Moment realizability and the validity of the Navier–Stokes equations for rarefied gas dynamics

C. David Levermore
Department of Mathematics and Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85721

William J. Morokoff
School of Mathematics, Institute of Advanced Study, Princeton, New Jersey 08540

B. T. Nadiga
Applied Theoretical and Computational Physics Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

(Received 20 February 1998; accepted 26 August 1998)

We present criteria for monitoring the validity of the Navier–Stokes approximation during the simulation of a rarefied gas. Our approach is based on an underlying kinetic formulation through which one can construct nondimensional non-negative definite matrices from moments of the molecular distribution. We then identify one such 3×3 matrix that can be evaluated intrinsically in the Navier–Stokes approximation. Our criteria are based on deviations of the eigenvalues of this matrix from their equilibrium value of unity. Not being tied to a particular benchmark problem, the resulting criteria are portable and may be applied to any Navier–Stokes simulation. We study its utility here by comparing stationary planar shock profiles computed using the Navier–Stokes equations with those computed using Monte Carlo simulations. © *1998 American Institute of Physics.* [S1070-6631(98)01312-9]

I. INTRODUCTION

The proper simulation of rarefied gases presents a formidable challenge to computational science. In situations where the gas is sufficiently dense that the mean free path of a particle (the average distance traveled between collisions) is much smaller than the macroscopic length scales of interest, fluid dynamical descriptions are valid. In that case the velocity distribution of particles will approach a local equilibrium parameterized by so-called fluid variables, typically the mass density, fluid velocity, and temperature. The evolution of these fluid variables will be governed either by the compressible Euler equations, which approximate the velocity distribution by a local equilibrium, or the compressible Navier–Stokes equations, which account for small deviations of the velocity distribution from a local equilibrium. Such fluid dynamical equations are routinely solved numerically to effectively model gases in a wide variety of applications.

However, when a gas is sufficiently rarefied that the mean free path is no longer much smaller than the macroscopic length scales, the deviation of the velocity distribution from a local equilibrium may become large and Navier–Stokes equations can yield momentum and energy fluxes that are inconsistent with non-negative distribution functions and that may even be wrong by orders of magnitude. In that case the gas can be modeled by a kinetic theory of dilute gases like that governed by the Boltzmann equation.[1] The gas is then described by single particle phase-space densities (one for each species) rather than fluid dynamical variables, and the evolution of these phase-space densities is then governed by kinetic equations. Such equations may be effectively solved via molecular dynamics or Monte Carlo methods[2] at low densities that are far outside the fluid dynamical regime. However, because of its phase-space description and numerical stiffness, the computational cost of doing so in regimes near the fluid dynamical limit becomes too prohibitive in both time and storage requirements to allow for general usage. An alternative to a full kinetic description is to use a moment closure,[3–5] but these too are significantly more expensive than a Navier–Stokes simulation.

Given this state of affairs, there is considerable incentive to employ the (compressible) Navier–Stokes approximation as far outside of the fluid dynamical regime as possible, and to develop hybrid numerical methods which couple the fluid regimes to the kinetic regimes. This would allow the expensive kinetic solvers to be used on as small a domain as possible. A typical approach is to solve the Navier–Stokes equations in the entire domain, and then identify regions of breakdown where it is suspected that the solution has physically unacceptable errors. Such a scheme creates the need for the development of criteria that can be monitored during the course of a Navier–Stokes simulation so as to indicate the magnitude of the errors being generated and, thereby, check the internal validity of the approximation. Indeed, the question of the validity of Navier–Stokes for the planar shock wave has been addressed quite extensively in the gas-dynamic literature over the last four decades.[6–9] Because the Navier–Stokes equations can be systematically recovered from the underlying kinetic description in the regime of small mean free paths by truncating either a Hilbert or Chapman–Enskog expansion, kinetic theory can provide the foundation for any such criterion. For example, one could try

 © 1998 American Institute of Physics

to use the first neglected term in the Chapman–Enskog expansion to estimate the error. That is not the approach taken here. Rather, we first present criteria based on the requirement that certain velocity moments generated by the velocity distribution of the truncated Chapman–Enskog expansion (which is generally not non-negative) be realizable by some non-negative velocity distribution. Such criteria are called *moment realizability criteria*. We then generalize this approach to more sensitive criteria which measure the deviation of the underlying distribution function from equilibrium.

Our approach will be presented for a single species monatomic gas in three dimensions. In that setting the Navier–Stokes equations for the evolution of the mass density $\rho = \rho(x,t) \geqslant 0$, the fluid velocity $u = u(x,t) \in \mathbb{R}^3$, and the temperature $\theta = \theta(x,t) \geqslant 0$ can be written as

$$\partial_t \rho + u \cdot \nabla_x \rho + \rho \nabla_x \cdot u = 0,$$

$$\partial_t u + u \cdot \nabla_x u + \frac{1}{\rho} \nabla_x (\rho \theta) = \frac{1}{\rho} \nabla_x \cdot \left(\mu \left[\nabla_x u + (\nabla_x u)^T - \frac{2}{3} \nabla_x \cdot u I \right] \right), \tag{1}$$

$$\tfrac{3}{2} (\partial_t \theta + u \cdot \nabla_x \theta) + \theta \nabla_x \cdot u = \frac{1}{2} \frac{\mu}{\rho} \mathrm{tr} \left(\left[\nabla_x u + (\nabla_x u)^T - \frac{2}{3} \nabla_x \cdot u I \right]^2 \right) + \frac{1}{\rho} \nabla_x \cdot (\kappa \nabla_x \theta),$$

where the viscosity $\mu = \mu(\rho, \theta) > 0$ and the heat conduction $\kappa = \kappa(\rho, \theta) > 0$ are prescribed. Here θ is related to the physical temperature T by $\theta = kT/m$ where k is the Boltzmann constant and m is the mass of a gas atom.

In this paper we apply ideas of moment realizability to derive criteria for the validity of the Navier–Stokes approximation. These criteria are given in terms of the eigenvalues of the nondimensional 3×3 matrix,

$$V_{NS} = I - \frac{\mu}{\rho \theta} \left[\nabla_x u + (\nabla_x u)^T - \frac{2}{3} \nabla_x \cdot u I \right] - \frac{2}{3} \frac{\kappa^2}{\rho^2 \theta^3} \nabla_x \theta (\nabla_x \theta)^T. \tag{2}$$

The eigenvalues of this so-called validity matrix can be readily computed at any mesh point during a Navier–Stokes simulation. Significant deviations of these eigenvalues from unity indicates large gradients and suggests that the assumption of a near equilibrium gas, necessary for the validity of the Navier–Stokes equations, may no longer be correct. On the other hand, the closer these eigenvalues are to unity, the more confidence one would have in the simulation. When and where one of these eigenvalues becomes negative, the fluxes computed from the velocity distribution used in the Navier–Stokes approximation are not realizable by any non-negative distribution, and have therefore become unphysical. This is the extreme case of breakdown. Depending on the application, the user may fine tune the size of acceptable deviations of the eigenvalues from unity to give a more sensitive criterion.

The remainder of the paper is laid out as follows. In Sec. II we first present the general theory of moment realizability criteria as derived from kinetic theory. We then show that for the Navier–Stokes approximation a moment realizability criterion may be expressed in terms of quantities that are intrinsic to the Navier–Stokes equations. In Sec. III we present a numerical comparison of the Navier–Stokes solution and the Monte Carlo solution for a stationary planar shock profile at various Mach numbers. We investigate the relationship of the validity matrix (2) to the breakdown of the Navier–Stokes solution, and thereby demonstrate the applicability of the proposed criterion. The detailed assumptions on the kinetic equation required for the derivation are presented in the Appendix.

II. THE MOMENT REALIZABILITY CRITERION

A. Fluid dynamical closures and moment realizability

We begin by deriving a general moment realizability criterion from the foundations of the kinetic theory of gases. This criterion is then applied to the Chapman–Enskog distribution function from which the Navier–Stokes equations are obtained, which thereby yields the criterion that V_{NS} given by (2) be non-negative definite. A comparison of the kinetic fluxes and velocity moments, from which the general criterion is composed, with their Navier–Stokes approximations may then be used as a guide for interpreting the moment realizability criterion and for analyzing the breakdown of the Navier–Stokes equations.

In this paper we will treat as illustrative the case of a gas composed of a single species of classical point particles contained within a fixed spatial domain $\Omega \subset \mathbb{R}^3$. Kinetic theories describe such a gas through the evolution of a non-negative density $F = F(t,x,v)$ over the single particle phase-space $\Omega \times \mathbb{R}^3$. We will denote the integral of any scalar or vector-valued function $f = f(v)$ over the 3-dimensional measure d_v^3 by $\langle f \rangle$; thus

$$\langle f \rangle = \int f(v) d_v^3.$$

The spatial densities of mass, momentum, and energy, respectively, are then recovered from the phase-space density F as

$$\langle F \rangle = \rho, \quad \langle vF \rangle = \rho u, \quad \langle \tfrac{1}{2} |v|^2 F \rangle = \tfrac{1}{2} \rho |u|^2 + \tfrac{3}{2} \rho \theta. \tag{3}$$

These relations define the mass density $\rho = \rho(x,t) \geqslant 0$, the fluid velocity $u = u(x,t) \in \mathbb{R}^3$, and the temperature $\theta = \theta(x,t) \geqslant 0$ in terms of F.

