INTERNATIONAL JOURNAL FOR NUMERICAL METHODS IN FLUIDS, VOL. 4, 477–496 (1984)

ON SOME OPTIMIZATION PROCEDURES FOR SHOCK LOCALIZATION

E. V. VOROZHTSOV AND N. N. YANENKO

Institute of Theoretical and Applied Mechanics, U.S.S.R. Academy of Sciences, Novosibirsk 630090, U.S.S.R.

SUMMARY

We consider a problem on shock wave localization in the numerical solution of one-dimensional unsteady problems of gas dynamics in Eulerian variables obtained on the basis of finite difference shock-capturing schemes. An optimization method for strong discontinuity localization proposed previously by Miranker and Pironneau is investigated by means of methods of classical variational calculus. This method may be difficult to implement when the entropy condition is included in the formulation of Miranker and Pironneau's optimization problem as an active constraint. In this connection we suggest an alternative optimization problem using artificial viscosity in the variational principle. It is shown theoretically that the application of such a variational principle yields a trajectory which coincides with the true discontinuity trajectory in the case of a shock wave moving at a constant speed. On the basis of this modification one more algorithm is proposed which reduces the shock localization problem to a problem of minimization of a univariate function. Numerical tests corroborate completely the theoretical conclusions. In particular, a higher shock localization accuracy is obtained on the basis of the proposed algorithms as compared to the original Miranker–Pironneau method.

KEY WORDS Gas Dynamics Finite Difference Schemes Shock Localization

1. INTRODUCTION

An analysis of trends in computational aerodynamics shows that finite difference shock-capturing schemes will continue to play an important role in the numerical simulations of gas flows with discontinuities, see for example recent surveys.[1,2] One characteristic feature of these methods consists in the fact that the amount of numerical information obtained as a result of a problem solution exceeds by several orders of magnitude the amount which is of real interest for a research worker.[3,4] Another feature of shock-capturing techniques is their low accuracy in the vicinity of strong discontinuities which are spread over several intervals of the computing mesh.[1,5,6] These circumstances give rise to two practical problems: one regarding the interpretation of the numerical results[4] and the other the increase in accuracy of a difference solution in the vicinity of shocks.[1] In this connection the questions of the development and the foundation of the algorithms for localization of strong discontinuities in the computational domain by the shock-capturing results are of present interest. A survey of a number of algorithms for localization of singularities which were applied empirically by various authors can be found in Reference 7. In References 7–11 the first theory is described which apparently justifies some methods for locating shock waves by the shock-capturing computation results of the one-dimensional gas dynamic problems. A 'global shock fitting method' was proposed in References 12 and 13. As a matter of fact this method is the method for localization of singularities in the solution. It reduces the shock localization

0271–2091/84/050477–20$02.00

Received 29 September 1982
Revised 18 January 1983

problem to an unconstrained minimization problem. Computational realization of the Miranker–Pironneau method has been accomplished[12,13] for the example of Burgers' equation. In References 14 and 15 this method was applied to shock localization in the numerical solutions of one-dimensional problems of filtration of a multiphase incompressible fluid. In Reference 16 Miranker and Pironneau's method has been extended to the case of localization of non-stationary one-dimensional shock waves in a gas when using shock-capturing schemes for the computation of the overall flow field.

The information on the shock location obtained with the aid of the above mentioned localization algorithms can be successfully applied for purposes of difference solution refinement in the vicinity of a discontinuity.[17,18] To determine the shock location within the zone of a smeared shock wave[17,18] the 'wave centre notion' was used, which coincides in its geometrical sense with the definition of the finite difference shock wave centre that we have given.[8,10] Note that this notion was the basic one in our theoretical studies.[7–11,19] Suppose that the shock wave under consideration is so intensive that the jumps of basic kinematic and thermodynamic quantities at the shock front are $O(1)$. Then in a sufficiently small neighbourhood of the smeared shock wave centre the finite difference solution obtained by the shock-capturing scheme has an error of the same order of smallness $O(1)$. This fact follows from numerous computation results.[7–11,19,20] If in this case the localization algorithm is applied consecutively at each time step as the first stage of the algorithm for difference solution refinement (similarly to References 17 and 18), then it is clear that if the accuracy of shock localization is low, the errors in the 'refined' difference solution in the shock vicinity and in the position of the discontinuity itself can reach values of the order $O(1)$ after a finite number of time steps in the case of uniform spatial computing mesh. As a result of this the picture of the flow under consideration will be completely distorted. Therefore, especially severe requirements on the shock localization accuracy should be imposed in the cases when the information on the shock location is to be used for purposes of subsequent refinement of the difference solution. In this connection it was of interest to study the shock localization accuracy in the case of application of the optimization approach proposed in References 12 and 13. Below we present the results of such a study and propose some modifications of the method[12,13] which are also subject to theoretical and numerical investigation.

2. AN ANALYSIS OF THE MIRANKER–PIRONNEAU METHOD

By analogy with References 12 and 13 let us consider Burger's equation

$$\partial u/\partial t + \partial\varphi(u)/\partial x = 0 \tag{1}$$

with the initial condition

$$u(x, 0) = u_0(x), \qquad x \in R, \qquad 0 \le t \le T \tag{2}$$

The function $\varphi(u)$ in (1) is assumed to be twice continuously differentiable, and $\varphi''(u) \neq 0$. The function $u_0(x)$ in (2) is assumed to have a discontinuity at a point x_0, $-\infty < x_0 < \infty$, and

$$u_0(x) = \begin{cases} U_1(x), & x < x_0 \\ U_2, & x \ge x_0 \end{cases} \tag{3}$$

where $U_2 = \text{const}$ and $U_1(x)$ is a continuous function, such that $U_1(x_0 - 0) - U_2 \ge K$, $K = \text{const} > 0$. As in References 12 and 13 let us now introduce two auxiliary Cauchy problems P^+ and P^- for the equation (1) with solutions u^+ and u^-, respectively. As initial values let us take the functions u_0^+ and u_0^- in $C^1(-\infty,\infty)$ such that

$$u_0^- \equiv u_0 \quad \text{for} \quad x < x_0$$
$$u_0^+ \equiv u_0 \quad \text{for} \quad x > x_0$$

Let $x=\zeta(t)$ be some approximation of the discontinuity trajectory $x=\xi(t)$ in the solution of the problem (1), (2). Let us introduce the function[12,13]

$$J(\zeta(t),t)=\{\varphi(u^-(\zeta(t),t))-\varphi(u^+(\zeta(t),t))\}/\{u^-(\zeta(t),t)-u^+(\zeta(t),t)\} \tag{4}$$

as well as the function

$$F(\zeta(t),\dot{\zeta}(t),t)=J(\zeta(t),t)-\dot{\zeta}(t) \tag{5}$$

where $\dot{\zeta}(t)\equiv \mathrm{d}\zeta(t)/\mathrm{d}t$. Now consider as in References 12 and 13 the non-negative functional

$$I(\zeta)=\int_0^T F^2(\zeta(t),\dot{\zeta}(t),t)\,\mathrm{d}t \tag{6}$$

The Rankine–Hugoniot condition implies the equality $I(\xi)=0$. Since at $t=0$ the discontinuity position is known according to (3), we can set $\zeta(0)=x_0$. The position of the abscissa $x=\zeta(T)$ is generally unknown (it is to be determined). In this connection it is natural to impose at the right end $t=T$ the transversality condition

$$2FF_{\dot{\zeta}}\,|_{t=T}=0 \tag{7}$$

where $F_{\dot{\zeta}}=(\partial/\partial\dot{\zeta})F(\zeta,\dot{\zeta},t)$.

Subsequently in Section 5 we shall present computational results which validate the use of the condition (7). Thus, let us consider the following variational problem for the functional (6):

$$I(\zeta)\to \min, \qquad \zeta(0)=x_0, \qquad 2FF_{\dot{\zeta}}|_{t=T}=0 \tag{8}$$

Regarding (3) let us set $u_0^+=U_2$. Taking into account the properties of equation (1) we have that $u^+(x,t)=U_2$. The function $u^-(x,t)$, being the solution of the auxiliary problem P^- may be determined similarly to References 12 and 13 as the solution obtained by a shock-capturing scheme (in References 12 and 13 the Lax–Wendroff scheme was used for this purpose). Thus in the solution $u^-(x,t)$ the shock front is approximated by a smeared shock wave zone occupying several mesh intervals on the x-axis.

