Convergence analysis of a finite difference scheme for viscous gas problems.

A. V. Popov
Computational Mathematics subdepartment
of Mechanics-Mathematics department.
Moscow State University.
Vorobyevy Gory, Moscow 119899, Russia.
E-mail: popov_av@nw.math.msu.su

Abstract

The boundary value problem for the system of equations describing a non-stationary flow of viscous compressible gas is considered. An implicit finite difference scheme approximating the problem is used. Estimates are proved for the numerical solution including the error, which appears when the systems of tridiagonal algebraic equations are solved numerically.

1 Introduction

The two-dimensional non-stationary flow of a viscous barotropic gas is described with the system of equations

$$\begin{aligned}
&\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho \mathbf{u}) = 0, \\
&\rho \left[\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u}, \nabla)\mathbf{u} \right] + \nabla p = L\mathbf{u} + \rho \mathbf{f}, \qquad (1.1) \\
&p = p(\rho),
\end{aligned}$$

where L is a linear elliptic operator

$$L\mathbf{u} \equiv \operatorname{div}(\mu \nabla \mathbf{u}) + \frac{1}{3}\nabla(\mu \operatorname{div}\mathbf{u}).$$

Here $\mu > 0$ is a constant of viscosity.

The values sought, i.e., the density ρ and the velocity vector $\mathbf{u}$, are functions of Euler variables

$$(t, \mathbf{x}) \in Q = [0, T] \times \Omega.$$

Besides, the equations contain two known functions p and $\mathbf{f}$.

Let the system (1.1) be complemented with initial and boundary conditions

$$\begin{aligned} &(\rho, \mathbf{u})|_{t=0} = (\rho_0, \mathbf{u}_0),\ \mathbf{x} \in \Omega, \\ &\mathbf{u}(t, \mathbf{x}) = \mathbf{0},\ (t, \mathbf{x}) \in [0, T] \times \partial\Omega. \end{aligned} \tag{1.2}$$

Numerical methods used for the numerical solution of problems of viscous compressible dynamics are as a rule not substained mathematically. See for example [1, 2, 3]. Mathematically justified results are obtained only for the case of a one-dimensional viscous compressible gas in mass Lagrange variables [4]. However, this technique can not be generalized to include the case of several space variables. The most one can do in this direction is to construct difference schemes and to obtain estimates of accuracy [5] for them assuming that a solution of the differential problem exists and is sufficiently smooth. A weakly stable difference scheme approximating the original problem was constructed in [6]. A solution of this scheme exists and is unique.

In this paper, the difference scheme suggested in [5] is considered. The estimates of the accuracy with regard to round-off computer errors is obtained when the sweep method is used to solve algebraic systems.

2 Difference Scheme

Let us consider the domain $\Omega = [0; X_1] \times [0; X_2]$ and in the domain Q introduce a mesh $Q_{\tau\mathbf{h}} = \bar{\omega}_\tau \times \bar{\omega}_\mathbf{h}$, where $\bar{\omega}_\tau = \{j\tau \,|\, j = 0, \dots, M\}$, $\tau M = T$ and $\bar{\omega}_\mathbf{h} = \{(i_1 h_1,\, i_2 h_2) \,|\, i_1 = 0, \dots, N_1;\, i_2 = 0, \dots, N_2\}$, $h_k N_k = X_k$, $k = 1, 2$. Below we use the following notations for subsets $\bar{\omega}_\mathbf{h}$

$$\begin{gathered} \gamma_1 = \{\mathbf{ih} \,|\, i_1 = N_1\}, \quad \gamma_{-1} = \{\mathbf{ih} \,|\, i_1 = 0\}, \quad \gamma_2 = \{\mathbf{ih} \,|\, i_2 = N_2\}, \\ \gamma_{-2} = \{\mathbf{ih} \,|\, i_2 = 0\}, \quad \gamma_\mathbf{h} = \gamma_1 \cup \gamma_{-1} \cup \gamma_2 \cup \gamma_{-2}, \quad \omega_\mathbf{h} = \bar{\omega}_\mathbf{h} \setminus \gamma_\mathbf{h}. \end{gathered}$$

The value of a mesh function g at a node (j, i_1, i_2) is denoted by $g^j_{i_1,i_2}$. The omission of indices at the function means that they are equal to

(j, i_1, i_2). For the values of functions at nodes neighboring to a node (j, i_1, i_2) we shall use the following notations

$$g^{j+1}_{i_1,i_2} = \hat{g}, \quad g^{j}_{i_1\pm 1,i_2} = g^{(\pm 1_1)}, \quad g^{j}_{i_1,i_2\pm 1} = g^{(\pm 1_2)}.$$

