19th AIAA Computational Fluid Dynamics
22 - 25 June 2009, San Antonio, Texas

AIAA 2009-3874

2009–3874

19th AIAA Computational Fluid Dynamics Conference, June 22 – 25, 2009, San Antonio, Texas

An Analysis of a Space-Time Discontinuous-Galerkin Method for Moment Equations and Its Solid-Boundary Treatment

Loc Khieu*

The University of Michigan, Ann Arbor, MI 48109–2140 USA

Yoshifumi Suzuki†

Desktop Aeronautics, Inc., Palo Alto, CA 94303–3310 USA

Bram van Leer‡

The University of Michigan, Ann Arbor, MI 48109–2140 USA

We present progress in the numerical and physical modeling of hyperbolic-relaxation systems, in particular, those obtained as moments of the Boltzmann equation and used to describe rarefied-gas flow. Such systems have many potential numerical advantages, mainly because there are no second or higher derivatives to be approximated. This avoids accuracy problems on adaptive unstructured grids, and the source terms, though often stiff, are only local; the compact stencils facilitate code parallelization.

The greater part of the paper deals with treating the stiff source terms in the equations, which drive the system towards equilibrium while changing its eigenstructure. The stiffness issue is solved by an implicit treatment of the source terms; for faithful modeling of the short-term and long-term physical processes we have developed a space-time discontinuous-Galerkin method, the DG(1)–Hancock method, based on Huynh's "upwind moment scheme." In this paper, detailed Fourier analyses of the method for 1-D and 2-D model equations are shown, and its comparison to semi-discrete, method-of-line approach are presented.

The second part of the paper deals with the accurate formulation of solid-wall boundary conditions for such systems, using detailed information about the molecular velocity distribution in the gas. The latest numerical results based on newly derived boundary conditions are presented.

I. Introduction

The research reported here originally was aimed at modeling all flows, except free-molecular flow, by systems of hyperbolic-relaxation equations, obtained as moments of the Boltzmann equation, and at developing efficient numerical methods for these.[1] Such systems have many potential numerical advantages, mainly because there are no second or higher derivatives to be approximated. This avoids accuracy problems on adaptive unstructured grids; furthermore, the source terms, though often stiff, are only local. In addition, the compact stencils facilitate code parallelization. A single code based on hyperbolic-relaxation equations

*Graduate Student Research Assistant, Department of Aerospace Engineering, 2001 François-Xavier Bagnoud Building, 1320 Beal Ave., Student Member.

†Research Scientist, 1900 Embarcadero Rd., Suite 101.

‡Professor, Department of Aerospace Engineering, 3025 François-Xavier Bagnoud Building, 1320 Beal Ave., Fellow.

could simulate flows up to intermediate Knudsen numbers, and would be preferable to a Navier–Stokes code if hybridization with a DSMC code were needed.

In this ambitious project, named "CFD by First-Order PDE's," one major problem arose that we have not yet solved: the accurate representation of shock structures. This makes the methodology currently unsuited for supersonic/hypersonic flows, in particular, for re-entry flows. But we have validated it for subsonic and transonic flows and are concentrating on applications to flows through and around MEMS devices.[2] A multitude of analyses and numerical testing have led us to recommending discontinuous-Galerkin methods for these systems, preferably with coupled space and time operators, as opposed to Runge–Kutta time-marching.[2,3]

In the first and largest part of this paper (Sections II and III) we deal with numerical difficulties caused by the presence of the stiff relaxation-type source terms in hyperbolic-relaxation equations. While their effect is unimportant for times short compared to the relaxation time (the regime of "frozen" physics), for large times these drive the system towards an equilibrium governed by genuinely disparate physics, corresponding to a completely different eigenstructure of the mathematical model.

We have solved the stiffness issue by an implicit treatment of the source terms. For faithful modeling of both the frozen and equilibrium physics, we have developed a space-time discontinuous Galerkin method, the DG(1)–Hancock method; it is based on Huynh's "upwind moment scheme.[4]" In Sections II and III we concentrate on the presentation and analysis of these techniques. Specifically, Fourier analyses of the DG(1)–Hancock and several semi-dicrete methods for both 1-D and 2-D linear hyperbolic-relaxation equations are conducted. Analyses show that the existence of grid size restriction in order to ensure the intended order of accuracy; DG(1)–Hancock is the least restrictive method among others. It is also shown that a two-dimensional extension of DG(1) method introduces an extra multidimensional dissipation error. Owing to its third-order accuracy with less function evaluations as a result of a space-time discretization, the DG(1)–Hancock method is not only accurate but efficient in comparison to other semi- and fully discrete finite-volume and semi-discrete discontinuous-Galerkin methods.

The second part of the paper (Section IV) deals with the accurate formulation of solid-wall boundary conditions for systems of moment equations describing rarefied gas dynamics. The boundary treatment uses detailed information about the molecular velocity distribution of the gas near the wall, as well as parameters of the boundary, i. e., temperature T^{w} and velocity $\mathbf{u}^{\mathrm{w}}$. It also takes into account the roughness of the solid boundary. The approach differs from that of Grad, which contains an inconsistency. We analyze this inconsistency and propose an alternative approach. To validate it we compute simple *linearized* Couette flow in two dimensions in combination with the *Gaussian* 10-moment physical description. This test is geometrically simple, yet physically adequate, and the existence of an analytical approximate solution makes this flow a good candidate for first try-out of the new boundary treatment.

II. Analysis for 1-D Linear Hyperbolic-Relaxation Equations

In this section, numerical methods including the DG(1)–Hancock method for hyperbolic-relaxation equations are investigated analytically. We shall now carry out a Fourier analysis of three methods applied to one-dimensional systems of linear hyperbolic-relaxation equations. The local truncation error of DG(1)–Hancock is compared to HR2–MOL and DG(1)–MOL, where HR2 stands for a high-resolution (second-order) finite-volume method. Fourier analyses show the superior accuracy of the DG(1)–Hancock method compared to that of the semi-discrete, method-of-lines approach. The analyses conducted here are strongly motivated by the work of Lowrie and Morel,[5] and Hittinger.[6]

II.A. Model Equations: Generalized Hyperbolic Heat Equations

II.A.1. Dimensional Form

The model equation we consider is the generalized hyperbolic heat equations (GHHE),[5–7]

$$\begin{aligned} \partial_t u + \partial_x v &= 0, \\ \partial_t v + a_F^2 \partial_x u &= -\frac{1}{\tau}(v - a_E u); \quad x \in \mathbb{R},\ t > 0, \end{aligned} \tag{1}$$

where $u(x,t) \in \mathbb{R}$ is the conserved variable and $v(x,t) \in \mathbb{R}$ is the flux of u. In vector form, $\mathbf{u} = [u, v]^T, \mathbf{f} = [v, a_F^2 u]^T$, and $\mathbf{s} = [0, a_E u - v]^T$ in

$$\partial_t \mathbf{u} + \partial_x \mathbf{f}(\mathbf{u}) = \frac{1}{\tau}\mathbf{s}(\mathbf{u}). \tag{2}$$

There are three *constant* parameters: $\tau > 0$ is a relaxation time, $a_F > 0$ is a frozen wave speed, and $a_E > 0$ is an equilibrium wave speed. For stability, $|a_E| \leq a_F$. More detailed mathematical descriptions can be found in the review paper by Natalini[8] and references therein.

The constant Jacobian matrix and its eigenvalues are as follows:

$$\mathbf{A} := \frac{\partial \mathbf{f}}{\partial \mathbf{u}} = \begin{pmatrix} 0 & 1 \\ a_F^2 & 0 \end{pmatrix} \quad \longrightarrow \quad \lambda_{1,2} := \mathrm{Eig}(\mathbf{A}) = \pm a_F. \tag{3}$$

Here, we insist that these three parameters have physical meaning; once the problem is described, these parameters are fixed. The above equations are constructed such that the frozen waves propagate at speed $\pm a_F$ in the beginning; these eventually decay. Simultaneously, equilibrium waves at speed $\pm a_E$ enter the model; one of the equilibrium waves with speed $-a_E$ is quickly damped out, and the other wave with speed a_E dominates the solution. Figure 1 describes these waves schematically. The right hand side of (1) represents the relaxation process, which always drives the non-equilibrium flux variable v to its equilibrium flux $a_E u$. A detailed dispersion analysis and the exact solution of the Riemann problem are presented by Hittinger.[6,9]

Let L be a length scale of interest, and a_F serve as a reference wave speed, then a reference time scale can be defined by $T := \dfrac{L}{a_F}$. Note that this is a particular choice of scaling: another reference time may be chosen. Since a_F is a fixed value, changing the length scale of interest affects the reference time. The GHHE

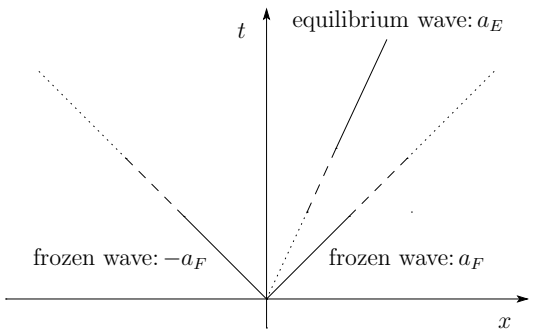

Figure 1. Initially, two frozen waves propagate with speed $\pm a_F$; they eventually decay. Meanwhile, the equilibrium wave with speed a_E arises and dominates the flow field in the long-time limit.

can be reduced to a smaller set of equations by a certain choice of T relative to the relaxation time τ, which really means choosing a certain length scale of interest.

When the time of interest is much smaller than the relaxation time ($T \ll \tau$), the relaxation process is not yet important, and the GHHE is reduced to the wave equation,

$$\begin{aligned} \partial_t u + \partial_x v &= 0, \\ \partial_t v + a_F^2 \partial_x u &\simeq 0, \end{aligned} \quad \longrightarrow \quad \partial_{tt} u - a_F^2 \partial_{xx} u = 0, \tag{4}$$

where the wave speeds are $\pm a_F$. This is the reduced form of the *frozen limit*.

On the other hand, when the time of interest is much larger than the relaxation time ($T \gg \tau$), the relaxation process is no longer negligible. Asymptotic expansion of u and v for small τ gives an advection-diffusion equation (the derivation for the particular scaling is given in [2, Appendix C]):

$$\partial_t u + a_E \partial_x u = \tau(a_F^2 - a_E^2)\partial_{xx} u + O\big(\tau^2\big)\,. \tag{5}$$

This is the reduced form in the *near-equilibrium limit*. Note that the leading diffusion coefficient $\tau(a_F^2 - a_E^2)$ always has a positive sign as long as $|a_E| \le a_F$; this property is called the sub-characteristic condition for stability.[10] There are two different physical processes included in this equation; the relative strength of the two parameters, advection speed a_E and diffusion coefficient $\epsilon(a_F^2 - a_E^2)$, decides which is the dominant physics. This will be discussed in more detail in a later section.

We further consider the time scale of interest T to be infinite; this is equivalent to letting $\tau \to 0$, so the relaxation process occurs instantaneously, and the above equation becomes a pure advection equation:

$$\partial_t u + a_E \partial_x u = 0, \tag{6}$$

where the wave speed is a_E. This is the reduced form of the GHHE in the *equilibrium limit*.

To summarize, let $\bar{t}$ be the dimensionless time normalized by the relaxation time τ such that

$$\bar{t} := \frac{T}{\tau} = \frac{L}{a_F \tau}. \tag{7}$$

The reduced equations of the GHHE corresponding to $\bar{t}$ are shown in Table 1. These forms can be seen as consecutive transformations of the GHHE in the time frame.

dimensionless time	assumption	reduced equation
$\bar{t} \ll 1$	frozen limit	$\partial_{tt} u - a_F^2 \partial_{xx} u = 0$
$\bar{t} \gg 1$	near-equilibrium limit	$\partial_t u + a_E \partial_x u \simeq \tau(a_F^2 - a_E^2)\partial_{xx} u$
$\bar{t} \to \infty$	equilibrium limit	$\partial_t u + a_E \partial_x u = 0$

Table 1. The reduced forms of the GHHE are listed in each limit. The characteristic of the GHHE changes with the time scale of interest.

