SIAM J. NUMER. ANAL.
Vol. 52, No. 1, pp. 207–219
© 2014 Society for Industrial and Applied Mathematics

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

NUMERICAL DEFECTS OF THE HLL SCHEME AND DISSIPATION MATRICES FOR THE EULER EQUATIONS*

YUE WANG† AND JIEQUAN LI‡

Abstract. The Harten–Lax–van Leer scheme is popularly used in the CFD community. However, oscillations are observed from shock tube problems for stiffened gases when the adiabatic index is greater than 3. To understand this phenomenon, the dissipation effect of the scheme is evaluated quantitatively in terms of dissipation matrices. We have proven and numerically demonstrated that lack of positive definiteness is the root of numerical defects.

Key words. Euler equations, HLL scheme, stiffened gases, numerical defects, dissipation matrices

AMS subject classifications. 76M12, 35Q31, 35L65

DOI. 10.1137/130917752

1. Introduction.

Godunov-type schemes play a dominant role in computing solutions of hyperbolic conservation laws of the form

$$U_t + F(U)_x = 0, \tag{1.1}$$

where U is a scalar or a vector of conservative quantities. One of key elements of these schemes is to construct numerical fluxes properly, which boils down to solving a local Riemann problem at each cell interface. Such a technique is termed *the Riemann solver* that has many exact or approximate variants [14]. One of them was proposed by Harten, Lax, and van Leer and the resulting scheme is the HLL scheme [5] that is extensively used due to a lot of preferable properties such as simplicity, robustness, and entropy satisfaction.

However, our study shows that such a robust scheme may produce local undershoots of numerical solutions. Consider (1.1) for the one-dimensional (1-D) compressible Euler equations

$$U = (\rho, \rho u, E)^\top, \quad F(U) = (\rho u, \rho u^2 + p, u(E+p))^\top, \tag{1.2}$$

where ρ is the density, u is the velocity, p is the pressure, $E = \rho(\frac{1}{2}u^2 + e)$ is the total energy, and e is the internal energy that links with ρ and p through the law of thermodynamics. In this paper we are interested in the case of stiffened gases, for which the equation of state (EOS) is

$$\rho e = \frac{p + \gamma p_\infty}{\gamma - 1}, \tag{1.3}$$

*Received by the editors April 19, 2013; accepted for publication (in revised form) November 22, 2013; published electronically January 28, 2014.
http://www.siam.org/journals/sinum/52-1/91775.html

†Institute of Applied Physics and Computational Mathematics, Beijing, 100094, People's Republic of China (yue.wang0828@gmail.com). This author was supported by National Foundation of Science in China with 91130021.

‡School of Mathematical Sciences, Beijing Normal University, Beijing, 100875, People's Republic of China (jiequan@bnu.edu.cn). This author was supported by NSFC (91130021,11371063,11031001), the Doctoral program from Educational Ministry (20130003110004), Innovation Funds from Beijing Normal University (2012LZD08), and an open project from the Institute of Applied Physics and Computational Mathematics, Beijing.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

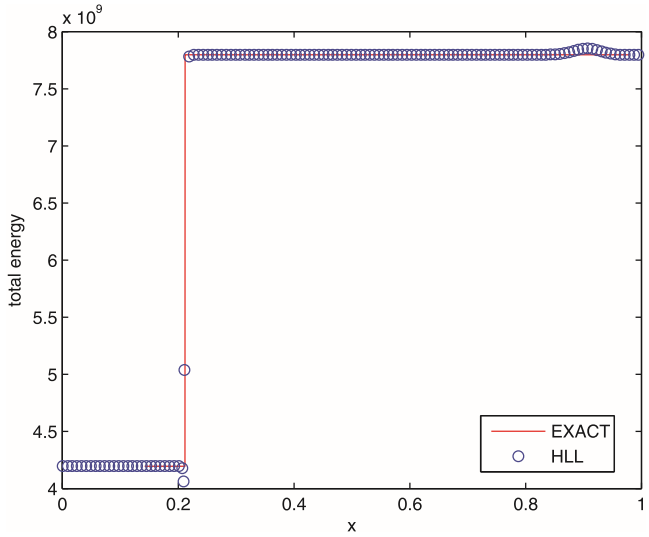

FIG. 1.1. *The numerical result computed by the HLL scheme is compared with the exact solution of the shock tube problem* (1.2)–(1.4). $N = 500$ *grid points are used but only* 100 *grids are shown. The CFL number is* $\nu = 0.5$. *Undershoots are observed near the shock wave.*

where p_∞ is a stiffness parameter and $\gamma > 1$ is the adiabatic index. The choice $p_\infty = 0$ recovers the case of ideal gases.

Under very high pressure, liquids and solids become compressible and behave like gases and the Euler equations (1.2) become a valid model for such flows. Then it is plausible to describe these materials under high pressure with reasonable accuracy by using the stiffened gas EOS [3, 4]. The values of γ and p_∞ can be used to describe the material properties of interest and can be determined from laboratory experiments via an empirical fit [7, 10]. For example, for water we have $\gamma = 4.4$ and $p_\infty = 6 \times 10^8$ [12] which will be applied to the following example.

The numerical test is about a shock tube problem. The initial data are so designed that there is only one shock emanating from an initial discontinuity, i.e.,

$$(1.4)\quad (\rho, u, p)(x, 0) = \begin{cases} (10^3,\ 2.5 \times 10^3,\ 10^9) & \text{for } \ 0 < x < 0.5, \\ (1.497 \times 10^3,\ -247.382,\ 2.372 \times 10^{10}) & \text{for } \ 0.5 < x < 1. \end{cases}$$

For this case, γ is chosen to be 4.4 and p_∞ is 6×10^8. Very high pressures are given initially. The numerical result is exhibited at $t = 5 \times 10^{-5}$ in Figure 1.1 using the HLL scheme. It is observed that there are undershoots near the shock. This phenomenon is against our common sense that the HLL scheme is thought to be very dissipative generally. Moreover, a second order calculation is carried out in the discussion section (see Figure 5.1) to further strengthen our observation, which shows that such defects cannot be effectively suppressed by the strategy of high order calculation and they do not result from the insufficiency of accuracy of the scheme.