The Euler–Lagrange equation corresponding to the functional (6) has the form

$$\frac{\mathrm{d}^2\zeta}{\mathrm{d}t^2}-\frac{\mathrm{d}\zeta}{\mathrm{d}t}\frac{\partial F}{\partial \zeta}-\frac{\partial F}{\partial t}-F\frac{\partial F}{\partial \zeta}=0 \tag{9}$$

Employing the formula (5) let us rewrite equation (9) as follows

$$\ddot{\zeta}-\frac{\partial J}{\partial t}-J\frac{\partial J}{\partial \zeta}=0 \tag{10}$$

By virtue of the construction of the solution $u^-(x,t)$ the relationship

$$\frac{\partial u^-}{\partial t}=-\varphi'(u^-)\frac{\partial u^-}{\partial x} \tag{11}$$

is valid. Making use of the the formulae (4) and (11) it is easy to find that

$$J_t + JJ_\zeta = -(u^- - U_2)^{-3}\frac{\partial u^-}{\partial x}(\zeta(t), t)\langle\{[\varphi(u^-) - \varphi(U_2)]/(u^- - U_2)\} - \varphi'(u^-)\rangle^2 \tag{12}$$

Let us now employ the formula

$$\varphi(U_2) = \varphi(u^-) + \varphi'(u^-)(U_2 - u^-) + \tfrac{1}{2}\varphi''(u^*)(U_2 - u^-)^2 \tag{13}$$

where $u^* \in [U_2, u^-(\zeta(t), t)]$. With the formulae (12) and (13) in view let us rewrite equation (10) in the form

$$\ddot{\zeta} = -(u^- - U_2)^{-1}\frac{\partial u^-}{\partial x}(\zeta(t), t)[\tfrac{1}{2}\varphi''(u^*)]^2 \tag{14}$$

When an approximate solution obtained by a shock-capturing scheme is used for $u^-(x, t)$, then in the smeared shock wave zone $\partial u^-/\partial x < 0$.[22] In addition, we use such a function $u^-(x, t)$ that $u^- > U_2$ in the smeared shock wave zone. Therefore we obtain from (14) that at the initial data (3) always $\ddot{\zeta} > 0$. This inequality means that the shock wave found as a result of the minimization of functional (6) always accelerates independently of the actual behaviour of the true discontinuity. Thus the solution of the variational problem (8) in the general case does not coincide with the true discontinuity trajectory in the solution of the problem (1), (2). This result may be formulated as

Theorem 1

If the following conditions are satisfied:

(a) In Burgers' equation (1) the function $\varphi(u)$ is twice continuously differentiable and $\varphi''(u) \neq 0$ at $u \neq 0$.
(b) The solution $u^-(x, t)$ of the auxiliary Cauchy problem for equation (1) satisfies the inequality $\partial u^-(x, t)/\partial x < 0$ in the smeared shock wave zone.
(c) The initial function $u_0(x)$ has the form (3) where

$$U_1(x) \in C^1(-\infty, x_0), \qquad U_2 = \text{const}, \qquad U_1(x_0 - 0) - U_2 \geq K, \qquad K = \text{const.} > 0$$

then the solution $\zeta(t)$ of the variational problem (8) where $I(\zeta)$ is determined by the formulae (4)–(6) satisfies the inequality $\ddot{\zeta}(t) > 0$ for $0 \leq t \leq T$ and at finite mesh sizes.

Remark.

Similarly to References 12 and 13 the variational problem (8) for the functional (6) does not contain the entropy condition

$$\varphi'(u(\xi(t) + 0, t)) \leq \dot{\xi}(t) \leq \varphi'(u(\xi(t) - 0, t))$$

in the form of active constraints. Note by the way that the inequality $u^- > U_2$ which we assumed in the proof of Theorem 1 and which follows from condition (b) implies satisfaction of the entropy condition at least in the case when $\varphi''(u) > 0$.[22] In this connection we note also that computational results presented in References 12 and 13 for the case of Burgers' equation (1) with $\varphi(u) = 0{\cdot}5\,u^2$ confirm our theoretical result. On the other hand the Miranker–Pironneau method may be difficult to implement when the entropy condition is included in the formulation of an optimization problem for the functional (6) in the form of two active constraints. In this connection we suggest and investigate in the next section an alternative optimization problem using artificial viscosity in the basic functional. This suggestion was inspired by the well-known fact that the entropy condition can be enforced by introducing artificial viscosity terms when calculating flows with shock waves, see for example References 5, 6 and 22.

Now a natural question arises on the size of the difference $|\xi(t)-\zeta(t)|$. Let us show that in the case when the Miranker–Pironneau functional (6) is used for shock localization in some model gas dynamic problem an exact formula can be obtained for the extremal $\zeta(t)$ providing the solution of the problem (8).

Consider a problem on the motion of a stationary shock wave in a gas. Following Reference 5 let us employ the one-dimensional gas dynamic equations in Eulerian variables in the presence of the artificial viscosity q of the form

$$q = ah^2\rho[\min(\partial u/\partial x, 0)]^2 \tag{15}$$

introduced additively into the pressure. In (15) ρ is the density of the gas, u is velocity, h is the step of a uniform computing mesh on the x-axis and a is a dimensionless constant, $a \geq 0$. Consider the progressive wave type solutions of the above mentioned equation system, that is the solutions depending only on the variable

$$y = x - DT - x_0 \tag{16}$$

where $D = \text{const.}$ is the speed of a steady shock wave and x_0 is an arbitrary constant. Then the equation system under consideration may be integrated once. As a result of this the following system is obtained

$$\rho(u-D) = C_1 = m \tag{17}$$

$$p + q + m(u-D) = C_2 \tag{18}$$

$$m\left[\frac{p}{\rho(\gamma-1)} + \tfrac{1}{2}(u-D)^2\right] + (p+q)(u-D) = C_3 \tag{19}$$

In (17)–(19) C_1, C_2, C_3 are integration constants and γ is a constant entering the equation of state

$$p = (\gamma-1)\rho\varepsilon \tag{20}$$

where p is pressure and ε is the specific internal energy.

Let us ascribe to the arbitrary constant x_0 entering into (16) the following physical meaning: let x_0 be the abscissa of the smeared shock wave centre in the solution of the system (15), (17)–(19) at $t = 0$, that is

$$\xi(0) = x_0 \tag{21}$$

Then the trajectory of the smeared shock wave centre at $t \geq 0$ is described by the equation

$$x = \xi(t) = Dt + x_0 \tag{22}$$

and thus it coincides with the steady shock wave trajectory. The quantities determining the state behind and before the shock front will be marked by subscripts '1' and '2', respectively. As in Reference 10 these states are assumed to be constant. Let $V(y) = 1/\rho$. As in Reference 10, let us find from the system (17)–(19) ρ, u, p, q as functions of the specific volume V. Assuming that the value of m in (17) is given, the constants C_2, C_3 are determined from the conditions $q(V_1)=q(V_2)=0$. Then ρ, u, p, q as functions of V are found as[10]

$$\rho=1/V, \qquad u=mV+D$$
$$p=(\gamma-1)\frac{m^2}{2}\left[\frac{\gamma+1}{\gamma-1}\frac{V_1V_2}{V}+V-\frac{\gamma+1}{\gamma}(V_1+V_2)\right] \tag{23}$$
$$q=\frac{\gamma+1}{2}\frac{m^2}{V}(V_2-V)(V-V_1)$$

Analogously to (4)–(6) let us introduce the functional $I(\zeta)$ by the formulae (24), (5) and (6) where

$$J(\zeta(t),t)=[(p+\rho u^2)(\zeta(t),t)-(p_2+\rho_2u_2^2)]/$$
$$[(\rho u(\zeta(t),t)-\rho_2u_2] \tag{24}$$

