + u^{j+1/2}, \quad y^{j+1} = z^{j+1} + u^{j+1},$$

$$\frac{z^{j+1/2} - z^j}{\tau} + a_1 z^{j+1/2} = -\psi_1^j,$$

$$\frac{z^{j+1} - z^{j+1/2}}{\tau} + a_2 z^{j+1} = -\psi_2^j, \quad j = 0, 1, \ldots,$$

$$z^0 = 0, \quad \psi_1^j = \frac{u^{j+1/2} - u^j}{\tau} + a_1 u^{j+1/2} - f_1^j,$$

$$\psi_2^j = \frac{u^{j+1} - u^{j+1/2}}{\tau} + a_2 u^{j+1} - f_2^j,$$

Introduciendo aquí

$$u^{j+1} = (u + \overset{\circ}{\tau u}/2)^{j+1/2} + O(\tau^2), \quad u^j = (u - \overset{\circ}{\tau u}/2)^{j+1/2} + O(\tau^2)$$

obtenemos

$$\psi_1^j = (\mathring{u}/2 + a_1 u - f_1)^{j+1/2} + O(\tau), \qquad \psi_2^j = (\mathring{u}/2 + a_2 u - f_2)^{j+1/2} + O(\tau). \tag{15}$$

De aquí se ve que $\psi_1^j = O(1)$, $\psi_2^j = O(1)$, sin embargo

$$\psi_1^j + \psi_2^j = O(\tau) \to 0 \text{ cuando } \tau \to 0. \tag{16}$$

Todos los razonamientos aducidos más arriba, a partir de (10), (11), (14), quedan en vigor si a_1 y a_2 representan las matrices o los operadores, y u, f, y son los vectores.

De este modo, el esquema (11), (12) aproxima el problema (9) en el sentido sumario (16) (tales esquemas se denominan *aditivos*).

Para demostrar la convergencia del esquema (11), (12) es menester obtener la estimación para el error $z^{j+1} = y^{j+1} - u^{j+1}$ en que se toma en consideración la propiedad (16) de la aproximación sumaria.

Pongamos

$$\psi_\alpha = \mathring{\psi}_\alpha + \psi_\alpha^*,$$

$$\mathring{\psi}_\alpha = (\mathring{u}/2 + a_\alpha u - f_\alpha)^{j+1/2}, \quad \psi_\alpha^* = O(\tau), \quad \alpha = 1, 2,$$

$$z^{j+1/2} = \eta_{j+1/2} + \xi_{j+1/2}, \quad z^{j+1} = \eta_{j+1} + \xi_{j+1},$$

donde η_{j+1}, ξ_{j+1} son las soluciones de los problemas

$$\eta_{j+1/2} = \eta_j + \tau\mathring{\psi}_1, \quad \eta_{j+1} = \eta_{j+1/2} + \tau\mathring{\psi}_2, \qquad j = 0, 1, \ldots, \quad \eta_0 = 0, \tag{17}$$

$$(1 + a_1\tau)\,\xi_{j+1/2} = \xi_j + \tau\widetilde{\psi}_1, \quad (1 + a_2\tau)\,\xi_{j+1} = \xi_{j+1/2} + \tau\widetilde{\varphi}_2, \qquad j = 0, 1, \ldots, \tag{18}$$

$$\xi_0 = 0,$$

$$\widetilde{\psi}_1^j = \psi_1^{*j} - a_1\tau\eta_{j+1/2}, \quad \widetilde{\psi}_2^j = \psi_2^{*j} - a_2\tau\eta_{j+1}. \tag{19}$$

De aquí encontramos $\eta_{j+1} = \eta_j + \tau\,(\mathring{\psi}_1^j + \mathring{\psi}_2^j) = \eta_j = \ldots = \eta_0 = 0$, es decir, $\eta_j = 0$ para cualquier $j = 0, 1, \ldots$, y $z^j = \xi_j$.

$$\eta_{j+1/2} = \tau\mathring{\psi}_1 = O(\tau), \qquad \widetilde{\psi}_\alpha = O(\tau). \tag{20}$$

De (16) obtenemos

$$|\xi_{j+1/2}| \leqslant |\xi_j| + \tau |\tilde{\psi}_1^j|,$$

$$|\xi_{j+1}| \leqslant |\xi_{j+1/2}| + \tau |\tilde{\psi}_2^j| \leqslant |\xi_j| + \tau (|\tilde{\psi}_1^j| + |\tilde{\psi}_2^j|),$$

de modo que resulta lícita la estimación

$$|z^{j+1}| \leqslant \sum_{k=1}^{j} \tau (|\tilde{\psi}_1^k| + |\tilde{\psi}_2^k|), \qquad (21)$$

de la cual proviene precisamente (en virtud de (17)) la convergencia del sistema aditivo (14) con la velocidad $O(\tau)$.

En lugar de (11) podemos tomar otro sistema de ecuaciones:

$$\frac{dv_{(1)}}{dt} + a_1 v_{(1)} = f_1(t), \quad t_j \leqslant t \leqslant t_{j+1}, \quad v_{(1)}(t_j) = v(t_j),$$

$$\frac{dv_{(2)}}{dt} + a_2 v_{(2)} = f_2(t), \quad t_j \leqslant t \leqslant t_{j+1}, \quad v_{(2)}(t_j) = v_{(1)}(t_{j+1}),$$

$$j = 0, 1, \ldots, \quad v_{(1)}(0) = u_0.$$

Como solución de este problema interviene la función

$$v(t) = v_{(2)}(t). \qquad (23)$$

A diferencia de (11), aquí ambas ecuaciones se interpretan en todo el segmento $t_j \leqslant t \leqslant t_{j+1}$, por lo cual la aproximación de dichas ecuaciones se realiza con el paso τ (y no con el paso $\tau/2$, como en el caso (11)) y da los mismos esquemas (14). Los dos métodos de reducción del problema (9) al sistema de problemas (11) ó (22) emplean una misma propiedad

$$a = a_1 + a_2 \qquad (24)$$

y la condición $f = f_1 + f_2$, la cual siempre puede satisfacerse.