In order to understand the inherence of such numerical defects, the modified equation approach is employed here to investigate the dissipation behavior of the scheme under investigation. The modified equation approach was first proposed in [15, 18] and then studied in various contexts such as conservation laws involving

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

shocks [2] and linear systems [9]. The past numerical scheme analysis based on the modified equation approach is typically described as heuristic only. However, the main purpose of this paper is to use this approach for the stability analysis of the scheme for the Euler equations and quantify the dissipative effect by introducing dissipative matrices. Although only the HLL scheme is studied in this paper, it is possible to use this approach for investigating other schemes quantitatively.

Our approach is the following. We take the HLL scheme and write it in a viscosity form

$$U_j^{n+1} = U_j^n - \frac{\lambda}{2}[F_{j+1}^n - F_{j-1}^n] + \frac{1}{2}[\widetilde{Q}_{j+\frac{1}{2}}^n (U_{j+1}^n - U_j^n) - \widetilde{Q}_{j-\frac{1}{2}}^n (U_j^n - U_{j-1}^n)], \tag{1.5}$$

where conventional notations are adopted: Δt is the time step, Δx is the spatial cell size, $t_n = n\Delta t$, $x_j = j\Delta x$, $n = 0, 1, 2 \ldots, j = 1, \ldots, N$, $\lambda = \Delta t/\Delta x$, U_j^n approximates $U(x_j, t_n)$, $F_j^n = F(U_j^n)$, and $\widetilde{Q}_{j+\frac{1}{2}}^n$ is the *numerical viscosity coefficient matrix*. Then the modified equation of (1.5) has the form

$$\widetilde{U}_t + F(\widetilde{U})_x = \frac{\Delta x}{2\lambda}[\tilde{\beta}(\widetilde{U}, \lambda)\widetilde{U}_x]_x, \tag{1.6}$$

where $\widetilde{U}$ is assumed to be a polynomial interpolated through the mesh values of U in the neighborhood of the mesh point (x_j, t_n). The second order and even higher order error terms are suppressed in (1.6) in the context of modified equations. The *dissipation matrix* $\tilde{\beta}(\widetilde{U}, \lambda)$ is defined as

$$\tilde{\beta}(\widetilde{U}, \lambda) = \widetilde{Q}(\widetilde{U}, \lambda) - \lambda^2 A^2(\widetilde{U}), \quad A(\widetilde{U}) = \frac{\partial F(\widetilde{U})}{\partial \widetilde{U}}. \tag{1.7}$$

The term in the right-hand side of (1.6) takes the dissipation effect, for which a necessary condition is the positive definiteness of $\tilde{\beta}(\tilde{U}, \lambda)$. As far as the HLL scheme is concerned, we find out that the positive definiteness of the dissipation matrix $\tilde{\beta}(\tilde{U}, \lambda)$ may be broken down, which occurs for cases with $\gamma > 3$. This explains defects observed in Figure 1.1. The tilde over U will be suppressed for notational simplicity if no confusion is caused.

The positive definiteness of $\widetilde{\beta}(U, \lambda)$ (the definition is provided in Appendix B) in (1.7) is not so obvious even for a very small ratio λ, since $\widetilde{Q}(U, \lambda)$ is not symmetric and the classical theory of symmetric matrices cannot be applied directly. In particular, the fact that all positive eigenvalues are positive does not imply the positive definiteness of asymmetric matrices. The following matrix is just a simple example:

$$A = \begin{pmatrix} 1 & 0 \\ -5 & 2 \end{pmatrix}, \; L = (x, y)^\top,$$
$$L^\top A L = x^2 - 5xy + 2y^2.$$

Then eigenvalues are positive 1 and 2, but the matrix A is not positive definite. Hence we have to analyze the dissipation matrix $\widetilde{\beta}(U, \lambda)$ in detail.

The positive definiteness of the dissipation matrix is particularly essential because this property not only provides the dissipation mechanism, but it enables the limit solution of (1.5) to satisfy the well–known entropy inequality as well,

$$V(U)_t + G(U)_x \le 0, \tag{1.8}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

in the sense of distributions, where $V(U)$ is the convex entropy function and $G(U)$ is the associated entropy flux satisfying $G'(U) = V'(U)F'(U)$ [6]. On the other hand, the modified equation analysis has not been done quantitatively for nonlinear systems before in literature, to our best knowledge, although this concept seems to be widely accepted and heuristically used in practice. Therefore, it is interesting to investigate the dissipation property of the HLL scheme for the compressible Euler equations.

This paper is organized in five sections. Besides the introduction section here, the derivation of modified equations for systems of conservation laws is provided in section 2. The general form of the dissipation matrices of the HLL scheme is evaluated in section 3. In section 4 the positive definiteness of the resulting dissipation matrices is analyzed for the Euler equations under certain constraints of γ. Finally we present a discussion in section 5.

2. A simple derivation of modified equations for systems. Although the modified equation was derived originally in [15] and extensively studied for scalar equations [2] and linear systems [9], the derivation for nonlinear systems seems never to have been done quantitatively before, to our knowledge. So we give a simple calculation below at least to make this study self-contained.

Consider conservative schemes for (1.1):

$$U_j^{n+1} = U_j^n - \lambda[F^n_{j+\frac{1}{2}} - F^n_{j-\frac{1}{2}}], \tag{2.1}$$

where $F^n_{j+\frac{1}{2}}$ is the numerical flux which is the approximation of the flux through $x = x_{j+\frac{1}{2}}$ over the time interval $[t_n, t_{n+1})$,

$$F^n_{j+\frac{1}{2}} \sim \frac{1}{\Delta t}\int_{t_n}^{t_{n+1}} F(U(x_{j+\frac{1}{2}}, t))dt. \tag{2.2}$$

As it is written in viscosity form (1.5), the numerical viscosity coefficient $\widetilde{Q}^n_{j+\frac{1}{2}}$ is defined as

$$\widetilde{Q}^n_{j+\frac{1}{2}}(U^n_{j+1} - U^n_j) := \lambda[F^n_{j+1} - 2F^n_{j+\frac{1}{2}} + F^n_j]. \tag{2.3}$$

Then we derive the modified equation using the viscosity form (1.5).