It is assumed in (24) that p, ρ, u is the solution of the system (23), (15) subject to initial condition[10]

$$V(0)=0{\cdot}5(V_1+V_2)$$

Then

$$V(y)=0{\cdot}5[V_1+V_2+(V_2-V_1)\sin by] \tag{25}$$

where

$$b=(1/h)[(\gamma+1)/(2a)]^{0.5} \tag{26}$$

Since we consider solutions of progressive wave type, the Euler–Lagrange equation (10) may be rewritten in the form

$$\ddot{\zeta}-(J-D)\frac{\mathrm{d}J}{\mathrm{d}y}=0 \tag{27}$$

With (24) and (16) in view let us introduce the function

$$v(t)=\zeta(t)-DT-x_0$$

Then equation (27) may be rewritten as follows

$$\ddot{v}-(J-D)\frac{\mathrm{d}J}{\mathrm{d}v}=0 \tag{28}$$

Employing the formulae (23) and (24) it is easy to find that

$$J(v)=\frac{1}{D}\left[D^2-\frac{m^2}{2}(\gamma+1)V_2(V-V_1)\right] \tag{29}$$

where

$$V=V(v)=0{\cdot}5[V_1+V_2+(V_2-V_1)\sin bv(t)] \tag{30}$$

Making use of the formulae (29) and (30) let us rewrite equation (28) in the form

$$\ddot{v}-bc(1+\sin bv)\cos bv=0 \tag{31}$$

where

$$c=[m^2(\gamma+1)V_2(V_2-V_1)/(4D)]^2 \tag{32}$$

and the constant b is determined by (26). Let us search for the solution of equation (31) in the interval $[0, T]$ which satisfies the boundary conditions

$$v(0)=0, \qquad J(v(T))-\dot{v}(T)-D=0 \tag{33}$$

Applying the substitution $\dot{v}=s(v)$ as well as substitutions

$$z=bv, \qquad \chi=tg(z/2), \qquad t_1=(\chi+1)/(\chi-1)$$
$$\sin^2\varphi=(a_1^2+a_2^2)/(a_2^2+t_1^2) \tag{34}$$

we can write the general solution of equation (31) as

$$t+A_2=\sigma(A_1)F(\varphi\backslash\alpha) \tag{35}$$

where A_1, A_2 are integration constants,

$$\sigma(A_1)=(2/b)[A_1(1+\lambda_1)(1+\lambda_2)(a_1^2+a_2^2)]^{-1/2}$$
$$\lambda_1=\nu+\sqrt{(\nu^2-\nu)}, \qquad \lambda_2=\nu-\sqrt{(\nu^2-\nu)}, \qquad \nu=c/A_1$$
$$a_1^2=(\lambda_1-1)/(\lambda_1+1), \qquad a_2^2=(1-\lambda_2)/(1+\lambda_2)$$
$$\alpha=\operatorname{arcctg}(a_1/a_2)=\operatorname{arcctg}\left[\left(\frac{2\sqrt{\nu}+\sqrt{(\nu-1)}}{2\sqrt{\nu}-\sqrt{(\nu-1)}}\right)^{0\cdot5}\right]$$

and $F(\varphi\backslash\alpha)$ is the incomplete elliptic integral of the first kind.[23] The solution (35) is valid for $\nu>1$. Employing the boundary condition $v(0)=0$, the chain of substitutions (34) and the formula (35) it is easy to find the constant A_2 as a function of the constant A_1:

$$A_2=2b^{-1}\{4A_1[\nu(\nu-1)]^{0\cdot5}\}^{-0\cdot5}F\left(\arcsin\left[\frac{2(\nu^2-\nu)^{0\cdot5}}{1+\nu+(\nu^2-\nu)^{0\cdot5}}\right]^{0\cdot5}\backslash\alpha\right)$$

Let us find the integration constant A_1 employing the second condition in (33). Let us introduce the notation

$$r(t)=b\{A_1\sqrt{[\nu(\nu-1)]}\}^{0\cdot5}(t+A_2) \tag{36}$$

Then the formula (35) may be rewritten with (34) in view as

$$v(t)=\frac{1}{b}[2\operatorname{arctg}((a_1^2+a_2^2)^{0\cdot5}\operatorname{ds} r(t))-\pi/2] \tag{37}$$

where by definition[23] $\operatorname{ds} r=\operatorname{dn} r/\operatorname{sn} r$, dn and sn being elliptic Jacobi functions.[23] Employing the formulae (29) and (30) let us write the transversality condition $J(v(T))-\dot{v}(T)-D=0$ in the form

$$\sqrt{c}[1+\sin(bv(T))]+v(T)=0 \tag{38}$$

Making use of the formulae (34) and (37) we find from the condition (38) the following equation for the determination of the constant A_1:

$$\operatorname{cn} r(T)=[(1-m)/m]^{0\cdot5} \tag{39}$$

where

$$m=\sin^2\alpha=a_2^2/(a_1^2+a_2^2)$$

Considering the behaviour of the left hand side and of the right hand side of equation (39) as functions of the constant A_1 it is not difficult to show that equation (39) has the single root

$A_1 = c$. However, the solution (35) becomes invalid at such a value of A_1. Therefore the case $A_1 = c$ should be treated separately. As a result of this treatment the solution of the equation (31) can easily be obtained in the form

$$v(t) = \frac{2}{b}\left\{-\frac{\pi}{4} - \operatorname{arctg}\left[\frac{1}{\pm b\sqrt{c}(t+A_2)}\right]\right\} \tag{40}$$

by using (34) where A_2 is an integration constant. From the condition $v(0) = 0$ we find that

$$A_2 = \mp\frac{1}{b\sqrt{c}}$$

and from the transversality condition we easily find that the '$-$' sign should be taken in formula (40) before $\sqrt{c}$. Thus in the case $A_1 = c$ the solution of equation (31) under conditions (33) is given by the formula

$$v(t) = \frac{2}{b}\left[-\frac{\pi}{4} + \operatorname{arctg}\left(\frac{1}{1+(b\sqrt{c})t}\right)\right]. \tag{41}$$

In the case of a shock moving from the left to right $m < 0$, $V_2 > V_1$ and then we get from (30) that $dV/dv > 0$ at $|v(t)| < \pi/(2b)$. Consequently the use of the solution (30) in the functional (6) is valid only at such values of $\zeta(t)$ that

$$|v(t)| \leq \pi/(2b) \tag{42}$$

Employing (41) it is easy to show that at any $a > 0$ in (15) the inequality (42) holds. It is easy to get an upper estimate for $|v(t)|$ from (41). Really, with (26) in view

$$|v(t)| \leq h(\pi/2)[2a/(\gamma+1)]^{0.5}, \qquad t \geq 0 \tag{43}$$

If, in particular

$$a \leq 2(\gamma+1)/(\pi^2) \tag{44}$$

then $|v(t)| \leq h$. Since at $t > 0$ the inequality

$$\operatorname{arctg}[(1+(b\sqrt{c})t)^{-1}] < \pi/4$$

holds, it follows from (41) that $v(t) < 0$. This means that at $t > 0$ the shock front calculated by the extremum of the functional (6), (29), (30) lags behind the true shock front. It is interesting to note that in the preceding example with Burgers' equation the situation was the opposite. Substituting the exact solution (41) into the functional (6), (29), (30) it is easy to find that $I(\zeta) = 0$. Although the deviation $|v(t)|$ becomes greater as t increases, under proper choice of the dimensionless coefficient a in (15) one can obtain that the inequality $|v(t)| \leq h$ will be valid for any $t \geq 0$. However, (44) yields values of the quantity a too small for the successful application of the pseudoviscosity (15) (cf. Reference 5). Thus in practice the value $|v(t)|$ can reach a magnitude of several intervals h. On the other hand it follows from the estimate (43) that the difference $v(t)$ tends to zero with the mesh size. However, in practical computations we always use finite mesh sizes, furthermore, the computing mesh is generally more crude in multidimensional computations.

3. MODIFICATION OF THE BASIC FUNCTIONAL

In this section we derive and study some modifications of the basic functional (4)–(6) with purpose of increasing the discontinuity localization accuracy when an optimization procedure is used which is based on such a functional.