The mean value of a mesh function at two neighboring nodes will be written in the following abbreviated form

$$g_{s_k,(i_1,i_2)} = \frac{1}{2}\left(g^{(+1_k)}_{i_1 i_2} + g_{i_1,i_2}\right), \quad g_{\bar{s}_k,(i_1,i_2)} = \frac{1}{2}\left(g_{i_1 i_2} + g^{(-1_k)}_{i_1,i_2}\right).$$

The following notations are used for operations of the difference differentiation

$$g_t = \frac{\hat{g} - g}{\tau}, \quad \partial_{x_k} g \equiv g_{x_k} = \frac{g^{(+1_k)} - g}{h_k},$$
$$\partial_{\bar{x}_k} g \equiv g_{\bar{x}_k} = \frac{g - g^{(-1_k)}}{h_k}, \quad \partial_{\overset{\circ}{x}_k} g \equiv g_{\overset{\circ}{x}_k} = \frac{g^{(+1_k)} - g^{(-1_k)}}{2h_k}.$$

The mesh functions $(H, \mathbf{V})$ give the approximate values of the functions $(\rho, \mathbf{u})$.

Let us consider the following difference scheme (E = indentity operator)

$$\prod_{k=1}^{2} \left[E + \tau A^1_k(V^j_k) + 3\tau^2 A^2_k(\mathbf{V}^j)\right] \hat{H} = H + \tau a^j_1, \tag{2.1}$$

$$\begin{aligned} &\left[E + \frac{\tau}{3}\left(V^j_1 \partial_{\overset{\circ}{x}_1} + \partial_{\overset{\circ}{x}_1} V^j_1 - 4\nu \partial_{x_1}\partial_{\bar{x}_1}\right)\right] \\ &\times \left[E + \frac{\tau}{2}\left(V^j_2 \partial_{\overset{\circ}{x}_2} + \partial_{\overset{\circ}{x}_2} V^j_2 - 2\nu \partial_{x_2}\partial_{\bar{x}_2}\right)\right] \hat{V}_1 \\ &= V_1 + \tau a^j_2 \quad \mathbf{x_i} \in \omega_{\mathbf{h}}, \quad \hat{V}_1 = 0, \quad \mathbf{x_i} \in \gamma_{\mathbf{h}}, \end{aligned} \tag{2.2}$$

$$\begin{aligned} &\left[E + \frac{\tau}{2}\left(V^j_1 \partial_{\overset{\circ}{x}_1} + \partial_{\overset{\circ}{x}_1} V^j_1 - 2\nu \partial_{x_1}\partial_{\bar{x}_1}\right)\right] \\ &\times \left[E + \frac{\tau}{3}\left(V^j_2 \partial_{\overset{\circ}{x}_2} + \partial_{\overset{\circ}{x}_2} V^j_2 - 4\nu \partial_{x_2}\partial_{\bar{x}_2}\right)\right] \hat{V}_2 \\ &= V_2 + \tau a^j_3 \quad \mathbf{x_i} \in \omega_{\mathbf{h}}, \quad \hat{V}_2 = 0 \quad \mathbf{x_i} \in \gamma_{\mathbf{h}}. \end{aligned} \tag{2.3}$$

Projections onto the mesh of functions given in the condition (1.2) are used as initial conditions

$$(H,\, \mathbf{V})|_{t=0} = (\rho_0,\, \mathbf{u}_0). \tag{2.4}$$

Operators $A_k^1(V_k^j)$ approximate the part of the difference operator divergence containing partial derivative on the direction x_k

$$A_k^1(V_k)W = \begin{cases} \left[(WV_k)_{\overset{\circ}{x}_k} + V_k W_{\overset{\circ}{x}_k}\right]/2, & \mathbf{x_i} \in \bar{\omega}_\mathbf{h} \setminus (\gamma_k \cup \gamma_{-k}), \\ (WV_k)_{x_k}/2, & \mathbf{x_i} \in \gamma_{-k}, \\ (WV_k)_{\bar{x}_k}/2, & \mathbf{x_i} \in \gamma_k. \end{cases} \tag{2.5}$$

Operators $A_k^2(V^j)$ are added for the stability of the difference scheme

$$A_k^2(\mathbf{V})W = \begin{cases} -(\Phi_k W_{x_k})_{\bar{x}_k}, & \mathbf{x_i} \in \bar{\omega}_\mathbf{h} \setminus (\gamma_k \cup \gamma_{-k}), \\ -2h_k^{-1}\Phi_k W_{x_k}, & \mathbf{x_i} \in \gamma_{-k}, \\ 2h_k^{-1}\Phi_k^{(-1_k)} W_{\bar{x}_k}, & \mathbf{x_i} \in \gamma_k. \end{cases} \tag{2.6}$$

The mesh functions Φ_1 and Φ_2 are given by

$$\Phi_1 = |\mathbf{V}|^2_{s_1 s_2 \bar{s}_2}, \quad \Phi_2 = |\mathbf{V}|^2_{s_1 \bar{s}_1 s_2}.$$

For the correctness of these definitions, the function $|\mathbf{V}^2|$ is continued beyond the boundary of the mesh $\bar{\omega}_\mathbf{h}$ in the odd way.