As a convention, we denote by $\tilde{U}(x,t)$ a polynomial interpolated through the mesh values of U in the neighborhood of the mesh point (x_j, t_n). Then it is substituted in the scheme (1.5) so that we can obtain

$$\tilde{U}_t + \frac{\Delta t}{2}\tilde{U}_{tt} + F(\tilde{U})_x = \frac{\Delta x}{2\lambda}[\widetilde{Q}(\tilde{U}, \lambda)\tilde{U}_x]_x + \mathcal{O}(\Delta x)^2 + \mathcal{O}(\Delta t)^2. \tag{2.4}$$

We continue to differentiate it with respect to the time variable t to obtain

$$\tilde{U}_{tt} + \frac{\Delta t}{2}\tilde{U}_{ttt} + F(\tilde{U})_{xt} = \frac{\Delta x}{2\lambda}[\widetilde{Q}(\tilde{U}, \lambda)\tilde{U}_x]_{xt} + \mathcal{O}(\Delta x)^2, \tag{2.5}$$

where Δt is replaced by Δx using the CFL constraint $\lambda S = \nu < 1$, where S is the maximum wave speed. Inserting (2.5) into (2.4) yields

$$\tilde{U}_t + F(\tilde{U})_x = \frac{\Delta x}{2\lambda}[\widetilde{Q}(\tilde{U}, \lambda)\tilde{U}_x + \lambda^2 F(\tilde{U})_t]_x + \mathcal{O}(\Delta x)^2. \tag{2.6}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

In order to eliminate the term of the time derivative $F(\tilde{U})_t$, we use (2.4) again to produce

$$F(\tilde{U})_t = -A^2(\tilde{U})\tilde{U}_x + \mathcal{O}(\Delta x). \tag{2.7}$$

We proceed and replace $F(\tilde{U})_t$ by $-A^2(\tilde{U})\tilde{U}_x$ in (2.6) by ignoring high order error terms. Then the modified equation (1.6) is derived.

Thus the dissipation matrix is expressed in terms of the numerical viscosity coefficient $\widetilde{Q}(U,\lambda)$ and the Jacobian matrix $A(U)$,

$$\tilde{\beta}(U,\lambda) = \widetilde{Q}(U,\lambda) - \lambda^2 A^2(U). \tag{2.8}$$

The main issue remains to discuss its positive definiteness for specific schemes.

3. Dissipation matrices of the HLL scheme for systems. In this section we recall the HLL scheme proposed in [5] and write the corresponding dissipation matrices for systems of hyperbolic conservation laws. The scheme can be written as (2.1) and the numerical flux $F^n_{j+\frac{1}{2}}$ is

$$F^n_{j+\frac{1}{2}} = \begin{cases} F^n_j & \text{for } S^L_{j+\frac{1}{2}} \geqslant 0, \\ F^n_{j+1} & \text{for } S^R_{j+\frac{1}{2}} \leqslant 0, \\ F_{hll} & \text{for } S^L_{j+\frac{1}{2}} < 0 < S^R_{j+\frac{1}{2}}, \end{cases} \tag{3.1}$$

where $S^L_{j+\frac{1}{2}} = S_L(U^n_j, U^n_{j+1})$ and $S^R_{j+\frac{1}{2}} = S_R(U^n_j, U^n_{j+1})$ denote the speeds of the leftmost and rightmost waves emanating from $(x,t) = (x_{j+\frac{1}{2}}, t_n)$, F_{hll} denotes

$$F_{hll} = \frac{S^R_{j+\frac{1}{2}} S^L_{j+\frac{1}{2}}(U^n_{j+1} - U^n_j) + S^R_{j+\frac{1}{2}} F^n_j - S^L_{j+\frac{1}{2}} F^n_{j+1}}{S^R_{j+\frac{1}{2}} - S^L_{j+\frac{1}{2}}}. \tag{3.2}$$

Here $S^L_{j+\frac{1}{2}}$ and $S^R_{j+\frac{1}{2}}$ are assumed to be continuous with respect to U^n_j and U^n_{j+1}.

In order to write out the dissipation matrix, we need to first define the numerical viscosity coefficient in (1.5) or (2.3). There are two cases.

1. *Nonintermediate cases.* For such cases, waves only move to one side of the interface. Thus we have $S^L_{j+\frac{1}{2}} \geqslant 0$ or $S^R_{j+\frac{1}{2}} \leqslant 0$ and

$$\widetilde{Q}^n_{j+\frac{1}{2}}(U^n_{j+1} - U^n_j) = \begin{cases} \lambda[F^n_{j+1} - F^n_j] & \text{for } S^L_{j+\frac{1}{2}} \geqslant 0, \\ -\lambda[F^n_{j+1} - F^n_j] & \text{for } S^R_{j+\frac{1}{2}} \leqslant 0. \end{cases} \tag{3.3}$$

2. *Intermediate cases.* For such cases, the interface is located between two waves, namely, $S^L_{j+\frac{1}{2}} < 0 < S^R_{j+\frac{1}{2}}$. Then the numerical viscosity coefficient $\widetilde{Q}^n_{j+\frac{1}{2}}$ is expressed as, after simple calculations,

$$\begin{aligned} &\widetilde{Q}^n_{j+\frac{1}{2}}(U^n_{j+1} - U^n_j) \\ &= \lambda\left[-\frac{2S^R_{j+\frac{1}{2}} S^L_{j+\frac{1}{2}}}{S^R_{j+\frac{1}{2}} - S^L_{j+\frac{1}{2}}}(U^n_{j+1} - U^n_j) + \frac{S^R_{j+\frac{1}{2}} + S^L_{j+\frac{1}{2}}}{S^R_{j+\frac{1}{2}} - S^L_{j+\frac{1}{2}}}(F^n_{j+1} - F^n_j)\right]. \end{aligned} \tag{3.4}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

As far as the modified equation (1.6) is concerned, we will use the Roe expression

$$\begin{aligned} F_{j+1}^n - F_j^n &= \hat{A}(U_{j+1}^n, U_j^n)[U_{j+1}^n - U_j^n], \\ \hat{A}(U_j^n, U_{j+1}^n) &= \int_0^1 A(\theta U_{j+1}^n + (1-\theta)U_j^n)d\theta, \end{aligned} \tag{3.5}$$

where $A(U)$ is the Jacobian of the flux function $F(U)$, as introduced in (1.7). Hence we write the explicit expression for the numerical viscosity matrix $\widetilde{Q}_{j+\frac{1}{2}}^n$,