II.A.2. Nondimensionalization of the 1-D GHHE

Choice of Scaling Parameters As seen in the previous section, the GHHE changes characteristics in different time scales of interest even though the equations themselves are linear. Thus, when we nondimensionalize the original equations (1), the specific choice of reference time t_0 affects the behavior of the equations significantly. Here, three different reference times are chosen for nondimensionalization. Let each symbol with subscript 0 serve as a reference parameter to nondimensionalize the variables, and the notation $\hat{(\cdot)}$ represent a dimensionless variable, then

$$\hat{t} := \frac{t}{t_0}, \quad \hat{x} := \frac{x}{x_0}, \quad \hat{u} := \frac{u}{u_0}, \quad \text{and} \quad \hat{v} := \frac{v}{v_0}. \tag{8}$$

Inserting these relations into (1) leads to

$$\partial_{\hat{t}} \hat{u} + \left(\frac{v_0/u_0}{x_0/t_0} \right) \partial_{\hat{x}} \hat{v} = 0, \tag{9a}$$

$$\partial_{\hat{t}} \hat{v} + a_F^2 \left(\frac{u_0/v_0}{x_0/t_0} \right) \partial_{\hat{x}} \hat{u} = -\frac{1}{\tau/t_0} \left(\hat{v} - a_E \frac{u_0}{v_0} \hat{u} \right). \tag{9b}$$

Assuming a unity wave speed in (9a), hence

$$\frac{v_0/u_0}{x_0/t_0} = 1 \quad \longrightarrow \quad \frac{v_0}{u_0} = \frac{x_0}{t_0}, \tag{10}$$

does not change the problem, and the above equations become

$$\begin{aligned} \partial_{\hat{t}} \hat{u} + \partial_{\hat{x}} \hat{v} &= 0, \\ \partial_{\hat{t}} \hat{v} + \left(\frac{a_F}{x_0/t_0} \right)^2 \partial_{\hat{x}} \hat{u} &= -\frac{1}{\tau/t_0} \left(\hat{v} - \frac{a_E}{x_0/t_0} \hat{u} \right). \end{aligned} \tag{11}$$

Now, the proper reference time t_0 and reference speed $\frac{x_0}{t_0}$ have to be chosen for the nondimensionalization. Available constant parameters are a_F [LT^{-1}], a_E [LT^{-1}], and τ [T]. Also, let L [L] be a length scale of interest, which may vary within a problem. As to a reference time, the obvious choice is $t_0 = \tau$; in this scaling, time is measured at a scale of the same order of the relaxation process. A next possible scaling is $t_0 = \frac{L}{a_F}$ where time is scaled by the traveling time of frozen waves. The equilibrium speed can be used as scaling when $a_E \neq 0$, thus $t_0 = \frac{L}{a_E}$. Note that $\frac{L}{a_F} \leq \frac{L}{a_E}$. Another nonintuitive choice is $t_0 = \frac{L^2}{\tau a_F^2}$.

As a reference speed $\frac{x_0}{t_0}$, both frozen speed a_F and equilibrium speed a_E are the obvious choices; the characteristic speed of relaxation $\frac{L}{\tau}$ might be a possible choice as well. The specific forms of each scaling are discussed in the next subsection under the assumption $u_0 = O(1)$.

II.A.3. Nondimensional Form

SYMMETRIC FROZEN-WAVE-SPEEDS MODEL Among the various nondimensionalization, we adopt the frozen-wave time scale ($t_0 = L/a_F$), for the following analysis. For simplicity, the notation $\hat{(\cdot)}$ in (11) is henceforth omitted, and our target model equations are written as

$$\begin{aligned} \partial_t u + \partial_x v &= 0, \\ \partial_t v + \partial_x u &= -\frac{1}{\epsilon}(v - ru). \end{aligned} \tag{12}$$

Here, u is the conserved variable, v is the flux of u, and $\epsilon > 0$ is a dimensionless relaxation time. In vector form, $\mathbf{u} = [u, v]^T, \mathbf{f} = [v, u]^T$, and $\mathbf{s} = [0, ru - v]^T$ in

$$\partial_t \mathbf{u} + \partial_x \mathbf{f}(\mathbf{u}) = \frac{1}{\epsilon}\mathbf{s}(\mathbf{u}), \tag{13}$$

where

$$\epsilon := \frac{\tau a_F}{L} > 0, \qquad r := \frac{a_E}{a_F}, \quad |r| \leq 1, \tag{14}$$

is the dimensionless relaxation time, and dimensionless equilibrium speed respectively.

This system has 'frozen' wave speeds ± 1 when relaxation is weak ($\epsilon \gg 1$); when relaxation dominates ($\epsilon \ll 1$), it reduces to the advection-diffusion equation,

$$\partial_t u + r\partial_x u = \epsilon(1 - r^2)\partial_{xx} u + O\left(\epsilon^2\right), \tag{15}$$

with an 'equilibrium' wave speed r. For stability, $|r| \leq 1$. Note that we have written this equation in a form that leads to an advection-dominated advection-diffusion in asymptotic limit. This is consistent with a focus on compressible, viscous flow. Other choices of scalings, such as diffusive scalings,[5,11] can lead to more strongly parabolic limits. Indeed, for $r = O(\epsilon)$, the scaling of Lowrie and Morel[5] is in effect a long-time, small-advective-flux limit.

ASYMMETRIC FROZEN-WAVE-SPEEDS MODEL Hittinger et al. show that the system (12) can be generalized to break symmetry in the frozen limit:[12]

$$\begin{aligned} \partial_t u + \partial_x v &= 0, \\ \partial_t v + c\partial_x u + (1 - c)\partial_x v &= -\frac{1}{\epsilon}(v - ru). \end{aligned} \tag{16}$$

The frozen wave speeds are thus $-c$ and 1, and the near-equilibrium form is

$$\partial_t u + r\partial_x u = \epsilon(1 - r)(c + r)\partial_{xx} u + O\left(\epsilon^2\right). \tag{17}$$

Note the modification to the diffusion rate. For stability, $-c \leq r \leq 1$. This model is used only for limited cases due to the complexity of analysis.

EXACT SOLUTION In the reduced equation of the GHHE (15), the exact eigenvalue of the spatial differentiation operator for the harmonic mode $\mathbf{u}(x, t) = \hat{\mathbf{u}}_0 e^{(ikx + \lambda t)}$ is given by

$$\lambda_{\text{exact}}^{\text{GHHE}} = -irk - \epsilon(1 - r^2)k^2 - 2i\epsilon^2 r(1 - r^2)k^3 + O\left(\epsilon^3\right). \tag{18}$$

Note that the above exact solution is indeed an infinite series. Conversely, the exact solution of the spatial differential operator of the genuine advection-diffusion equation runs only up to $O(\epsilon^2)$, thus

$$\lambda_{\text{exact}}^{\text{adv-diff}} = -irk - \epsilon(1-r^2)k^2. \tag{19}$$

II.B. Difference Operators and Their Properties

Various discretization methods are applied to the linear hyperbolic-relaxation equations (12), and a Fourier analysis is conducted to uncover those properties. By this we can show, to a given order in ϵ, if a scheme captures the advection-dominated advection-diffusion limit (15) with second-order accuracy in Δx. Similar analyses have been done using modified differential equations,[5,7] though not always using the same scaling and limit. Furthermore, the analysis here also considers temporal discretization to reveal an issue of both spatial and temporal stiffness inherent in the system.

II.B.1. Operator-Splitting Method

At first, to demonstrate an extra difficulty arising due to the stiff source term, operator splitting is adopted in the time integrator.[13,14] This splitting decouples the time evolution of the flux and source terms, allowing us to compute these independently. The great advantage of this method, particularly for hyperbolic-relaxation equations, is that the source term, which yields exponential damping, can be integrated exactly. In order to isolate the error introduced by the operator splitting, we eliminate the spatial discretization error by taking the flux derivative from the exact solution. Thus, the operator-split update operator for (12) takes the form

$$\begin{aligned} \mathbf{u}^{(1)} &= e^{\frac{\Delta t}{2}\frac{1}{\epsilon}\mathbf{Q}}\mathbf{u}^n, \\ \mathbf{u}^{(2)} &= e^{-ik\Delta t\mathbf{A}}\mathbf{u}^{(1)}, \\ \mathbf{u}^{n+1} &= e^{\frac{\Delta t}{2}\frac{1}{\epsilon}\mathbf{Q}}\mathbf{u}^{(2)}, \end{aligned} \tag{20}$$

where

$$\mathbf{A} := \frac{\partial \mathbf{f}}{\partial \mathbf{u}} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{Q} = \begin{pmatrix} 0 & 0 \\ r & -1 \end{pmatrix}. \tag{21}$$

After some algebra, the local truncation error in the low-frequency limit is found to be

$$\begin{aligned} \text{LTE}_{\text{splitting}} &= \left[\frac{(1+e^{\Delta t/\epsilon})(1-r^2)\Delta t}{2(1-e^{\Delta t/\epsilon})} + \epsilon(1-r^2)\right]k^2 + O(k^3) \\ &\simeq -\frac{1-r^2}{12}\frac{\boxed{\Delta t^2}}{\epsilon}k^2 = -\frac{(1-r^2)\nu^2}{12}\frac{\boxed{\Delta x^2}}{\epsilon}k^2, \end{aligned} \tag{22}$$

where the Courant number, ν, based on the frozen wave speeds is defined by

$$\nu := 1\frac{\Delta t}{\Delta x}. \tag{23}$$

The above equation shows that the splitting is second-order in space and time. However, since the above error is in the k^2-term, an extra numerical dissipation is added to the physical dissipation $-\epsilon(1-r^2)k^2$

in (18); this leads to an incorrect diffusion coefficient. To ensure the physical dissipation is dominant, the following inequality has to be satisfied:

$$\frac{(1-r^2)\nu^2}{12}\frac{\Delta x^2}{\epsilon} \ll \epsilon(1-r^2) \quad \longrightarrow \quad \nu\Delta x \ll \epsilon. \tag{24}$$

Particularly, when the near-equilibrium limit ($\epsilon \ll 1$) is considered, the time step and grid size are severely restricted such that

$$\Delta t = \nu\Delta x \propto \epsilon, \tag{25}$$

otherwise the excessive numerical dissipation damps all waves in the domain.

The above example shows that straightforward decoupling of the flux and source term leads to an accurate method in the near-equilibrium limit only when (25) is satisfied. In order to overcome this severe restriction, coupling between flux and source term is necessary; for instance, an MOL with several stages, or a fully discrete method in which the flux has strong coupling with the source term.

II.B.2. HR–MOL Method

A semi-discrete high-resolution Godunov method (HR–MOL), particularly, the second-order method, is considered. The HR2–MOL method applied to the asymmetric GHHE equation (16) has the following generic form:

$$\frac{\partial\bar{\mathbf{u}}_j(t)}{\partial t} = -\frac{1}{\Delta x}\left(\hat{\mathbf{f}}_{j+1/2} - \hat{\mathbf{f}}_{j-1/2}\right) + \frac{1}{\epsilon}\mathbf{s}\left(\bar{\mathbf{u}}_j\right), \tag{26}$$

where $\hat{\mathbf{f}}_{j\pm1/2}$ denotes the approximate flux at interfaces $j \pm 1/2$. We will take this to be the upwind flux

$$\hat{\mathbf{f}}_{j+1/2}\left(\mathbf{u}_L, \mathbf{u}_R\right) = \mathbf{A}^+\mathbf{u}_L + \mathbf{A}^-\mathbf{u}_R, \tag{27}$$

where, if $\mathbf{\Lambda}$ is the diagonal matrix of eigenvalues of $\mathbf{A}$, $\mathbf{A}^\pm = \mathbf{R}\mathbf{\Lambda}^\pm\mathbf{L}$. In the case of an asymmetric system (16),

$$\mathbf{A}^+ = \frac{1}{1+c}\begin{pmatrix} c & 1 \\ c & 1 \end{pmatrix}, \quad \mathbf{A}^- = \frac{c}{1+c}\begin{pmatrix} -1 & 1 \\ c & -c \end{pmatrix}. \tag{28}$$