Veamos, como un ejemplo, la ecuación de conductibilidad térmica

$$\frac{\partial u}{\partial t} = Lu + f(x, t), \qquad x = (x_1, x_2), \qquad (25)$$

$$Lu = \Delta u = L_1 u + L_2 u, \quad l_\alpha u = \frac{\partial^2 u}{\partial x_\alpha^2}, \quad \alpha = 1, 2,$$

L_1 y L_2 son los operadores «unidimensionales». La resolución de la ecuación

$$\frac{\partial v_{(\alpha)}}{\partial t} = L_\alpha v_{(\alpha)} + f_\alpha, \tag{26}$$

será, evidentemente, un problema más sencillo que la resolución de la ecuación (25). Las condiciones $L = L_1 + L_2$, $f = f_1 + f_2$ garantizan la aproximación sumaria para un esquema que se obtiene como resultado de la aproximación corriente, por ejemplo, con ayuda de un esquema de dos capas con pesos de cada una de las ecuaciones del sistema

$$\frac{dv_{(1)}}{dt} = L_1 v_{(1)} + f_1, \quad t_j \leqslant t \leqslant t_{j+1}, \quad v^j_{(1)} = v^j,$$

$$\frac{dv_{(2)}}{dt} = L_2 v_{(2)} + f_2, \quad t_j \leqslant t \leqslant t_{j+1}, \quad v^j_{(2)} = v^{j+1}_{(1)}, \quad v^{j+1} = v^{j+1}_{(2)}.$$

De resultas obtenemos un esquema aditivo, un esquema unidimensional local o bien un esquema de fisión

$$\begin{aligned}
&\frac{y^{j+1/2} - y^j}{\tau} = \Lambda_1 (\sigma_1 y^{j+1/2} + (1 - \sigma_1) y^j) + \varphi^j_1, \quad x \in \omega_h, \\
&\frac{y^{j+1} - y^{j+1/2}}{\tau} = \Lambda_2 (\sigma_2 y^{j+1} + (1 - \sigma_2) y^{j+1/2}) + \varphi^j_2, \\
&\qquad\qquad x \in \omega_h, \ j = 0, 1, \ \ldots, \\
&y^0 = u_0(x), \qquad x \in \omega_h, \\
&y^{j+1/2}|_{\gamma_h} = \mu^{j+1/2}, \qquad y^{j+1}|_{\gamma_h} = \mu^{j+1}.
\end{aligned} \tag{27}$$

Aquí $\Lambda_1 y = y_{\bar{x}_1 x_1}$, $\Lambda_2 y = y_{\bar{x}_2 x_2}$. Los parámetros σ_1 y σ_2 se determinan partiendo de las condiciones de estabilidad y aproximación. Por ejemplo, cuando $\sigma_1 = \sigma_2 = 1$, obtenemos un esquema con adelanto

$$\begin{aligned}
&\frac{y^{j+1/2} - y^j}{\tau} = \Lambda_1 y^{j+1/2} + \varphi^j_1, \\
&\frac{y^{j+1} - y^{j+1/2}}{\tau} = \Lambda_2 y^{j+1} + \varphi^j_2, \ j = 0, 1, \ \ldots
\end{aligned}$$

Al sustituir aquí $y^j = z^j + u^j$, $y^{j+1/2} = z^{j+1/2} + (u^j + u^{j+1})/2$, $y^{j+1} = z^{j+1} + u^{j+1}$, obtenemos para el error z

las ecuaciones

$$\frac{z^{j+1/2}-z^j}{\tau}=\Lambda_1 z^{j+1/2}+\psi_1^j,$$

$$\frac{z^{j+1}-z^{j+1/2}}{\tau}=\Lambda_2 z^{j+1}+\psi_2^j,$$

donde u es la solución del problema de partida (25), ψ_1 y ψ_2 son los residuos

$$\psi_1^j=\Lambda_1\frac{u+\hat{u}}{2}-\frac{1}{2}\frac{\hat{u}-u}{\tau}+\varphi_1,\quad \psi_2^j=\Lambda_2\hat{u}-\frac{1}{2}\frac{\hat{u}-u}{\tau}+\varphi_2,$$

$$\hat{u}=u^{j+1},\qquad u=u^j.$$

De aquí se ve que $\psi_1 = O(1)$, $\psi_2 = O(1)$, es decir, cada una de las ecuaciones (27) no aproxima, tomada separadamente, la ecuación (25). Tomemos la suma de residuos

$$\psi=\psi_1+\psi_2=\Lambda_1\frac{u+\hat{u}}{2}+\Lambda_2\hat{u}-\frac{\hat{u}-u}{\tau}+\varphi_1+\varphi_2=$$

$$=(L_1+L_2)\,\bar{u}-\frac{\partial\bar{u}}{\partial t}+\varphi_1+\varphi_2+O(\tau+|h|^2),$$

donde $\bar{u}=u^{j+1/2}$. Tomando en consideración la ecuación (25) para $t=t_{j+1/2}$, obtendremos

$$\psi=\varphi_1+\varphi_2-f^{j+1/2}+O(\tau+|h|^2)=O(\tau+|h|^2)$$

$$|h|^2=h_1^2+h_2^2,$$

siempre que

$$\varphi_1+\varphi_2=f^{j+1/2}+O(\tau^2).$$

Esto puede ser conseguido suponiendo, por ejemplo,

$$\varphi_1=0,\ \varphi_2=f^{j+1/2}\ \text{ o bien }\ \varphi_1=\varphi_2=f^j/2.$$

Se puede mostrar que el esquema (27) converge uniformemente con la velocidad

$O(\tau+|h|^2)$, es decir, $\|y^{j+1}-u^{j+1}\|_C=O(\tau+|h|^2)$.