The mesh function a_1^j is defined by the formula

$$a(H,\mathbf{V}) = \begin{cases} -\left(HV_{1\overset{\circ}{x}_1} + HV_{2\overset{\circ}{x}_2}\right)/2, & \mathbf{x_i} \in \omega_\mathbf{h}, \\ -\{HV_{kx_k} - h_k[H\left(V^{(+1_k)}_{kx_k\bar{x}_k} - V^{(+2_k)}_{kx_k\bar{x}_k}/2\right) \\ + (HV_k)^{(+1_k)}_{x_k\bar{x}_k} - (HV_k)^{(+2_k)}_{x_k\bar{x}_k}/2]\}/2, & \mathbf{x_i} \in \gamma_{-k}, \\ -\{HV_{k\bar{x}_k} + h_k[H\left(V^{(-1_k)}_{kx_k\bar{x}_k} - V^{(-2_k)}_{kx_k\bar{x}_k}/2\right) \\ + (HV_k)^{(-1_k)}_{x_k\bar{x}_k} - (HV_k)^{(-2_k)}_{x_k\bar{x}_k}/2]\}/2, & \mathbf{x_i} \in \gamma_k. \end{cases}$$

The mesh function ν is a constant at the time step j and equal $\nu^j = \mu\left(\min\limits_{\mathbf{x_i}\in\bar{\omega}_\mathbf{h}} H^j\right)^{-1}$.

The operators $V_k\partial_{\overset{\circ}{x}_m}$ and $\partial_{\overset{\circ}{x}_m} V_k$ are defined in the following way

$$V_k\partial_{\overset{\circ}{x}_m}(W) = V_k W_{\overset{\circ}{x}_m}, \quad (\partial_{\overset{\circ}{x}_m} V_k)W = (V_k W)_{\overset{\circ}{x}_m}.$$

The functions a_2^j and a_3^j are given by

$$a_2^j = V_1 V_{2\overset{\circ}{x}_2}/2 - H^{-1} p_{\overset{\circ}{x}_1} + f_1 + (3H)^{-1}\left(\mu V_{2\overset{\circ}{x}_2}\right)_{\overset{\circ}{x}_1}$$
$$+4/3[\mu H^{-1} V_{1x_1\bar{x}_1} - \nu V_{1x_1\bar{x}_1}] + \mu H^{-1} V_{1x_2\bar{x}_2} - \nu V_{1x_2\bar{x}_2},$$
$$a_3^j = V_2 V_{1\overset{\circ}{x}_1}/2 - H^{-1} p_{\overset{\circ}{x}_2} + f_2 + (3H)^{-1}\left(\mu V_{1\overset{\circ}{x}_1}\right)_{\overset{\circ}{x}_2}$$
$$+4/3[\mu H^{-1} V_{2x_2\bar{x}_2} - \nu V_{2x_2\bar{x}_2}] + \mu H^{-1} V_{2x_1\bar{x}_1} - \nu V_{2x_1\bar{x}_1}.$$

The solution of the problem (2.1)-(2.4) is reduced to the computation of the mesh functions $(\hat{H}, \hat{\mathbf{V}})$ at the time step $j+1$ using the already known functions $(H, \mathbf{V})$ from the previos time step j. It is done for time steps with numbers $j = 0, \ldots, M-1$.

Equations (2.1)-(2.3) form a system of algebraic equations, from which we can find the unknown functions $(\hat{H}, \hat{\mathbf{V}})$. This system consists of three subsystems (determined by equations (2.1)-(2.3) respectively), which are linear, if they are solved consecutively.

Let us consider the question of existence and uniqueness of the difference solution. It is evidently that the function H must be positive. It is proved in [5] that this property of the differential solution remains for the difference solution under certain conditions.

To determine the unknown function $\hat{H}$ in the nodes of the mesh $\bar{\omega}_{\mathbf{h}}$ we obtain from equation (2.1) the system of algebraic equations

$$A_1 A_2 \hat{H} = a,$$

where the square matrix A_k $(k = 1, 2)$ has size $(N_1+1)(N_2+1) \times (N_1 + 1)(N_2+1)$. The matrix A_k is block-diagonal and the each block has size $(N_k+1) \times (N_k+1)$ and is a tridiagonal matrix, which is a sum of three matrices. The first matrix is diagonal: the elements in the first and the last lines are equal 1/2, the others are equal 1. The second matrix is skew-symmetric, and the third matrix is symmetric and nonnegative. As a consequence of this structure, each block is nonsingular, such that the whole matrix A_k is nonsingular.