The evolution of F is usually governed by a kinetic equation of the form

$$\partial_t F + v \cdot \nabla_x F = \mathcal{C}(F), \tag{4}$$

where the collision operator $F \mapsto \mathcal{C}(F)$ acts only on the v dependence of F locally at each (t,x). The exact nature of the domain Ω as well as any associated boundary conditions plays a minor role in what follows, and so will remain un-

specified. We consider the class of collision operators $\mathcal{C}(F)$ that satisfy the assumptions precisely stated in the appendix. This class includes classical Boltzmann, Fokker–Planck, and BGK collision operators. The point here is the universality with respect to the collision operator of the moment realizability criteria based on the matrix (2). This matrix may be used to assess the validity of the Navier–Stokes approximation for any such kinetic equation, the difference being only that the various collision operators lead to different transport coefficients (viscosity and thermal conductivity). In Sec. III we present a comparison between the Boltzmann equation for rarefied gas of hard spheres and its Navier–Stokes approximation.

The collisions between the gas particles are assumed to conserve mass, momentum, and energy. Mathematically, this means that the collision operator $\mathcal{C}$ must satisfy

$$\langle \mathcal{C}(f)\rangle = 0, \quad \langle v\,\mathcal{C}(f)\rangle = 0, \quad \langle |v|^2 \mathcal{C}(f)\rangle = 0, \tag{5}$$

for every f for which these expressions are defined. Consequently, the solutions of the kinetic equation (4) must formally satisfy the local conservation laws,

$$\begin{aligned} &\partial_t \langle F\rangle + \boldsymbol{\nabla}_x \cdot \langle v\, F\rangle = 0, \\ &\partial_t \langle v\, F\rangle + \boldsymbol{\nabla}_x \cdot \langle v v^T F\rangle = 0, \\ &\partial_t \langle \tfrac{1}{2}|v|^2 F\rangle + \boldsymbol{\nabla}_x \cdot \langle \tfrac{1}{2}|v|^2 v F\rangle = 0, \end{aligned} \tag{6}$$

corresponding to mass, momentum, and energy conservation, respectively.

Fluid dynamical descriptions of the gas are those in which the phase-space density F is approximated in terms of the fluid variables ρ, u, and θ, the evolution of which is then governed by the local conservation laws (6). The spatial mass, momentum, and energy densities of (6) are given in terms of the fluid variables by the defining relations (3) while the corresponding fluxes can be expressed as

$$\langle v v^T F\rangle = \rho u u^T + \rho\theta(I + \bar{A}), \tag{7a}$$

$$\langle \tfrac{1}{2}|v|^2 v F\rangle = \tfrac{1}{2}\rho|u|^2 u + \tfrac{5}{2}\rho\theta(I+\bar{A})u + \rho\theta^{3/2}\bar{B}, \tag{7b}$$

where the nondimensional stress $\bar{A}$ and the nondimensional heat flux $\bar{B}$ take on values of a traceless symmetric matrix in $\mathbb{R}^{3\times 3}$ and a vector in $\mathbb{R}^3$, respectively. A fluid dynamical closure can be specified by evaluating $\bar{A}$ and $\bar{B}$ for an approximate expression for F given in terms of ρ, u, and θ and their derivatives.

The simplest such closure makes the approximation that F has the form of the unique local Maxwellian density,

$$\mathcal{E}(\rho,u,\theta) = \frac{\rho}{(2\pi\theta)^{3/2}} \exp\left(-\frac{|v-u|^2}{2\theta}\right), \tag{8}$$

for $(\rho(t,x), u(t,x), \theta(t,x))^T \in \mathbb{R}_+ \times \mathbb{R}^3 \times \mathbb{R}_+$; this is consistent with (3). The nondimensional stress $\bar{A}$ and the nondimensional heat flux $\bar{B}$ defined in (7) vanish identically when $F = \mathcal{E}$, thereby reducing (6) to the compressible Euler equations. Because many flows have significant stress and heat flux, the Euler solutions are not always accurate.

More detailed fluid dynamic systems such as the Navier–Stokes equations are obtained by considering approximate distribution functions F which are perturbations of the local Maxwellian $\mathcal{E}$. Such approximations can be derived systematically by the Chapman–Enskog expansion, which will be sketched in Sec. II B. The proximity of F to $\mathcal{E}$ suggests that $\bar{A}$ and $\bar{B}$ may be expressed in terms of F as

$$\bar{A} = \frac{1}{\rho}\langle A F\rangle, \quad \bar{B} = \frac{1}{\rho}\langle B F\rangle, \tag{9}$$

where the nondimensional traceless symmetric 3×3-matrix A and the nondimensional 3-vector B are given by

$$A = \frac{(v-u)(v-u)^T}{\theta} - \frac{1}{3}\frac{|v-u|^2}{\theta} I, \tag{10a}$$

$$B = \left(\frac{|v-u|^2}{2\theta} - \frac{5}{2}\right)\frac{v-u}{\theta^{1/2}}. \tag{10b}$$

The quantities $\bar{A}$ and $\bar{B}$ may therefore be thought of as the average of A and B with respect to F. The entries of A and B are second and third degree Hermite polynomials over $\mathbb{R}^3$ that are orthogonal to the span of 1, v, and $|v|^2$ with respect to the unit weight $\mathcal{E}/\rho$. As we will see in Sec. II B, A and B also arise naturally through the Chapman–Enskog expansion.

While the density F must be non-negative, approximations to it need not be. When such an approximation is used for a fluid dynamical closure, care must be taken to insure that unphysical errors are not introduced. For example, from (9) one can consider the identity,

$$\frac{1}{\rho\theta}\langle (v-u)(v-u)^T F\rangle = I + \bar{A}\,. \tag{11}$$

The expression on the left is manifestly a non-negative definite matrix for any non-negative F. If an approximation for F is used that yields an expression on the right that is not non-negative definite then the approximation has become so bad that there is no distribution of particles that is consistent with the predicted values of the momentum flux computed from the fluid variables. The matrix $I+\bar{A}$ is a simple example of a so-called *moment realizability* criterion for fluid dynamical closures. If the approximation to F is a perturbation of the local equilibrium $\mathcal{E}$, the $I+\bar{A}$ may be viewed as a perturbation of the equilibrium value, which is the identity matrix. Moment realizability means that the perturbation is not so large that the matrix $I+\bar{A}$ is no longer non-negative definite.

More general moment realizability criteria may be derived from the elementary observation that if F is non-negative then so is the quantity $\langle \phi^2 F\rangle$ for any function $\phi = \phi(v)$. This observation is applied to the case when ϕ is an arbitrary polynomial spanned by a given set of polynomials in the components of v. More precisely, if $\mathbf{m} = \mathbf{m}(v)$ denotes a column vector of M given polynomials then applying our observation to $\phi = \mathbf{a}^T\mathbf{m}$ for an arbitrary $\mathbf{a} \in \mathbb{R}^M$ shows that

$$\mathbf{a}^T\langle \mathbf{m}\mathbf{m}^T F\rangle \mathbf{a} = \langle (\mathbf{a}^T\mathbf{m})^2 F\rangle \geqslant 0, \quad \text{for every } \mathbf{a} \in \mathbb{R}^M. \tag{12}$$

We conclude from this that if F is non-negative and not identically zero almost everywhere (so that $\rho>0$), then for every $\mathbf{m}$ the $M\times M$ matrix,

$$\mathbf{M}=\frac{1}{\rho}\langle \mathbf{m}\mathbf{m}^T F\rangle \quad \text{is non-negative definite.} \tag{13}$$

More generally, we could consider $\mathbf{m}$ to be the infinite column vector comprised of all monomials in the components of v provided we interpret (13) to mean that (12) holds for every $\mathbf{a}$ with only a finite number of nonzero components.[10]

In the following we will assume that both ρ and θ, defined by (3), are strictly positive. This is the most basic moment realizability criterion; an approximating distribution function for which this is not true is certainly not close to a thermodynamic equilibrium, and the associated fluid dynamic equations would not be valid.

Now consider two matrices $\mathbf{M}$ and $\mathbf{M}'$ that are defined by two vectors $\mathbf{m}$ and $\mathbf{m}'$ such that the polynomial elements in v of each vector span the same space. It follows that $\mathbf{M}$ may be transformed into $\mathbf{M}'$ by a change of basis operation, and thus both matrices will be non-negative definite over the same region of moment space. Therefore they give the same moment realizability criterion. Moreover, it is also possible to transform $\mathbf{M}$ into a block diagonal matrix without changing the criterion. The question of non-negative definiteness is then reduced to that of the block diagonal elements.

The analysis may be simplified by selecting the polynomial elements of $\mathbf{m}$ to be nondimensional and orthonormal with respect to the normalized Maxwellian equilibrium distribution $\mathcal{E}/\rho$. It is readily checked that applying (13) to the vector

$$\mathbf{m}=\left(1,\frac{v-u}{\theta^{1/2}}\right)^T$$

leads to the criterion that the matrix $I+\bar{A}$ be non-negative definite. As discussed above, the same result is obtained if $\mathbf{m}$ is taken to be the monomial vector $\mathbf{m}=(1,v)^T$.

