Maximum Entropy Reconstruction for Gas Dynamics

Thesis by
Dustin Phillip Summy

In Partial Fulfillment of the Requirements for the
degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2017
Defended May 2nd, 2017

Dustin Phillip Summy
ORCID: 0000-0002-6383-0621

ACKNOWLEDGEMENTS

I extend heartfelt thanks to my advisor, Professor Dale Pullin, for his guidance and support through my investigation into what became an interestingly complex problem. It is difficult to imagine having completed this work without his experience and his intuition for finding a path towards progress through a tangle of equations and figures.

I thank the other members of my defense committee, Profs. Austin, Blanquart, and Meiron, for their suggestions and input which have greatly improved the presentation of this work.

I must thank GALCIT's excellent administrators and assistants who have always been critical to navigating the sea of procedures, forms, deadlines, room reservations, and protocols that is a graduate's career at Caltech. Most helpful to me personally have been Barbara, Dimity, and Christine, who are indispensable members of the GALCIT community.

I thank my many good friends and classmates at Caltech, among them Esteban, Teja, Wouter, Akshay, Noah, Namiko, and Marcello who, in addition to being friends of the highest caliber, have helped me directly, in some way or another, with questions or problems related to this work. They were helped in their endeavors to make Caltech a survivable encounter by many others, among them Hilary, Carlos, Thomas, Reeve, Ryan, Jason, Susie, Melissa, and Jonathan. All of their contributions to my life are likely countable but nevertheless infinite.

I thank my grandfather, Jay Boy, for helping to set me on the path towards science and engineering with surprising patience, subscriptions to various magazines, and the occasional blackboard lecture in the shed. I would be remiss not to mention my grandmother, whom we actually call *Mamaw Moo*, whose voice has been a judge's gavel, calling "Order!" on many a needed occasion.

I thank my mother, Andrea, my father, Todd, and my brother, Austin, for their support of every kind and their love which never ends—no matter how frequently I am too busy to return their phone calls. Each day I am thankful to have had such excellent support, high expectations, and constant love from them throughout my life.

I thank my best friend, confidant, co-conspirator, and fiancée, Ananda. There is little that can be put into writing that can express how much I draw upon our relationship

for strength, comfort, and motivation. I will forever be indebted for the way she has supported me when deadlines approached, or when hunger or thirst threatened to tear me from my desk while I wrote this dissertation, or on any given day when I simply needed her beside me.

Finally, while I don't think it entirely appropriate to thank our pets, given how much Ananda would want it and how much she has done for me, the least I can do for her is to say thank you for everything, Hubble and The Cheat.

ABSTRACT

We present a method for selecting a unique and natural probability distribution function (PDF) which satisfies a given number of known moments and apply it for use in the closure of moment-based schemes for approximately solving the Boltzmann equation in gas dynamics.

The method used for determining the PDF is the Maximum Entropy Reconstruction (MER) procedure, which determines the PDF with maximum entropy which satisfies a given set of constraining moments. For the five-moment truncated Hamburger moment problem in one dimension, the MER takes the form of the exponential of a quartic polynomial. This implies a bimodal structure which gives rise to a small-amplitude packet of PDF-density sitting quite far from the mean. This is referred to as the Itinerant Moment Packet (IMP). It is shown by asymptotic analysis that the IMP gives rise to a solution that, in the space of constraining moments, is singular along a line emanating from, but not including, the point representing thermodynamic equilibrium. We use this analysis of the IMP to develop a numerical regularization of the MER, creating a procedure we call the Hybrid MER (HMER). Compared with the MER, the HMER is a significant improvement in terms of robustness and efficiency while preserving accuracy in its prediction of other important distribution features, such as higher order moments.

We apply the one-dimensional HMER to close a fourth order moment system derived from the Boltzmann equation by using a specific set of moment constraints which allow the full, three-dimensional velocity PDF to be treated as a product of three independent, one-dimensional PDFs. From this system, we extract solutions to the problem of spatially homogeneous relaxation and find excellent agreement with a standard method of solution. We further apply this method to the problem of computing the profile within a normal shock wave, and find that solutions exist only within a finite shock Mach number interval. We examine the structure of this solution and find that it has interesting behavior connected to the singularity of the MER and the IMP. Comparison is made to standard solution methods. It is determined that the use of the MER in gas dynamics remains uncertain and possible avenues for further progress are discussed.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

Number *Page*

Chapter 1

INTRODUCTION

1.1 Gas Dynamics Away from Equilibrium

The dynamics of fluid flows are, for situations of common experience, well-modeled by the Navier-Stokes equations or the Euler equations [4]. Even many complex engineering flow problems, such as those encountered in commercial aviation [3], combustion engine design [32], and weather and climate forecasting [31], can be represented, at least in principle, by recourse to these equations, augmented with necessary features to handle chemistry, turbulence, and computational cost. However, both simple and complex flows well-described by such physical models deal necessarily with mild gradients of flow properties (*e.g.*, velocity and temperature) relative to the scales associated with the microscopic behavior of the gas [6]. There exist many types of flows in which this is no longer the case and recourse must be made to a more fundamental description of a gas in order to properly account for its behavior.

When are flow gradients considered "large" with respect to microscopic scales in the gas? The best measure of this is given by the Knudsen number,

$$Kn \equiv \frac{\lambda}{L}, \tag{1.1}$$

where λ is the mean free path of a gas molecule between collisions and L is a representative length scale of the flow. In the limit $Kn \to 0$, particles undergo many collisions before moving any appreciable distance in the flow, and so the flow takes on its equilibrium state at every point, which corresponds to the Euler equations. For small but finite Kn, the Navier Stokes equations are appropriate. When $Kn \gtrapprox 0.1$, as a rough rule, the flow will be significantly affected by its underlying molecular nature, and any accurate model must account for this [6]. Such flows can arise due to gas rarefaction, as in reentry flows [34], or due to flow scale reduction, whether by physical problem size as in microflows [17] or by gradient steepening as in the interiors of shock waves [38] and hypersonic boundary layers [48]. In cases of extremely large Kn, the flow's behavior is completely dominated by the molecular model, and particle or statistical methods are best used to capture this [13].

1.2 Equilibrium and Non-Equilibrium

The basic problem with the extension of the Navier-Stokes equations into the mid-high Kn regime is that fluxes in the Navier-Stokes equations are modeled as functions of state variables in a way that is consistent with near-equilibrium gas behavior but, in this regime, the gas may persist for a non-trivial fraction of the domain of interest in a significantly non-equilibrium state. Equilibrium is characterized by a distribution of molecular velocities corresponding to the famous Maxwell-Boltzmann distribution, whose probability distribution function (PDF), $f_E(c)$, is given by

$$f_E(c) = \frac{4\pi c^2}{(2\pi RT)^{3/2}} e^{\frac{-c^2}{2RT}}, \tag{1.2}$$

where c is the peculiar speed of a molecule, given by $c = |\boldsymbol{\xi} - \boldsymbol{u}|$, with $\boldsymbol{\xi}$ the velocity and $\boldsymbol{u}$ the bulk or mean velocity of a molecule in this distribution, R is the specific gas constant, given by $R = k/m$ with k the Boltzmann constant and m the mass of a molecule, and T is the equilibrium thermodynamic temperature [6]. Only the speed distribution of the molecules is given, as the direction is uniformly isotropic. It can be shown, as in [6] and [12], that the use of (1.2) as a model for the velocity distribution function everywhere in a gas leads to the Euler equations, while using a slight perturbation of (1.2) known as the Chapman-Enskog distribution can be seen to be consistent with the Navier-Stokes equations. This explains both the success of these near-equilibrium methods as well as their failure in non-equilibrium settings where these distributions are not good approximations of the true velocity distribution. Important controllers of flow evolution, such as energy and momentum fluxes, are miscalculated by these approaches due to a limited representation of the possible flow configurations [37].

The Mott-Smith Model

A good link between the kinetic models which will be the subject of this work and the near-equilibrium models useful in low-Kn regimes is the model of Mott-Smith [45] for molecular velocity distribution functions in the interior of shock waves. Mott-Smith's idea was to use as a model distribution a summed pair of Maxwellian distributions corresponding to the upstream and downstream equilibrium distributions, unequally weighted by a parameter which varied through the shock wave such that at the upstream and downstream points only the corresponding distribution function was used. Notionally, the idea is a linear interpolation between two equilibria. This results in strongly bimodal distributions in the interiors of strong

shock waves which do a better job of approximating the actual distribution than the Navier-Stokes [6].

1.3 The Boltzmann Equation

While an improvement over Navier-Stokes in the specific case of the shock wave example, Mott-Smith's model is not very general since it relies on *a priori* knowledge of the boundary states in a particular flow. General methods for gas dynamics problems in the mid-high-Kn regime must rely on the Boltzmann equation, discussed at some length in [6], [12], and [15] and introduced in the next chapter. The Boltzmann equation describes the evolution of the velocity distribution function in space, time, and in terms of the molecular velocity (often called velocity-space); this makes it a seven-dimensional, partial-differential equation. It accounts for molecular movement by velocity, and for intermolecular interference via isolated, binary collision events. The restriction to binary collisions is an assumption, valid when the gas is dilute, in the sense that intermolecular spacing is large compared with the interaction range. In cases where this assumption does not hold, ternary and even more complex collisions must be considered (as they are in dense gas kinetic theory, see [12]). The effect of collisions on the distribution is a result of the consideration of collisions between particles of one velocity with particles of every other velocity in the distribution, in every possible collision geometry, and thus takes the form of an integral, known as the collision integral, making the Boltzmann equation an integro-differential equation.

The high dimensionality and presence of the collision integral, which is itself of high dimension, in the Boltzmann equation make it very computationally expensive, so much so that little practical calculation has been made using it in full form, though useful theoretical results can be derived from it, including the aforementioned derivation of the Navier-Stokes and Euler equations from kinetic theory [6] [12], boundary conditions near a wall in rarefied flows which demonstrate the existence of a slip velocity at the wall as opposed to the no-slip condition standard in near-equilibrium flows [21], and results on relaxation to equilibrium from an initially disturbed state [33]. Computational results are limited to the simplest cases, such as the shock structure problem, which has its dimensionality reduced considerably due to symmetry, [46], [40].

The two main methods by which progress is made towards performing practical numerical calculations for problems where the Boltzmann equation is required are

moment methods and direct simulation Monte Carlo (DSMC).

1.4 Moment Methods

Moment methods are closer in spirit to the familiar Navier-Stokes equations, which can actually be derived as a special case of a moment method [6], [15]. The basic idea behind moment methods, described in detail in Chapter 5, is to use the Boltzmann equation to derive conservation-like equations for various moments of the velocity distribution function. It will be seen that when doing this it is impossible to form a closed system of equations for any set of chosen moments, since there will always be a dependence on higher order moments to compute the fluxes, as well as a dependence on the full velocity distribution function which appears in the collision integral. The problem is to develop a scheme by which to reconstruct an approximate velocity PDF which can be used to compute the unclosed terms. An example can be found in [19], wherein calculated moments define a velocity PDF construction using Hermite polynomials with a Maxwell-Boltzmann distribution as a weight, which can then be used to compute necessary closing terms. The moment approach seems straightforward, but usually gives rise to artificial features in the shock structure solution for high Mach numbers due to limited characteristic speeds captured by the closure [55]. Recently, some interest has developed in the use of more natural closure schemes, such as the Principle of Maximum Entropy [24], [25]. Examples of applications to kinetic theory include [35], [36], [41], and [42]. McDonald developed approximations to the Maximum Entropy closure which were functional, but these schemes have been either impractical or intentionally more approximate than would seem necessary. In this work we will show how difficult features of the true form of this closure scheme and its interaction with the moment equations likely explain why this is the case.

1.5 Direct Simulation Monte Carlo

The dominant approach to solving the Boltzmann equation is known as direct simulation Monte Carlo [8]. The basic method is relatively straightforward, as laid out in [6]. In this scheme, information about the solution is carried in the form of a small number of simulated molecules which represent a huge number of real molecules. The simulated molecules can move and collide with other simulated molecules, and averages are taken over time or over multiple instances of a simulation to compute mean values of important flow properties. The procedure involves generating random initial conditions for the particles according to initial

and boundary conditions of the problem at hand and then accurately simulating the physics for each particle thereafter, with randomness again playing a role in the computation of the effect of collisions where particle collision geometry parameters (mainly the scattering angle and transfers of energy between kinetic and internal modes) are randomized. When implemented carefully, this approach can achieve high accuracy in terms of the actual averages of flow properties and their comparison with solutions of the Boltzmann equation [54], although computing them accurately requires many samples in order to accumulate a high signal-to-noise ratio in the computed results [30]. This results in calculations that are inherently more expensive than equivalently-sophisticated Navier-Stokes methods when they are also valid, although DSMC of course is capable of working accurately throughout the Knudsen number range, even if it would be prohibitively expensive, due to the huge numbers of collisions required, to perform calculations deep into the low-Kn regime where Navier-Stokes is accurate [47].

DSMC is extremely robust, owing to its explicitly physical nature which makes it inherently stable; a solution will be produced regardless of the satisfaction of the usual constraints like the CFL number [14]. DSMC does, however, require that certain convergence criteria be met. Specifically, collision cells (used for sorting particles into local groups for collision selection) should be significantly smaller than the mean free path, and cells must be populated with a sufficient number of molecules, usually around ten or fewer per cell, depending on the specific application [7], [47]. One stumbling block for DSMC appears to be the correct handling of angular momentum in collisions [44]. Since particles are chosen from a range of positions within a collision cell, collisions can transfer angular momentum non-conservatively, even as linear momentum and energy are conserved. With care, it seems, this can be mitigated somewhat, at the cost of a more elaborate collision scheme [53].

1.6 Objectives

In the the present work, our objectives are focused on moment methods and specifically the Maximum Entropy closure. We will examine the Maximum Entropy Principle and develop from it a suitably robust scheme for reconstructing a probability distribution function from a small number of moments. We will see in the next chapter that the Boltzmann H-theorem suggests a link between such Maximum Entropy Reconstructions and gas dynamics, and we will proceed to develop a moment method closure based on our Maximum Entropy Reconstruction. Finally, we

will analyze the usefulness of the closure by applying it to canonical problems of gas dynamics and examining features of the solution.

Chapter 2

KINETIC THEORY

2.1 The Boltzmann Equation

The Boltzmann Equation is both a fundamental result and starting point in gas-kinetic theory. For the case of a single-species monatomic gas with no body force, Bird [6] gives its form as

$$\frac{\partial(\rho f)}{\partial t} + \xi_i \frac{\partial(\rho f)}{\partial x_i} = \frac{\rho^2}{m} \int_\Omega \int_0^\pi \int_0^{2\pi} \left(f_1' f' - f_1 f \right) \sigma c_r \, \sin(\chi) d\epsilon d\chi \boldsymbol{d\xi_1}, \quad (2.1)$$

Where the integral of the velocity distribution function, $f(\boldsymbol{\xi}, \boldsymbol{x}, t)$, is normalized to unity over the space Ω of all atomic velocities, $\boldsymbol{\xi}$. $\rho(\boldsymbol{x}, t)$ is the mass density at position $\boldsymbol{x}$ and time t. $\sigma(\boldsymbol{c_r}, \epsilon, \chi)$ is the cross section for a collision between particles with relative velocity $\boldsymbol{c_r} = \boldsymbol{\xi_1} - \boldsymbol{\xi}$ and a given deflection of the relative velocity of particles at impact given by the spherical angles for azimuth, ϵ, and colatitude, χ, relative to the original relative velocity vector. The subscript 1 indicates that a quantity is in reference to its value for a particle with velocity $\boldsymbol{\xi_1}$ and the primed quantities are evaluated with post-collision atomic velocity, $\boldsymbol{\xi'}$ or $\boldsymbol{\xi_1'}$. The details of the calculation of post-collision velocities are determined by the specific molecular interaction model chosen for the problem, with common examples including hard-sphere interaction and power-law force potentials.

Physically, the left hand side of (2.1) represents transport of the velocity distribution by the motion of particles, and the right hand side represents the effect on the distribution of binary collisions between particle pairs. The collision term in this case is specified for the regime of dilute gases where binary collisions dominate and collision pair velocities are uncorrelated, *i.e.* the assumption of "molecular chaos" [6].

Interpretation of the Collision Integral

The left hand side is uncontroversially derived from kinematic and statistical considerations, but the right hand side is more nuanced. It was used by Boltzmann himself to prove his famous H-theorem [10], which was considered by some to be a sort of proof of the Second Law of Thermodynamics concerning ever-increasing

entropy. In the absence of spatial gradients, the H-theorem shows clearly that the entropy, given in [24] by

$$S \equiv -\int_{\Omega} f(\boldsymbol{\xi}) \log f(\boldsymbol{\xi}) d\boldsymbol{\xi}, \tag{2.2}$$

is ever-increasing towards its equilibrium value in a gas (the actual function Boltzmann used is inconsequentially different from this expression) [6], [15]. It wasn't until much later that it was determined that the H-theorem is descriptive but not causative in relation to the Second Law, a result of investigating Loschmidt's paradox [39].

The paradox arises when considering that the underlying physical processes which the collision integral accounts for are completely time reversible, *i.e.* reversing all particle velocities at time t and proceeding with the calculation until time $2t$ should result in all particles returning to their initial configuration. The H-theorem, on the other hand, predicts an ever-increasing entropy, regardless of the form of the velocity distribution function. This results in a contradiction since entropy clearly must decrease if particles reverse their paths under conservative forces while entropy is increasing.

Upon examination, the products $f_1' f'$ and $f_1 f$ appear in the collision integral due to the molecular chaos assumption. Under this useful simplification, the velocities of collision partners are independent, meaning that the joint density function for the probability of finding a given particle pair, $f_2(\boldsymbol{c}, \boldsymbol{c_1})$, can be factorized into a product of the single-particle distribution function for each velocity, $f(\boldsymbol{c}) f(\boldsymbol{c_1})$. This of course cannot be strictly true in general once collisions have taken place in a gas, as previous collision partners must have some interdependence due to their interaction under conservation laws. This reveals the theoretical flaw in using the H-theorem as an explanation of the Second Law; by discarding correlational information after each collision, entropy increases irreversibly, but *if correlations are preserved* states can be constructed in which entropy decreases, since no information is lost and the original state can be recalled from any future state simply by reversing all particle velocities. The gas state will then evolve exactly backwards along its original trajectory, meaning that if entropy was increasing previously, it must now be decreasing. It is the molecular chaos assumption which enforces irreversibility by simply "forgetting" the required information to construct the reverse case.

Nevertheless, while much mental effort can be expended debating the interpretation of this form of the collision integral, its usefulness cannot be denied in practice. Our concerns in this work will be with the solution of the Boltzmann equation for practical purposes more than with its ultimate validity and meaning, so we accept the form (2.1).

2.2 Direct Solution Methods for the Boltzmann Equation

Direct Numerical Simulation (DNS)

There are very few examples of practically-sized DNS computations, owing to the fact that the Boltzmann equation has solutions which exist in a seven-dimensional space (physical space, particle velocity space, and time), ensuring that computational effort scales sharply with increasing simulation size and resolution. Moreover, the collision integral which makes up the right hand side is, in its general form, a five-dimensional integral which must be calculated for each point in that seven-dimensional space. Naturally, such extreme computational demands render infeasible all but the simplest solutions.

One such problem which is meaningful but simple enough to be amenable to numerical simulation is the normal shock structure. Some solutions exist for fully numerical investigations of this problem, *e.g.* [46]. The solution is enabled due to its one-dimensional (in space) and axi-symmetric (in velocity) nature, as well as being a steady solution. This results in a three-dimensional problem. The collision integral is often also simplified in this case taking advantage of the symmetry in velocity space in a careful way.

Direct Simulation Monte-Carlo (DSMC)

Currently, the most widely used method for non-equilibrium flows is Direct Simulation Monte-Carlo (DSMC) [6]. The method essentially consists of the simulation of the dynamics of a small fraction (very roughly of order 10^{-20} in a typical calculation) of the gas molecules in a flow, including collisions, free flight, and surface interactions. The molecules are embedded in a grid of cells, and the properties of molecules in each cell are averaged to estimate gas properties within the cell. Free flight of particles is simple to solve directly, and the effect of collisions between particles is implemented via a probabilistic approach, wherein molecules are randomly selected to undergo a collision with other nearby molecules. As in the Boltzmann equation, only binary collisions are considered in most implementations, as the interest is usually in dilute gases where these interactions strongly dominate

more complicated and rare collisions of three or more particles.

This simple, bookkeeping approach allows for several attractive advantages, such as simple schemes to ensure conservation laws are satisfied at each step, which implies robustness. The main driver of computational cost is the method of handling the collisions, which demands that the cell size used for sorting particles into local groups for choosing collision partners must remain significantly smaller than the mean free path (typically the mean free path $\approx 3\Delta x$). The requirement that each cell be populated with a sufficient number of molecules to accurately model the collision process then implies a very high cost of computation [47].

The method has been very successful for certain classes of problems, particularly stable, steady flows for which time-averaging may be employed to converge the result from the statistically noisy instantaneous simulation. Stable but unsteady flows are also amenable to ensemble-averaging, whereby multiple instances of the same simulation may be averaged. In these cases, storage requirements may be as low as ten simulated molecules per cell or fewer with more sophisticated sampling techniques, and the simulation may be iterated as long as desired to achieve a well-resolved solution. In the case of unstable flows however, such methods are no longer applicable in general and statistics must be gathered instantaneously (the stochastic noise in the simulation will cause divergence of different instances of the solution) [47], [51] . Such cases require larger populations of simulated molecules in each cell if smooth results are desired. In any case, solutions are ultimately limited to converge statistically, or approximately as the inverse square root of the number of simulated molecules [30].

2.3 Conservation Equations

Boltzmann's H-theorem, mentioned previously, is one example of the many useful results which can be derived from (2.1). The most general, and a first step in many other important results, is a form for conservation equations for quantities which represent expected values calculable from $f(\boldsymbol{\xi})$.

Following Section 3.3 of [6] (see also [12] and [15]), we begin by multiplying both sides of (2.1) by $\psi(\boldsymbol{\xi}, \boldsymbol{x}, t)$ and integrating over all $\boldsymbol{\xi}$, the transport equation for moments of the velocity distribution function may be written as

$$\frac{\partial \langle \psi \rangle}{\partial t} + \frac{\partial \langle \xi_j \psi \rangle}{\partial x_j} - \left\langle \frac{\partial \psi}{\partial t} \right\rangle - \left\langle \xi_j \frac{\partial \psi}{\partial x_j} \right\rangle = \Psi, \tag{2.3}$$

where the bracket operator is defined as

$$\langle \psi \rangle \equiv \rho \int_{\Omega} \psi(\boldsymbol{\xi}, \boldsymbol{x}, t) f(\boldsymbol{\xi}) \, \boldsymbol{d\xi}, \tag{2.4}$$

and Ψ is the moment collision integral,

$$\Psi = \int_{\Omega} \int_{\Omega} \int_{0}^{\pi} \int_{0}^{2\pi} \frac{\rho^2}{m} \psi \left(f_1' f' - f_1 f \right) \sigma c_r \sin\chi \, d\epsilon d\chi \boldsymbol{d\xi_1 d\xi}. \tag{2.5}$$

This can be simplified to a more instructive form if we take advantage of a pair of symmetries in (2.5). The first is the equivalence of interchanging $\boldsymbol{\xi_1}$ with $\boldsymbol{\xi}$, representing the fact that the two particles are indistinguishable. We call this the "partner symmetry". The second is more interesting, relying on the existence of inverse collisions (reversing particle velocities and repeating a collision before reversing velocities again returns particles to their original state) to allow the interchange of $\boldsymbol{\xi_1'}$ with $\boldsymbol{\xi'}$. We call this the "collision symmetry". To use these, we first expand (2.5) into

$$\begin{aligned} \Psi = & \int_{\Omega} \int_{\Omega} \int_{0}^{\pi} \int_{0}^{2\pi} \frac{\rho^2}{m} \psi f_1' f' \sigma c_r \sin\chi \, d\epsilon d\chi \boldsymbol{d\xi_1 d\xi} \\ & - \int_{\Omega} \int_{\Omega} \int_{0}^{\pi} \int_{0}^{2\pi} \frac{\rho^2}{m} \psi f_1 f \sigma c_r \sin\chi \, d\epsilon d\chi \boldsymbol{d\xi_1 d\xi}. \end{aligned} \tag{2.6}$$

We then use the collision symmetry on the first of these two integrals before recombining to arrive at

$$\Psi = \frac{\rho^2}{m} \int_{\Omega} \int_{\Omega} \int_{0}^{\pi} \int_{0}^{2\pi} \left(\psi' - \psi \right) f_1 f \sigma c_r \sin\chi \, d\epsilon d\chi \boldsymbol{d\xi_1 d\xi}. \tag{2.7}$$

This makes it clear that the integral calculates the average difference in the quantity ψ caused by collisions in a gas represented by f. We then use the partner symmetry on (2.7) to write

$$\Psi = \frac{\rho^2}{m} \int_{\Omega} \int_{\Omega} \int_{0}^{\pi} \int_{0}^{2\pi} \left(\psi_1' - \psi_1 \right) f f_1 \sigma c_r \sin\chi \, d\epsilon d\chi \boldsymbol{d\xi_1 d\xi}. \tag{2.8}$$

Finally, we average the two equivalent expressions (2.7) and (2.8) to arrive at our point,

$$\Psi = \frac{1}{2}\frac{\rho^2}{m}\int_\Omega\int_\Omega\int_0^\pi\int_0^{2\pi}\left(\psi' - \psi + \psi_1' - \psi_1\right) f f_1 \sigma c_r \sin\chi \, d\epsilon d\chi \boldsymbol{d\xi_1 d\xi}. \quad (2.9)$$

The form of (2.9) makes it readily apparent that the collision term vanishes for any quantity conserved in collisions, that is any ψ for which $\psi + \psi_1 = \psi^{'} + \psi_1^{'}$. Examples are the functions $\psi = 1$, $\psi = \xi_i$, and $\psi = \xi_i\xi_i$, representing mass, momentum, and energy conservation. Noting this and letting $\psi = 1$ in (2.3) then yields the mass conservation equation,

$$\frac{\partial\rho}{\partial t} + \frac{\partial\rho u_i}{\partial x_i} = 0, \quad (2.10)$$

letting $\psi = \xi_i$ yields the momentum conservation equations,

$$\frac{\partial\rho u_i}{\partial t} + \frac{\partial\left(p_{ij} + \rho u_i u_j\right)}{\partial x_j} = 0, \quad (2.11)$$

where $u_i \equiv \frac{1}{\rho}\langle\xi_i\rangle$ and $p_{ij} \equiv \left\langle c_i c_j \right\rangle$, with $c_i \equiv \xi_i - u_i$ denoting the peculiar velocity, and letting $\psi = \frac{1}{2}\xi_i\xi_i$ yields the energy equation,

$$\frac{\partial\left(\frac{1}{2}\rho u_j u_j + \rho e\right)}{\partial t} + \frac{\partial\left(q_i + \frac{1}{2}p_{ij}u_j + \rho e u_i + \rho u_i u_j u_j\right)}{\partial x_i} = 0, \quad (2.12)$$

where $q_i = \left\langle \frac{1}{2}c_i \, c_j c_j \right\rangle$ represents the heat flux and $e = \left\langle \frac{1}{2}c_i c_i \right\rangle$ represents the internal energy, in this case consisting only of kinetic thermal energy.

We notice immediately a point which will be a theme in this work—terms appear, in this case the terms p_{ij}, and q_i, for which we do not have equations. Closing the system requires some method of connecting these quantities to the primary variables ρ, u_i, and e.

2.4 Moment Methods

Basic Approach

The procedure used to derive (2.10), (2.11), and (2.12) may be used to derive a transport equation for any moment, of any order or type, of the velocity distribution by an appropriate choice of ψ. While any moment may be used to generate such conservation equations, the quantities of interest in a solution to the Boltzmann

equation are usually particular moments of the velocity distribution corresponding to physically meaningful quantities (*e.g.*, the velocity, u_i, pressure tensor, p_{ij}, and heat flux vector, $q_i \equiv \left\langle \frac{1}{2} c_i c^2 \right\rangle$). It is easy to imagine an ideal method which computes only these physically relevant quantities, using equations of the type (2.3). This approach is appealing in its economy; the engineer is less often interested in, say, the exact velocity distribution or the Fourier coefficients of the distribution function than in the drag coefficient.

However, these equations, as seen in our example of the mass (2.10), momentum (2.11), and energy (2.12) conservation laws, will invariably include references to quantities involving higher-order moments than those requested by choice of ψ, such as p_{ij} and q_i. This is due to the flux term in (2.3), which includes an extra factor of ξ_i in the expression for the flux in the i-direction. Additionally, the collision terms, which in all cases apart from those represented by (2.10), (2.11), and (2.12) will be present, depend on the full velocity distribution function, f. This results in a closure problem, wherein the unknown higher-order moments and the collision terms, potentially even f itself, must be connected in some way to the values of the primary variables [9].

Closure Schemes

A closure scheme is an algorithm for determining the collision terms and unclosed moments from the computed moments. One physically intuitive approach would be to simply choose a canonical distribution and match it to the known moments. For the case of known density, velocity, and total kinetic energy, a Maxwellian velocity distribution is an appropriate choice, which can be shown to result in a system of equations for those quantities equivalent to the Euler equations [29], having no heat flux or shear stress supported. Breaking up the kinetic energy into the six pressure tensor terms in conjunction with the density and velocity defines a Gaussian distribution, which includes shear stress but, being symmetric in all directions about the mean, does not account for heat flux via diffusion [36]. Both of these reconstructions are, as will be shown later, actually special cases of the Maximum Entropy closure scheme [35].

Some previous works attempted to improve upon these simple closures by proposing some *ansatz* form for the distribution function, often projecting it onto a set of orthogonal polynomials. Grad's original attempt used Hermite polynomials [19] with a Maxwellian weight to match certain higher-order moments in the hopes of

capturing more detail in the distribution function. The famous Chapman-Enskog distribution results from expanding the Boltzmann equation as a series in terms of a non-equilibrium parameter and truncating that series to produce a distribution which depends on the local velocity and temperature gradients in the gas as well as the conserved moments [9]. This approach has been quite successfully used in theoretical endeavors, making itself useful in connecting the microscopic description of gases captured by the Boltzmann equation to the macroscopic description represented by the Navier-Stokes equations. It is a well known result that, taking the appropriate limits, the Chapman-Enskog form of the distribution reduces the Boltzmann equation to the Navier Stokes equations, providing a clear way to connect bulk gas properties such as viscosity and thermal conductivity to microscopic parameters of the gas such as collision cross section and molecular mass [12].

A major problem with these polynomial fit approaches is that they are not guaranteed to be positive, and choosing such a form for the velocity distribution represents *ad hoc* assumptions which are not driven by the information present in the flow. This can lead to non-physical solutions. As a specific example, for the case of a normal shock wave the Grad 13-moment system with the Hermite closure displays discontinuities in quantities in the shock profile. In [55], Weiss explains that this is due to the limited wave speeds represented by the closure, forcing sub-shocks to accommodate disparate solutions up- and downstream with no way of communicating.

The Principle of Maximum Entropy as a Closure Scheme

All closure schemes need not be *ad hoc*, however, at least in the view of some. In [24], Jaynes propounds quite convincingly the idea that the Principle of Maximum Entropy should be the natural way to proceed in cases where limited information is present. The Principle of Maximum Entropy is that the least-biased estimate (a "working value", one could say) of a probability distribution is that distribution from the set of candidate distributions satisfying all known information (the constraints) and which has maximum entropy within that set. Entropy here is defined in the one-dimensional case as the functional

$$S(f) \equiv -\int_{\Omega} f(x) \log f(x)\, dx, \tag{2.13}$$

where Ω is again used as a symbol for the full domain of the problem, *e.g.*, the real line. The use of this principle is motivated similarly to that of Bayesian reasoning, owing to many interesting mathematical similarities, although the two approaches

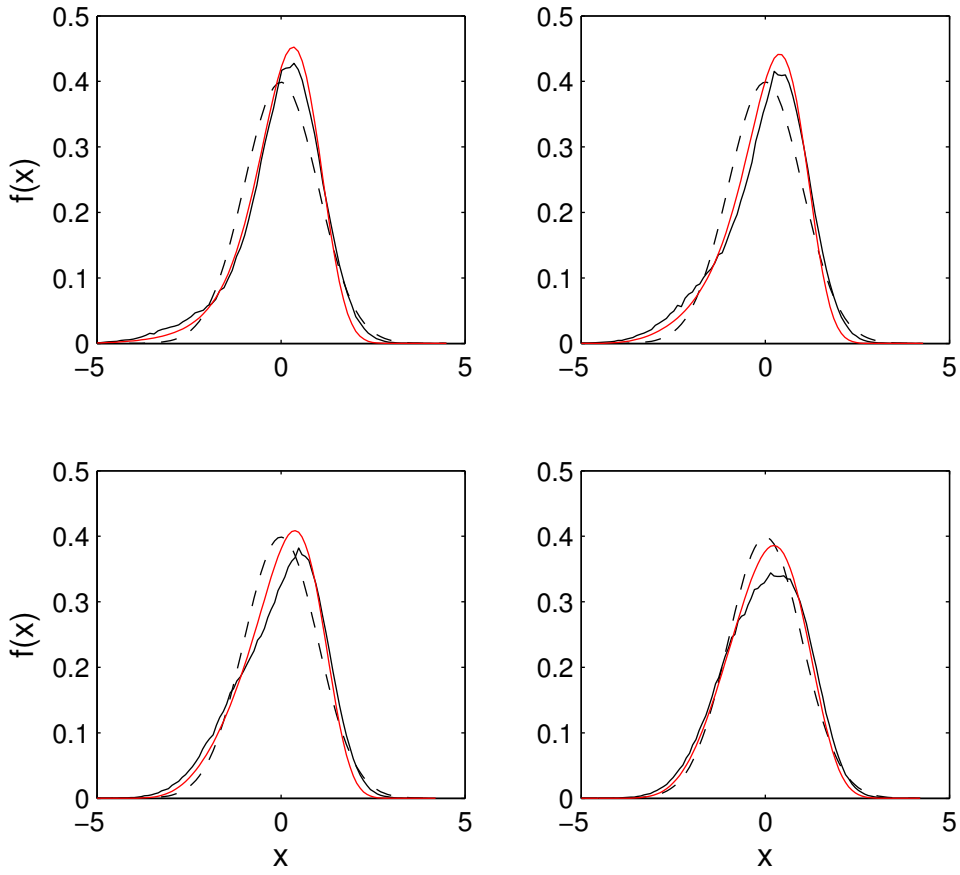

Figure 2.1: Comparing sample PDFs from various points in the interior of a Mach 2.5 normal shock wave. Black: DSMC. Black-dashed: Gaussian fit. Red: Five-Moment Maximum Entropy Reconstruction

are different philosophically and in their focus. Bayes' theorem requires a prior belief, while the Maximum Entropy Principle estimates a probability distribution given incomplete knowledge [25].