Equations (2.2) and (2.3) lead to systems of algebraic equations, which have the following structure

$$B_1^k B_2^k \hat{V}^k = b_k, \quad k = 1, 2.$$

The matrixes B_l^k are a sum of a positive and symmetric matrix and a skew-symmetric matrix. Hence, the matrices B_l^k are nonsingular.

Thus the difference scheme always has a unique solution, which can be found by the sweep method.

The constructed difference scheme is the subject of the following theorem on the estimation of the numerical integration error.

Theorem. Let the following conditions be satisfied for the differential problem (1.1)-(1.2):
(a) there exists an unique classical solution;
(b) $(\rho, \mathbf{u}) \in C^{4,2}(Q)$, $grad(\rho) \in C^{3,2}(Q)$, $grad(\mathbf{u}) \in C^{3,1}(Q)$, where $C^{p,q}(Q)$ is a class of functions having derivatives of order p in spatial variables $\mathbf{x}$ and of order q in the time variable t in the domain Q;
(c) $\rho(t, \mathbf{x}) \geq m > 0$ for $(t, \mathbf{x}) \in Q$;
(d) the function p is twice-differentiable in the domain $\rho > 0$, $\|\mathbf{f}\|_{C(Q)} < \infty$.

Then there exist quantities $\tau_{\max}$, $\mathbf{h}_{\max}$, K and C which depend on
(i) the size of the domain Q;
(ii) the norms of the exact solution to a differential problem in the classes of functions indicated in the condition (b) and the constant m from the condition (c);
(iii) the constant μ and the properties of functions p and $\mathbf{f}$;
(iiii) the constant ε - a relative accuracy of arithmetic operations.
These quantities are such that the mesh norms of the difference between the difference solution and the projection of the exact solution to the differential problem on the mesh $Q_{\tau\mathbf{h}}$ obey the following estimate

$$\max_{j \leq M} \|(H^j, \mathbf{V}^j) - (\rho^j, \mathbf{u}^j)\|_1^2 + \tau \sum_{i=1}^{M} (|\mathbf{V}^i - \mathbf{u}^i|_2^2 + \|\mathbf{V}_{\bar{t}}^i - \mathbf{u}_{\bar{t}}^i\|^2) \leq$$
$$\leq C \left(\left(\tau + h_1^2 + h_2^2\right)^2 + (\phi_c)^2 \right),$$

where

$$\tau \leq \tau_{\max}, \quad \mathbf{h} \leq \mathbf{h}_{\max}, \quad \tau \leq K|ln|\mathbf{h}|^2|^{-1/2},$$

and the value ϕ_c is given by

$$\phi_c = \frac{\varepsilon}{\tau} \left(\left(\frac{1}{h_1} + \frac{1}{h_2} \right) \prod_{k=1}^{2} \left(1 + \frac{\tau}{h_k} + \frac{\tau^2}{h_k^2} + \frac{\tau^3}{h_k^3} \right) + \prod_{k=1}^{2} \left(1 + \frac{\tau}{h_k^2} \right) \right).$$

This theorem is proved under the assumption that the sweep algorithm is used for solving the systems of linear equations. During the proof we assume that the value ϕ_c satisfies the condition $\phi \leq K|ln|\mathbf{h}|^2|^{-0,5}$.

The mesh norms used in the theorem are

$$\|g\| = \sqrt{h_1 h_2 \sum_{\mathbf{i}\in\bar{\omega}_{\mathbf{h}}} g_{\mathbf{i}}}, \quad \|g\|_2^1 = \sqrt{\|g\|^2 + \|g_{x_1}\|^2 + \|g_{x_2}\|^2},$$

$$|g|_2 = \sqrt{\|g_{x_1\bar{x}_1}\|^2 + \|g_{x_1 x_2}\|^2 + \|g_{x_2\bar{x}_2}\|^2}.$$

The remaining part of this paper deals with the proof of this theorem.

3 The perturbation of the system of linear equations

N.S.Bakhvalov suggested an approach to investigate this problem in [7]. He analyzes the system of equations

$$(E + \tau B)\,\hat{\mathbf{U}} = \mathbf{U} + \tau\mathbf{b},$$

where $B = B^* > 0$ is a tridiagonal matrix of the difference differentiation operator of the second order.

It is proved that using the sweep method with the relative accuracy of arithmetic operations ε we get a vector $\hat{U}^*$, which satisfies the system of equations

$$(E + \tau B + \tau B_\varepsilon)\,\hat{\mathbf{U}}^* = (1 + \xi)(\mathbf{U} + \tau\mathbf{b}).$$

The matrix B_ε is tridiagonal. For its elements the following estimate is correct

$$|b_{ij}^\varepsilon| \leq C\left(\frac{1}{\tau} + \frac{1}{h^2}\right)\varepsilon,$$

and values of the function ξ have an order of ε.