Let us write the Euler equation system for an inviscid, compressible, non-heat-conducting gas as follows

$$\partial w/\partial t + \partial \varphi(w)/\partial x = 0 \tag{45}$$

where

$$w = \begin{pmatrix} \rho \\ \rho u \\ \rho E \end{pmatrix}, \qquad \varphi(w) = \begin{pmatrix} \rho u \\ p + \rho u^2 \\ \rho u + \rho u E \end{pmatrix}, \qquad E = \varepsilon + u^2/2 \tag{46}$$

The system (45), (46) is completed by the equation of state (20). Let us approximate (45), (46) by a conservative difference scheme of the rth order of accuracy ($1 \le r$). As was shown previously,[10] the first differential approximation (f.d.a.)[22,24] of such a difference scheme is representable in divergence form[10]

$$\partial w/\partial t + \partial \varphi(w)/\partial x = \partial Q(x, t)/\partial x \tag{47}$$

In a more detailed form the vector $Q(x, t)$ may be written as

$$Q(x, t) = \tilde{Q}(w(x, t), h\, \partial w(x, t)/\partial x, \ldots, h^r\, \partial^r w(x, t)/\partial x^r, x, t, h, \tau) \tag{48a}$$

and

$$Q = O(h^r) + O(\tau^r) \tag{48b}$$

In (48) τ is the time step and h is the step of the uniform computing mesh on the x-axis. In accordance with (46), (47) let us introduce the notations

$$w = \{w_1, w_2, w_3\}, \qquad \varphi = \{\varphi_1, \varphi_2, \varphi_3\}, \qquad Q = \{Q_1, Q_2, Q_3\}$$

where

$$w_1 \equiv \rho, \qquad w_2 \equiv \rho u, \qquad w_3 \equiv \rho E, \qquad \varphi_1 = \rho u, \qquad \varphi_2 = p + \rho u^2, \qquad \varphi_3 = pu + \rho u E$$

As in the foregoing section let us consider the progressive wave type solutions of the system (47), that is the solutions depending upon the variable y defined by (16). Then the system (47) may be integrated once:

$$-Dw(y) + \varphi(w) + C = Q(y) \tag{49}$$

where C is a constant vector. As in References 7, 10 and 19 we search for the solution of the system (49) which satisfies the conditions

$$w(y) = \begin{cases} W_1, & y \to -\infty \\ W_2, & y \to +\infty \end{cases} \tag{50}$$

where W_1 and W_2 are constant vectors satisfying the Rankine–Hugoniot conditions. Then the vector C in (49) should satisfy the requirements

$$C = DW_1 - \varphi(W_1) = DW_2 - \varphi(W_2) \tag{51}$$

As was shown earlier[25,19] the exact solution of the problem (49), (50) describes well the actual behaviour of the difference solution in the zone of a smeared strong discontinuity. For example, the corresponding relative error obtained in Reference 19 did not exceed 4 per cent for shock waves of finite intensity. Let in (50)

$$W_j = \{W_{j1}, W_{j2}, W_{j3}\}, \qquad j = 1, 2$$

Employing the formula (51) let us rewrite equation (49) for the kth component ($1 \le k \le 3$) as

$$[\varphi_k(w) - \varphi_k(W_2) - Q_k(y)]/[w_k(y) - W_{2k}] - D = 0 \tag{52}$$

Let us now consider, on the basis of (52), the functional

$$I_k(\zeta)=\int_0^T\left[\frac{\varphi_k(w(\zeta(t),t))-\varphi_k(W_2)-Q_k(\zeta(t),t)}{w_k(\zeta(t),t)-W_{2k}}-\dot{\zeta}(t)\right]^2 \mathrm{d}t \tag{53}$$

Theorem 2

If (a) the first differential approximation of a difference scheme approximating the system (45), (46) is representable in the form

$$\frac{\partial w}{\partial t}+\frac{\partial \varphi(w)}{\partial x}=\frac{\partial}{\partial x}\left[B(w, w_x, h, \tau)\frac{\partial w}{\partial x}\right] \tag{54}$$

where B is a 3×3 matrix whose elements have an order of smallness $O(h^r)+O(\tau^r)$, $r\geq 1$, and (b) the system (54) possesses the smooth solution $w=w(y)$, $y=x-Dt-x_0$ which is unique up to translation along the x-axis, and which satisfies the conditions (50) where W_1, W_2 are constant vectors satisfying the Rankine–Hugoniot conditions across the steady shock wave moving at a speed D, then the solution of the variational problem

$$I_k(\zeta)\rightarrow \min, \qquad \zeta(0)=x_0 \tag{55}$$

$$\frac{\varphi_k(w(\zeta(T),T))-\varphi_k(W_2)-Q_k(\zeta(T),T)}{w_k(\zeta(T),T)-W_{2k}}-\dot{\zeta}(T)=0 \tag{56}$$

in the case of a steady shock wave coincides with the exact shock trajectory.

Proof. By analogy with (4)–(6) let us introduce the notations

$$J_k(\zeta(t),t)=[\varphi_k(w(\zeta(t),t))-\varphi_k(W_2)-Q_k(\zeta(t),t)]/[w_k(\zeta(t),t)-W_{2k}] \tag{57}$$

$$F_k(\zeta(t),\dot{\zeta}(t),t)=J_k(\zeta(t),t)-\dot{\zeta}(t)$$

$$I_k(\zeta)=\int_0^T F_k^2(\zeta(t),\dot{\zeta}(t),t)\,\mathrm{d}t \tag{58}$$

Then the Euler–Lagrange equation corresponding to the functional (58) has the form

$$\ddot{\zeta}-\frac{\partial J_k}{\partial t}-J_k\frac{\partial J_k}{\partial \zeta}=0 \tag{59}$$

Since we have that $w=w(y)$, equation (59) may be rewritten by analogy with (28) as

$$\ddot{\zeta}-(J_k-D)\frac{\mathrm{d}J_k}{\mathrm{d}y}(\zeta(t)-Dt-x_0)=0 \tag{60}$$

It is easy to find from (57) with (52) in view that

$$J_k(y)=D \tag{61}$$

In regard to (61) it is easy to obtain from (60) that $\ddot{\zeta}(t)=0$, therefore, the general solution of equation (60) has the form

$$\zeta(t)=c_1t+c_2 \tag{62}$$

where c_1, c_2 are integration constants. It follows from the condition $\zeta(0)=x_0$ that $c_2=x_0$. To determine c_1 let us make use of the transversality condition (56). It follows from (52) that

$$Q_k(y)=\varphi_k(w)-\varphi_k(W_2)-Dw_k+DW_{2k} \tag{63}$$

Substituting (63) into (56) we obtain that $\dot{\zeta}(T)=D$. Therefore it follows from the transversality condition (56) that $c_1=D$ in (62).

Example 1

Consider again the system (17)–(19). In this case

$$Q(w)=\{0,-q,-qu\}$$

where q, u can be uniquely determined as functions of the specific volume V by means of the formulae (23). In its turn the formula (25) provides a smooth and unique transition from the state '1' behind the shock front to the state '2' before the front in the case of the quadratic artificial viscosity (15) (see also Reference 5). Thus in this case the conditions of Theorem 2 are satisfied. Earlier[9,10] we have shown that the accuracy of the shock front localization by max q depends substantially on the values of dimensionless coefficients entering the expression for q. In the case of shock localization on the basis of the functional (53) the localization result does not depend on the form of q, as follows from the proof of Theorem 2. It is only required for q to ensure a unique and smooth transition from one state to another in the steady shock wave. There is no doubt that this is a positive property of the functional (53).

Example 2

Consider the 'breakdown of discontinuity' scheme proposed by Godunov.[26] In the case $u>c>0$, where c is the speed of sound, the f.d.a. of this scheme is representable in the form[19]

$$\frac{\partial w}{\partial t}+\frac{\partial \varphi(w)}{\partial x}=\frac{\partial}{\partial x}\left[B(w,h,\tau)\frac{\partial w}{\partial x}\right] \tag{64}$$

where

$$B=\frac{h}{2}A-\frac{\tau}{2}A^2, \qquad A=\partial\varphi(w)/\partial w \tag{65}$$

Substituting $w=w(y)$, $y=x-Dt-x_0$ into (66) we can find[19] that

$$\frac{dw}{dy}=q(w)=\frac{1}{\det B}B^*(w)[\varphi(w)-Dw+C] \tag{66}$$

where B^* is the adjoint of the matrix B

$$\det B=\prod_{j=1}^{3}\left(\frac{h}{2}\lambda_j-\frac{\tau}{2}\lambda_j^2\right) \tag{67}$$