After inserting the difference forms into the original ODE (26) and some algebra, the semi-discrete method can be written in the compact form:

$$\frac{\partial\bar{\mathbf{u}}_j(t)}{\partial t} = \left(\mathbf{N}_{\mathrm{HR2}} + \frac{1}{\epsilon}\mathbf{Q}\right)\bar{\mathbf{u}}_j(t), \tag{29}$$

where $\bar{\mathbf{u}}_j = [\bar{u}_j, \bar{v}_j]^T$; the difference operator of the flux derivative $\mathbf{N}_{\mathrm{HR2}}$ is given by

$$\mathbf{N}_{\mathrm{HR2}} = -\frac{1}{4(1+c)\Delta x}\left(\boldsymbol{\mathcal{A}}^{2+}\mathbf{D}^{2+} + \boldsymbol{\mathcal{A}}^{+}\mathbf{D}^{+} + \boldsymbol{\mathcal{A}}^{-}\mathbf{D}^{-} + \boldsymbol{\mathcal{A}}^{2-}\mathbf{D}^{2-}\right), \tag{30}$$

where

$$\boldsymbol{\mathcal{A}}^+ = \begin{pmatrix} -2c & 1+3c \\ c(1+3c) & 1-3c^2 \end{pmatrix}, \quad \boldsymbol{\mathcal{A}}^- = \begin{pmatrix} 2c & 3+c \\ c(3+c) & 3-c^2 \end{pmatrix}$$

$$\boldsymbol{\mathcal{A}}^{2+} = \begin{pmatrix} c & -c \\ -c^2 & c^2 \end{pmatrix}, \quad \boldsymbol{\mathcal{A}}^{2-} = \begin{pmatrix} c & 1 \\ c & 1 \end{pmatrix}, \quad \mathbf{D}^\pm = \delta^\pm\mathbf{I}, \quad \mathbf{D}^{2\pm} = (\delta^\pm)^2\mathbf{I}.$$

In order to obtain the eigenvalues of the HR2 spatial discretization, the quadratic characteristic equation,

$$\det\left(\mathbf{N}_{\mathrm{HR2}} + \frac{1}{\epsilon}\mathbf{Q} - \lambda\mathbf{I}\right) = 0, \tag{31}$$

is solved. For simplicity, we only present the case when $c = 1$; the two roots have following form:

$$\lambda_{\mathrm{HR2}}^{(1),(2)} = \frac{(1-\cos\beta)^2}{2\Delta x} + \frac{1}{2\epsilon}\left[1 \mp \sqrt{1 - 2ir\frac{\epsilon}{\Delta x}(3-\cos\beta)\sin\beta + \left(\frac{\epsilon}{\Delta x}(3-\cos\beta)\sin\beta\right)^2}\right]. \tag{32}$$

Spatial Accuracy We expand the trigonometric factors in (32) for the long wave-length limit $\beta \ll 1$, and then expand the square root. The results are

$$\begin{aligned}\lambda_{\mathrm{HR2}}^{(1)} = &-\frac{ir}{\Delta x}\beta - \frac{\epsilon(1-r^2)}{\Delta x^2}\beta^2 - \left[\frac{ir}{12\Delta x} + \frac{2i\epsilon^2 r(1-r^2)}{\Delta x^3}\right]\beta^3 \\ &- \left[\frac{1}{8\Delta x} + \frac{\epsilon(1-r^2)}{6\Delta x^2} + \frac{\epsilon^2(1-r^2)(1-5r^2)}{\Delta x^4}\right]\beta^4 + O\big(\beta^5\big)\,,\end{aligned} \tag{33a}$$

$$\lambda_{\mathrm{HR2}}^{(2)} = -\frac{1}{\epsilon} + \frac{ir}{\Delta x}\beta + O\big(\beta^2\big)\,. \tag{33b}$$

The second root (33b) exhibits rapid exponential decay for $\epsilon \ll 1$, while the first root (33a) does not, thus the latter $\lambda_{\mathrm{HR2}}^{(1)}$ is the dominant behavior in this asymptotic limit. The spatial discretization error is obtained by replacing β by the wave number $k := \dfrac{\beta}{\Delta x}$, then

$$\begin{aligned}\lambda_{\mathrm{HR2}}^{(1)} - \lambda_{\mathrm{exact}}^{\mathrm{GHHE}} = &\underbrace{-\frac{ir}{12}\boxed{\Delta x^2}}_{\text{dispersion error}}\ k^3 \underbrace{-\left[\frac{1}{8}\boxed{\Delta x^3} + \frac{\epsilon(1-r^2)}{6}\Delta x^2\right]}_{\text{dissipation error}} k^4 + O\big(k^5\big)\,, \\ \lambda_{\mathrm{HR2}}^{(2)} - \lambda_{\mathrm{exact}}^{\mathrm{GHHE}} = &-\frac{1}{\epsilon} + 2irk + O\big(k^2\big)\,,\end{aligned} \tag{34}$$

where the exact spatial differential operator $\lambda_{\mathrm{exact}}^{\mathrm{GHHE}}$ is given in (18). In the first equation, both dispersive (k^3-term) and dissipative (k^4-term) errors are present; the dispersive error is second-order in Δx, and since the correct limit shows no dispersion, these numerical dispersion errors can not be confused with any physical dispersion. However, the leading dissipative error term, $-\dfrac{1}{8}\Delta x^3 k^4$, does not scale with ϵ, and so can compete with the physical dissipation $-\epsilon\,(1-r^2)k^2$ in (18) if the relaxation scale is unresolved ($\Delta x \gg \epsilon$). For the physical dissipation to dominate,

$$\text{(dominant numerical dissipation)} \ll \text{(physical dissipation)},$$

thus,

$$\frac{1}{8}\Delta x^3 k^4 \ll \epsilon(1-r^2)k^2. \tag{35}$$

Solving for Δx leads to the threshold grid size $\Delta h_{\mathrm{HR2}}^*$:

$$\Delta x \ll 2\left[\frac{\epsilon(1-r^2)}{k^2}\right]^{1/3} = 2\left(\frac{r}{k^2\,Pe}\right)^{1/3}, \tag{36}$$

where the Peclet number, Pe, is defined by $Pe := \dfrac{r}{\epsilon(1-r^2)}$, then,

$$\Delta h_{\mathrm{HR2}}^* := 2\left(\frac{r}{k^2\,Pe}\right)^{1/3}. \tag{37}$$

Hence, the HR2 scheme does not attain the asymptotic limit to second-order in Δx with Δx independent of ϵ.

For the HR2 scheme, the result (36) is well known. It appears for the $r = 0$ case in previous studies,[5,7] although not necessarily in our scaling. The form (34) for $r \neq 0$ for the HR2 scheme can be obtained from Eq. (3.17) in Jin and Levermore [7, p. 461] or Eq. (31) in Lowrie and Morel [5, p. 420]. In the scaling of the latter work, the term that leads to (36) is actually divergent since it goes like $\dfrac{\Delta x^3}{\epsilon}$ (due to the time dilation of this scaling), but it still leads to the constraint (36). Jin and Levermore claimed that the grid size restriction can be removed by averaging the frozen and equilibrium fluxes.[14] However, we find that the grid restriction still exists in their method. The detailed analysis is described in [2, Appendix D].

It is interesting to compare (36) with a direct discretization of the asymptotic equation (15) using the Rusanov flux function and slope reconstruction. The diffusion term is discretized using a three-point, second-order central-difference approximation. From a Fourier analysis, the eigenvalue of the scheme is

$$\lambda_{\text{HR2}}^{\text{adv-diff}} - \lambda_{\text{exact}}^{\text{adv-diff}} = -\frac{ir}{12}\boxed{\Delta x^2}\,k^3 - \left[\frac{1}{8}\boxed{\Delta x^3} - \frac{\epsilon(1-r^2)}{12}\Delta x^2\right]k^4 + O\big(k^5\big)\,, \tag{38}$$

for $\Delta x \ll 1$. We see that this has the same fourth-order numerical dissipation term as the HR2 discretization of GHHE (34). This discretization will have the same restriction (36) on Δx to ensure that the physical dissipation is dominant. When this restriction is satisfied, the above equation shows that the HR2 method is second-order in space owing to Δx^2 in the k^3-term.

Spatial-Temporal Accuracy In our previous analysis, we only consider the spatial discretization of the HR2 method under the assumption that the flux and source term are discretized at the same time level.[12] However, a typically ODE solver for a stiff problem discretizes the flux and source terms at different time levels due to the implicit treatment of the source term. Furthermore, the system (12) possesses both spatial and temporal stiffness, thus, analyzing a fully discrete form of any method is necessary. A great number of stiff ODE solvers have been proposed in the last few decades.[11,14–19] Among these methods, we adopt the implicit-explicit (IMEX) Runge–Kutta methods originally developed by Ascher et al.[20,21] for hyperbolic-parabolic equations, and later extended to hyperbolic-relaxation equations by Pareschi and Russo.[22] The methods treat the flux term explicitly by a strong-stability-preserving (SSP) method, and the source term by an L-stable diagonally implicit Runge–Kutta method (DIRK). The authors developed a family of second and third-order methods. Here, as an example, we adopt the IMEX–SSP2(3,3,2) method; both explicit and implicit methods require three stages, and overall accuracy is second-order. The actual update formulas are the following:

$$\begin{aligned}
\mathbf{u}^{(1)} &= \mathbf{u}^n + \frac{\Delta t}{4\epsilon}\mathbf{s}(\mathbf{u}^{(1)}),\\
\mathbf{u}^{(2)} &= \mathbf{u}^n - \frac{\Delta t}{2}\partial_x\mathbf{f}(\mathbf{u}^{(1)}) + \frac{\Delta t}{4\epsilon}\mathbf{s}(\mathbf{u}^{(2)}),\\
\mathbf{u}^{(3)} &= \mathbf{u}^n - \frac{\Delta t}{2}\left[\partial_x\mathbf{f}(\mathbf{u}^{(1)}) + \partial_x\mathbf{f}(\mathbf{u}^{(2)})\right] + \frac{\Delta t}{3\epsilon}\left[\mathbf{s}(\mathbf{u}^{(1)}) + \mathbf{s}(\mathbf{u}^{(2)}) + \mathbf{s}(\mathbf{u}^{(3)})\right],\\
\mathbf{u}^{n+1} &= \mathbf{u}^n - \frac{\Delta t}{3}\left[\partial_x\mathbf{f}(\mathbf{u}^{(1)}) + \partial_x\mathbf{f}(\mathbf{u}^{(2)}) + \partial_x\mathbf{f}(\mathbf{u}^{(3)})\right] + \frac{\Delta t}{3\epsilon}\left[\mathbf{s}(\mathbf{u}^{(1)}) + \mathbf{s}(\mathbf{u}^{(2)}) + \mathbf{s}(\mathbf{u}^{(3)})\right].
\end{aligned} \tag{39}$$

The intermediate formulas for the derivation of the local truncation error are omitted here; only the final result is presented. The local truncation error of the HR2–IMEX method becomes

$$\mathrm{LTE}_{\mathrm{HR2IMEX}} = \left[\underbrace{-\frac{ir}{12}\left(1+(r\nu)^2\right)\boxed{\Delta x^2}}_{\text{dominant dispersion error}} - \frac{i\epsilon\, r(1-r^2)\nu}{6}\Delta x \right] k^3 - \left[\underbrace{\frac{1}{8}\left(1-\frac{r}{3}(r\nu)^3\right)\boxed{\Delta x^3}}_{\text{dominant dissipation error}} + \frac{\epsilon\,(1-r^2)}{6}\Delta x^2 \right] k^4. \quad (40)$$

Under the assumption of near-equilibrium we have $r = O(1)$ and $\epsilon \ll 1$; when the physical dissipation dominates over the dissipation error,

$$\frac{1}{8}\left(1-\frac{r}{3}(r\nu)^3\right)\Delta x^3 k^4 \ll \epsilon\,(1-r^2)k^2, \quad (41)$$

then the method is second-order in both space and time owing to the dominant dispersion error in the k^3-term. Note that the remaining terms in (40) are guaranteed to be always smaller than the physical dispersion and dissipation owing to ϵ in their coefficients. Finally, the threshold grid size for the fully discrete method is given by

$$\Delta h^*_{\mathrm{HR2IMEX}} := 2\left[\frac{\epsilon(1-r^2)}{\left(1-\frac{r}{3}(r\nu)^3\right)k^2}\right]^{1/3}. \quad (42)$$

II.B.3. DG–MOL Method

The DG(1)–MOL scheme for (2) has the form

$$\frac{\partial \bar{\mathbf{u}}_j(t)}{\partial t} = -\frac{1}{\Delta x}\left(\hat{\mathbf{f}}_{j+1/2} - \hat{\mathbf{f}}_{j-1/2}\right) + \frac{1}{\epsilon}\mathbf{s}\left(\bar{\mathbf{u}}_j\right), \quad (43a)$$

$$\frac{\partial \overline{\Delta \mathbf{u}}_j(t)}{\partial t} = -\frac{6}{\Delta x}\left(\hat{\mathbf{f}}_{j+1/2} + \hat{\mathbf{f}}_{j-1/2} - 2\mathbf{f}(\bar{\mathbf{u}}_j)\right) + \frac{1}{\epsilon}\mathbf{s}\left(\overline{\Delta \mathbf{u}}_j\right), \quad (43b)$$

where the upwind flux function becomes

$$\hat{\mathbf{f}}_{j+1/2}\left(\mathbf{u}_L, \mathbf{u}_R\right) := \hat{\mathbf{f}}_{j+1/2}\left(\bar{\mathbf{u}}_j + \frac{1}{2}\overline{\Delta \mathbf{u}}_j, \bar{\mathbf{u}}_{j+1} - \frac{1}{2}\overline{\Delta \mathbf{u}}_{j+1}\right). \quad (44)$$

The first update equation (43a) with (44) is precisely the HR2 method (modulo limiting) where $\dfrac{\overline{\Delta \mathbf{u}}_j}{\Delta x}$ is the slope in cell j. For the HR2 method, the differences $\overline{\Delta \mathbf{u}}_j$ are approximated at each step by differencing neighboring cell-averaged values $\bar{\mathbf{u}}_{j\pm1}$, whereas in the DG(1) method, the slopes evolve as additional variables.

