



M. DRYJA (Warszawa)

Metoda Galerkina przemiennych kierunków dla quasiliniowych równań parabolicznych*

(Praca przyjęta do druku 23.11.1977)

1. Wstęp. W pracy rozważane jest zagadnienie początkowo-brzegowe, z jednorodnymi warunkami brzegowymi Dirichleta dla quasiliniowego równania parabolicznego, rozpatrywanego w walcu $Q_T = \bar{\Omega} \times [0, T]$, gdzie Ω jest obszarem ograniczonym na płaszczyźnie. Dla tego zagadnienia w klasie metod wariacyjno-różnicowych poszukujemy metody bezwarunkowo stabilnej oraz takiej, której koszt realizacji na maszynie cyfrowej za pomocą znanych algorytmów numerycznie stabilnych byłby proporcjonalny do liczby punktów, w których poszukujemy rozwiązania przybliżonego.

Metodą wariacyjno-różnicową nazywamy tutaj metodę powstałą z połączenia metod siatek i Galerkina, w szczególności metody elementu skończonego, przy czym metoda siatek jest stosowana względem zmiennej czasowej, a metoda Galerkina — względem zmiennych przestrzennych. Metodą, która spełnia powyższe wymagania, jest metoda Galerkina przemiennych kierunków (ADG), mająca wiele wspólnego z metodą ADI, w szczególności ze schematami różnicowymi z rozszczepiającym się operatorem ([1], [10]). Pracą, która zapoczątkowała ten kierunek badań, jest praca [2] Douglasa i Dupont'a z 1971 r. W pracy tej metoda ADG jest sformułowana i zbadana dla równań parabolicznych i hiperbolicznych, również nieliniowych, ale z nieliniowościami ograniczonymi (słabymi), rozpatrywanych w kostce. W pracy [3] metoda ta jest uogólniona na obszar Ω o kształcie litery L. W niniejszej pracy metodę ADG uogólniamy na obszary dowolne i na pewne równania z nieliniowościami nieograniczonymi (silnymi). Uogólnienia te są przedstawione na przykładzie zagadnienia (2.1), (2.2) z rozwiązaniami uogólnionymi (zob. (3.1)–(3.3)). Podane są oszacowanie ogólne błędu metody (tw. 1) oraz szybkość zbieżności metody przy konkretnym wyborze funkcji bazowych postaci iloczynu funkcji daszkowych jednej zmiennej (tw. 2).

* Artykuł ten jest sprawozdaniem z pracy wykonanej w ramach tematu 4(1)1 problemu międzyresortowego I.1.

Przedstawiony algorytm realizacji metody w przypadku gdy Ω jest prostokątem lub ma kształt litery L jest prostszy od algorytmów przedstawionych dla tych obszarów w pracach [2], [3].

2. Zagadnienie różniczkowe. Rozważamy zagadnienie początkowo-brzegowe dla quasilineowego równania parabolicznego postaci ⁽¹⁾

$$(2.1) \quad \frac{\partial u}{\partial t} - \sum_{i=1}^2 \frac{d}{dx_i} a_i(x, t, u, D_1 u, D_2 u) + a_0(x, t, u, D_1 u, D_2 u) = f(x, t),$$

$$x = (x_1, x_2) \in \Omega, \quad t \in (0, T],$$

$$(2.2) \quad u(x, 0) = u_0(x), \quad x \in \Omega, \quad u(x, t) = 0, \quad x \in \partial\Omega, \quad t \in [0, T],$$

gdzie Ω jest obszarem ograniczonym w R^2 , a $\partial\Omega$ jego brzegiem przedziałami gładkim. Dalej będziemy korzystać z następujących oznaczeń:

$$a_i(x, t, u, D_1 u, D_2 u) \equiv a_i(x, t, u, \nabla u) \equiv a_i(x, t, p'), \quad i = 0, 1, 2,$$

gdzie $p' \equiv (p_0, p_1, p_2) \equiv (p_0, p)$, $p = (p_1, p_2)$. Zakładamy, że współczynniki a_i , $i = 0, 1, 2$, spełniają następujący warunek:

\exists stałe $\gamma_0, \gamma_1 > 0$ takie, że $\forall (x, t) \in Q_T = \bar{\Omega} \times [0, T]$, $p' \in R^3$, $\forall \xi \in R^2$, $\xi \neq 0$

$$(2.3) \quad \gamma_0 \sum_{i=1}^2 \xi_i^2 \leq \sum_{i,j=0}^2 \frac{\partial a_i}{\partial p_j} \xi_i \xi_j \leq \gamma_1 \sum_{i=1}^2 \xi_i^2.$$

Warunek (2.3) jest mocniejszy od warunku jednostajnej eliptyczności operatora różniczkowego: sumowanie w formie dwuliniowej występuje od zera. Założenie to jest potrzebne po to, by forma dwuliniowa odpowiadająca temu operatorowi była H_0^1 -eliptyczna (zob. lemat 1). Dodajmy, że dla niektórych zagadnień wystarczy zakładać tylko jednostajną eliptyczność (zob. uwagę 2 ze str. 11). Oprócz (2.3) będziemy jeszcze zakładać, że funkcje $a_i(x, t, p')$, $i = 0, 1, 2$, są ciągle i mają pochodne ciągle w $Q_T \times R^3$ oraz że

$$(2.4) \quad |a_i(x, t, 0, 0)| \leq M, \quad \left| \frac{\partial a_i}{\partial x_k} \right| \leq M, \quad \left| \frac{\partial a_i}{\partial p_j} \right| \leq M, \quad \left| \frac{\partial a_i}{\partial t} \right| \leq M,$$

dla $i, j = 0, 1, 2$ $\forall (x, t) \in Q_T$ i $\forall p' \in R^3$. Równanie (2.1), którego współczynniki spełniają warunki (2.3) i (2.4), będziemy nazywać *równaniem z nieliniowością ograniczoną* [4]. Zagadnienia z nieliniowościami nieograniczonymi będziemy rozpatrywać w ustępie 8.

3. Postać wariacyjna zagadnienia różniczkowego. Dla $f \in L^2(0, T; L^2(\Omega))$, $u_0 \in L^2(\Omega)$ znaleźć funkcję $u: [0, T] \rightarrow H_0^1(\Omega)$ taką, że

$$(3.1) \quad u \in L^2(0, T; H_0^1(\Omega)), \quad \frac{du}{dt} \in L^2(0, T; L^2(\Omega)),$$

⁽¹⁾ Tutaj da_i/dx_i oznacza pochodną złożoną funkcji a_i względem zmiennej x_i .

$$(3.2) \quad \frac{d}{dt}(u(t), v) + a(t, u, \nabla u; v) = (f(t), v) \quad \forall v \in H_0^1(\Omega), t \in (0, T],$$

$$(3.3) \quad u(0) = u_0,$$

gdzie (\cdot, \cdot) jest iloczynem skalarnym w $L^2(\Omega)$,

$$(3.4) \quad a(t, u, \nabla u; v) \equiv \int_{\Omega} \left\{ \sum_{i=1}^2 a_i(x, t, u, \nabla u) D_i v + a_0(x, t, u, \nabla u) v \right\} d\Omega.$$

Zagadnienie (3.1)–(3.3) przy założeniach (2.3), (2.4) ma jednoznaczne rozwiązanie (zob. [5], [6]).

4. Zagadnienie przybliżone. Naszym celem jest skonstruowanie metody Galerkinia zmiennych kierunków dla zagadnienia (3.1)–(3.3). W związku z tym wprowadzimy kilka oznaczeń. Niech ω_τ będzie siatką na $[0, T]$ postaci

$$\omega_\tau \equiv \{t: t = n\tau, n = 0, \dots, N, N\tau = T\}.$$

Niech V_m będzie m -wymiarową podprzestrzenią H_0^1 o własności:

$$D_1 D_2 u \in L^2(\Omega) \quad \text{dla } \forall u \in V_m.$$

Zadanie przybliżone w metodzie Galerkinia zmiennych kierunków formułujemy w następujący sposób:

znaleźć taką funkcję $U: \omega_\tau \rightarrow V_m$, że

$$(4.1) \quad (U_t^n, v) + \theta\tau(\nabla U_t^n, \nabla v) + \theta^2\tau^2(D_1 D_2 U_t^n, D_1 D_2 v) + a^n(U^n, \nabla U^n; v) = (f^n, v),$$

$$\forall v \in V_m, n = 0, \dots, N-1,$$

$$(4.2) \quad (U^0, v) = (u_0, v) \quad \forall v \in V_m,$$

gdzie θ jest parametrem dodatnim, który wybierzemy później, zaś

$$U^n \equiv U(n\tau), \quad U_t^n \equiv (U^{n+1} - U^n)/\tau, \quad a^n(U^n, \nabla U^n; v) = a(n\tau, U^n, \nabla U^n; v).$$

5. Oszacowanie błędu. Niżej podamy ogólne oszacowanie błędu metody (4.1)–(4.2), na podstawie którego ustalamy szybkość zbieżności konkretnej metody, np. metody elementu skończonego, w zależności od parametrów określających tę metodę. Do wykazania takiego oszacowania wykorzystujemy podejście stosowane w pracach [7], [2] i innych.