(3.6)

$$\widetilde{Q}_{j+\frac{1}{2}}^n = \begin{cases} \lambda\hat{A}(U_j^n, U_{j+1}^n) & \text{for } S_{j+\frac{1}{2}}^L \geqslant 0, \\ -\lambda\hat{A}(U_j^n, U_{j+1}^n) & \text{for } S_{j+\frac{1}{2}}^R \leqslant 0, \\ \lambda\left[-\frac{2S_{j+\frac{1}{2}}^R S_{j+\frac{1}{2}}^L}{S_{j+\frac{1}{2}}^R - S_{j+\frac{1}{2}}^L} I + \frac{S_{j+\frac{1}{2}}^R + S_{j+\frac{1}{2}}^L}{S_{j+\frac{1}{2}}^R - S_{j+\frac{1}{2}}^L}\hat{A}(U_j^n, U_{j+1}^n)\right] & \text{for } S_{j+\frac{1}{2}}^L < 0 < S_{j+\frac{1}{2}}^R, \end{cases}$$

where I is the identity matrix. Then we proceed to express the dissipation matrix $\tilde{\beta}(U,\lambda)$ in (1.7), by sending U_j^n and U_{j+1}^n to U,

$$\begin{aligned} \tilde{\beta}(U,\lambda) &= \tilde{Q}(U,\lambda) - \lambda^2 A^2(U) \\ &= \begin{cases} \lambda A(U) - \lambda^2 A^2(U) & \text{for } S_L \geqslant 0, \\ -\lambda A(U) - \lambda^2 A^2(U) & \text{for } S_R \leqslant 0, \\ \lambda[-\frac{2S_R S_L}{S_R - S_L} I + \frac{S_R + S_L}{S_R - S_L} A(U)] - \lambda^2 A^2(U) & \text{for } S_L < 0 < S_R, \end{cases} \end{aligned} \tag{3.7}$$

where $S_L := S_L(U,U)$, $S_R := S_R(U,U)$, and

$$\tilde{Q}(U,\lambda) = \begin{cases} \lambda A(U) & \text{for } S_L \geqslant 0, \\ -\lambda A(U) & \text{for } S_R \leqslant 0, \\ \lambda[-\frac{2S_R S_L}{S_R - S_L} I + \frac{S_R + S_L}{S_R - S_L} A(U)] & \text{for } S_L < 0 < S_R. \end{cases} \tag{3.8}$$

It should be noticed that the positive definiteness of $\tilde{\beta}(U,\lambda)$ is not obvious even for the first case of (3.7) due to the asymmetry of the dissipation matrix, although all of the eigenvalues of $\lambda A(U)(I - \lambda A(U))$ are positive under the CFL constraint. In the next section we will provide details of the matrices for the Euler equations and analyze the positive definiteness based on Proposition B.2 (see Appendix B).

4. Dissipation matrices of the HLL scheme for the Euler equations. Before the discussion of the dissipation matrices of the HLL scheme for the Euler equations, some notation is presented for simplicity of presentation. Since there is a factor λ both in $\tilde{\beta}(U,\lambda)$ and $\tilde{Q}(U,\lambda)$ in terms of (3.7) and (3.8), we denote $\beta(U,\lambda) = \tilde{\beta}(U,\lambda)/\lambda$ and $Q(U) = \widetilde{Q}(U,\lambda)/\lambda$. Then we only need to analyze the positive definiteness of

$$\beta(U,\lambda) = Q(U) - \lambda A^2(U).$$

It should be noticed again that the dissipation matrices are usually not symmetric. Thus the analysis of the positive definiteness of $\beta(U,\lambda)$ becomes much more complicated. However, a necessary condition to guarantee the positive definiteness of asymmetric matrices is that all ordered principal minors of the concerned matrices are greater than zero (see [8, 16, 17] and Appendix B).

Let us apply the HLL scheme to the Euler equations (1.1)–(1.2). The Jacobian $A_E(U) = \frac{\partial F(U)}{\partial U}$ of (1.2) is

$$A_E(U) = \begin{pmatrix} 0 & 1 & 0 \\ \frac{1}{2}(\gamma-3)u^2 & (3-\gamma)u & \gamma - 1 \\ uc^2/(1-\gamma) + \frac{1}{2}(\gamma-2)u^3 & c^2/(\gamma-1) + \frac{1}{2}(3-2\gamma)u^2 & \gamma u \end{pmatrix}, \tag{4.1}$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

where the subscript E refers to the Eulerian frame and $c = \sqrt{\gamma(p+p_\infty)/\rho}$ is the sound speed. The eigenvalues of $A_E(U)$ are

$$\lambda_1 = u - c, \quad \lambda_2 = u, \quad \lambda_3 = u + c. \tag{4.2}$$

The Roe averaged eigenvalues [11] are used here for the estimates of the maximum velocities for the HLL scheme, that is,

$$S^L_{j+\frac{1}{2}} = S_L(U^n_j, U^n_{j+1}) = \hat{u} - \hat{c}, \ S^R_{j+\frac{1}{2}} = S_R(U^n_j, U^n_{j+1}) = \hat{u} + \hat{c}, \tag{4.3}$$

where $\hat{u}$ and $\hat{c}$ are the Roe averaged particle and sound speeds, respectively,

$$\hat{u} = \frac{\sqrt{\rho^n_j}u^n_j + \sqrt{\rho^n_{j+1}}u^n_{j+1}}{\sqrt{\rho^n_j} + \sqrt{\rho^n_{j+1}}}, \ \hat{c} = \left[(\gamma-1)\left(\hat{H} - \frac{1}{2}\hat{u}^2\right)\right]^{1/2},$$

with the enthalpy $H = (E+p)/\rho$ approximated as

$$\hat{H} = \frac{\sqrt{\rho^n_j}H^n_j + \sqrt{\rho^n_{j+1}}H^n_{j+1}}{\sqrt{\rho^n_j} + \sqrt{\rho^n_{j+1}}}.$$

It is easy to verify that if $U^n_j = U^n_{j+1} = U$, S_L and S_R return to

$$S_L = u - c, \quad S_R = u + c. \tag{4.4}$$

In the following we will be concerned with the signs of the ordered principal minors of $\beta(U,\lambda)$. We denote by $\|\beta_1\|$, $\|\beta_2\|$, and $\|\beta_3\|$ the ordered principal minors of β and by β_{ij} the entry of the ith row and jth column, respectively.