The naturalness of this closure scheme is attractive, as it should avoid imposing artificial restrictions on the form of the distribution based on the construction chosen. To illustrate, Figure 2.1 shows PDFs of normalized, shock-normal velocity, computed with DSMC by the method of Bird [6], described later in more detail. These PDFs are sampled from various points in the interior of a normal shock wave, and compared with their Gaussian and five-moment Maximum Entropy Reconstruction approximations. The Maximum Entropy Reconstruction captures the general shape of the underlying PDF using only its third and fourth moments. The procedure to compute the Maximum Entropy Reconstruction will be described fully in Chapter 3.

Initially, some early attempts at using the Maximum Entropy closure to extend the simple Euler and Gaussian moment closures to higher orders encountered some difficulty with its use. Particularly, it was shown that the closure is singular in a subspace of its domain of definition, and moreover that that singular region included the state of thermal equilibrium on its border [26]. Attempts to avoid this singularity by use of different constraining moments developed problems of their own, eliminating basic features of the solution such as Galilean invariance [29]. Other attempts to regularize the procedure and avoid the singularity without loss of fundamental solution features met with some success, but also included their own discontinuities and sub-shocks due to low characteristic speeds [41]. Eventually, McDonald developed a method involving an analytic closing flux function inspired by but not closely matching the form of the closing flux obtained from using the Maximum Entropy Principle [42]. This scheme does produce smooth shock wave profiles owing to its embrace of the singular behavior, and therefore infinite characteristic speeds, of the closure near equilibrium. There is to date, however, no convincing solution available to the problem using the closure scheme derived directly from the actual Maximum Entropy. Later, in Section 6.4, we will show why this is the case based on our own investigations into the Maximum Entropy Principle and its preferred form of the velocity distribution.

2.5 The Maximum Entropy Reconstruction (MER) as a Closure

The natural closure based on Bayesian reasoning for a moment method is to use the Principle of Maximum Entropy to compute an approximate distribution function which recognizes our knowledge of the true distribution (represented by the moments and their conservation equations) and incorporates no additional, *ad hoc* constraints on the distribution function. The principle of maximum entropy is, stated plainly:

> *The least biased estimate of a distribution satisfying a number of constraints is the distribution with maximum entropy from the set of distributions which satisfy all constraints.* [24]

This approach, suitably specified to the problem at hand, is useful in uncertainty estimation, and in estimation in the face of uncertainty, in many fields, including robotics [11], economics [16], and imaging [20]. A general mathematical derivation of the Maximum Entropy Recontruction (MER) procedure, given a set of moment constraints, is found in [52].

We do not pretend to justify here the validity of the Maximum Entropy Principle, but the following is a notional explanation of the logic behind the choice: the principle rests on the idea that entropy is a measure of the uncertainty, or perhaps one could say the "mundaneness" of a distribution function. Functions with sharp spikes and many peaks tend to have low entropy, while broad, smooth functions tend to have higher entropy. If we accept this for the time being without argument (not to say there is none to be had!), it then makes intuitive sense that we should choose the distribution with maximum entropy satisfying our present knowledge of the actual distribution. Suppose we have some information about a one-dimensional distribution—say, its first three velocity moments in x. Call $\hat{f}_3$ the maximum entropy distribution matching these constraints. Now imagine we are given additional information, namely the value of the fourth and fifth moments, $\langle x^3 \rangle$ and $\langle x^4 \rangle$. If we then compute the maximum entropy distribution according to our principle with this new information, $\hat{f}_5$, then the entropy (or uncertainty, or "mundaneness") of our distribution, $S(\hat{f}_5) \leq S(\hat{f}_3)$, with equality only in the case that the two moments were already those predicted by $\hat{f}_3$, in which case we gained no information to update our estimate. Had we not chosen the maximum entropy distribution for our candidate distribution but rather chosen some *ad hoc* function, then it is possible that there would exist a function, $\hat{f}_5$, which matches the five constraints and has more entropy than our estimate of the function given only three constraints. With our understanding of the meaning of the entropy, it is inconsistent for our estimate of the distribution to become less specialized with increasing information.

Despite this and other attractive properties of using the Maximum Entropy Principle to deal with uncertainty, one may be concerned, given the difficulties had when applying it to gas dynamics [29], that the principle may not always provide a suitable closure. It has been shown by Junk, in [26] and in [27], that a suitable MER always exists for any set of moment constraints which match any positive distribution function, except for a zero-volume subspace of otherwise valid moments which exists only in the infinite domain case. Junk's work demonstrates that this solutionless subspace, while problematic, is not pervasive within the domain of possible problems, and comparison with results by McDonald [42] seems to indicate that, at least in practice, this subspace is avoided naturally.

The computation of the MER with appropriate constraints is fairly numerically intensive, owing to the need to compute many integrals over the problem domain in order to follow the iterative solution procedure outlined in [52]. Note that in gas

dynamics, the velocity distribution is in general a three-dimensional function, meaning that the problem domain is $\mathbb{R}^3$ and the required integrals are three-dimensional. This can make the calculation of the MER impractical, and is a major reason that McDonald chose an analytical model for the closing flux over using the true MER [42]. Slight improvements can be made to the procedure to improve performance, but to date there seems to be no known way to truly avoid the ballooning computational cost as the velocity distribution becomes higher dimensional [1]. In Section 4.2 we will propose a way to avoid three-dimensional integrals entirely while still computing a fully valid MER for a given set of moments.

The final feature of the MER procedure which is of note is the behavior of the closing flux as the singular subspace is approached. Junk explores this in passing, going as far as to sketch the mechanism by which the singular subspace arises, namely the inability of the entropy functional to control the presence of tiny perturbations in the distribution function at extreme distances from the mean of the distribution [28]. These small perturbations take the form of tiny, rapidly-moving (in terms of how they move as the moments vary) packets in the one-dimensional, five-moment reconstruction case. These packets have a negligible effect on the lowest order moments, but a significant effect on the third and fourth moments, and are dominant when computing the fifth moment, which is required as the closing flux in gas dynamics applications of this order. Their presence and significance at large x means that the MER Problem must truly be solved on the infinite domain if it is to be accurate, and not on some finite-width interval around the origin, as there is important information about the function at large x which, if discarded, corrupts the solution, especially when computing the closing moments. In Section 2.4 we present an analysis of this feature of the solution, and will show that the magnitude and location of this small perturbation can be accurately predicted analytically, allowing for a great increase in the ease of computing the full MER.

Chapter 3

THE MAXIMUM ENTROPY RECONSTRUCTION PROBLEM

3.1 Statement of the problem

Background: The Problem of Moments

The Maximum Entropy Reconstruction (MER) Problem is related to the general problem of moments, detailed in [2] and [50], which is to determine whether a given sequence, μ_i, represents a sequence of moments of a probability distribution function (PDF), $f(x)$, *i.e.*,

$$\mu_i = \int_\Omega x^i f(x) dx, \qquad i = (0, 1, 2, ...), \qquad f(x) \geq 0. \tag{3.1}$$

To avoid repetition and to emphasize generality of results, we will use Ω to refer to the relevant problem domain. The canonical one-dimensional problems in this area are divided by the domain of support of the PDF; the Hamburger moment problem deals with PDFs on the real line, $x \in (-\infty, \infty)$, the Stieltjes moment problem deals with the half-line, $x \in [0, \infty)$, and the Hausdorff moment problem deals with the closed interval, $x \in [0, 1]$. The problem can, of course, be extended to higher dimensions, but much of the theory and formulation remains unchanged.

More practically interesting is the truncated moment problem, where only a finite number of moments are known and a corresponding distribution function sought [2]. While in general there exist an infinite number of PDFs which satisfy a valid finite set of moment constraints, useful results can be derived from this rather ill-posed problem. The main thrust of the analysis is to define the allowable set of moment constraints for which an answer is possible.

Consider a system like (3.1), but with a finite number, N, of moments provided, $i \in (0, 1, 2, ..., N-1)$. Consider the matrix $J_{ij} = \mu_{i+j-2}$, which can be written as

$$J_{ij} = \int_\Omega v_i v_j f(x) dx, \qquad v_i = x^{i-1}, \qquad i \in \left(1, 2, ..., \frac{N-1}{2}\right). \tag{3.2}$$

The matrix of moment integrands, $v_i v_j$, is positive definite, owing to its form as the dyadic product of a vector with itself [23], and this implies that J_{ij} is therefore positive definite if $f(x)$ is always positive. Therefore, a sequence of moments

μ_i only admits valid solutions if $J_{ij} = \mu_{i+j-2}$ is positive definite. By performing row reduction on J_{ij} to reduce it to an upper triangular matrix, we realize that each diagonal component of the matrix yields a constraint on the admissible set of moments [42]. For the case of $N = 5$ this gives the transformation

$$\boldsymbol{J} \to \begin{pmatrix} 1 & \mu_1 & \mu_2 \\ 0 & \mu_2 - \mu_1^2 & \mu_3 - \mu_1\mu_2 \\ 0 & 0 & -\mu_2^2 + \mu_4 - \frac{(\mu_3 - \mu_1\mu_2)^2}{\mu_2 - \mu_1^2} \end{pmatrix}, \tag{3.3}$$

which must have uniformly positive diagonal entries. It should be noted that adding additional moments to $\boldsymbol{J}$ does not change the previous diagonal entries, so all constraints derived from this 5-moment case apply to higher-moment cases as well. The first diagonal entry is automatically positive by definition since $f(x)$ has unit weight, the second entry's positivity requires that the variance is positive, but the third and higher entries yield quadratic inequalities in pairs of consecutive moments, starting with

$$\mu_4 > \mu_2^2 + \frac{(\mu_3 - \mu_1\mu_2)^2}{\mu_2 - \mu_1^2}. \tag{3.4}$$

Later it will be convenient to refer to this expression for the special case where the moments are normalized such that $(\mu_0, \mu_1, \mu_2) = (1, 0, 1)$, which is

$$\mu_4 > 1 + \mu_3^2. \tag{3.5}$$

This relation defines strict limits on the allowable constrained moments which are applicable even under more specific forms of the moment problem, as shown in the parabolic, outer boundary of Figure 3.1.

Another useful result can be found specifically for the Hausdorff moment problem. Examine the form

$$x^n (1 - x)^m . \tag{3.6}$$

It is clear that over the range $x \in [0, 1]$ (3.6) is positive semi-definite. Computing the expected value of this form for a given positive $f(x)$, therefore, is guaranteed to produce a positive result. This gives us the moment constraint

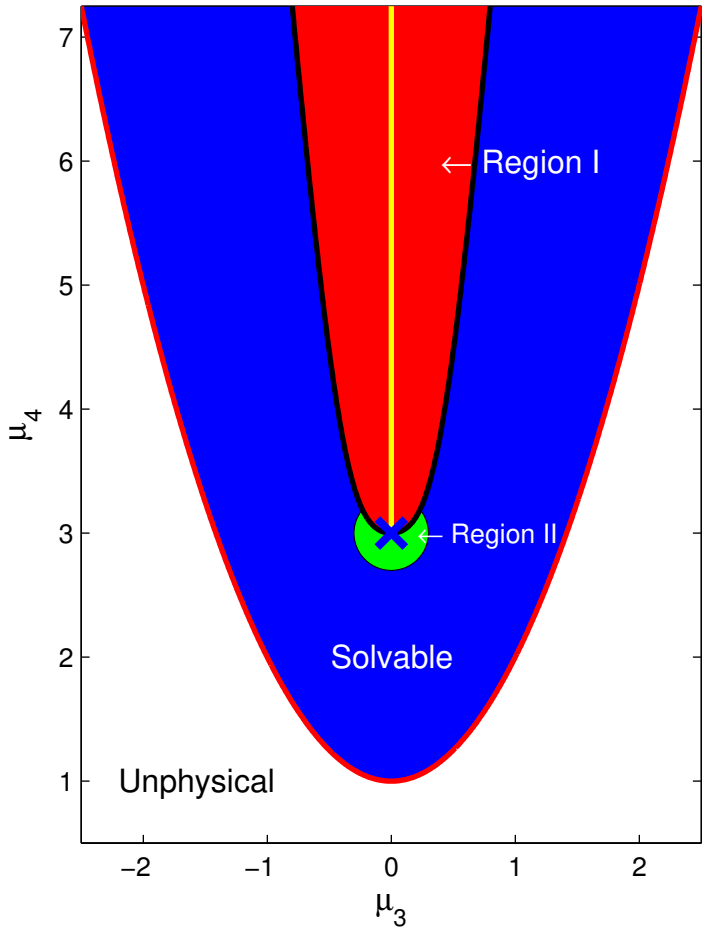

Figure 3.1: Restrictions on the solvability and features of the MER problem in (μ_3, μ_4)-space. Blue region: Moments for which the problem can be solved numerically. Red region: Singular asymptotic solution valid (Region I). Green region: Regular asymptotic solution approximately valid (Region II). White region: Unphysical moment space. Red boundary: Physical realizability boundary. Thick black boundary: Region I empirical boundary. Yellow line: Junk singular subspace. Blue X: Equilibrium point.

$$(-1)^m \left(\Delta^m \mu\right)_n = \int_0^1 x^n \left(1 - x\right)^m dx \geq 0 \tag{3.7}$$

with $(\Delta\mu)_n \equiv \mu_{n+1} - \mu_n$. This condition is called "complete monotonicity" and is an important constraint in the Hausdorff version of the moment problem [43].

Statement of the MER Problem

We first address the problem of the MER in one independent variable on the real line. Let $f(x)$, $-\infty < x < \infty$ be a PDF. Typically this could be the PDF of velocities

in a one-dimensional velocity space with $x \rightarrow c$ where c is the molecular velocity relative to the mean. The problem at hand is to reconstruct an analytic approximation to $f(x)$ subject to a specified set of moment constraints denoted by the vector $\boldsymbol{\mu}$ with components μ_i, $i = 0, 1, 2, ..., N-1$, where N is the total number of moments. In the Hamburger moment problem case, N should be odd so that the highest moment is even.

In this work, we will take the entropy to be the Shannon entropy [49],

$$S(f) \equiv -\int_{\Omega} f(x) \log [f(x)]\, dx. \tag{3.8}$$

The most general statement of the MER Problem is as follows [43]:

> Find $f(x)$ of the set of probability distribution functions with first N moments given by μ_i, $i = 0, 1, 2, ..., N-1$, which has maximum entropy.

The only known properties of the distribution are the computed moments, $\boldsymbol{\mu}$, so the maximum entropy principle calls for the maximum-entropy distribution $\hat{f}$ such that

$$\int_{\Omega} \boldsymbol{\psi}(x)\, \hat{f}(x)\, dx = \boldsymbol{\mu}, \tag{3.9}$$

where $\boldsymbol{\psi}(x)$ is a vector of moment kernels. In the simplest case, $\psi_i(x) = x^i$ (the notation is cleaner if a convention is adopted that vector components are indexed beginning with 0), though this is not necessary. In fact, using certain forms for $\boldsymbol{\psi}(x)$ can greatly improve the numerical conditioning of the reconstruction algorithm as we shall see later.

3.2 The Maximum Entropy Distribution

Using the method of Lagrange and following the derivation by [52] for constraint optimization problems, define

$$H(f, \boldsymbol{\lambda}) \equiv S(f) + \lambda_i \left(\int_{\Omega} \psi_i(x) f(x) dx - \mu_i \right), \tag{3.10}$$

summation implied, with $\boldsymbol{\lambda}$ the vector of Lagrange multipliers of the same length as $\boldsymbol{\mu}$. This function is so constructed that its maximum in f and $\boldsymbol{\lambda}$ will coincide with both the maximum of the entropy functional (3.8) and the satisfaction of the

moment constraints (3.9). We begin by finding a form for $\hat{f}(x)$, the maximal form of $f(x)$, by seeking a fixed point in H with respect to $f(x)$. We take the f functional derivative of H and set it to zero to obtain

$$\frac{\delta H}{\delta f} = -1 - \log \hat{f} + \lambda_i \psi_i(x) = 0. \tag{3.11}$$

This results in a form for the function itself,

$$\hat{f}(x) = \exp\left(\sum_{i=0}^{N-1} \lambda_i \psi_i(x) - 1\right). \tag{3.12}$$

Since we require a constraint on the total probability density, we enforce $\psi_0 = 1$ always and absorb the -1 term in the exponent into λ_0, leading to the cleaner expression with equivalent properties,

$$\hat{f}(x) = e^{\lambda_i \psi_i(x)}, \tag{3.13}$$

with summation over like subscripts implied.

Finding a fixed point for each λ_i simply returns the constraint satisfaction requirement:

$$\int_\Omega \psi_i e^{\lambda_j \psi_j} dx = \mu_i, \qquad i = 0, 1, 2, ..., N-1. \tag{3.14}$$

The reconstruction depends on the coefficients λ_i, which may be determined by solving the N nonlinear equations in (3.14). Even in this general case with unspecified domain and moments, when solvable, this problem is guaranteed to have a unique solution and can be shown [26] to be convex in the λ_i, making the Newton-Raphson method [22] an effective choice for finding solutions in practice.

Centered, Scaled Moments

The canonical version of the reconstruction problem has the simple form of $\psi_i = x^i$, or monomial moments. This form is useful in order to analyze the properties of the solution, and any moment problem using polynomial moments can be converted to a problem of this form simply by computing the relevant monomial moments from the polynomial moments and imposing them as constraints.

We require that $f(x)$ is a unit-normalized PDF, so $\mu_0 = 1$. Further, the mean, μ_1, and second moment, μ_2, can be set to zero and unity, respectively and without loss of generality, by shifting and scaling the independent variable. If we let ξ be the independent variable for some general PDF, $f_\xi(\xi)$ with moments m_i, $i = 0, 1, 2, ..., N-1$, the problem is then shifted into centered, scaled coordinates by means of the transformation

$$\xi = \sigma x + m_1, \qquad x = \frac{\xi - m_1}{\sigma}, \tag{3.15}$$

where $\sigma \equiv +\sqrt{m_2 - m_1^2}$, recalling that $m_0 = 1$. This allows the computation of the centered, scaled moments from the general moments and visa-versa by inserting the appropriate choice from (3.15) into the definitions of the moments and expanding the resulting expressions.

$$\begin{aligned} m_i &= \int_\Omega \xi^i f_\xi(\xi) d\xi = \int_\Omega (\sigma x + m_1)^i f_\xi(\sigma x + m_1)\sigma dx = \int_\Omega (\sigma x + m_1)^i f(x) dx \\ &= \int_\Omega \sum_{k=0}^{i} \binom{i}{k} (\sigma x)^k m_1^{i-k} f(x) dx \\ &= \sum_{k=0}^{i} \binom{i}{k} \sigma^k \mu_k m_1^{i-k}, \end{aligned} \tag{3.16}$$

$$\begin{aligned} \mu_i &= \int_\Omega x^i f(x) dx = \int_\Omega \frac{(\xi - m_1)^i}{\sigma^i} f\left(\frac{\xi - m_1}{\sigma}\right) \frac{d\xi}{\sigma} = \int_\Omega \frac{(\xi - m_1)^i}{\sigma^i} f_\xi(\xi)\, d\xi \\ &= \int_\Omega \sum_{k=0}^{i} \binom{i}{k} \frac{\xi^k (-m_1)^{i-k}}{\sigma^i} f_\xi(\xi) d\xi \\ &= \sum_{k=0}^{i} \binom{i}{k} \frac{m_k (-m_1)^{i-k}}{\sigma^i}. \end{aligned} \tag{3.17}$$

One can confirm that (3.17) gives $(\mu_0, \mu_1, \mu_2) = (1, 0, 1)$ for any possible m_i.

Consider for now the truncated Hamburger moment problem in the case $N = 5$, and take $\boldsymbol{m} = (m_0, m_1, m_2, m_3, m_4) \rightarrow \boldsymbol{\mu} = (1, 0, 1, \mu_3, \mu_4)$ following rescaling. The reconstruction problem may then be expressed as finding the five λ_i such that

$$\int_{-\infty}^{\infty} x^i e^{\lambda_j x^j} dx - \mu_i = 0, \qquad i = 0, 1, 2, 3, 4. \tag{3.18}$$

Solutions will be discussed in the plane defined by (μ_3, μ_4). A special case is $\boldsymbol{\mu} = (1, 0, 1, 0, 3)$, which gives the standard normal or thermal equilibrium solution $f_E(x)$, where $\boldsymbol{\lambda} = (-\log(2\pi)/2, 0, -1/2, 0, 0)$,

$$f_E(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}. \tag{3.19}$$

The result can be checked by hand. It is a degenerate case which reduces to a three-moment maximum entropy problem for $(\mu_0, \mu_1, \mu_2) = (1, 0, 1)$, which is analytically tractable owing to the ready integrals for Gaussian moments. This reduction in order happens because 0 and 3 are the third and fourth order moments, respectively, of the standard normal distribution, which is the maximum entropy solution for any three-moment maximum entropy problem (scaled and centered). We subsequently refer to $(\mu_3, \mu_4) = (0, 3)$ as the thermal equilibrium point, E. Note that a vector distribution with each component having a velocity distribution like (3.19), with the same average variance in all three components, gives rise to the Maxwell-Boltzmann velocity distribution.

The truncated Hamburger moment problem has a solution (that is, there exists *some* distribution matching the given moments) so long as the target moments correspond to those of some possible positive distribution. The maximum entropy version of this problem, however, is not uniquely solvable for a zero-volume set of moments within this domain. In the case of $N = 5$, this situation occurs when $\mu_3 = 0$ and $\mu_4 > 3$, which is a ray extending from the equilibrium point in the (μ_3, μ_4) plane [26].

This singularity arises in the Hamburger moment problem because of the unbounded domain, and is fundamental to the behavior of the maximum entropy problem. Intuitively, the problem arises because increasing the fourth moment of a distribution of the form (3.13) with $N = 5$, while maintaining zero odd moments, means that the tails of the distribution must be increased at large x. However, since the exponential function is monotonic, $\lambda_2 x^2$ is dominated only by $\lambda_4 x^4$ at large x. This means that increasing λ_2 will not have the desired effect, as λ_4 must decrease in order to control the growth of both moments (and λ_4 must always be non-positive or the distribution will become unbounded). Decreasing λ_4 necessarily reduces the weight of the tails of the distribution and prevents any growth of μ_4 relative to μ_2 beyond a certain point.

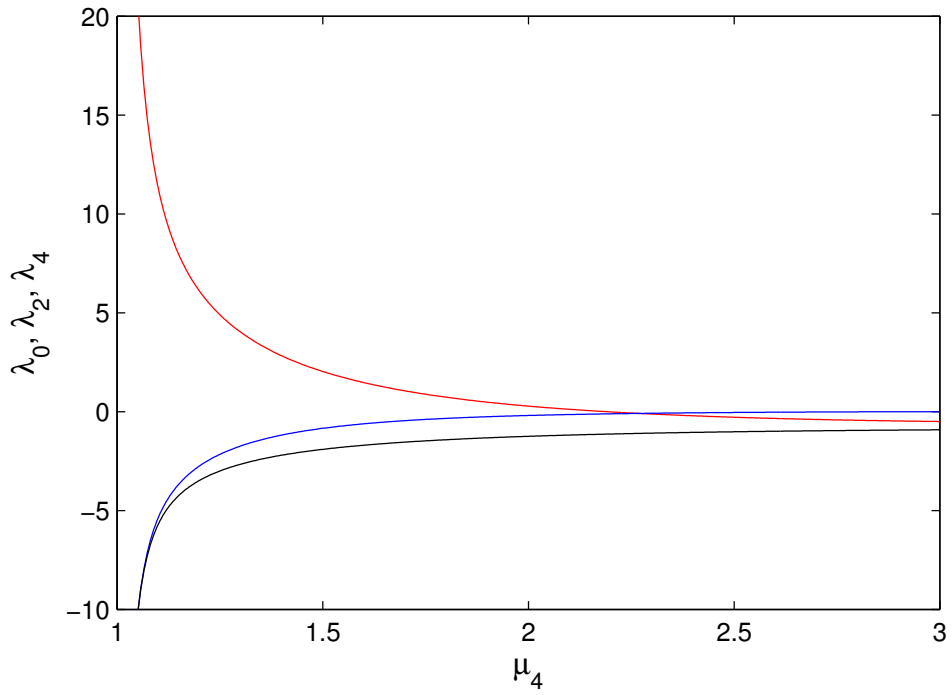

Figure 3.2: λ_0 (black), λ_2 (red), and λ_4 (blue) values of the MER on the infinite domain vs. μ_4. $\mu_3 = 0$ is held constant. The values diverge as μ_4 approaches 1.

Figure 3.2 shows how the range of μ_4 is bounded between 1 and 3 for even distributions of the type (3.13). It can be seen that as μ_4 tends to 3, λ_4 tends to zero.

This singular line would not be so significant if the solution were well-behaved elsewhere, but in fact it will be seen that as the singular line is approached, moments of higher order than the constraining moments become unbounded. Given that this behavior causes extreme numerical instability, limiting the robustness of any algorithm designed to solve the Maximum Entropy moment problem, and the region of higher order moment unboundedness extends to the neighborhood of the equilibrium point, it becomes important to understand the nature of this singularity in order to fully tackle the problem.

The analysis of the maximum entropy solution near this singular line is more conveniently done by working with $\boldsymbol{p} = \boldsymbol{\lambda} - \boldsymbol{\lambda}_E$, the deviation in $\boldsymbol{\lambda}$ from the equilibrium solution, $\boldsymbol{\lambda}_E$. These become

$$p_0 = \lambda_0 + \frac{1}{2}\log(2\pi), \quad p_1 = \lambda_1, \quad p_2 = \lambda_2 + \frac{1}{2}, \quad p_3 = \lambda_3, \quad p_4 = \lambda_4, \qquad (3.20)$$

and (3.14) becomes

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x^i e^{-\frac{x^2}{2}} e^{p_j x^j} dx = \mu_i, \quad i = 0, 1, 2, ..., N-1. \tag{3.21}$$

For given (μ_3, μ_4), (3.21) gives five nonlinear equations for $\boldsymbol{p}$. We define the maximum entropy polynomial (MPE) by

$$g(p_i, x) = -\frac{1}{2}\log(2\pi) + p_0 + p_1 x + (p_2 - \frac{1}{2})x^2 + p_3 x^3 + p_4 x^4, \tag{3.22}$$

being the natural logarithm of the MER, (3.13). An acceptable solution requires $p_4 < 0$ (except at equilibrium which is a degenerate case where $\boldsymbol{p} = 0$) so that the reconstructed $\hat{f}$ vanishes exponentially as x approaches $\pm\infty$ and all moments are finite. The extrema of the fourth-order polynomial $g(x)$ must then consist of either a single, global maximum, or else two maxima and one minimum. Since the exponential function is strictly monotonic and always positive, this implies that $\hat{f}$ must also be either unimodal or bimodal. This detail will become important as the problem is analyzed further.

3.3 Numerical Solutions

Using a quadrature scheme for the x-integrations together with a suitably robust Newton-Raphson scheme, (3.21) can be solved numerically to high accuracy within a finite region of the moment space for the given problem. As an illustration, a simple MATLAB code was written to solve the MER problem for general sets of monomial moments. The integrals required to compute the Jacobian for the Newton's step were performed using a composite Simpson's rule [56] on the infinite domain, mapped onto a finite interval with the change of variables $x = \tan\theta$, using 1200 integration points. The Newton's step is made robust by employing a backtracking scheme, whereby unacceptably high errors or undefined components in the Jacobian are detected and the Newton's increment for λ_i is reduced by a factor of 5 and the solution error is recalculated. Figure 3.3 shows a sequence of such solutions for increasing numbers of moment constraints. In this figure, a true distribution supplies the target moments and is constructed as a summation of two Gaussian expressions, one with weight 0.4, mean -0.93394, and standard deviation 0.45, the other with weight 0.6, mean 0.62263, and standard deviation 0.75. The moments to eighth order are $\boldsymbol{\mu} = (1, 0, 1, 0.2243, 2.2222, 1.4699, 7.5929, 9.2924, 35.895)^T$. Reconstructions including all moments to second, fourth, sixth, and eighth order are shown. In

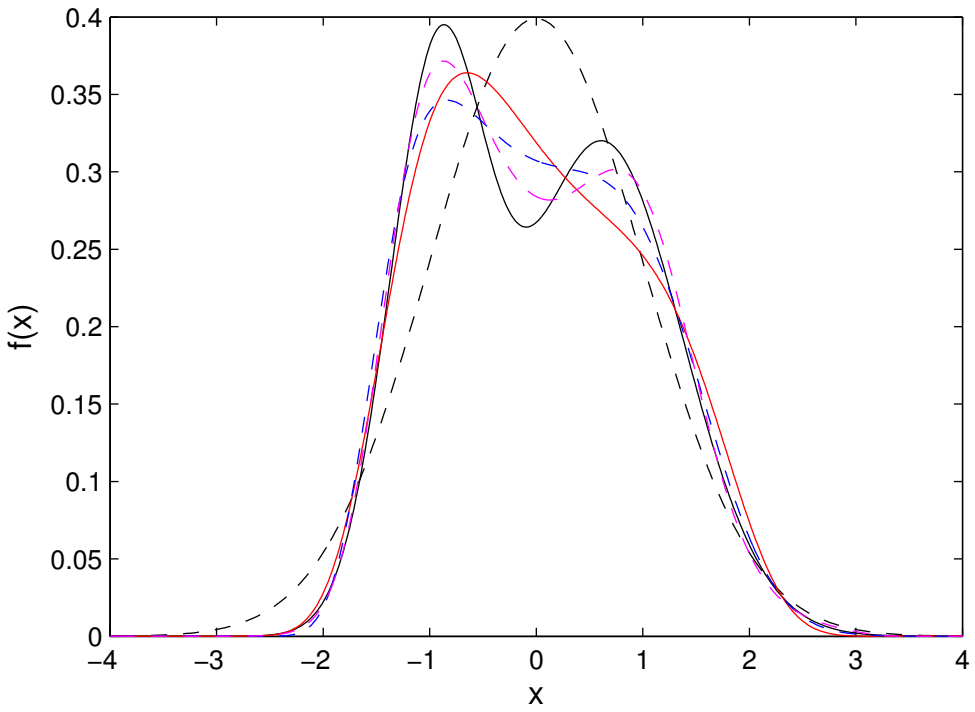

Figure 3.3: Example of increasing fidelity of the MER with increasing numbers of moments.Black line: True function from which moments are calculated to serve as constraints in the MER. Other lines represent the MER using the first N moments; black-dashed: $N = 3$, red: $N = 5$, blue-dashed: $N = 7$, magenta-dashed: $N = 9$.

this benign and well-behaved case, the higher-order reconstructions demonstrate increasing fidelity of approximation to the true function with increasing numbers of moments, but we shall see that naive reconstruction procedures such as these are of little practical use owing to the numerical difficulty of the reconstruction in a particular region of (μ_3, μ_4)-space.

When $|\mu_3|$ becomes small with fixed $\mu_4 > 3$, it is found that naive solution schemes become increasingly ill-conditioned in the sense that the distribution becomes extremely sensitive to the values of λ_i, to such a degree that numerical error precludes finding a suitable solution. A partial reason for this is apparent from Figure 3.4. As $|\mu_3|$ reduces, the MEP, $g(p_i, x)$, obtained with the simple MATLAB MER code, shows two maxima and a deep minimum. The black line represents the naive numerical solution, and the dashed lines represent its approximation by the Hybrid Maximum Entropy Reconstruction (HMER) which we will describe in detail at the end of the following chapter. In the MEP, one maximum is always near $x = 0$ while the second (or left maximum for $\mu_3 < 0$) moves to increasingly large x. Further, the

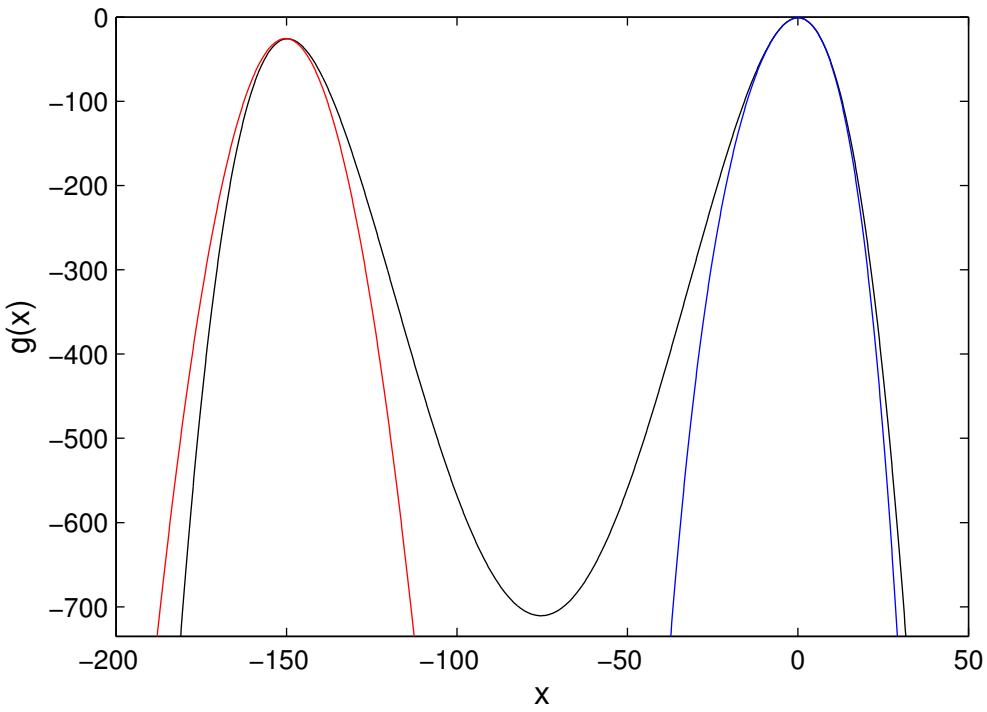

Figure 3.4: Black: True solution, Red: Hybrid Model IMP, Blue: Hybrid Model Central Component. A particular case of the five-moment MER Problem with $\mu_3 = -0.04$ and $\mu_4 = 3.01$. The Hybrid model, HMER, and the IMP are explained in detail in Chapter IV.

size of the left peak in $\hat{f}(x)$ becomes progressively smaller. Without some understanding of the nature of $g(p_i, x)$ in this region, the only way to produce solutions is to use a very large number of integration points, so that sufficient range and resolution is achieved to capture the contribution to the required moments of the region around each maximum. This leads to an inevitable numerical ill-conditioning of the problem as literally infinite range and resolution are required using this approach as $\mu_3 \to 0$. This results in an unacceptable computational burden and degraded accuracy as the singular line is approached, necessitating a better understanding of the solution in this region.

This bimodal character of $\hat{f}(x)$ near the singular line, with peaks that are widely separated in x leads to numerical values of higher-order moments $\mu_5, \mu_6 \cdots$ that become increasingly large when $|\mu_3| \to 0$ at fixed $\mu_4 > 3$. This is also noticed by Junk [28], and will be demonstrated and discussed in more detail subsequently. The interpretation is clear: the capability of the ME reconstructed $\hat{f}(x)$ to produce given (μ_3, μ_4) with $(\mu_0, \mu_1, \mu_2) = (1, 0, 1)$ fixed is severely constrained by the properties of

the fourth-order $g(p_i, x)$. With the given moment constraints, the available solution is of a wide-separation, bimodal form containing peaks of increasingly disparate amplitude. This character suggests a singular limit when $\mu_3 \to 0$ with $\mu_4 > 3$.