Rozpatrujemy najpierw zadanie pomocnicze:

znaleźć dla każdego $t \in [0, T]$ taką funkcję $W \in V_m$, że

$$(5.1) \quad a(t, W, \nabla W; v) = a(t, u, \nabla u; v) \quad \forall v \in V_m(\Omega),$$

gdzie $a(t, u, \nabla u; v)$ traktujemy jako zadaną formę na rozwiązaniu u zagadnienia (3.1)–(3.3).

Przechodzimy do sformułowania twierdzenia o szacowaniu błędu metody (4.1), (4.2).

TWIERDZENIE 1. Niech będą spełnione założenia (2.3), (2.4) oraz niech $f \in C(0, T; L^2(\Omega))$, $du/dt \in C(0, T; L^2(\Omega))$

$$(5.2) \quad \frac{du}{dt} \in L^2(0, T; H^2(\Omega)).$$

Wtedy dla $\theta > \theta_0 > 0$ ma miejsce następujące oszacowanie:

$$(5.3) \quad \max_n \|U^n - u^n\| \leq M \left\{ \max_n \|u^n - W^n\| + \left(\tau \sum_{n=0}^{N-1} \left\| \left(\frac{du}{dt} \right)^n - W^n \right\|^2 \right)^{1/2} + \right. \\ \left. + \tau \|\nabla(U^0 - W^0)\| + \tau^2 \|D_1 D_2(U^0 - W^0)\| + \tau \right\},$$

gdzie $\theta_0 = c_3^2/2\varepsilon$, c_3 — stała z nierówności (5.5), $\varepsilon > 0$, $\|\cdot\|$ jest normą w L^2 , a u^n , U^n , W^n są rozwiązaniami odpowiednio zadań (3.1)–(3.3), (4.1), (4.2) i (5.1), $M = M(u) > 0$.

Dowód twierdzenia 1 poprzedzimy trzema lematami.

LEMAT 1. Jeśli spełniony jest warunek (2.3), to ma miejsce następująca nierówność:

$$(5.4) \quad a(t, u, \nabla u; u-v) - a(t, v, \nabla v; u-v) \geq c_0 \|\nabla(u-v)\|^2, \\ u, v \in H_0^1, \quad t \in [0, T],$$

gdzie $c_0 = \text{const} > 0$.

D o w ó d. Niech q_i , $i = 0, 1, 2$, oznaczają następujące wielkości:

$$q_0 = su + (1-s)v, \quad q_i = sD_i u + (1-s)D_i v, \quad i = 1, 2,$$

gdzie $u, v \in H_0^1(\Omega)$, $s \in [0, 1]$. Rozpatrzmy wyrażenie

$$I_i \equiv a_i(x, t, u, D_1 u, D_2 u) - a_i(x, t, v, D_1 v, D_2 v).$$

Za pomocą prostych przekształceń sprawdzamy, że mają miejsce następujące równości ($i = 0, 1, 2$):

$$I_i = \int_0^1 \frac{d}{ds} a_i(x, t, q_0, q_1, q_2) ds = (u-v) \int_0^1 \frac{\partial a_i(x, t, q_0, q_1, q_2)}{\partial q_0} ds + \\ + (D_1 u - D_1 v) \int_0^1 \frac{\partial a_i(x, t, q_0, q_1, q_2)}{\partial q_1} ds + \\ + (D_2 u - D_2 v) \int_0^1 \frac{\partial a_i(x, t, q_0, q_1, q_2)}{\partial q_2} ds.$$

Wykorzystując te równości, mamy

$$(5.4') \quad a(t, u, \nabla u; u-v) - a(t, v, \nabla v; u-v) = \\ = \sum_{i,j=1}^2 \int_{\Omega} \left(\int_0^1 \frac{\partial a_i}{\partial q_j} ds \right) D_j(u-v) D_i(u-v) + \left(\int_0^1 \frac{\partial a_i}{\partial q_0} ds \right) (u-v) D_i(u-v) +$$

$$+ \left(\int_0^1 \frac{\partial a_0}{\partial q_i} ds \right) (u-v) D_i (u-v) + \left(\int_0^1 \frac{\partial a_0}{\partial q_0} ds \right) (u-v)^2 \Big\} d\Omega.$$

Stąd na podstawie warunku (2.3) otrzymujemy (5.4).

U w a g a. Przy założeniu tylko jednostajnej eliptyczności operatora, po lewej stronie (5.4) pojawi się jeszcze wyraz $-c_1 \|u-v\|^2$, $c_1 > 0$, przy założeniu (2.4), co widać z postaci (5.4'). Zakładając dodatkowo, że

$$c_1 \text{diam}(\Omega)^2 \leq c_0 - \delta, \quad \delta > 0,$$

widzimy, że w tym przypadku również ma miejsce nierówność (5.4).

LEMAT 2. *Jeśli spełniony jest warunek (2.4), to ma miejsce następujące oszacowanie:*

$$(5.5) \quad |a(t, u, \nabla u; u-w) - a(t, v, \nabla v; u-w)| \leq c_3 \|\nabla(u-w)\| \|\nabla(u-v)\|,$$

$$\forall w, u, v \in H_0^1(\Omega), \quad t \in [0, T],$$

gdzie $c_3 = \text{const} > 0$.

D o w ó d. Nierówność (5.5) wynika z (5.4'). Rzeczywiście, równość (5.4') w tym przypadku ma postać

$$\begin{aligned} & a(t, u, \nabla u; u-w) - a(t, v, \nabla v; u-w) = \\ & = \sum_{i,j=1}^2 \int_{\Omega} \left\{ \left(\int_0^1 \frac{\partial a_i}{\partial q_j} ds \right) D_j(u-v) D_i(u-w) + \left(\int_0^1 \frac{\partial a_i}{\partial q_0} ds \right) (u-v) D_i(u-w) + \right. \\ & \quad \left. + \left(\int_0^1 \frac{\partial a_0}{\partial q_i} ds \right) D_i(u-v) \cdot (u-w) + \left(\int_0^1 \frac{\partial a_0}{\partial q_0} ds \right) (u-v)(u-w) \right\} d\Omega. \end{aligned}$$

Stosując teraz nierówności Schwarz'a, Cauchy'ego i Friedrichsa otrzymujemy (5.5).

LEMAT 3. *Niech będą spełnione założenia twierdzenia 1. Wtedy dla $Z^n = U^n - W^n$, gdzie U^n, W^n są rozwiązaniami odpowiednio zadań (4.1), (4.2) i (5.1), ma miejsce następujące oszacowanie:*

$$\begin{aligned} (5.6) \quad \max_n \{ \|Z^n\|^2 + \tau \|\nabla Z^n\|^2 + \tau^2 \|D_1 D_2 Z^n\|^2 \} & \leq \\ & \leq M \left\{ \|Z^0\|^2 + \tau \|\nabla Z^0\|^2 + \tau^2 \|D_1 D_2 Z^0\|^2 + \right. \\ & \quad \left. + \sum_{n=0}^{N-1} \tau \left\| \left(\frac{du}{dt} \right)^n - W_t^n \right\|^2 + \tau^2 \|\nabla W_t^n\|^2 + \tau^2 \|D_1 D_2 W_t^n\|^2 \right\}, \end{aligned}$$

gdzie u jest rozwiązaniem (3.1)–(3.3), $M = M(u) > 0$.