De los ejemplos aducidos se ve que el método de aproximación sumaria permite realizar la partición de los problemas complejos en una sucesión de problemas más sencillos y simplificar considerablemente la resolución de los problemas multidimensionales de la física matemática.

Anexo

Algoritmo de marcha y método de reducción para resolver sistemas de ecuaciones lineales con matriz tridiagonal

En varias aplicaciones se encuentran problemas que conducen a la resolución de los sistemas de ecuaciones algebraicas lineales especiales (con una matriz enrarecida que cuenta con muchos elementos nulos) de orden superior. Los sistemas de tal género surgen al realizar una aproximación de diferencias de las ecuaciones elípticas o bien al utilizar esquemas implícitos para la ecuación de conductibilidad térmica.

Al aproximar en el cap. IV una ecuación diferencial corriente de segundo orden en el molde tripuntual, hemos obtenido una ecuación en diferencias de segundo orden la cual representa un sistema de ecuaciones algebraicas lineales de $(N-1)$-ésimo orden ($N-1$ es el número de nodos interiores) con la matriz tridiagonal. Con el objeto de resolver dicho sistema se ha construido en el § 3, cap. I un método cuya realización requiere $O(N)$ operaciones aritméticas.

En el cap. VI hemos obtenido, aproximando la ecuación de Poisson bidimensional en un molde pentapuntual, un esquema de diferencias, a la cual corresponde el sistema de ecuaciones algebraicas lineales con matriz pentadiagonal de orden $N=(N_1-1)(N_2-1)$, donde N_1-1, N_2-1 es el número de nodos interiores según cada dirección. Al partir el vector de incógnitas en bloques, cada uno de los cuales contenga N_1-1 elementos, obtendremos una inscripción del sistema con matriz tridiagonal de bloques, con la particularidad de que el número de bloques en la matriz citada es igual a N_2-1. Para tal sistema hemos estudiado en el § 2, cap. VI el método de separación de variables con la estimación $O(N \log N)$ para el número de operaciones. Cuando los sistemas semejantes se resuelven varias veces se hace muy importante que los algoritmos computacionales sean económicos.

Más abajo construiremos un método directo para resolver sistemas especiales con la matriz triagonal, el cual exige sólo $O(N)$ operaciones tanto en el caso en que los elementos de la matriz son escalares, como en el caso de la matriz de bloques.

1. Estudiemos al principio un caso en que los elementos de la matriz son escalares. Escribamos el sistema con una matriz tridiagonal en forma de un problema de diferencias tripuntual:

$$-y_{i-1}+Cy_i-y_{i+1}=F_i, \quad 1\leqslant i\leqslant N-1, \quad y_0=0, \quad y_N=0, \quad (1)$$

donde C es un número, y supongamos que $N=2k+1$. Si escribimos la ecuación en diferencias de segundo orden (1) en forma de las rela-

ciones recurrentes

$$y_{i+1} = Cy_i - y_{i-1} - F_i, \quad i \geqslant 1, \quad y_0 = 0, \tag{2}$$

no será difícil notar que todas las incógnitas y_i pueden ser encontradas sucesivamente por la fórmula (2), si calculamos y_1 de tal o cual modo. En este caso, cualquier y_i se expresará linealmente en términos de y_0 e y_1. Todo lo dicho nos permite escribir para cualquier $i \geqslant 1$ una correlación

$$y_{i+1} = \alpha_i y_1 - \beta_{i-1} y_0 - p_i \tag{3}$$

con los coeficientes α_i, β_i, p_i que por ahora quedan indeterminados. Si ponemos

$$\alpha_0 = 1, \quad \beta_{-1} = 0, \quad p_0 = 0, \tag{4}$$

entonces (3) se verifica también para $i = 0$. Así pues, la solución del problema se buscará en la forma (3) para cualquier $i \geqslant 0$.

Anotando (1) en forma de las relaciones recurrentes

$$y_{i-1} = Cy_i - y_{i+1} - F_i, \quad i \leqslant N - 1, \quad y_N = 0 \tag{5}$$

y razonando análogamente, llegamos a que la solución del problema (1) para cualquier $i \leqslant N$ se puede buscar en la forma

$$y_{i-1} = \xi_{N-i} y_{N-1} - \eta_{N-i-1} y_N - q_{N-i}, \tag{6}$$

si ponemos

$$\xi_0 = 1, \quad \eta_{-1} = 0, \quad q_0 = 0. \tag{7}$$

Observemos que si y_{N-1} queda determinado, todos los y_i se pueden calcular sucesivamente según la fórmula (5).

Hallemos y_1 e y_{N-1}. Con este fin determinemos los coeficientes α_i, β_i, ξ_i, η_i, p_i, q_i. Al comparar (2) y (3) para $i = 1$, y (5) y (6), para $i = N - 1$, obtendremos

$$\alpha_1 = \xi_1 = C, \quad \beta_0 = \eta_0 = 1, \quad p_1 = F_1, \quad q_1 = F_{n-1}. \tag{8}$$

Encontremos ahora las fórmulas recurrentes para determinar los coeficientes buscados. Sustituyamos (3), al igual que las expresiones para y_i e y_{i-1} que se desprenden de (3):

$$y_i = \alpha_{i-1} y_1 - \beta_{i-2} y_0 - p_{i-1}, \quad y_{i-1} = \alpha_{i-2} y_1 - \beta_{i-3} y_0 - p_{i-2},$$

en la ecuación (1). Obtendremos

$$-(\alpha_{i-2} - C\alpha_{i-1} + \alpha_i)\, y_1 + (\beta_{i-2} - C\beta_{i-2} + \beta_{i-1})\, y_0 + \\ + p_{i-2} - Cp_{i-1} + p_i = F_i, \quad i \geqslant 2.$$