In what follows it is shown that separate analytic expansions can be constructed in two regions near the moment axes. The regions are illustrated in Figure 3.1. The first, denoted Region I, lies adjacent to that part of the co-ordinate axis $|\mu_3| << 1,\ \mu_4 > 3$, while the second, presently denoted Region II, additionally requires $|\mu_4 - 3| < \mu_3^2$.

Chapter 4

ASYMPTOTIC SOLUTION TO THE MER PROBLEM

4.1 Two-PDF Model in Region I

We first consider the asymptotic solution for the MER when $\mu_3 \to 0$ with $\mu_4 > 3$ (Region I). This will be seen to be singular. Solutions are anticipated to be antisymmetric in the odd-indexed p_i variables and also in odd-indexed moments, and so presently it is sufficient to discuss $\mu_3 < 0$. Motivated by inspection of many numerical solutions similar to Figure 3.4 showing bimodal distributions, our approach is to build a two-PDF model for $\hat{f}(x)$ of the form

$$\hat{f}(x) = \epsilon\, f_1 + (1-\epsilon) f_0. \tag{4.1}$$

In (4.1), f_0 will be modeled as a perturbed Gaussian centered on $x = 0$, approximating the central maximum, and f_1 as a shifted Gaussian approximation to the left-hand maximum in $g(p_i, x)$ shown in Figure 3.4. The left-hand maximum center is at x_1 (to be determined in this analysis) and its contribution to the zeroth-moment of $\hat{f}(x)$ is $\epsilon \ll 1$, subsequently referred to as its amplitude (also to be determined). The analysis will seek analytical approximations to x_1, ϵ , the p_i and moments $\mu_i, i > 4$ as functions of (μ_3, μ_4) for $|\mu_3| \ll 1$. Numerical solutions suggest, and this is later verified, that asymptotic solutions in the sense of $|\mu_3| \to 0$, $\mu_4 > 3$ exist with, to leading order, $p_0, p_2, p_4 = O(\mu_3^2)$ and $p_1, p_3 = O(\mu_3)$.

4.2 Position of the Itinerant Moment Packet (IMP)

The left peak is centered on the local maximum of $g(x)$ defined by the left-most (for $\mu_3 < 0$) or right-most ($\mu_3 > 0$) x for which

$$\frac{dg}{dx} = p_1 + 2\,(p_2 - \tfrac{1}{2})\,x + 3\,p_3\,x^2 + 4\,p_4\,x^3 = 0. \tag{4.2}$$

Numerical investigation suggests that the location, x_1, is nearly proportional to μ_3^{-1}. Since the left maximum rapidly carries important moment information towards infinity in a compact packet as $\mu_3 \to 0$, and because of its small amplitude and disproportionate role in causing the difficulties encountered when naively solving the MER problem, we aptly name it the Itinerant Moment Packet (IMP). Proceeding with the *ansatz* (which will be verified later) that $x_1 \propto \mu_3^{-1}$ as μ_3 becomes small, a

dominant balance argument suggests that p_1 and p_2 in (4.2) may be neglected which gives the estimate of the IMP's position as

$$x_1 = \frac{-3\,p_3 + \sqrt{9\,p_3^2 + 16\,p_4}}{8\,p_4}. \tag{4.3}$$

A more precise analysis including the neglected p_1 and p_2 gives a more complicated but consequentially indifferent expression for x_1 in terms of all p_i, so for the sake of neatness we will proceed with the approximation (4.3).

The osculating parabola approximating the IMP is given by

$$g_p(p_i, x) = g_1 + \frac{1}{2}\,g_1^{''}\,(x - x_1)^2, \tag{4.4}$$

where $g_1 = g(x = x_1)$ and $g_1^{''} = d^2g/dx^2(x = x_1) < 0$. Near $x = x_1$, we then have

$$\epsilon f_1(p_i, x) = \exp[g_p(p_i, x], \quad \epsilon = \exp[g_1(p_i)]\,\sqrt{-\frac{2\,\pi}{g_1^{''}(p_i)}}. \tag{4.5}$$

Because $|x_1| \gg 1$ and is therefore many times larger than the variance of $f_1(p_i, x)$, we will approximate the contribution from $f_1(p_i, x)$ to the k-th moment of $\hat{f}$ about $x = 0$ as $\epsilon\,x_1^k$. Henceforth, for the present analysis, we need be concerned only with $\epsilon(p_i)$.

The quantities g_1 and $g_1^{''}$ are obtained by direct substitution of (4.3) into both (3.13) and its second derivative, and are respectively

$$\begin{aligned} g_1(p_i) = p_0 &- \frac{27p_3^4}{512p_4^3} - \frac{9p_3^2}{64p_4^2} + \frac{9p_2p_3^2}{32p_4^2} - \frac{1}{16p_4} + \frac{p_2}{4p_4} - \frac{3p_1p_3}{8p_4} + \frac{9p_3^3\sqrt{9p_3^2 + 16p_4}}{512p_4^3} \\ &+ \frac{p_3\sqrt{9p_3^2 + 16p_4}}{32p_4^2} - \frac{3p_2p_3\sqrt{9p_3^2 + 16p_4}}{32p_4^2} + \frac{p_1\sqrt{9p_3^2 + 16p_4}}{8p_4} - \frac{1}{2}\,\log[2\,\pi], \end{aligned} \tag{4.6}$$

$$g_1^{''}(p_i) = \frac{9p_3^2 + 16(1 + p_2)p_4 - 3p_3\sqrt{9p_3^2 + 16p_4}}{8p_4}. \tag{4.7}$$

It can be seen that the complicated form of (4.6) contains terms of order p_4^{-1}. Since we anticipate $p_2, p_4 = O(\mu_3^2)$ these terms, unless canceled, will lead to the unacceptably singular form $\epsilon \sim \exp[\mu_3^{-2}]$, $\mu_3 \to 0$. This can be mollified by

requiring the leading order relation $p_4 = (-1/2 + \alpha)\, p_3^2$. Substituting into (4.6) and expanding to order α then gives

$$g_1 \sim p_0 + \frac{p_2}{p_3^2} + \frac{p_1}{p_3} + \frac{\alpha(1 + 8p_2 + 4p_1 p_3)}{p_3^2} - \tfrac{1}{2}\, \log[2\,\pi]. \tag{4.8}$$

The singularity is relieved by the choice $\alpha = 0$. This gives the leading-order closure relation

$$p_4 = -\tfrac{1}{2}\, p_3^2 + HOT. \tag{4.9}$$

Using (4.9) in (4.3) we find that $x_1 = p_3^{-1}$ to leading order. Surprisingly, it will be seen that this result together with (4.9) is all that is required from the analysis of f_1 to obtain the leading-order behavior when $\mu_3 \to 0$.

Moment equations

We now address the contribution to moments from f_0 modeled as a perturbed Gaussian. First, f_0 is written in the form

$$f_0(p_p, x) = \frac{1}{\sqrt{2\,\pi}} \exp\left(-\tfrac{1}{2}\, x^2\right)\, \exp(Z) \quad , Z = \sum_{j=0}^{J} p_j\, x^j. \tag{4.10}$$

We seek a perturbation expansion of the p_i of the form

$$\begin{aligned} p_0 &= a_0\, \mu_3^2 + \cdots, \\ p_1 &= a_1\, \mu_3 + b_1\, \mu_3^3 + \cdots, \\ p_2 &= a_2\, \mu_3^2 + \cdots, \\ p_3 &= a_3\, \mu_3 + b_3\, \mu_3^3 + \cdots, \\ p_4 &= a_4\, \mu_3^2 + \cdots, \\ \epsilon &= e_4\, \mu_3^4 + e_6\, \mu_3^6 + \cdots, \end{aligned} \tag{4.11}$$

where the a_i, b_i are functions of μ_4 to be determined. For p_0, p_2 and p_4 higher-order terms are expected to be of the form $\mu_3^4\,(b_i + c_i\, \log(|\mu_3|))$, $i = 0, 2, 4$. These are omitted here because they do not affect the present expansion to order(μ_3^3). The given form for ϵ will be later shown to be required for self-consistency. The reconciliation of the algebraic expansion ϵ with the apparent exponentially small form suggested by the second of (4.5) will be discussed subsequently and will shown to associated with an order $\mu_3^4\, \log[|\mu_3|]$ correction to p_4.

Next, the quantity $\exp(Z)$ in (4.10) is expanded as a Taylor series in Z about $Z = 0$ to third order and $Z(p_i, x)$ is then substituted, followed by further substitution of the

expansion (4.11). This gives an expression that represents the model f_0 as a Gaussian times a high-order polynomial in x in which the coefficients are combinations of powers of the a_i, b_i and μ_3. The third-order expansion can be shown to be sufficient to capture all contributions to order μ_3^3. We note in passing that this expansion has not been proven to be convergent for some range of μ_3 nor can it be guaranteed to give positive f_0. Nonetheless, k-th-order moments of the model f_0 exist for any $k \geq 0$ and can be determined analytically by integration in $(-\infty < x < \infty)$. The calculated moments for $k = 0, \cdots 4$ may now be determined by integration and added to the contributions from f_1 discussed above to give the five moment equations, as power series in μ_3, to order μ_3^3

$$(a_0 + \tfrac{1}{2}\, a_1^2 + a_2 + 3a_1\, a_3 + \tfrac{15}{2}\, a_3^2 + 3a_4)\, \mu_3^2 = 0, \tag{4.12}$$

$$\begin{aligned}(a_1 + 3\, a_3)\, \mu_3 + (a_0 a_1 + \tfrac{1}{2}\, a_1^3 + 3\, a_1\, a_2 + 3a_0\, a_3 + \tfrac{15}{2}\, a_1^2 a_3 + 15\, a_2\, a_3 \\ + \tfrac{105}{2}\, a_1\, a_3^2 + \tfrac{315}{2}\, a_3^3 + 15a_1\, a_4 + 105\, a_3\, a_4 + b_1 + 3\, b_3)\mu_3^3 \\ + (e_4\, \mu_3^4 + e_6\, \mu_3^6)(a_3\, \mu_3 + b_3\, \mu_3^3)^{-1} = 0,\end{aligned} \tag{4.13}$$

$$\begin{aligned}(a_0 + \tfrac{3}{2}\, (a_1^2 + 2\, a_2 + 10\, a_1\, a_3 + 35\, a_3^2 + 10\, a_4))\mu_3^2 \\ + (e_4\, \mu_3^4 + e_6\, \mu_3^6)(a_3\, \mu_3 + b_3\, \mu_3^3)^{-2} = 0,\end{aligned} \tag{4.14}$$

$$\begin{aligned}(-1 + 3\, a_1 + 15\, a_3)\, \mu_3 + (3\, a_0\, a_1 + \tfrac{15}{2}\, a_1^3 + 15\, a_1\, a_2 + 15\, a_0\, a_3 + \tfrac{105}{2}\, a_1^2\, a_3 + 105\, a_2\, a_3 \\ + \tfrac{945}{2}\, a_1\, a_3^2 + \tfrac{3465}{2}\, a_3^3 + 105\, a_1\, a_4 + 945\, a_3\, a_4 + 3\, b_1 + 15\, b_3)\, \mu_3^3 \\ + (e_4\, \mu_3^4 + e_6\, \mu_3^6)(a_3\, \mu_3 + b_3\, \mu_3^3)^{-3} = 0,\end{aligned} \tag{4.15}$$

$$\begin{aligned}3 - \mu_4 + (3\, a_0 + \tfrac{15}{2}\, (a_1^2 + 2\, a_2 + 14\, a_1\, a_3 + 63\, a_3^2 + 14\, a_4))\, \mu_3^2 \\ + (e_4\, \mu_3^4 + e_6\, \mu_3^6)\, (a_3\, \mu_3 + b_3\, \mu_3^3)^{-4} = 0.\end{aligned} \tag{4.16}$$

In the above, the last term in the last four moment equations are the contributions from the IMP evaluated as $\epsilon\, x_1^k = \epsilon / p_3^k$, $k = 1, \cdots 4$, using $x_1 = p_3^{-1}$. The equation set contains cubic nonlinearities in the coefficients and are not closed. Nonetheless they can be solved simply and efficiently.

First, equating the $O(\mu_3)$ term to zero in (4.13) immediately gives

$$a_1 = -3a_3. \tag{4.17}$$

Next, the last term in (4.15) is expanded as a Taylor Series in μ_3 and terms of order μ_3 retained. The first order term in μ_3 in (4.15) then gives

$$e_4 = -a_3^3\, (-1 + 6a_3). \tag{4.18}$$

Substituting the expressions for a_1 and e_4 into (4.16) and again expanding the last term in a Taylor series to order μ_3^2 then gives, equating the order one term to zero,

$$a_3 = 1/(\mu_4 + 3). \tag{4.19}$$

This then gives

$$e_4 = (\mu_4 - 3)\,(\mu_4 + 3)^{-4}. \tag{4.20}$$

Substituting the results in hand, (4.17) through (4.20), into the order μ_3^2 terms in (4.12),(4.14), (4.16), and into the order μ_3^3 terms in (4.13),(4.15) and equating these to zero then leads to

$$a_0 + a_2 + 3\,a_4 + \frac{3}{(\mu_4 + 3)^2} = 0, \tag{4.21}$$

$$b_1 + 3\,b_3 + a_2\,\frac{6}{\mu_4 + 3} + a_4\,\frac{60}{\mu_4 + 3} + \frac{51 + \mu_4}{(\mu_4 + 3)^3} = 0, \tag{4.22}$$

$$a_0 + 3\,a_2 + 15\,a_4 + \frac{18 + \mu_4}{(\mu_4 + 3)^2} = 0, \tag{4.23}$$

$$3\,b_1 + b_3\,(8 - \mu_4) + a_0\,\frac{6}{\mu_4 + 3} + a_2\,\frac{60}{\mu_4 + 3} + a_4\,\frac{630}{\mu_4 + 3} + e_6\,(\mu_4 + 3)^3 + \frac{720}{(\mu_4 + 3)^3} = 0, \tag{4.24}$$

$$3\,a_0 + 15\,a_2 + 105\,a_4 + 4\,b_3\,(9 - \mu_4^2) + e_6\,(\mu_4 + 3)^4 + \frac{225}{(\mu_4 + 3)^2} = 0. \tag{4.25}$$

These are five linear equations for the six unknowns $(a_0, a_2, a_4, b_1, b_3, e_6)$. It can be seen that a consistent power series expansion is possible only if ϵ takes the form given in the last of (4.11). Any alternative form, for example a leading-order term of lower order, can be shown lead to unbalanced terms of zeroth order. Further, the order μ_3^4 as the leading order in the IMP amplitude, considering (4.2) placing the IMP at a location that scales as μ_3^{-1}, is consistent with the form of perturbation used by Junk to demonstrate the existence of the singular region [28]. Note that where IMP-like packets of equivalently scaling amplitude and position are used to explain the existence of the singular line in Region I, Junk supplied no justification or precise prediction of these packets.

Equations (4.21-4.25) can now be solved using the closure relation (4.9). This gives

$$a_4 = -\frac{(\mu_4 + 3)^{-2}}{2}. \tag{4.26}$$

Solving (4.21-4.25) and substituting into (4.11) provides the solution

$$
\begin{aligned}
p_0 &= \frac{6+\mu_4}{2\,(\mu_4+3)^2)}\,\mu_3^2 + \cdots \\
p_1 &= -\frac{3}{\mu_4+3}\,\mu_3 + \frac{189-33\mu_4+2\mu_4^2}{(\mu_4+3)^4}\,\mu_3^3 + \cdots \\
p_2 &= -\frac{9+\mu_4}{2\,(\mu_4+3)^2}\,\mu_3^2 + \cdots \\
p_3 &= \frac{1}{\mu_4+3}\,\mu_3 + \frac{3\,(5\,\mu_4-19)}{(3+\mu_4)^4}\,\mu_3^3 + \cdots \\
p_4 &= -\frac{\mu_3^2}{2(3+\mu_4)^2} + \cdots \\
\epsilon &= \frac{\mu_4-3}{(\mu_4+3)^4}\,\mu_3^4 + \frac{6\,(57-84\,\mu_4+11\,\mu_4^2)}{(\mu_4+3)^7}\,\mu_3^6 + \cdots
\end{aligned}
\tag{4.27}
$$

An example of these values used to predict the full distribution function is shown in Figure 4.1. Excellent agreement is found near the central component of the distribution, but near the IMP location there are necessarily many cancellations, degrading accuracy of an approximate approach such as this. The distribution is very sensitive to small changes in the p_i, as the relative error from the true p_i in this case is less than $3 \times 10^{-}3$ for all values of p_i predicted, despite the significant resulting error in the distribution function itself. It is interesting to note, however, that the model for the central component qualitatively resembles the full solution near the IMP, despite this portion of the distribution having negligible influence on the central component's moments. The IMP model nicely approximates the left peak of the distribution.

Moments $\mu_k, k \geq 5$ are given by ϵ/p_3^k with no contribution from f_0. To leading order these are

$$\mu_k = \mu_3^{4-k}\,(\mu_4-3)(\mu_4+3)^{k-4}, \quad k = 5, \cdots \tag{4.28}$$

These are singular when $\mu_3 \to 0$. The odd-order moments change sign as μ_3 passes through zero. Note in particular the fifth-order moment,

$$\mu_5 = \frac{\mu_4^2-9}{\mu_3}, \tag{4.29}$$

which will become important later in this work. This moment is singular and multivalued around the point E. If approaching E along a line with angle ω with respect to the μ_3-axis, then μ_5 becomes

$$\mu_5 = \tan\omega\,(\mu_4+3), \tag{4.30}$$

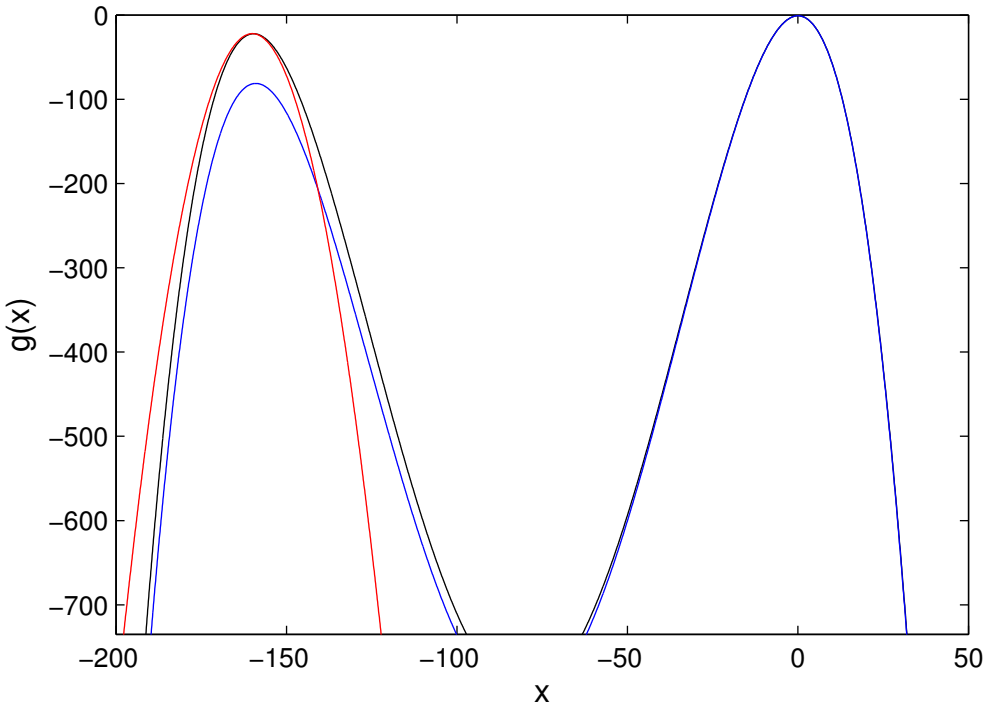

Figure 4.1: Comparison of the logarithm of the full MER solution with the asymptotic solution given by (4.27) in Region I with $\mu_3 = -0.04$ and $\mu_4 = 3.01$. Black: Full MER solution. Red: IMP model Blue: Central component of Region I asymptotic solution.

which clearly has a different limit at E for different ω.

Asymptotic solution for f_1

The solution is completed by obtaining the asymptotic form of f_1, defined by (4.4) and (4.5) when $\mu_3 \to 0$. The quantity $g_1^{''}$ is estimated by substituting the leading-order expressions for the p_i from (4.27) and $x_1 = 1/p_3$ (see discussion following (4.9)) directly into the second derivative of g defined by (3.22), followed by Taylor series expansion in μ_3 about $\mu_3 = 0$. The algebra can be done with Mathematica and gives $g_1^{''} = -1 + O(\mu_3^2)$. Thus, the IMP component of the solution has unit variance.

The quantity $g_1 = g(x = x_1)$ is more delicate and requires care. First, the p_i are written as $p_i = p_{i0} + b_i$ where the p_{i0} are given by the respective first terms in each of (4.27) and the b_i are higher-order corrections, two of which (b_1, b_3) are given in (4.27). It is then assumed that b_0, b_2, and b_4 are at worst of order $\mu_3^4 \log[|\mu_3|]$. When $g_1 = g(x = x_1)$ is then expanded for small $|\mu_3|$ and then the limit $\mu_3 \to 0$

examined, it is found that only surviving terms are contained in

$$g_1 = b_4 \frac{(\mu_4 + 3)^4}{\mu_3^4} - \frac{15 + \mu_4 + \log[2\,\pi]}{2} + \cdots \tag{4.31}$$

This suggests the choice

$$b_4 = \frac{\mu_3^4}{(\mu_4 + 3)^4} \;\; (\alpha \; \log[-\mu_3] + \log[B(\mu_4)] + K(\mu_4)), \quad \mu_3 < 0 \tag{4.32}$$

The constant α, and the functions $B(\mu_4)$, $K(\mu_4)$ are now chosen such that ϵ given by the second of (4.5) (with $g_1^{''} = -1$) matches the leading order μ_3^4 term of the last equation of (4.27). This is satisfied by the choice

$$\alpha = 4, \quad B = \frac{\mu_4 - 3}{(\mu_4 + 3)^4}, \quad K = \frac{15 + \mu_4}{2}. \tag{4.33}$$

The final form for the IMP is then a normalized Gaussian centered on $x = x_1$

$$\epsilon f_1(p_i, x) = \frac{(\mu_4 - 3)\mu_3^4}{(\mu_4 + 3)^4} \frac{1}{\sqrt{2\,\pi}} \mathrm{e}^{-\frac{1}{2}(x - x_1)^2}, \tag{4.34}$$

and b_4 is given by

$$b_4 = \frac{\mu_3^4}{(\mu_4 + 3)^4} \left(4 \; \log(-\mu_3) + \log(\mu_4 - 3) - 4 \; \log(\mu_4 + 3) + \tfrac{1}{2}\,(15 + \mu_4)\right). \tag{4.35}$$

An example of the resulting two-PDF model of the MER in Region I can be seen in Figure 3.4. The two-PDF model with the IMP matches quite well over the ranges of x where the true solution is not extremely attenuated. Capturing the IMP analytically allows for a treatment of the near-equilibrium behavior that is much better conditioned by avoiding the appearance of singular moments in the Reconstruction Jacobian.

It can be seen from (4.35) that b_4 is logarithmically singular when $\mu_4 \longrightarrow 3$ from above with μ_3 fixed. This suggests that, for any small but finite $|\mu_3|$, a different approach is required for sufficiently small $|\mu_4 - 3|$. This is supported by numerical solutions very near the line $\mu_4 = 3$, that indicate that here $g(p_i, x)$ defined by (3.22) has only one extremum that is near $x = 0$. We call this region of (μ_3, μ_4)-space near equilibrium with unimodal maximum entropy distributions Region II, as shown in Figure 3.1.

4.3 Perturbation Expansion in Region II

Here we use a double expansion in both μ_3 and $\eta \equiv \mu_4 - 3$. The analysis is simplified by the absence of an IMP at a distance, but is complicated by the inclusion of a zero-th order term in the expansion for the even-subscripted p_i which represents the solution for $\mu_3 = 0$, where $\mu_4 < 3$ or $\eta < 0$. The perturbation expansion takes the form

$$\begin{aligned} p_0 &= A_0 + B_0\,\mu_3^2 + \cdots, \\ p_1 &= A_1\,\mu_3 + B_1\,\mu_3^3 + \cdots, \\ p_2 &= A_2 + B_2\,\mu_3^2 + \cdots, \\ p_3 &= A_3\,\mu_3 + B_3\,\mu_3^3 + \cdots, \\ p_4 &= A_4 + B_4\,\mu_3^2 + \cdots, \end{aligned} \tag{4.36}$$

where here the A_i, B_i coefficients are expected to be functions of η. The analysis proceeds as previously. First $\exp(Z)$ in (4.10) is expanded as a Taylor series in Z followed by substitution of Z defined by the second of (4.10), and then the p_i given by (4.36). This allows the moment integrals to be evaluated. The moment equations can then be constructed up to order μ_3, μ_3^3 for the odd-order moment equations and to order $\mu_3^0 = 1$ and μ_3^2 for the three even-order moments. Equating the respective coefficients of powers of μ_3 to zero in each of the five moment equations then gives ten equations for the ten coefficients $A_i, B_i, i = 0, \ldots, 4$. These are quadratically nonlinear but are more complex than the similar equation set obtained previously for Region I owing to the inclusion of zeroth-order terms in the even-indexed p_i coefficients in (4.36). As a consequence, without further simplification, these cannot be directly solved sequentially as was done above.

Progress can be made by use of a second expansion in terms of η of the form

$$\begin{aligned} A_0 &= a_{01}\,\eta + a_{02}\,\eta^2 \cdots, \\ A_2 &= a_{21}\,\eta + a_{22}\,\eta^2 \cdots, \\ A_0 &= a_{41}\,\eta + a_{42}\,\eta^2 \cdots. \end{aligned} \tag{4.37}$$

This allows sequential solution in terms of powers of η first for the a_{ij} defined in (4.37), then for $A1$, $A3$ and finally for the B_i. The calculation is straightforward but algebraically cumbersome and is not reproduced in detail. The solutions for the p_i

are

$$
\begin{aligned}
p_0 &= \tfrac{1}{8}(\mu_4-3) - \tfrac{1}{8}(\mu_4-3)^2 + \mu_3^2\left(-\tfrac{5}{24} + \tfrac{145}{96}(\mu_4-3) - \tfrac{3865}{384}(\mu_4-3)^2\right) + \cdots, \\
p_1 &= \mu_3\left(-\tfrac{1}{2} + \tfrac{7}{12}(\mu_4-3) - \tfrac{193}{96}(\mu_4-3)^2\right) + \mu_3^3\left(-\tfrac{55}{24} + \tfrac{1645}{72}(\mu_4-3) - \tfrac{208145}{1152}(\mu_4-3)^2\right) + \cdots, \\
p_2 &= -\tfrac{1}{4}(\mu_4-3) + \tfrac{7}{24}(\mu_4-3)^2 + \mu_3^2\left(\tfrac{1}{2} - \tfrac{179}{48}(\mu_4-3) + \tfrac{2375}{96}(\mu_4-3)^2\right) + \cdots, \\
p_3 &= \mu_3\left(\tfrac{1}{6} - \tfrac{1}{4}(\mu_4-3) + \tfrac{83}{96}(\mu_4-3)^2\right) + \mu_3^3\left(\tfrac{73}{72} - 10(\mu_4-3) + \tfrac{272315}{3456}(\mu_4-3)^2\right) + \cdots, \\
p_4 &= \tfrac{1}{24}(\mu_4-3) - \tfrac{1}{16}(\mu_4-3)^2 + \mu_3^2\left(-\tfrac{1}{8} + \tfrac{83}{96}(\mu_4-3) - \tfrac{2155}{384}(\mu_4-3)^2\right) + \cdots.
\end{aligned}
\tag{4.38}
$$

Higher-order moments can now be calculated. The fifth moment is

$$
\mu_5 = \mu_3\left(10 + 10(\mu_4-3) - \tfrac{65}{2}(\mu_4-3)^2\right) + \mu_3^3\left(-35 + \tfrac{1425}{4}(\mu_4-3) - \tfrac{203825}{72}(\mu_4-3)^2\right) + \cdots. \tag{4.39}
$$

The resulting expression demonstrates that in Region II, higher-order moments are non-singular due to the absence of the IMP.

Summary of Asymptotic Analysis

The results of our efforts are asymptotically correct solutions for the MER problem near the singular line (singular solution in Region I) and near equilibrium but away from the singular line (regular solution in Region II). A major issue with the results in terms of applying them to a generally valid solution is that there is no asymptotically correct solution which interpolates between the two regions, and there is no analysis of the exact boundary between the two regions or the way the two solution forms interpolate themselves along this boundary.

Nevertheless, we can make practical progress by defining an approximate but useful boundary and interpolation scheme for Region I which can analytically predict and account for the difficult-to-calculate IMP solution, while handling the central component of the solution numerically. Most useful in the following work then will be the results for μ_5 and the IMP in Region I, represented by x_1 and ϵ given in (4.34). In the following, we will examine the accuracy of the predictions for the IMP's contribution and incorporate them into a more robust version of the MER.

4.4 The Hybrid Maximum Entropy Reconstruction (HMER) Method

With the model of the IMP in Region I, it is possible to design a robust algorithm for computing the MER for a given set of five moments which avoids the problems

encountered with resolving integrals on the infinite domain. To facilitate this, we seek to determine an approximate boundary and an interpolation function for the IMP's contribution to the moments when they are found to be in Region I. Within this boundary we will model the IMP as in (4.34). Our approach is to subtract the contribution of the IMP from the given moments and solve numerically only for the central portion of the distribution. Since this central distribution is compact, we can approximate it with a solution on a finite domain, eliminating the resolution issues encountered with the infinite domain and making the algorithm robust. Since we combine analytic and numerical solutions to achieve a more robust result, we call this procedure the Hybrid Maximum Entropy Reconstruction (HMER). We detail the procedure throughout the following sections.

4.5 Approximation of the Region I Boundary

The Region I boundary can be identified with the location of a very sharp transition in the slope of the closing moment μ_5 with respect to μ_3. Figure 4.2 shows how the fifth moment of the full, infinite domain reconstruction behaves as a function of μ_3, in addition to the Region I and Region II asymptotic solutions for μ_5 and the HMER result of the following development. The solution shows a rapid change of derivative as it transitions from being dominated by the singular IMP to the regular central component. When μ_3 is sufficiently large, the distribution ceases to be bimodal and the central component of the solution takes over. The full numerical solution becomes badly conditioned as μ_3 approaches zero, resulting in noticeable noise-like deviations at small μ_3, but the HMER gives a smooth variation in μ_5 for any valid μ_3 or μ_4.

This region boundary appears to become asymptotic to a line at large μ_3, and is well fit by a quartic, hyperbola-like curve. It is difficult to see how to determine the curve analytically, as the asymptotic solutions in Regions I and II become poor approximations farther from equilibrium. For this reason, an empirical fit which seems to give good agreement with the location of the transition in comparing μ_5 is used, given by

$$(\mu_4 + 3)^4 = 6^4 \left(1 + 16\mu_3^4 + \frac{8}{5}\mu_3^2\right). \tag{4.40}$$

An interesting note is that this form of the region boundary intersects the physical realizability boundary given by

$$\mu_4 = \mu_3^2 + 1, \tag{4.41}$$

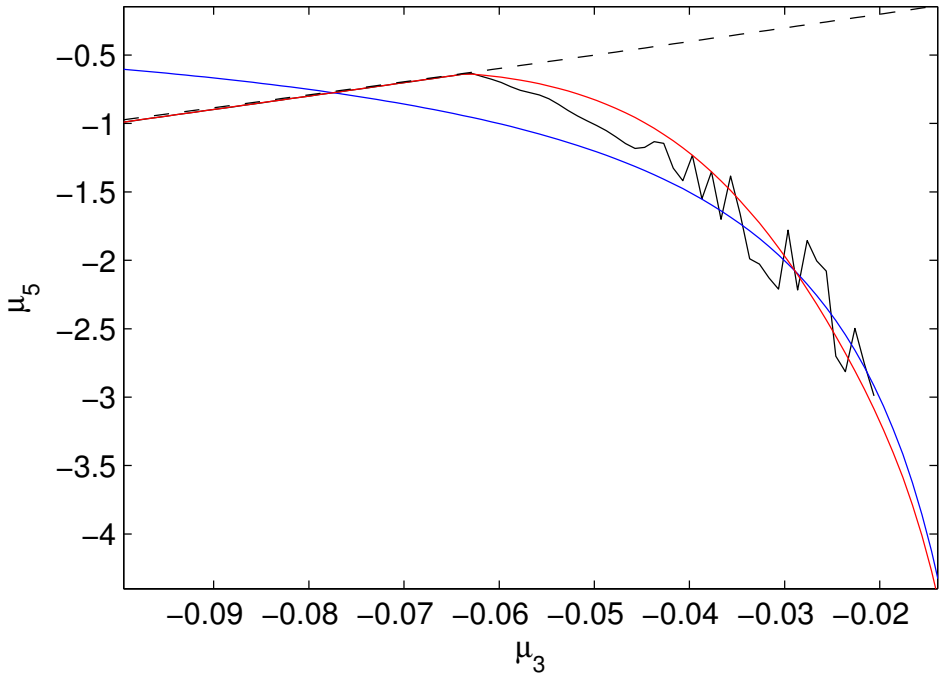

Figure 4.2: Comparison of HMER (Red) with Region I (Blue), Region II (Black dashed) asymptotic solutions, and numerical MER on the infinite domain (Black) for $\mu_4 = 3.01$. The full numerical result ends at $\mu_3 = -0.02$ because it frequently fails at small μ_3.

at approximately $(\pm 11.659, 136.93)$, implying that the IMP is always important for very high values of μ_4. This is little more than a curiosity in the context of this work, however, as this level of non-equilibrium is quite extreme and unlikely to be encountered in the types of flows we will examine.

4.6 The Itinerant Moment Packet (IMP) Model

Having identified the region in which the IMP should be included in our calculation of the MER, we can now apply the appropriate model from the Region I asymptotic solution, given by (4.34) and illustrated in Figure 3.4. For clarity, we again note that ϵ and x_1 are

$$\epsilon = \frac{\mu_4 - 3}{x_1^4}, \qquad x_1 = \frac{\mu_4 + 3}{\mu_3}. \tag{4.42}$$

The contributions of the IMP to the constrained moments are approximated by

$$\mu_{i,B} = \epsilon x_1^i, \tag{4.43}$$

and can be subtracted from the constrained moments to determine the moments of

the central portion of the distribution, to be solved numerically, as

$$\mu_{i,C} = \mu_i - \mu_{i,B}. \tag{4.44}$$

This serves as a first approximation, but as the contribution of the IMP in this scheme is non-zero at the Region I boundary, using the IMP as is would cause a discontinuity in the solution in (μ_3, μ_4)-space, both in the determined values of λ_i and in μ_5. This is undesirable as it is both inaccurate (no such discontinuity appears in the true MER solution) and presents problems when attempting to couple this reconstruction to a scheme for solving conservation equations, essentially creating a discontinuity in calculated fluxes. This is alleviated by including an attenuation function in the calculation of $\mu_{i,B}$,

$$\mu_{i,B} = w(\mu_3, \mu_4)\, \epsilon x_1^i. \tag{4.45}$$

This $w(\mu_3, \mu_4)$ should ideally be a correction factor to cause our scheme to perfectly match the value of μ_5 with the full distribution, but as no analytical solution exists for μ_5 as a function of μ_3 and μ_4 in general, the best that can be done is to use an approximate expression.