D o w ó d. Odejmując od obydwu stron równania (4.1) trzy pierwsze wyrazy lewej strony (4.1), w których zamiast U_t^n występuje W_t^n , otrzymujemy

$$\begin{aligned} (5.7) \quad (Z_t^n, v) + \theta \tau (\nabla Z_t^n, \nabla v) + \theta^2 \tau^2 (D_1 D_2 Z_t^n, D_1 D_2 v) + a^n(U^n, \nabla U^n; v) = \\ = (f^n, v) - (W_t^n, v) - \theta \tau (\nabla W_t^n, \nabla v) - \theta^2 \tau^2 (D_1 D_2 W_t^n, D_1 D_2 v). \end{aligned}$$

Równanie (3.2) bierzemy na n -warstwie i odejmujemy stronami od (5.7). Wtedy, po uwzględnieniu ciągłości względem t funkcji $du/dt, f$ i a , otrzymujemy

$$(5.8) \quad (Z_t^n, v) + \theta \tau (\nabla Z_t^n, \nabla v) + \theta^2 \tau^2 (D_1 D_2 Z_t^n, D_1 D_2 v) + A^n(U^n, u^n, v) = \\ = \left(\left(\frac{du}{dt} \right)^n - W_t^n, v \right) - \theta \tau (\nabla W_t^n, \nabla v) - \theta^2 \tau^2 (D_1 D_2 W_t^n, D_1 D_2 v), \quad \forall v \in V_m,$$

gdzie

$$A^n(U^n, u^n, v) = a^n(U^n, \nabla U^n; v) - a^n(u^n, \nabla u^n; v).$$

Przyjmując $v = Z^{n+1}$ oszacujemy najpierw wyraz A^n . Wykorzystując oczywistą tożsamość $Z^{n+1} = Z^n + \tau Z_t^n$, równanie (5.1) oraz oszacowania z lematów 1 i 2, mamy

$$(5.9) \quad A^n(U^n, u^n, Z^{n+1}) = A^n(U^n, W^n, Z^n) + \tau A^n(U^n, W^n, Z_t^n) \geq \\ \geq (c_0 - \varepsilon_1) \|\nabla Z^n\|^2 - (c_3^2/4\varepsilon_1) \tau^2 \|\nabla Z_t^n\|^2.$$

Przyjmując w równaniu (5.8) $v = 2Z^{n+1}$, sumując je względem n od 0 do $k-1$, wykorzystując oszacowanie (5.9) i następującą tożsamość:

$$2(v_t^n, v^{n+1}) = (v^n, v^n)_t + \tau (v_t^n, v_t^n)$$

oraz ε -nierówność

$$ab \leq \frac{\varepsilon}{2} a^2 + \frac{1}{2\varepsilon} b^2, \quad \varepsilon > 0, \quad a, b \in R,$$

do przekształcenia niektórych wyrazów (5.8), otrzymujemy

$$(5.10) \quad (1 - \varepsilon_2) \|Z^k\|^2 + \tau \left(\theta - \frac{\varepsilon_3}{2} \right) \|\nabla Z^k\|^2 + \tau^2 \theta^2 (1 - \varepsilon_4) \|D_1 D_2 Z^k\|^2 + \\ + 2\tau \sum_{n=0}^{k-1} \left\{ \left(c_0 - \varepsilon_1 - \frac{\varepsilon_3}{2} \right) \|\nabla Z^n\|^2 - \frac{\varepsilon_2}{2} \|Z^n\|^2 + \frac{\tau^2}{2} \left(\theta - \frac{c_3^2}{2\varepsilon_1} \right) \|\nabla Z_t^n\|^2 \right\} \leq \\ \leq \tau \sum_{n=0}^{k-1} \left\{ \frac{1}{\varepsilon_2} \left\| \left(\frac{du}{dt} \right)^n - W_t^n \right\|^2 + \frac{\theta^2 \tau^2}{\varepsilon_3} \|\nabla W_t^n\|^2 + \theta^2 \tau^2 \left(\frac{1}{\varepsilon_4} \|D_1 D_2 W_t^n\|^2 + \right. \right. \\ \left. \left. + \varepsilon_4 \|D_1 D_2 Z^n\|^2 \right) \right\} + \|Z^0\|^2 + \tau \theta \|\nabla Z^0\|^2 + \theta^2 \tau^2 \|D_1 D_2 Z^0\|^2.$$

Wybierając odpowiednie $\varepsilon_i, i = 1, \dots, 4$ oraz $\theta > \theta_0 = c_3^2/2\varepsilon_1$ i stosując nierówność Gronwalla, z nierówności (5.10) otrzymujemy (5.6), co należało wykazać.

D o w ó d t w i e r d z e n i a 1. Ponieważ $du/dt \in L^2(0, T, H^2(\Omega))$, więc

$$\tau \sum_{n=0}^{N-1} (\|D_1 D_2 W_t^n\|^2 + \|\nabla W_t^n\|^2) \leq M.$$

Wykorzystując ten fakt, nierówność (5.6) możemy przepisać w postaci

$$(5.11) \quad \max_n \|Z^n\| \leq M \left\{ \left(\tau \sum_{n=0}^{N-1} \left\| \left(\frac{du}{dt} \right)^n - W_t^n \right\| \right) + \|Z^0\| + \right. \\ \left. + \tau \|\nabla Z^0\| + \tau^2 \|D_1 D_2 Z^0\| + \tau \right\}.$$

Stosując teraz nierówność trójkąta otrzymujemy

$$(5.12) \quad \max_n \|U^n - u^n\| \leq \max_n \|U^n - W^n\| + \max_n \|W^n - u^n\| \leq \\ \leq M \left\{ \left(\tau \sum_{n=0}^{N-1} \left\| \left(\frac{du}{dt} \right)^n - W_t^n \right\|^2 \right)^{1/2} + \max_n \|u^n - W^n\| + \|Z^0\| + \right. \\ \left. + \tau \|\nabla Z^0\| + \tau^2 \|D_1 D_2 Z^0\| + \tau \right\}.$$

Z nierówności (5.12), po uwzględnieniu nierówności

$$\|Z^0\| \leq \|u^0 - W^0\|$$

wynikającej z warunku początkowego (4.2), otrzymujemy oszacowanie (5.3), co należało wykazać.

U w a g a 1. Założenie w twierdzeniu 1 o ciągłości funkcji f można osłabić. Gdy $f \in L^2(0, T; L^2(\Omega))$, wtedy za f^n przyjmujemy uśrednienie całkowe na odcinku np. $[t_n, t_{n+1}]$. W tym przypadku oszacowanie (5.3) będzie jeszcze zawierać błąd aproksymacji funkcji f względem t .

U w a g a 2. W twierdzeniu 1 warunek (2.3) można zastąpić warunkiem jednostajnej eliptyczności, np. w przypadkach gdy a_0 w równaniu (2.1) jest postaci $\operatorname{div} f(x, t, u)$ (zob. [16]) lub $\partial^2 a_0 / \partial x_i \partial p_i$, $i = 1, 2$, są ograniczone lub $c_1 \operatorname{diam}(\Omega)^2 \leq \leq c_0 - \delta$, $\delta > 0$ (zob. uwagę ze str. 9 i [17]).

6. Wybór funkcji bazowych w metodzie ADG. Przechodzimy do wyboru funkcji bazowych, przy których wyznaczenie U^n w każdym kroku czasowym sprowadza się do rozwiązania dwóch układów równań algebraicznych o macierzy trójkątnej. Tymi funkcjami będą iloczyny funkcji daszkowych jednej zmiennej.

Aby określić te funkcje, konstruujemy siatkę Ω_h na obszarze Ω w następujący sposób. Niech R_h^2 będzie siatką prostokątną na R^2 postaci

$$R_h^2 := \{x: x = (x_1, x_2) = (ih_1, jh_2), h_1, h_2 > 0, i, j \text{ — całkowite}\}.$$

Za Ω_h przyjmujemy zbiór postaci

$$\Omega_h := \{x: x \in R_h^2, I_i^\pm x, I_j^\pm x \in \bar{\Omega}, i, j = 1, 2, i \neq j\},$$

gdzie $I_i^\pm x = x \pm h_i e_i$, $e_1 = (1, 0)$, $e_2 = (0, 1)$. Niech $\Omega_{\square, h}$ oznacza obszar składający się z prostokątów „siatki” o bokach h_1 i h_2 , taki, że $\Omega_h \subset \Omega_{\square, h} \subset \bar{\Omega}$.

Zbiór punktów siatkowych leżących na brzegu obszaru $\Omega_{\square, h}$ oznaczymy symbolem Γ_h .