Para que estas igualdades sean idénticas con i cualquiera es suficiente hacer para $i \geqslant 2$

$$p_i = Cp_{i-1} - p_{i-2} + F_i, \tag{9}$$

$$\alpha_i = C\alpha_{i-1} - \alpha_{i-2}, \quad \beta_{i-1} = C\beta_{i-2} - \beta_{i-3}. \tag{10}$$

Análogamente, haciendo uso de (6) y (1), obtendremos para $i \leqslant N-2$ las relaciones recurrentes

$$q_{N-i} = Cq_{N-i-1} - q_{N-i-2} + F_i,$$

$$\xi_{N-i} = C\xi_{N-i-1} - \xi_{N-i-1}, \quad \eta_{N-i-1} = C\eta_{N-i-2} - \eta_{N-i-3}.$$

Al sustituir aquí $N-i$ por i, tenemos para $i \geqslant 2$ las fórmulas siguientes

$$q_i = Cq_{i-1} - q_{i-2} + F_{N-i}, \tag{11}$$

$$\xi_i = C\xi_{i-1} - \xi_{i-2}, \quad \eta_{i-1} = C\eta_{i-2} - \eta_{i-3}. \tag{12}$$

Así pues, las fórmulas (4), (7)—(12) determinan por completo los coeficientes buscados. Al cotejar (10) y (12) bajo las condiciones (4), (7), (8), llegamos a que $\beta_i = \eta_i = \xi_i = \alpha_i$ para $i \geqslant 0$. De este modo, las fórmulas (3), (6) toman la forma

$$y_{i+1} = \alpha_i y_1 - \alpha_{i-1} y_0 - p_i, \quad i \geqslant 0, \tag{13}$$

$$y_{i-1} = \alpha_{N-i} y_{N-1} - \alpha_{N-i-1} y_N - q_{N-i}, \quad i \leqslant N, \tag{14}$$

donde

$$p_i = Cp_{i-1} - p_{i-2} + F_i, \quad i \geqslant 2, \quad p_0 = 0, \quad p_1 = F_1, \tag{15}$$

$$q_i = Cq_{i-1} - q_{i-2} + F_{N-i}, \quad i \geqslant 2, \quad q_0 = 0, \quad q_1 = F_{N-1} \tag{16}$$

$$\alpha_i = C\alpha_{i-1} - \alpha_{i-2}, \quad i \geqslant 2, \quad \alpha_0 = 1, \quad \alpha_1 = C. \tag{17}$$

Hallemos ahora y_1 e y_{N-1}. Con este fin pongamos en (13) $i = k$, y en (14) $i = k+2$. Teniendo presente que $N = 2k+1$, tendremos

$$y_{k+1} = \alpha_k y_1 - \alpha_{k-1} y_0 - p_k, \qquad y_{k+1} = \alpha_{k-1} y_{N-1} - \alpha_{k-2} y_N - q_{k-1}.$$

Restando la primera igualdad de la segunda, obtendremos una ecuación respecto de y_1 e y_{N-1}:

$$\alpha_{k-1} y_{N-1} - \alpha_k y_1 + \alpha_{k-1} y_0 - \alpha_{k-2} y_N = q_{k-1} - p_k. \tag{18}$$

Obtengamos una ecuación más para y_1 e y_{N-1}, suponiendo $i = k-1$ en (13) y $i = k+1$ en (14), y sustrayendo la segunda ecuación de la primera

$$-\alpha_k y_{N-1} + \alpha_{k-1} y_1 - \alpha_{k-1} y_0 + \alpha_{k-1} y_N = p_{k-1} - q_k. \tag{19}$$

Teniendo presente que $y_0 = y_N = 0$, sumemos y sustrayamos (18) y (19). Obtendremos un esquema equivalente

$$\begin{aligned} (\alpha_{k-1} - \alpha_k)\,(y_{N-1} + y_1) &= q_{k-1} - p_k + p_{k-1} - q_k, \\ (\alpha_{k-1} + \alpha_k)\,(y_{N-1} - y_1) &= q_{k-1} - p_k - p_{k-1} + q_k; \end{aligned} \tag{20}$$

al resolver dicho sistema hallaremos los valores buscados de y_1 e y_{N-1}:

$$\begin{aligned} y_1 &= (\alpha_{k-1}^2 - \alpha_k^2)^{-1} [\alpha_k (q_{k-1} - p_k) + \alpha_{k-1} (p_{k-1} - q_k)], \\ y_{N-1} &= (\alpha_{k-1}^2 - \alpha_k^2)^{-1} [\alpha_{k-1} (q_{k-1} - p_k) + \alpha_k (p_{k-1} - q_k)]. \end{aligned} \tag{21}$$

De este modo, el algoritmo de resolución del problema (1) consiste en calcular por las fórmulas (15)—(17) los coeficientes p_{k-1}, p_k, q_{k-1}, q_k, α_{k-1}, α_k, por las formulas (21) los valores de y_1, y_{N-1}, por la fórmula (2) las incógnitas y_i, $i = 2, 3, \ldots, k$ y por la fórmula (5) para $i = N-2, N-3, \ldots, k+1$ con y_0, y_N dados e y_1, y_{N-1} calculados. El algoritmo descrito recibió el nombre de *algoritmo de marcha*. Es fácil calcular que para su realización se necesitan aproximadamente $8N$ operaciones. Podemos mostrar que si $C \neq 2 \cos m\pi/N$, m es un número entero, entonces el problema (1) es resoluble para cualquier miembro segundo y $\alpha_{k-1}^2 \neq \alpha_k^2$. Por consiguiente en este caso las fórmulas (21) no contienen la operación de división por cero.