4.1. The nonintermediate cases. For nonintermediate cases, we assume

$$S_L = u - c \geqslant 0, \quad S_R = u + c \geqslant 0. \tag{4.5}$$

Then all waves move to the right. For the case that all waves move to the left, the analysis can be analyzed in parallel. Here the dissipation matrix computed by (3.7) is

$$\begin{aligned}\beta(U,\lambda) &= (\beta_{ij}) = A_E(U)(I - \lambda A_E(U)) \\ &= \begin{pmatrix} \frac{1}{2}\lambda(3-\gamma)u^2 & 1-\lambda(3-\gamma)u & (1-\gamma)\lambda \\ \beta_{21} & -\lambda c^2 + (3-\gamma)u + 3\lambda(\gamma-2)u^2 & (\gamma-1)(1-3\lambda u) \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix},\end{aligned} \tag{4.6}$$

where β_{21}, β_{31}, β_{32}, and β_{33} are

$$\begin{aligned}\beta_{21} &= \lambda u c^2 + \frac{1}{2}(\gamma-3)u^2 + \frac{1}{2}\lambda(7-3\gamma)u^3, \\ \beta_{31} &= \frac{1}{2}\lambda\frac{\gamma+3}{\gamma-1}c^2u^2 - \frac{uc^2}{\gamma-1} + \frac{1}{2}(\gamma-2)u^3 + \frac{1}{4}\lambda(9-5\gamma)u^4, \\ \beta_{32} &= -\frac{2\lambda}{\gamma-1}c^2u + \frac{c^2}{\gamma-1} + \frac{1}{2}(3-2\gamma)u^2 - \frac{1}{2}\lambda(7-5\gamma)u^3, \\ \beta_{33} &= -\lambda c^2 + \gamma u + \frac{1}{2}\lambda(3-5\gamma)u^2.\end{aligned}$$

Readers should take caution here that the matrix $\beta(U,\lambda)$ is not symmetric. The positiveness of eigenvalues implied by (4.5) does not guarantee the positive definiteness

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

of $\beta(U,\lambda)$. Hence we will have to compute the determinants of all the upper left submatrices, i.e., the ordered principal minors. They are

(4.7)

(a) $\|\beta_1(U,\lambda)\| = \frac{1}{2}\lambda(3-\gamma)u^2,$

(4.8)

(b) $\|\beta_2(U,\lambda)\| = \frac{1}{2}(3-\gamma)u^2(1-2\lambda u)(1-\lambda u) + (u^2-c^2)\lambda u\left[1-\frac{1}{2}(3-\gamma)\lambda u\right],$

(4.9)

(c) $\|\beta_3(U,\lambda)\| = u(u^2-c^2)[(1-2\lambda u)(1-\lambda u) + \lambda^2(1-\lambda u)(u^2-c^2)].$

Obviously, $\|\beta_1\| \le 0$ if $\gamma \ge 3$. Since all of the ordered principal minors (4.7)–(4.9) being greater than zero is a necessary condition to guarantee the positive definiteness of $\beta(U,\lambda)$, the following conclusions are easily derived.

THEOREM 4.1. *Assume that the CFL condition*

$$\lambda(c+|u|) < \frac{1}{2} \tag{4.10}$$

is satisfied; then

1. *if the adiabatic index $\gamma \geqslant 3$, the dissipation matrix $\beta(U,\lambda)$ (4.6) is not positive definite;*
2. *if the adiabatic index $\gamma \in (1,3)$, under the nonintermediate assumption (4.5), all the ordered principal minors (4.7)–(4.9) of the dissipation matrix $\beta(U,\lambda)$ (4.6) are positive.*

In addition, for the sonic cases $u^2 = c^2$, the dissipation matrix $\beta(U,\lambda)$ (4.6) is not positive definite since $\|\beta(U,\lambda)\| \equiv 0$.

4.2. The intermediate cases. For the intermediate cases, S_L and S_R computed by (4.3) are

$$S_L = u - c < 0, \quad S_R = u + c > 0. \tag{4.11}$$

The dissipation matrix is derived from (3.7),

$$\begin{aligned}\beta(U,\lambda) = (\beta_{ij}) &= \frac{1}{c}(c^2-u^2)I + \frac{u}{c}A_E(U) - \lambda A_E^2(U)\\ &= \begin{pmatrix} c - \frac{u^2}{c} - \frac{1}{2}\lambda(\gamma-3)u^2 & \frac{u}{c} - \lambda(3-\gamma)u & (1-\gamma)\lambda \\ \beta_{21} & \beta_{22} & (\gamma-1)\frac{u}{c}(1-3\lambda c) \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix},\end{aligned} \tag{4.12}$$

where β_{21}, β_{22}, β_{31}, β_{32}, and β_{33} are

$$\begin{aligned}\beta_{21} &= \lambda u c^2 + \frac{1}{2}(\gamma-3)\frac{u^3}{c} + \frac{1}{2}\lambda(7-3\gamma)u^3,\\ \beta_{22} &= -\lambda c^2 + c + (2-\gamma)\frac{u^2}{c} + 3\lambda(\gamma-2)u^2,\\ \beta_{31} &= \frac{1}{2}\lambda\frac{\gamma+3}{\gamma-1}c^2u^2 - \frac{u^2c}{\gamma-1} + \frac{1}{2}(\gamma-2)\frac{u^4}{c} - \frac{1}{4}\lambda(5\gamma-9)u^4,\\ \beta_{32} &= -\frac{2\lambda}{\gamma-1}c^2u + \frac{uc}{\gamma-1} + \frac{1}{2}(3-2\gamma)\frac{u^3}{c} - \frac{1}{2}\lambda(7-5\gamma)u^3,\\ \beta_{33} &= -\lambda c^2 + c + (\gamma-1)\frac{u^2}{c} + \frac{1}{2}\lambda(3-5\gamma)u^2.\end{aligned}$$

Then the following theorem can be proved.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

THEOREM 4.2. *Assume that the CFL condition* (4.10) *is satisfied. Then the following conclusions are drawn:*

1. *if* $\gamma \geq 3$*, it is possible that the dissipation matrix* (4.12) *is not positive definite;*
2. *if the adiabatic index* $\gamma \in (1, 3)$*, the ordered principal minors of the dissipation matrix* $\beta(U,\ \lambda)$ (4.12) *are positive under the intermediate assumption* (4.11).