Z każdym punktem $x = (ih_1, jh_2) \in \Omega_h$ wiążemy funkcję postaci

$$\varphi_{ij}(x_1, x_2) = \varphi_i(x_1)\varphi_j(x_2),$$

gdzie $\varphi_i(x_1)$, $\varphi_j(x_2)$ są funkcjami daszkowymi, tzn.

$$\varphi_i(x_1) = \varphi\left(\frac{x_1}{h_1} - i\right), \quad \varphi_j(x_2) = \varphi\left(\frac{x_2}{h_2} - j\right),$$

a

$$\varphi(s) = \{1 - |s| \text{ dla } |s| \leq 1 \text{ i } 0 \text{ dla } |s| > 1\}.$$

Za przestrzeń V_m , którą dalej będziemy oznaczać przez V_h , przyjmujemy

$$V_h = \text{lin}\{\varphi_{ij}\}_{(ih_1, jh_2) \in \Omega_h}.$$

Funkcje przestrzeni V_h są określone w $\Omega_{\square, h}$ i $V_h(\Omega_{\square, h}) \subset H_0^1(\Omega_{\square, h})$. Aby otrzymać zawieranie się $V_h(\Omega_{\square, h})$ w $H_0^1(\Omega)$, funkcje przestrzeni V_h określone w $\Omega_{\square, h}$ przedłużamy zerem na $\Omega/\Omega_{\square, h}$.

Podamy teraz oszacowanie błędu metody (4.1), (4.2) dla tak określonej przestrzeni V_h .

TWIERDZENIE 2. *Niech będą spełnione założenia twierdzenia 1 oraz*

$$(6.1) \quad \frac{d^2 u}{dt^2} \in L^2(0, T; L^2(\Omega)),$$

gdzie u jest rozwiązaniem (3.1)–(3.3). Wtedy dla $\theta > \theta_0 > 0$ mamy

1° jeśli $\Gamma_h \subset \partial\Omega$, to

$$(6.2) \quad \max_n \|U^n - u^n\| \leq M(\tau + h^2);$$

2° jeśli $\Gamma_h \not\subset \partial\Omega$, to

$$(6.3) \quad \max_n \|U^n - u^n\| \leq M(\tau + h),$$

gdzie U^n jest rozwiązaniem (4.1), (4.2), $h = \max\{h_1, h_2\}$, $M \neq M(h, U^n) > 0$.

D o w ó d. Rozpatrzmy najpierw przypadek, gdy $\Gamma_h \subset \partial\Omega$. Na podstawie twierdzenia 1 oszacowanie błędu rozważanej metody sprowadza się do oszacowania wyrazów prawej strony (5.3). Rozważmy najpierw zadanie pomocnicze (5.1). Z lematów 1 i 2 wynika, że zadanie (5.1) ma jednoznaczne rozwiązania oraz

$$(6.4) \quad \|u^n - W^n\|_{H^1} \leq M \inf_{v \in V_h} \|u^n - v\|_{H^1}, \quad M > 0, \quad n = 0, \dots, N-1.$$

Biorąc za v funkcję $\tilde{u}^n \in V_h$, gdzie \tilde{u}^n jest interpolacją funkcji u^n , z nierówności (6.4) na podstawie twierdzenia o błędzie takiej interpolacji (zob. np. [11]) otrzymujemy

$$(6.5) \quad \|u^n - W^n\|_{H^1} \leq Mh \|u^n\|_{H^2}.$$

W pracy [17] podane jest oszacowanie (nierówność (3.4))

$$\|u^n - W^n\|_{L^2(\Omega)} \leq Mh \|u^n - W^n\|_{H^1(\Omega)},$$

z którego i z (6.5) otrzymujemy

$$(6.6) \quad \|u^n - W^n\|_{L^2(\Omega)} \leq Mh^2 \|u^n\|_{H^2(\Omega)}.$$

Przechodzimy do oszacowania wyrazu

$$(6.7) \quad I = \left(\tau \sum_{n=0}^{N-1} \left\| \left(\frac{du}{dt} \right)^n - W_t^n \right\|^2 \right)^{1/2}.$$

Bezpośrednio sprawdzimy, że

$$(6.8) \quad \left(\frac{du}{dt} \right)^n = u_t^n + \delta^n, \quad \text{gdzie} \quad \|\delta^n\|_{L^2} = O(\tau).$$

Z lematu (3.1) i nierówności (3.9) zawartych w pracy [17] wynika, że $\forall t \in [0, T]$

$$\left\| \frac{d}{dt}(u - W) \right\|_{L^2(\Omega)} \leq Mh^2 \left\{ \|u\|_{H^2(\Omega)} + \left\| \frac{du}{dt} \right\|_{H^2(\Omega)} \right\}.$$

Z tego oszacowania, z (6.8) i nierówności trójkąta otrzymujemy

$$(6.9) \quad I \leq M(\tau + h^2).$$

Podstawiając (6.5), (6.6) i (6.9) do prawej strony (5.3) otrzymujemy oszacowanie (6.2).

Pozostaje rozpatrzyć przypadek, gdy $\Gamma_h \not\subset \partial\Omega$. Funkcje przestrzeni V_h przedłużamy zerem na $\Omega/\Omega_{\square, h}$. Rozwiązanie U^n równania (4.1) przy tak określonej przestrzeni V_h jest elementem $H_0^1(\Omega)$. Dalej, dla różnicy $U^n - u^n$ określonej już w $\bar{\Omega}$ pozostaje prawdziwe oszacowanie (5.3) z twierdzenia 1. Teraz wybierzemy funkcję v występującą w prawej stronie nierówności (6.4). Niech $v = \hat{u}^n$ będzie interpolacją funkcji \bar{u}^n równej u^n dla $x = \Omega_h$ (przy ustalonym n), a dla pozostałych x równej zeru. Uwzględniając powyższe, otrzymujemy

$$(6.10) \quad \|u^n - u^n\|_{H^1(\Omega)} \leq \|u^n\|_{H^1(\Omega/\Omega_{\square, h})} + \|u^n - \hat{u}^n\|_{H^1(\Omega_{\square, h})}.$$

Do oszacowania pierwszego wyrazu lewej strony (6.10) wykorzystujemy nierówność z [14] postaci

$$(6.11) \quad \|u\|_{L^2(\Omega_\delta)} \leq \delta^{1/2} \|u\|_{H^1(\Omega)},$$

gdzie Ω_δ jest „paskiem” przybrzegowym obszaru Ω szerokości δ . Stąd i z wyżej wspomnianego twierdzenia o aproksymacji otrzymujemy

$$(6.12) \quad \|u^n - \hat{u}^n\|_{H^1(\Omega)} \leq Mh^{1/2} \|u^n\|_{H^2(\Omega)} + h \|u^n\|_{H^2(\Omega)}.$$

Podobnie dowodzimy, że

$$(6.13) \quad \|u^n - u^n\|_{L^2} \leq M \{ h \|u^n\|_{H^2} + h^2 \|u^n\|_{H^2(\Omega)} \}$$

oraz

$$(6.14) \quad I \leq M(\tau + h).$$

Podstawiając (6.12)–(6.14) do nierówności (5.3) otrzymujemy (6.3), co kończy dowód twierdzenia 2.

7. Algorytm. Przechodzimy do określenia układu równań algebraicznych odpowiadającego zadaniu (4.1) (4.2) przy wyżej wybranych funkcjach bazowych. Następnie podamy algorytm rozwiązywania tego układu.

I tak, rozwiązania $U^n \in V_h$ dla ustalonego n poszukujemy w postaci

$$(7.1) \quad U^n = \sum_{(i,j) \in \Omega_h} \xi_{ij}^n \varphi_{ij}, \quad \xi_{ij}^n \in R,$$

gdzie $(i,j) \in \Omega_h$ tutaj i dalej oznacza, że $(ih_1, jh_2) \in \Omega_h$ i sumowanie w (7.1) przebiega wszystkie punkty Ω_h . Podstawiając (7.1) do równania (4.1) otrzymujemy

$$(7.2) \quad \sum_{(i,j) \in \Omega_h} \{ \xi_{ij}^n (\varphi_{ij}, \varphi_{pq}) + \theta \tau (\nabla \varphi_{ij}, \nabla \varphi_{pq}) + \theta^2 \tau^2 (D_1 D_2 \varphi_{ij}, D_1 D_2 \varphi_{pq}) \} = F_{pq}^n,$$

$(p, q) \in \Omega_h,$

gdzie

$$F_{pq}^n \equiv (f^n, \varphi_{pq}) - a^n(U^n, \nabla U^n; \varphi_{pq}).$$

Przekształcimy iloczyny skalarne występujące po lewej stronie (7.2) z wykorzystaniem postaci funkcji bazowych. Otóż

$$\begin{aligned} \sum_{(i,j) \in \Omega_h} (\varphi_{ij}, \varphi_{pq}) &= \sum_{r=-1}^1 \sum_{s=-1}^1 (\varphi_{p+r, p+s}, \varphi_{pq}) = \\ &= \sum_{r=-1}^1 \sum_{s=-1}^1 (\varphi_{p+r}, \varphi_p)_{x_1} (\varphi_{q+s}, \varphi_q)_{x_2} = \\ &= \sum_{r=-1}^1 (\varphi_{p+r}, \varphi_p)_{x_1} \sum_{s=-1}^1 (\varphi_{q+s}, \varphi_q)_{x_2}, \end{aligned}$$

gdzie

$$\begin{aligned} (\varphi_p, \varphi_p)_{x_1} &= \int_{(p-1)h_1}^{(p+1)h_1} \varphi_p^2(x_1) dx_1 = \frac{2}{3} h_1, \\ (\varphi_{p+1}, \varphi_p)_{x_1} &= \int_{ph_1}^{(p+1)h_1} \varphi_{p+1} \varphi_p dx_1 = \frac{1}{6} h_1, \\ (\varphi_{p-1}, \varphi_p)_{x_1} &= \int_{(p-1)h_1}^{ph_1} \varphi_{p-1} \varphi_p dx_1 = \frac{1}{6} h_1; \end{aligned}$$

podobnie określamy $(\cdot, \cdot)_{x_2}$.