The expression used for $w(\mu_3, \mu_4)$ must satisfy the following conditions:

1. $w(\mu_3, \mu_4)$ must equal zero on the Region I boundary for continuity.
2. $w(\mu_3, \mu_4)$ must have a positive slope in μ_3 and μ_4 on the Region I boundary to simulate the "kink" seen in the full solution for μ_5.
3. $w(\mu_3, \mu_4)$ must approach 1 as $\mu_3 \to 0$ and as $\mu_4 \to \infty$ as the IMP dominates the solution for μ_5 in these limits.
4. $w(\mu_3, \mu_4)$ must approach 1 more rapidly as $(\mu_3, \mu_4) \to (0, 3)$, to accurately simulate the behavior of μ_5 in the full solution.

A reasonable choice that satisfies these terms is

$$w(\mu_3, \mu_4) = 1 - \exp\left[\frac{1}{2}(1 - \frac{\mu_4 - 3}{\mu_I - 3})\right], \tag{4.46}$$

with μ_I given by the μ_4 coordinate of the Region I boundary at the given μ_3, which can be solved for from the expression for the Region I boundary curve, (4.40), as

$$\mu_I - 3 = 6\left(+\sqrt[4]{1 + \frac{2}{5}(2\mu_3)^2 + (2\mu_3)^4} - 1\right), \tag{4.47}$$

using the positive fourth root. w is clearly zero on the Region I boundary (where $\mu_4 = \mu_I$) and approaches unity as $\mu_4 - 3$ becomes large relative to $\mu_I - 3$. As μ_3 approaches zero, μ_I approaches 3, so that for any $\mu_4 > 3$, the denominator in (4.46) approaches zero, causing w to approach 1. This also has the effect of increasing the rate at which w transitions from zero to unity as equilibrium is approached. Thus, the expression satisfies the terms outlined above. The detailed parameters are chosen to give a simple form and good fit of (4.45) to the value of μ_5 calculated by the full reconstruction.

Central Component of the Distribution

Having computed the required moments of the central component of the distribution with (4.44), the central portion of the distribution can be computed in a straightforward and robust way. Because the central portion of the distribution is always compact, a finite and consistent domain can be chosen for the required integrations. The domain is chosen such that the error in computing the highest required moment, μ_{2N-2}, due to truncating the domain to finite size is less than the error due to the discretization of the integral, estimated using a standard normal distribution.

The pure exponential form of the MER eliminates the possibility of using high-order global quadratures for the computation of the moments, as the distribution necessarily departs from any chosen weighting function in a way not well approximated by a polynomial factor. Therefore, composite, low-order quadratures are preferred. The composite Simpson's rule is a simple and effective choice. In this work, both Simpson's rule and standard Mathematica integration routines are used for computation of these integrals and the domain of solution for the central component is $[-14, 14]$.

The procedure for computing the central portion of the distribution consists of choosing an initial guess for the λ_i and proceeding with Newton's method until converging on a solution which matches the constraint moments. The required Jacobian is given in (3.2) as J_{ij}. The Newton's step is

$$\boldsymbol{\lambda_{n+1}} = \boldsymbol{\lambda_n} + \boldsymbol{J^{-1}}\left(\boldsymbol{\mu} - \boldsymbol{\mu_n}\right). \tag{4.48}$$

It is useful to improve the Newton's method by detecting large excursions due to the sensitive nature of the distribution function which can cause overflows leading

to program failure. The remedy is simply to include a backtracking function which catches large excursions and overflows in the calculation of the Jacobian or in the error at each step. When detected, this procedure backtracks along the previous increment in λ_i by reducing the increment magnitude by a constant factor, 5 in this work, and recalculating the offending terms. Occasionally this procedure must be iterated in very sensitive regions of the solution.

One further improvement can be made to improve the numerical stability of the method and increase convergence speed and hence overall computational efficiency. If the constraint moments are converted to moments of Chebyshev polynomials of equivalent order (specifically Chebyshev polynomials of the first kind, or so-called T polynomials), the reconstruction procedure is much better conditioned since the Chebyshev polynomials are much more independent over the domain compared to simple powers of x. The Chebyshev polynomials used are given by a recurrence relation:

$$\begin{aligned} T_0(x) &= 1, \\ T_1(x) &= x, \\ T_{n+1}(x) &= 2xT_n(x) - T_{n-1}. \end{aligned} \tag{4.49}$$

The transformation from standard moments, μ_i, to Chebyshev moments, τ_i, is given by

$$\tau_n = \int_\Omega T_n\left(\frac{x}{W}\right)\frac{dx}{W}, \tag{4.50}$$

where W is the chosen width of the solution domain for the central component, typically $W = 10$ to 14. Essentially, μ_i is used in place of x^i in the Chebyshev polynomials. The inverse of this transformation can be calculated by constructing the transformation matrix from (4.50) and taking its inverse.

While transforming the problem into its equivalent in Chebyshev moments requires some additional complexity in transforming the problem back and forth between Chebyshev and standard moments, but in testing it has shown to be an overall efficiency increase, speeding up typical HMER calculations by a factor of up to two. The cause of this speedup is the mollification of ill-conditioned Jacobian matrices, which has the effect of increasing the accuracy of each Newton's step and reducing the number of steps required overall. While this represents a useful time-saving technique, it is not necessary to compute the HMER robustly if highly non-

equilibrium solutions, such as those found near the physical realizability boundary, are not sought.

4.7 Results

The HMER procedure outlined here is an excellent approximation to the full numerical reconstruction. Figures 4.3, 4.4, 4.5, and 4.6 show how the HMER scheme compares with the full MER and the asymptotic solution in Region I. The full numerical MER becomes badly conditioned as μ_3 approaches zero, resulting in noticeable deviation in predicted higher moments at small μ_3, even though p_i values are seemingly smooth, because of the extreme sensitivity of higher moments to the values of some p_i. In fact, the full numerical solution on the infinite domain often fails for $|\mu_3|$ below some limit determined by the location of the quadrature points in the numerical integrations required for its computation. The hybrid scheme, on the other hand, is extremely robust owing to the elimination of the need for numerically computing the shape of the highly-sensitive IMP. Compare Figures 4.3, 4.4, and 4.5 with the transition point (the "kink") in Figure 4.6 to see that disturbances in some hybrid p_i values are due to the onset of the weighting function which interpolates along the boundary of Region I. Specifically, the even p_i display the greatest disturbance. This is not problematic since the values of the closing moments remain in good agreement with the infinite domain MER solution, but could possibly be improved upon with a different model of the IMP's behavior near the Region I boundary.

Figures 4.2 and 4.7 show a comparison of the two numerical methods, full and Hybrid, with Region I and II analytics. The HMER neatly interpolates between the two analytic solutions in a similar manner as the full numerical solution, but is completely smooth and unaffected by numerical difficulties present in attempting the full solution (the numerical noise in the full MER is evident in panel A of Figure 4.7), even in regions in which the full numerical solution fails completely (the black line in Figure 4.2 ends at $\mu_3 = -0.02$). This smoothness and robustness allows computations to be performed which would be impossible without addressing the analytic structure of the IMP, specifically allowing the procedure to handle a case in which a gas approaches or departs equilibrium from within Region I.

In practice, the only known limits to producing solutions with this procedure have been caused by the fact that, as the constrained moments approach the boundary of physical realizability, the MER degenerates into two very sharp peaks. In fact, in the

limit as the boundary is reached, the actual solution is a pair of delta functions. This causes issues with the fixed resolution of the numeric integration procedures, and eventually the singular nature of the near-delta functions produces problems even with adaptive integration schemes like Mathematica's in-built routines. Fortunately, this happens only very close to the physical realizability boundary. In the case of using the MER as part of a gas dynamic closure scheme, this region is not expected to be encountered. As an example, the lower limit for μ_4 in even distributions ($\mu_3 = 0$) is 1, and solutions can easily be calculated down to $\mu_4 = 1.0005$ before ill-conditioned Jacobians become problematic.

The full procedure is also relatively fast, considering its task. Even with an intentionally poor initial guess, such as a uniform distribution, convergence to machine precision usually takes fewer than seven steps. In practice, in gas dynamics usage, a very good initial guess is available from the previous time step or a neighboring cell or grid point. With such initial guesses, convergence usually takes only two or three iterations. The full numerical MER solution procedure on the infinite domain takes much longer, both due to the greater number of points needed to achieve sufficient resolution near the IMP, and due to the ill-conditioning encountered when in Region I.

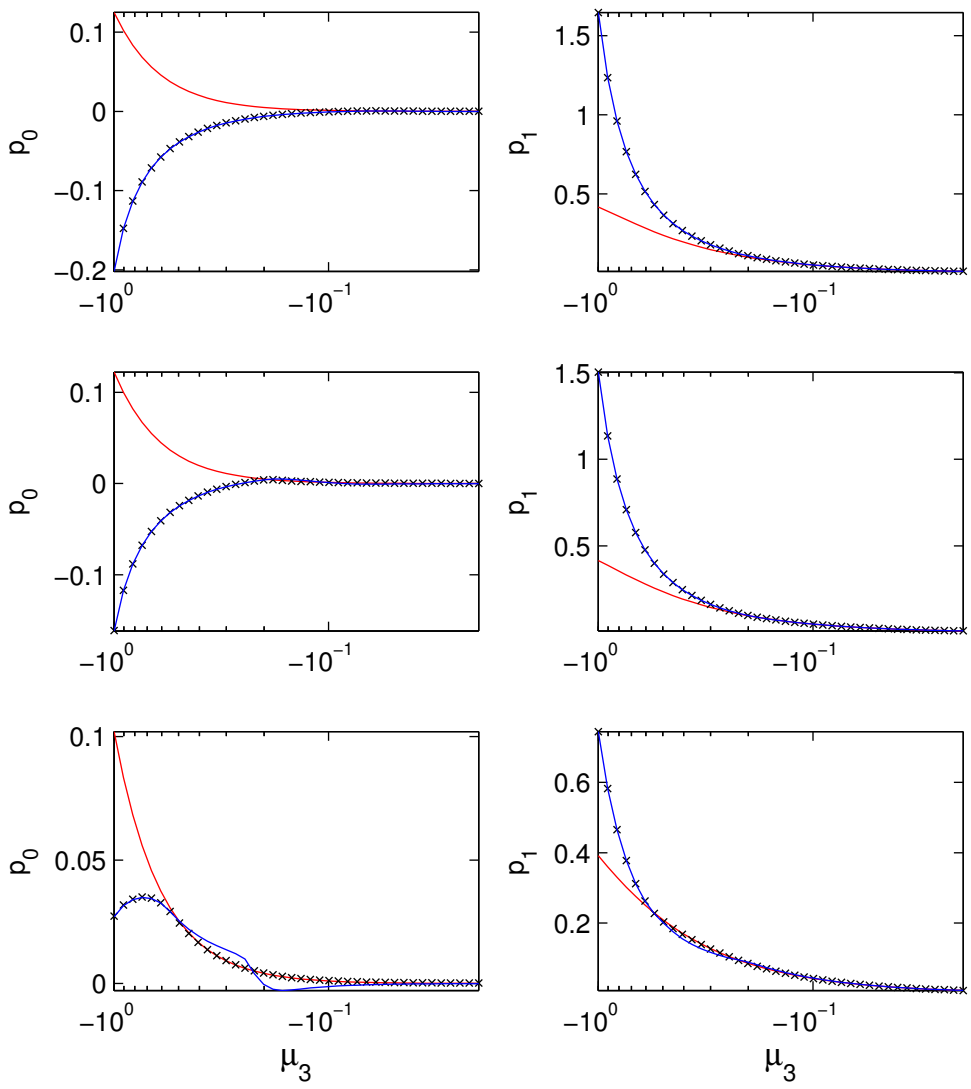

Figure 4.3: Comparison of p_0 and p_1 values at three values of μ_4. From top to bottom row, $\mu_4 = 3.01$, $\mu_4 = 3.1$, and $\mu_4 = 4$. Red lines: analytic result in (4.27), blue lines: HMER, black symbols: full MER on the infinite domain. μ_3 axes are logarithmic.

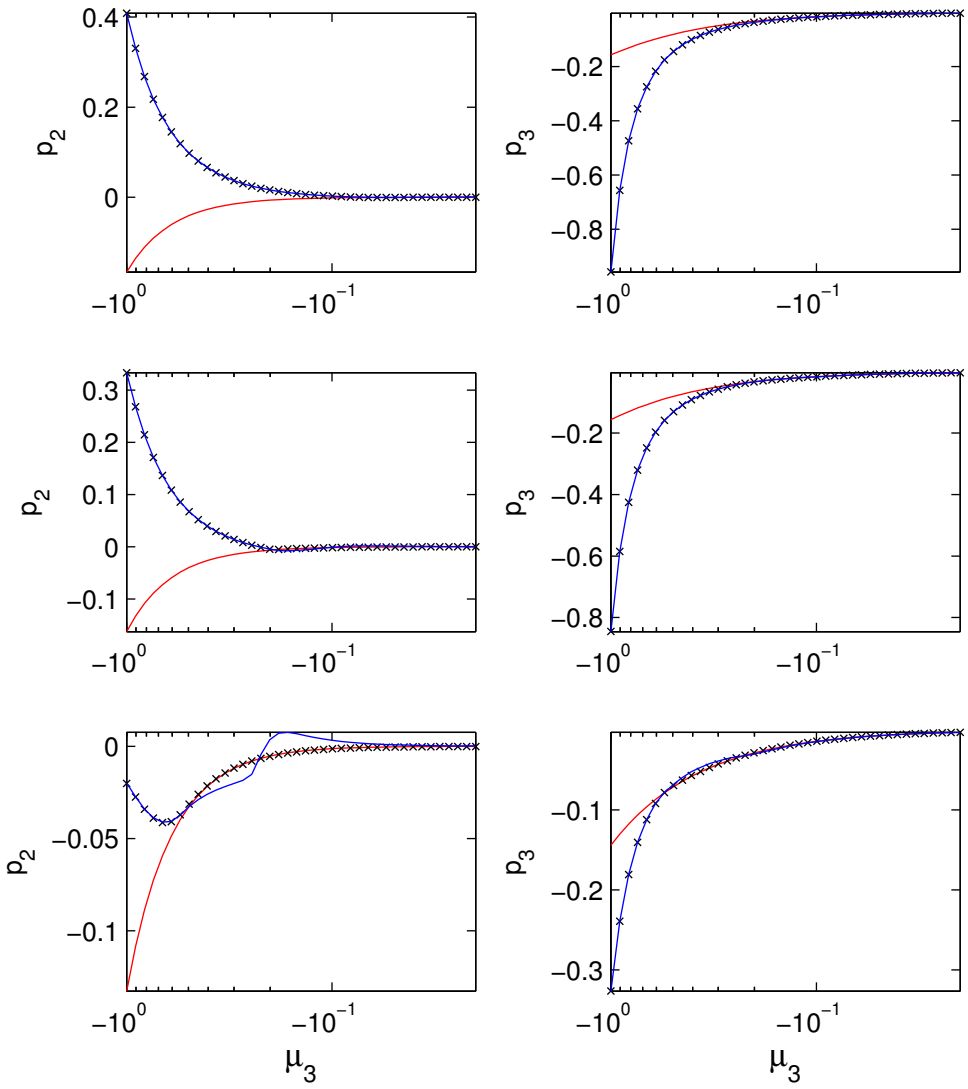

Figure 4.4: Comparison of p_2 and p_3 values at three values of μ_4. From top to bottom row, $\mu_4 = 3.01$, $\mu_4 = 3.1$, and $\mu_4 = 4$. Red lines: analytic result in (4.27), blue lines: HMER, black symbols: full MER on the infinite domain. μ_3 axes are logarithmic.

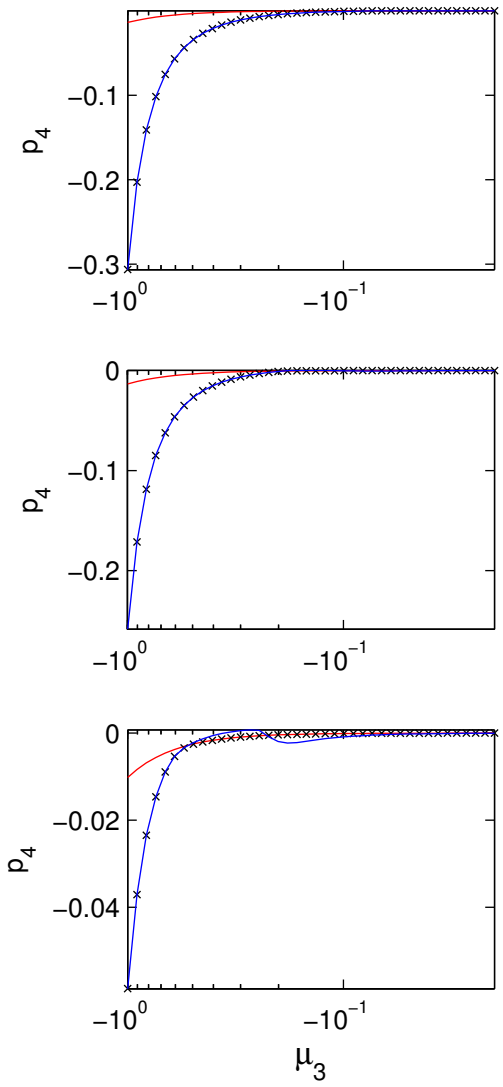

Figure 4.5: Comparison of p_4 values at three values of μ_4. From top to bottom row, $\mu_4 = 3.01$, $\mu_4 = 3.1$, and $\mu_4 = 4$. Red lines: analytic result in (4.27), blue lines: HMER, black symbols: full MER on the infinite domain. μ_3 axes are logarithmic.

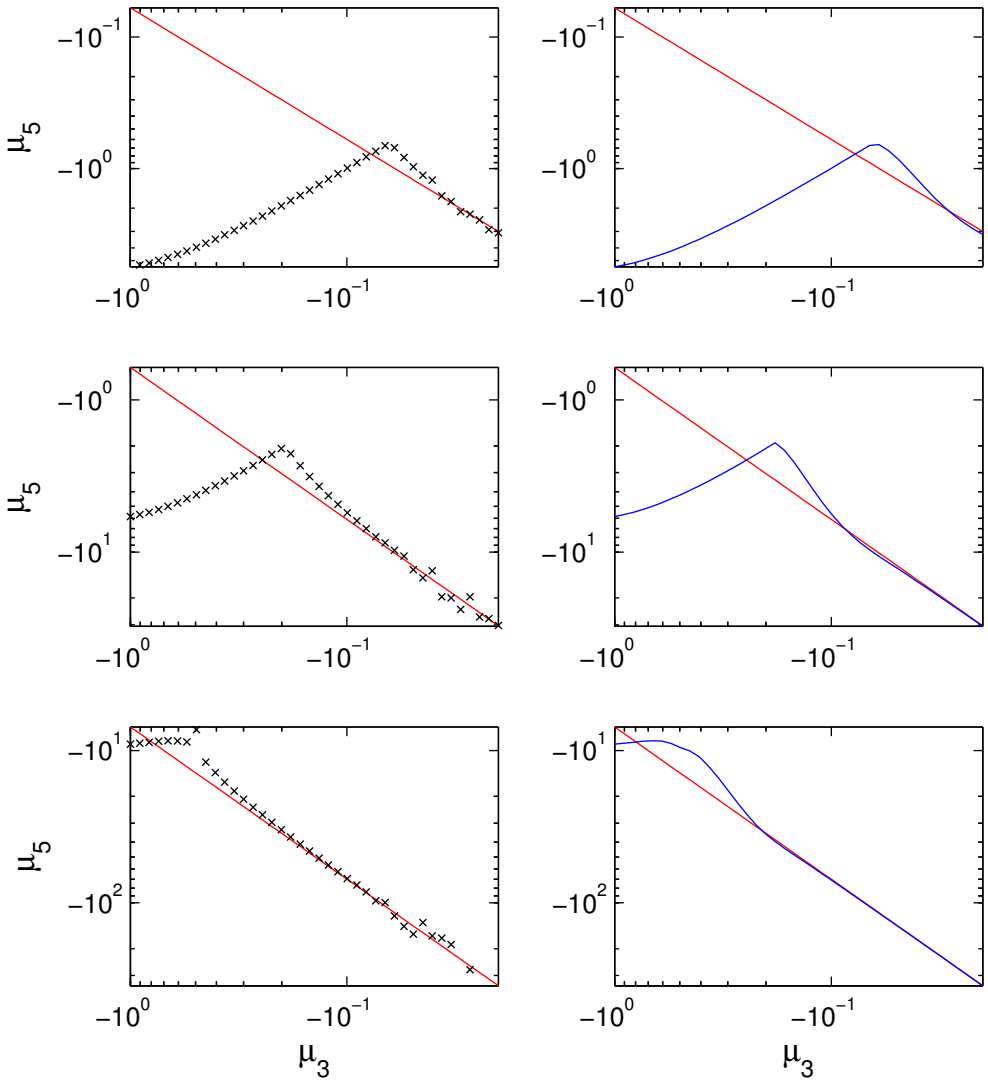

Figure 4.6: Comparison of first unconstrained moment, μ_5 at three values of μ_4. From top to bottom row, $\mu_4 = 3.01$, $\mu_4 = 3.1$, and $\mu_4 = 4$. Red lines: analytic result in 4.45, blue lines: HMER, black symbols: full MER on the infinite domain. μ_3 axes are logarithmic.

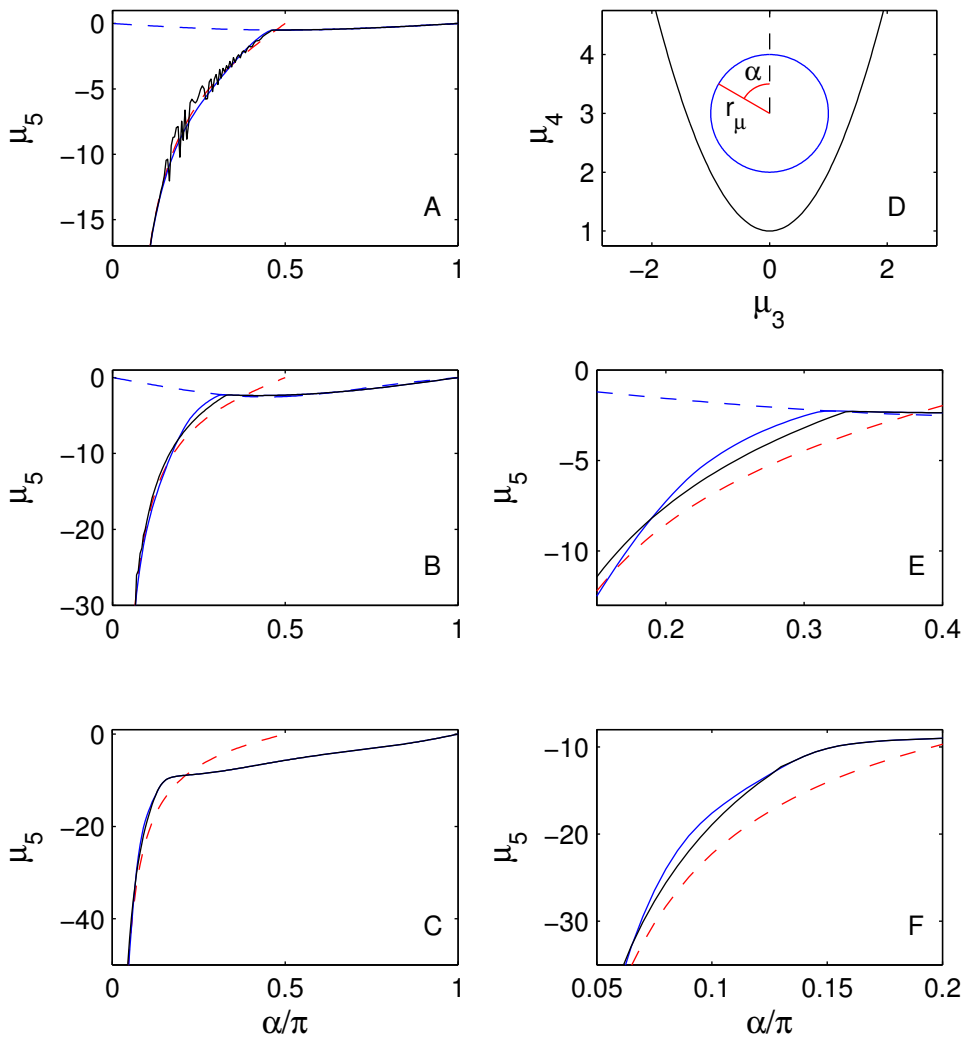

Figure 4.7: Comparison of first unconstrained moment, μ_5, on a circle around the equilibrium point in the (μ_3, μ_4)-plane (blue circle in panel D). The radius of the circle, r_μ is different for each panel: A: 0.05, B and E: 0.25, C and F: 1.3. The angle, α is measured from the singular line as illustrated in panel D. Panels E and F are details of panels B and C, respectively. In all panels except D, Black: full MER on the infinite domain, Blue solid line: HMER, Blue dashed line: Region II analytic solution, Red dashed line: Region I analytic solution. Note that the bottom panel does not include the Region II solution as it is well outside of its domain of validity here.

Chapter 5

A MAXIMUM ENTROPY GAS-DYNAMIC CLOSURE

5.1 The Boltzmann and BGK Equations

Recall that the Boltzmann Equation for a single-species monatomic gas with no body force (2.1) was developed in Section 1.1.

The Bhatnagar-Gross-Krook, or BGK, collision model first proposed in [5] is a simplifying approximation of the collision integral:

$$\frac{\rho^2}{m}\iiint_{\Omega}\int_0^{\pi}\int_0^{2\pi}\left(f_1' f' - f_1 f\right)\sigma c_r \sin\chi\, d\epsilon\, d\chi\, \boldsymbol{d\xi_1} \approx \frac{\rho}{\tau}\left(f_E - f\right). \tag{5.1}$$

Here, τ is a relaxation timescale which can be connected to bulk fluid parameters in the near-equilibrium limit by use of the Chapman-Enskog Distribution to be

$$\tau \equiv \frac{\mu}{\rho R T}, \tag{5.2}$$

where μ is the dynamic viscosity of the gas, not to be confused with a standard moment. For a hard sphere gas, [6] gives the viscosity as

$$\mu = \frac{5\rho\hat{\lambda}}{16}\sqrt{2\pi R T}, \tag{5.3}$$

where $\hat{\lambda}$ is the equilibrium mean free path of the gas, not to be confused with the Lagrange multipliers in the MER, which is proportional to the inverse of the density, ρ. This means that for a hard sphere gas

$$\mu \propto \sqrt{T}, \tag{5.4}$$

which will be useful later when computing τ in practice.

Using the BGK collision model in place of the full Boltzmann collision integral in (2.3) yields

$$\Psi_i^j \approx \frac{m_{i,e}^x - m_i^x}{\tau}, \tag{5.5}$$

where $m^x_{i,e}$ is the thermodynamic equilibrium value of m^x_i. When the BGK collision model is used in place of the integral collision operator from the original Boltzmann equation, the equation being solved is referred to as the BGK equation, given by

$$\frac{\partial(\rho f)}{\partial t} + \xi_i \frac{\partial(\rho f)}{\partial x_i} = \frac{\rho}{\tau}\left(f_E - f\right). \tag{5.6}$$

The full transport equation for an arbitrary moment in $\boldsymbol{\xi}$ in the BGK approximation becomes

$$\frac{\partial \langle\psi\rangle}{\partial t} + \frac{\partial \langle\xi_i\psi\rangle}{\partial x} = \frac{\langle\psi\rangle_E - \langle\psi\rangle}{\tau}. \tag{5.7}$$

5.2 Moment Equations

In order to relate these equations to our understanding of the MER problem, it is necessary to define

$$m_{ijk} \equiv \left\langle \xi_1^i \xi_2^j \xi_3^k \right\rangle. \tag{5.8}$$

and

$$n_{ijk|p} \equiv \left\langle \xi_1^i \xi_2^j \xi_3^k \xi_p \right\rangle = m_{(i+\delta_{p1})(j+\delta_{p2})(k+\delta_{p3})}. \tag{5.9}$$

These are related to the centered, scaled moments (from here on referred to as standard moments) by

$$\begin{aligned}
\mu_{ijk} &= \left\langle x^i y^j z^k \right\rangle = \frac{1}{\rho}\left\langle \left(\frac{\xi_1 - u_1}{\sigma_1}\right)^i \left(\frac{\xi_2 - u_2}{\sigma_1}\right)^j \left(\frac{\xi_3 - u_3}{\sigma_1}\right)^k \right\rangle \\
&= \frac{1}{\rho\sigma_1^{i+j+k}} \sum_{m=0}^{i}\sum_{n=0}^{j}\sum_{p=0}^{k} \binom{i}{m}\binom{j}{n}\binom{k}{p} \left\langle \xi_1^m \xi_2^n \xi_3^p \right\rangle (-u_1)^{i-m} (-u_2)^{j-n} (-u_3)^{k-p} \\
&= \frac{1}{\rho\sigma_1^{i+j+k}} \sum_{m=0}^{i}\sum_{n=0}^{j}\sum_{p=0}^{k} \binom{i}{m}\binom{j}{n}\binom{k}{p} m_{mnp} (-u_1)^{i-m} (-u_2)^{j-n} (-u_3)^{k-p}.
\end{aligned} \tag{5.10}$$

The inverse of this transformation is

$$\begin{aligned} m_{ijk} &= \left\langle \xi_1^i \xi_2^j \xi_3^k \right\rangle = \rho \left\langle (\sigma_1 x + u_1)^i (\sigma_1 y + u_2)^j (\sigma_1 z + u_3)^k \right\rangle \\ &= \rho \sum_{m=0}^{i} \sum_{n=0}^{j} \sum_{p=0}^{k} \binom{i}{m} \binom{j}{n} \binom{k}{p} \sigma_1^{m+n+p} \left\langle x^m y^n z^p \right\rangle u_1^{i-m} u_2^{j-n} u_3^{k-p} \\ &= \rho \sum_{m=0}^{i} \sum_{n=0}^{j} \sum_{p=0}^{k} \binom{i}{m} \binom{j}{n} \binom{k}{p} \sigma_1^{m+n+p} \mu_{mnp} u_1^{i-m} u_2^{j-n} u_3^{k-p}. \end{aligned} \tag{5.11}$$

The equations for the evolution of these moments from (2.3) are

$$\frac{\partial m_{ijk}}{\partial t} + \frac{\partial n_{ijk|p}}{\partial x_p} = \Psi. \tag{5.12}$$

These equations reveal the dependence of the fluxes of the highest order m_{ijk} on even higher moments, such as $n_{ijk|1} = m_{(i+1)jk}$, and, in the case of the full Boltzmann collision operator, the dependence of the collision terms on the full velocity distribution function. This results in a closure problem, wherein the system remains unclosed unless some method is imposed for determining the unclosed fluxes and collision terms from the primary moments.

5.3 The Hybrid Maximum Entropy Reconstruction Closure (HMERC) Scheme

In our case, we propose the use of the HMER to approximate the full velocity distribution function, which is then used to compute the unclosed fluxes and the collision terms (in the case of the Boltzmann collision operator only). This approach has the benefit of being fully extensible to any set of moments, including non-polynomial moments such as trigonometric moments or any generalized functions, though at present we focus on polynomial moments.

Additionally, as discussed in Section 3.2, the constraints on the domain of the moments are forgiving. Indeed, a solution to the maximum entropy problem exists for any set of moments that could possibly arise from a physical probability distribution, save for a zero-volume subset extending from, but not including, equilibrium [26]. Thus, so long as the closure scheme does not drive the system into an unphysical state or into a zero-density subset of physical states, the scheme should be able to proceed. We further expect that the absence of *ad hoc* or *ansatz* forms of the distribution will provide some benefit in reducing the paths for artificial features to develop in the solution, and indeed it will be seen that small discontinuities, or

sub-shocks, present in methods using such schemes are not present when using the HMER, though the solution is not without its own artifacts.

The collision terms are only unclosed in the full Boltzmann collision integral, in which case they may be calculated using the HMER distribution by any desired method, such as direct integration or Monte Carlo integration. The fluxes appear, of course, in derivative form, so the flux Jacobian may be required depending on the scheme used. It arises from use of the chain rule in calculating (for the x-flux only, in this case)

$$\frac{d\boldsymbol{n}_1}{dx} = \frac{\partial \boldsymbol{n}_1}{\partial \boldsymbol{m}} \frac{d\boldsymbol{m}}{dx} = \frac{\partial \boldsymbol{n}_1}{\partial \boldsymbol{\nu}} \frac{\partial \boldsymbol{\nu}}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\lambda}}{\partial \boldsymbol{\mu}} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{m}} \frac{d\boldsymbol{m}}{dx}, \tag{5.13}$$

where $\boldsymbol{\nu}$ here signifies the centered-scaled moments required to compose the uncentered moments $\boldsymbol{n}_1$. This looks daunting, but some terms become clear upon further inspection. The transformation Jacobians from n_{ijk} to ν_{ijk} and from μ_{ijk} to m_{ijk} are simply moment conversion matrices as in (5.10), and the Jacobian of $\boldsymbol{\lambda}$ with respect to $\boldsymbol{\mu}$ is the inverse of the Jacobian of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\lambda}$, which is (3.2), and so can be directly computed from the HMER. The final part is the Jacobian of fluxes ν with respect to λ, which is another moment Jacobian similar to (3.2). Many of the terms in this Jacobian should already be computed during the reconstruction procedure and so do not require integration, but some will be higher order and will require further integrations. When using the HMER, this Jacobian has both analytically and numerically calculated parts, so it is best to compute these two components separately. Once each component is constructed, the full flux Jacobian is available for calculation/estimation of eigenvalues and eigenvectors or other requirements of a specific numerical scheme.

Later it will become necessary to examine the one-spatial-dimensional case with moments arising from functions of a single velocity component and assumed symmetry about the x-axis; we will call this the velocity-factorizable moments case. Simplifying (2.3) for this case gives

$$\frac{\partial m_i^j}{\partial t} + \frac{\partial n_i^j}{\partial x} = \Psi_i^j, \tag{5.14}$$

where $n_i^j \equiv \left\langle \xi_j^i \xi_1 \right\rangle$. Here $j \in (\xi_x, \xi_r)$ is used to differentiate ξ_x and $\xi_r = \sqrt{\xi_y^2 + \xi_z^2}$ moments.