λ_1, λ_2, λ_3 are eigenvalues of the matrix A,

$$\lambda_1=u-c, \qquad \lambda_2=u, \qquad \lambda_3=u+c \tag{68}$$

By virtue of the Courant–Friedrichs–Lewy stability condition

$$(\tau/h)\max_y(|u|+c)<1 \tag{69}$$

we obtain, with (67), (68) in view, that $\det B\neq 0$. Thus the formula (66) uniquely determines

dw/dy. Consequently one of the necessary conditions for uniqueness of the solution $w(y)$ of the problem (70), (65), (50)

$$\varphi(w)-\varphi(W_2)-Dw+DW_2=B(w,h,\tau)\frac{dw}{dy} \tag{70}$$

is satisfied at $u>c>0$. It follows from (67) and (68) that at the point where $u=c$ $\det B=0$. Then from the assumption on the boundedness of $|dw/dy|$ at this point we obtain with regard to (66) that at the point $u-c=0$ a saddle-like singularity of the solution $w(y)$ of the system (66) with boundary conditions (50) takes place (see also Reference 19). Detailed investigation of the qualitative behaviour of the integral curves in the vicinity of this saddle-like singularity and in the subdomains where $c>u>0$, $u>c>0$ represents a complicated mathematical problem whose analysis goes beyond the scope of the present paper. Below, in Section 5, we present some results of practical application of the functional (53) with $k=1$. In this case the expression for $Q_1(\zeta,t)$ has the form[27]

$$Q_1(\zeta,t)=\begin{cases}-\frac{\tau}{2}(p+\rho u^2)_x+\frac{h}{2}\left[\frac{1}{c}\left(1-\frac{u}{c}\right)p_x+u\rho_x+\frac{u}{c}\rho u_x\right], & 0<u<c\\ -\frac{\tau}{2}(p+\rho u^2)_x+\frac{h}{2}(\rho u)_x, & u\geq c>0\end{cases} \tag{71}$$

It follows from these examples that, at least in the case of a stationary shock wave, the use of the functional (53) provides a result coinciding with the exact discontinuity trajectory. This fact constitutes the advantage of the functional (53) over the functional (4)–(6).

4. SHOCK LOCALIZATION ON THE BASIS OF FUNCTION MINIMIZATION

It follows from the construction of the basic functional (4)–(6) or (53) that it is necessary to store in the computer memory the values of the quantities u, p, ρ, ε found as the solution of finite difference equations approximating the Euler equation system (45), (46) in a domain $\Omega(T)$ of the (x,t) plane. To estimate the minimal size of this domain along the x-axis at $T=N\tau$ where N is a positive integer, let us make use of the stability condition (69) and Zemplén's theorem.[22] Really, in virtue of this theorem and the condition (69) the shock front cannot propagate a distance exceeding h during the time τ. Let x_0 be the discontinuity abscissa at $t=0$. Then the domain $\Omega(T)$ includes all those (x,t) points for which the following inequalities are satisfied

$$x_0-nh\leq x\leq x_0+nh,\qquad n=1,\ldots,N$$

One can get rid of the requirement of additional storage if instead of the problem on minimization of the functional (53) the minimization problem for some function $F(\zeta(t))$ is considered for required moments of time t. Let us consider the following function on the basis of (53):

$$\begin{aligned}F(\zeta(t^{n+1}))=\sum_{k=1}^{3}\propto_k\{[&\varphi_k(w^-(\zeta^{n+1},t^{n+1}))-\varphi_k(w^+(\zeta^{n+1},t^{n+1}))\\ &-Q_k(w^-,h\partial w^-/\partial x,\ldots,h^r\partial^r w^-/\partial x^r,x,t,h,\tau)]/\\ &[w_k^-(\zeta^{n+1},t^{n+1})-w_k^+(\zeta^{n+1},t^{n+1})]-\dot{\zeta}_{\tau_1}(t^{n+1})\}^2\end{aligned} \tag{72}$$

where $\zeta^{n+1} \equiv \zeta(t^{n+1})$, $t^{n+1} = (n+1)\tau$, w^-, w^+ are the solutions of the auxiliary problems P^- and P^+, respectively (cf. Section 2 and also Reference 16), $\dot{\zeta}_{\tau_1}(t^{n+1})$ is some difference approximation of the derivative $\dot{\zeta}(t^{n+1})$ employing the step $\tau_1 = \beta\tau$, $\beta \geq 1$. α_1, α_2, α_3 are penalty constants, $\alpha_k \geq 0$, $\alpha_1 + \alpha_2 + \alpha_3 = 1$. The simplest approximation for $\dot{\zeta}_{\tau_1}(t^{n+1})$ is as follows

$$\dot{\zeta}_{\tau_1}(t^{n+1}) = [\zeta(t^{n+1}) - \zeta(t^{n+1} - \tau_1)]/\tau_1 \tag{73}$$

Let us briefly consider the question of the limit trajectory $\zeta(t)$ to which the solution of the problem

$$F(\zeta(t^{n+1})) \to \min, \qquad 0 \leq n \leq N$$
$$\zeta(0) = x_0 \tag{74}$$

tends as $\tau \to 0$, $h \to 0$. Let in (72) $w_k^+ = W_{2k} = \text{const.}$ and let the function (72) be applied for stationary shock wave localization. Let $w(y)$, where y where y is defined according to (16), be the exact solution of the problem (49), (50). As has been shown[19,25] the difference solution $w_h^-(y)$ approximates well the solution $w(y)$ in the zone of a smeared shock wave at finite h, τ. Suppose now that this approximation property holds also at $\tau \to 0$, $h \to 0$, and

$$h^m \mid \partial^m w_h^-/\partial x^m - d^m w(y)/dy^m \mid \leq C_1 h^{\beta_1} + C_2 \tau^{\beta_2}, \qquad m = 0, 1, \ldots, r \tag{75}$$

where β_1, β_2 are positive constants depending on the order r of the finite difference scheme employed. If the simplest difference approximations are used for the x-derivatives entering into Q_k (for example a one-sided difference for $\partial w_h/\partial x$), then it is clear that

$$|(Q_k(w_h^-, h\, \partial w_h^-/\partial x, \ldots) - Q_k(w(y), h\, dw/dy, \ldots)| \leq C_3 h^{\beta_3} + C_4 \tau^{\beta_4} \tag{76}$$

where

$$\beta_3 = r + \min(1, \beta_1), \qquad \beta_4 = r + \beta_2$$

with the definition (48) in view. When using (73) we have that

$$[\zeta(t^{n+1}) - \zeta(t^{n+1} - \tau_1)]/\tau_1 = \dot{\zeta}(t^{n+1}) + O(\beta\tau) \tag{77}$$

In regard to the estimates (75)–(77) and the formula (52) it is easy to see that

$$F(\zeta(t^{n+1})) = \sum_{k=1}^{3} \alpha_k [D - \dot{\zeta}(t^{n+1}) + f_{1k} h^{\beta_{5k}} + f_{2k} \tau^{\beta_{6k}}]^2 \tag{78}$$

where β_{5k}, β_{6k} are positive constants,

$$\beta_{5k} = \beta_{5k}(\beta_1, \beta_3), \beta_{6k} = \beta_{6k}(\beta_2, \beta_4, \beta)$$

$f_{jk} = f_{jk}(\tilde{y}) = O(1)$, $j = 1, 2$, $k = 1, 2, 3$, $\tilde{y} = \zeta(t^{n+1}) - Dt - x_0$. It follows from (78) that

$$\lim_{h \to 0, \tau \to 0} F(\zeta(t^{n+1})) = [D - \dot{\zeta}(t^{n+1})]^2$$

Let $\delta\zeta$ be the error in determining the discontinuity abscissa when some method of numerical minimization of the function $F(\zeta(t^{n+1}))$ is used and let δF be the corresponding error in the value of the function F. Then

$$\delta F = O((\delta\zeta)^2) \tag{79}$$

according to general considerations presented in Reference 28. On the other hand it is well known that the width of the zone of discontinuity 'smearing' is $O(h)$.[5,6,22] Therefore $\delta\zeta \to 0$

as $h \to 0$ and then $\delta F \to 0$ according to (79). At each finite value of h, τ the procedure for numerical minimization of the function F (72) finds the minimal possible value of the function F. It is obvious that this minimal value is $F=0$ (in practice we obtained the values $10^{-9} \leq F \leq 10^{-7}$ at chosen h, τ and at chosen accuracy of the determination of the minimum abscissa). Therefore the limit trajectory $\zeta(t)$ satisfies the equation

$$[\dot{\zeta}(t^{n+1}) - D]^2 = 0 \tag{80}$$

Since the value $\zeta(0) = x_0$ is given, it can be easily seen from (80) that the limit trajectory $\zeta(t)$ coincides with the true discontinuity trajectory.