Wprowadźmy pomocnicze operatory różnicowe postaci

$$(7.3) \quad \begin{aligned} \bigwedge_p^{(1)} &\equiv \sum_{r=-1}^1 (\varphi_{p+r}, \varphi_p)_{x_1} I_1^r = \frac{1}{6} h_1 (I_1^1 + 4I_1^0 + I_1^{-1}), \\ \bigwedge_q^{(2)} &\equiv \sum_{s=-1}^1 (\varphi_{q+s}, \varphi_q)_{x_2} I_2^s = \frac{1}{6} h_2 (I_2^1 + 4I_2^0 + I_2^{-1}), \end{aligned}$$

gdzie $I_1^r v_{ij} = v_{i+r,j}$, $I_2^s v_{ij} = v_{i,j+s}$. Wykorzystując je, pierwszy wyraz lewej strony (7.2) zapisujemy w postaci

$$(7.4) \quad \sum_{(i,j) \in \Omega_h} \xi_{ij,t}^n(\varphi_{ij}, \varphi_{pq}) = \bigwedge_p^{(1)} \bigwedge_q^{(2)} \xi_{pq,t}^n.$$

Podobne postępowanie drugi wyraz lewej strony (7.2) doprowadza do postaci

$$(7.5) \quad \sum_{(i,j) \in \Omega_h} \theta \tau \xi_{ij,t}^n(\nabla \varphi_{ij}, \nabla \varphi_{pq}) = \theta \tau (\tilde{\bigwedge}_p^{(1)} \bigwedge_q^{(2)} + \tilde{\bigwedge}_q^{(2)} \bigwedge_p^{(1)}) \xi_{pq,t}^n,$$

gdzie

$$\tilde{\bigwedge}_p^{(1)} \equiv \sum_{r=-1}^1 (D_1 \varphi_{p+r}, D_1 \varphi_p)_{x_1} I_1^r = -\frac{1}{h_1} (I_1^1 - 2I_1^0 + I_1^{-1}),$$

$$\tilde{\bigwedge}_q^{(2)} = \sum_{s=-1}^1 (D_2 \varphi_{q+s}, D_2 \varphi_q)_{x_2} I_2^s = -\frac{1}{h_2} (I_2^1 - 2I_2^0 + I_2^{-1}).$$

Ostatni wyraz lewej strony (7.2) daje się przedstawić w postaci

$$(7.6) \quad \sum_{(i,j) \in \Omega_h} \theta^2 \tau^2 (D_1 D_2 \varphi_{ij}, D_1 D_2 \varphi_{pq}) \xi_{pq,t}^n = \theta^2 \tau^2 \tilde{\bigwedge}_p^{(1)} \tilde{\bigwedge}_q^{(2)} \xi_{pq,t}^n.$$

Podstawiając (7.4)–(7.6) do równania (7.2), otrzymujemy schemat różnicowy, który daje się zapisać w postaci

$$(7.7) \quad (\bigwedge_p^{(1)} + \theta \tau \tilde{\bigwedge}_p^{(1)}) (\bigwedge_q^{(2)} + \theta \tau \tilde{\bigwedge}_q^{(2)}) \xi_{pq,t}^n = F_{pq}^n, \quad (p, q) \in \Omega_h,$$

$$(7.8) \quad \xi_{pq}^n = 0, \quad (p, q) \in \Gamma_h.$$

Przechodzimy teraz do określenia algorytmu rozwiązywania schematu (7.7), (7.8). Algorytm ten będzie istotnie zależał od kształtu obszaru Ω . Rozpatrzmy kilka przypadków.

P r z y p a d e k 1. *Obszar Ω jest prostokątem postaci*

$$\Omega = (0, l_1) \times (0, l_2).$$

Na $\bar{\Omega}$ konstruujemy siatkę Ω_h tak, aby $\Gamma_h \subset \partial\Omega$ ($\partial\Omega$ — brzeg obszaru Ω). Niech

$$\Omega_h = \{x = (ih_1, jh_2), i = 1, \dots, N_1, j = 1, \dots, N_2, h_p(N_p + 1) = l_p, p = 1, 2\}$$

W tym przypadku rozwiązywanie (7.7), (7.8) dla ustalonych n i obliczonych prawych stron F_{pq}^n sprowadza się do trzech etapów.

I etap: rozwiązanie N_2 układów ($q = 1, \dots, N_2$)

$$(\bigwedge_p^{(1)} + \theta \tau \tilde{\bigwedge}_p^{(1)}) v_{pq}^n = F_{pq}^n, \quad p = 1, \dots, N_1,$$

$$v_{pq}^n \equiv (\bigwedge_q^{(2)} + \theta \tau \tilde{\bigwedge}_q^{(2)}) \xi_{pq,t}^n = 0; \quad p = 0; p = N_1 + 1.$$

II etap: rozwiązanie N_1 układów ($p = 1, \dots, N_2$)

$$(\bigwedge_q^{(2)} + \theta \tau \tilde{\bigwedge}_q^{(2)}) \xi_{pq,t}^n = v_{pq}^n, \quad q = 1, \dots, N_1,$$

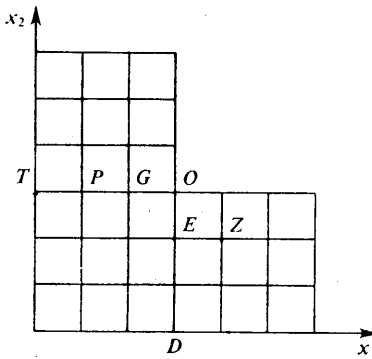
$$\xi_{pq,t}^n = 0, \quad q = 0, q = N_2 + 1.$$

III etap: wyznaczenie ξ_{pq}^{n+1} ze wzoru

$$\xi_{pq}^{n+1} = \xi_{pq}^n + \tau \xi_{pq,t}^n, \quad p = 1, \dots, N_1, \quad q = 1, \dots, N_2.$$

Każdemu z układów w etapach I i II odpowiada układ równań algebraicznych z macierzą trójprzekątniową. Do rozwiązywania tych układów wykorzystujemy algorytm faktoryzacji („progonki”) numerycznie stabilny (zob. np. [8], str. 618), którego realizacja na maszynie cyfrowej kosztuje liczbę działań arytmetycznych proporcjonalną do liczby niewiadomych. Przy takiej realizacji koszt rozwiązania schematu różnicowego (7.7), (7.8) wynosi $O(N \times N_1 \times N_2)$ działań, jeżeli koszt obliczania prawych stron jest tego samego rzędu. Tak więc algorytm ten jest optymalny w sensie minimalizacji liczby działań arytmetycznych potrzebnych do realizacji metody na maszynie cyfrowej.

Przypadek II. Obszar Ω ma kształt litery L (zob. rys. 1)



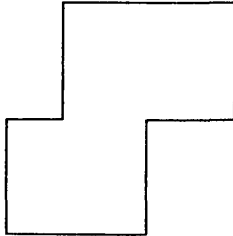
Rys. 1

Siatkę Ω_h konstruujemy tak, aby $\Gamma_h \subset \partial\Omega$. W tym przypadku algorytm z przypadku I nie daje się bezpośrednio zastosować, a to z tego względu, że w punkcie O nie możemy obliczyć v_{pq}^n , czyli wartości brzegowej potrzebnej do realizacji etapu I, bo nie znamy $\xi_{pq,t}^n$ w punkcie E . Dlatego modyfikujemy powyższy algorytm, stosując podejście Djakonova [9] wykorzystywane w schematach różnicowych z rozszczepiającym się operatorem. Modyfikacja ta jest następująca. Realizujemy etap I dla wszystkich q (po warstwach poziomych) z wyjątkiem q odpowiadającego punktom siatkowym na warstwie TO . Następnie przechodzimy do realizacji etapu II, ale tylko dla tego p , które odpowiada warstwie pionowej DO . To pozwala nam wyznaczyć $\xi_{pq,t}^n$ wśród innych w punkcie E . Teraz wracamy do realizacji etapu I na opuszczonej warstwie TO , a następnie do etapu II, obliczając $\xi_{pq,t}^n$ na pozostałych warstwach pionowych. Etap III jest taki sam jak w przypadku prostokąta.