The equations may also be written in quasi-linear form, which is of use for many numerical schemes which require calculation of the flux-Jacobian,

$$\frac{\partial m_i^j}{\partial t} + A_{ik}^j \frac{\partial m_k^j}{\partial x} = \Psi_i^j, \tag{5.15}$$

where the flux-Jacobian is

$$A_{ik}^j = \frac{\partial n_i^j}{\partial m_k^j} = \frac{\partial n_i^j}{\partial \nu_m^j} \frac{\partial \nu_m^j}{\partial \lambda_n^j} \frac{\partial \lambda_n^j}{\partial \mu_p^j} \frac{\partial \mu_p^j}{\partial m_k^j}. \tag{5.16}$$

Some rows of $\boldsymbol{A^j}$ are simple, since for $i < N_j - 1$, N_j being the number of moments in direction j, n_i^j is equal to m_{i+1}^j. Also, it will be seen that when using this choice of moments, the HMER procedure gives the simplified flux for ξ_r moments as $n_i^{\xi_r} = u m_i^{\xi_r}$. Other rows, such as $A_{5k}^{\xi_x}$ for the truly unclosed term $n_5^{\xi_x} = \left\langle \xi^5 \right\rangle$, are general and cannot be simplified but must be computed, either numerically or by the analytic expression

$$A_{ik}^j = \frac{\partial n_i^j}{\partial \nu_m^j} \frac{\partial \nu_m^j}{\partial \lambda_n^j} \frac{\partial \lambda_n^j}{\partial \mu_p^j} \frac{\partial \mu_p^j}{\partial m_k^j} = \frac{\partial n_i^j}{\partial \nu_m^j} \mu_{m+n}^j \left(\boldsymbol{J^{j^{-1}}} \right)_{np} \frac{\partial \mu_p^j}{\partial m_k^j},$$

where $\boldsymbol{J^j}$ is the Jacobian of the reconstruction's moments $\boldsymbol{\mu^j}$ with respect to $\boldsymbol{\lambda}^j$.

To illustrate the structure of the flux-Jacobian for the velocity-factorizable moments case, it is simplest to expand it in matrix form for each j, leaving in derivative form the entries which have no simple expression.

Factorizable A^x and A^y:

$$\boldsymbol{A}^x = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ \frac{\partial n_5^x}{\partial m_1^x} & \frac{\partial n_5^x}{\partial m_2^x} & \frac{\partial n_5^x}{\partial m_{30}} & \frac{\partial n_5^x}{\partial m_{40}} & \frac{\partial n_5^x}{\partial m_{50}} \end{pmatrix}, \quad \boldsymbol{A}^y = \begin{pmatrix} u & 0 \\ 0 & u \end{pmatrix}. \tag{5.17}$$

Note that, although the form of (5.17) seems simple, solving for the eigenvalues analytically still requires the solution of a fifth-order polynomial and must be done numerically.

5.4 Relaxation to Equilibrium

As an illustration of the use of the HMER closure scheme, we examine the case of a spatially homogeneous gas undergoing relaxation to equilibrium from an initially non-equilibrium state. The obvious modification to (2.3) is that the spatial derivative term is uniformly zero, so fluxes are not needed. The distribution evolves solely due to the influence of the collision integral, which involves the velocity distribution function, f. Since f is not known, we impose the HMER closure, which uses the HMER to approximate the needed values of f based on the constraining moments. In this case we will focus on symmetric distributions with $\mu_i = 0$ for all odd i, so we can avoid any issue with the singular subspace so long as we maintain μ_4 below the singular line. This will not test the analytic portion of the HMER procedure, but it will demonstrate the MER as a closure scheme where it is well-behaved.

Finding a Suitable Test Problem

Using the full Boltzmann collision operator, there is no analytic solution in general. For the special case of a radially symmetric velocity distribution with

$$\sigma c_r = \kappa, \tag{5.18}$$

where κ is a constant, c_r is again the relative speed of the two molecules, and σ is the same collision cross section from (2.1), an analytic solution is possible as detailed in [33]. However, the solution for the fourth moment $\mu_4 = \langle c^4 \rangle$, and indeed all moments of even power, can be computed analytically, assuming an isotropic scattering. The final differential equations are closed, *i.e.*, the equation for μ_{2n} depends on μ_{2n-2}, μ_{2n-4}, and so on to μ_0, and so no closure scheme is needed. Since the closure scheme is unnecessary, this is not a good test of the HMER closure scheme and no comparison can be made with an analytic result.

Instead, we can examine the case for a hard sphere collision cross-section with

$$\sigma_T = 4\pi\sigma = \text{constant}, \tag{5.19}$$

where σ_T is the total cross section. To demonstrate that this problem is unclosed, it is necessary to begin with the collision integral for μ_4, given by (2.9) with $\psi = c^4$. Without loss of generality, we impose that the density, $\rho = 1$, and the velocity variance, $\sigma_c^2 = \mu_2 = 3$. We non-dimensionalize the time, t, by introducing the time parameter $\tau = \nu_c t$, where

$$\nu_c \equiv \frac{4}{\sqrt{\pi}} \sigma_T n \sqrt{\frac{\mu_2}{3}}, \tag{5.20}$$

which can be identified with Equation (4.53) in [6] as the mean frequency of collisions for particles in an equilibrium gas of hard spheres. Also, by symmetry, only moments which arise from the speed of a particle, not its velocity in any direction, need be considered. With these modifications, the equation for the change in μ_i with τ is

$$\frac{d\mu_i}{d\tau} = \frac{\sqrt{\pi}}{8} \iiint_{-\infty}^{\infty} \iiint_{-\infty}^{\infty} \int_0^{2\pi} \int_0^{\pi} \Delta\left(c^i\right) f f_1 \frac{c_r}{4\pi} \sin\chi \, d\chi \, d\epsilon \boldsymbol{dc_1 dc}, \tag{5.21}$$

where χ is the angle of deflection of the relative velocity vector during collision, ϵ is the angle between the plane in which this deflection occurs and a reference plane, and

$$\Delta\left(c^i\right) \equiv (c')^i - c^i + (c_1')^i - c_1^i. \tag{5.22}$$

We can simplify the integrals by noting a few features of (5.21).

Following Section 2.5 in [6], to compute post collision speeds for hard sphere collisions we need only the mean and relative particle speeds, c_m and c_r, and the angle between them following the collision. Using vector addition, the mean and relative speeds are given by

$$\begin{aligned} c_m^2 &= \frac{c^2}{4} + \frac{c_1^2}{4} + \frac{cc_1}{2}\cos\theta, \\ c_r^2 &= c^2 + c_1^2 - 2cc_1\cos\theta, \end{aligned} \tag{5.23}$$

with θ the angle between $\boldsymbol{c}$ and $\boldsymbol{c_1}$. Note that, since the directions of $\boldsymbol{c}$ and $\boldsymbol{c_1}$ are independent and isotropic, we can choose θ as a standard polar angle without loss of generality.

The post collision speeds are then given by

$$\begin{aligned} (c')^2 &= c_m^2 + \frac{c_r^2}{4} + c_m c_r \cos\chi, \\ \left(c_1'\right)^2 &= c_m^2 + \frac{c_r^2}{4} - c_m c_r \cos\chi. \end{aligned} \tag{5.24}$$

We note that post collision scattering is isotropic and does not affect c_m or c_r, so we can choose χ to be the angle between $\boldsymbol{c_m}$ and $\boldsymbol{c_r}$ without loss of generality.

Combining (5.24) with (5.23), inserting into (5.21), converting the velocity integrals to spherical coordinates, and performing the trivial integrals (over the three

azimuthal angles and over the polar angle of $\boldsymbol{c}$) which do not affect the integrand reduces the differential equation for μ_i to

$$\frac{d\mu_i}{d\tau} = \frac{\sqrt{\pi}}{8}(2\pi)^2 \int_0^\infty \int_0^\infty \int_0^\pi \int_0^\pi \Delta\left(c^i\right) f f_1 c_r c^2 c_1^2 \sin\chi \sin\theta d\chi d\theta dc_1 dc. \quad (5.25)$$

Note that the post collision speed expressions from (5.24), and therefore θ and χ, appear in the expanded form of Δ. Expanding the integrand and performing the angular integrals analytically leaves, in the case $i = 4$, the double integral

$$\frac{d\mu_4}{d\tau} = \frac{\sqrt{\pi^5}}{210} \int_0^\infty \int_0^\infty (A(+) - A(-)) f f_1 c^2 c_1^2 dc_1 dc, \quad (5.26)$$

where

$$A(\pm) \equiv \left(13c^2 \pm 30cc_1 + 13c_1^2\right) \frac{|c \mp c_1|^5}{cc_1}. \quad (5.27)$$

The expression A is clearly not directly a function of μ_4, being of sixth-order in velocity, so the problem is unclosed and the MER can be used to represent f for the purposes of numerically computing the integral.

For comparison, we use DSMC's collision procedure to calculate the evolution of μ_4. Essentially, the gas is represented by a number N_p of representative particles, in our case $N_p = 2 \times 10^5$. Each is given a random initial velocity with an initial distribution different from equilibrium. For each time step, a number of particles are chosen to undergo collisions, the scattering angle is randomly determined, and the particle velocities are updated. After each time step, the average value of μ_4 is computed over all particles, and the simulation proceeds.

The choice of particles for collisions at each time step is made according to Bird's "NTC" method, described in Section 11.1 of [6]. The method uses an acceptance rejection scheme which we describe simplified for hard-sphere molecules and the homogeneous relaxation case. First, a number, N_c, of collision partners are chosen at random, where

$$N_c = \frac{1}{2} N_p^2 F_N \sigma_T (c_r)_{\max} \frac{\Delta t}{V_c}. \quad (5.28)$$

Here, F_N is the ratio of real particles to simulated particles in the cell, $(c_r)_{\max}$ is the maximum encountered value of relative particle speed, Δt is the time step, and V_c is the spatial volume under consideration. Clearly, if n is the number density, then

$$F_N = \frac{nV_c}{N_p}. \quad (5.29)$$

The time step can be written in terms of the step in the time parameter $\tau = \nu_c t$, using the expression (5.20) and noting that $\mu_2 = 3$ as

$$\Delta t = \frac{\sqrt{\pi}\Delta\tau}{4\sigma_T n}. \tag{5.30}$$

The value of $(c_r)_{\max}$ is updated in the event that a larger value of c_r is encountered than the currently used value, and because of this, its initial value can be set to a reasonable value without affecting the results. In our simulations we use an initial value of $(c_r)_{\max} = 6\sqrt{2}$.

Using (5.29) and (5.30) in (5.28), we get

$$N_c = \frac{\sqrt{\pi}}{8} N_p (c_r)_{\max} \Delta\tau \tag{5.31}$$

as the number of collision partners. In order to choose collision pairs with the proper joint distribution, each of these potential collisions is then either accepted and performed on the particles or ignored, with the probability of acceptance given by $c_r/(c_r)_{\max}$.

The H-theorem ensures that equilibrium is the end state of the gas. At equilibrium, the even μ_i are

$$\mu_{2n,E} = (2n+1)!! \equiv \frac{(2n+1)!}{2^n n!} = 1, 3, 15, 105, 945, \ldots. \tag{5.32}$$

The initial state can be chosen from a constructed velocity distribution with radial symmetry, such as a shell or a sphere (a three-dimensional "top-hat" distribution). In our case, we use a pair of shells, with a PDF given by

$$f(c) = \frac{5}{24}\left(\delta\left(c - \sqrt{\frac{6}{5}}\right) + \delta\left(c - 3\sqrt{\frac{2}{5}}\right)\right)c^2, \tag{5.33}$$

where $\delta(c)$ is the Dirac delta function. It can be verified that this represents an isotropic gas with one quarter of particles having speed $c = \sqrt{6/5}$ and the remainder having speed $c = 3\sqrt{2/5}$, and this is indeed the method of assigning particle speeds in the DSMC calculation. The fourth moment of (5.33) represents the initial condition for the MERC case and is $\mu_4(\tau = 0) = 252/25$.

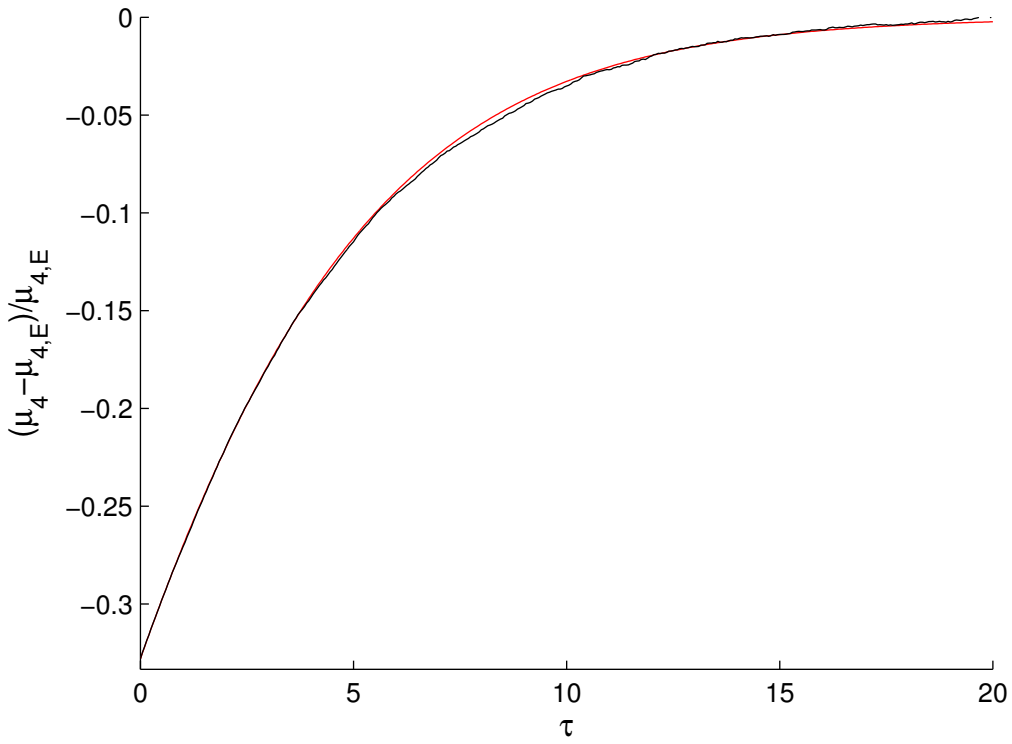

Figure 5.1: Relaxation comparison of the five-moment, spherically-symmetric MER closure with the DSMC prediction for the case of hard spheres. Black lines: Normalized DSMC μ_4. Red lines: Normalized MER solution for μ_4

HMER Results

To compare, we institute a simple Euler step to handle the time integration in (5.26). Given the solution's decaying-exponential form, this should ensure stability for a small enough time step. In our example, we use $\Delta\tau = 0.05$. The integration of the collision integral itself is handled by the same composite Simpson's rule which is used in the MER, with 100 equally spaced points between speed zero and ten, above which the likelihood of a collision is far too small for such particles to have an appreciable effect on the overall collision integral. We emphasize that this procedure results in a direct numerical solution for the full Boltzmann equation with no approximation, apart from the MER closure itself.

The case for $\mu_0 = 1$, $\mu_2 = 3$, and $\mu_4(0) = 252/25$ for the MER is shown in Figure 5.1. The results demonstrate excellent agreement between the MER closure scheme and DSMC solutions, especially considering the crude initial approximation of the distribution function by the MER, illustrated in Figure 5.2.

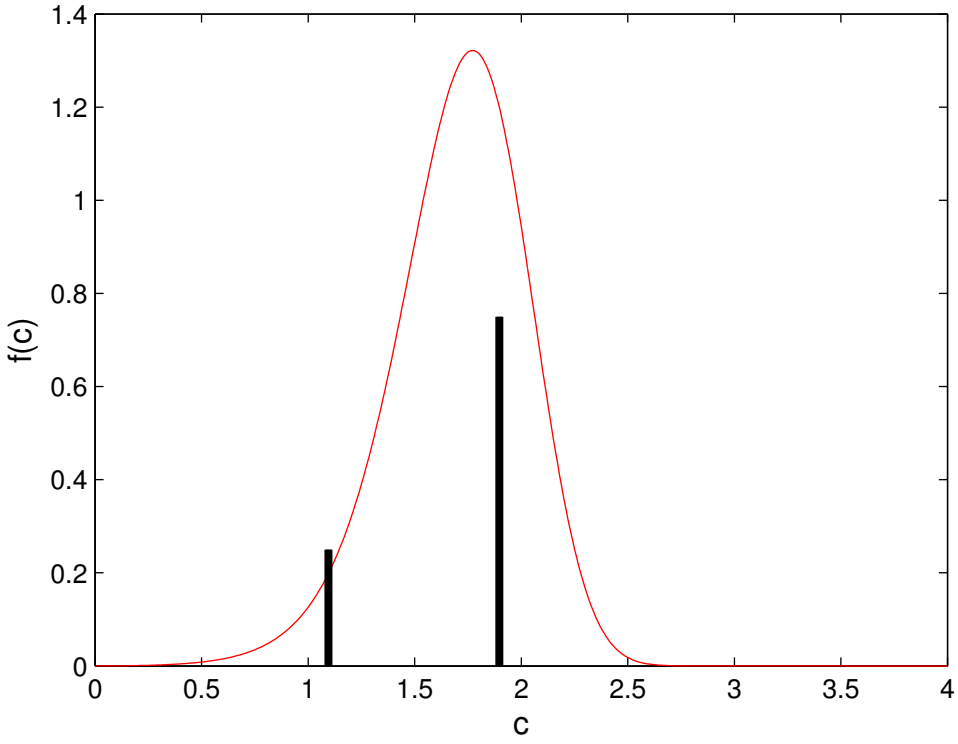

Figure 5.2: Initial distribution of relaxation example in Figure 5.1. Black bars: Percentage of particles in shell in true initial distribution. Red line: MER representation of true initial distribution.

Chapter 6

THE PLANAR SHOCK WAVE

6.1 One-dimensional Shock Wave Equations

In the case of a planar shock wave, the velocity distribution can be assumed to be axisymmetric in velocity space about the normal vector to the shock plane. All moments should be constructed of shock-normal, or streamwise, coordinate velocity ξ and shock-parallel, or lateral, coordinate radial velocity ζ. We will consider moments to fourth-order. We will use z to represent the spatial coordinate in the axis of symmetry direction, x to represent scaled, centered axial velocity, and r to represent scaled, centered radial velocity.

$$x = \frac{\xi - u}{\sigma_1}, \qquad r = \frac{\zeta}{\sigma_1}. \tag{6.1}$$

σ_1 here is the velocity variance in the shock-normal direction.

Further, in the frame of reference which moves with the shock, we seek a steady solution, so that all time derivatives vanish. What remains is a balance between the flux terms and the collision terms. Due to the use of the MER, it is useful to here to work with the primitive variables

$$\boldsymbol{U} = (\rho, u, \sigma_1, \mu_{30}, \mu_{40}, \mu_{02}, \mu_{12}, \mu_{22}, \mu_{04})^T, \tag{6.2}$$

with $\mu_{ij} \equiv \left\langle x^i r^j \right\rangle$. These moments represent all polynomial moments to fourth-order in velocity which are not trivially zero in axisymmetric flow. Their evolution equations are obtained by taking the corresponding uncentered moment integrands (the expression in angle brackets, with uncentered, unscaled velocity, $\xi^i r^j$ for μ_{ij}) for ψ in (5.7) to first arrive at

$$\frac{\partial m_{ij}}{\partial t} + \frac{\partial n_{ij}}{\partial z} = \frac{m_{ij,E} - m_{ij}}{\tau}, \tag{6.3}$$

where $n_{ij} = m_{(i+1)j}$ and $m_{ij,E}$ is the thermodynamic equilibrium value of the moment m_{ij} determined by ρ, u, and $T = \sigma_1\sqrt{(1 + \mu_0 2)/(3R)}$.

We then use (5.10) to expand these uncentered moments in terms of our chosen primitive variables and eliminate time derivatives to obtain

$$\frac{d}{dz}\left(\rho u\right) = 0, \tag{6.4}$$

$$\frac{d}{dz}\left(\rho\left(\sigma_1^2 + u^2\right)\right) = 0, \tag{6.5}$$

$$\frac{d}{dz}\left(\rho\left(u^3 + 3u\sigma_1^2 + \mu_{30}\sigma_1^3\right)\right) = \frac{2\rho}{3}\frac{\sigma_1^2\left(\mu_{02} - 1\right)}{\tau}, \tag{6.6}$$

$$\frac{d}{dz}\left(\rho\left(u^4 + 6u^2\sigma_1^2 + 4u\mu_{30}\sigma_1^3 + \mu_{40}\sigma_1^4\right)\right) = 2\rho\sigma_1^2\frac{2u\left(\mu_{02} - 1\right) + \mu_{30}\sigma_1}{\tau}, \tag{6.7}$$

$$\begin{aligned}\frac{d}{dz}&\left(\rho\left(u^5 + 10u^3\sigma_1^2 + 10u^2\mu_{30}\sigma_1^3 + 5u\mu_{40}\sigma_1^4 + \mu_{50}\sigma_1^5\right)\right)\\ &= \frac{\rho\sigma_1^2}{3}\frac{12u^2\left(\mu_{02} - 1\right) - 12u\mu_{30}\sigma_1 + \sigma_1^2\left(1 + 4\mu_{02} + 4\left(\mu_{02}\right)^2 - 3\mu_{40}\right)}{\tau},\end{aligned} \tag{6.8}$$

$$\frac{d}{dz}\left(\rho\left(u\mu_{02}\sigma_1^2 + \mu_{12}\sigma_1^3\right)\right) = \frac{\rho}{3}\frac{\sigma_1^2\left(1 - \mu_{02}\right)}{\tau}, \tag{6.9}$$

$$\frac{d}{dz}\left(\rho\left(u^2\mu_{02}\sigma_1^2 + 2u\mu_{12}\sigma_1^3 + \mu_{22}\sigma_1^4\right)\right) = \frac{\rho\sigma_1^2}{3}\frac{u\left(1 - \mu_{02}\right) - 3\mu_{12}\sigma_1}{\tau}, \tag{6.10}$$

$$\begin{aligned}\frac{d}{dz}&\left(\rho\left(u^3\mu_{02}\sigma_1^2 + 3u^2\mu_{12}\sigma_1^3 + 3u\mu_{22}\sigma_1^4 + \mu_{32}\sigma_1^5\right)\right)\\ &= \frac{\rho\sigma_1^2}{9}\frac{3u^2\left(1 - \mu_{02}\right) - 18u\mu_{12}\sigma_1 + \sigma_1^2\left(\left(1 + 2\mu_{02}\right)^2 - 9\mu_{22}\right)}{\tau},\end{aligned} \tag{6.11}$$

$$\frac{d}{dz}\left(\rho\left(u\mu_{04}\sigma_1^4 + \mu_{14}\sigma_1^5\right)\right) = \frac{\rho}{3}\frac{\sigma_1^4\left(\left(1 + 2\mu_{02}\right)^2 - 3\mu_{04}\right)}{\tau}. \tag{6.12}$$

The equations are in conservative form with source terms arising due to non-conservative collision integrals. The unclosed terms which must be related to the primary variables in (6.2) are μ_{50}, μ_{32}, and μ_{14}.

Boundary Conditions

The boundary conditions are given at $z = \pm\infty$ by the shock-jump relations for a given Mach number, M, and ratio of specific heats, $\gamma = \frac{5}{3}$ for monatomic gases. In this context it is convenient to normalize such that the upstream conditions are equilibrium ($\mu_{30} \to 0$ and $\mu_3^4 \to 3$) with density, temperature (T), and most probable mean speed ($\beta^{-1} \equiv \sqrt{2RT}$) equal to unity at $x = -\infty$. Under these conditions, the specific gas constant $R = \frac{1}{2}$, and the boundary conditions can be summarized as

$$\rho\left(z \to -\infty\right) = 1, \tag{6.13}$$

$$\rho\left(z \to +\infty\right) = \frac{M^2\left(\gamma+1\right)}{2 + M^2\left(\gamma-1\right)} = \frac{4M^2}{3+M^2}, \tag{6.14}$$

$$u\left(z \to -\infty\right) = M\sqrt{\frac{\gamma}{2}} = M\sqrt{\frac{5}{6}}, \tag{6.15}$$

$$u\left(z \to +\infty\right) = \frac{u\left(z \to -\infty\right)}{\rho\left(z \to +\infty\right)} = \sqrt{\frac{5}{6}}\frac{3+M^2}{4M}, \tag{6.16}$$

$$\sigma_1^2\left(z \to -\infty\right) = \frac{1}{2}, \tag{6.17}$$

$$\begin{aligned}\sigma_1^2\left(z \to +\infty\right) &= \frac{\left(2\gamma M^2 - (\gamma-1)\right)\left((\gamma-1)M^2+2\right)}{2M^2\left(\gamma+1\right)^2} \\ &= \frac{\left(5M^2-1\right)\left(M^2+3\right)}{32M^2},\end{aligned} \tag{6.18}$$

$$\mu_{30}\left(z \to \pm\infty\right) = 0, \tag{6.19}$$

$$\mu_{40}\left(z \to \pm\infty\right) = 3, \tag{6.20}$$

$$\mu_{04}\left(z \to \pm\infty\right) = 8, \tag{6.21}$$

$$\mu_{02}\left(z \to \pm\infty\right) = 2, \tag{6.22}$$

$$\mu_{12}\left(z \to \pm\infty\right) = 0, \tag{6.23}$$

$$\mu_{22}\left(z \to \pm\infty\right) = 2. \tag{6.24}$$

At this point, the equations, along with the HMER closure for the terms μ_{50}, μ_{32}, and μ_{14}, form a closed system of nine equations which can be solved. The basic

idea is to take an initial state for the primitive variables in (6.2), use this to calculate the HMER and thus obtain the three unclosed terms, μ_{50}, μ_{32}, and μ_{14}, compute derivatives and then solve to update the primitive variables for the next time step (in an unsteady solution) or the next grid point (in a steady state solution).

However, we can simplify the equations further. Since there is no source term in the continuity and momentum equations, (6.4) and (6.5), these can be integrated directly. The similar source terms in the two energy equations, (6.6) and (6.9), differ by a factor of -2, meaning the source terms in these two equations can be eliminated and the resulting equation integrated directly. These three integrals of motion correspond to mass, momentum, and energy conservation in the shock problem.

Using these results, three equations and three unknowns can be eliminated by integration, giving

$$\rho(u) = \frac{k_1}{u}, \tag{6.25}$$

$$\sigma_1(u) = +\sqrt{\frac{k_2}{\rho(u)} - u^2}, \tag{6.26}$$

$$\mu_{02}(u, \mu_{30}, \mu_{12}) = \frac{\frac{k_3}{\rho} - \mu_{12}\sigma_1^3 - \mu_{30}\sigma_1^3 - 3u\sigma_1^2 - u^3}{u\sigma_1^2}, \tag{6.27}$$

with the k_i being constants of motion, determinable at the upstream boundary condition as

$$k_1 = \rho_0 u_0, \tag{6.28}$$

$$k_2 = k_1^2 + \frac{1}{2}, \tag{6.29}$$

$$k_3 = k_1 (k_2 + 2). \tag{6.30}$$

Here, ρ_0 and u_0 are the upstream density and velocity, respectively.

With these simplifications, only six unknowns remain: $(u, \mu_{30}, \mu_{40}, \mu_{12}, \mu_{22}, \mu_{04})$. The necessary equations are (6.6), (6.7), (6.8), (6.10), (6.11), and (6.12). The unclosed moments all remain as before.

Note that all canonical variables (ρ, $T = \frac{2}{3}\sigma_1^2 (1 + \mu_{02})$) are parameterized by the velocity, u, while all other variables refer only to the characterization of the shape of the MER and the difference from equilibrium.

6.2 The Velocity-Factorizable Moment Approach

While (6.6) through (6.12) with the elimination equations and the HMER closure would represent a closed system, they rely on a multidimensional version of the HMER which has not been developed. We must specialize the shock wave problem to a case which can use the one-dimensional HMER we have developed in Chapter 3. We can do this by choosing a special set of moments which we call Velocity-Factorizable Moments (VFM).

First, take the example of computing a plane shock wave profile. This problem has symmetries which make all odd moments arising from transverse velocity components identically zero. We may choose two separate velocity integrand vectors, call them $\boldsymbol{\psi}_x(\xi_1)$ and $\boldsymbol{\psi}_r(\xi_2)$. We define

$$\psi_{x,i} \equiv \xi^i, \qquad i \in (0, 1, 2, 3, 4), \tag{6.31}$$

and

$$\psi_{r,j} \equiv \zeta^j, \qquad j \in (2, 4), \tag{6.32}$$

so that $\boldsymbol{\psi}_x$ contains powers of ξ_1 alone, and $\boldsymbol{\psi}_r$ contains even powers of ξ_2 alone. With this choice of moments, the maximum entropy reconstruction distribution (in the standardized moment context with centered, scaled moments) is

$$\hat{f}(x, r) = e^{\lambda_{x,i} x^i + \lambda_{r,2} r^2 + \lambda_{r,4} r^4}. \tag{6.33}$$

Since the terms in the exponential depend only on either x or r, but never both, the distribution is factorizable into an x and an r distribution which are independent. In this case, each set of Lagrange multipliers, $\boldsymbol{\lambda}_x$ and $\boldsymbol{\lambda}_y$, may be solved for independently using the constraint equations

$$\langle \boldsymbol{\psi}_x \rangle = \boldsymbol{m}_x, \tag{6.34}$$

and

$$\langle \boldsymbol{\psi}_r \rangle = \boldsymbol{m}_r, \qquad \langle 1 \rangle = 1. \tag{6.35}$$

Lastly, the flux term for each moment is calculated thus: For the moment $m_{x,i}$, the flux is simply the subsequent moment, $m_{x,(i+1)}$, except in the case of the highest

moment, $m_{x,5}$ in which case it is calculated from $\left\langle {\xi_1}^5 \right\rangle$. In the case of the moments $\boldsymbol{m}_y$, due to the separability of the integral, the fluxes are simply $u\boldsymbol{m}^y$.

This approach reduces the computational effort of the MER closure by eliminating expensive three-dimensional integrals from the reconstruction process. The only difference from the fully three-dimensional, mixed moment MER is the guarantee of independence of each velocity component in the reconstructed velocity distribution. This would seem to guarantee a flow with zero shear everywhere, making use of this method for other than very simple one-dimensional flows dubious, however the shear stresses may be included as moments in a VFM formulation of the problem if the distribution reconstruction is performed in the principal frame of the pressure tensor.

In order to use this approach, the moments μ_{200}, μ_{020}, μ_{002}, μ_{110}, μ_{011}, and μ_{101} must all be calculable from the constraint moments chosen. These form the pressure tensor as

$$\begin{pmatrix} \mu_{200} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{020} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{002} \end{pmatrix}, \tag{6.36}$$

which should remain positive definite if the solution is physical. We then calculate the eigenvalues and eigenvectors of this tensor. Since the pressure tensor is real and symmetric, the eigenvectors, suitably normalized, form a rotation matrix which allows us to rotate velocity components into the principle frame of the gas, where the pressure tensor becomes diagonal with the eigenvalues representing the principal stresses. In this frame shear is zero and once all moments are rotated into this frame, interpreting higher-order moments as flux vectors, we can perform the MER with our factorized approach, remembering to rotate the resulting flux vectors back into the original coordinate system.

To illustrate, take a fully three-dimensional flow and consider velocity-factorizable moments up to fourth order. After converting to the standard moments, these are

$$\mu_{j,i} = \left\langle x_j^i \right\rangle, \qquad j \in (1, 2, 3), \qquad i \in (0, 1, 2, 3, 4), \tag{6.37}$$

plus the three shear stresses

$$\mu_{110} = \langle x_1 x_2 \rangle, \qquad \mu_{011} = \langle x_2 x_3 \rangle, \qquad \mu_{101} = \langle x_3 x_1 \rangle. \tag{6.38}$$

The pressure tensor is

$$\boldsymbol{P} = \begin{pmatrix} \mu_{1,2} & \mu_{110} & \mu_{101} \\ \mu_{110} & \mu_{2,2} & \mu_{011} \\ \mu_{101} & \mu_{011} & \mu_{3,2} \end{pmatrix}, \tag{6.39}$$

and it can be decomposed into a diagonal matrix and a transformation matrix as

$$\boldsymbol{P} = \boldsymbol{Q} \begin{pmatrix} \mu'_{1,2} & 0 & 0 \\ 0 & \mu'_{2,2} & 0 \\ 0 & 0 & \mu'_{3,2} \end{pmatrix} \boldsymbol{Q^{-1}}, \tag{6.40}$$

where $\boldsymbol{Q}$ is the matrix of unit eigenvectors of P. Interpreting the three moments of each order as components of a flux vector, we can use Q to rotate each flux into the principal frame by

$$\begin{pmatrix} \mu'_{1,i} \\ \mu'_{2,i} \\ \mu'_{3,i} \end{pmatrix} = \boldsymbol{Q} \begin{pmatrix} \mu_{1,i} \\ \mu_{2,i} \\ \mu_{3,i} \end{pmatrix}, \tag{6.41}$$

for each $i \in (1, 3, 4)$. We then solve for the maximum entropy reconstruction using the rotated moments and then de-rotate the required fluxes back into the original frame as

$$\begin{pmatrix} \mu_{1,5} \\ \mu_{2,5} \\ \mu_{3,5} \end{pmatrix} = \boldsymbol{Q^T} \begin{pmatrix} \mu'_{1,5} \\ \mu'_{2,5} \\ \mu'_{3,5} \end{pmatrix}. \tag{6.42}$$

If the full distribution function is required, say, to compute the collision terms in the full Boltzmann collision integral, it can be computed by simply pre-rotating the requested velocities into the principal frame similarly to (6.41).

This approach offers a massively reduced computational cost, by reducing the integrals required for computing the reconstruction Jacobian from three-dimensional to one-dimensional. The drawback of this approach is a loss of cross-correlational

information regarding the distribution function, represented in the normal shock case by the mixed moments μ_{12} and μ_{22}. These moments appear in the fluxes of their x-inferior neighbor moments. Their most important role is in representing interactions between the shape of the distribution in x and the flux of radial velocity energy, represented by μ_{02}. It is important to note that μ_{02} features prominently in the collision terms for all equations (see the right hand sides of (6.6) through (6.12)).

6.3 The Velocity-Factorizable Planar Shock Wave

Velocity-Factorizable Moment (VFM) Equations

For the case of a planar shock wave, the full set of equations to be solved in the VFM case, using the same notation and procedure as in the axi-symmetric case, are obtained by combining Equations (5.7) and (5.11) with $\rho \equiv m_0$, $u \equiv \frac{m_{10}}{\rho}$, $\sigma_1^2 \equiv \frac{m_{20}}{\rho} - u^2$, and $\sigma_2^2 \equiv \mu_{02}\sigma_1^2$ to find

$$\frac{d}{dz}\left(\rho u\right) = 0, \tag{6.43}$$

$$\frac{d}{dz}\left(\rho\left(\sigma_1^2 + u^2\right)\right) = 0, \tag{6.44}$$

$$\frac{d}{dz}\left(\rho\left(u^3 + 3u\sigma_1^2 + \mu_{30}\sigma_1^3\right)\right) = \frac{2\rho}{3}\frac{\sigma_2^2 - \sigma_1^2}{\tau}, \tag{6.45}$$

$$\frac{d}{dz}\left(\rho\left(u^4 + 6u^2\sigma_1^2 + 4u\mu_{30}\sigma_1^3 + \mu_{40}\sigma_1^4\right)\right) = \frac{2\rho u\left(\sigma_2^2 - \sigma_1^2\right) - \rho\mu_{30}\sigma_1^3}{\tau}, \tag{6.46}$$

$$\begin{aligned}\frac{d}{dz}&\left(\rho\left(u^5 + 10u^3\sigma_1^2 + 10u^2\mu_{30}\sigma_1^3 + 5u\mu_{40}\sigma_1^4 + \mu_{50}\sigma_1^5\right)\right)\\ &= \frac{4\rho u^2\left(\sigma_2^2 - \sigma_1^2\right) - 4\rho u\mu_{30}\sigma_1^3 + \frac{\rho}{3}\left(\sigma_1^4 + 4\sigma_1^2\sigma_2^2 + 4\sigma_2^4\right) - \rho\mu_{40}\sigma_1^4}{\tau},\end{aligned} \tag{6.47}$$

$$\frac{d}{dz}\left(\rho u\sigma_2^2\right) = \frac{\rho}{3}\frac{\sigma_1^2 - \sigma_2^2}{\tau}, \tag{6.48}$$

$$\frac{d}{dz}\left(\rho u\mu_{04}\sigma_2^4\right) = \frac{\rho}{3}\frac{\sigma_1^4 + 4\sigma_1^2\sigma_2^2 + 4\sigma_2^4 - 3\mu_{04}\sigma_2^4}{\tau}. \tag{6.49}$$

For the hard sphere molecular model, assuming the upstream equilibrium mean free path is the unit length in z, we obtain from (5.2), (5.3), and (5.4)

$$\tau \equiv \frac{5}{8}\sqrt{\frac{\pi}{2}}\frac{1}{\rho\sigma}, \tag{6.50}$$

where σ^2 is the equilibrium velocity variance, $3\sigma^2 \equiv \sigma_1^2\left(1+\mu_{02}^2\right)$.