The first part of Theorem 4.2 is easy to derive because $\|\beta_1\| = (\beta)_{11} < 0$ for some specific c and u (see the examples in section 5 below). Details of the proof for the second part of Theorem 4.2 will be provided in Appendix A.

5. Discussions. In this last section we discuss the dissipation property of the celebrated HLL scheme, as shown in Theorems 4.1 and 4.2. It turns out that the adiabatic index γ plays a crucial role in determining the dissipation effect of the HLL scheme for the Euler equations. Once $\gamma > 3$, the positive definiteness is immediately violated. The numerical example in Figure 1.1 supports this conclusion for the intermediate cases. In fact, for the intermediate numerical test (1.2)–(1.4), we have $c = 2653.3$ and $u = 2500$ for the left side of the initial discontinuity. The CFL number is chosen as $\nu = 0.5$ so that $\lambda = \nu/(c + |u|)$. We can compute that $\|\beta_1\| = (\beta)_{11} = -3.363 \times 10^5 < 0$. This means that the HLL scheme is short of dissipation and some numerical defects must be present. As a result, the numerical results in Figure 5.1 show that undershoots in energy are observed on the left side of the shock wave.

The following example is provided to support the analysis of the nonintermediate cases and the result is displayed in Figure 5.1(b). The initial data are designated by

$$(\rho, u, p)(x, 0) = \begin{cases} (10^3,\ 3 \times 10^3,\ 10^9) & \text{for } \ 0 < x < 0.5, \\ (1505.21,\ 91.83,\ 2.62 \times 10^{10}) & \text{for } \ 0.5 < x < 1. \end{cases} \tag{5.1}$$

In this case, the adiabatic index γ is 4.4 and $p_\infty = 6 \times 10^8$. It can be computed that $c^2 < u^2$ for the left side of the initial discontinuity. As $\gamma > 3$, it is obvious that in (4.7), $\|\beta_1\| = \beta_{11} = \frac{1}{2}\lambda(3 - \gamma)u^2 < 0$, which violates the positive definiteness of the dissipation matrix $\beta(U, \lambda)$. This explains the numerical results in Figure 5.1(b) where undershoots are still observed on the left side of the shock wave besides a deficient

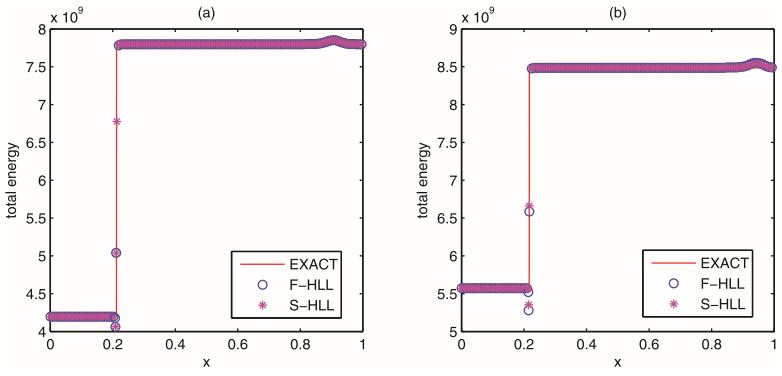

FIG. 5.1. *The numerical results computed by the first order and second order HLL schemes are compared with the exact solutions of the shock tube problems* (1.2) *and* (1.4) *(see* (a)*) and* (1.2) *and* (5.1) *(see* (b)*). $N = 500$ grid points are used but only 100 grids are shown. F-HLL represents the results for the first order HLL scheme, and S-HLL stands for the results for the second order extension of the HLL scheme with a standard limiter modification.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

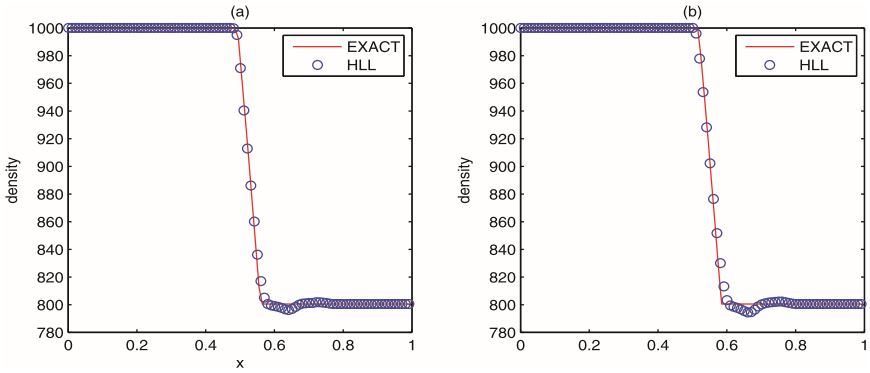

FIG. 5.2. *The rarefaction wave problems* (1.2) *and* (5.2) (a) *and* (1.2) *and* (5.3) (b). *The numerical results are shown at* $t = 5 \times 10^{-5}$. 500 *grid points are used but only* 100 *grids are shown.*

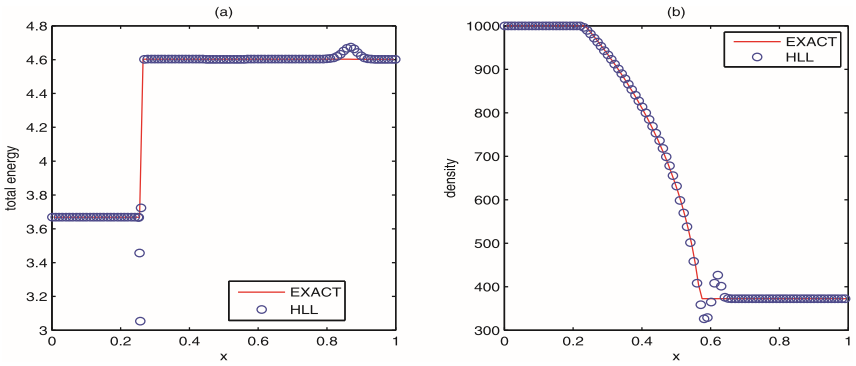

FIG. 5.3. *The shock wave* (1.2) *and* (5.4) (a) *at time* $t = 0.08$ *and rarefaction wave* (1.2) *and* (5.5) (b) *at time* $t = 0.1$ *with an ideal gas EOS.* 500 *grid points are used but only* 100 *grids are shown.*

bump. Moreover, in Figure 5.1 it is obvious that high order extension of the HLL scheme cannot suppress the oscillations ultimately.