It is these slopes, whether computed or self-evolving, that are responsible for providing second-order spatial accuracy in the flux evaluation. It is also these slopes that provide the distinction between the two schemes.

For length scales much larger than the relaxation length scale $a\tau$, the flux discretization must approximate the coupling between the two hyperbolic and relaxation operators. For an HR method, the flux function is

based solely on the original hyperbolic operator and each slope purely on the initial data. In contrast, the DG method *simultaneously* updates the solution average and slope under the influence of the source.

It is interesting to conduct a Fourier analysis of the one-dimensional DG(1) method (43) for the asymmetric system (16) as $\epsilon \to 0$. Following the previous analysis, take $\mathbf{u}_j = [\bar{\mathbf{u}}_j, \overline{\Delta \mathbf{u}}_j]^T$, then the difference form of the DG(1) method can be written as

$$\frac{\partial \mathbf{u}_j(t)}{\partial t} = \left(\mathbf{N}_{\mathrm{DG}(1)} + \frac{1}{\epsilon} \mathbf{Q}_{\mathrm{DG}(1)} \right) \mathbf{u}_j(t). \tag{45}$$

Here, the difference operator of the flux discretization is given by

$$\mathbf{N}_{\mathrm{DG}(1)} = \boldsymbol{\mathcal{A}}^+ \mathbf{D}^+ + \boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{A}}^- \mathbf{D}^-, \tag{46}$$

where

$$\boldsymbol{\mathcal{A}}^+ = \frac{1}{(1+c)\Delta x} \begin{pmatrix} c & -c & -\frac{c}{2} & \frac{c}{2} \\ -c^2 & c^2 & \frac{c^2}{2} & -\frac{c^2}{2} \\ 6c & -6c & -3c & 3c \\ -6c^2 & 6c^2 & 3c^2 & -3c^2 \end{pmatrix}, \quad \boldsymbol{\mathcal{A}}^- = \frac{1}{(1+c)\Delta x} \begin{pmatrix} -c & -1 & -\frac{c}{2} & -\frac{1}{2} \\ -c & -1 & -\frac{c}{2} & -\frac{1}{2} \\ 6c & 6 & 3c & 3 \\ 6c & 6 & 3c & 3 \end{pmatrix}, \tag{47a}$$

$$\boldsymbol{\mathcal{C}} = \frac{1}{(1+c)\Delta x} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -12c & -6(1-c) \\ 0 & 0 & -6c(1-c) & -6(1+c^2) \end{pmatrix}, \quad \mathbf{Q}_{\mathrm{DG}(1)} = \begin{pmatrix} \mathbf{Q} & 0 \\ 0 & \mathbf{Q} \end{pmatrix}. \tag{47b}$$

In order to obtain eigenvalues of the DG(1) spatial discretization, we compute the characteristic equation,

$$\det \left(\mathbf{N}_{\mathrm{DG}(1)} + \frac{1}{\epsilon} \mathbf{Q}_{\mathrm{DG}(1)} - \lambda \mathbf{I} \right) = 0, \tag{48}$$

and take the equilibrium $\epsilon = 0$ to obtain a quadratic equation in λ. (This will yield the leading-order term in λ.) This is, for a Fourier mode,

$$(\Delta x \lambda)^2 - 2 \left[\left(\frac{2c + r(1-c)}{1+c} \right) (2 + \cos\beta) - ir \sin\beta \right] \Delta x \lambda + 6r \left[r(1-\cos\beta) + \left(\frac{2c + r(1-c)}{1+c} \right) i \sin\beta \right] = 0. \tag{49}$$

Spatial Accuracy The DG(1) scheme in compact form is given in (45); we restrict ourselves here to the special case of $c = 1$. Since this is a 4×4 system, the characteristic polynomial is of degree four. The leading-order behavior is given by the quadratic equation (49). Using these to find the terms of $O(\epsilon)$, and, at each order, expanding out the trigonometric terms for $\beta \ll 1$, we find

$$\begin{aligned} \lambda^{(1)}_{\mathrm{DG}(1)} &= & -\frac{ir}{\Delta x}\beta & & -\frac{\epsilon(1-r^2)}{\Delta x^2}\beta^2 + O\left(\beta^3\right), \\ \lambda^{(2)}_{\mathrm{DG}(1)} &= -\frac{1}{\epsilon} & +\frac{ir}{\Delta x}\beta & & +\frac{\epsilon(1-r^2)}{\Delta x^2}\beta^2 + O\left(\beta^3\right), \\ \lambda^{(3)}_{\mathrm{DG}(1)} &= -\frac{6}{\Delta x} & +\frac{3ir}{\Delta x}\beta & +\frac{\Delta x - 9\epsilon(1-r^2)}{\Delta x^2}\beta^2 & + O\left(\beta^3\right), \end{aligned}$$

$$\lambda^{(4)}_{\mathrm{DG}(1)} = -\frac{1}{\epsilon} - \frac{6}{\Delta x} - \frac{3ir}{\Delta x}\beta + \frac{\Delta x + 9\epsilon(1-r^2)}{\Delta x^2}\beta^2 + O\big(\beta^3\big)\,.$$

The last three roots all exhibit rapid exponential decay for $\Delta x \ll 1$ and $\epsilon \ll 1$, while the first root does not. Since $\lambda^{(1)}_{\mathrm{DG}(1)}$ is the dominant root in the asymptotic limit, we continue expanding it; then the spatial discretization error becomes

$$\lambda^{(1)}_{\mathrm{DG}(1)} - \lambda^{\mathrm{GHHE}}_{\mathrm{exact}} = -\frac{1}{72}\left[\Delta x^3 - \frac{1-r^2}{\Delta x + 6\epsilon}\Delta x^4\right] k^4 + O\big(k^5\big)\,, \tag{50}$$

where we have made the substitution $k = \dfrac{\beta}{\Delta x}$. Since we are considering the near-equilibrium limit $\epsilon \ll 1$, expanding the above error with respect to ϵ provides

$$\lambda^{(1)}_{\mathrm{DG}(1)} - \lambda^{\mathrm{GHHE}}_{\mathrm{exact}} = -\frac{1}{72}\left[r^2\,\boxed{\Delta x^3} + 6\epsilon(1-r^2)\Delta x^2 - 36\epsilon^2(1-r^2)\Delta x\right] k^4 + O\big(\epsilon^3 k^4, k^5\big)\,, \tag{51}$$

Again, we find a third-order numerical dissipation term independent of ϵ in the k^4-term that can compete with the physical second-order dissipation. The criterion for the physical dissipation to dominate is

$$\frac{1}{72} r^2 \Delta x^3 k^4 \ll \epsilon(1-r^2)k^2. \tag{52}$$

Solving for Δx leads to the threshold grid size $\Delta h^*_{\mathrm{DG}(1)}$:

$$\Delta x \ll 2\left[\frac{9\epsilon(1-r^2)}{r^2k^2}\right]^{1/3} = 2\left(\frac{9}{rk^2\,Pe}\right)^{1/3}, \tag{53}$$

thus,

$$\Delta h^*_{\mathrm{DG}(1)} := 2\left(\frac{9}{rk^2\,Pe}\right)^{1/3}, \tag{54}$$

which is a factor of $\left(\dfrac{9}{r^2}\right)^{1/3}$ larger than for the HR2 scheme shown in (37). When rescaled, this is the same result as Eq. (32) in Lowrie and Morel [5, p. 420] which was obtained from a modified differential-equation analysis.

We directly discretize the advection-diffusion limit (15) using the DG(1) scheme with the Rusanov flux function; this is equivalent to the HLL1 flux with $c = 1$. The diffusion term is discretized using the recently developed 'recovery method'.[23,24] The eigenvalues of spatial discretization from the Fourier analysis are

$$\begin{aligned}
\lambda^{(1),\ \text{adv-diff}}_{\mathrm{DG}(1)} - \lambda^{\text{adv-diff}}_{\mathrm{exact}} &= -\frac{r^2\Delta x^4}{36\,[2\Delta x + 5\epsilon(1-r^2)]}k^4 + O\big(k^5\big)\,,\\
\lambda^{(2),\ \text{adv-diff}}_{\mathrm{DG}(1)} - \lambda^{\text{adv-diff}}_{\mathrm{exact}} &= -\frac{6}{\Delta x} - \frac{15\epsilon(1-r^2)}{\Delta x^2} + O(k)\,.
\end{aligned} \tag{55}$$

The dominant eigenvalue, the first equation, can be further expanded in terms of ϵ since we are assuming the near-equilibrium limit $\epsilon \ll 1$; then

$$\lambda^{\text{adv-diff}}_{\mathrm{DG}(1)} - \lambda^{\text{adv-diff}}_{\mathrm{exact}} = -\frac{1}{72}\left[r^2\,\boxed{\Delta x^3} - \frac{5\epsilon r^2(1-r^2)}{2}\Delta x^2\right] k^4 + O\big(\epsilon^2 k^4, k^5\big)\,.$$

Again, the dominant numerical dissipation is of precisely the same form as for the GHHE discretization (51).

Finally, we note that, for $r = O(\epsilon)$ with $\epsilon \ll 1$, the DG(1) scheme exhibits an interesting property. In this case, the fourth-order numerical dissipation term in (51) becomes higher-order in ϵ, and the constraint (53) on

Δx is removed. Thus, the DG(1) scheme should converge with second-order accuracy with Δx independent of ϵ, since the higher-even-order terms are too small to compete with the physical dissipation. This case is included in the diffusive limit considered by Lowrie and Morel,[5] and our result agrees with theirs when one accounts for the time dilation of their scaling.

Spatial-Temporal Accuracy Following the procedure used earlier, the local truncation error of the DG(1)–MOL combined with the IMEX–SSP2(3,3,2) is found to have following form:

$$\mathrm{LTE}^{(1)}_{\mathrm{DG(1)IMEX}} = -\Bigg[\underbrace{\frac{ir(r\nu)^2}{12}\boxed{\Delta x^2}}_{\text{dominant dispersion error}} + \frac{i\epsilon\, r(1-r^2)\nu}{6}\Delta x\Bigg] k^3 - \Bigg[\underbrace{\frac{r\big(r-3(r\nu)^3\big)}{72}\boxed{\Delta x^3}}_{\text{dominant dissipation error}} + \frac{\epsilon(1-r^2)}{12}\Delta x^2\Bigg] k^4. \quad (56)$$

Just as for the HR2 method, if the physical dissipation is dominant over the dissipation error,

$$\frac{r\big(r-3(r\nu)^3\big)}{72}\Delta x^3 k^4 \ll \epsilon(1-r^2)k^2, \quad (57)$$

the method is second-order in space and time. Solving the above equation for Δx leads to the threshold grid size:

$$\Delta h^*_{\mathrm{DG(1)IMEX}} := 2\left[\frac{9\epsilon\,(1-r^2)}{r\big(r-3(r\nu)^3\big)k^2}\right]^{1/3}. \quad (58)$$