The criterion we advocate is based on a matrix obtained by applying (13) to the vector,

$$\mathbf{m}=\left(1,\frac{v-u}{\theta^{1/2}},\left(\frac{2}{3}\right)^{1/2}\left(\frac{|v-u|^2}{2\theta}-\frac{3}{2}\right)\right)^T. \tag{14}$$

Again, the elements have been chosen to be orthonormal to $\mathcal{E}/\rho$. This choice of $\mathbf{m}$ is equivalent to using the monomial vector $\mathbf{m}=(1,v,|v|^2)$. This leads to the smallest matrix of the form (13) which contains all the densities and fluxes of the local conservation laws. The resulting criterion is therefore a natural candidate for assessing the validity of approximate distribution functions. For the choice of $\mathbf{m}$ in (14) the moment realizability criterion is based on the 5×5 matrix,

$$\mathbf{M}=\frac{1}{\rho}\begin{pmatrix} \langle F\rangle & 0 & 0 \\ 0 & \left\langle \frac{(v-u)(v-u)^T}{\theta}F\right\rangle & \left\langle \left(\frac{2}{3}\right)^{1/2}\left(\frac{|v-u|^2}{2\theta}-\frac{3}{2}\right)\frac{v-u}{\theta^{1/2}}F\right\rangle \\ 0 & \left\langle \left(\frac{2}{3}\right)^{1/2}\left(\frac{|v-u|^2}{2\theta}-\frac{3}{2}\right)\frac{v-u}{\theta^{1/2}}F\right\rangle^T & \left\langle \frac{2}{3}\left(\frac{|v-u|^2}{2\theta}-\frac{3}{2}\right)^2 F\right\rangle \end{pmatrix}.$$

It follows from (3) and (9) that the matrix $\mathbf{M}$ may be expressed as

$$\mathbf{M}=\begin{pmatrix} 1 & 0 & 0 \\ 0 & I+\bar{A} & (\frac{2}{3})^{1/2}\bar{B} \\ 0 & (\frac{2}{3})^{1/2}\bar{B}^T & \bar{C} \end{pmatrix}. \tag{15}$$

Here $\bar{C}$ is the nondimensional scalar given by

$$\bar{C}=\frac{1}{\rho}\langle CF\rangle, \quad C=\frac{2}{3}\left(\frac{|v-u|^2}{2\theta}-\frac{3}{2}\right)^2. \tag{16}$$

Upon assuming that $\bar{C}>0$ and defining the matrix

$$\mathbf{Q}=\begin{pmatrix} 1 & 0 & 0 \\ 0 & I & 0 \\ 0 & -\left(\frac{2}{3}\right)^{1/2}\frac{\bar{B}^T}{\bar{C}} & 1 \end{pmatrix}, \tag{17}$$

it follows that

$$\mathbf{Q}^T\mathbf{M}\mathbf{Q}=\begin{pmatrix} 1 & 0 & 0 \\ 0 & I+\bar{A}-\frac{2\bar{B}\bar{B}^T}{3\bar{C}} & 0 \\ 0 & 0 & \bar{C} \end{pmatrix}. \tag{18}$$

Hence, when $\bar{C}>0$ the non-negative definiteness of $\mathbf{M}$ is equivalent to the non-negative definiteness of the nondimensional 3×3 matrix,

$$V=I+\bar{A}-\frac{2\bar{B}\bar{B}^T}{3\bar{C}}. \tag{19}$$

As mentioned above all choices of the vector $\mathbf{m}(v)$ whose elements span the space of polynomials in v correspond to equivalent moment realizability criteria defined by the non-negative definiteness of the matrix (13). In fact, any operation which does not change the sign of the eigenvalues of the matrix (13) may be performed without changing the criterion. This includes the change of basis operations such

as (18), as well as taking odd powers of the symmetric matrix, etc. For determining non-negative definiteness, we advocate the use of (19) as the simplest matrix which includes all the densities and fluxes of the conservation laws.

More general criteria for assessing the validity of fluid dynamic equations are suggested by the form of the matrix **M** (15). Because the vector **m** in (14) was chosen to have orthonormal elements with respect to integration against the Maxwellian equilibrium $\mathcal{E}/\rho$, it follows that at equilibrium the matrix **M** must be the identity matrix. For fluid equations derived from a perturbative expansion around the Maxwellian, we may view **M** as a perturbation of the identity matrix. Large deviations of **M** from I indicate large deviations from equilibrium. The moment realizability criterion described above states that the fluid dynamics description has broken down when the perturbation is so large that **M** is no longer non-negative definite.

The generalization of moment realizability is to choose **m** as in (14) and to monitor the deviation of the eigenvalues of **M** from their equilibrium value 1. When the deviation becomes ''too large,'' the fluid description is said to have broken down. What constitutes too large may be problem and user dependent. Thus the size of acceptable deviations of the eigenvalues should be regarded as a tuning parameter.

We may also regard the matrix V in (19) as a perturbation of the identity. Because the matrix **Q** in (17) is not orthonormal, the matrices V and **M** will not share the same eigenvalues. Thus assessing breakdown with V requires a different tuning than **M**. Because it is smaller and somewhat simpler, we advocate using the matrix V of (19) for the generalized criterion. The investigation of the planar shock profile in Sec. III is based on this matrix.

Finally we remark that a generalized criterion may be based on the deviations of the matrix **M** in (13) from its equilibrium value for any choice of the vector **m**. However, each choice would require a separate tuning strategy, in contrast to the original non-negative definiteness criterion, which was equivalent for all **m** with the same span in v. In order for the eigenvalues to have a physical interpretation, it is necessary that the elements of **m** be nondimensional. Also, it is most natural to choose **m** so that v and u appear only as the combination $v-u$. This will guarantee that the matrix **M** is Galilean invariant.

B. The Navier–Stokes approximation

Small deviations of F from its local equilibrium (8) can be computed systematically using a Chapman–Enskog analysis of the scaled kinetic equation,

$$\partial_t F + v \cdot \nabla_x F = \frac{1}{\epsilon}\mathcal{C}(F),$$

where ϵ is a formal nondimensional parameter that gives the order of the mean free path over the macroscopic gradient length scales.[1,11] The Chapman–Enskog expansion of F in powers of ϵ has the form

$$F = \mathcal{E}(\rho,u,\theta)(1+\epsilon\, g^{(1)} + \epsilon^2 g^{(2)} + \cdots),$$

where the leading order is the local equilibrium given by (8). This expansion may be carried out in a fairly general setting provided the collision operator $\mathcal{C}$ and its linearization about $\mathcal{E}$ satisfy the assumptions given in the appendix. The result of doing so to first order in ϵ is that the distribution function is approximated by

$$F_{NS} = \mathcal{E}\left(1 - \frac{1}{2}\mathcal{L}_{\mathcal{E}}^{-1}A : \left[\nabla_x u + (\nabla_x u)^T - \frac{2}{3}\nabla_x \cdot u I\right] - \frac{1}{\theta^{1/2}}\mathcal{L}_{\mathcal{E}}^{-1}B \cdot \nabla_x \theta\right). \tag{20}$$

Here the formal parameter ϵ has been dropped, the matrix-valued function A and vector-valued function B are those defined in (10), and the operator $\mathcal{L}_{\mathcal{E}}^{-1}$ is the pseudo-inverse of the linearized collision operator $\mathcal{L}_{\mathcal{E}}$ discussed in (A8)–(A13) of the Appendix. This is known as the Navier–Stokes closure because when (20) is substituted into (9), the result is

$$\bar{A}_{NS} = \frac{1}{\rho}\langle A F_{NS}\rangle = -\frac{\mu}{\rho\theta}\left[\nabla_x u + (\nabla_x u)^T - \frac{2}{3}\nabla_x \cdot u I\right], \tag{21a}$$

$$\bar{B}_{NS} = \frac{1}{\rho}\langle B F_{NS}\rangle = -\frac{\kappa}{\rho\theta^{3/2}}\nabla_x \theta. \tag{21b}$$

Here the viscosity $\mu = \mu(\rho,\theta)$ and the heat conduction $\kappa = \kappa(\rho,\theta)$ are given by[11]

$$\mu = \tfrac{1}{10}\theta\langle A : \mathcal{E}\mathcal{L}_{\mathcal{E}}^{-1}A\rangle, \quad \kappa = \tfrac{1}{3}\theta\langle B \cdot \mathcal{E}\mathcal{L}_{\mathcal{E}}^{-1}B\rangle.$$

The positivity and boundedness of μ and κ are consequences of the positivity and boundedness of $\mathcal{L}_{\mathcal{E}}^{-1}$. The functional independence of μ and κ on u is a consequence of Galilean invariance. Whenever $\mathcal{C}$ is homogeneous of degree two in F (like for the classical Boltzmann collision operator) then both $\mathcal{L}_{\mathcal{E}}$ and $\mathcal{E}$ will be homogeneous of degree one in ρ, whereby μ and κ will be independent of ρ as well.

The moment realizability criterion matrix (19) may now be evaluated for the Navier–Stokes distribution function (20). $\bar{A}_{NS}$ and $\bar{B}_{NS}$ are given by (21), so it remains only to determine $\bar{C}_{NS}$. This is evaluated as follows. Because the function $C(v)$ defined in (16) is a function only of the magnitude of $v-u$, it is invariant under the orthogonal transformations defined in (A11). Making use of the symmetry of $\mathcal{L}_{\mathcal{E}}^{-1}$ under these orthogonal transformations (A13), we see that the matrix $\langle C\mathcal{E}\mathcal{L}_{\mathcal{E}}^{-1}A\rangle$ and the vector $\langle C\mathcal{E}\mathcal{L}_{\mathcal{E}}^{-1}B\rangle$ are invariant under the orthogonal transformations. For the matrix this means it must be a multiple of the identity, while for the vector this means it must be identically zero. However, because the matrix A is traceless, then the matrix $\langle C\mathcal{E}\mathcal{L}_{\mathcal{E}}^{-1}A\rangle$ must be so too. It follows that

$$\langle C\mathcal{E}\mathcal{L}_{\mathcal{E}}^{-1}A\rangle = 0, \quad \langle C\mathcal{E}\mathcal{L}_{\mathcal{E}}^{-1}B\rangle = 0.$$

This may be used to evaluate (16) for the Navier–Stokes closure (20). The result is

$$\bar{C}_{NS} = \frac{1}{\rho}\langle C F_{NS}\rangle = \frac{1}{\rho}\langle C\mathcal{E}\rangle = 1. \tag{22}$$

Using $\bar{A}_{NS}$ and $\bar{B}_{NS}$ from (21) and $\bar{C}_{NS}$ from (22) to evaluate (19) gives the Navier–Stokes moment realizability criterion matrix (2).