We note here that the first five of these equations are identical to (6.4) through (6.8). Here, μ_{50} is an additional unknown term which must be computed from the HMER. Including this there are eight unknowns, ρ, u, σ_1, σ_2, μ_{30}, μ_{40}, μ_{50}, and μ_{04} in these seven equations, (6.43) through (6.49).

Boundary Conditions

The boundary conditions are given at $z = \pm\infty$ by the shock-jump relations for a given Mach number, M, and ratio of specific heats, $\gamma = \frac{5}{3}$ for monatomic gases. In this context it is convenient to normalize such that the upstream conditions are equilibrium ($\mu_{30} \to 0$ and $\mu_{40} \to 3$) with density, temperature, T, and most probable mean speed ($\beta^{-1} \equiv \sqrt{2RT}$) equal to unity at $x = -\infty$. Under these conditions, the specific gas constant $R = \frac{1}{2}$, and the boundary conditions are identical to (6.13) through (6.22).

These equations can be solved as a system of seven, but are more conveniently reduced to three equations by first eliminating three variables. The first step is to use the same approach as in the full axisymmetric case, taking advantage of the conservation of mass, momentum, and energy to directly integrate (6.43), (6.44), and a combination of (6.45) and (6.48). The elimination rules are

$$\rho(u) = \frac{k_1}{u}, \tag{6.51}$$

$$\sigma_1(u) = +\sqrt{\frac{k_2}{\rho(u)} - u^2}, \tag{6.52}$$

$$\mu_{02}(u,\mu_{30}) = \frac{\frac{k_3}{\rho(u)} - \mu_{30}\sigma_1^3(u) - 3u\sigma_1^2(u) - u^3}{u\sigma_1^2(u)}. \tag{6.53}$$

The k_i are again constants of motion, determinable at the upstream boundary condition as

$$k_1 = \rho_0 u_0, \tag{6.54}$$

$$k_2 = k_1^2 + \frac{1}{2}, \tag{6.55}$$

$$k_3 = k_1 (k_2 + 2), \tag{6.56}$$

where ρ_0 and u_0 are again the upstream density and velocity, respectively.

The remaining equations are (6.45), (6.46), (6.47), and (6.49). Due to the independence of the x-velocity distribution from the r-velocity distribution, however, it can be seen that (6.49) is decoupled, since μ_{04} appears nowhere in any of the other three equations, and also cannot affect the value of μ_{50} produced by the reconstruction since μ_{50} depends on the x-velocity distribution and μ_{04} affects only the r-velocity distribution.

Therefore, the VFM approach reduces the problem to solving equations (6.45), (6.46), and (6.47) for the unknowns u, μ_{30}, and μ_{40}, using the HMER to close the system by giving μ_{50} as a function of the other three unknowns.

6.4 Perturbations Around Equilibrium

Motivation

Recall from Figure 3.1 that the equilibrium point is on the boundary of a singular subspace flanked by asymptotically infinite values of the closing flux. In the case which we have best analyzed, that of the five-moment (fourth-order) one-dimensional reconstruction, the closing moment near this singular subspace is dominated by the IMP's contribution,

$$\mu_{5,B} = \frac{\mu_4^2 - 9}{\mu_3}. \tag{6.57}$$

This term, as part of the flux in (6.47), must be differentiated. Notice that partial derivatives of (6.57) with respect to μ_4 and μ_3 are both singular as $\mu_3 \to 0$, resulting in unbounded terms in the flux Jacobian along any approach to equilibrium in Region I. This results in one or more unbounded eigenvalues in the flux Jacobian and therefore unbounded wave speeds and an inherently unstable calculation in unsteady simulations.

Because of this inevitable ill-conditioning, and because perturbed solutions will be needed for boundary conditions to produce a shock profile, since equilibrium is a

fixed point, if we are to examine solutions which approach or depart equilibrium with $\mu_{40} > 3$, it is necessary to analyze the behavior of the system in this region. Ideally, we would be able to understand the behavior of the system close to equilibrium well enough to compute valid boundary conditions that can be imposed sufficiently far from equilibrium to avoid the otherwise inevitable unstable behavior whenever equilibrium is approached from within Region I. We will proceed with an examination of the solution near equilibrium from a perturbation perspective and show that there exists only a finite Mach number window in which a consistent solution can be found that departs from the upstream equilibrium point producing a steady shock solution.

Perturbative Form of the Equations

Beginning with the reduced equations for a planar shock wave we have (6.45), (6.46), and (6.47) in the VFM case. We recall the boundary conditions, Equations (6.13)-(6.22). Both the full and velocity-factorizable sets of equations are of the form

$$\boldsymbol{A}(\boldsymbol{U})\frac{d\boldsymbol{U}}{dz} = \boldsymbol{F}(\boldsymbol{U}). \tag{6.58}$$

In order to proceed with the analysis, the HMER closure must be approximated by an analytical form for the closing moment, μ_{50}. This is approximated by the asymptotic results for both Region I and for Region II. In Region I, the IMP is dominant in the calculation of μ_{50} as equilibrium is approached, so its analytic form is substituted as an approximation for μ_{50}. The relation is

$$\mu_{50} = \frac{\mu_{40}^2 - 9}{\mu_{30}}, \tag{6.59}$$

which is singular and requires care in analysis. In Region II, there is no IMP and the closure to leading order is the more straightforward

$$\mu_{50} = 10\mu_{30}. \tag{6.60}$$

We substitute into (6.45), (6.46), and (6.47) perturbation expressions for each of the

three unknowns,

$$\begin{aligned} u(z) &= u_e + \epsilon u_p(z), \\ \mu_{30}(z) &= \mu_{30|e} + \epsilon \mu_{30|p}(z), \\ \mu_{40}(z) &= \mu_{40|e} + \epsilon \mu_{40|p}(z), \end{aligned} \tag{6.61}$$

with values with subscript e denoting an equilibrium point, upstream or downstream, and values with subscript p denoting perturbation from the equilibrium state. The perturbation parameter ϵ is taken to be very small, and Taylor expansions are performed for each term. We then collect terms in the equations for each order of ϵ. For the Region II case, the ϵ^0 term is automatically satisfied, but for the Region I case there is a residual term which must be set to zero, given by

$$\frac{\mu_{30|p}(z)\mu'_{40|p}(z) - \mu_{40|p}(z)\mu'_{30|p}(z)}{\mu^2_{30|p}(z)} = 0. \tag{6.62}$$

Here, $(\,)' \equiv \frac{d(\,)}{dz}$. This requires that

$$\mu_{40|p}(z) = \frac{\nu_{40}}{\nu_{30}} \mu_{30|p}(z), \tag{6.63}$$

where ν_{30} and ν_{40} are constants whose meaning will become clear later. The solution must leave equilibrium along a straight line in the (μ_{30}, μ_{40}) plane, consistent with a real-eigenvalued solution of a linearized ODE around equilibrium. If it does so with positive $\mu_{40|p}$, it will be in Region I, and with negative $\mu_{40|p}$, in Region II, which is inconsistent with the presumption of using Region I closure. Therefore, in Region I, we seek solutions with real eigenvalues and eigenvectors which point the solution toward positive $\mu_{40|p}$ when u is perturbed in the direction of the other boundary condition. It must be so, then, that we seek solutions in the Region II case which point towards negative $\mu_{40|p}$ when u is perturbed in the direction of the other boundary condition.

The first order in perturbation quantities gives a full system of equations to be solved for the perturbation values. For Region I, those equations are

$$\begin{pmatrix} 50\sqrt{\pi}M\left(9-5M^2\right) & 75M\sqrt{2\pi} & 0 \\ \sqrt{30\pi}\left(9+30M^2-25M^4\right) & 20M^2\sqrt{15\pi} & 15M\sqrt{\pi} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}\begin{pmatrix} u_p'(z) \\ \mu_{30|p}'(z) \\ \mu_{40|p}'(z) \end{pmatrix}$$
$$= \begin{pmatrix} 80\sqrt{30}\left(M^2-1\right) & -16\sqrt{15} & 0 \\ 240M\left(M^2-1\right) & -48M\sqrt{2} & 0 \\ 160\sqrt{30}\left(5M^4-2M^2-3\right) & -96\sqrt{15}\left(5M^2+1\right) & -240M \end{pmatrix}\begin{pmatrix} u_p(z) \\ \mu_{30|p}(z) \\ \mu_{40|p}(z) \end{pmatrix}, \tag{6.64}$$

where

$$A_{31} = 500M\left(9-25M^4\right) + 30\sqrt{15\pi}\left(9-25M^2\right)\frac{\mu_{40|p}}{\mu_{30|p}}, \tag{6.65}$$

$$A_{32} = 1250\sqrt{2\pi}M^3 - 75M\sqrt{2\pi}\frac{\mu_{40|p}^2}{\mu_{30|p}^2} - 30\sqrt{15\pi}\left(9-25M^2\right)\frac{u_p\mu_{40|p}}{\mu_{30|p}^2}, \tag{6.66}$$

$$A_{33} = 125\sqrt{30\pi}M^2 + 150M\sqrt{2\pi}\frac{\mu_{40|p}}{\mu_{30|p}} + 30\sqrt{15\pi}\left(9-25M^2\right)\frac{u_p}{\mu_{30|p}}. \tag{6.67}$$

A_{31}, A_{32}, and A_{33} contain terms which become singular as $\mu_{30|p} \to 0$ if (6.63) is not imposed. To continue in Region I, we seek exponential solutions of the type

$$\begin{pmatrix} u(z) \\ \mu_{30|p}(z) \\ \mu_{40|p}(z) \end{pmatrix} = \begin{pmatrix} 1 \\ \nu_{30} \\ \nu_{40} \end{pmatrix} e^{\Lambda z}, \tag{6.68}$$

with v_{ij} the eigenvector component of a linearized solution around the equilibrium state corresponding to μ_{ij}. Clearly this satisfies the zero condition with real solutions if there are real solutions in ν_{30}, ν_{40}, and Λ. The resulting equations for these quantities once (6.68) is inserted into (6.64) are

$$\Lambda\left(\frac{\nu_{30}}{2\sqrt{2}} + \frac{1}{6}\left(9-5M^2\right)\right) + \frac{4\nu_{30}}{5\sqrt{15\pi}M} = \frac{4\sqrt{\frac{2}{15\pi}}\left(M^2-1\right)}{M}, \tag{6.69}$$

$$\frac{4}{5}\sqrt{\frac{2}{\pi}}\nu_{30} + \Lambda\left(\frac{\nu_{40}}{4} + \frac{-25M^4+30M^2+9}{2\sqrt{30}M} + \sqrt{\frac{5}{3}}\nu_{30}M\right) = \frac{4\left(M^2-1\right)}{\sqrt{\pi}}, \tag{6.70}$$

$$\Lambda\left(-\frac{5}{6}\left(5M^4-9\right)+\frac{25\nu_{30}M^2}{6\sqrt{2}}+\frac{M\nu_{40}^2+\sqrt{\frac{6}{5}}\nu_{40}\left(9-25M^2\right)}{4M\nu_{30}\sqrt{2}}+\frac{5}{4}\sqrt{\frac{5}{6}}\nu_{40}M\right)$$
$$+\frac{2\nu_{40}}{5\sqrt{\pi}}+\frac{4\sqrt{\frac{3}{5\pi}}\nu_{30}\left(5M^2+1\right)}{5M}=-\frac{4\sqrt{\frac{2}{15\pi}}\left(5M^4-2M^2-3\right)}{M}. \tag{6.71}$$

The equations (6.69), (6.70), and (6.71) are three nonlinear, inhomogeneous equations for the three unknowns, ν_{30}, ν_{40}, and Λ. The exact solution of this system can be obtained algebraically, but the resulting expression for the unknowns is a highly complicated function of M and is better examined graphically. Following the same steps above with the Region II closure, and with both closures with downstream equilibrium values for u_e, gives similar perturbation equations for exponential solutions in those cases. Using the Region II closure, no order zero term survives and (6.62) is not required, but seeking solutions which do not enter Region I still restricts us to seeking real-valued solutions to the resulting equations.

The equations for Region I downstream perturbation eigenvectors and eigenvalues are produced by repeating the above analysis with the downstream boundary values used in place of upstream values and are given by

$$\Lambda\left(\frac{B^3}{\sqrt{2}M\left(32M^2+96\right)}\nu_{30}+\frac{5}{3M}-\frac{1}{M^3}\right)+\frac{B^2\left(5M^2-1\right)}{5\sqrt{15\pi}\left(M^2+3\right)^2}\nu_{30}$$
$$=\frac{16\sqrt{\frac{2}{15\pi}}B\left(1-M^2\right)}{\left(M^2+3\right)^2}, \tag{6.72}$$

$$\Lambda\left(\frac{1}{64}\sqrt{\frac{5}{3}}B^3\nu_{30}M+\frac{\nu_{40}\left(M^2+3\right)\left(1-5M^2\right)^2}{256M^2}\right.$$
$$\left.+\frac{175M^4+90M^2-153}{16\sqrt{30}M}\right)$$
$$+\frac{B^2\nu_{30}\left(5M^2-1\right)M^2}{10\sqrt{2\pi}\left(M^2+3\right)}=\frac{4B\left(1-M^2\right)M^2}{\sqrt{\pi}\left(M^2+3\right)}, \tag{6.73}$$

$$\Lambda\left(\frac{25B^3\nu_{30}\left(M^2+3\right)}{1536\sqrt{2}}+\frac{\frac{B^5\nu_{40}^2M^2}{1024\sqrt{2}\left(M^2+3\right)}+\frac{\sqrt{\frac{3}{5}}B^3\nu_{40}M\left(5M^2-21\right)}{64\left(M^2+3\right)}}{\nu_{30}}\right.$$
$$\left.+\frac{5\left(M^2+3\right)\left(55M^4-30M^2-9\right)}{96M^2}+\frac{5\sqrt{\frac{5}{6}}\nu_{40}\left(M^2+3\right)^2\left(1-5M^2\right)^2}{1024M^3}\right)$$
$$+\frac{\sqrt{\frac{3}{5\pi}}B^2\nu_{30}M\left(25M^4+30M^2-7\right)}{40\left(M^2+3\right)}+\frac{B\nu_{40}\left(1-5M^2\right)^2}{160\sqrt{\pi}}$$
$$=\frac{4\sqrt{\frac{2}{15\pi}}BM\left(3+2M^2-5M^4\right)}{M^2+3}, \tag{6.74}$$

where

$$B=\sqrt{5M^2-\frac{3}{M^2}+14}. \tag{6.75}$$

The equations for Region II upstream perturbation eigenvectors and eigenvalues are the same except for (6.71), which now contains the Region II closure relation and is

$$\Lambda\left(\frac{5}{6}\left(9-5M^4\right)+\frac{5\nu_{30}\left(5M^2+3\right)}{6\sqrt{2}}+\frac{5}{4}\sqrt{\frac{5}{6}}\nu_{40}M\right)$$
$$+\frac{2\nu_{40}}{5\sqrt{\pi}}+\frac{4\sqrt{\frac{3}{5\pi}}\nu_{30}\left(5M^2+1\right)}{5M}=\frac{4\sqrt{\frac{2}{15\pi}}\left(5M^4-2M^2-3\right)}{M}. \tag{6.76}$$

Finally, the equations for Region II downstream perturbation eigenvectors and eigenvalues are the same as (6.72), (6.73), and (6.74) for the Region I case, except again for (6.74) which contains the Region II closure relation and is now

$$\Lambda\left(\frac{5B^3\nu_{30}\left(5M^2+3\right)}{384\sqrt{2}}+\frac{5\left(M^2+3\right)\left(55M^4-30M^2-9\right)}{96M^2}\right.$$
$$\left.+\frac{5\sqrt{\frac{5}{6}}\nu_{40}\left(M^2+3\right)^2\left(1-5M^2\right)^2}{1024M^3}\right)+\frac{\sqrt{\frac{3}{5\pi}}B^2\nu_{30}M\left(25M^4+30M^2-7\right)}{40\left(M^2+3\right)}$$
$$+\frac{B\nu_{40}\left(1-5M^2\right)^2}{160\sqrt{\pi}}=\frac{4\sqrt{\frac{2}{15\pi}}BM\left(3+2M^2-5M^4\right)}{M^2+3}. \tag{6.77}$$

Figures 6.1 and 6.2 show solutions (real only) for the eigenvalue, Λ, as a function of Mach number for the various cases. The departure angle in radians is $\omega = \mathrm{Arg}(\pm\nu_{30} \pm i\nu_{40})$, where $\mathrm{Arg}(re^{i\theta}) = \theta \bmod 2\pi$ is the principal argument function and the $\pm$ sign is positive for downstream equilibrium and negative for upstream equilibrium, as u perturbations from equilibrium values are positive downstream and negative upstream. Solutions which use Region I analytics but depart from equilibrium in the negative-μ_{40} direction are inconsistent and represent false solutions, as do Region II solutions which depart towards positive-μ_{40}. We seek solutions with positive Λ at the upstream point and negative Λ at the downstream point in order to give decaying behavior in the corresponding limits.

Figure 6.1 reveals that at very low Mach numbers there is a Region II solution which is consistent for the upstream boundary, but most importantly there is the Region I solution which has a divergent eigenvalue at nearly $M = 2$. The actual value is approximately $M_C = 1.992$. Beyond this Mach number there exist no consistent solutions which depart from upstream equilibrium, or equivalently, there are no solutions for $M > M_C$ which decay as $z \to -\infty$. Notice that while the eigenvalue diverges and returns from negative infinity, the departure angle is smooth, as the eigenvectors do not vary strongly in this region. This, unfortunately, indicates that there is no solution for this set of equations which satisfies the upstream boundary condition above the critical Mach number.

Figure 6.2 reveals that there are consistent Region II solutions at the downstream bondary for all Mach numbers in this range, and there are consistent Region I solutions up to a Mach number of approximately 1.53. In practice, the Region I

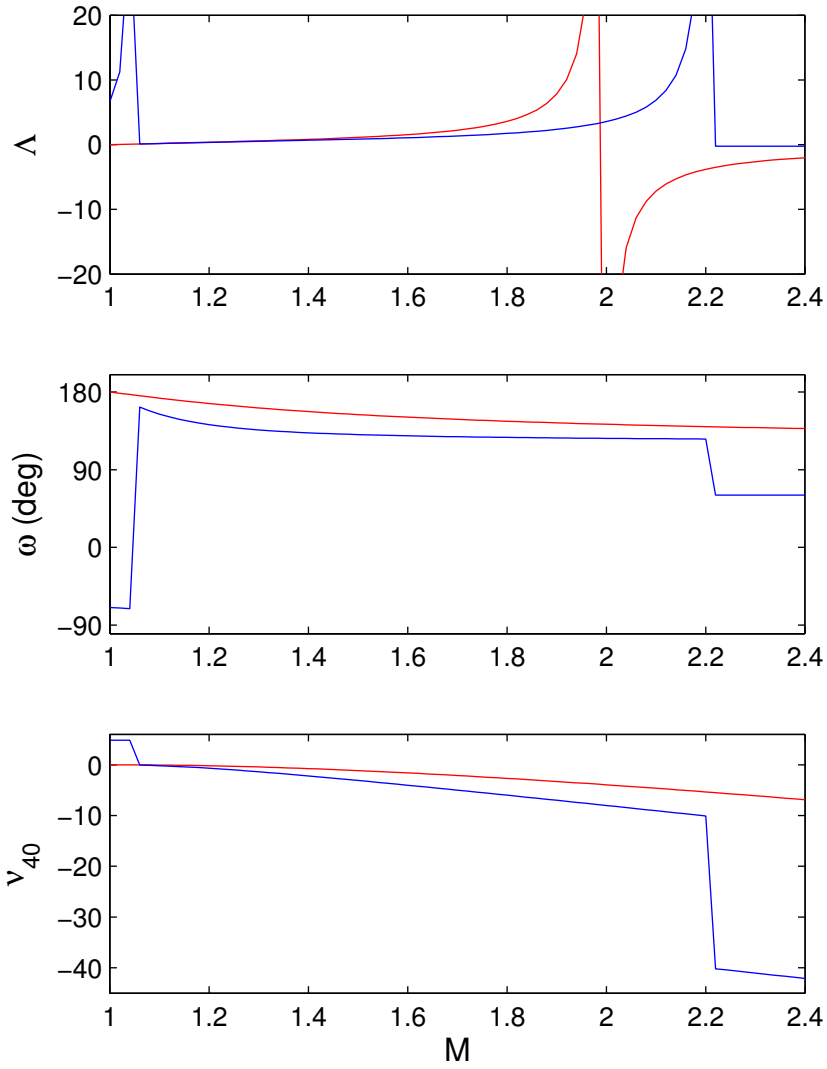

Figure 6.1: Upstream equilibrium linearized solutions for Region I (red) and Region II (blue). Shown are the most positive eigenvalues (in this case, these are the only positive, real eigenvalues), their departure angle from equilibrium in the (μ_{30}, μ_{40}) plane in degrees (assuming negative velocity perturbation), and the value of ν_{40}, which should be negative for consistency in Region I and positive for consistency in Region II.

solutions are preferred by the numerical solution as they can connect with incoming solutions from the upstream boundary which pass through a peculiar singularity.

6.5 Mach Number Solution Window

It is an unfortunate revelation of these results that for Mach numbers above $M_C = 1.992$ in the VFM case no consistent solution exists which corresponds to departure from equilibrium in a steady state. This is certainly a major fault of the method,

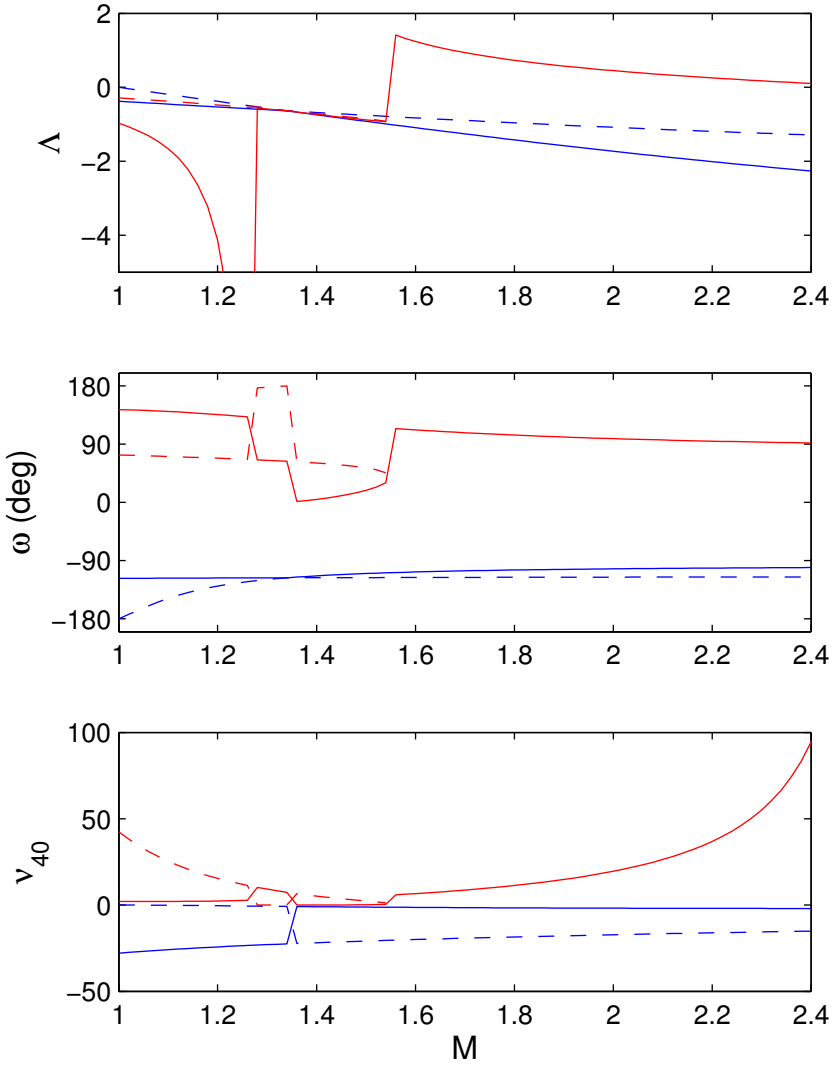

Figure 6.2: Downstream equilibrium linearized solutions for Region I (red) and Region II (blue). Shown are solutions for the most negative eigenvalues (solid lines) and the second-most negative eigenvalues (dashed lines). The departure angle from equilibrium in the (μ_{30}, μ_{40}) plane is in counterclockwise degrees from the μ_{30}-axis and positive velocity perturbations are assumed from the downstream equilibrium point). ν_{40} is the eigenvector component for μ_{40}, which should be positive for consistency in Region I and negative for consistency in Region II.

effectively eliminating it (in current form) as a general method for gas dynamic simulation. Nevertheless, the result raises interesting questions.

Why does a method based on what would seem the naturally-assumed velocity distribution not allow physical solutions? The MER is the most unbiased estimate of the velocity distribution, and in some sense is linked to the Boltzmann H-Theorem. In fact, given the H-theorem, one would expect the maximum entropy distribution to be exactly the distribution achieved in an actual gas if higher moments could somehow be physically constrained analogously to those constrained by physical conservation laws, since the collision integral in that case would always be working to bring the distribution to its maximum entropy form given the constraints.

Nevertheless, we have shown that the MER has the peculiar IMP, which seems entirely artificial. It is possible that our choice of third and fourth order moments as additional constraints on the distribution itself is the artificial element. It is difficult to understand how this could be the case however, as it seems the most natural choice of additional constraints, and [27] shows that polynomial moments are necessary to preserve certain physical features of the moment system.

One can conceive of a thought experiment wherein the moments to fourth order of velocity distribution are very accurately measured everywhere in a gas and then the MER is used to compute the velocity distribution everywhere. From this distribution we compute the fluxes and effect of collisions and advance a simulation forward some time. Given that we have shown good evidence for the non-existence of steady solutions for certain cases, this approach would likely not yield the correct or even any steady state solution in all cases. This would seem to indicate that there is some information not being carried by the constrained moments that is vital to the physical behavior of the system. What is this information and where is it stored in the distribution? These and other puzzles we leave to future work on this problem.

It is possible that the solution window is an artifact of the VFM approach itself. In eliminating the two cross-moments, μ_{12} and μ_{22} from the solution, it is possible that the solution is constrained in an unnatural manner. A thorough investigation of this, however, must wait until a similar asymptotic analysis can be made to give a sensible estimate of the additional closing fluxes, μ_{32} and μ_{14}, in the axisymmetric MER.

6.6 Factorized Plane Shock Wave Solution

In the case of a planar shock wave with velocity-factorizable moments, the equations to be solved are (6.45), (6.46), and (6.47), with (6.25), (6.26), and (6.53). The ultimate boundary conditions are given by (6.13), (6.14), (6.17), (6.18), (6.23), (6.24), (6.19), and (6.20), however, since the actual boundary conditions are fixed points, and also due to the singularity in the maximum entropy reconstruction, it is necessary to start with an approximate boundary condition derived from the form (6.68) which is some distance from the equilibrium point in state space. In this case, the solution chosen at the upstream boundary corresponds to the most positive eigenvalue, Λ, with eigenvectors forcing the solution into Region I. The initial z location in (6.68) is our choice, and we choose such that $\exp(\Lambda z) = 0.0001$. This provides us enough standoff distance in our computational domain to see the upstream tail of the shock clearly.

To give an example, for a shock with $M = 1.5$, the appropriate solution is approximately

$$\begin{pmatrix} u(0) \\ \mu_{30}(0) \\ \mu_{40}(0) \end{pmatrix} = -0.0001 \begin{pmatrix} 1 \\ 2.30378 \\ -1.11807 \end{pmatrix}. \tag{6.78}$$

The system is a set of three non-linear ODEs which can be solved using standard numerical packages. In all following examples, Mathematica's NDSolve routine is used with absolute and relative error tolerances fixed at 10^{-8}. A Mathematica version of the HMER procedure detailed in Chapter IV was written to provide the unclosed terms. Figure 6.3 demonstrates the shock profile generated by the method at $M = 1.5$. The factorized method shows an eye-smooth profile with no subshocks and a peak in streamwise temperature as predicted by theory. Notice that the profiles for ρ and σ_2 are very similar. This is because in the velocity-factorizable case, only the collision terms and bulk transport affect σ_2 due to the lack of cross-correlation between shock-normal and shock-parallel velocities.

6.7 The Pseudo-equilibrium Point

Although the shock profile in Figure 6.3 looks smooth to the eye, it was not obtained by direct integration from the upstream boundary condition to the downstream condition. There exists a nearly imperceptible discontinuity in the derivatives of the profile at approximately $z = 1.7$. These discontinuities are visible in the profiles of

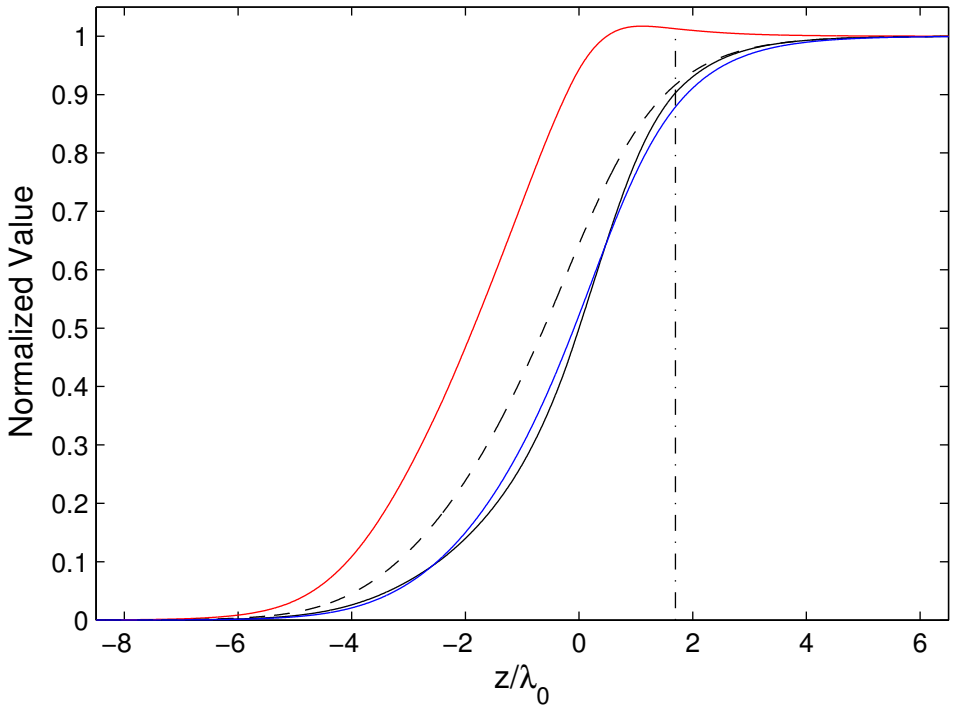

Figure 6.3: Shock profile for $M = 1.5$ using the HMERC. Black line: ρ. Black dashed line: T. Red line: σ_1. Blue line: σ_2. All values are normalized over their range from upstream to downstream quantities. The dashed-dotted line indicates the location of a defect in the solution called the pseudo-equilibrium point.

the velocity derivative plotted in Figure 6.4. The point at which this occurs coincides with a premature return to equilibrium for μ_{30} and μ_{40}, best visualized by Figure 6.5, which shows μ_{40} plotted against μ_{30} over the full shock wave. In this (μ_{30}, μ_{40})-plane plot, the trajectory follows the upper portion of the left lobe, and then returns to equilibrium along the lower portion before departing and returning along almost the same vector in the right-half plane. Comparing profiles of μ_{30}, μ_{40}, and u in Figure 6.6 reveals that at the point that pseudo-equilibrium is reached in terms of μ_{30} and μ_{40}, u has yet to reach its downstream equilibrium state, driving the solution away from equilibrium again. In light of this, we call this point the pseudo-equilibrium point. It will be shown that the behavior of this point is connected to the singular structure of the IMP. Careful handling of this point is important to obtain a full shock profile, as any attempt at direct integration will likely fail as it reaches this point due to the singular nature of the closing flux.