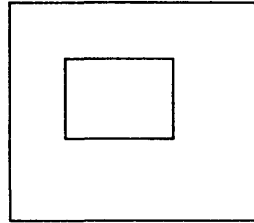
Koszt realizacji metody w tym przypadku również jest proporcjonalny do liczby niewiadomych równej liczbie punktów siatki Ω_h . Powyższy algorytm przenosi się tylko na pewne obszary składające się z prostokątów o bokach równoległych do osi współrzędnych (przykłady na rys. 2 i 3).

Na rys. 5 przedstawiony jest obszar, dla którego powyższy algorytm nie daje się zrealizować. Również na ogół nie możemy go wykorzystać w obszarze dowolnym Ω , aproksymując ten obszar obszarem składającym się z prostokątów.

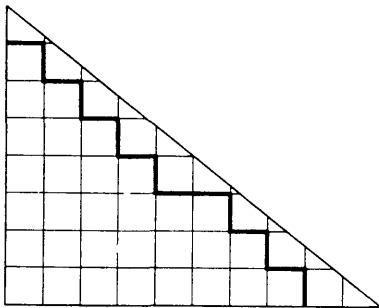
Niżej przedstawimy algorytm rozwiązywania (7.8), (7.9), który możemy stosować w dowolnych obszarach. Oparty jest on na tzw. algorytmie z macierzą „po-



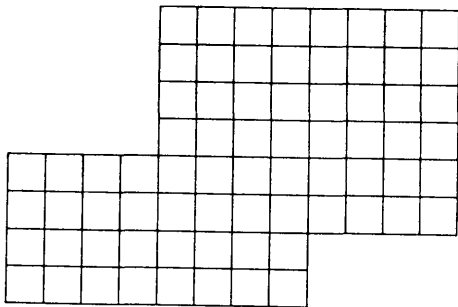
Rys. 2



Rys. 3



Rys. 4



Rys. 5

jemnościową” (*a capacitance matrix*, [18]) rozwiązywania układów równań algebraicznych. Najpierw przedstawimy ten algorytm dla dowolnego układu

$$(7.9) \quad Ax = y$$

gdzie A jest macierzą nieosobliwą wymiaru $M \times M$, a x i y są wektorami M -wymiarowymi. Niech K będzie zbiorem o postaci

$$K := \{1, 2, \dots, N\},$$

a S jego podzbiorem, $S \subset K$, zawierającym p elementów. Niech B oznacza macierz nieosobliwą również wymiaru $M \times M$, różniącą się od macierzy A tylko p wierszami, dokładniej: j -ty wiersz $A = j$ -ty wiersz B , $\forall j \in (K \setminus S)$.

Niech \tilde{x} będzie rozwiązaniem układu

$$B\tilde{x} = y.$$

Rozwiązania x układu (7.9) poszukujemy w postaci

$$(7.10) \quad x = \tilde{x} + \sum_{i \in S} \beta_i g_i,$$

gdzie $\beta_i \in \mathbb{R}$, g_i są wektorami.

Bezpośrednio sprawdzamy, że g_i są rozwiązaniami układów

$$Bg_i = e_i, \quad i \in S,$$

gdzie e_i jest wektorem mającym składową i równą 1, pozostałe zaś równe zero, a wektor $\beta = \{\beta_i\}_{i \in S}$ jest rozwiązaniem układu

$$C\beta = \gamma,$$

gdzie $C = \{(Ag_i)_j\}$, $i, j \in S$, $\gamma = \{(y)_j - (A\tilde{x})_j\}$, $j \in S$. Macierz C wymiaru $p \times p$, nazywana macierzą „pojemnościową”, jest nieosobliwa, co wynika z nieosobliwości macierzy A i B . Mając wyznaczone wektory g_i i β rozwiązanie x obliczamy albo z (7.10), albo rozwiązując układ

$$Bx = y + \sum_{i \in S} \beta_i e_i;$$

to ostatnie jest wygodniejsze ze względu na pamięć maszyny.

Schemat postępowania jest więc następujący:

1. Rozwiązać układy $Bg_i = e_i$, $i \in S$,
2. Obliczyć $C = \{(Ag_i)_j\}$, $i, j \in S$,
3. Rozwiązać $B\tilde{x} = y$,
4. Rozwiązać $C\beta = \gamma$,
5. Rozwiązać $Bx = y + \sum_{i \in S} \beta_i e_i$.

Przechodzimy teraz do wykorzystania powyższego algorytmu do rozwiązania schematu różnicowego (7.7), (7.8). Najpierw rozpatrzmy przypadek, gdy Ω jest obszarem takim jak na rys. 1.

Schemat różnicowy (7.7), (7.8) napiszemy w postaci operatorowej (macierzowej). W związku z tym wprowadzamy przestrzeń $L_h^2(\Omega_h)$ — przestrzeń Hilberta funkcji siatkowych określonych na Ω_h :

$$(u, v)_{L_h^2(\Omega_h)} = \sum_{x \in \Omega_h} h_1 h_2 u(x)v(x).$$

W przestrzeni $L_h^2(\Omega_h)$ schemat (7.7), (7.8) zapisujemy w następującej postaci operatorowej:

$$(7.11) \quad A\xi_r^n = F^n, \quad n = 0, \dots, N-1,$$

$$(7.12) \quad \xi^0 = \psi,$$

gdzie $(Av)(x) = (\bar{A}v)(x)$, $x \in \Omega_h$, zaś $\bar{A}: L_h^2(\bar{\Omega}_h) \rightarrow L_h^2(\Omega_h)$ ma postać

$$(\bar{A}v)(x) = \{(\wedge_p^{(1)} + \theta\tau \tilde{\wedge}_p^{(1)}) (\wedge_q^{(2)} + \theta\tau \tilde{\wedge}_q^{(2)})v(x), \\ x = (ph_1, qh_2) \in \Omega_h; v(x) = 0, x \in \Gamma_h\},$$

$$\bar{\Omega}_h = \Omega_h U\Gamma_h, \quad (\xi_r^n)(x) = \xi_r^n(x) = \xi_{pq,t}^n, \quad (F^n)(x) = F^n(x), \quad x \in \Omega_h.$$

Funkcję siatkową ψ wyznaczamy z warunku początkowego (4.2); za ψ możemy przyjąć $(\psi)(x) = u_0(x)$, jeśli tylko u_0 jest funkcją ciągłą, co oznacza, że U^0 jest interpolacją funkcji u_0 opartą na funkcjach bazowych φ_{ij} .

Zauważmy, że operator A nie jest rozszczepialny. Za operator B przyjmujemy

$$(7.13) \quad B = B_1 B_2 \equiv (\bar{B}_1 + \theta\tau\tilde{B}_1) (\bar{B}_2 + \theta\tau\tilde{B}_2),$$

gdzie

$$(\bar{B}_1 v)(x) = (\bigwedge_p^{(1)} v)(x), \quad (\tilde{B}_1 v)(x) = (\tilde{\bigwedge}_p^{(1)} v)(x), \quad x \in \Omega_h;$$

podobnie określamy \bar{B}_2 i \tilde{B}_2 .

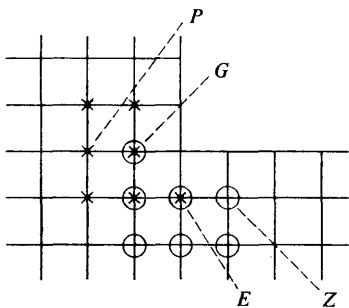
Operatory B , B_i , jak i wyżej określony operator A , są samosprężone i dodatnio określone w $L_h^2(\Omega_h)$.