Next, numerical solutions to another two Riemann problems are shown in Figure 5.2. The initial data of these two tests are provided as

$$(\rho, u, p)(x, 0) = \begin{cases} (10^3,\ 2.5 \times 10^3,\ 10^9) & \text{for } 0 < x < 0.5, \\ (800.49,\ 2991.62,\ 10^6) & \text{for } 0.5 < x < 1, \end{cases} \tag{5.2}$$

$$(\rho, u, p)(x, 0) = \begin{cases} (10^3,\ 3 \times 10^3,\ 10^9) & \text{for } 0 < x < 0.5, \\ (800.49,\ 3491.62,\ 10^6) & \text{for } 0.5 < x < 1, \end{cases} \tag{5.3}$$

respectively. These two Riemann problems together with those in Figures 1.1 and 5.1 are all constructed based on the data of the liquid side of a two-phase shock tube problem. Thus the adiabatic index is selected to be $\gamma = 4.4$ and $p_\infty = 6 \times 10^8$. We observe that the solutions wiggle at the tails of the rarefaction waves.

Furthermore, oscillations are also observed in the numerical tests (see Figure 5.3) with an ideal gas EOS ($p_\infty = 0$) and $\gamma > 3$, although it may be physically meaningless.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

The initial values of two Riemann problems are provided as

$$(\rho,u,p)(x,0)=\begin{cases}(7,\ 1,\ 1.01) & \text{for } 0<x<0.5,\\ (9.1429,\ 0.0578,\ 27.5225) & \text{for } 0.5<x<1,\end{cases} \tag{5.4}$$

$$(\rho,u,p)(x,0)=\begin{cases}(10^3,\ 0,\ 10^3) & \text{for } 0<x<0.5,\\ (372.759,\ 0.8362,\ 1) & \text{for } 0.5<x<1,\end{cases} \tag{5.5}$$

with the same adiabatic index $\gamma=7$ and $p_\infty=0$.

To summarize the above analysis of numerical examples, this study has several implications:

1. In past studies, dissipation properties of numerical schemes are mostly illustrated in terms of scalar equations, while the dissipation is rarely evaluated quantitatively for nonlinear systems although there are various explanations from different points of view (see [13] and references therein). Hence this raises the issue of the dissipation properties of various schemes when applied to systems in practice.
2. High order accuracy does not help suppress the defects. In Figure 5.1 we carry out the second order extension of the HLL scheme for the examples (1.2) and (1.4) (a) and (1.2) and (5.1) (b). It is observed that the undershoots and bump could not be eliminated. Hence the deficient phenomena are due to the intrinsic insufficiency of numerical dissipation.
3. Once a scheme is short of necessary dissipation, the scheme should be modified. The Lax–Friedrichs scheme has the largest numerical viscosity with a uniform dissipation matrix. The acoustic approximation technique by Dukowicz in [1] has a favorable dissipation property. Nevertheless, it is quite subtle to make such modifications if the dissipation property of a scheme is not quantified.
4. The present study is restricted to stiffened gases. The system of Euler equations with a general EOS is of particular interest such as Mie–Grueisen in solid mechanics and Chaplygin gases in astrophysics. Due to the limitation of the rigorous analysis, the further extension has to be left for future study.

Appendix A. The proof of Theorem 4.2. In this appendix, details of the proof for the second part of Theorem 4.2 are presented. That part is about the intermediate case, $c^2>u^2$. As $1<\gamma<3$ and the CFL condition (4.10) holds, the ordered principal minors of $\beta(U,\lambda)$ are estimated as follows.

(a) The first ordered principal minor of $\beta(U,\lambda)$ is equal to

$$\|\beta_1\|=\beta_{11}=c-\frac{u^2}{c}-\frac{1}{2}\lambda(\gamma-3)u^2=\frac{1}{c}\left[(c^2-u^2)+\frac{1}{2}\lambda(3-\gamma)u^2c\right]>0.$$

(b) The second ordered principal minor of $\beta(U,\lambda)$ can be computed as

$$\begin{aligned}\|\beta_2\|&=c^2(1-\lambda c)+(\gamma-1)\frac{u^2}{c^2}\left[u^2\left(\frac{1}{2}-\lambda c\right)+c^2\left(1-\frac{3}{2}\lambda c\right)\right]\\ &\quad+\frac{1}{2}(3-\gamma)\lambda u^2(-2c+\lambda c^2+\lambda u^2).\end{aligned} \tag{A.1}$$

For fixed c and u, it is linear with respect to γ. It suffices to consider the cases $\gamma=1$ and $\gamma=3$.

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

For $\gamma = 1$, we have $\|\beta_2\| = c^2 - \lambda c^3 - 2\lambda u^2 c + \lambda^2 u^2 c^2 + \lambda^2 u^4$. If $u > 0$, we arrive at

$$\|\beta_2\| \geqslant c^2[1 - 2\lambda(c+u)] + \lambda^2 u^2 c^2 + \lambda^2 u^4 > 0.$$

Otherwise, if $u \leq 0$, we have

$$\|\beta_2\| \geqslant c^2[1 + 2\lambda(u-c)] + \lambda^2 u^2 c^2 + \lambda^2 u^4 > 0.$$

For $\gamma = 3$, it is easy to show

$$\|\beta_2\| = \frac{1}{c^2}(c^2 - u^2)[(c^2 - u^2)(1 - \lambda c) + \lambda u^2 c] > 0.$$

Thus the linearity in γ implies that $\|\beta_2\|$ is positive for $1 < \gamma < 3$.