The HMER at this point is a Gaussian in the shock normal direction. We note again the multivalued nature of the closing flux, μ_{50}, at the point E. Consider a solution

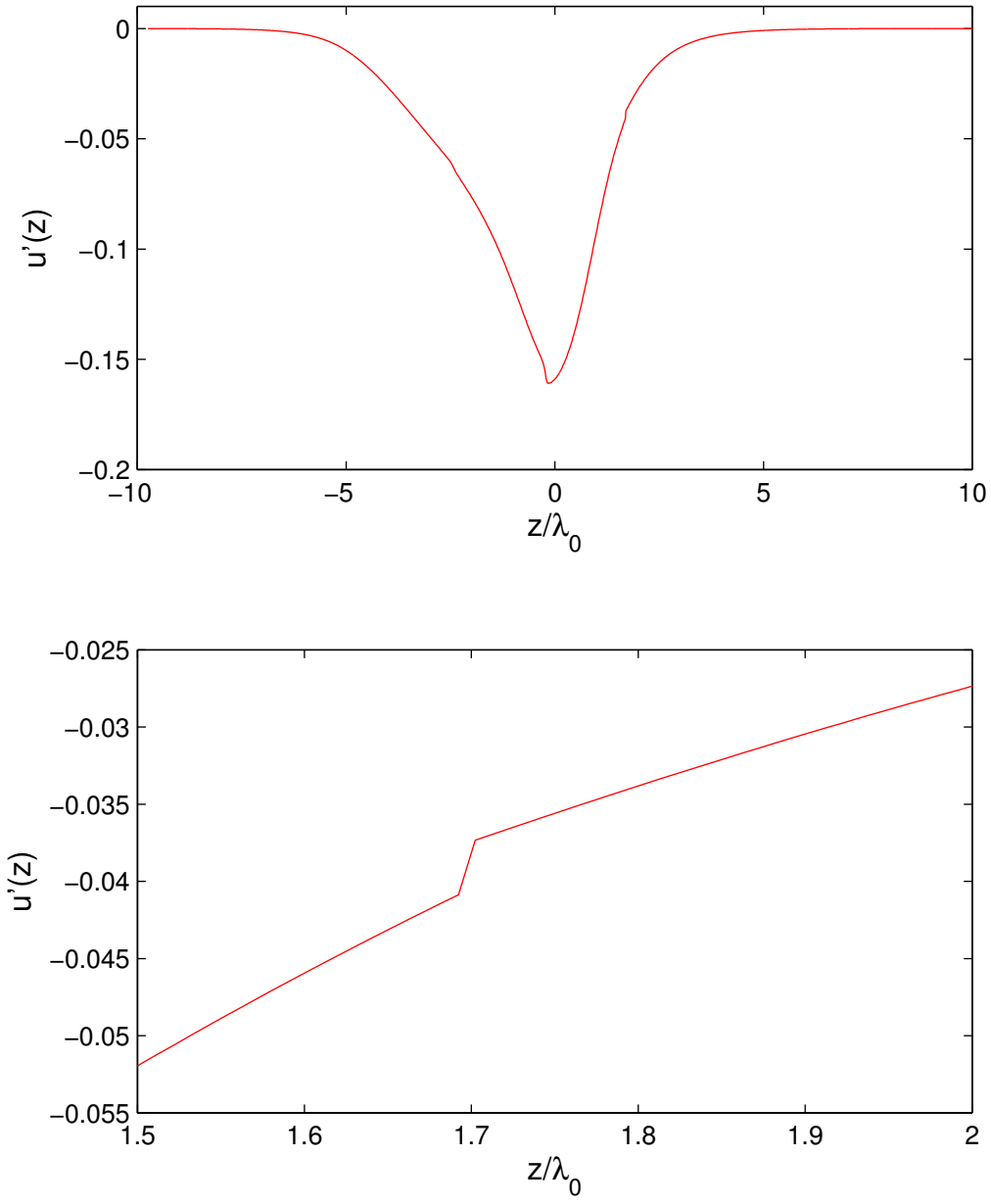

Figure 6.4: Top: $u'(z)$ in the $M = 1.5$ shock profile. Bottom: Detail of top figure around the pseudo-equilibrium point. Note the discontinuity at $z \approx 1.7$.

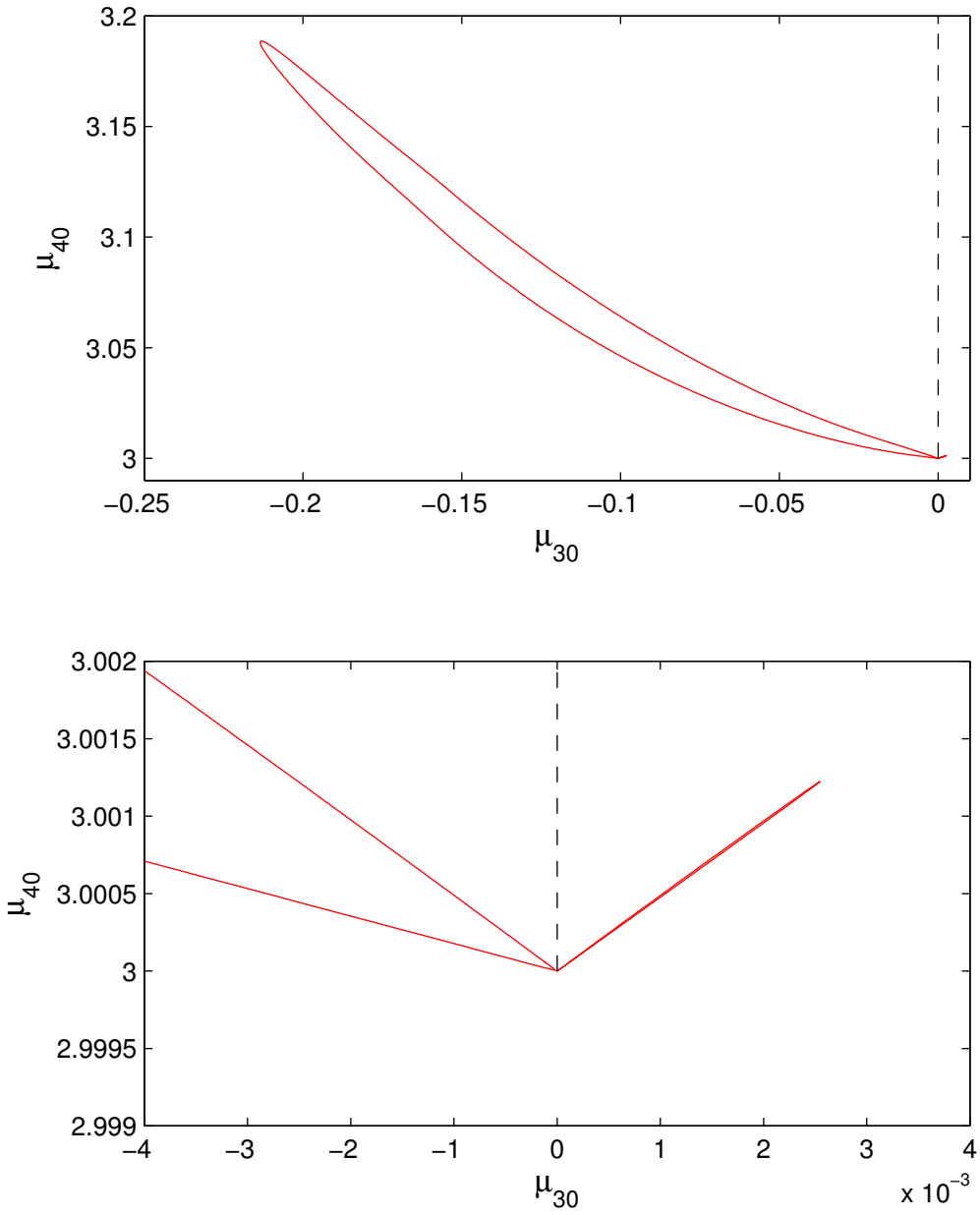

Figure 6.5: Top: Shock profile for $M = 1.5$ in (μ_{30}, μ_{40})-space. Bottom: Detail of top figure. The solution begins on the top-left curve, then returns to pseudo-equilibrium along the bottom-left curve before again leaving and returning to equilibrium along the right loop, which is too thin to display at this scale.

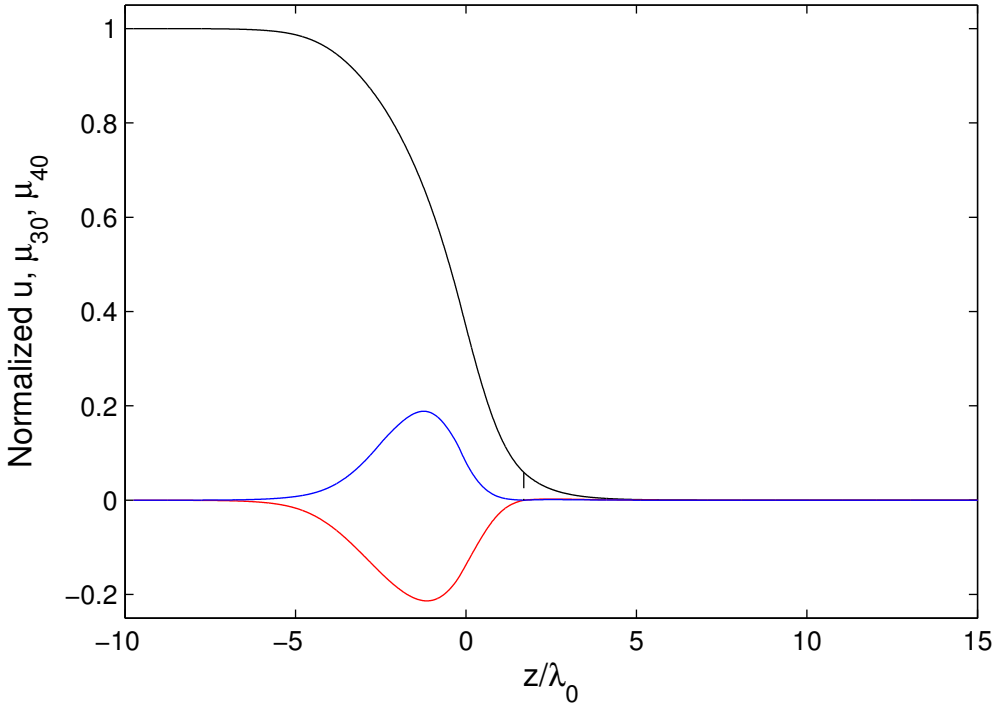

Figure 6.6: Top: Shock profiles for $M = 1.5$. Black line: u. Red line: μ_{30}. Blue line: μ_{40}. The dashed-dotted line indicates the location of the pseudo-equilibrium point.

on a trajectory consistent with (6.62) in Region I as it approaches or departs the equilibrium point. Plugging in the form (6.63) to the IMP contribution to the closing flux, (6.57), gives

$$\mu_{50,B} = \frac{(\mu_{40} - 3)(\mu_{40} + 3)}{\mu_{30}} = \frac{\nu_{40}}{\nu_{30}}\left(\frac{\nu_{40}}{\nu_{30}}\mu_{30} + 6\right). \tag{6.79}$$

As $\mu_{30} \to 0$, this approaches a constant value of $6\tan\omega$, where ω is the departure angle. This means that, if a solution enters the point $(0, 3)$ and then leaves, it must either do so in Region II, develop a discontinuity in the closing flux, or depart along exactly the same path by which it arrived. Clearly the third option is impossible as it would imply the ODE drives the system both inward and outward along the same trajectory at a single point in state space. Also, the first option is immaterial as the arrival at the pseudo-equilibrium point will be seen to occur within Region I in all cases. This means that there is a necessary discontinuity in the closing flux across the pseudo-equilibrium point, visible in Figure 6.7, and, except for the unlikely and, in practice, unobserved case of the derivative of μ_{50} remaining continuous across the pseudo-equilibrium point, this causes a discontinuity in the derivatives of the state

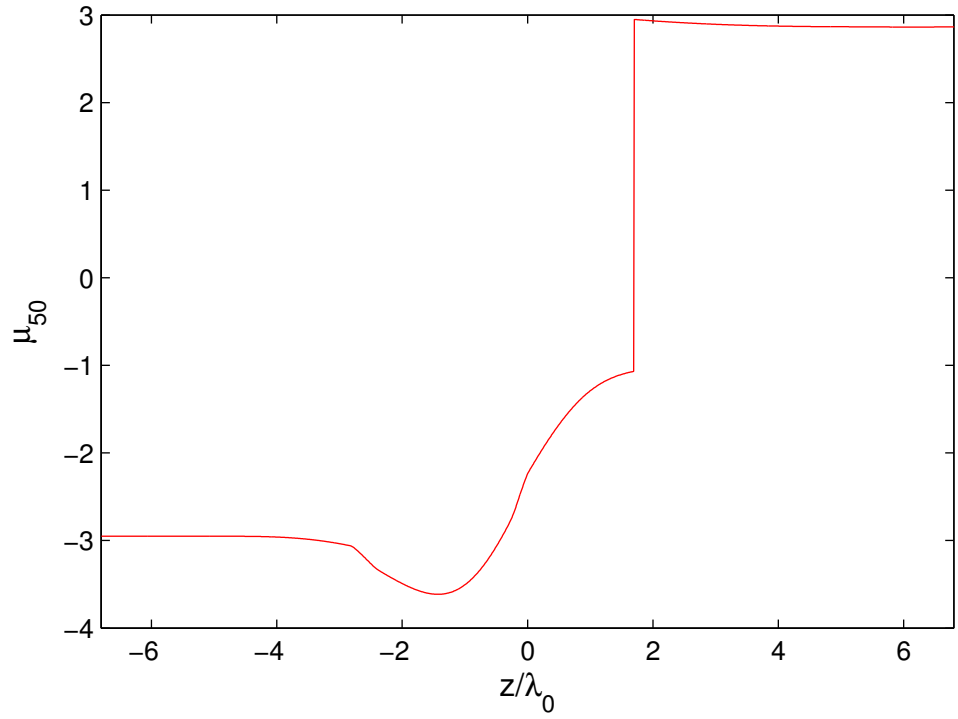

Figure 6.7: Top: Shock profile for $M = 1.5$ in μ_{50}.

variables. Unlike those discontinuities discussed in [55], however, this discontinuity is only in the derivatives of the primitive variables, and appears to occur for a quite different reason than the lack of sufficiently high wave speeds, as wave speeds on approach to this point are unbounded. It occurs because of the multivaluedness of the closing flux caused by the IMP.

The most intuitive method to handle the existence of the pseudo-equilibrium point, since it cannot be integrated through directly, is to begin integrating backwards from the downstream boundary, choosing an approximate boundary condition based on the linearized solution in (6.68), and continue backward until meeting the same pseudo-equilibrium point, hoping that the values of u from both solutions will match. In fact, this is easily done once the correct downstream boundary condition is chosen, which turns out to be the Region I solution which departs into the first quadrant, the $(+\mu_{30}, +\mu_{40})$ direction. A problem with this matching procedure is that the values of u do not match precisely at the pseudo-equilibrium point, meaning not only its derivative but its value is discontinuous there. Attempts have been made to subtly adjust the approximate boundary conditions in order to match both u and u' at the pseudo-equilibrium point, but unfortunately the value of u' at this point is almost insensitive to the choice of boundary condition and little improvement can be had

beyond laboriously matching the values of u on both sides of the pseudo-equilibrium point.

Another more crude but simpler and arguably more effective method is to detect when the solution has reached the pseudo-equilibrium point, then reflect the final solution point about the μ_{40} axis, scale the distance from the E point, $(0, 3)$, to some acceptable value, and then continue the integration from the new point. This ensures u remains continuous, and it is found in practice that the basin of attraction for the final downstream equilibrium point is quite large, meaning that the solution will find its way back to equilibrium once u approaches its downstream value. This reflection method is used in the examples shown. Using this scheme, profiles can be constructed for any Mach number in the valid range.

This feature of the solution is perplexing, but does seem to have some possible motivation. Consider again the basic moment equations for the shock wave, irrespective of the closure used. Three variables were eliminated, ρ, σ_1, and μ_{02}, but these need not have been our choice. We could have kept μ_{02} and instead eliminated u, reducing the unknowns to the normalized moments μ_{02}, μ_{30}, and μ_{40}. These have the same values at both boundaries, which means that our eliminating expression for u must be multi-valued. Each value of u exists on a different branch of the eliminating expression, but if we are to have a continuous shock profile, there must be some point within the shock at which some of these branches overlap. This suggests that within a steady shock there are already one or more special points at which the eliminating expression for u has repeated roots. It is hypothesized that the pseudo-equilibrium point is related to such a point.

6.8 Comparison With Standard Methods

Lastly, we should compare the Maximum Entropy Method to other well-established methods. We first demonstrate shock profiles for $M = 1.3, 1.5$, and 1.7. Figure 6.8 shows u, μ_{30}, and μ_{40} profiles, Figure 6.9 shows shock profile in (μ_{30}, μ_{40})-space, Figure 6.10 shows u' profiles, and Figure 6.11 shows the closing flux, μ_{50}. We comment that the pseudo-equilibrium point moves downstream as M decreases, thus it is not visible in the $M = 1.3$ profile, since u has already sufficiently relaxed to its downstream value at the pseudo-equilibrium point that the solution can be considered complete. Figures 6.9 and 6.10 show that the profile for $M = 1.7$ contains significant artifacts other than the pseudo-equilibrium point. These are due to crossing the Region I boundary, (4.40), at which point there is a discontinuity in

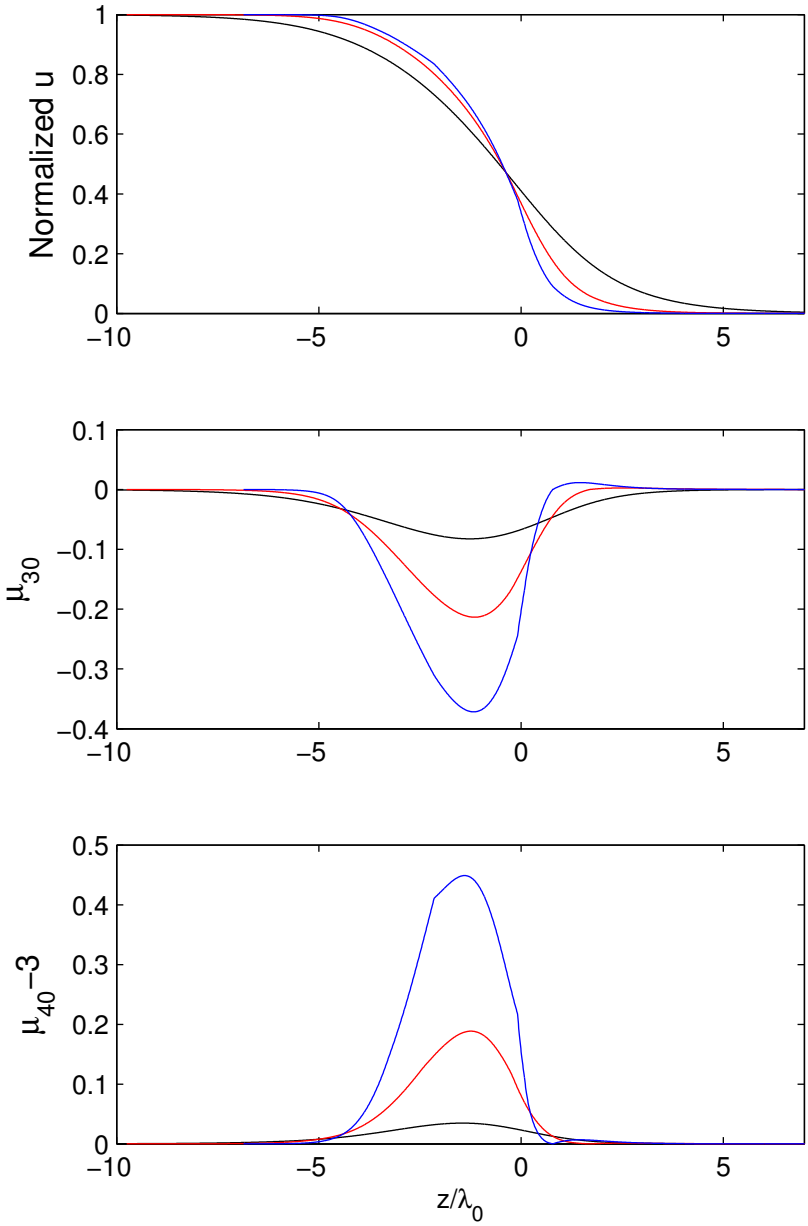

Figure 6.8: From top: Shock profiles of normalized u, μ_{30}, and μ_{40}. Black lines: $M = 1.3$. Red lines: $M = 1.5$. Blue lines: $M = 1.7$.

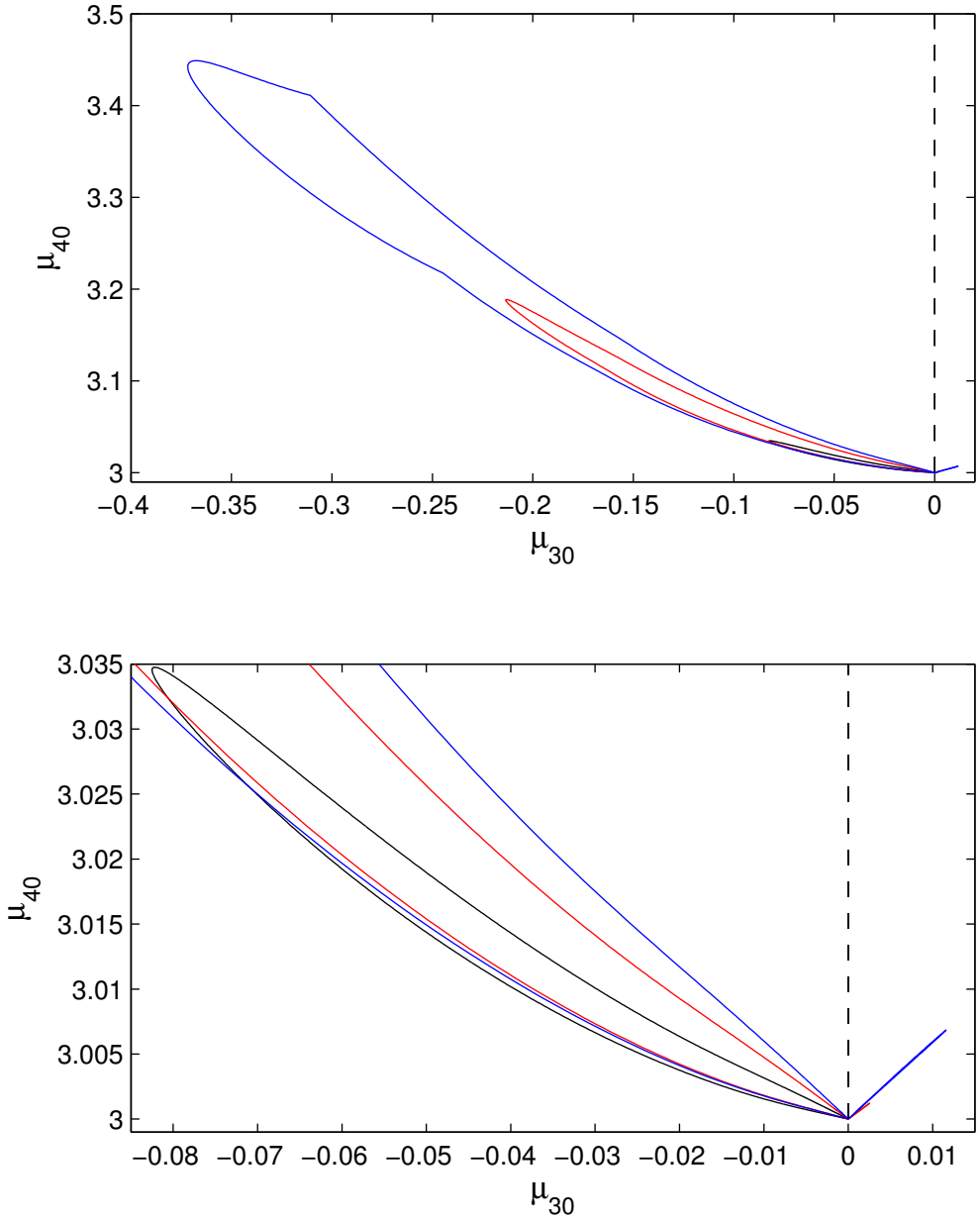

Figure 6.9: $M = 1.3$ (black), 1.5 (red), and 1.7 (blue) shock profiles in (μ_{30}, μ_{40}) plane. Black-dashed line: singular line extending from the equilibrium point, $\mu_{40} = 3$. The bottom figure is a detail of the top, showing the $M = 1.3$ profile and the small excursion on the right half plane after passing through the pseudo-equilibrium point for the other Mach number profiles.

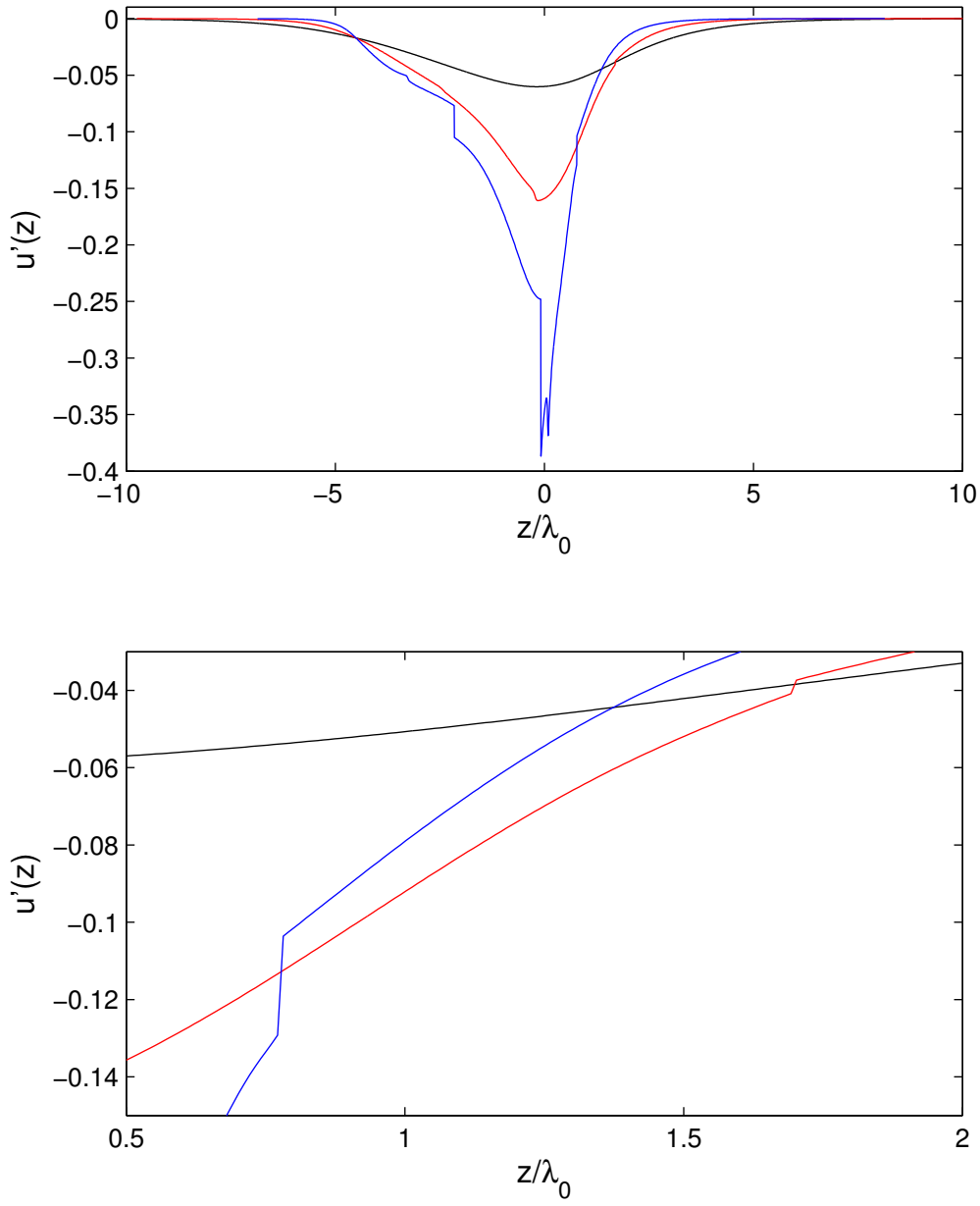

Figure 6.10: $M = 1.3$ (black), 1.5 (red), and 1.7 (blue) shock profiles of $u'(z)$ and detail. Attention is drawn to the lower plot, showing that the derivative of velocity has a slight discontinuity as the solution passes through the pseudo-equilibrium point.

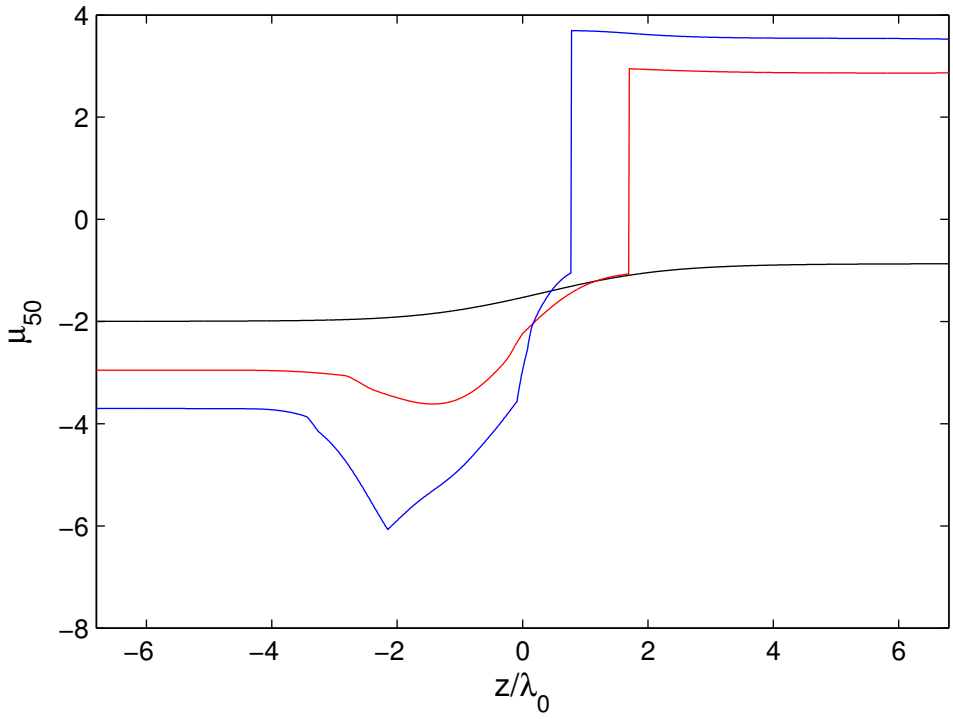

Figure 6.11: Mach 1.3 (black), 1.5 (red), and 1.7 (blue) profiles of the closing flux, μ_{50}, as calculated by the HMER.

the derivative of μ_{50} by design, in order to match the HMER and MER values and behavior of μ_{50} at this interface. We now look to compare these solutions to those produced by other methods.

For shock waves and many other non-equilibrium gas flows, DSMC is well suited. Due to the low Mach number range wherein solutions may be constructed using the Maximum Entropy Method, Navier–Stokes is another point of comparison as it is still somewhat accurate at predicting broad features of weak shock waves. When using DSMC, we follow the example of [6] and use the simple but robust method outlined there. Particle trajectories are simulated through the shock wave, including movement and collisions, and are used to collect statistical information about the flow. Hard sphere molecules are used to match the viscosity exponent from our expression for the BGK relaxation time (6.50). The domain for $M = 1.5$ was 65 times the upstream mean free path length, with 390 evenly spaced collision cells. The average number of molecules in the simulation was 18525, and this was maintained by using a version of Bird's STABIL algorithm. STABIL shifts all particles forward or backward when the total number of particles in the domain diverges from the intended mean by a small amount, in this case the number was

chosen such that the usual shift required was approximately one quarter of a cell. STABIL is necessary to avoid the shock "smearing" due to a random walk in its position over the averaging window. The simulation was allowed to relax to its steady state over 10 mean transit times through the domain of an average particle, and then was averaged over 20 mean transit times for smoothness.

Apart from the DSMC method as presented in the text, a subtle modification was required to tailor the method to very weak shocks; the number of particles escaping the domain through the upstream boundary, usually an extremely rare event at high Mach numbers, becomes significant as the Mach number becomes close to 1 and must be tracked so that the total can be added to the number of particles emitted from the upstream boundary. Neglecting this balancing flux of particles makes little difference at Mach numbers of 2 and higher, but at low Mach numbers the effect of lost particles can reduce the inflow by a few percent or more.

The Navier-Stokes solution is simpler. The problem reduces to an ODE [18] and can be simulated directly using a numerical integrator such as MATLAB's ode45 routine, which was the choice in this case.

Solutions are compared in Figures 6.12-6.14. Maximum Entropy clearly underestimates shock thickness compared to the other two methods, but produces a smooth profile (at least to the eye, recall the discussion of the pseudo-equilibrium point and resulting discontinuous derivatives of flow properties, such as in Figure 6.10). The reason for the thin shock profile using the Maximum Entropy Method is likely because of the limited power of the distribution to accommodate information about both the central and IMP components of the distribution. Extension to higher orders should help, but the analytical work necessary to understand the shape of the distribution at higher order is significantly more challenging than for fourth order as explored in Chapter 2.

Navier-Stokes does not produce meaningful predictions of the higher moments used in the HMERC, but comparison can be made with DSMC. Figures 6.15, **??**, and **??** show how the higher moments compare with DSMC for $M = 1.5$. Qualitatively, it is clear that there is no pseduo-equilibrium point in the DSMC calculation, as μ_{40} drops below its equilibrium value of 3 while $\mu_{30} < 0$. Because of this, μ_{50} computed using the HMERC is necessarily substantially different from its DSMC value due to reaching a finite limit at equilibrium, unlike the zero value seen in DSMC. Also, μ_{50} is not discontinuous. μ_{50} computed with DSMC begins to deviate from its equilibrium value much earlier than the HMERC solution, has different

behavior in the interior of the shock wave, and following the discontinuity at the pseudo-equilibrium point, the derivative of μ_{50} changes sign and no longer agrees even qualitatively with the DSMC solution. The singularity in the HMERC is linked to all of these features.

In Figures 6.18-6.20, velocity distribution functions are compared at equivalent density in the Maximum Entropy and DSMC shock solutions. The Maximum Entropy method does a good job of approximating the PDF to the eye, but slight differences compared with DSMC become apparent in the relative error plot. These subtle differences are due to the central component of the distribution being unable to fully accommodate the information present in the moments because of the sensitive IMP, which satisfies a non-trivial portion of the two important moments, μ_{30} and μ_{40}. The IMP carries away all of the excess μ_{40} and a fraction of the value of μ_{30}, resulting in the central component of the distribution being significantly more Gaussian than the DSMC distribution for small x. Essentially, satisfying the demands of the IMP has robbed some of the adaptive power of the central component. If an analytic version of the singular behavior of higher order moment reconstructions could be attained, better approximation of the central region of the velocity distribution function would be possible, as more degrees of freedom would be available to the solution after satisfying the singular portion. In effect, the MER behaves more like a reduced-order Maximum Entropy central distribution with an IMP.

6.9 A Note on Computational Cost

The HMERC method has a significant computational burden associated with the cost of performing the HMER procedure, which involves computing, for each Newton's step, a number of moment integrals proportional to the order of the reconstruction to compute the Jacobian of the constraint equations. This immediately means that the method cannot be competitive with the Navier-Stokes equations in terms of speed. Indeed, the Navier-Stokes solutions for the various shock waves presented in this chapter can be produced almost instantaneously, while the HMERC solutions take on the order of a few minutes. However, Navier-Stokes is unable to handle problems like the relaxation problem. For these explicitly non-equilibrium problems, other, more expensive methods are required.