El algoritmo de marcha descrito arriba puede ser empleado también en un caso en que C es una matriz cuadrada, F_1 son los vectores prefijados, e y_1, los vectores buscados. Ha de notarse que el problema de Dirichlet de diferencias para la ecuación de Poisson (véase el cap. VI) sobre una red rectangular uniforme según cualquier dirección, introducida en el rectángulo, puede ser escrito en la forma (1). En este caso los valores de la función reticular buscada correspondientes a la i-ésima fila son los componentes del vector, mientras que la matriz C es tridiagonal y su orden es igual al número de filas interiores de la red.

Sea M el orden de la matriz C. Entonces los vectores p_i, q_i son de dimensión M y para calcular p_{k-1}, q_{k-1}, p_k, q_k según las fórmulas (15), (16) se exigirán $O(MN)$ operaciones. Es evidente que el mismo número de operaciones se necesitarán también para encontrar los vectores y_i, $2 \leqslant i \leqslant N-2$ según las fórmulas (2), (5). Veamos ahora la cuestión sobre el cálculo de y_1 e y_{N-1}.

De la fórmula (17) se deduce que α_k es un polinomio de grado k de C, además, si C es un número, entonces α_k será un polinomio algebraico, y si C es una matriz, α_k será un polinomio matricial. Para un polinomio que satisface la relación recurrente (17) existe una representación explícita: $\alpha_k = U_k(C/2)$, donde $U_k(x)$ es el polinomio de Chébishev de segunda especie de grado k:

$$U_k(x) = \begin{cases} \dfrac{\operatorname{sen}(k+1)\arccos x}{\operatorname{sen}\arccos x}, & |x| \leqslant 1, \\ \dfrac{(x+\sqrt{x^2-1})^{k+1} - (x-\sqrt{x^2-1})^{k+1}}{2\sqrt{x^2-1}}, & |x| \geqslant 1. \end{cases}$$

Haciendo uso de la expresión explícita para α_k, $k \geqslant 0$, y teniendo presente que α_k es un polinomio cuyo grado mayor tiene el coeficiente unidad, podemos obtener los siguientes desarrollos:

$$\begin{aligned} \alpha_k - \alpha_{k-1} &= \prod_{l=1}^{k} \left(C - 2\cos\frac{(2l-1)\pi}{2k+1} E \right), \\ \alpha_k + \alpha_{k-1} &= \prod_{l=1}^{k} \left(C - 2\cos\frac{2l\pi}{2k+1} E \right). \end{aligned} \tag{22}$$

Recurriendo a (22) y (23), construyamos el siguiente algoritmo para hallar y_1 e y_{N-1}:

$$v_0 = p_k - q_{k-1} - p_{k-1} + q_k, \quad w_0 = q_{k-1} - p_k - p_{k-1} + q_k,$$

$$\left(C - 2\cos\frac{(2l-1)\pi}{2k+1}E\right) v_l = v_{l-1},$$

$$\left(C - 2\cos\frac{2l\pi}{2k+1}E\right) w_l = w_{l-1}, \quad l = 1, 2, \ldots, k, \tag{23}$$

$$y_1 = 0{,}5\,(v_k - w_k), \quad y_{N-1} = 0{,}5\,(v_k + w_k).$$

Por cuanto cada uno de los sistemas (23) cuenta con una matriz tridiagonal (el número de tales sistemas es $2k$) y puede ser resuelto por el método de factorización realizando $O(M)$ operaciones, entonces para encontrar y_1 e y_{N-1} se necesitarán $O(NM)$ operaciones aritméticas.

Así pues, para resolver el sistema (1) con la matriz triagonal se ha construido un método en el que el número de operaciones aritméticas es proporcional al número de incógnitas.

Fijémonos en que el algoritmo de marcha construido puede ser numéricamente inestable. En efecto, si el número C satisface la condición $|C| > 2$, entonces para el algoritmo resulta característico el crecimiento del error, exponencial según N, puesto que entre las raíces de la ecuación característica $q^2 - Cq + 1 = 0$ se tiene una que en módulo es superior a la unidad. La inestabilidad del mismo tipo tiene lugar también cuando la matriz C tiene valores propios que son superiores en módulo a 2. Para los problemas de esta índole está construida actualmente una variante del algoritmo de marcha estable en aquel sentido que el error crece, al crecer N, según la ley potencial.

2. **Método de reducción.** En algunos casos, al resolver los sistemas de ecuaciones algebraicas lineales con una matriz tridiagonal, es de mucha importancia la exactitud de la solución obtenida. El análisis de las fórmulas del método de factorización que se emplea para resolver los sistemas citados muestra que la fuente del error puede radicar en las fórmulas para calcular los coeficientes de factorización. Estas fórmulas contienen la operación de división por una diferencia de las magnitudes que son próximas en su valor. Más abajo se dará a conocer el método de reducción para resolver dichos sistemas, privado de la deficiencia mencionada.

Así pues, supongamos que se requiere hallar la solución de un problema de diferencias tripuntual

$$-a_i y_{i-1} + c_i y_i - b_i y_{i+1} = f_i, \quad 1 \leqslant i \leqslant N-1, \tag{24}$$

$$y_0 = 0, \quad y_N = 0,$$

donde $c_i = a_i + b_i + d_i$, $a_i > 0$, $b_i > 0$, $d_i \geqslant 0$, $N = 2^n$. La idea del método de reducción consiste en que del sistema (24) se eliminan consecutivamente las incógnitas con números impares primeramente, y a continuación con los números múltiplos de 2, etc.