II.B.4. DG–Hancock Method

The DG(1)–Hancock method for the 1-D GHHE has the form:[2,3]

$$\begin{aligned}
\bar{\mathbf{u}}_j^{n+1/3} &= \bar{\mathbf{u}}_j^n - \frac{\Delta t}{3}\frac{1}{\Delta x}\left[\hat{\mathbf{f}}^{n+1/6}_{j+1/2} - \hat{\mathbf{f}}^{n+1/6}_{j-1/2}\right] + \frac{\Delta t}{3}\frac{1}{\epsilon}\left[\frac{5}{4}\mathbf{Q}\bar{\mathbf{u}}_j^{n+1/3} - \frac{1}{4}\mathbf{Q}\bar{\mathbf{u}}_j^{n+1}\right],\\
\bar{\mathbf{u}}_j^{n+1} &= \bar{\mathbf{u}}_j^n - \frac{\Delta t}{\Delta x}\left[\hat{\mathbf{f}}^{n+1/2}_{j+1/2} - \hat{\mathbf{f}}^{n+1/2}_{j-1/2}\right] + \frac{\Delta t}{\epsilon}\left[\frac{3}{4}\mathbf{Q}\bar{\mathbf{u}}_j^{n+1/3} + \frac{1}{4}\mathbf{Q}\bar{\mathbf{u}}_j^{n+1}\right],
\end{aligned} \quad (59a)$$

$$\begin{aligned}
\overline{\Delta\mathbf{u}}_j^{n+1/3} &= \overline{\Delta\mathbf{u}}_j^n - \frac{\Delta t}{3}\frac{6}{\Delta x}\left[\hat{\mathbf{f}}^{n+1/6}_{j+1/2} + \hat{\mathbf{f}}^{n+1/6}_{j-1/2} - \frac{2}{\Delta x\Delta t}\overline{\mathbf{f}(\mathbf{u}_j^{n+1/6})}\right] + \frac{\Delta t}{3}\frac{1}{\epsilon}\left[\frac{5}{4}\mathbf{Q}\overline{\Delta\mathbf{u}}_j^{n+1/3} - \frac{1}{4}\mathbf{Q}\overline{\Delta\mathbf{u}}_j^{n+1}\right],\\
\overline{\Delta\mathbf{u}}_j^{n+1} &= \overline{\Delta\mathbf{u}}_j^n - \frac{\Delta t}{\Delta x}6\left[\hat{\mathbf{f}}^{n+1/2}_{j+1/2} + \hat{\mathbf{f}}^{n+1/2}_{j-1/2} - \frac{2}{\Delta x\Delta t}\overline{\mathbf{f}(\mathbf{u}_j^{n+1/2})}\right] + \frac{\Delta t}{\epsilon}\left[\frac{3}{4}\mathbf{Q}\overline{\Delta\mathbf{u}}_j^{n+1/3} + \frac{1}{4}\mathbf{Q}\overline{\Delta\mathbf{u}}_j^{n+1}\right].
\end{aligned} \quad (59b)$$

After some algebra, the local truncation error of the dominant eigenvalue is found to be given by

$$\mathrm{LTE}^{(1)}_{\mathrm{DG(1)Ha}} = -\Bigg[\underbrace{\frac{r}{72}\left(\frac{r\big(1-(r\nu)^2\big)^2}{1-r^2\nu} - 3r\nu(1-\nu)\right)\boxed{\Delta x^3}}_{\text{dominant dissipation error}} + \frac{\epsilon\,(1-r^2)}{12(1-r^2\nu)^2} \times \left(1+\frac{1}{3}r^2 + \frac{\nu^2}{6}\big(2r^2(r^2-9) + 3r^4\nu + r^4(9-7r^2)\nu^2 + 3r^6\nu^3\big)\right)\Delta x^2\Bigg] k^4. \quad (60)$$

Note that the leading error is a k^4-term, hence a dissipation, whereas in other methods possess a leading dispersion error. The threshold grid size to guarantee the method be third-order accurate is

$$\Delta h^*_{\mathrm{DG(1)Ha}} := 2\left[\frac{9\epsilon\,(1-r^2)(1-r^2\nu)}{r^2\big(1-3\nu+(3+r^2)\nu^2-3r^2\nu^3+(r\nu)^4\big)k^2}\right]^{1/3}. \tag{61}$$

II.C. Dominant Dispersion/Dissipation Error in 1-D

In summary, the local truncation error of each method is listed for comparison.

semi-discrete methods:

$$\mathrm{LTE}_{\mathrm{HR2MOL}} = \left[c_3\left(1+(r\nu)^2\right)+\frac{1}{6}\tilde{c}_3\nu\right]k^3 \;+\left[\;c_4\left(\frac{1}{r}-\frac{1}{3}(r\nu)^3\right)+\frac{1}{6}\tilde{c}_4\right]k^4, \tag{62a}$$

$$\mathrm{LTE}^{(1)}_{\mathrm{DG(1)MOL}} = \left[c_3(r\nu)^2 \qquad\quad +\frac{1}{6}\tilde{c}_3\nu\right]k^3+\left[\frac{1}{9}c_4\left(r-3(r\nu)^3\right) \quad +\frac{1}{12}\tilde{c}_4\right]k^4, \tag{62b}$$

fully discrete methods:

$$\mathrm{LTE}^{(1)}_{\mathrm{DG(1)Ha}} = \left[\frac{1}{9}c_4\left(\frac{r\big(1-(r\nu)^2\big)^2}{1-r^2\nu}-3r\nu(1-\nu)\right)+\frac{\tilde{c}_4}{12(1-r^2\nu)^2}\right.$$
$$\left.\times\left(1+\frac{1}{3}r^2+\frac{\nu^2}{6}\big(2r^2(r^2-9)+3r^4\nu+r^4(9-7r^2)\nu^2+3r^6\nu^3\big)\right)\right]k^4, \tag{62c}$$

where the coefficients of errors attributed to the explicit flux discretization are

$$c_3 = -\frac{ir}{12}\boxed{\Delta x^2}, \quad c_4 = -\frac{r}{8}\boxed{\Delta x^3}, \tag{63a}$$

and the implicit source-term errors have coefficients

$$\tilde{c}_3 = -i\epsilon r(1-r^2)\boxed{\Delta x}, \quad \tilde{c}_4 = -\epsilon(1-r^2)\boxed{\Delta x^2}. \tag{63b}$$

The above equations show that HR2–MOL and DG(1)–MOL have a *first-order* error $\sim k^3$, as $\tilde{c}_3 = O(\epsilon\Delta x)$. However, numerically, this error term is not pronounced in the near-equilibrium limit since $\epsilon \ll 1$, thus the dominant error of the methods stem from $c_3 = O\big(\Delta x^2\big)$. Similarly, the dominant error of DG(1)–Hancock is $c_4 = O\big(\Delta x^3\big)$.

As it was described previously, when $r = O(\epsilon)$ with $\epsilon \ll 1$, the DG(1) method reveals uniform spatial convergence. To demonstrate this, we simply set $r = 0$, then the local truncation error of each method becomes

$$\mathrm{LTE}_{\mathrm{HR2MOL}} = \left[-\frac{1}{8}\boxed{\Delta x^3}-\frac{1}{6}\epsilon\Delta x^2\right]k^4, \tag{64a}$$

$$\mathrm{LTE}^{(1)}_{\mathrm{DG(1)MOL}} = -\frac{1}{12}\epsilon\Delta x^2\,k^4, \tag{64b}$$

$$\mathrm{LTE}^{(1)}_{\mathrm{DG(1)Ha}} = -\frac{1}{12}\epsilon\Delta x^2\,k^4. \tag{64c}$$

All dispersions, $O\big(k^3\big)$-terms, have disappeared, and the dominate error is the dissipation. The dominant dissipation errors of both DG(1) methods are proportional to ϵ, while HR2 methods have the term, $\frac{1}{8}\Delta x^3$,

which is independent of ϵ. Hence, DG(1) methods lose their grid size restrictions, but HR2 methods still need to satisfy the following inequality:

$$\frac{1}{8}\Delta x^3 k^4 \ll \epsilon k^2 \quad \longrightarrow \quad \Delta x \ll 2\left(\frac{\epsilon}{k^2}\right)^{1/3}, \tag{65}$$

to guarantee physical dissipation is dominant. Since the leading errors of HR2 methods are proportional to Δx^3, we expect third-order convergence on coarse grids when $r = 0$.

III. Analysis for 2-D Linear Hyperbolic-Relaxation Equations

III.A. Model Equations for Two-Dimensional Problem

In two dimensions we consider the simple system:

$$\partial_t u + \partial_x v + \partial_y w = 0, \tag{66a}$$

$$\partial_t v + \partial_x u + r\partial_y w = -\frac{1}{\epsilon}(v - ru), \tag{66b}$$

$$\partial_t w + s\partial_x v + \partial_y u = -\frac{1}{\epsilon}(w - su), \tag{66c}$$

where v and w are the fluxes in the x- and y-directions, respectively. The above equations can be written in vector form:

$$\partial_t \mathbf{u}(\boldsymbol{x}, t) + \partial_x \mathbf{f}(\mathbf{u}) + \partial_y \mathbf{g}(\mathbf{u}) = \frac{1}{\epsilon}\mathbf{s}(\mathbf{u}); \quad \boldsymbol{x} \in \mathbb{R}^2,\ t > 0, \tag{67}$$

with linear fluxes and source, $\mathbf{f}(\mathbf{u}) = \mathbf{A}\mathbf{u}$, $\mathbf{g}(\mathbf{u}) = \mathbf{B}\mathbf{u}$, and $\mathbf{s}(\mathbf{u}) = \mathbf{Q}\mathbf{u}$, where

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & s & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & r \\ 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{Q} = \frac{1}{\epsilon}\begin{pmatrix} 0 & 0 & 0 \\ r & -1 & 0 \\ s & 0 & -1 \end{pmatrix}. \tag{68}$$

The near-equilibrium limit is formally

$$\partial_t u + r\partial_x u + s\partial_y u = \epsilon\left[(1 - r^2)\partial_{xx} u + (1 - s^2)\partial_{yy} u\right] + O\left(\epsilon^2\right), \tag{69}$$

with the equilibrium wave speeds r and s in the x- and y-directions, respectively. For a harmonic mode with wave vector $\mathbf{k} = (k_x, k_y)$, a stability criterion in the near-equilibrium limit is found by insisting that the second-order derivative terms are dissipative; mathematically, this is $|rk_x + sk_y| \le |\mathbf{k}|$. Due to the complexity of the analysis, we restrict the discussion to a uniform grid with unit aspect ratio ($\Delta h := \Delta x = \Delta y$), and the wave frequencies in the x- and y-directions are the same, thus $\alpha = \beta$ and $k_x = k_y = k$. Based on these assumptions, the exact solution of the reduced equation (69) in the near-equilibrium limit is

$$\lambda_{\text{exact}}^{\text{GHHE}} = -i(r + s)k - \epsilon(2 - r^2 - s^2)k^2 - i\epsilon^2(r + s)(3 - 2r^2 + rs - 2s^2)k^3 + O\left(\epsilon^3\right). \tag{70}$$

III.B. Difference Operators and Their Properties in 2-D

In this section, due to the complexity of multidimensional Fourier analysis, we restrict ourselves to semi-discrete methods, namely, HR2–MOL and DG(1)–MOL methods.

III.B.1. HR–MOL Method

Following the previous analysis, the eigenvalues of the spatial discretization are obtained by solving the characteristic equations:

$$\det\left(\mathbf{N}_{\mathrm{HR2}}+\frac{1}{\epsilon}\mathbf{Q}-\lambda\mathbf{I}\right)=0. \tag{71}$$

Taking the low-frequency limit, under the assumptions of $\Delta x=\Delta y=\Delta h$ and $\alpha=\beta$, leads to asymptotic eigenvalues,

$$\begin{aligned}\lambda^{(1)}_{\mathrm{HR2}} &= -\frac{i(r+s)}{\Delta h}\beta-\frac{\epsilon(2-r^2-s^2)}{\Delta h^2}\beta^2 \\ &\quad -\left[\frac{i(r+s)}{12\Delta h}+\frac{i\epsilon^2(r+s)(3+2r^2-rs+2s^2)}{\Delta h^3}\right]\beta^3+O\big(\beta^4\big)\end{aligned} \tag{72a}$$

$$\lambda^{(2)}_{\mathrm{HR2}} = -\frac{1}{\epsilon}+\frac{ir}{\Delta h}\beta+O\big(\beta^2\big)\,, \tag{72b}$$

$$\lambda^{(3)}_{\mathrm{HR2}} = -\frac{1}{\epsilon}+\frac{is}{\Delta h}\beta+O\big(\beta^2\big)\,, \tag{72c}$$

where the first eigenvalue represents the dominant wave in the asymptotic limit, and the other waves damp quickly since the leading errors are large negative real. The spatial discretization error corresponding to the dominant wave is derived by replacing the wave frequency by the wave number, then

$$\lambda^{(1)}_{\mathrm{HR2}}-\lambda^{\text{2D-GHHE}}_{\text{exact}} = -\frac{i(r+s)}{12}\boxed{\Delta h^2}\,k^3-\left[\frac{1}{4}\boxed{\Delta h^3}+\frac{\epsilon(2-r^2-s^2)}{6}\Delta h^2\right]k^4+O\big(k^5\big)\,. \tag{73}$$