III. APPLICATION TO PLANAR SHOCKS

A. The shock-structure problem

Because of its geometric simplicity and absence of solid boundaries, we consider the structure of a planar shock layer as a test case for investigating the utility of the moment realizability criterion. The underlying kinetic equation is taken to be the Boltzmann equation for rarefied gas dynamics, and the transport coefficients for the Navier–Stokes approximation are those obtained from the Boltzmann collision operator.[1] While much of the following discussion is specific to the planar shock problem, it should be remembered that any criterion based on matrix (2) is quite portable and may be used to indicate the breakdown of fluid dynamics in any setting. Other cases where breakdown is likely to occur include flows near boundaries and flows computed from unphysical initial data.

An important attribute of the matrix (2) is that it is composed of terms related to both stress and heat flux. This is the natural extension of (11), which only considers stress, and is the simplest moment realizability condition which also includes the heat flux. As we shall see in the analysis of the results below, the most significant fluid dynamic breakdown in the shock layer occurs when the Navier–Stokes equations are no longer able to accurately model $\bar{B}$ and $\bar{C}$, the elements which contribute to the term in the criterion involving heat flux. The criterion also captures a less severe breakdown related to the stress which occurs at a different location in the shock. This illustrates the advantage of this portable criterion. One need not know a priori which effect or combination of effects reflects the dominant deviation from fluid dynamics. For a specific geometry, a detailed study such as we present here for the shock layer may be carried out to fine tune the interpretation of the criterion.

The planar shock is characterized by upstream (subscript U) and downstream (subscript D) equilibria which are connected by a profile along which the fluid densities vary. In the stationary coordinate frame of the shock, the conservation laws dictate that the mass, momentum and energy fluxes, M, P, and ME, must be constant along the profile. In particular, this means that the upstream and the downstream equilibria may be connected through the Rankine-Hugoniot relations

$$\begin{aligned} M &\equiv \rho_U u_U = \rho_D u_D, \\ P &\equiv p_U + \rho_U u_U^2 = p_D + \rho_D u_D^2, \\ E &\equiv e_U + \frac{1}{2}u_U^2 + \frac{p_U}{\rho_U} = e_D + \frac{1}{2}u_D^2 + \frac{p_D}{\rho_D}. \end{aligned} \tag{23}$$

Here $p=\rho\theta$ is the pressure given by the ideal gas equation of state and $e=\frac{3}{2}\theta$ is specific internal energy for a monatomic gas.

A stationary planar shock may be described by two parameters, the Mach number Ma, which scales the dependent variables, and the mean free path ℓ, which scales the independent position variable. The local Mach number is defined at each point in the shock profile by

$$\mathrm{Ma} = \frac{|u|}{\sqrt{\frac{5}{3}\theta}}. \tag{24}$$

This is bounded above by Ma_U, its value at the upstream equilibrium. The local Mach number decreases through the shock and achieves its smallest value at the downstream equilibrium, where it lies between 1 and $1/\sqrt{5}$ (these are the limiting values for an infinitely weak and an infinitely strong shock). With the help of (23), the ratios of the downstream to upstream density, velocity, and temperature may be expressed solely as functions of Ma_U by

$$\begin{aligned} \frac{\rho_D}{\rho_U} &= \frac{4\mathrm{Ma}_U^2}{\mathrm{Ma}_U^2+3} \\ \frac{u_D}{u_U} &= \frac{\mathrm{Ma}_U^2+3}{4\mathrm{Ma}_U^2} \\ \frac{\theta_D}{\theta_U} &= \frac{(\mathrm{Ma}_U^2+3)(5\,\mathrm{Ma}_U^2-1)}{16\mathrm{Ma}_U^2}. \end{aligned} \tag{25}$$

Thus the shock strength is specified by the upstream Mach number, which is typically referred to simply as the Mach number of the shock. For an infinitely weak shock $\mathrm{Ma}_U = 1$; it increases without bound as the shock strength increases.

The mean free path is defined as the average distance a gas molecule travels between collisions. Because the changes through the shock profile occur due to collisions, this is a natural choice for the length scale. For a gas of hard sphere particles, the mean free path may be rigorously defined in terms of the particle diameter. This diameter may in turn be related to the viscosity of the gas, leading to the formula

$$\ell = \frac{16}{5}\,\frac{\mu}{\rho\sqrt{2\pi\theta}}. \tag{26}$$

Because the viscosity of a hard sphere gas is proportional to the square root of the temperature, it can be seen that the mean free path is independent of temperature and is inversely proportional to density. For softer interactions than hard sphere collisions, the mean free path is not rigidly defined. Often (26) is used, so that the collision model is expressed only through the temperature dependence of the viscosity. An alternate approach is to define a mean collision time (related to the mean free path and thermal velocity) by evaluating the integral over velocity space of the loss term of the Boltzmann equation for a Maxwellian distribution. For a given collision cross section (related to the molecular interaction, chosen with an appropriate cut off), a mean free path of the form (26) will result. However, the constant may be different. This is essentially the approach used by Bird[2] to define the mean free path for the Variable Hard Sphere model. He assumes that viscosity satisfies a power law rela-

tion with temperature and shows that the initial constant depends on the temperature exponent. (The constant varies from 16/5 for hard spheres to 2 for Maxwell-type molecules.) In the following, the numerical results will be presented for hard sphere molecules with lengths scaled according to (26) evaluated at the upstream density.

It is commonly accepted that the Navier–Stokes equations model most aspects of weak shock waves accurately. Previous work[2,7,12] has indicated that Navier–Stokes should be valid up to Mach numbers ranging from 1.5 to 2. For strong shocks, the presence of large gradients and the highly nonequilibrium state of the gas suggest that this is a regime where the applicability of the Navier–Stokes equations becomes questionable. Most efforts to check the validity of the Navier–Stokes approximation for the shock structure problem have compared Navier–Stokes profiles to profiles obtained from kinetic equations. A notable exception is the work of Liepmann, Narasimha, and Chahine[7] in which *a priori* estimates of the regimes of failure of the Navier–Stokes equations are made based on heuristic arguments about the size of $\bar{A}$ (for the case of the steady planar shock) in the Navier–Stokes approximation. They argued that the Navier–Stokes solution should remain a good approximation in the high-pressure region of the shock layer up to approximately the point of maximum stress for *all* shock strengths. Bird[13] and Boyd, Chen, and Candler[9] suggest related criteria based on a local Knudsen. The moment realizability criteria we study in this paper are in effect a reflection and extension of the same physical considerations as in Liepmann *et al.*,[7] as any other such characterization would be. The arguments here are, however, made more precise by a rigorous derivation of the criteria. They also lead to criteria applicable to general flows and suggest extensions to other fluid dynamics approximations (such as the Burnett equations or extensions of the Navier–Stokes equations that introduce multiple temperatures[9]).

The analysis of the planar shock layer is presented as follows. First the Navier–Stokes equations are reduced to a system of nondimensional ODEs[6] and solved. We then look at the eigenvalues of the moment realizability criterion matrix (2) in terms of the Knudsen number (a measure of deviation from fluid dynamic equilibrium) to gain insight into where in the shock and at what Mach number the criterion would indicate breakdown. Finally, after a brief description of the Direct Simulation Monte Carlo (DSMC) method used to solve the Boltzmann equation, we compare the computed profiles, nondimensional fluxes, and criterion eigenvalues to establish the nature of the Navier–Stokes breakdown.

B. The Navier–Stokes profile

In the stationary frame the Navier–Stokes equations (1) can be reduced to a pair of autonomous ordinary differential equations for the velocity and temperature. This is done by using the constancy of the mass flux, $M=\rho u$, to eliminate the density ρ in favor of velocity u in the momentum and energy conservation laws. The resulting equations are

$$\frac{4}{3}\mu\frac{du}{dx}=M\frac{1}{u}(u^2+\theta)-P,$$
$$\kappa\frac{d\theta}{dx}=M\left(\frac{3}{2}\theta-\frac{1}{2}u^2+\frac{P}{M}u\right)-ME. \tag{27}$$

Here M, P, and E are the constants appearing in the Rankine–Hugoniot relations (23). For a monatomic gas either composed of hard spheres or of point particles with a repulsive power-law intermolecular potential, kinetic theory calculations show that $\kappa=\frac{15}{4}\mu$, so that the Prandtl number $(=\frac{5}{2}\mu/\kappa)$ is constant at $\frac{2}{3}$. Moreover, it can be shown that the viscosity coefficient μ is independent of density and has a power-law dependence on temperature: $\mu/\mu_U=(\theta/\theta_U)^s$. (The exponent s is 0.5 for hard sphere molecules and 1.0 for Maxwell molecules.) For real monatomic gases the Prandtl number is only *very weakly* dependent on temperature and so we expect our results to be indicative of the situation for real monatomic gases.

We follow the classical procedure of Gilbarg and Paolucci[6] to compute the Navier–Stokes shock structure. The uniform states upstream and downstream of the shock ($x=-\infty$ and $x=+\infty$, respectively) are two equilibrium solutions of the differential equations (27). A linear stability analysis of these equilibria shows that the upstream state is an unstable node and that the downstream state is a saddle point. With that configuration of equilibria, by picking an initial point close to the downstream equilibrium state in the stable eigendirection and evolving that point backwards in space under the differential equations (27), one can arrive arbitrarily close to the upstream equilibrium state. The resulting solution curve gives the shock structure. (A fourth order Runge–Kutta scheme was used for the numerical integration.)