Thus we have shown that in the case of a stationary shock wave the minimization problem (74) and the problem (58), $1 \leq k \leq 3$, where $I_k(\zeta)$ is determined by (53), are equivalent in the sense that their solutions coincide in the limit as $h \to 0$, $\tau \to 0$. This means that at sufficiently small h, τ the solutions of these problems will be close to each other.

5. COMPUTATIONAL RESULTS

It was interesting to check by direct computation the validity of the analytic solution (41) obtained in Section 2 in the process of the analysis of Miranker and Pironneau's method. In accordance with considerations preceding the derivation of the formula (41) we have considered the problem of the motion of a one-dimensional stationary shock wave. The solution profiles of the system (17)–(19), (15) at each fixed $t \geq 0$ were set in accordance with the exact solution (23), (25). In the equation of state (20) we have taken $\gamma = 2$ as in Reference 10. The shock wave intensity was specified by imposing the following values (the meaning of the subscripts is the same as in Section 2): $p_1 = 5$, $p_2 = 1$, $u_2 = 0$, $\rho_2 = 1$. The shock wave trajectory was observed in the region $0 \leq x \leq 100h$, $0 \leq t \leq 300\tau$ where $h = 0{\cdot}02$, $\tau = 0{\cdot}002$. The width X of the 'smeared' shock wave zone in the case of pseudoviscosity (15) is given by the formula[5]

$$X = \pi h[2a/(\gamma+1)]^{0{\cdot}5} \tag{81}$$

from which we find at $a = 3$, $\gamma = 2$: $X \cong 4{\cdot}44h$. As was noted previously by several authors,[6,19,29,30] in some cases the shock wave is spread over more than 10 cells. To simulate this situation and to examine the shock localization accuracy in this case we have used in (15) also the value $a = 15$. In this case we find from (81) that $X \cong 9{\cdot}95h$. A technique for shock front localization by maximum of pseudoviscosity q was analysed in Reference 10 and it was shown that in the case of the quadratic viscosity (15) the abscissa x_* of the point of max q is placed to the left from the true abscissa x_f of the shock front, and

$$(x_* - x_f)/h = -[2a/(\gamma+1)]^{0{\cdot}5} \arcsin[(\sqrt{V_2} - \sqrt{V_1})/(\sqrt{V_2} + \sqrt{V_1})] \tag{82}$$

At $a = 3$, $\gamma = 2$, $V_1 = 0{\cdot}5$, $V_2 = 1$ we find from (82): $(x_* - x_f)/h = -0{\cdot}243847$; similarly, at $a = 15$, $\gamma = 2$, $V_1 = 0{\cdot}5$, $V_2 = 1$ we obtain $(x_* - x_f)/h = -0{\cdot}545259$.

In the numerical realization of the Miranker–Pironneau optimization approach the basic functional (6) was minimized by the steepest descent method as in References 12 and 13. The integral (6) as well as the integrals and the x-derivatives entering in formulae of the descent method presented in References 12 and 13 were approximated by the formulae of the order of accuracy $O(\tau) + O(h)$ as in References 12–16. The profiles of the quantities u, p, ρ in the zone of smeared shock wave which are necessary for computer implementation of the minimization algorithm[12,13] were stored for the moments of time $t_j = j\beta\tau$ where we have taken $\beta = 6$. By this computer memory was economized which was necessary for the above u,

p, ρ profiles. The zeroth approximation $\zeta_0(t)$ was set by using the fact that in the zone of a 'smeared' shock wave the inequality $\partial u/\partial x<0$ takes place. With this in mind we have assumed $\zeta_0(t)$ at fixed t to be equal to the abscissa of that end point of the smeared shock wave zone which corresponds to the state behind the shock front. This abscissa was determined by analogy with Reference 31. In some cases we arbitrarily moved this abscissa to the left by a distance which reached $10h$ for large values of t (this displacement was accomplished by the formula $\bar{\zeta}_0(t)=x_0+0{\cdot}7(\zeta_0(t)-x_0)$, that is only for $t>0$). The convergence of the steepest descent method described in References 12 and 13 took place also in this case. Four to twelve iterations were needed to reach an abscissa localization accuracy of the order $10^{-3}h$.

To locate the shock front by $\max q$ we determined the cell of the grid on the x-axis where $\max q$ was reached at fixed t, and an approximate abscissa $\zeta(t)$ of the shock front point was set as an abscissa of the centre of this cell. In Figures 1–5 the dimensionless time t/τ is measured along the abscissa axis, and the error

$$[x_f(t)-\zeta(t)]/h \tag{83}$$

is measured along the ordinate axis. In (83) $\zeta(t)$ is the shock front abscissa determined with the aid of some optimization algorithm and $x_f(t)$ is the shock front abscissa computed in accordance with the exact solution. It follows from Figures 1 and 2 that the trajectory $\zeta(t)=x$ obtained by numerical minimization of the Miranker–Pironneau functional (24), (5), (6) agrees very well with the analytic solution (41) at different values of the coefficient 'a' in (15). An especially good agreement between the numerical solution and the exact one (41) takes place at $a=15$, see Figure 2. We explain this by the fact that with increase in the coefficient a in (15) the profiles in the shock wave zone become more smooth, and as a consequence of this the size of truncation errors caused by the use of one-sided differences for $\partial J(\zeta(t),t)/\partial x$ in References 12–16 and of the formula of rectangles for (6) diminishes. Similar considerations explain the increase in discontinuity localization accuracy by means of the new functional (53), $k=2$, as the coefficient a in (15) increases. It should be noted that

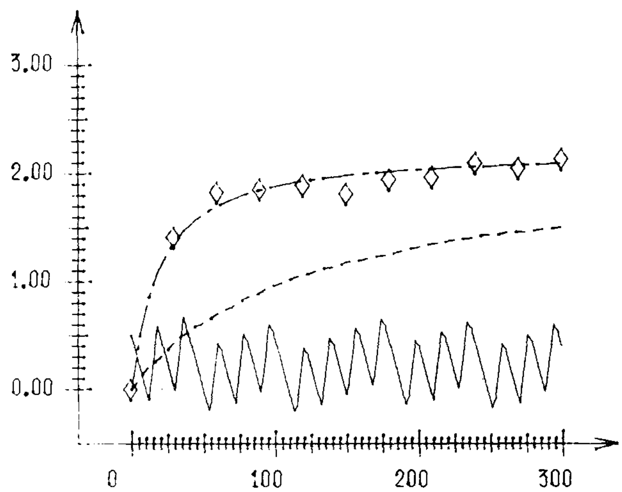

Figure 1. Localization error (83) as a function of time: $a=3$ in (15); ——— localization by $\max q$ where q is pseudoviscosity (15); – – – the use of the functional (53) with $k=2$, $Q_2=-q$; –·–·– the function $-v(t)/h$ where $v(t)$ is the function (41); ◇◇◇ the use of the original Miranker–Pironneau algorithm

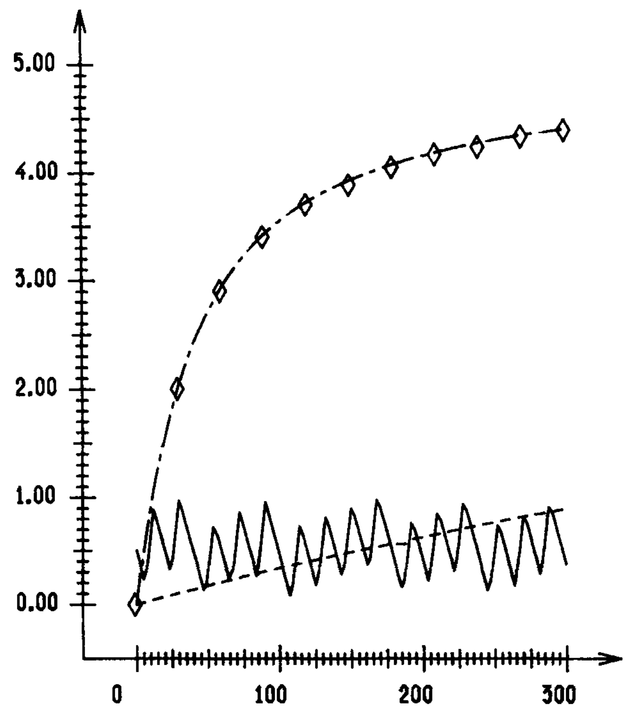

Figure 2. The same notations as in Figure 1: $a = 15$ in (15)

the use of the functional (53) provides a higher localization accuracy than in the case of the functional (24), (5), (6) (see Figures 1 and 2).