Thus, the spatial discretization error of the dominant wave is second-order in space. In order to ensure the physical dissipation is dominant in the near-equilibrium limit $\epsilon\ll 1$, the following relation has to be satisfied:

$$\frac{1}{4}\Delta h^3k^4\ll\epsilon(2-r^2-s^2)k^2. \tag{74}$$

Solving for Δh results in the threshold grid size:

$$\Delta h_{\mathrm{HR2}}\ll\left[\frac{4\epsilon\,(2-r^2-s^2)}{k^2}\right]^{1/3}. \tag{75}$$

III.B.2. DG–MOL Method

We look for roots of the characteristic polynomial

$$\det\left(\mathbf{N}_{\mathrm{DG(1)}}+\frac{1}{\epsilon}\mathbf{Q}_{\mathrm{DG(1)}}-\lambda\mathbf{I}\right)=0, \tag{76}$$

under the assumption $\alpha=\beta$. As in the 1-D case, we assume a power-series form for the eigenvalue, and solve for the eigenvalue order-by-order in β. We find the following eigenvalues:

$$\lambda^{(1)}_{\mathrm{DG(1)}} = -i\left(\frac{r}{\Delta x}+\frac{s}{\Delta y}\right)\beta-\epsilon\left(\frac{1-r^2}{\Delta x^2}+\frac{1-s^2}{\Delta y^2}\right)\beta^2+O\big(\beta^3\big)\,, \tag{77a}$$

$$\lambda^{(2)}_{\mathrm{DG(1)}} = -\frac{6}{\Delta x}+O(\beta)\,,\qquad \lambda^{(3)}_{\mathrm{DG(1)}} = -\frac{6}{\Delta y}+O(\beta)\,, \tag{77b}$$

$$\lambda^{(4)}_{\mathrm{DG(1)}} = -\frac{6}{\Delta x}-\frac{1}{\epsilon}+O(\beta)\,,\quad \lambda^{(5)}_{\mathrm{DG(1)}} = -\frac{6}{\Delta y}-\frac{1}{\epsilon}+O(\beta)\,, \tag{77c}$$

$$\lambda^{(6,7,8,9)}_{\mathrm{DG(1)}} = -\frac{1}{\epsilon}+O(\beta)\,. \tag{77d}$$

To simplify the analysis, we further assume a uniform grid, $\Delta x = \Delta y = \Delta h$, then the spatial-discretization error is obtained by comparing the dominant eigenvalue to the exact solution:

$$\begin{aligned}\lambda^{(1)}_{\mathrm{DG}(1)} - \lambda^{\text{2D-GHHE}}_{\text{exact}} &= -\left[\frac{1}{9}\Delta h^3 - \frac{2-r^2-s^2}{72(\Delta h + 6\epsilon)}\Delta h^4\right] k^4 + O\big(k^5\big) \\ &= -\frac{1}{72}\left[\Big(\underbrace{6}_{\text{multi-D error}} + r^2 + s^2\Big)\boxed{\Delta h^3} + 6\epsilon(2-r^2-s^2)\Delta h^2\right] k^4 + O\big(\epsilon^2 k^4, k^5\big)\,.\end{aligned} \tag{78}$$

The above equation shows that in the near-equilibrium limit $\epsilon \ll 1$, the dominant error is $O\big(\Delta h^3\big)$, thus the DG(1) spatial discretization is third-order in space. To ensure the physical dissipation is dominant, the mesh size has the following constrain:

$$\frac{6 + r^2 + s^2}{72}\Delta h^3 k^4 \ll \epsilon(2 - r^2 - s^2)k^2. \tag{79}$$

Solving for Δh leads to

$$\Delta h_{\mathrm{DG}(1)} \ll 2\left[\frac{9\epsilon\,(2-r^2-s^2)}{(6+r^2+s^2)k^2}\right]^{1/3}. \tag{80}$$

As pointed out in a Fourier analysis of the DG(1)–MOL method for the 2-D advection equation [2, p. 169], the two-dimensional DG(1) discretization contains a multidimensional error, $-\frac{1}{12}\Delta h^3$, in the $O\big(k^4\big)$-term. Since the upwind flux for the GHHE in the near-equilibrium limit is equivalent to the direct discretization of the advection equation with the Rusanov flux ($q_x = q_y = 1$), the above error really comes from the term $-\frac{1}{24}(q_x + q_y)\Delta h^3$. This extra multidimensional error eliminates the uniform-convergence property which the DG(1) method possess in one dimensional problem with a certain scaling.

In the 1-D case, for the specific scaling where the equilibrium speed is $r = O(\epsilon)$, the DG(1) method does not have any grid size restriction to achieve second-order accuracy. However, due to the multidimensional error independent of the equilibrium wave speeds r, s, there is always a grid size restriction even in the case where $r = O(\epsilon)$ with $\epsilon \ll 1$.

III.C. Dominant Dispersion/Dissipation Error

In summary, the local truncation errors of spatial discretization methods, HR2 and DG(1), are listed for comparison:

$$\mathrm{LTE}_{\mathrm{HR2}} = c_3\,k^3 + \left[\;c_4\,\frac{2}{r+s} \qquad\qquad + \;\frac{1}{6}\tilde{c}_4\right] k^4, \tag{81a}$$

$$\mathrm{LTE}_{\mathrm{DG}(1)} = \qquad\qquad \left[\frac{1}{9}c_4\left(\frac{6}{r+s} + \frac{r^2+s^2}{r+s}\right) + \frac{1}{12}\tilde{c}_4\right] k^4, \tag{81b}$$

where the coefficients of errors attributed to the flux discretization are

$$c_3 = -\frac{i(r+s)}{12}\boxed{\Delta h^2}, \quad c_4 = -\frac{r+s}{8}\boxed{\Delta h^3}, \tag{82a}$$

and the source-error coefficients are

$$\tilde{c}_4 = -\epsilon(2 - r^2 - s^2)\boxed{\Delta h^2}. \tag{82b}$$

Note that the above errors are merely the spatial discretization errors; unlike in the 1-D analysis given by (62) on page 15, the temporal errors are not considered. The HR2 method has a leading second-order dispersion error, whereas the DG(1) method is third-order accurate as long as grids are coarse, hence $\Delta h \gg O(\epsilon)$.

When we consider the case $r = s = 0$, the above truncation errors become

$$\mathrm{LTE}_{\mathrm{HR2}} = -\frac{1}{4}\boxed{\Delta h^3} - \frac{1}{3}\epsilon\Delta h^2, \tag{83a}$$

$$\mathrm{LTE}_{\mathrm{DG(1)}} = -\frac{1}{12}\boxed{\Delta h^3} - \frac{1}{6}\epsilon\Delta h^2. \tag{83b}$$

Unlike in one dimension, the DG(1) method possesses a dominant error independent to ϵ. Hence, both spatial discretizations yield grid-size restrictions to ensure the physical dissipation is dominant.

IV. Solid-boundary treatments

In this section we first discuss Grad's[25] approach to formulating boundary conditions for a gas described by a system of moments of Boltzmann's equation. We demonstrate an inconsistency in his boundary treatment and propose alternative treatments.

IV.A. Grad's formulation and its inconsistency

For a system of moment equations in extended hydrodynamics, Grad [25, p. 379] proposed a classical method to formulate wall-boundary conditions based on Maxwell's kinetic boundary condition.[26] In this approach the *velocity-distribution function* or, for short, *distribution function*, $f(v_x, v_y, v_z)$ of particles in the Knudsen layer is a linear combination of the distribution functions of the incoming particles, $f^-(v_x, v_y, v_z)$, and the reflected particles, $f^+(v_x, v_y, v_z)$. The total velocity vector of a particle $\mathbf{v} = \{v_x, v_y, v_z\}$ is the sum of the particle's thermal velocity $\mathbf{c}$ and the average fluid velocity $\mathbf{u}$, that is $\mathbf{v} = \mathbf{c} + \mathbf{u}$. The superscript "$-$" indicates that particles in this class travel in the direction opposite to the boundary normal, making their normal velocity component negative; similarly, the superscript "+" indicates particles moving in the direction of the normal, thus having positive normal velocity. In what follows the normal direction will be denoted by subscript "y." By definition, $f^-(v_x, v_y, v_z) = 0$ for $v_y > 0$ and $f^+(v_x, v_y, v_z) = 0$ for $v_y < 0$.

For any moment model, the velocity distribution of particles in the flow domain far away from the boundary is in a known form, hence the form of f^- is also known. The distribution function f^+ is more complicated because particles in this class experience collisions with the boundary. One possibility is that the particle is reflected specularly after collision; this type of collision reverses the normal component of the particle's momentum, everything else remains unchanged. Thus, the distribution function for this class is the mirror image of the one before collision, $f^-(v_x, v_y, v_z)$ with respect to the plane $(v_y = 0)$; hence $f^+_{\mathrm{specular}} = f^-(v_x, -v_y, v_z)$. Due to roughness of the boundary surface, there is a possibility that particles experiences enough collisions at the boundary to reach equilibrium before being reflected back into the flow; they may then be assumed to have acquired a *Maxwellian* distribution function, $f^+_{\mathrm{diffusive}} = f^{\mathrm{w}}_{\mathrm{M}}(v_x, v_y, v_z)$. This type of reflection is called diffusive reflection.

The overall distribution function in the Knudsen layer is expressed as:

$$\begin{aligned} f(v_x, v_y, v_z) = f^- + f^+ &= f^- + \left[\sigma f^+_{\text{diffusive}} + (1-\sigma) f^+_{\text{specular}}\right] \\ &= f^-(v_x, v_y, v_z) + \left[\sigma f^{\text{w}}_{\text{M}}(v_x, v_y, v_z) + (1-\sigma) f^-(v_x, -v_y, v_z)\right]. \end{aligned} \tag{84}$$

Here $\sigma \in [0,1]$, the accommodation factor, expresses how likely a particle will be diffusively reflected after collision with the boundary; $f^{\text{w}}_{\text{M}}(v_x, v_y, v_z)$ carries information about the temperature T^{w} and velocity $\mathbf{u}^{\text{w}}$ of the wall in the following form

$$f^{\text{w}}_{\text{M}}(v_x, v_y, v_z) = \frac{C\,\rho}{(2\pi R T^{\text{w}})^{3/2}} \exp\left[-\frac{(\mathbf{v}-\mathbf{u}^{\text{w}})\cdot(\mathbf{v}-\mathbf{u}^{\text{w}})}{2RT^{\text{w}}}\right]. \tag{85}$$

Without loss of generality it may be assumed that the solid boundary does not move in any direction, i. e., $\mathbf{u}^{\text{w}} = 0$. The coefficient C is to be determined by the boundary condition of non-penetration, or zero normal mass-flux at the wall:

$$\iiint_{-\infty}^{+\infty} v_y f(v_x, v_y, v_z)\, dv_x\, dv_y\, dv_z \equiv \langle v_y f \rangle = 0. \tag{86}$$

To calculate any *macroscopic* or *average* quantity in the Knudsen layer, the corresponding moment of the full distribution function given by (84) must be taken:

$$\begin{aligned} \langle w(\mathbf{v}) f) \rangle = &\iint_{-\infty}^{+\infty} \left[\int_{-\infty}^{0} w(\mathbf{v})\, f^-(v_x, v_y, v_z)\, dv_y\right] dv_x\, dv_z \\ &+ \sigma \iint_{-\infty}^{+\infty} \left[\int_{0}^{+\infty} w(\mathbf{v})\, f^{\text{w}}_{\text{M}}(v_x, v_y, v_z)\, dv_y\right] dv_x\, dv_z \\ &+ (1-\sigma) \iint_{-\infty}^{+\infty} \left[\int_{0}^{+\infty} w(\mathbf{v})\, f^-(v_x, -v_y, v_z)\, dv_y\right] dv_x\, dv_z. \end{aligned} \tag{87}$$

The choice of weight function $w(v)$ determines which macroscopic quantity will be calculated. For example, tangential velocity components are calculated by using $w(\mathbf{v}) = v_i$, $i = \{x, z\}$; the pressure tensor P_{ij} is calculated with $w(\mathbf{v}) = v_i v_j$, $(i,j) = \{x, y, z\}$; etc.