For this one dimensional problem, the matrix V in (2) is diagonal and the question of its non-negative definiteness reduces to that of the non-negativity of

$$V_{11}=1-\frac{4}{3}\frac{\mu}{\rho\theta}\partial_x u-\frac{2}{3}\frac{\kappa^2}{\rho^2\theta^3}(\partial_x\theta)^2,$$
$$V_{22}=V_{33}=1+\frac{2}{3}\frac{\mu}{\rho\theta}\partial_x u. \tag{28}$$

It is worthwhile to consider the criteria (28) in a slightly different light. Here we will introduce the Knudsen number, a nondimensional parameter which is often used as a measure of the degree of rarefaction of a gas (and thus the degree of deviation from equilibrium). The Knudsen number is closely related to the expansion parameter used in the Chapman–Enskog procedure. A frequently cited heuristic is that the Navier–Stokes equations will be valid for flows in which the maximum Knudsen number is in the range 0.01 to 0.1. If (26) is used to relate viscosity to the mean free path, then (28) may be written as

$$V_{11}=1+k_\rho\left(\frac{\Delta\rho}{\rho}\right)\mathrm{Ma}\,\mathrm{Kn}_\rho-k_\theta\left(\frac{\Delta\theta}{\theta}\right)^2(\mathrm{Kn}_\theta)^2,$$
$$V_{22}=1-\frac{1}{2}k_\rho\left(\frac{\Delta\rho}{\rho}\right)\mathrm{Ma}\,\mathrm{Kn}_\rho. \tag{29}$$

Here $\Delta\rho$ and $\Delta\theta$ are the density and temperature jumps across the shock, Kn_ρ and Kn_θ are the local Knudsen numbers defined in terms of length scales associated with the density and temperature gradients as

$$\mathrm{Kn}_\rho=l\frac{1}{\Delta\rho}\frac{d\rho}{dx},\quad \mathrm{Kn}_\theta=l\frac{1}{\Delta\theta}\frac{d\theta}{dx},$$

and k_ρ and k_θ are the constants

$$k_\rho=\sqrt{\frac{250}{432}}\pi\approx 1.35,\quad k_\theta=\frac{1875}{1024}\pi\approx 5.75.$$

The values of k_ρ and k_θ may vary somewhat if a definition of the mean free path other than (26) is chosen for non-hard-sphere molecules. The (global) Knudsen number for a shock is generally taken to be the maximum value of the local Knudsen number. It has been observed that the shock thickness (the length scale associated with the Knudsen number) for stronger shocks has only a weak dependence on the Mach number, indicating that the Knudsen number itself will not vary extensively for strong shocks. Up to a slight difference in the constant, the quantity $k_\rho\mathrm{Ma}(\Delta\rho/\rho)\mathrm{Kn}_\rho$ is the same criterion proposed by Bird.[13] There it was suggested that the Navier–Stokes approximation would be valid when this quantity was less than 0.05.

Based on (29) it is possible to make several predictions concerning the breakdown of moment realizability. First, toward the downstream end, the local Mach number will be small, as will the ratios $\Delta\rho/\rho$ and $\Delta\theta/\theta$. This indicates that with regard to moment realizability, Navier–Stokes should give reasonable results downstream of a certain point in the shock, suggested by Liepmann *et al.*[7] to be the point of maximum stress. Our computational results suggest that for the temperature profile this point is approximately the ''equal area'' point, while for the density profile it is around two mean free paths further downstream. A second observation is that toward the upstream side, the local Mach number will be on the order of Ma_U, while $\Delta\theta/\theta$ will be on the order of Ma_U^2, as can be seen from (25). This indicates that the heat flux term will play the largest role in the breakdown for the shock profile.

C. Comparison with the Boltzmann profile

The Boltzmann shock profile is computed using the Direct Simulation Monte Carlo (DSMC) method developed by Bird.[2] The DSMC method is a particle simulation with alternating convection and collision steps. Particles convect independently over a time step Δt. They are then grouped into cells of size Δx. Pairs of particles within each cell undergo collisions with a probability proportional to their relative velocities (or a power thereof) and Δt.

For the current simulation of the steady planar shock wave, the spatial domain was one dimensional, while the fully three dimensional velocity space was retained. The cell size, which was uniform throughout the domain, was chosen to be $\Delta x=\frac{1}{20}l_U$, where l_U is the upstream mean free path. The time step was then chosen to be $\Delta t=\frac{1}{10}\Delta x/\sqrt{2\theta_U}$. For the Mach numbers considered, the cell size always satisfied $\Delta x\leqslant l/5$ everywhere in the domain, while the fastest particles (those with velocities on the order of $|u_U|+3\sqrt{3\theta_U}$) travel approximately one cell in one time step. These choices are well within the guidelines for the DSMC method.

The collision step was computed using Bird's NTC algorithm with the Variable Hard Sphere collision model. This method uses a Monte Carlo rejection technique which requires an upper bound on the relative velocities of all pairs of particles in a cell. The bound was set at $10\sqrt{\theta_D}$, where θ_D is the downstream (larger) temperature.

The stationary shock profile was computed with a time dependent simulation starting from initial data consisting of the domain divided equally into the two constant states. Initially, the upstream cells contained 200 simulation particles. The initial particle velocities were sampled from a Maxwellian distribution with the density, stream velocity, and temperature of the associated constant state. Incoming particles at both ends were sampled from the corresponding flux weighted Maxwellian distributions. Additionally, a ''shock stabilizing'' scheme was used at the boundary to ensure that neither too few or too many particles left the domain in a given time step. This essentially held the shock in a fixed position, minimizing the smearing out effect associated with a random drift of the shock position. The number of cells used varied with the Mach number (the weakest shock requiring the widest domain) and collision model. The simulation was run until a steady state was reached. The velocity averages were then collected by running the simulation another 15 000 to 30 000 time steps and sampling the cell data every 3^{rd} time step.

We now present the numerical results for the computation of stationary planar shock profiles for Mach numbers 1.2, 2, 3, 4, and 5. Mach 1.2 is a typical weak shock which is often presented in the literature.[2,9] Mach 2 is often considered a borderline value for the validity of Navier–Stokes. Mach 3, 4, and 5 correspond to intermediate strength shocks for which the validity of Navier–Stokes would generally be questioned because F would be expected to differ significantly from that given by (20). Thus it is over this range that we expect to observe the breakdown of the moment realizability criterion.

The results presented here are for a hard sphere gas for which the viscosity and thermal conductivity are proportional to the square root of the temperature. The mean free path of a hard sphere gas is given by (26). The same computations were carried out for Maxwell molecules for which viscosity and thermal conductivity are proportional to the temperature. With regard to the moment realizability criterion and the magnitude of the nondimensional fluxes, the results were essentially identical to the hard sphere case. Because hard spheres and Maxwell molecules represent the two extremes of molecular interaction, this indicates that the following results should be valid for real gases which may be modeled as monatomic, nonreacting particles.

TABLE I. Eigenvalues of validity matrices V_{NS} and V_{BE} within the shock profile at the point of maximum deviation from 1. Note that at a given Mach number, the position of the maximum deviation of each quantity is not necessarily the same.

Mach #	$(V_{11}^{NS})_{\max}$	$(V_{11}^{BE})_{\max}$	$(V_{22}^{NS})_{\max}$	$(V_{22}^{BE})\max$
1.2	1.02	1.02	0.99	0.99
2.0	1.15	1.13	0.89	0.89
3.0	0.57	1.14	0.79	0.80
4.0	−0.61	0.72	0.72	0.74
5.0	−2.35	0.60	0.68	0.71

A summary of the results is shown in Tables I and II. Table I gives the eigenvalues of the validity matrix V_{NS} (computed from the Navier–Stokes solution) and the validity matrix V_{BE} [computed with DSMC according to (19)] at the point of each one's maximum deviation from the equilibrium value 1. It can be seen that for the Navier–Stokes solution by Mach 3 there is significant deviation of V_{11} from 1, while at Mach 4 the non-negativity of V_{11} has been violated.

More insight can be gained by directly comparing the Navier–Stokes solution to the DSMC solution. The maximum relative difference of the densities, fluxes, and eigenvalues is shown in Table II. For $Q=(\rho,\theta,1/\bar{C},V_{11},V_{22})$, this is defined by

$$[Q]_{\max}=\max\left|\frac{Q_{NS}(x)-Q_{BE}(x)}{Q_{BE}(x)}\right|,\quad x\in\mathbb{R}.$$

For $Q=(\bar{A}_{11},\bar{B}_1)$, a slightly different relative measure is used to avoid difficulties when these quantities go to zero at equilibrium. In this case we set

$$[Q]_{\max}=\frac{\max|Q_{NS}(x)-Q_{BE}(x)|}{\max|Q_{BE}(x)|},\quad x\in\mathbb{R}.$$

Because of relatively large statistical fluctuations in the Monte-Carlo results for $\bar{A}_{11}$ and $\bar{B}_1$ at Mach 1.2, the corresponding values of $[\bar{A}_{11}]_{\max}$ and $[\bar{B}_1]_{\max}$ in Table II are rather large. This is of no concern, however, because the values of $\bar{A}_{11}$ and $\bar{B}_1$ themselves are quite small at Mach 1.2 as can be inferred from the corresponding values of V_{11} and V_{22} in Table I.