Ustalimy teraz różnice między operatorami A i B . Rozpisując te operatory i opuszczając indeksy p i q przy operatorach oznaczonych symbolem \bigwedge , mamy

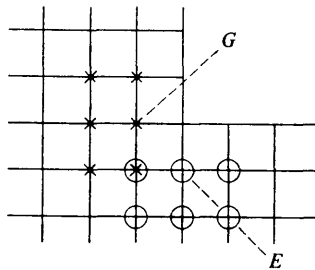
$$(Av)(x) = (\bigwedge^{(1)} \bigwedge^{(2)} + \theta\tau(\tilde{\bigwedge}^{(1)} \bigwedge^{(2)} + \bigwedge^{(1)} \tilde{\bigwedge}^{(2)} + \theta^2\tau^2 \tilde{\bigwedge}^{(1)} \tilde{\bigwedge}^{(2)})v(x),$$

$$(Bv)(x) = ((\bar{B}_1 \bar{B}_2 + \theta\tau(\tilde{B}_1 \bar{B}_2 + \bar{B}_1 \tilde{B}_2) + \theta^2\tau^2 \tilde{B}_1 \tilde{B}_2)v)(x).$$

Postacie tych operatorów różnią się w punktach G i E (rys. 1); z definicji powyższych operatorów wynika, że w punktach G i E (rys. 6 i 7) iloczyny operatorów występujących w A są określone na 7-punktowym otoczeniu, a w operatorze B — na 6-punktowym.



Rys. 6



Rys. 7

\times — punkty otoczenia punktu G , a \odot — punktu E .

Schemat rozwiązywania (7.11) (dla ustalonego n) oparty na algorytmie z macierzą „pojemnościową” jest następujący:

1. Rozwiązać układy

$$Bg_G = e_G, \quad Bg_E = e_E,$$

gdzie

$$(g_G)(x) = \{1 \text{ dla } x = G \text{ i } 0 \text{ dla } x \neq G\},$$

$$(g_E)(x) = \{1 \text{ dla } x = E \text{ i } 0 \text{ dla } x \neq E\}.$$

2. Obliczyć elementy macierzy $C = \{c_{ij}\}_{i,j=1,2}$:

$$c_{11} = (Ag_G)(G), \quad c_{12} = (Ag_G)(E),$$

$$c_{21} = (Ag_E)(G), \quad c_{22} = (Ag_E)(E).$$

3. Rozwiązać układ

$$B\tilde{\xi}_i^n = F^n.$$

4. Rozwiązać układ

$$C\beta = \gamma,$$

gdzie $\beta = (\beta_1, \beta_2)$, $\gamma = (\gamma_1, \gamma_2)$; $\gamma_1 = (F^n)(G) - (A\tilde{\xi}_i^n)(G)$, $\gamma_2 = (F^n)(E) - (A\tilde{\xi}_i^n)(E)$.

5. Rozwiązać układ

$$B\xi_i^n = F^n + \beta_1 e_G + \beta_2 e_E.$$

Rozwiązywanie układów z operatorem (macierzą) B sprowadza się do rozwiązywania dwóch układów z operatorami B_1 i B_2 , którym odpowiadają macierze trójkątne. Koszt realizacji jest więc rzędu M , gdzie M oznacza liczbę punktów siatki Ω_h . Zauważmy, że w przedstawionym schemacie nie ma potrzeby rozwiązywać całkowicie układów z macierzą B w etapach 1 i 3; rozwiązania g_G, g_E i $\tilde{\xi}_i^n$ są potrzebne jedynie do obliczenia macierzy C i wektora γ , wystarczy więc je obliczyć w punktach otoczenia 7-punktowego punktów G i E (zob. rys. 6). Na przykład, aby obliczyć g_G w wymienionych punktach, wystarczy rozwiązać układ $B_1 v = e_G$ tylko na warstwie poziomej, na której leży punkt G , a następnie rozwiązać układ $B_2 e_G = v$ na warstwach pionowych zawierających punkty E, G i P (zob. rys. 6); na warstwie pionowej zawierającej punkt Z mamy $e_G = 0$, gdyż $v = 0$ na tej warstwie.

Zauważmy jeszcze, że w przedstawionym algorytmie etapy 1 i 2 nie zależą od n (od warstwy czasowej), więc realizujemy je tylko jeden raz, a obliczoną macierz C należy zapamiętać.

P r z y p a d e k III. *Obszar Ω jest dowolny.* W tym przypadku Ω aproksymujemy obszarem $\Omega_{\square, h}$, składającym się z prostokątów o bokach h_1 i h_2 . Na obszarze $\Omega_{\square, h}$ konstruujemy siatkę Ω_h , która pokrywa się z siatką określoną w ustępie 6 (zob. rys. 4). Tak więc zadanie przybliżone daje się sprowadzić do schematu różnicowego postaci (7.7), (7.8). Następnie schemat ten sprowadzamy do postaci operatorowej (7.11), (7.12) z operatorem A . Za operator B przyjmujemy tak jak i wyżej, operator postaci

$$B = B_1 B_2 \equiv (\bar{B}_1 + \theta \tau \tilde{B}_1) (\bar{B}_2 + \theta \tau \tilde{B}_2).^{-1}$$

Operatory A i B różnią się w punktach sąsiednich punktów „narożnych” — takich jak punkt 0 na rys. 1 i podobnych do niego. Dalej postępujemy według wyżej przedstawionego schematu. W ogólnym przypadku macierz C będzie wymiaru $2p \times 2p$, gdzie p jest liczbą punktów „narożnych”. Układy z macierzą C rozwiązujemy metodą Gaussa. Ponieważ C nie zależy od n , więc przedstawiamy ją w postaci LU , gdzie L, U są macierzami trójkątnymi, odpowiednio dolną i górną. Koszt tego przedstawienia jest rzędu p^3 działań arytmetycznych.

Obliczmy koszt wyznaczenia rozwiązania przybliżonego ξ^n , $n = 1, \dots, N$. W przypadku gdy liczba punktów „narożnych” p nie zależy od kroków siatki, koszt ten jest rzędu MN , gdzie M jest liczbą punktów siatki Ω_h . W pozostałych przypadkach jest on rzędu $MN + p^3$. W najgorszym przypadku p jest rzędu \sqrt{M} — liczby punktów brzegowych siatki Γ_h . Przyjmując, że N jest proporcjonalne do \sqrt{M} , tzn. liczba punk-

tów siatkowych po t jest tego samego rzędu co liczba punktów po h_1 lub po h_2 , mamy, że koszt realizacji metody jest rzędu MN , a więc przedstawiony algorytm jest optymalny w sensie minimalizacji liczby działań arytmetycznych.

U w a g a. W przedstawionym algorytmie rozwiązywania (7.8), (7.9) jako operator B można wybrać operator, który różni się od operatora A tylko w p -punktach, gdzie p jest liczbą punktów „narożnych”, np. w przypadku obszaru z rys. 1 tylko w punkcie G .

8. Zagadnienia z nieliniowościami nieograniczonymi. Klasa zagadnień (2.1), (2.2) spełniających założenia (2.3), (2.4), tzn. z nieliniowościami ograniczonymi, jest dość wąska; wiele zagadnień nieliniowych ważnych w praktyce nie spełnia tych założeń i dla tych zagadnień nie możemy zagwarantować zbieżności metody ADG rozważanej w poprzednich ustępach. Istnieje jednak możliwość rozszerzenia metody ADG na pewną klasę zagadnień z nieliniowościami nieograniczonymi, wykorzystując podejście z pracy [4], które niżej przedstawimy.

Rozważamy zagadnienie (2.1), (2.2). Zakładamy, że w miejsce warunku (2.3) spełniony jest następujący warunek:

∃ funkcje $\gamma_0(\tau)$, $\gamma_1(\tau)$, $0 \leq \tau < +\infty$ dodatnie i ciągłe wraz z pierwszą pochodną, takie, że $\forall (x, t) \in Q_T = \bar{\Omega} \times [0, T]$, $\forall p' \in R^3$ $\forall \xi \in R^2$, $\xi \neq 0$

$$(8.1) \quad \gamma_0(|p|) \sum_{i=1}^2 \xi_i^2 \leq \sum_{i,j=0}^2 \frac{\partial a_i}{\partial p_j} \xi_i \xi_j \leq \gamma_1(|p|) \sum_{i=1}^2 \xi_i^2.$$

Zamiast założenia (2.4) przyjmujemy następujące: niech funkcje $a_i(x, p')$, $i = 0, 1, 2$, będą ciągłe wraz z pierwszymi pochodnymi w $Q_T \times R^3$ oraz

$$(8.2) \quad |a_i(x, t, 0, 0)| \leq M, \quad \left| \frac{\partial a_i}{\partial x_k} \right| \leq M, \quad \left| \frac{\partial a_i}{\partial t} \right| \leq M$$

dla $i, j = 0, 1, 2$, $k = 1, 2$, $\forall (x, t) \in Q_T$, $0 < M = \text{const}$. Jak widzimy, w (8.2) nie zakłada się ograniczoności $\partial a_i / \partial p_j$, co miało miejsce w (2.4).