To demonstrate the inconsistency in Grad's derivation of boundary conditions, the aforementioned process is applied to the 10-moment model which uses an *anisotropic ellipsoidal* distribution function, or *Gaussian* distribution function [27, p. 58]. For this moment model, $f^-(v_x, v_y, v_z)$ has the following form:

$$f^-(v_x, v_y, v_z) = \frac{\rho}{(2\pi)^{3/2}\Delta^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{v}-\mathbf{u})^T \cdot \left(\frac{\mathbf{P}}{\rho}\right)^{-1} \cdot (\mathbf{v}-\mathbf{u})\right], \tag{88}$$

where $\Delta = |\mathbf{P}/\rho|$. From (86) the value of the unknown coefficient C is found to be $C = \sqrt{\frac{P_{yy}}{\rho R T^{\text{w}}}}$. Other flow quantities inside the Knudsen layer can be computed directly from the distribution function by evaluating the corresponding integrals, e. g.,

$$\begin{aligned} \rho u_x \equiv \langle v_x f \rangle &= -\frac{2(2-\sigma)}{\sigma}\sqrt{\frac{\rho}{2\pi P_{yy}}}\, P_{xy} \\ \Rightarrow P^{\text{1st}}_{xy} &= -\frac{\sigma}{2(2-\sigma)}\sqrt{\frac{2\pi P_{yy}}{\rho}}\, \rho u_x, \end{aligned} \tag{89a}$$

$$P_{xy}^{2nd} \equiv \langle c_1 c_2 f \rangle \underset{\text{per (86)}}{\overset{u_1=0}{=}} \langle v_x v_y f \rangle = -\frac{2\sigma}{2-\sigma}\sqrt{\frac{P_{yy}}{2\pi\rho}}\,\rho u_x. \tag{89b}$$

Note that the expressions for the viscous shear-stress component P_{xy} obtained from the first-order moment (89a) and from the second-order moment (89b) are different; their ratio is $P_{\mathbf{xy}}^{\mathbf{1st}}/P_{\mathbf{xy}}^{\mathbf{2nd}} = \frac{\boldsymbol{\pi}}{\mathbf{2}}$, instead of *unity*. Thus, when one needs to impose a boundary condition on P_{xy}, which expression should be used? This is the aforementioned inconsistency in Grad's formulation of solid-boundary conditions, which was also recognized by Grad himself [25, p. 380].

Grad suggested a remedy for this situation in [25, p. 381] based on considering the characteristic structure of the moment equations and the behavior of the full distribution function's moments in the limit case of pure specular reflection, i. e., $\sigma = 0$. In this limit, (84) shows that the distribution function $f(v_x, v_y, v_z)$ is an even function of v_y, hence any *odd-order* v_y-moment of $f(v_x, v_y, v_z)$ has to be *zero*. Furthermore, for the 2-D 10-moment model, a characteristic analysis shows that in space-time there are two characteristic cones coming off the wall, therefore only two wall-boundary conditions are needed, although the system has seven independent variables: $[\rho, u_x, u_y, P_{xx}, P_{xy}, P_{yy}, P_{zz}]^{\mathrm{T}}$. One condition is the non-penetration condition (86); as the second one we require that $P_{xy} = 0$ when $\sigma = 0$, since P_{xy} is the only remaining odd-order moment in v_y. We observe that P_{xy}^{2nd} as computed from $\langle c_x c_y f \rangle$ indeed vanishes when $\sigma = 0$, satisfying Grad's second condition. However, so does the expression (89a), derived from $\langle v_x f \rangle$. Thus, Grad's alleged remedy fails to remove the inconsistency for $\sigma \neq 0$.

IV.B. An alternative approach to solid-boundary conditions

The aforementioned inconsistency happens because *local* quantities in the Knudsen layer are utilized in the distribution function $f^-(v_x, v_y, v_z)$. This makes that local average quantities will appear on the right-hand side of (87), while its left-hand side, by definition, is a local average quantity. Therefore, when (87) is used to compute any average quantity, that very quantity will appear on both sides of the expression, making it an equation to be solved for the quantity of interest. This is technically equivalent to imposing additional conditions on the distribution function (84) in the Knudsen layer. It requires the distribution function to have a sufficient number of *degree-of-freedom* or *dof*, in order to avoid mathematical inconsistency.

Reviewing the specific case of the *Gaussian* 10-moment model, the distribution function $f(v_x, v_y, v_z)$ indeed contains local average quantities inside $f^-(v_x, v_y, v_z)$ given by (88). There is only one *dof*, in the form of the coefficient C in $f_{\mathrm{M}}^{\mathrm{w}}$, while there are at least two conditions to be imposed on $f(v_x, v_y, v_z)$: *non-penetration*, and equality of P_{xy}^{1st} and P_{xy}^{2nd}. Therefore, there are more restrictions than *dof*. The value of C was determined by the *non-penetration* condition (86); the second condition was not satisfied.

Here we propose an alternative approach. Firstly, we would like to eliminate the appearance of additional constraints caused by the calculation of macroscopic quantities as described above. To achieve this, we will assume from now on that the distribution function $f(v_x, v_y, v_z)$ contains only macroscopic quantities from *outside* the Knudsen layer, denoted by the superscript "$\sim$." Thus, the right-hand side of (87) no longer contains any local macroscopic quantities of interest. Eqn. (87) now is just a recipe for computing quantities

inside the Knudsen layer in terms of quantities outside Knudsen layer, and parameters of the solid boundary. This makes physical sense and removes the inconsistency problem.

Secondly, we suggest to impose a new condition on $f(v_x, v_y, v_z)$ besides *non-penetration* one; it is a *normalization* condition:

$$\langle f/\rho \rangle \equiv \left\langle \widehat{f} \right\rangle = 1. \tag{90}$$

This requires further explanation. The distribution function $f(v_x, v_y, v_z)\, d\mathbf{v}$ expresses the mass density of particles having a total velocity in the range of $\mathbf{v} \to \mathbf{v} + d\mathbf{v}$, at an arbitrary space-time location. On the other hand, the distribution function $\widehat{f} = f/\rho$ carries *no information* about mass density; $\widehat{f}(v_x, v_y, v_z)\, d\mathbf{v}$ expresses the *probability* to find a particle having its velocity in the range $\mathbf{v} \to \mathbf{v} + d\mathbf{v}$, at an arbitrary space-time location. All particles in the flow field have a real velocity vector; thus there is 100% certainty to find a particle with velocity in the range of $\mathbf{v}$ from $-\infty$ to $+\infty$. This leads to making the *normalization* condition (90) a fundamental requirement.

In order for this second condition to be satisfied, an additional *dof* in the form of a coefficient C_1 is introduced into the expression for $f^-(v_x, v_y, v_z)$; the coefficient C in $f_{\mathrm{w}}^{\mathrm{M}}(v_x, v_y, v_z)$ is renamed C_2. The coefficient C_1 must also appear in the expression for $f^+_{\mathrm{specular}} = f^-(v_x, -v_y, v_z)$, since specular reflection does not change anything but the sign of the normal component of momentum.

We illustrate the above approach again on the basis of the *Gaussian* 10-moment model. The distribution function in the Knudsen layer becomes

$$\widehat{f}(v_x, v_y, v_z) = \widehat{f^-}(v_x, v_y, v_z) + \left[\sigma \widehat{f}_{\mathrm{M}}^{\mathrm{w}}(v_x, v_y, v_z) + (1-\sigma)\, \widehat{f^-}(v_x, -v_y, v_z)\right]; \tag{91}$$

with

$$\widehat{f^-}(v_x, v_y, v_z) = \frac{C_1}{(2\pi)^{3/2} \widetilde{\Delta}^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{v} - \widetilde{\mathbf{u}})^T \cdot \left(\frac{\widetilde{\mathbf{P}}}{\widetilde{\rho}}\right)^{-1} \cdot (\mathbf{v} - \widetilde{\mathbf{u}})\right], \tag{92}$$

and

$$\widehat{f}_{\mathrm{M}}^{\mathrm{w}}(v_x, v_y, v_z) = \frac{C_2}{(2\pi R T^{\mathrm{w}})^{3/2}} \exp\left[-\frac{(\mathbf{v} - \mathbf{u}^{\mathrm{w}}) \cdot (\mathbf{v} - \mathbf{u}^{\mathrm{w}})}{2 R T^{\mathrm{w}}}\right]. \tag{93}$$

Using conditions (86) and (90), the values of C_1 and C_2 become

$$C_1 = \frac{2}{2 - \sigma + \sigma\, r}, \tag{94a}$$

$$C_2 = \frac{2r}{2 - \sigma + \sigma\, r}, \tag{94b}$$

with $r = \sqrt{\widetilde{P}_{yy}/\widetilde{\rho} R T^{\mathrm{w}}} > 0$ under all circumstances. Because $\sigma \leq 1$, both coefficients are positive. Next, tangetial velocity u_x (instead of momentum ρu_x) and shear stress P_{xy} are calculated by (87) with $f(v_x, v_y, v_z)$ replaced by $\widehat{f}(v_x, v_y, v_z)$:

$$u_x = \left\langle v_x \widehat{f} \right\rangle = (2-\sigma)\, C_1 \left[\frac{\widetilde{u}_x}{2} - \frac{\widetilde{P}_{xy}}{\sqrt{2\pi \widetilde{\rho}\, \widetilde{P}_{yy}}}\right], \tag{95a}$$

$$P_{xy} = \left\langle \rho\, v_x v_y \widehat{f} \right\rangle = \sigma C_1 \rho \left[\frac{\widetilde{P}_{xy}}{2\widetilde{\rho}} - \sqrt{\frac{\widetilde{P}_{yy}}{2\pi\widetilde{\rho}}}\, \widetilde{u}_x\right]. \tag{95b}$$

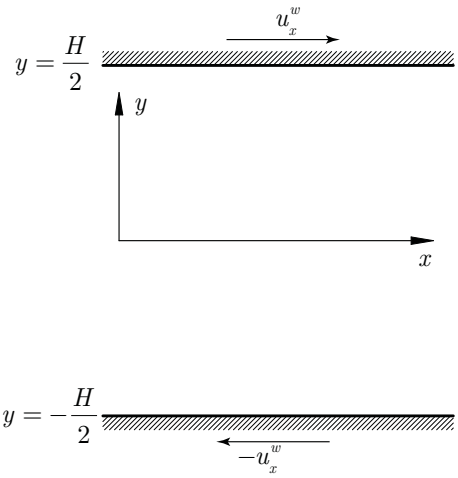

Figure 2. Couette flow geometry and coordinate system.

These new results show that there now is *coupling* between (95a) and (95b), unlike before when we had *inconsistency* between (89a) and (89b). Furthermore, (95b) shows that $P_{xy} = 0$ at $\sigma = 0$ as expected. Thus, our alternative approach indeed does not create inconsistency, and satisfies another essential condition.

When solving a moment system numerically, it suffices to compute *fluxes* at a solid boundary. These fluxes can be calculated by integrating the distribution function with appropriate weighting, as used in [28, p. 132]. From a moment system, the vector of fluxes in x_l-direction can be calculated directly by the following integral,

$$\mathbf{F}_l = \left\langle \rho v_l \mathbf{W}(\mathbf{v}) \widehat{f} \right\rangle, \tag{96}$$

where $\mathbf{W}(\mathbf{v})$ is the weight vector. Its value depends on which moment model is being studied. For example, for the 10-moment model the value of $\mathbf{W}(\mathbf{v})$ is $[1, v_i, v_i v_j]^{\mathrm{T}}$, and for the 13-moment model it is $[1, v_i, v_i v_j, v_i v_j v_j]^{\mathrm{T}}$. In the Knudsen layer, $\widehat{f}$ is constructed as above, and one can use (96) to calculate the required fluxes at the solid boundary in every direction. In[28] the authors used this *flux-based* boundary condition in an analytical study of a 14-moment model using Levermore's closure.[29] They successfully proved this combination has a *unique* solution. We will use the same flux-based technique.

IV.C. Numerical validation by *linearized* Couette flow

To validate the proposed alternative boundary treatment, a simple 2-D *linearized* Couette flow is solved for both continuum and transitional flows. Geometry and coordinate system for the flow are shown in Figure 2. There exists an analytical approximate solution to this problem, Lees's solution; its brief description and comparison with the Navier–Stokes solution is given in [30, p. 429].