Escribamos tres ecuaciones del sistema (24) que vienen una tras otra con los números $i-1$, i, $i+1$, donde i es un número par.

$$-a_{i-1}y_{i-2}+(a_{i-1}+b_{i-1}+d_{i-1})\,y_{i-1}-b_{i-1}y_i=f_{i-1}, \tag{25}$$

$$-a_iy_{i-1}+(a_i+b_i+d_i)\,y_i-b_iy_{i+1}=f_i, \tag{26}$$

$$-a_{i+1}y_i+(a_{i+1}+b_{i+1}+d_{i+1})\,y_{i+1}-b_{i+1}y_{i+2}=f_{i+1}. \tag{27}$$

Multiplicando la ecuación (25) por $\alpha_i^{(1)}=a_i\,(a_{i-1}+b_{i-1}+d_{i-1})^{-1}$, la ecuación (27) por $\beta_i^{(1)}=b_i\,(a_{i+1}+b_{i+1}+d_{i+1})^{-1}$ y sumando las ecuaciones obtenidas con (26) llegamos a que

$$-a_i^{(1)}y_{i-2}+(a_i^{(1)}+b_i^{(1)}+d_i^{(1)})\,y_i-b_i^{(1)}y_{i+2}=f_i^{(1)},$$

$$i=2,\ 4,\ 6,\ \ldots,\ N-2,\quad y_0=0,\quad y_N=0, \tag{28}$$

donde $a_i^{(1)}=\alpha_i^{(1)}a_{i-1}$, $b_i^{(1)}=\beta_i^{(1)}b_{i+1}$, $d_i^{(1)}=\alpha_i^{(1)}d_{i-1}+d_i+\beta_i^{(1)}d_{i+1}$, $f_i^{(1)}=\alpha_i^{(1)}f_{i-1}+f_i+\beta_i^{(1)}f_{i+1}$. Si las incógnitas con números pares quedan halladas (ellas satisfacen el sistema (28)), entonces las demás incógnitas se determinarán según la fórmula

$$y_i=\frac{f_i+a_iy_{i-1}+b_iy_{i+1}}{a_i+b_i+d_i},\quad i=1,\ 3,\ 5,\ \ldots,\ N-1,$$

Es evidente que el proceso descrito de eliminación de las incógnitas puede ser aplicado al sistema (28), del cual serán eliminadas en el segundo paso las incógnitas con los números múltiplos de 2, pero no múltiplos de 4. Como resultado del l-ésimo paso del proceso de eliminación obtendremos un sistema

$$-a_i^{(l)}y_{i-2^l}+(a_i^{(l)}+b_i^{(l)}+d_i^{(l)})\,y_i-b_i^{(l)}y_{i+2^l}=f_i^{(l)}, \tag{29}$$

$$i=2^l,\ \ 2\cdot2^l,\ \ 3\cdot2^l,\ \ \ldots,\ N-2^l,\ \ y_0=0,\ \ y_N=0,$$

donde

$$\begin{aligned}
a_i^{(l)}&=\alpha_i^{(l)}a_{i-2^{l-1}}^{(l-1)},\quad b_i^{(l)}=\beta_i^{(l)}b_{i+2^{l-1}}^{(l-1)},\\
d_i^{(l)}&=\alpha_i^{(l)}d_{i-2^{l-1}}^{(l-1)}+d_i^{(l-1)}+\beta_i^{(l)}d_{i+2^{l-1}}^{(l-1)},\\
f_i^{(l)}&=\alpha_i^{(l)}f_{i-2^{l-1}}^{(l-1)}+f_i^{(l-1)}+\beta_i^{(l)}f_{i+2^{l-1}}^{(l-1)},\\
\alpha_i^{(l)}&=a_i^{(l-1)}\left(a_{i-2^{l-1}}^{(l-1)}+b_{i-2^{l-1}}^{(l-1)}+d_{i-2^{l-1}}^{(l-1)}\right)^{-1},\\
\beta_i^{(l)}&=b_i^{(l-1)}\left(a_{i+2^{l-1}}^{(l-1)}+b_{i+2^{l-1}}^{(l-1)}+d_{i+2^{l-1}}^{(l-1)}\right)^{-1},
\end{aligned} \tag{30}$$

$$i=2^l,\ \ 2\cdot2^l,\ \ 3\cdot2^l,\ \ \ldots,\ N-2^l,\ \ l\geqslant1.$$

Aquí se usan las designaciones $a_i^{(0)}=a_i$, $b_i^{(0)}=b_i$, $d_i^{(0)}=d_i$, $f_i^{(0)}=f_i$.

El proceso de eliminación se dará por terminado en el $(n-1)$-ésimo paso, cuando el sistema (29) se componga de una sola ecuación

respecto de la incógnita $y_{N/2} = y_{2^{n-1}}$. De esta ecuación encontraremos

$$y_{2^{n-1}} = \frac{f_{2^{n-1}}^{(n-1)} + a_{2^{n-1}}^{(n-1)} y_0 + b_{2^{n-1}}^{(n-1)} y_N}{a_{2^{n-1}}^{(n-1)} + b_{2^{n-1}}^{(n-1)} + d_{2^{n-1}}^{(n-1)}}, \quad y_0 = y_N = 0. \tag{31}$$

Las incógnitas restantes se determinarán según las fórmulas

$$y_l = \frac{f_l^{(l)} + a_l^{(l)} y_{l-2^l} + b_l^{(l)} y_{l+2^l}}{a_l^{(l)} + b_l^{(l)} + d_l^{(l)}}, \quad l = 2^l, \quad 3 \cdot 2^l, \quad 5 \cdot 2^l, \ \ldots, \ N - 2^l, \tag{32}$$

donde $l = n-2, \ n-3, \ \ldots, \ 0, \ y_0 = y_N = 0$. Observemos que la fórmula (32) incluye la fórmula (31) para $l = n-1$.