Równanie (2.1) z warunkami (8.1), (8.2) przyjęto nazywać *równaniem z nieliniowościami nieograniczonymi*.

Dalej będziemy zakładać, że rozwiązania (2.1), (2.2) spełniają następujący warunek

$$(8.3) \quad (|u|^2 + |\nabla u|^2)^{1/2} \leq M.$$

Klasa zagadnień, spełniających ten warunek, jest określona w [6]. Okazuje się, że jeśli rozwiązania zagadnienia (2.1), (2.2) z nieliniowościami nieograniczonymi spełniają warunek (8.3), to dla zagadnienia (2.1) (2.2) można skonstruować zagadnienie pomocnicze z nieliniowościami ograniczonymi, równoważne (2.1), (2.2), tzn. rozwiązanie nowego zagadnienia pokrywa się z rozwiązaniem (2.1), (2.2). Mając skonstruowane zagadnienie pomocnicze rozwiązujemy je metodą ADG opisaną w poprzednich ustępach i rozwiązanie przybliżone zagadnienia pomocniczego przyjmujemy za rozwiązanie przybliżone zagadnienia wyjściowego. Rozważania

z poprzednich ustępów gwarantują zbieżność tego rozwiązania do zadania wyjściowego oraz dają oszacowanie błędu.

Przechodzimy do konstrukcji zagadnienia pomocniczego dla (2.1), (2.2).

Niech $\xi(s)$ oznacza funkcję określoną na $[0, \infty)$ postaci

$$\xi(s) = \begin{cases} 1, & s \leq 1, \\ 0, & s \geq 2, \end{cases}$$

a na odcinku $[1, 2]$ tak określoną, aby funkcja $\xi(s)$ na $[0, \infty)$ była nierosnąca, nieujemna, ciągła wraz z drugimi pochodnymi. Wprowadzimy jeszcze jedną funkcję

$$\theta(s) = c\gamma_1^2(s)/\gamma_0(s),$$

gdzie $\gamma_i(s)$ są funkcjami z warunku (8.1), a c jest pewną stałą dodatnią.

Mając te funkcje określamy $\tilde{a}_i(x, t, p')$ na podstawie $a_i(x, t, p')$ w następujący sposób:

$$\tilde{a}_i(x, t, p') = \begin{cases} a_i(x, t, p') \xi \left(\left| \frac{p'}{M} \right|^2 \right) + \theta(|p'|^2) p_i \left(1 - \xi \left(\left| \frac{p'}{M} \right|^2 \right) \right) & \text{dla } |p'| \leq 2M, \\ \theta(4M^2) p_i & \text{dla } |p'| \geq 2M, \end{cases}$$

gdzie $i = 0, 1, 2$, a M jest stałą z warunku (8.3). Określamy zagadnienie pomocnicze: znaleźć taką $u(x, t)$, żeby

$$(8.4) \quad \frac{\partial u}{\partial t} = \sum_{i=1}^2 \frac{d}{dx_i} \tilde{a}_i(x, t, u, D_1 u, D_2 u) + \tilde{a}_0(x, t, u, D_1 u, D_2 u) + f(x, t),$$

$$x \in \Omega, t \in (0, T]$$

$$(8.5) \quad u(x, 0) = u_0(x), \quad u(x, t) = 0 \quad x \in \partial\Omega, t \in (0, T].$$

Dowodzi się, że współczynniki $\tilde{a}_i(x, t, p')$ równania (8.4) przy założeniach (8.1)–(8.3) spełniają warunki (2.3), (2.4) (zob. [4]).

Zagadnienie (8.4), (8.5) formułujemy w postaci wariacyjnej (3.1)–(3.3), gdzie w miejscu $a_i(x, t, p')$ występują $\tilde{a}_i(x, t, p')$. Można wykazać, że rozwiązanie tak sformułowanego zagadnienia wariacyjnego pokrywa się z rozwiązaniem zagadnienia (3.1)–(3.3). Tak sformułowane zagadnienie wariacyjne pomocnicze jest wyjściowe dla konstrukcji metody ADG.

9. Uogólnienia. 1. Powyższe rozważania dają się przenieść na zagadnienia dla równań hiperbolicznych rzędu II, układów równań parabolicznych i hiperbolicznych rzędu I i II oraz dla układów paraboliczno-hiperbolicznych (zob. [12], [13]).

2. Rozważana metoda przenosi się na przypadek $n > 2$, przy czym jest ona optymalna tylko dla obszarów składających się z n -wymiarowych prostokątów.

3. Analogiczne wyniki można uzyskać także dla pewnych schematów trójwarstwowych i schematów dwuwarstwowych innej postaci niż (4.1)–(4.2). W szcze-

gólności, dla pewnych zagadnień można konstruować metodę ADG podwyższonej aproksymacji względem t , np. schemat Crank-Nicolsona.

4. Za funkcje bazowe w ADG można wybierać iloczyny funkcji jednej zmiennej, z których każda jest wielomianem stopnia wyższego niż 1, np. iloczyny wielomianów stopnia 2 jednej zmiennej, co daje nam szybszą zbieżność metody.

5. Powyższe rozważania dają się przenieść na zagadnienia z warunkami początkowo-brzegowymi, w których warunki brzegowe mogą być II i III rodzaju lub ich kombinacjami z warunkiem Dirichleta.

Prace cytowane

- [1] Е. Г. Дьяков, *Разностные методы решения краевых задач*, II вып., Москва 1973.
- [2] I. Douglas, Jr., T. Dupont, *Alternating direction Galerkin methods on rectangles*, Proc. Symposium on Numerical Solution PDE, II (B. Hubbard, ed.), Academic Press, New York (1971), str. 133–214.
- [3] J. Dendy, G. Fairweather, *Alternating direction Galerkin methods for parabolic and hyperbolic problems on rectangular polygons*, SIAM J. Numer. Anal. 12.2 (1975).
- [4] В. Я. Ривкинд, Н. Н. Уралцева, *Априорные оценки и их применение для приближенного метода решения параболических уравнений*, ДАН СССР 2 (1969), str. 271–274.
- [5] I. L. Lions and E. Magenes, *Problèmes aux limites non homogènes et applications*, Paris 1968 (jest tłumaczenie w języku rosyjskim).
- [6] О. А. Ладыженская, В. А. Солонников, Н. Н. Уралцева, *Линейные и квазилинейные уравнения параболического типа*, Москва 1967.
- [7] M. F. Wheeler, *A priori L_2 error estimates for Galerkin approximations to parabolic PDE*, SIAM J. Numer. Anal., 10 (1973), str. 723–759.
- [8] Н. С. Бахвалов, *Численные методы*, т. I, Москва 1973.
- [9] Е. Г. Дьяконов, *О применении разностных расщепляющихся операторов*, ЭВМ и МФ 3.2 (1963), str. 385–388.
- [10] М. Дрыя, *Абсолютная устойчивость разностных схем с расщепляющимся оператором для параболических и гиперболических уравнений в выуклых областях*, ЖВМ и МФ, 15.4 (1975), str. 966–976.
- [11] G. Strang, G. J. Fix, *An analysis of the finite element method*, Prentice-Hall, New York 1973.
- [12] М. Дрџа, *Difference and finite-element methods for the dynamical problem of thermo-diffusion in an elastic solid*, Archives of Mechanics 29, 1 (1977), str. 81–96.
- [13] — *Alternating direction Galerkin methods for the system of hyperbolic equations of first order*, SIAM J. Numer. Anal. 15.1 (1978).
- [14] В. П. Ильин, *Некоторые неравенства в функциональных пространствах и их применение к исследованию сходимости вариационных процессов*, Тр. Мат. ин-та СССР 53 (1959).
- [15] Ph. Ciarlet, *Numerical analysis of the finite element methods*, 1977.
- [16] M. Zlamal, *Finite element methods for nonlinear parabolic equations*, R.A. I. R.O. Numerical Analysis 11.1 (1977), str. 93–107.
- [17] O. Axelsson, *Error estimates for Galerkin methods for quasilinear parabolic and elliptic differential equations in divergence form*, Numer. Math. 28 (1977), str. 1–14.
- [18] J. Bunch, D. Rose, *Sparse matrix computations*, Academic Press, 1976.