Again, the *Gaussian* 10-moment model is used as physical model in this problem. In 3-D, it is a system of ten hyperbolic-relaxation PDEs; it describes the time and space evolution of ten macroscopic quantities: density ρ, average velocity u_i, and components of the pressure tensor P_{ij}. The pressure tensor is also expressed as a combination of pressure $p = P_{jj}/3$ and non-equilibrium quantities p_{ij}, which relate to shear

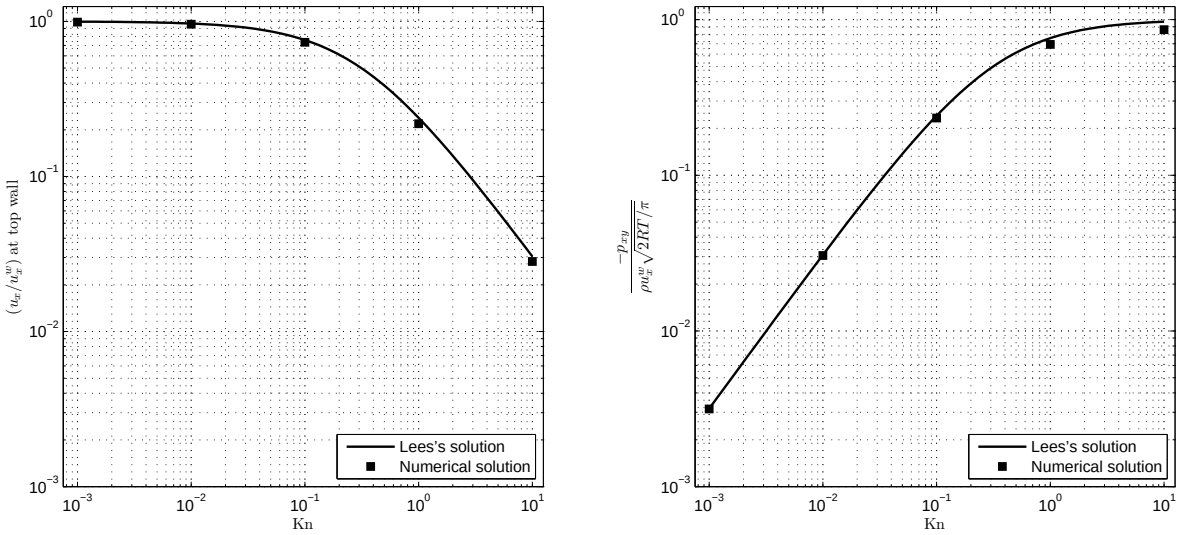

Figure 3. Numerical results for *linearized* Couette flow with *adiabatic* and *fully-diffusive* wall boundary conditions.

stresses τ_{ij} in the Navier–Stokes description by $p_{ij} = -\tau_{ij}$. The 3-D form is

$$\frac{\partial}{\partial t}\begin{bmatrix} \rho \\ \rho u_i \\ \rho u_i u_j + p_{ij} + p\,\delta_{ij} \end{bmatrix} + \frac{\partial}{\partial x_k}\begin{bmatrix} \rho u_k \\ \rho u_i u_k + P_{ik} \\ \rho u_i u_j u_k + (u_i P_{jk} + u_j P_{ik} + u_k P_{ij}) \end{bmatrix} = -\frac{1}{\tau}\begin{bmatrix} 0 \\ 0 \\ P_{ij} - p\,\delta_{ij} \end{bmatrix}. \quad (97)$$

For this 2-D flow problem, it reduces to a system of only seven unknowns: $[\rho, u_x, u_y, p, p_{xx}, p_{xy}, p_{yy}]^{\mathrm{T}}$. The system is solved assuming argon gas at 273K, adiabatic and fully diffusive wall conditions, and very small Mach number, $\mathrm{M}_{\mathrm{wall}} = 0.01$, to satisfy condition (98) for linearizability [30, p. 424]

$$\frac{\left(u_y^{\mathrm{w}}\right)^2}{2RT_w} \ll 1. \quad (98)$$

Because our current concern is validity of the boundary approach, this flow problem is solved by a first-order finite-volume method with HLL Riemann solver,[31] instead of DG(1)–Hancock analyzed in previous sections.

Numerical results for normalized velocity and shear stress are plotted against Lees's analytical approximation in Figure 3. The solutions agree with each other very well quantitatively, up to the limit of transitional and free-molecular flow. For the higher Knudsen numbers the numerical velocity and shear-stress values fall below the curve of Lees's solution. However, the same happens for a numerical solution based on Boltzmann's equation, with BGK approximation of the source terms; see [30, p. 433, Fig. 7]. It appears that for $\mathrm{Kn} \approx 1$ the results of the moment approach are closer to the full Boltzmann results than Lees's solution is. This shows that our new boundary treatment does a good job in describing strong non-equilibrium effects next to a solid boundary. Obviously, further validation and comparisons are necessary.

V. Conclusions

This paper deals with the numerical treatment of hyperbolic-relaxation systems such as result from taking moments of Boltzmann's equation for fluid flow. The main conclusions are:

1. Several discretization methods including DG(1)–Hancock, and semi-discrete finite-volume and DG methods are applied to linear hyperbolic-relaxtion systems for a Fourier analysis. Analyses show that the existence of grid size restriction in order to ensure the intended order of accuracy; DG(1)–Hancock is the least restrictive method among others. It is also shown that a two-dimensional extension of DG(1) method introduces an extra multidimensional dissipation error. Owing to its third-order accuracy with less function evaluations as a result of a space-time discretization, the DG(1)–Hancock method is not only accurate but efficient in comparison to other semi- and fully discrete finite-volume and discontinuous-Galerkin methods.

2. An inconsistency in Grad's solid-boundary treatment for moment systems can be removed; the resulting alternative boundary treatment performs well in a numerical validation based on *linearized* Couette flow.

References

[1]Van Leer, B., "Computational Fluid Dynamics: science or toolbox?" *15th AIAA Computational Fluid Dynamics Conference*, Anaheim, California; USA, 2001, AIAA Paper 2001-2520.

[2]Suzuki, Y., *Discontinuous Galerkin Methods for Extended Hydrodynamics*, Ph.D. thesis, The University of Michigan, 2008.

[3]Suzuki, Y. and Van Leer, B., "An analysis of the upwind moment scheme and its extension to systems of nonlinear hyperbolic-relaxation equations," *18th AIAA Computational Fluid Dynamics Conference*, Miami, Florida; USA, June 25–28, 2007, AIAA 2007-4468.

[4]Huynh, H. T., "An upwind moment scheme for conservation laws," *Computational Fluid Dynamics 2004: Proceedings of the Third International Conference on Computational Fluid Dynamics, ICCFD3, Toronto, 12–16 July 2004*, edited by C. Groth and D. W. Zingg, Springer-Verlag, Berlin, 2006, pp. 761–766.

[5]Lowrie, R. B. and Morel, J. E., "Methods for hyperbolic systems with stiff relaxation," *International Journal for Numerical Methods in Fluids*, Vol. 40, No. 3-4, 2002, pp. 413–423.

[6]Hittinger, J. A., *Foundations for the Generalization of the Godunov Method to Hyperbolic Systems with Stiff Relaxation Source Terms*, Ph.D. thesis, The University of Michigan, 2000.

[7]Jin, S. and Levermore, C. D., "Numerical schemes for hyperbolic conservation laws with stiff relaxation terms," *Journal of Computational Physics*, Vol. 126, No. 2, 1996, pp. 449–467.

[8]Natalini, R., "Recent results on hyperbolic relaxation problems," *Analysis of Systems of Conservation Laws*, edited by H. Freistühler, Monographs and Surveys in Pure and Applied Mathematics, Volume 99, Chapman & Hall/CRC, Boca Raton, 1998, pp. 128–198.

[9]Hittinger, J. A. F. and Roe, P. L., "Asymptotic analysis of the Riemann problem for constant coefficient hyperbolic systems with relaxation," *Zeitschrift für Angewandte Mathematik und Mechanik*, Vol. 84, No. 7, 2004, pp. 452–471.

[10]Liu, T.-P., "Hyperbolic conservation laws with relaxation," *Communications in Mathematical Physics*, Vol. 108, No. 1, 1987, pp. 153–175.

[11]Naldi, G. and Pareschi, L., "Numerical schemes for hyperbolic systems of conservation laws with stiff diffusive relaxation," *SIAM Journal on Numerical Analysis*, Vol. 37, No. 4, 2000, pp. 1246–1270.

[12]Hittinger, J. A. F., Suzuki, Y., and Van Leer, B., "Investigation of the discontinuous Galerkin method for first-order PDE approaches to CFD," *17th AIAA Computational Fluid Dynamics Conference*, Toronto, Ontario; Canada, June 6–9, 2005, AIAA Paper 2005-4989.

[13]Arora, M. and Roe, P. L., "Issues and strategies for hyperbolic problems with stiff source terms," *Barriers and Challenges in Computational Fluid Dynamics*, edited by V. Venkatakrishnan, M. D. Salas, and S. R. Chakravarthy, ICASE/LaRC Interdisciplinary Series in Science and Engineering, Kluwer Academic Publishers, Dordrecht, 1998, pp. 139–154.

[14]Jin, S., "Runge–Kutta methods for hyperbolic conservation laws with stiff relaxation terms," *Journal of Computational Physics*, Vol. 122, No. 1, 1995, pp. 51–67.

[15]Caflisch, R. E., Jin, S., and Russo, G., "Uniformly accurate schemes for hyperbolic systems with relaxation," *SIAM Journal on Numerical Analysis*, Vol. 34, No. 1, 1997, pp. 246–281.

[16]Liotta, S. F., Romano, V., and Russo, G., "Central schemes for balance laws of relaxation type," *SIAM Journal on Numerical Analysis*, Vol. 38, No. 4, 2000, pp. 1337–1356.

[17]Tyson, R., Stern, L. G., and LeVeque, R. J., "Fractional step methods applied to a chemotaxis model," *Journal of Mathematical Biology*, Vol. 41, No. 5, 2000, pp. 455–475.

[18]Hairer, E. and Wanner, G., *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*, Springer Series in Computational Mathematics 14, Springer-Verlag, Berlin, second revised ed., 1996.

[19]Lambert, J. D., *Numerical Methods for Ordinary Differential Systems: The Initial Value Problem*, John Wiley & Sons, New York, 1991.

[20]Ascher, U. M., Ruuth, S. J., and Wetton, B. T. R., "Implicit-explicit methods for time-dependent partial differential equations," *SIAM Journal on Numerical Analysis*, Vol. 32, No. 3, 1995, pp. 797–823.

[21]Ascher, U. M., Ruuth, S. J., and Spiteri, R. J., "Implicit-explicit Runge–Kutta methods for time-dependent partial differential equations," *Applied Numerical Mathematics*, Vol. 25, No. 2-3, 1997, pp. 151–167.

[22]Pareschi, L. and Russo, G., "Implicit-explicit Runge–Kutta schemes and applications to hyperbolic systems with relaxation," *Journal of Scientific Computing*, Vol. 25, No. 1, 2005, pp. 129–155.

[23]Van Leer, B. and Nomura, S., "Discontinuous Galerkin for diffusion," *17th AIAA Computational Fluid Dynamics Conference*, Toronto, Ontario; Canada, June 6–9, 2005, AIAA Paper 2005-5108.

[24]Van Leer, B., Lo, M., and Van Raalte, M., "A discontinuous Galerkin method for diffusion based on recovery," *18th AIAA Computational Fluid Dynamics Conference*, Miami, Florida; USA, June 25–28, 2007, AIAA Paper 2007-4083.

[25]Grad, H., "On the kinetic theory of rarefied gases," *Communications on Pure and Applied Mathematics*, Vol. 2, No. 4, 1949, pp. 331–407.

[26]Maxwell, J. C., "On Stresses in Rarefied Gases Arising from Inequalities of Temperature," *Philosophical Transactions of the Royal Society of London*, Vol. 170, 1879, pp. 231–256.

[27]Brown, S. L., *Approximate Riemann Solvers for Moment Models of Dilute Gases*, Ph.D. thesis, The University of Michigan, 1996.

[28]Le Tallec, P. and Perlat, J. P., "Boundary conditions and existence results for Levermore's moments system," *Mathematical Models and Methods in Applied Sciences*, Vol. 10, No. 1, 2000, pp. 127–152.

[29]Levermore, C. D., "Moment closure hierarchies for kinetic theories," *Journal of Statistical Physics*, Vol. 83, No. 5-6, 1996, pp. 1021–1065.

[30]Vincenti, W. G. and Kruger, Jr., C. H., *Introduction to Physical Gas Dynamics*, Krieger Publishing Company, Malabar, Florida, 1986.

[31]Harten, A., Lax, P. D., and Van Leer, B., "On Upstream Differencing and Godunov-Type Schemes for Hyperbolic Conservation Laws," *SIAM Review*, Vol. 25, No. 1, 1983, pp. 35–61.