The tables show that at Mach 2 the deviation of the Navier–Stokes quantities from the Boltzmann quantities is acceptably small. The Navier–Stokes fluid densities ρ and θ are accurate to within 5%, while the deviation of the more sensitive nondimensional heat flux $\bar{B}$ is close to the border of acceptability at around 20%. The most sensitive quantity being measured is $\bar{C}^{-1}$, which is related to a fourth velocity moment. Given that the Navier–Stokes equations have no mechanism built in to model the fluxes of the third moments, it is not surprising that the constant value $\bar{C}_{NS}^{-1}$ is associated with the largest error. In fact, a deviation of only 40% suggests that the true distribution function is still reasonably close to equilibrium. This is confirmed by the fact that the eigenvalues of the validity matrix are fairly close to the equilibrium value of 1.

TABLE II. The maximum relative difference between the Navier–Stokes shock profile and Boltzmann shock profile for densities, fluxes, and eigenvalues of the validity matrix.

Mach #	$[\rho]_{\max}$	$[\theta]_{\max}$	$[\bar{A}]_{\max}$	$[\bar{B}]_{\max}$	$[\bar{C}^{-1}]_{\max}$	$[V_{11}]_{\max}$	$[V_{22}]_{\max}$
1.2	0.01	0.01	0.18	0.19	0.01	0.004	0.002
2.0	0.05	0.04	0.13	0.21	0.40	0.03	0.02
3.0	0.10	0.13	0.22	0.30	1.91	0.36	0.05
4.0	0.12	0.24	0.36	0.47	4.36	1.84	0.11
5.0	0.13	0.38	0.47	0.64	7.59	4.87	0.18

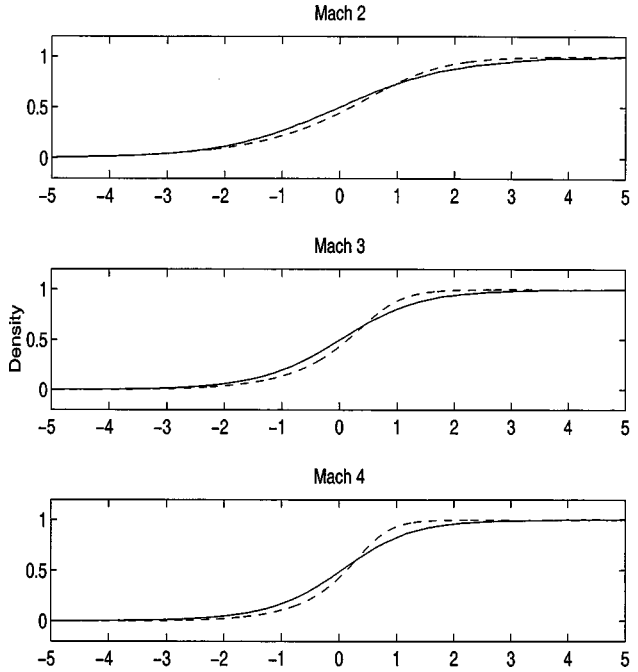

FIG. 1. Scaled density profiles. Solid line: DSMC; dashed line: Navier–Stokes.

By Mach 3, the deviations of the fluid densities are above 10% and the heat flux is off by 30%. This can be correlated to the first eigenvalue of the validity matrix being significantly different from 1. By Mach 4, the moments given by the Navier–Stokes equations are no longer realizable by any non-negative distribution function. This corresponds to the larger deviations of the Navier–Stokes temperature, stress, and heat flux relative to the kinetic solution.

Further information about the breakdown within the shock structure is obtained by examining the shock profiles. Figure 1 compares the scaled density profiles of the Navier–Stokes and Boltzmann solutions for Mach numbers 2, 3, and 4. Figure 2 compares the scaled temperature profile. The scaling is done in the standard way, defining the scaled quantity Q_s for $Q=(\rho,\theta)$ by

$$Q_s=\frac{Q-Q_U}{Q_D-Q_U}.$$

The matrix eigenvalues V_{11} and V_{22} are given in Figs. 3 and 4. Finally, Fig. 5 shows the corresponding nondimensional fluxes. The x-axis is scaled by the upstream mean free path. The value $x=0$ corresponds to the ''equal area'' point of the density profile. This is defined as the point x^* where

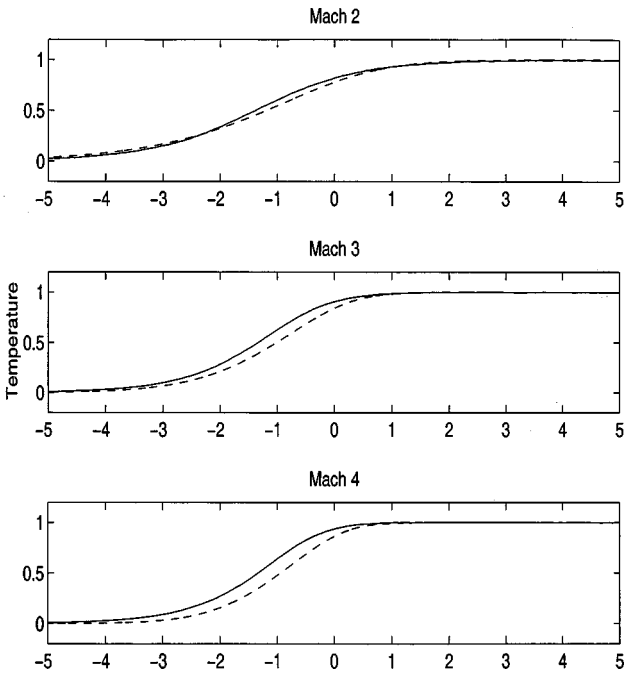

FIG. 2. Scaled temperature profiles. Solid line: DSMC; dashed line: Navier–Stokes.

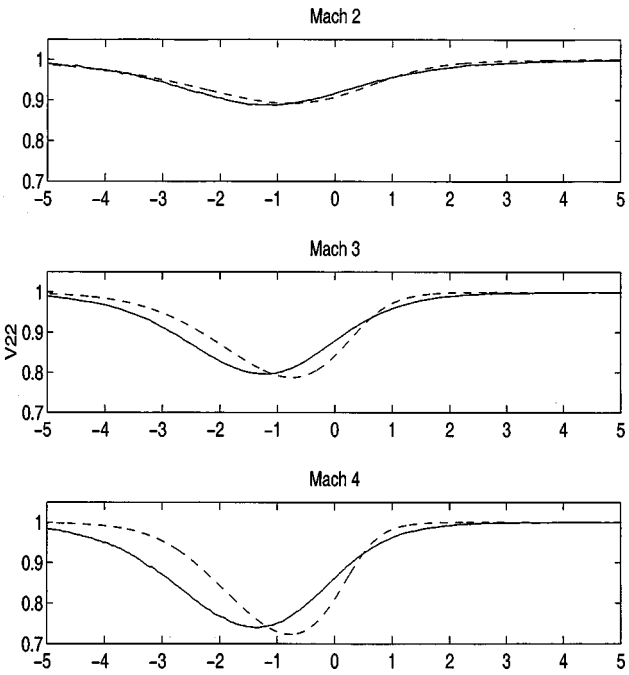

FIG. 4. V_{22} profiles. Solid line: DSMC; dashed line: Navier–Stokes.

$$\int_{-\infty}^{x^*}(\rho(x)-\rho_U)\,dx=\int_{x^*}^{\infty}(\rho_D-\rho(x))\,dx\,.$$

While the positioning of the shock is arbitrary, the choice of the equal area point is natural.

The temperature graphs and the heat flux graphs (B in Fig. 5) support the idea that the breakdown occurs between Mach 2 and 3. At Mach 2, while there are differences, the position of the temperature profile and the shape of the heat flux curve are essentially the same for both Navier–Stokes and DSMC. However, by Mach 3, the DSMC temperature profile is clearly leading the Navier–Stokes profile. Also, the heat flux curve is considerably wider for the DSMC case. The validity matrix eigenvalue V_{11} captures these effects and as a result shows a large deviation from the equilibrium value 1.

Another important observation which can be made from the shock profiles is that for the temperature and flux profiles, the major deviations appear to occur upstream from the

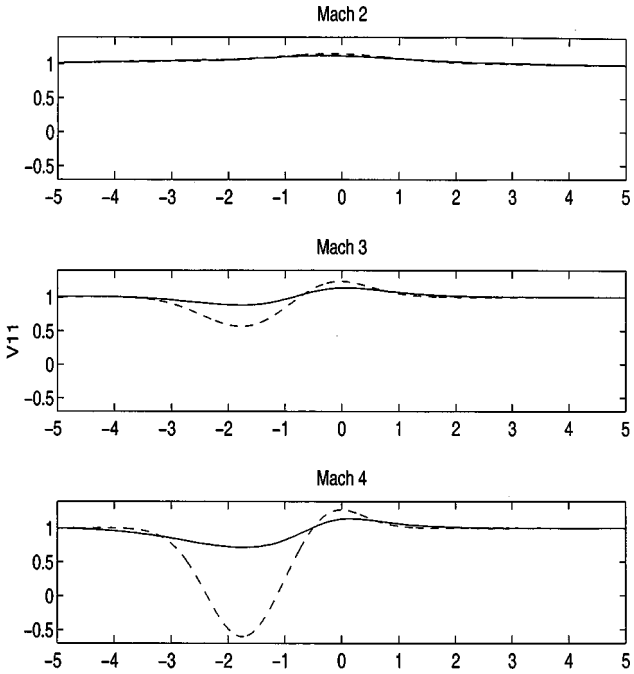

FIG. 3. V_{11} profiles. Solid line: DSMC; dashed line: Navier–Stokes.