Let us use this approach to investigate the system of equations

$$\left(E + \tau A_k^1(V_k^j) + 3\tau^2 A_k^2(\mathbf{V}^j)\right)\mathbf{R} = \mathbf{g},$$

where the operators $A_k^1(V_k^j)$ and $A_k^2(V^j)$ are defined in (2.5)-(2.6).

The solving of this system can be reduced to the solving of the series of the systems

$$A\mathbf{H} = \mathbf{F}, \tag{3.1}$$

where the matrix A and the vector $\mathbf{F}$ are the following

$$\begin{pmatrix} 4h^2+2\tau^2 b_0 & 2\tau h a_0 - 2\tau^2 b_0 & 0 & \dots\dots\dots\dots\dots\dots & 0 \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ \dots\ 0 & -\tau h a_{j-1} - \tau^2 b_{j-1} & 4h^2+\tau^2(b_j+b_{j-1}) & \tau h a_j - \tau^2 b_j & 0\ \dots \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ 0 & \dots\dots\dots\dots\dots & 0 & -2\tau h a_{N-1} - 2\tau^2 b_{N-1} & 4h^2+2\tau^2 b_{N-1} \end{pmatrix},$$

$$a_j = V_{kj} + V_{kj}^{(+1_k)}, \quad b_j = 12\Phi_{kj}, \quad \mathbf{F}_{i_k} = 4h^2 \mathbf{g}_{i_1 i_2}.$$

The solution of this system by the sweep method can be reduced to the following formulas

$$\alpha_0 = \frac{-\tau h a_0 + \tau^2 b_0}{2h^2 + \tau^2 b_0}, \qquad \beta_0 = \frac{F_0/2}{2h^2+\tau^2 b_0},$$

$$\alpha_j = \frac{-\tau h a_j + \tau^2 b_j}{4h^2 + \tau^2(b_j + b_{j-1}) - \alpha_{j-1}(\tau h a_{j-1} + \tau^2 b_{j-1})}, \qquad j = 1, \dots, N-1,$$

$$\beta_j = \frac{F_j + \beta_{j-1}(\tau h a_{j-1} + \tau^2 b_{j-1})}{4h^2 + \tau^2(b_j + b_{j-1}) - \alpha_{j-1}(\tau h a_{j-1} + \tau^2 b_{j-1})}, \qquad j = 1, \dots, N-1,$$

$$H_N = \frac{F_N/2 + \beta_{N-1}(\tau h a_{N-1} + \tau^2 b_{N-1})}{4h^2 + \tau^2(b_N + b_{N-1}) - \alpha_{N-1}(\tau h a_{N-1} + \tau^2 b_{N-1})},$$

$$H_k = \alpha r_{k+1} + \beta_k, \qquad k = N-1, \dots, 0. \tag{3.2}$$

The elements a_k, b_k $(k = 0, \dots, N-1)$ and F_k $(k = 0, \dots, N)$ are restricted above by a constant C independent on N. In this case the estimate for α_0 is evidently

$$|\alpha_0| \leq C\frac{\tau}{h} + 1.$$

Let us now prove that

$$|\alpha_j| \leq C\frac{\tau}{h} + 1 \tag{3.3}$$

for any j. The proof will be carried out by induction.

First we prove that the denominator of the fraction α_j can be represent as a sum

$$4h^2 + \tau^2 b_j + \gamma_j,$$

where $\gamma_j \geq 0$.

For α_0 it is evident. Let us assume that this representation is correct for all α_j with $j \leq J$. Therefore we must prove that

$$\tau^2 b_J - \alpha_J(\tau h \alpha_J + \tau^2 b_J) \geq 0.$$

According to the assumption we have

$$\alpha_J = \frac{-\tau h a_J + \tau^2 b_J}{4h^2 + \tau^2 b_J + \gamma_J}.$$

From this equality it follows that

$$\tau^2 b_J + \frac{\tau^2 h^2 a_J^2 - \tau^4 b_J^2}{4h^2 + \tau^2 b_J + \gamma_J} \geq 0.$$

Thus we have proved that $\gamma_{J+1} \geq 0$. Therefore $\gamma_j \geq 0$ for any j and the estimate (3.3) is evident.