At the same input data as for Figures 1 and 2 we have used the localization method based on minimization of the function (72). The minimization of the univariate function (72) was performed on the BESM-6 computer by using a standard subroutine MNGGR entering in the mathematical software of BESM-6. The application of the Fibonacci method for finding a local minimum of the function is an essential element of the subroutine MNGGR.[32] Corresponding results for different values of a in (15) are presented in Figure 3. It is easy to

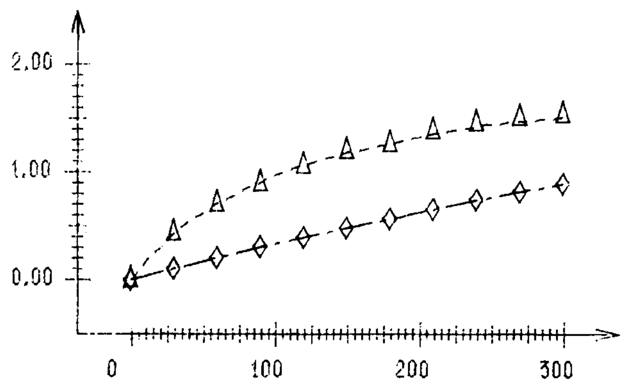

Figure 3. The use of the functional (53) with $k = 2$, $Q_2 = -q$: --- $a = 3$ in (15); -·-·- $a = 15$ in (15). Minimization of the function (72), (73) by Fibonacci method: △△△ $a = 3$ in (15), ◇◇◇ $a = 15$ in (15)

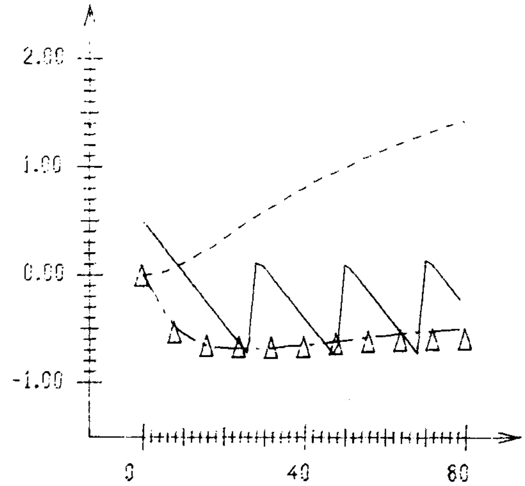

Figure 4. A problem on stationary shock wave: ——— localization by max $|Q_1|$ where Q_1 is the function (71); –·–·– the use of the functional (53) with $k = 1$, Q_1 is computed by (71); $\Delta\Delta\Delta$ minimization of th function (72), (73) by the golden section method; – – – original Miranker–Pironneau algorithm

see that the localization results on the basis of the minimization of the function (72) and on the basis of the functional (53) with $k = 2$, practically coincide. This agrees with theoretical conclusion of Section 4.

In Figure 4 the results of stationary shock wave localization obtained from the difference solution by the 'breakdown of discontinuity' scheme[26,33] are shown. In this case $h = 0{\cdot}05$, $\tau = 0{\cdot}008$, $\gamma = 2$, $p_1 = 5$, $p_2 = 1$, $u_2 = 0$, $\rho_2 = 1$ and the time interval was $0 \leq t \leq 80\tau$. It is seen from Figure 4 that the error (83) becomes greater as t increases in the case of application of the Miranker–Piranneau functional (6). The first results on the theoretical foundation of the shock localization technique based on a maximum of the approximation viscosity of a difference scheme are contained in Reference 19. In the case of the scheme of Reference 26 one of the components of the vector of the leading term of approximate viscosity is given by

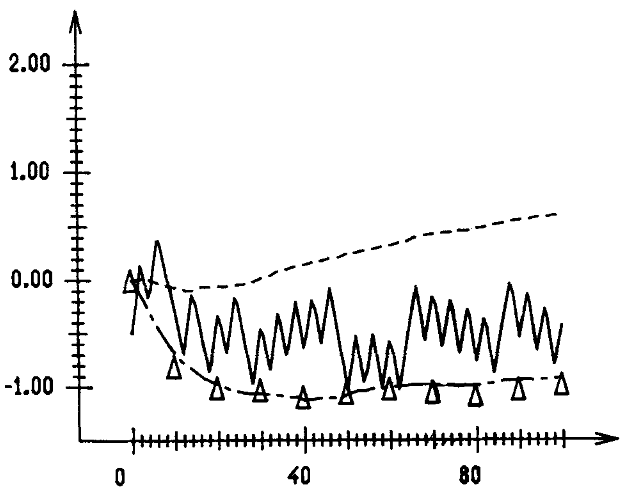

Figure 5. The same notations as in Figure 4. The problem on the propagation of an unsteady shock wave in an inhomogeneous atmosphere

the formula (71). The difference approximation of the function (71) employing one-sided differences for u_x, p_x, ρ_x represents a piecewise-constant function; more precisely, it is constant within each interval $(j-1)h \leq x \leq jh$, $j=1, 2, \ldots$. In this connection the shock front obtained by $\max|Q_1|$ will propagate stepwise, which is seen in Figures 1, 2, 4 and 5. However, if the front abscissa searched for enters explicitly in the localization algorithm employed, as takes place in the case of the functionals (6), (53) and the function (72), then the above abrupt changes in the located discontinuity abscissa vanish despite the discrete character of the difference solution on which basis the shock front was localized. This feature of localization techniques (6), (53) and (72) is easily seen in Figures 1–5. A qualitative explanation of this behaviour can be given using the example of the function (72). Let us take in (72), (73) two such values ζ_1, ζ_2 of the continuous variable $\zeta(t^{n+1})$ that $\zeta_1 \neq \zeta_2$ and, in addition,

$$\varphi_k(w^-(\zeta_1, t^{n+1})) = \varphi_k(w^-(\zeta_2, t^{n+1}))$$
$$Q_k|_{\zeta=\zeta_1} = Q_k|_{\zeta=\zeta_2}$$

Then it is easy to be convinced with the aid of (72), (73) that $F(\zeta_1) \neq F(\zeta_2)$. Minimization results of the function (72) shown in Figures 4 and 5 are obtained by the golden section method.[28,34] Only 17 evaluations of the function $F(\zeta^{n+1})$ were needed to compute the abscissa ζ^{n+1} with an error $0{\cdot}01h$.

Although the theoretical foundation of localization techniques described in the foregoing section was carried out only for the case of a stationary shock wave, it was interesting to see whether the new localization techniques described above are applicable to problems with non-stationary shock waves. We have considered a problem on a self-similar shock wave caused by an impulsive plane impact and propagating in an inhomogeneous atmosphere. The detailed formulation and the solution of this problem are contained in Chapter 12 of Reference 35. Here we only note that the profiles of the exact solution in the region behind the front are such that at $t \geq 0$ $\partial u/\partial x > 0$, $\partial p/\partial x > 0$, $\partial \rho/\partial x > 0$; in addition, $\partial \rho/\partial x > 0$ also in the undisturbed medium before the front. Corresponding profiles of the exact solution as well as the finite difference solution obtained by the 'breakdown of discontinuity scheme'[26,33] are presented in Reference 16. In this problem the exact abscissa x_f of the shock front at $t \geq 0$ is computed by the formula

$$x_f(t) = 6 \ln (t + t_0)$$

where t_0 is determined from the condition $x_0 = 6 \ln t_0$, x_0 is the given abscissa of shock front at $t=0$. In computations by the scheme[26] we have used the values $h = 0{\cdot}1$, $\tau = 0{\cdot}007$ as in Reference 16 which corresponded to the Courant number $K \cong 0{\cdot}4$. It follows from Figure 5 that the optimization algorithms proposed in Sections 3 and 4 are applicable also in the case of a non-stationary shock wave. It should be noted that in the case of both stationary (Figures 1, 2 and 4) and non-stationary (Figure 5) shock waves the inclusion in an optimization localization procedure of the information on the approximation and artificial viscosity moves the trajectory $\zeta(t)$ to the right in comparison with the original Miranker–Pironneau algorithm.[12,13]