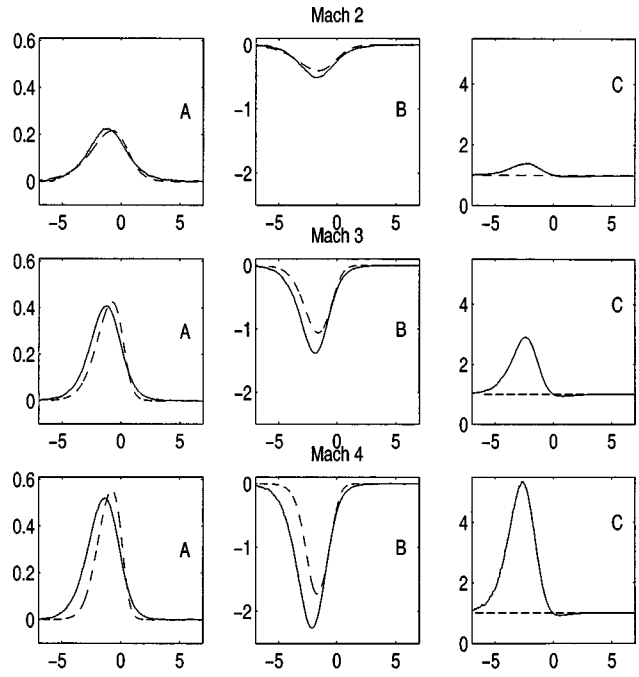

FIG. 5. Nondimensional flux profiles. Solid line: DSMC; dashed line: Navier–Stokes.

equal area point ($x=0$). Particularly for the fluxes $\bar{B}$ and $\bar{C}$, downstream of this point, the DSMC and Navier–Stokes solutions are nearly identical. This is reflected by the eigenvalue V_{11} which reaches its largest deviation from 1 in the upstream region. A second region of deviation, associated primarily with the stress tensor, occurs closer to the equal area point. This effect is signaled by both V_{11} and V_{22}. Again, these plots suggest that the deviations become significant between Mach 2 and 3.

Based on the results presented here for a planar shock, it seems that a reasonable value for the tuning parameter to be used with the eigenvalues of the validity matrix V of (2) would be 25% deviation from the equilibrium value 1. We stress that the choice of this parameter depends on the application and the needs of the user. For other flow geometries, benchmark computations should be run to fine tune the criterion.

IV. CONCLUSIONS

In this paper we have presented a matrix which can be used to assess the validity of the Navier–Stokes approximation for rarefied gas flows. The matrix is derived from moment realizability conditions for the underlying velocity distribution function and is the simplest such matrix which includes terms related to both stress and heat flux.

This matrix, composed of terms intrinsic to the Navier–Stokes equations and given in (2), should be non-negative definite when continuum fluid dynamics are valid. Numerical experiments comparing Navier–Stokes solutions and Direct Simulation Monte Carlo solutions for a stationary planar shock profile indicate that breakdown of the Navier–Stokes solution begins when the eigenvalues of this matrix deviate significantly from unity. When the eigenvalues become negative, corresponding to fluxes which are not realizable by any non-negative velocity distribution function, the predicted Navier–Stokes fluxes deviate by at least 25% from the kinetic equation fluxes. The level of deviation of the Navier–Stokes approximation which is acceptable of course depends on the user and the application. Thus the criterion given here should be viewed as a tool which can be calibrated for use in a specific application.

An advantage of the proposed criterion based on this matrix is that it is portable and may be applied to any Navier–Stokes calculation where breakdown is suspected. This includes not only shock waves, but also boundary layers, initial layers, and other regions of strongly nonequilibrium flow. The inclusion of both stress and heat flux effects is also an important advantage. Previous work has suggested using either one or the other. However, the breakdown of the Navier–Stokes approximation of these fluxes is not necessarily correlated, so that having one criterion which incorporates both effects is of considerable interest.

APPENDIX: PROPERTIES OF THE COLLISION OPERATOR

In this appendix we present a detailed summary of the assumptions on and the properties of the collision operator $\mathcal{C}(F)$ which appears in (4). Particular attention is paid to the linearized collision operator $\mathcal{L}_{\mathcal{E}}$ and its pseudo-inverse $\mathcal{L}_{\mathcal{E}}^{-1}$, which appears in the Navier–Stokes approximation (20) obtained from the Chapman–Enskog expansion. The properties of $\mathcal{L}_{\mathcal{E}}^{-1}$ described here make it possible to evaluate the stress and heat flux for the Navier–Stokes equations, as well as the quantity $\bar{C}$, defined in (16). This allows the moment realizability matrix (19) to be evaluated for the Navier–Stokes closure in terms which are intrinsic to the Navier–Stokes equations, yielding (2).

The collision operator $\mathcal{C}$ is assumed to be defined over a domain $\mathcal{D}(\mathcal{C})$ that is contained within the cone of non-negative functions of v. It will be assumed that $\mathcal{C}$ possesses three properties, discussed below, that relate to dynamical conservation, dissipation, and symmetry, and which are satisfied by a wide variety of classical collision operators.

First, the operator $\mathcal{C}$ is assumed to have 1, v and $|v|^2$ as locally conserved quantities. As stated in (5), this means

$$\begin{aligned} &\langle \mathcal{C}(f)\rangle = 0, \quad \langle v\,\mathcal{C}(f)\rangle = 0, \\ &\langle |v|^2 \mathcal{C}(f)\rangle = 0, \quad \text{for every } f \in \mathcal{D}(\mathcal{C}). \end{aligned} \tag{A1}$$

Moreover, it is assumed that every locally conserved quantity is a linear combination of these three, so that for any $g = g(v)$ the following statements are equivalent:

$$\begin{aligned} &\text{(i)} \quad \langle g\,\mathcal{C}(f)\rangle = 0, \quad \text{for every } f \in \mathcal{D}(\mathcal{C}); \\ &\text{(ii)} \quad g \in \mathbb{E} \equiv \operatorname{span}\{1, v_1, v_2, v_3, |v|^2\}. \end{aligned} \tag{A2}$$

The relations (A1) represent the physical laws of mass, momentum, and energy conservation during collisions while (A2) states that there are no other conservation laws.

Second, the operator $\mathcal{C}$ is assumed to satisfy the local dissipation relation

$$\langle \log f\,\mathcal{C}(f)\rangle \leqslant 0, \quad \text{for every } f \in \mathcal{D}(\mathcal{C}). \tag{A3}$$

The quantity on the left of (A3) is the so-called local entropy dissipation rate. The local equilibria of $\mathcal{C}$ are assumed to be characterized by the vanishing of the local entropy dissipation rate and to be given by the class of Maxwellian densities, i.e., those of the form

$$f = \mathcal{E}(\rho, u, \theta) \equiv \frac{\rho}{(2\pi\theta)^{3/2}} \exp\left(-\frac{|v-u|^2}{2\theta} \right), \tag{A4}$$

for some $(\rho, u, \theta) \in \mathbb{R}_+ \times \mathbb{R}^3 \times \mathbb{R}_+$. More precisely, for every $f \in \mathcal{D}(\mathcal{C})$ the following statements are assumed to be equivalent:

$$\begin{aligned} &\text{(i)} \quad \langle \log f\,\mathcal{C}(f)\rangle = 0, \\ &\text{(ii)} \quad \mathcal{C}(f) = 0, \\ &\text{(iii)} \quad f \text{ is a Maxwellian density given by (A4).} \end{aligned} \tag{A5}$$

These assumptions about $\mathcal{C}$ merely abstract some of the consequences of Boltzmann's celebrated H-theorem.

Third, the operator $\mathcal{C}$ is assumed to commute with translational and orthogonal transformations. Specifically, given any $f = f(v)$ then for every vector $w \in \mathbb{R}^3$ and for every orthogonal matrix $O \in \mathbb{R}^{3\times 3}$ define transformed functions $\mathcal{T}_w f$ and $\mathcal{T}_o f$ by

$$\mathcal{T}_u f = \mathcal{T}_u f(v) \equiv f(v-u), \quad \mathcal{T}_o f = \mathcal{T}_o f(v) \equiv f(O^T v). \quad \text{(A6)}$$

It is assumed that if f is in $\mathcal{D}(\mathcal{C})$ then so are $\mathcal{T}_u f$ and $\mathcal{T}_o f$ with

$$\mathcal{T}_u \mathcal{C}(f) = \mathcal{C}(\mathcal{T}_u f), \quad \mathcal{T}_o \mathcal{C}(f) = \mathcal{C}(\mathcal{T}_o f). \quad \text{(A7)}$$

This relation reflects the Galilean invariance of the microscopic collisional dynamics and implies that the kinetic equation (4) formally retains Galilean invariance.

In addition to the three properties of $\mathcal{C}$ discussed above, it will be assumed that for a given and absolute equilibrium $\mathcal{E}$ (independent of space and time) the linearization of the kinetic equation (4) about $F = \mathcal{E}$ has analogous structural features. Formally setting $F = \mathcal{E}(1+\delta g)$ into (4) and dropping terms of order δ^2 results in the linearization

$$\partial_t g + v \cdot \nabla_x g + \mathcal{L}_\mathcal{E}\, g = 0,$$

where $\mathcal{L}_\mathcal{E}$ is defined to act linearly on $g = g(v)$ by

$$\mathcal{L}_\mathcal{E}\, g \equiv -\frac{1}{\mathcal{E}} D\mathcal{C}(\mathcal{E})\mathcal{E}g \equiv -\frac{1}{\mathcal{E}} \frac{\partial}{\partial \delta} \mathcal{C}(\mathcal{E}(1+\delta g))\bigg|_{\delta=0}. \quad \text{(A8)}$$

The operator $\mathcal{L}_\mathcal{E}$ is assumed to be self-adjoint with domain $\mathcal{D}(\mathcal{L}_\mathcal{E})$ dense in $\mathbb{H}_\mathcal{E}$, the Hilbert space with the weighted inner product $(g|h)_\mathcal{E} = \langle g\, \mathcal{E}\, h\rangle$.