Así pues, aplicándose el método de reducción, en el paso directo se calculan $a_i^{(l)}$, $b_i^{(l)}$, $d_i^{(l)}$, $f_i^{(l)}$ según las fórmulas (30) para $l = 1, 2, \ldots, n-1$, y en el paso inverso se halla la solución buscada por la fórmula (32) para $l = n-1, \ n-2, \ \ldots, \ 0$. Ha de ser notado que el método no requiere una memoria complementaria, puesto que las magnitudes $a_i^{(l)}$, $b_i^{(l)}$, $d_i^{(l)}$, $f_i^{(l)}$ pueden ser dispuestas en los lugares respectivos de $a_{i-2^l}^{(l-1)}$, $b_{i-2^l}^{(l-1)}$, $d_i^{(l-1)}$, $f_i^{(l-1)}$. Para realizar el método se necesitan $12N$ adiciones, $8N$ multiplicaciones y $3N$ divisiones.

Bibliografía

1. Bakhválov N. S. Numérical Methods, Ed. Mir, M., 1978
2. Березин И. С., Жидков Н. П. Методы вычислений. М., Наука, 1966, ч. I; Физматгиз, 1962, ч. 2
(Berézin I. S., Zhidkov N. P. Métodos de los cálculos).
3. Воеводин В. В. Численные методы алгебры; теория и алгоритмы. М., Наука, 1966
(Voevodin V. V. Métodos numéricos del álgebra; teoría y algoritmos).
4. Годунов С. К., Рябенький В. С. Разностные схемы. М., Наука, 1977
(Godunov S. K., Riábenki V. S. Esquemas de diferencias).
5. Калиткин Н. Н. Численные методы. М., Наука, 1978
(Kalitkin N. N. Métodos numéricos).
6. Ляшко И. И., Макаров В. Л., Скоробогатько А. А. Методы вычислений. Киев, Высшая школа, 1977
(Liashkó I. I., Makárov V. L., Skorobogatko A. A. Métodos de los cálculos).
7. Марчук Г. И. Методы вычислительной математики. М., Наука, 1977
(Marchuk G. I. Métodos de las matemáticas de cálculo).
8. Никольский С. М. Квадратурные формулы. М., Наука, 1979
(Nikolski S. M. Fórmulas de cuadratura).
9. Самарский А. А. Теория разностных схем. М., Наука, 1977
(Samarski A. A. Teoría de los esquemas de diferencias).
10. Самарский А. А., Андреев В. Б. Разностные методы для эллиптических уравнений. М., Наука, 1976
(Samarski A. A., Andréiev V. B. Métodos de diferencias para las ecuaciones elípticas).
11. Самарский А. А., Гулин А. В. Устойчивость разностных схем. М., Наука, 1973
(Samarski A. A., Gulin A. V. Estabilidad de los esquemas de diferencias).
12. Samarski A. A., Nikoláev E. S. Métodos de solución de las ecuaciones reticulares, Ed. Mir, M., 1983.
13. Самарский А. А., Попов Ю. П. Разностные методы газовой динамики. М., Наука, 1980
(Samarski A. A., Popov Yu. P. Métodos de diferencias de la dinámica de los gases).

14. Tíjonov A. N., Samarski A. A. Ecuaciones de la física matemática, Ed. Mir, M., 1983.
15. Фаддеев Д. К., Фаддеева В. Н. Вычислительные методы линейной алгебры. М., Наука, 1972
(Faddéiev D. K., Faddéieva V. N. Métodos computacionales del álgebra lineal).
16. Яненко Н. Н. Метод дробных шагов решения многомерных задач математической физики. Новосибирск, Наука, 1967
(Yanenko N. N. Método de pasos fraccionarios para resolver problemas multidimensionales de la física matemática).

Lista de designaciones

$\omega_N = \{t;\ t = 0, 1, \ldots, N\}$, red con nodos de números enteros

$\omega_h = \{x_t = th,\ h = 1/N,\ 0 \leqslant t \leqslant N\}$, red uniforme de paso h en el segmento $[0, 1]$.

h, paso de la red ω_h

$y_t = y(x_t) = y(t)$, valor de la función reticular en el t-ésimo nodo de la red

$\hat{\omega}_h$, red no uniforme

$h_t = x_t - x_{t-1}$, paso de la red no uniforme $\hat{\omega}_h$:

$\hbar_t = \frac{1}{2}(h_t + h_{t+1})$

$v_{t_1 t_2} = v(x_{t_1}^1, x_{t_2}^2)$, valor de la función reticular bidimensional en el nodo (t, j)

$v_{t_1 t_2}^n = v(x_{t_1}^1, x_{t_2}^2, t_n)$, valor de la función reticular en el nodo (t, j) en la n-ésima capa temporal

$v_{tj}^{n+1} = \hat{v}$, valor de la función reticular bidimensional en el nodo (t, j) sobre la $(n+1)$-ésima capa

$\Delta y_t = y_{t+1} - y_t$, diferencia derecha en el t-ésimo nodo

$\nabla y_t = y_t - y_{t-1}$, diferencia izquierda en el t-ésimo nodo

$\delta y_t = \frac{1}{2}(\nabla y_t + \Delta y_t)$, diferencia central en el t-ésimo nodo

$\Delta^2 y_{t+1} = \Delta(\Delta y_{t+1}) = \Delta(\Delta y_t)$, diferencia de segundo orden

$y_{x,t} = (y_{t+1} - y_t)/h$, derivada de diferencias derecha en el nodo x_t

$y_{\bar{x},t} = (y_t - y_{t-1})/h$, derivada de diferencias izquierda en el nodo x_t

$y_{\mathring{x},t} = (y_{t+1} - y_{t-1})/(2h)$, derivada de diferencias central en el nodo x_t

$y_{\bar{x}x,t} = (y_{t+1} - 2y_t + y_{t-1})/h^2$, segunda derivada de diferencias

H, espacio de Hilbert

(y, v), producto escalar de los elementos $y, v \in H$, $\| y \| = \sqrt{(y, y)}$