(c) The determinant of $\beta(U, \lambda)$ is

$$\|\beta(U,\lambda)\| = c(c^2 - u^2)\left[1 - \lambda\left(2c + \frac{u^2}{c}\right) + \lambda^2 c^2 + \lambda^2 u^2(1 - \lambda c)\right] + \lambda^3 u^4 (c^2 - u^2).$$

If $u > 0$, $|\beta(U,\lambda)|$ is estimated as

$$\|\beta(U,\lambda)\| \geqslant c(c^2 - u^2)[1 - \lambda(2c+u) + \lambda^2 c^2 + \lambda^2 u^2(1 - \lambda c)] + \lambda^3 u^4 (c^2 - u^2) > 0.$$

Otherwise, we have

$$\|\beta(U,\lambda)\| \geqslant c(c^2 - u^2)[1 - \lambda(2c-u) + \lambda^2 c^2 + \lambda^2 u^2(1 - \lambda c)] + \lambda^3 u^4 (c^2 - u^2) > 0.$$

Thus we complete the proof of the second part of the theorem.

Appendix B. The positive definiteness of real asymmetric matrices. In the analysis of the modified equations (1.6), it is customary to require the symmetry of the dissipation matrix in the course of discussion of positive definiteness, and often concentrate on the set denoted by

$$P_n = \{L_{n\times n} | X^\top L X > 0,\ \forall X_{n\times 1} \neq 0,\ L^\top = L\}.$$

There are plenty of equivalence propositions to determine the positive definiteness of symmetric matrices.

However, the dissipation matrix $\tilde{\beta}(U,\lambda)$ is usually not symmetric for most numerical schemes. The theory for symmetric matrices no longer holds for asymmetric cases which will make our analysis much more difficult. Based on our demand, the definition of the positive definiteness for $n \times n$ real matrices should be extended as follows.

DEFINITION B.1 (see [8]). *An $n \times n$ real matrix L, where n is a positive integer, is positive definite if $(X, LX) = X^\top L X > 0$ for all nonzero column vectors $X \in \Re^n$. We shall denote the set of all such matrices (which is a subset of all $n \times n$ real matrices) as $\tilde{P}_n$, i.e.,*

$$\tilde{P}_n = \{L_{n\times n} | X^\top L X > 0,\ \forall X_{n\times 1} \neq 0\}.$$

Denote $\|L_i\|$ as the ith ordered principle minor of L. From [8, 16, 17], the following proposition is applied for our analysis in the following.

PROPOSITION B.2. *A necessary condition to guarantee a real matrix L is positive definite, i.e., $L \in \tilde{P}_n$, is that $\|L_i\|$ is positive for any $i = 1, 2, \ldots, n$.*

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Acknowledgments. Yue Wang appreciates Professor Shuanghu Wang for his kind guidance. Jiequan Li also thanks the support from the Key Laboratory in Institute of Applied Physics and Computational Mathematics, Beijing.

REFERENCES

[1] J. K. Dukowicz, *A general, non-iterative Riemann solver for Godunov's method*, J. Comput. Phys., 61 (1985), pp. 119–137.
[2] J. Goodman and A. Majda, *The validity of the modified equation for nonlinear shock waves*, J. Comput. Phys., 58 (1985), pp. 336–348.
[3] S. K. Godounov, A. Zabrodine, M. Ivanov, A. Kraiko, and G. Prokopov, *Résolution numérique des problèmes multidimensionnels de la dynamique des gaz*, Editions Mir, Moscow, 1979.
[4] F. Harlow and A. Amsden, *Fluid Dynamics*, Technial report LA-4700, Los Alamos National Laboratory, Los Alamos, NM, 1971.
[5] A. Harten, P. D. Lax, and B. van Leer, *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Rev., 25 (1983), pp. 35–61.
[6] A. Harten, P. D. Lax, C. D. Levermore, and W. J. Morokoff, *Convex Entropies and Hyperbolicity for General Euler Equations*, SIAM J. Numer. Anal., 35 (1998), pp. 2117–2127.
[7] K. S. Holian, *T-4 Handbook of Material Properties Data Bases*, Technical report LA-10160-MS, Los Alamos National Laboratory, Los Alamos, NM, 1984.
[8] C. R. Johnson, *Positive definite matrices*, Amer. Math. Monthly, 77 (1970), pp. 259–264.
[9] J. O. Langseth and R. J. LeVeque, *A wave propagation method for three-dimensional hyperbolic conservation laws*, J. Comput. Phys., 165 (2000), pp. 126–166.
[10] S. P. Marsh, *LASL Shock Hugoniot Data*, University of California Press, Berkeley, CA, 1980.
[11] P. L. Roe, *Approximate Riemann solvers, parameter vectors, and difference schemes*, J. Comput. Phys., 43 (1981), pp. 357–372.
[12] R. Saurel and R. Abgrall, *A simple method for compressible multifluid flows*, SIAM J. Sci. Comput., 21 (1999), pp. 1115–1145.
[13] H. Z. Tang, *On the sonic point glitch*, J. Comput. Phys., 202 (2005), pp. 507–532.
[14] E. F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*, Springer, Berlin, 1997.
[15] R. F. Warming and B. J. Hyett, *The modified equation approach to the stability and accuracy of finite difference methods*, J. Comput. Phys., 14 (1974), pp. 159–179.
[16] H. Yang, *Universal positive definiteness and the partial ordering of generalized inverses of a matrix*, J. Systems Sci. Math. Sci., 13 (1993), pp. 270–275.
[17] X. Yang and S. Li, *Some properties on universal positive definite matrices*, J. Xiangtan Normal Univ., 21 (2000), pp. 45–48.
[18] N. N. Yanenko and Y. I. Shokin, *First differential approximation method and approximate viscosity of difference schemes*, Phys. Fluids, 12 (1969), pp. 28–33.

Downloaded 02/12/14 to 130.203.94.99. Redistribution subject to SIAM license or copyright; see http://www.siam.org/journals/ojsa.php

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.

View publication stats