We shall assume that $\mathcal{L}_\mathcal{E}$ possesses properties that are linearized versions of the conservation and dissipation properties (A1)–(A5) of $\mathcal{C}$. Specifically, we shall assume that

$$\langle g\, \mathcal{E}\, \mathcal{L}_\mathcal{E}\, g\rangle \geqslant 0, \quad \text{for every } g \in \mathcal{D}(\mathcal{L}_\mathcal{E}), \quad \text{(A9)}$$

and that for every $g \in \mathcal{D}(\mathcal{L}_\mathcal{E})$ the following statements are assumed to be equivalent:

$$\text{(i)} \quad \langle g\, \mathcal{E}\, \mathcal{L}_\mathcal{E} g\rangle = 0,$$
$$\text{(ii)} \quad \mathcal{L}_\mathcal{E}\, g = 0, \quad \text{(A10)}$$
$$\text{(iii)} \quad g \in \mathbb{E}.$$

Notice that (A9) implies that (i) and (ii) in (A10) are always equivalent, so the only assertion in (A10) is the equivalence of (iii) to the others.

In addition, we shall assume that $\mathcal{L}_\mathcal{E}$ satisfies the Fredholm alternative $\mathcal{R}(\mathcal{L}_\mathcal{E}) = \mathbb{E}^\perp$, where $\mathcal{R}(\mathcal{L}_\mathcal{E})$ denotes the range of $\mathcal{L}_\mathcal{E}$. The Fredholm alternative states that the equation $\mathcal{L}_\mathcal{E}\, g = f$ has a solution if and only if $f \in \mathbb{E}^\perp$, in which case there is a unique solution in $\mathbb{E}^\perp$ denoted by $g = \mathcal{L}_\mathcal{E}^{-1} f$. Let $\mathcal{P}_\mathcal{E}$ be the orthogonal projection of $\mathbb{H}_\mathcal{E}$ onto $\mathbb{E}$, which acts on $g = g(v)$ as

$$\mathcal{P}_\mathcal{E}\, g = \frac{1}{\rho}\Bigg[\langle \mathcal{E} g\rangle + \frac{(v-u)\langle (v-u)\mathcal{E}g\rangle}{\theta} + \left(\frac{|v-u|^2}{2\theta} - \frac{3}{2}\right)\frac{2}{3}\left\langle \left(\frac{|v-u|^2}{2\theta} - \frac{3}{2}\right)\mathcal{E}g \right\rangle\Bigg].$$

The operator $\mathcal{L}_\mathcal{E}^{-1}$ is the unique bounded pseudo-inverse of $\mathcal{L}_\mathcal{E}$ such that

$$\mathcal{L}_\mathcal{E}\mathcal{L}_\mathcal{E}^{-1} = \mathcal{I} - \mathcal{P}_\mathcal{E}, \quad \mathcal{L}_\mathcal{E}^{-1}\mathcal{L}_\mathcal{E} \subset \mathcal{I} - \mathcal{P}_\mathcal{E},$$

where $\mathcal{I} - \mathcal{P}_\mathcal{E}$ is the orthogonal projection of $\mathbb{H}_\mathcal{E}$ onto $\mathbb{E}^\perp$. The dissipation relation (A9) ensures that the quadratic forms associated with $\mathcal{L}_\mathcal{E}$ and $\mathcal{L}_\mathcal{E}^{-1}$ enjoy the positivity properties:

$$\langle g\, \mathcal{E}\, \mathcal{L}_\mathcal{E}\, g\rangle > 0, \quad \text{for every nonzero } g \in \mathbb{E}^\perp \cap \mathcal{D}(\mathcal{L}_\mathcal{E});$$
$$\langle g\, \mathcal{E}\, \mathcal{L}_\mathcal{E}^{-1} g\rangle > 0, \quad \text{for every nonzero } g \in \mathbb{E}^\perp.$$

This means that $\mathcal{L}_\mathcal{E}$ and $\mathcal{L}_\mathcal{E}^{-1}$ are positive definite over $\mathbb{E}^\perp$.

Finally, fix a local equilibrium $\mathcal{E} = \mathcal{E}(\rho, u, \theta)$. Every orthogonal matrix $O \in \mathbb{R}^{3\times 3}$ defines the transformation

$$\mathcal{O}_o \equiv \mathcal{T}_u \mathcal{T}_o \mathcal{T}_u^{-1}, \quad \text{(A11)}$$

where $\mathcal{T}_u$ and $\mathcal{T}_o$ are defined by (A6). It is easily checked that $\mathcal{O}_o \mathcal{E} = \mathcal{E}$ and that $\mathcal{O}_o$ is an orthogonal transformation over $\mathbb{H}_\mathcal{E}$. Linearizing the commutation relations (A7) about $\mathcal{E}$ shows for every orthogonal matrix $O \in \mathbb{R}^{3\times 3}$ that if $g \in \mathcal{D}(\mathcal{L}_\mathcal{E})$ then so is $\mathcal{O}_o g$ and

$$\mathcal{O}_o \mathcal{L}_\mathcal{E}\, g = \mathcal{L}_\mathcal{E}\, \mathcal{O}_o g. \quad \text{(A12)}$$

In other words, the operator $\mathcal{L}_\mathcal{E}$ commutes with the orthogonal transformations $\mathcal{O}_o$. This symmetry is usually applied in one of two ways. First, specializing (A12) to the case $O = -I$ shows that the space of all functions that are even (odd) in $v-u$ is invariant under $\mathcal{L}_\mathcal{E}$; this is the so-called parity symmetry of $\mathcal{L}_\mathcal{E}$. Second, because the measure d_v^3 is invariant under action by any $\mathcal{O}_o$ it follows that

$$\langle g\, \mathcal{E}\, \mathcal{L}_\mathcal{E}\, h\rangle = \langle \mathcal{O}_o(g\, \mathcal{E}\, \mathcal{L}_\mathcal{E}\, h)\rangle = \langle (\mathcal{O}_o g)\mathcal{E}\, \mathcal{L}_\mathcal{E}\, \mathcal{O}_o h\rangle,$$
$$\langle g\, \mathcal{E}\, \mathcal{L}_\mathcal{E}^{-1} h\rangle = \langle \mathcal{O}_o(g\, \mathcal{E}\, \mathcal{L}_\mathcal{E}^{-1} h)\rangle = \langle (\mathcal{O}_o g)\, \mathcal{E}\, \mathcal{L}_\mathcal{E}^{-1} \mathcal{O}_o h\rangle, \quad \text{(A13)}$$

for every orthogonal matrix $O \in \mathbb{R}^{3\times 3}$ and any functions g and h for which the expressions make sense. This symmetry simplifies the evaluation of many tensors that appear in the Chapman–Enskog expansion.

ACKNOWLEDGMENTS

C.D.L. was partially supported by the AFOSR under Grant No. F49620-92-J-0054 at the University of Arizona. W.J.M. was supported by the National Science Foundation (NSF) through a NSF Postdoctoral Fellowship while at the University of Arizona. Some of this work was carried out while C.D.L. was visiting the Center for Nonlinear Studies (CNLS) at the Los Alamos National Laboratory, which is supported in part by the Department of Energy (DOE) under Grant No. W7404-ENG-36, and while C.D.L. and W.J.M. were visiting the Institute for Advanced Study (IAS) under a grant from the Alfred P. Sloan Foundation.

[1] C. Cercignani, *The Boltzmann Equation and Its Applications*, Applied Mathematical Sciences **67** (Springer-Verlag, New York, 1988).
[2] G. A. Bird, *Molecular Gas Dynamics and the Direct Simulation of Gas Flows* (Oxford University Press, Oxford, 1994).
[3] H. Grad, ''The profile of a steady plane shock,'' Commun. Pure Appl. Math. **5**, 257 (1952).
[4] C. D. Levermore, ''Moment closure hierarchies for kinetic theories,'' J. Stat. Phys. **83**, 1021 (1996).
[5] C. D. Levermore and W. J. Morokoff, ''The Gaussian moment closure for gas dynamics,'' in preparation.
[6] D. Gilbarg and D. Paolucci, ''The structure of shock waves in the continuum theory of fluids,'' J. Rat. Mech. Anal. **2**, 617 (1953).

[7]H. W. Liepmann, R. Narasimha, and M. T. Chahine, ''Structure of a plane shock layer,'' Phys. Fluids **5**, 1313 (1962).
[8]G. C. Pham-Van-Diep, D. A. Erwin, and E. P. Muntz, ''Testing continuum descriptions of low-Mach number shock structures,'' J. Fluid Mech. **232**, 403 (1991).
[9]I. Boyd, G. Chen, and G. Candler, ''Predicting failure of the continuum fluid equations in transitional hypersonic flows,'' Phys. Fluids **7**, 210 (1995).
[10]M. Reed and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-Adjointness* (Academic, New York, 1975), pp. 145–146.
[11]C. Bardos, F. Golse, and D. Levermore, ''Fluid dynamic limits of kinetic equations I: Formal derivations,'' J. Stat. Phys. **63**, 323 (1991).
[12]D. A. Erwin, G. C. Pham-Van-Diep, and E. P. Muntz, ''Nonequilibrium gas flows I: A detailed validation of Monte Carlo direct simulation for monatomic gases,'' Phys. Fluids A **3**, 697 (1991).
[13]G. A. Bird, ''Breakdown of translational and rotational equilibrium in gaseous expansion,'' AIAA J. **8**, 1998 (1970).