Let us use the sweep method with the relative accuracy of arithmetic operations ε. Then the formulas (3.2) changes in the following way

$$\begin{gathered}
\alpha_0^* = \frac{D_0 + \delta_0}{E_0 + p_0}, \quad \beta_0^* = \frac{F_0/2 + r_0}{E_0 + p_0}, \\
\alpha_j^* = \frac{D_j + \delta_j}{E_j - \alpha_{j-1}^* Q_{j-1} + p_j}, \quad \beta_j^* = \frac{F_j + Q_{j-1}\beta_{j-1}^* + r_j}{E_j - \alpha_{j-1}^* Q_{j-1} + p_j}, \\
j = 1, \ldots, N-1, \\
H_N^* = \frac{F_N/2 + Q_{N-1}\beta_{N-1}^* + r_N}{E_N - \alpha_{N-1}^* Q_{N-1} + p_N}, \\
H_k^* = (\alpha_k^* H_{k+1}^*(1 + \kappa_{k+1}) + \beta_k^*)(1 + \omega_{k+1}), \quad k = N-1, \ldots, 0.
\end{gathered} \tag{3.4}$$

The following notations are used in (3.4)

$$D_j = -\tau h a_j + \tau^2 b_j, \quad j = 0, \ldots, N-1,$$

$$E_0 = 2h^2 + \tau^2 b_0, \quad E_j = 4h^2 + \tau^2(b_j + b_{j-1}), \quad j = 1, \ldots, N,$$

$$Q_j = \tau h a_j + \tau^2 b_j, \quad j = 0, \ldots, N-1,$$

$$\delta_j = \delta_j^1 \tau h a_j + \delta_j^2 \tau^2 b_j, \quad j = 0, \ldots, N-1,$$

$$\begin{gathered}
p_0 = 2p_0^1 h^2 + p_0^2 \tau^2 b_0, \quad p_j = 4p_j^1 h^2 + p_j^2 \tau^2 b_j + p_j^3 \tau^2 b_{j-1} \\
+ p_j^4 \alpha_{j-1}^* \tau h a_{j-1} + p_j^5 \alpha_{j-1}^* \tau^2 b_{j-1}, \quad j = 1, \ldots, N,
\end{gathered}$$

$$r_0 = r_0^1 F_0, \quad r_j = r_j^1 F_j + \beta_{j-1}^* \sigma_j,$$

$$\sigma_j = r_j^2 \tau h a_{j-1} + r_j^3 \tau^2 b_{j-1}, \quad j = 1, \dots, N.$$

In these formulas and throughout we denote by δ_l^k, p_l^k, r_l^k, κ_k and ω_k $(k, l = 0, 1, \dots)$ the values of an order of ε.

Let us denote by β^{**} the expression

$$\beta_k^{**} = \beta_k^* + \kappa_{k+1} \alpha_k^* H_{k+1}^* + \omega^{k+1} H_k^* \tag{3.5}$$

where $\omega^k = \dfrac{\omega_k}{1 + \omega_k}$. Thus we have

$$H_k^* = \alpha_k^* H_{k+1}^* + \beta_k^{**}, \quad k = 0, \dots, N-1. \tag{3.6}$$

Now let us find a function F^* such that formulas (3.2) will remain correct if α, β and F are substituted by α^*, β^{**} and F^* with the only change: the value p_j is added to the denominator of values α_j, β_j and H_N and the value δ_j will be added to the numerator of the value α_j.

Let us determine the expression for F_0^*. For this purpose we change β_0^* in (3.4) according to (3.5)

$$\beta_0^{**} = \frac{F_0/2 + r_0}{E_0 + p_0} + \kappa_1 \alpha_0^* H_1^* + \omega^1 H_0^*.$$

From this expression we have that

$$F_0^* = F_0/2 + r_0 + (D_0 + \delta_0) \kappa_1 H_1^* + (E_0 + p_0) \omega^1 H_0^*.$$

Analogously it can be obtained

$$\begin{gathered} F_k^* = F_k + r_k^1 F_k - Q_{k-1} \left(\kappa_k \alpha_{k-1}^* H_k^* + \omega^k H_{k-1}^* \right) \\ + \sigma_k \left(\beta_{k-1}^{**} - \kappa_k \alpha_{k-1}^* H_k^* - \omega^k H_{k-1}^* \right) + (D_k + \delta_k) \kappa_{k+1} H_{k+1}^* \\ + (E_k - \alpha_{k-1}^* Q_{k-1} + p_k) \omega^{k+1} H_k^*, \quad k = 1, \dots, N-1, \end{gathered}$$

$$\begin{gathered} F_N^* = F_N/2 + r_N^1 F_N - Q_{N-1} \left(\kappa_N \alpha_{N-1}^* H_N^* + \omega^N H_N \right) \\ + \sigma_N \left(\beta_{N-1}^{**} - \kappa_N \alpha_{N-1}^* H_N^* - \omega^N H_{N-1} \right). \end{gathered}$$