6. CONCLUSIONS

We have carried out an investigation of shock wave localization accuracy both for Miranker and Pironneau's method[12,13] and for some of its modifications. It is shown that the methods of classical variational calculus can be successfully applied for theoretical analysis of the original algorithm[12,13] as well as its modification (53). In particular, the results of this analysis of the method[12,13] indicate a non-coincidence of the extremal trajectory obtained as a result of minimization of the basic functional (5), (6), (24) with the true discontinuity trajectory at finite mesh sizes. At the same time an analysis of the functional (53) yields a coincidence of the solution of corresponding variational problem with the exact trajectory of a steady shock wave. These theoretical results were confirmed by practical computations. In particular, the use of the functional (53) and the function (72), (73) yields more accurate localization results as compared to the original algorithm.[12,13]

REFERENCES

1. M. G. Hall, 'Computational fluid dynamics—a revolutionary force in aerodynamics', *AIAA 5th Computational Fluid Dyn. Conf., Palo Alto, Calif., 1981,* New York 176–188 (1981).
2. D. R. Chapman, 'Trends and pacing items in computational aerodynamics', in W. C. Reynolds and R. W. MacCormack (Eds) *Seventh International Conf. Numer. Methods in Fluid Dynamics. Proceedings, Stanford 1980* Lecture Notes in Physics, 141, Springer-Verlag, Berlin, 1981, pp. 1–11.
3. J. Von Neumann, *Theory of Self-Reproducing Automata,* University of Illinois Press, 1966.
4. V. V. Rusanov, 'Processing and analysis of computation results for multidimensional problems of aerodynamics', in H. Cabannes and R. Temam (Eds) *Third International Conf. Numer. Methods in Fluid Dynamics. Proceedings, Vol. 1,* Paris, 1972 Lecture Notes in Physics, 18 Springer-Verlag, Berlin, 1973, pp. 154–162.
5. R. D. Richtmyer and K. W. Morton, *Difference Methods for Initial-Value Problems,* 2nd edn, Wiley Interscience, New York, 1967.
6. P. J. Roache, *Computational Fluid Dynamics,* Hermosa, Albuquerque, New Mexico, 1976.
7. N. N. Yanenko and E. V. Vorozhtsov, 'Differential analysers of strong discontinuities in one-dimensional gas flow', in C. Taylor and K. Morgan (Eds) *Computational Techniques in Transient and Turbulent Flow,* Vol. 2, Pineridge Press Ltd, Swansea, 1981, pp. 59–96.
8. N. N. Yanenko, E. V. Vorožcov and V. M. Fomin, 'Differential analysers of shock waves', *Soviet Math. Dokl.,* **17,** 358–362 (1976).
9. E. V. Vorozhtsov, V. M. Fomin, N. N, Yanenko, 'Differential analysers of shock waves. Applications of the theory', *Čisl. Metody Mekhaniki Splošnoi Sredy,* **7**(6), 8–23 (1976).
10. V. M. Fomin, E. V. Vorozhtsov and N. N. Yanenko, 'Differential analysers of shock waves: Theory', *Computers and Fluids,* **4,** 171–183 (1976).
11. N. N. Yanenko, V. M. Kovenya, V. D. Lisejkin, V. M. Fomin and E. V. Vorozhtsov, 'On some methods for the numerical simulation of flows with complex structure', *Computer Methods in Appl. Mechanics and Engineering,* **17/18** (Pt III), 659–671 (1979).
12. W. L. Miranker, O. Pironneau, 'A global shock fitting method', *Rapport de Recherche No. 123,* IRIA, 1975.
13. W. L. Miranker and O. Pironneau, 'An example of a global shock fitting method', *An International Journal Computers & Mathematics with Applications,* **2,** 63–71 (1976).
14. Z. Usakov, 'on the localization of a discontinuity in the numerical solution of the Buckley–Leverett problem', *Čisl. Metody Mekhaniki Splošnoi Sredy,* **10**(6), 141–149 (1979).
15. Z. Usakov, 'On a numerical experiment on discontinuity localization in the solution of the filtration problem for a three-phase incompressible fluid, in A. N. Konovalov (Ed.) *Čislennkye Metody Rešeniia Zadač Fil'trazii Mnogofaznoĭ Nesžimaiemoĭ Žiskosti. Trudy IV Vsesoiusnogo Seminara* Novosibirsk, ITPM SO AN SSSR, 1980, pp. 241–245.
16. E. V. Voroszhtsov, V. M. Krepkii, Z. Usakov, 'On an optimization approach to problems of shock wave localization by shock-capturing computational results', *Čisl. Metody Mehaniki Splošnoi Sredy,* **12**(4), 30–47 (1981).
17. T. H. Chong, 'A variable mesh finite difference method for solving a class of parabolic differential equations in one space variable', *SIAM J. Numer. Anal.,* **15,** 835–857 (1978).
18. B. L. Lohar and P. C. Jain, 'Variable mesh cubic spline technique for N-wave solution of Burgers' equation', *J. Comput. Phys.,* **39,** 433–442 (1981).
19. E. V. Vorozhtsov and N. N. Yanenko, 'On some algorithms for shock wave recognition by shock-capturing computational results', *Computers and Fluids,* **8,** 313–326 (1980).
20. E. V. Vorozhtsov, 'Numerical tests of differential analysers of shock waves', *Čisl. Metody Mehaniki Splošnoi Sredy,* **8**(2), 12–27 (1977).
21. L. Ya. Zlaf, *Variational Calculus and Integral Equations. Handbook,* 2nd edn., 'Nauka', Moscow, 1970.
22. B. L. Rozhdestvensky and N. N. Yanenko, *Systems of Quasilinear Equations and Their Applications to Gas Dynamics,* 2nd edn, 'Nauka', Moscow, 1978.
23. M. Abramowitz and I. Stegun (Eds.), *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables,* National Bureau of Standards Applied Mathematics Series 55, 1964.

24. Yu. I. Shokin, *Differential Approximation Method*, 'Nauka', Novosibirsk, 1979.
25. A. Lerat, 'Numerical shock structure and nonlinear corrections for difference schemes in conservation form', in H. Cabannes, M. Holt and V. Rusanov (Eds.) *Sixth International Conf. Numer. Methods in Fluid Dynamics. Proceedings, Tbilisi 1978*, Lecture Notes in Physics, 90, Springer-Verlag, Berlin, 1979, pp. 345–351.
26. S. K. Godunov, A. V. Zabrodin, G. P. Prokopov, 'The difference scheme for two-dimensional unsteady problems in gas dynamics and the calculation of flows with a detached shock wave', *Zhurn. Vych. Mat. i Mat. Fiz.*, **1,** 1020–1050 (1961).
27. E. V. Vorozhtsov and N. N. Yanenko, 'On the theory of differential analysers of contact discontinuities in one-dimensional flows', *Computers and Fluids*, **9,** 1–32 (1981).
28. N. N. Kalitkin, *Numerical Methods*, 'Nauka', Moscow, 1978.
29. A. N. Minailos, 'On the importance of monotonicity of finite difference schemes in shock-capturing methods', *Zhurn. Vych. Mat. i Mat. Fiz.*, **17,** 1058–1063 (1977).
30. M. L. Wilkins, 'Calculation of elastic-plastic flow', in B. Alder, S. Fernbach and M. Rotenberg (Eds) *Methods in Computat. Phys. Vol. 3*, Academic Press, New York and London, 1964, pp. 211–263.
31. V. M. Fomin, E. V. Vorozhtsov and N. N. Yanenko, 'On the properties of curvilinear shock waves "smearing" in calculations by the particle-in-cell method', *Computers and Fluids*, **7,** 109–121 (1979).
32. D. I. Batishchev, *Search Methods of Optimal Design*, 'Sovetskoie Radio', Moscow, 1975.
33. M. Holt, *Numerical Methods in Fluid Dynamics*, Springer, Berlin, 1977.
34. Ph. E. Gill, W. Murray and M. H. Wright, *Practical Optimization*, Academic Press, New York, 1981.
35. Ya. B. Zel'dovich and Yu. P. Raizer, *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*, Vols 1 and 2 (Translated from Russian), Academic Press, New York, 1967.