E, operador unidad

A^*, operador conjugado del operador A

A^{-1}, operador inverso al operador A

$A > 0$, operador positivo

$A \geqslant 0$, operador no negativo

$A \geqslant \delta E$, $\delta > 0$, operador definido positivo

$\| y \|_A = \sqrt{(Ay, y)}$, $y \in H$, norma energética

Espacio de funciones reticulares:

$\Omega_{N+1} = \{y_i,\ i = 0, \ldots, N\}$

$\overset{*}{\Omega}_{N-1} = \{y_i,\ i = 0, \ldots, N;\ y_0 = 0,\ y_N = 0\}$

$\overset{\circ}{y}_i$, función del $\overset{*}{\Omega}_{N+1}$

$\Omega_N^* = \{y_i,\ i = 0, 1, \ldots, N-1\}$

$\overline{\Omega_N} = \{y_i,\ i = 1, 2, \ldots, N\}$

Productos escalares y normas en la red:

$$(y, v) = \sum_{i=1}^{N-1} y_i v_i h, \quad \| y \| = \sqrt{(y, y)}$$

$$(y, v] = \sum_{i=1}^{N} y_i v_i h, \quad \| y] \| = \sqrt{(y, y)}$$

$$\| y \|_C = \max_{x_i \in \overline{\omega}_h} | y(x_i) | = \max_{0 \leqslant i \leqslant N} | y(x_i) |$$

Indice alfabetico

A nuetros 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, 1-110, GSP, URSS.

УДК 519.6=60

Александр Андреевич *Самарский*

ВВЕДЕНИЕ В ЧИСЛЕННЫЕ МЕТОДЫ

Контрольный редактор С. Я. Калашник
Редактор Н. А. Стальнова
Художники С. А. Бычков, Г. В. Чучелов
Художественный редактор Е. И. Подмарькова
Технический редактор В. П. Сизова
Корректор Г. А. Макарова

ИБ № 5146

Сдано в набор 22.05.85. Подписано к печати 19.12.85.
Формат 84 × 108¹/₃₂. *Бумага типографская № 1.*
Гарнитура обыкновенная. Печать высокая.
Объем 4,88 бум. л. Усл. печ. л. 15,05. Усл. кр.-отт. 16,92.
Уч.-изд. л. 15,05. Изд. № 19/3488. Тираж 13 380 экз.
Зак. 0242. Цена 1 р. 55 к.

ИЗДАТЕЛЬСТВО «МИР»
129820, Москва, И-110, ГСП, 1-й Рижский пер., 2.

Ордена Трудового Красного Знамени
Московская типография № 7 «Искра революции»
Союзполиграфпрома Государственного комитета СССР
по делам издательств, полиграфии и книжной торговли.
103001, Москва, Трехпрудный пер., 9.

С $\frac{1702070000-024}{056(01)-86}$ 146-86, ч. 2

Goloviná L.

ALGEBRA LINEAL Y SUS APLICACIONES

(3ª edición)

No obstante su pequeño volumen, el libro contiene los problemas fundamentales del curso de álgebra lineal, así como sus distintas aplicaciones, incluyendo la investigación de las curvas y superficies de segundo orden, la noción sobre tensores y otros problemas.

En el libro se exponen los conceptos primordiales referentes a los espacios lineales y euclidianos, y a las transformaciones lineales; además se estudian problemas sobre vectores y se obtiene la forma canónica de las matrices de las transformaciones autoconjugada y otrogonal en el espacio euclidiano, dándose ejemplos básicos de la teoría de las formas cuadráticas.

Como suplemento a las formas cuadráticas, se examina la teoría general de las líneas y superficies de segundo orden. Dos capítulos están consagrados a las transformaciones de Lorentz y nociones fundamentales de la teoría especial de la relatividad. En el capítulo sobre grupos, además de las definiciones principales, se incluye una selección de ejemplos.

Un mérito evidente del libro es la acertada elección del material, en el cual se han examinado problemas que no entran en el programa de los estudiantes de especialidades no matemáticas, pero que son de cierto interés para éstos. Con ello, las nociones indispensables previas y el nivel de la exposición son tales que, al leer el texto, los estudiantes no encuentran ninguna dificultad.

La obra está destinada a estudiantes y profesores de centros de enseñanza superior. También les es de gran utilidad a los ingenieros que deseen conocer las nociones fundamentales del álgebra lineal, empleando una fuente que no exige información previa de matemáticas superiores.

Kagán V.

LOBACHEVSKI

En este libro se narra de la vida y de las actividades sociales y científicas del eminente matemático ruso N. I. Lobachevski.
Se relata en detalle sobre la niñez y adolescencia del gran sabio, sobre su familia, profesores del liceo que influyeron en la formación del carácter y la concepción del mundo del científico. El autor presta especial atención a los años estudiantiles de Lobachevski en la Universidad de Kazán, y en particular, a sus estudios de las matemáticas.
Gran parte de la obra está dedicada a la actividad científica de Lobachevski. El lector conocerá el manual de estudio titulado "Geometría", donde por primera vez la geometría absoluta fue separada de la euclidiana. Se trata en detalle la creación de la geometría no euclidiana, se analiza el trabajo "Geometría, investigación de la teoría de las líneas paralelas". Un apartado importante del libro se destina a los años más fructíferos de la obra de Lobachevski (1827—1846), en que fuera rector de la Universidad de Kazán. En este período él publicó sus trabajos más notables: "Sobre los principios de la geometría", "Nuevos principios de la geometría con la teoría completa de las paralelas", etc., que son analizados. El siguiente apartado ha sido dedicado a la relación de sus contemporáneos hacia las ideas de Lobachevski, al desarrollo de estas ideas en los años ulteriores, a la aplicación de la geometría de Lobachevski a otras partes de las matemáticas, así como a la mecánica, física y cosmología.
Este libro resultará sin duda de interés a estudiantes, maestros, profesores y a toda persona aficionada a las matemáticas.