Let us change β^{**} according to (3.6) in formulas for F^*

$$F_0^* = F_0/2 + r_0 + (D_0 + \delta_0) \kappa_1 H_1^* + (E_0 + p_0) \omega^1 H_0^*,$$

$$\begin{gathered} F_k^* = F_k + r_k^1 F_k - Q_{k-1} \left(\kappa_k \alpha_{k-1}^* H_k^* + \omega^k H_k \right) \\ + \sigma_k \left(H_{k-1}^* - \alpha_{k-1}^* H_k^* - \kappa_k \alpha_{k-1}^* H_k^* - \omega^k H_{k-1}^* \right) + (D_k + \delta_k) \kappa_k H_k^* \\ + \left(E_k - \alpha_{k-1}^* Q_{k-1} + p_k \right) \omega^{k+1} H_k^*, \quad k = 1, \dots, N-1, \end{gathered} \tag{3.7}$$

$$F_N^* = F_N/2 + r_N^1 F_N - Q_{N-1}\left(\kappa_N \alpha_{N-1}^* H_N^* + \omega^N H_N^*\right)$$
$$+\sigma_N\left(H_{N-1}^* - \alpha_{N-1}^* H_N^* - \kappa_N \alpha_{N-1}^* H_N^* - \omega^N H_{N-1}^*\right).$$

Note, that

$$|\alpha_j^*| \leq C\left(\frac{\tau}{h}+1\right), \quad j=0,\ldots,N-1. \tag{3.8}$$

This estimate can be proved similarly to the inequality (3.3) under the additional condition: the value $\varepsilon(h^{-2}+1)$ is sufficiently small.

So we have proved that the solution of the system (3.1) obtained by the formulas (3.2) with the relative accuracy of arithmetic operations ε is an exact solution of the system

$$(A + A_{p\delta})\,\mathbf{H}^* = \mathbf{F}^*, \tag{3.9}$$

where $A_{p\delta}$ is a two-diagonal matrix. The elements of diagonals of $A_{p\delta}$ are the values of functions p and δ.

It follows from formulas (3.7) that the system (3.9) can be rewritten in the following form

$$(A + A^*)\,\mathbf{H}^* = \mathbf{F} + \mathbf{F}^{**},$$

where $\mathbf{F}_k^{**} = r_k^1 \mathbf{F}_k$ and A^* is a tridiagonal matrix.

From the estimate (3.8) and formulas (3.7) the estimate for elements of A^* follows

$$|a_{ij}| \leq C\left(\left(\tau h + \tau^2\right)\left(1+\frac{\tau}{h}\right) + h^2\right)\varepsilon.$$

Returning to the original problem we have

$$\left(E + \tau A_k^1(V_k^j) + 3\tau^2 A_k^2(\mathbf{V}^j)\right) + A_{k\varepsilon})\mathbf{R}^* = (1+\xi)\mathbf{g}.$$

The following estimates are valid for the operator $A_{k\varepsilon}$ and the function ξ

$$\|A_{k\varepsilon}\mathbf{u}\| \leq C\left(\frac{\left(\tau h + \tau^2\right)\left(1+\dfrac{\tau}{h}\right)}{h^2} + 1\right)\varepsilon\|\mathbf{u}\|,$$

$$\|\xi\| \leq C\cdot\varepsilon.$$

4 Stability of the difference schemes with round-off computer errors

Let us consider a difference scheme, which is reduced to the solving of the following equation at the each time step

$$B_1 B_2 \mathbf{x} = \mathbf{b}, \tag{4.1}$$

where B_1 and B_2 are tridiagonal matrixes.

The problem can be easily reduced to the system of equations

$$\begin{cases} B_1 \mathbf{y} = \mathbf{b}, \\ B_2 \mathbf{x} = \mathbf{y}. \end{cases} \tag{4.2}$$

Then, let it is known, that the result of the solving of the equation

$$B_k \mathbf{z}_k = \mathbf{d}_k$$

is the function $\mathbf{z}_k^*$, which is an exact solution of the equation

$$(B_k + B_k^*) z_k^* = (1 + \xi_k) \mathbf{d}_k.$$

Thus solving the system (4.2) we obtain two functions $\mathbf{x}^*$ and $\mathbf{y}^*$, which are an exact solution of the system

$$\begin{cases} (B_1 + B_1^*) \mathbf{y}^* = (1 + \xi_1) \mathbf{b}, \\ (B_2 + B_2^*) \mathbf{x}^* = (1 + \xi_2) \mathbf{y}^*. \end{cases} \tag{4.3}$$

The system (4.3) can be reduced to the one equation

$$(B_1 B_2 + B^*) \mathbf{x}^* = \mathbf{b} + \mathbf{b}^*, \tag{4.4}$$

where

$$B^* = B_1^* B_2 + B_1 B_2^* + B_1^* B_2^*,$$

$$\mathbf{b}^* = \xi_1 \mathbf{b} + (B_1 + B_1^*) \left(\frac{\xi_2}{1 + \xi_2} (B_2 + B_2^*) \mathbf{x}^* \right).$$

Let us consider the concrete problem, where the equation (4.1) is realized by (2.1). In this case we obtain the following estimates using the results of the section 3

$$
\begin{aligned}
\|B^*\mathbf{u}\| \le C\Bigg(\varepsilon\left(1+\frac{\tau}{h_2}+\frac{\tau^2}{h_2^2}\right)\left(\frac{(\tau h_1+\tau^2)\left(1+\dfrac{\tau}{h_1}\right)}{h_1^2}+1\right)+ \\
+\varepsilon\left(1+\frac{\tau}{h_1}+\frac{\tau^2}{h_1^2}\right)\left(\frac{(\tau h_2+\tau^2)\left(1+\dfrac{\tau}{h_2}\right)}{h_2^2}+1\right)+ \\
+\varepsilon^2\prod_{k=1}^{2}\left((\tau h_k+\tau^2)\left(1+\frac{\tau}{h_k^2}\right)\frac{1}{h_k}+1\right)\Bigg)\|\mathbf{u}\|, \\
\|\mathbf{b}^*\| \le C\varepsilon\Bigg(\|\mathbf{b}\|+ \\
+\prod_{k=1}^{2}\left(1+\frac{\tau}{h_k}+\frac{\tau^2}{h^2}+\varepsilon\left((\tau h_k+\tau^2)\left(1+\frac{\tau}{h_k}\right)\frac{1}{h^2}+1\right)\right)\Bigg)\|\mathbf{x}^*\|.
\end{aligned}
\tag{4.5}
$$

Analogously the estimates for the equations (2.2)–(2.3) can be obtained

$$
\begin{aligned}
\|B^*\mathbf{u}\| \le C\varepsilon\left(1+\frac{\tau}{h_1^2}\right)\left(1+\frac{\tau}{h_2^2}\right)\|\mathbf{u}\|, \\
\|\mathbf{b}^*\| \le C\varepsilon\left(\|\mathbf{b}\|+\left(1+\frac{\tau}{h_1^2}\right)\left(1+\frac{\tau}{h_2^2}\right)\|\mathbf{x}^*\|\right).
\end{aligned}
\tag{4.6}
$$

Thus the solution, which we obtain using arithmetic operations with relative accuracy ε, satisfies equations (2.1)–(2.3) only approximately. Therefore these equations must be substituted by the equations (4.4). After that the difference scheme can be investigated by the method of energy inequalities applied in [5], using the estimates (4.5)–(4.6). In this case we obtain the estimate of the additional error of the numerical integration which depends on parameters of the difference scheme τ and $\mathbf{h}$ and the relative accuracy ε. The final result is formulated in the theorem in the subsection 2.

Acknowledgements. The author is grateful to Prof. N.S.Bakhvalov for fruitful discussions and Prof. Per Christian Hansen for support and for his useful comments concerning this work.

This work is a part of Joint Research Project "High performance Computing in Numerical Simulation", founded by INTAS within a framework of Cooperation Agreement INTAS-93-377-EXT and partly supported by

Russian Science Foundation for Basic Research (96-01-01394).

References

[1] David A.Kopriva, *Multidomain Spectral Solution of Compressible Viscous Flows.* J.Comput.Phyz. (1994) 115, N 1, 184-189.

[2] J.P.Collins, P.Colella, H.M Glaz, *An Implicit-Explicit Eulerian Godunov Scheme for Compressible Flow.* J.Comput.Phyz. (1995) 116, N 2, 195-211.

[3] Francois Renaud, Serge Gauthier, *A Dynamical Pseudo-Spectral Domain Decomposition Technique: Application to Viscous Compressible Flows.* J.Comput.Phyz. (1997) 131, N 1, 89-108.

[4] A.A.Amosov and A.A.Zlotnik, *A difference scheme for the equations of motion of a viscous heat-conducting gas, its properties and error estimates "in the large".* Sov.Math.Docl. (1985) 32, 397-402.

[5] A.V.Popov, *A study of cost-effective finite difference scheme for the system of equations for two-dimensional flow of a viscous barotropic gas.* Sov.J.Mumer.Mat.Modelling (1990) 5, 395-417.

[6] G.M.Kobelkov and A.G.Sokolov, *On finite difference schemes for viscous barotropic compressible gas problems.* Sov. J. Mumer. Mat. Modelling (1994) 9, 223-229.

[7] N.S.Bakhvalov, *About accumulation of computed error during numerical solving differential equations.* Collection of works of Computing Center of MSU (1962) 1, 47-68.