When compared with the other non-equilibrium method used here, DSMC, the question of desired solution fidelity becomes important, as crudely averaged DSMC solutions that are marginally converged in the statistical sense can be performed quickly

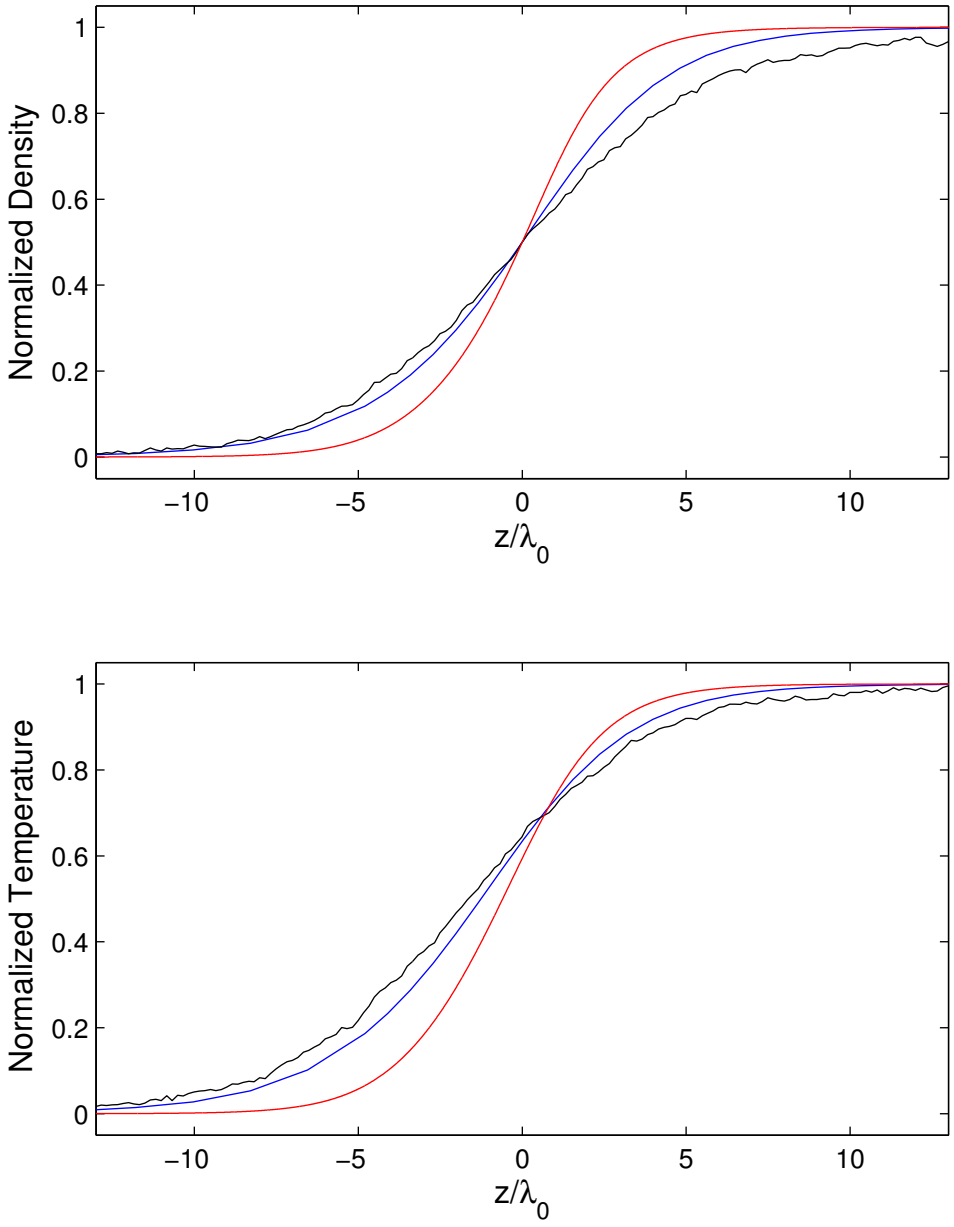

Figure 6.12: Normalized $M = 1.3$ shock density and temperature profiles. Black: DSMC, blue: Navier-Stokes, red: HMERC.

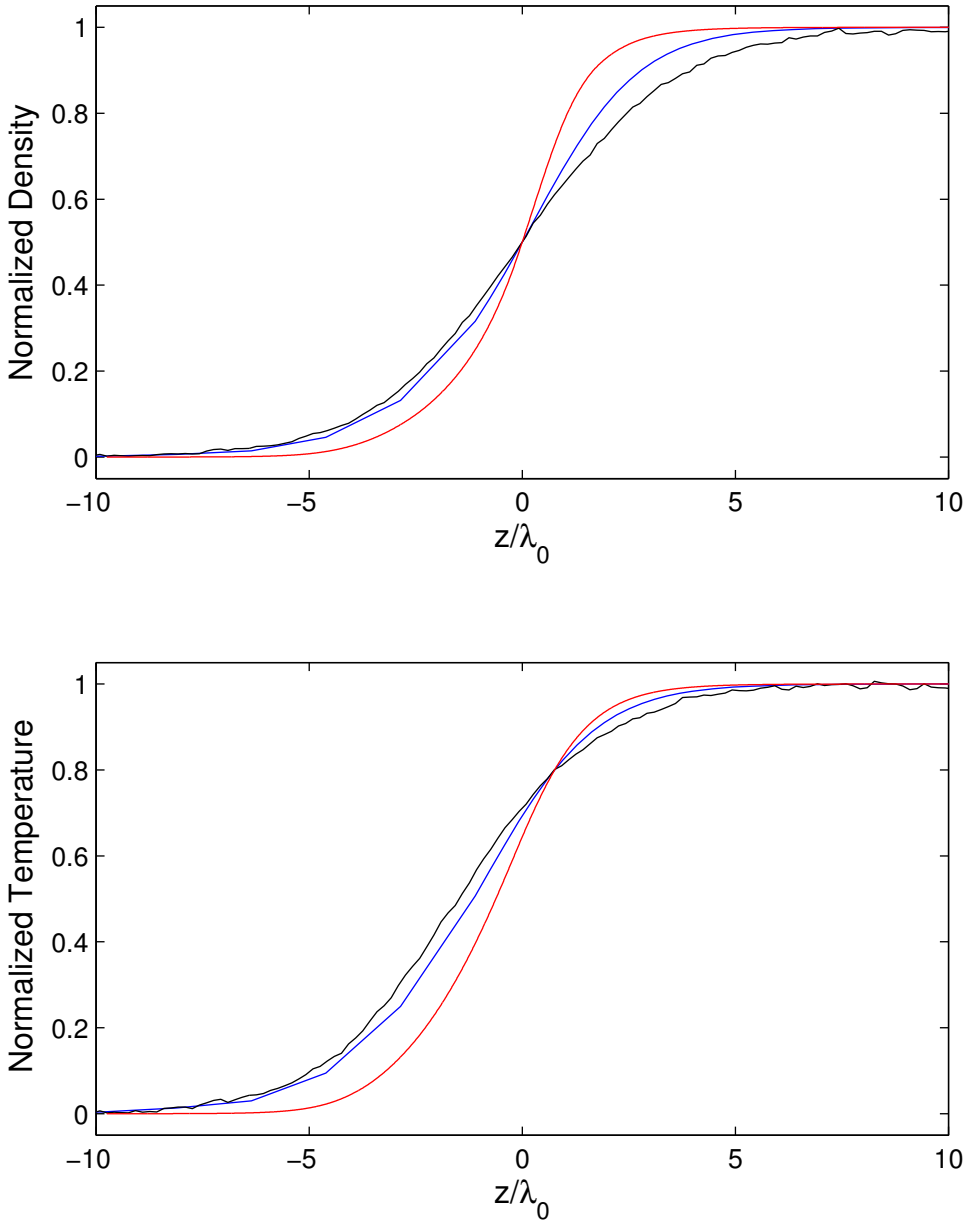

Figure 6.13: Normalized $M = 1.5$ shock density and temperature profiles. Black: DSMC, blue: Navier-Stokes, red: HMERC.

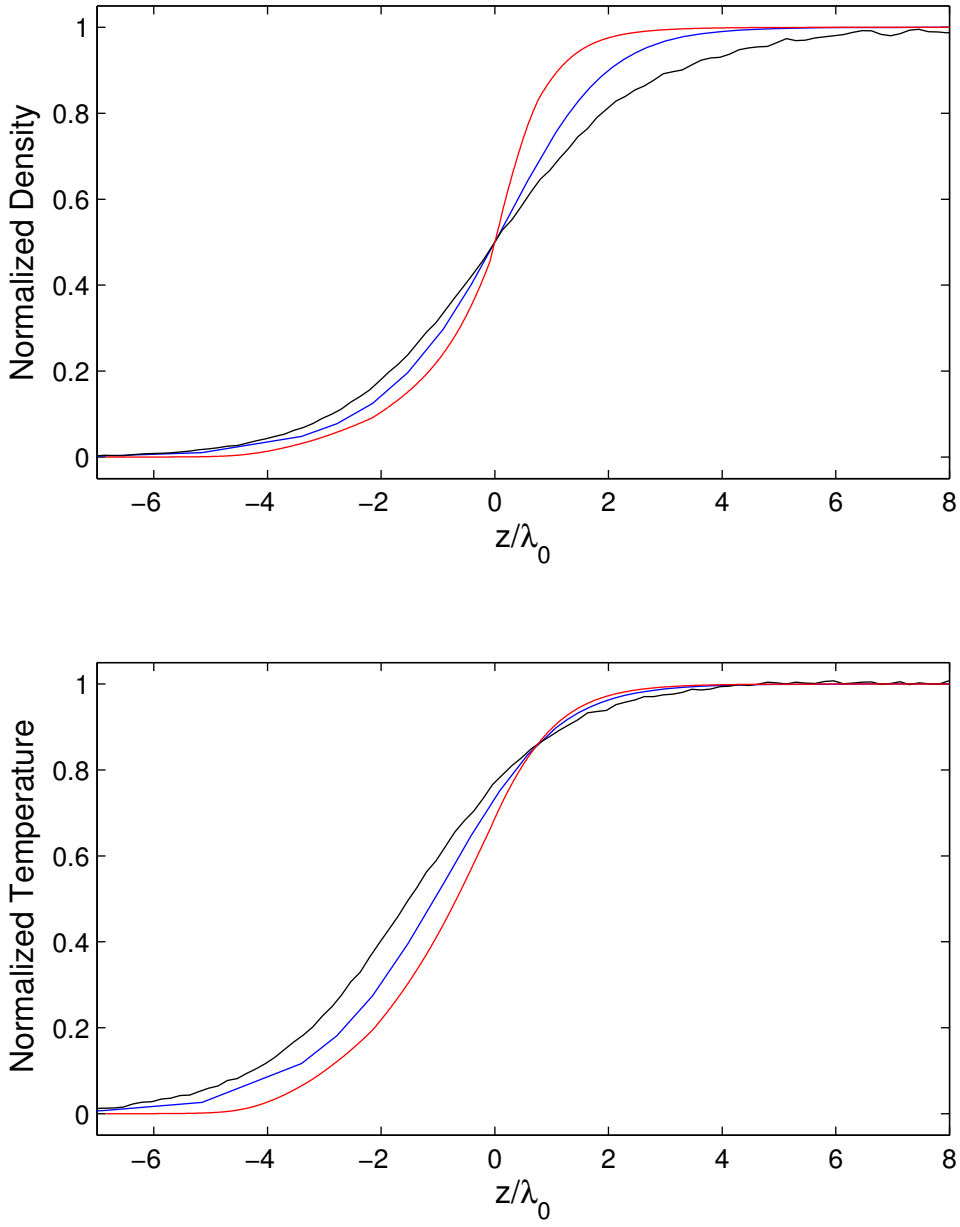

Figure 6.14: Normalized $M = 1.7$ shock density and temperature profiles. Black: DSMC, blue: Navier-Stokes, red: HMERC.

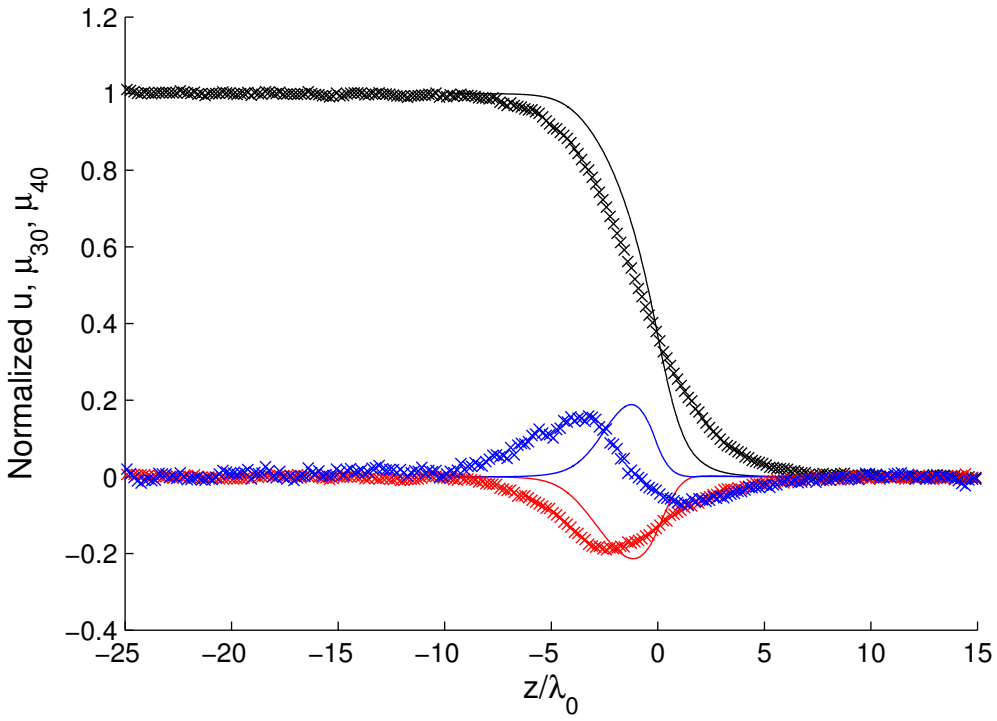

Figure 6.15: Comparison of $M = 1.5$ primitive solution variables u (black, normalized), μ_{30} (red), and $\mu_{40} - 3$ (blue), between HMERC (lines) and DSMC (black) solutions.

compared with HMERC solutions, but the HMERC method produces smooth PDFs and moments, whereas such marginal DSMC computations would be noisy and carry poor statistical information about the higher moments and the tails of the velocity PDFs. A relatively smooth DSMC solution with comparable information content to an HMERC solution would be very intensive, as vast numbers of particles must be sampled in order to obtain converged statistics about particles in the tails of the velocity distribution, which have an outsized effect on higher moments relative to their population. Such well-converged DSMC solutions take much longer than HMERC, on the order of hours for the shock wave solutions and several minutes for the relaxation problem, compared to minutes and seconds for HMERC, respectively. It is important to note that more sophisticated DSMC procedures do exist which enhance computational efficiency, but it is not believed that the reduction in cost is of the order required to make DSMC solutions comparable to HMERC when higher moments and PDF tails are of interest.

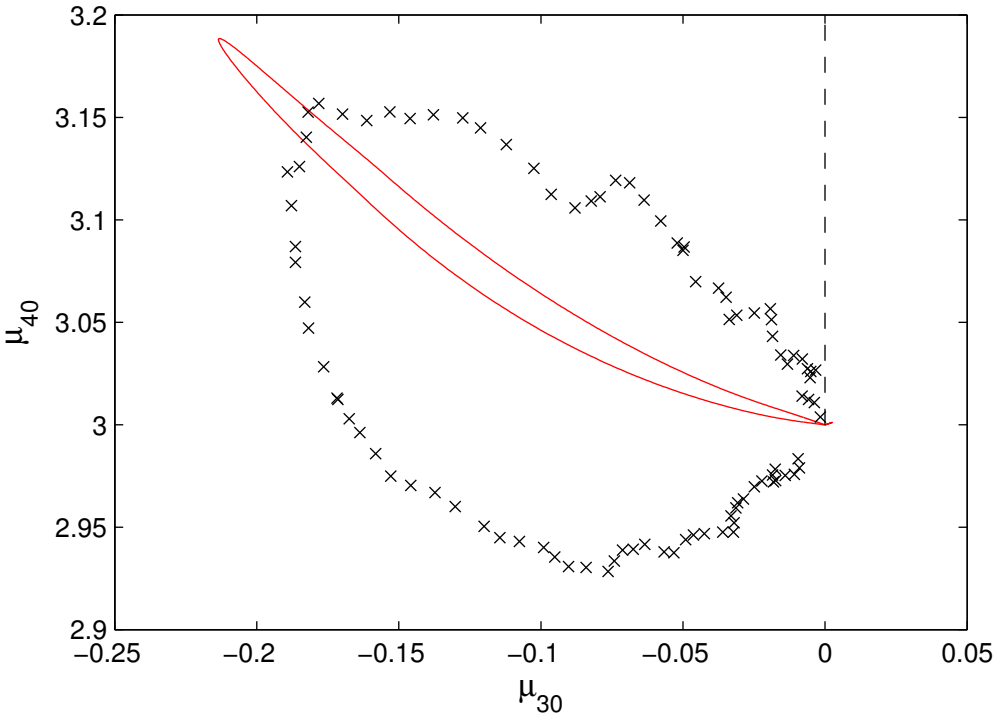

Figure 6.16: Comparison of $M = 1.5$ shock profiles in the (μ_{30}, μ_{40})-plane. Red Line: HMERC, Black Dashed Line: Singular line, Symbols: DSMC.

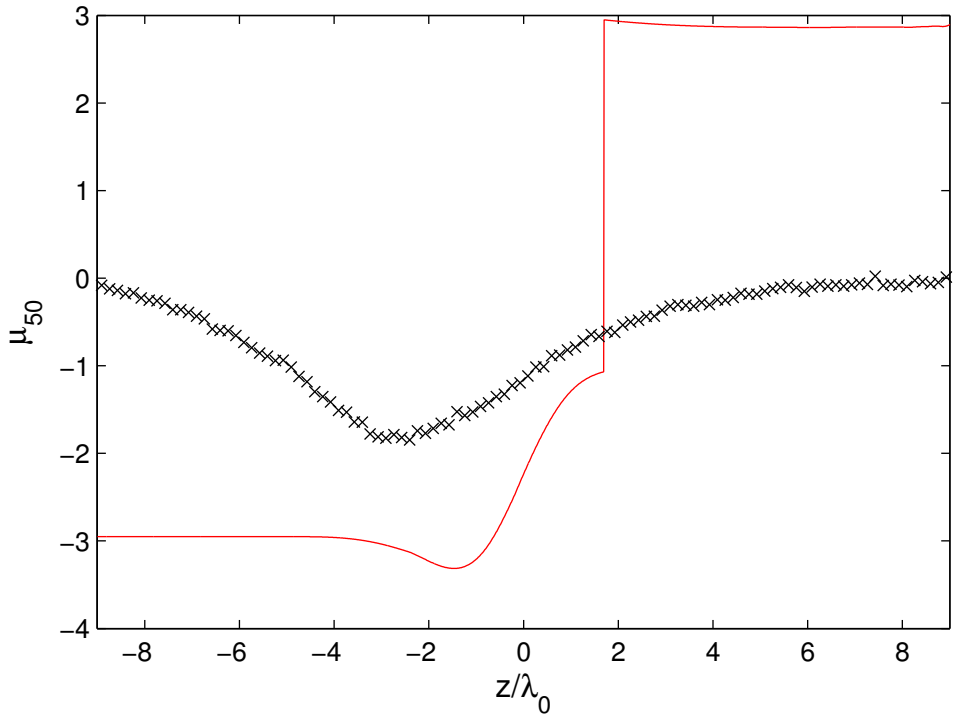

Figure 6.17: Comparison of $M = 1.5$ shock profiles of the closing moment, μ_{50}. Red Line: HMERC, Symbols: DSMC.

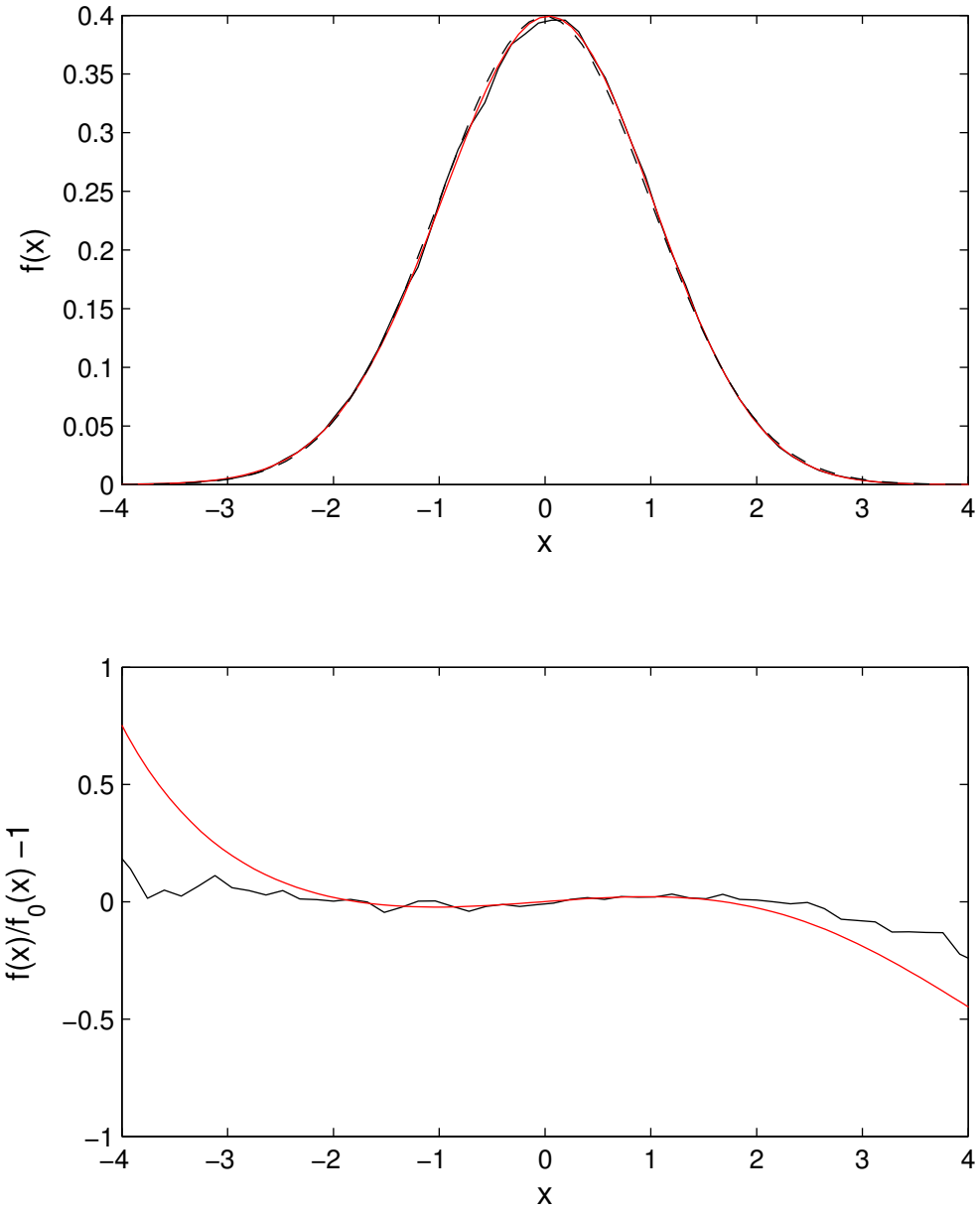

Figure 6.18: Velocity distributions comparison at a point within the $M = 1.3$ shock. Black: DSMC, Red: HMERC. Black Dashed: Maxwell-Boltzmann Distribution. Due to the varying shock thicknesses, the z location of the comparison point in each case is chosen at points with equal density. Normalized ρ here is 0.4893. Upper figure: Velocity distribution function values. Lower figure: Relative deviation from the Maxwell-Boltzmann distribution.

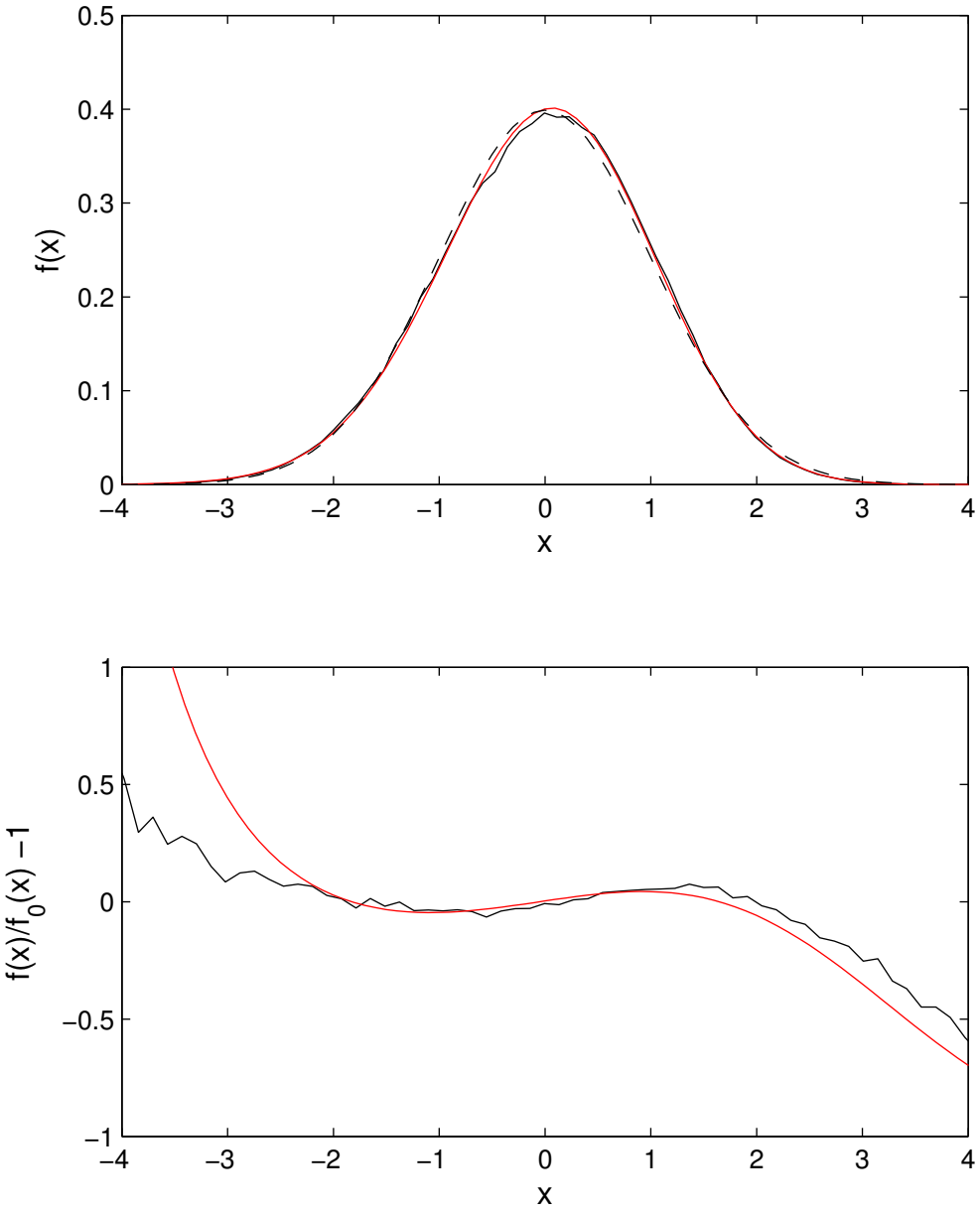

Figure 6.19: Velocity distributions comparison at a point within the $M = 1.5$ shock. Black: DSMC, Red: HMERC. Black Dashed: Maxwell-Boltzmann Distribution. Due to the varying shock thicknesses, the z location of the comparison point in each case is chosen at points with equal density. Normalized ρ here is 0.4893. Upper figure: Velocity distribution function values. Lower figure: Relative deviation from the Maxwell-Boltzmann distribution.

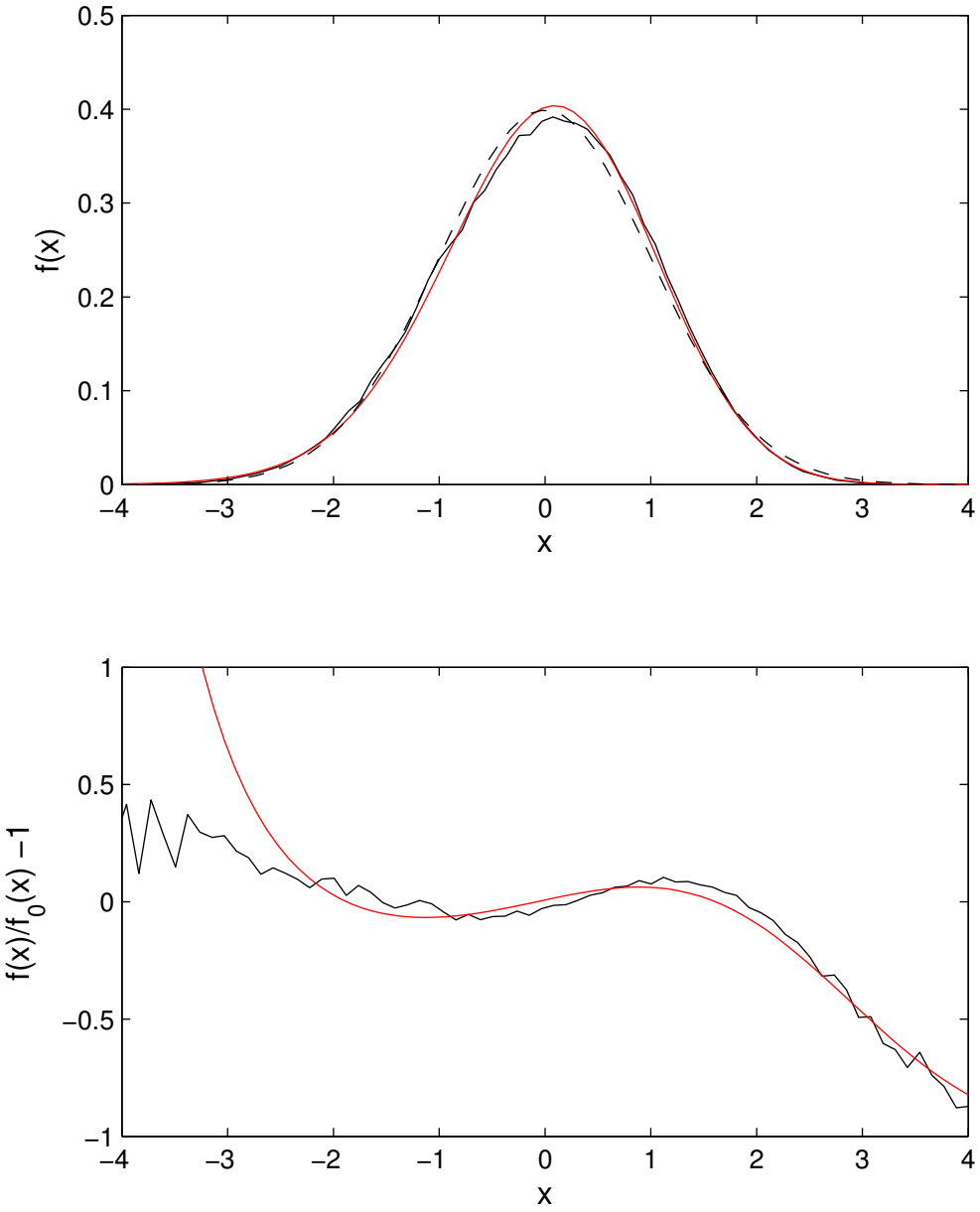

Figure 6.20: Velocity distributions comparison at a point within the $M = 1.7$ shock. Black: DSMC, Red: HMERC. Black Dashed: Maxwell-Boltzmann Distribution. Due to the varying shock thicknesses, the z location of the comparison point in each case is chosen at points with equal density. Normalized ρ here is 0.4893. Upper figure: Velocity distribution function values. Lower figure: Relative deviation from the Maxwell-Boltzmann distribution.

Chapter 7

CONCLUSIONS

We have considered the Maximum Entropy Reconstruction (MER) as a solution to the problem of approximating a distribution function from a finite number of known moments. We focused on the case using polynomial moments of order four and lower of a one-dimensional probability distribution function (PDF). It was known [26] that the MER on the real line has a singular line, of zero measure in terms of the prescribed moments but extending in moment space from the case corresponding to thermodynamic equilibrium, in which there is no solution to the MER problem. Near the singular line there is a region, which we call Region I, in which the MER is bimodal and the effect of the singular line is dominant. We constructed an asymptotically valid analytic solution to the singular MER problem in Region I, which predicts the existence of an extremely small-amplitude, near-Gaussian component of the MER solution which runs away to infinity as the singular line is approached. We call this component the Itinerant Moment Packet (IMP). The IMP is found to dominate the computed values of higher order moments, predictions of the MER based on the given values of the lower order moments. We used this IMP model to construct a hybrid analytic-numerical model of the MER, which we call the Hybrid Maximum Entropy Reconstruction (HMER). The HMER is found to accurately reproduce the predictions of higher order moments by the MER, and to significantly improve upon the MER in terms of robustness and computational efficiency.

We used the HMER to examine the problem of relaxation to equilibrium in a spherically symmetric, hard-sphere gas. The HMER was used to represent the velocity PDF in the calculation of the collision integral of the Boltzmann equation to high accuracy, providing a near perfect match when compared to the widely-accepted Direct Simulation Monte-Carlo (DSMC) method of Bird [6].

We also examined the use of the HMER as a closure scheme for moment-based approaches to solving the Boltzmann equation. We call this the HMER closure (HMERC). We determined that the one-dimensional HMER was sufficient to construct a closure scheme for the full three-dimensional moment equations if an appropriate choice of moments is used, and if the HMER is performed in a rotated

frame of reference given by the principal frame of the gas based on the local stress tensor. We call the appropriate choice of moments velocity-factorizable moments (VFM).

We employed the HMERC and the VFM approach in the problem of computing the profile of a normal shock wave. Perturbation solutions were constructed near upstream and downstream equilibrium, and it was found that solutions exist only for a finite range of Mach numbers. In the VFM case, solutions do not exist above a critical Mach number of approximately 1.992. When solutions do exist they are continuous, but the derivatives are not. There exists a point which we call the pseudo-equilibrium point where the solution returns to equilibrium in terms of its moments in the shock-normal direction, but the shock-parallel components of the solution have not reached their downstream equilibrium values yet. Since the equilibrium point is strongly singular and the closing flux is multi-valued, this causes a discontinuity in the closing flux across this point, which leads to a discontinuity in the derivatives of the state variables. We developed a method to integrate through this pseudo-equilibrium point, but the discontinuities are preserved and represent a subtle defect in the shock profile solution.

We compared the HMERC results for the shock wave with solutions from DSMC and Navier-Stokes, and found that, while the HMERC produces many qualitatively correct features in the shock profile, the overall shock thickness is under-predicted compared to both alternative methods.

7.1 Asymptotic Analysis of the MER

In Section 2.4 we developed a model for the MER. In Region II, being the region near equilibrium but away from the singular line defined by $\mu_3 = 0$ and $\mu_4 > 3$, the reconstruction is well-behaved and continuous even in the limit of equilibrium as $(\mu_3, \mu_4) \to (0, 3)$. In Region I, being the region of moment space near the singular line, the true solution was found to be bimodal and to consist of a central component, well-approximated by a perturbed Gaussian sitting very near $x = 0$, and an Itinerant Moment Packet (IMP) with rapidly decreasing amplitude and rapidly increasing distance from the origin. The model was found to be self-consistent and to solve the MER problem to leading order in the limit as $\mu_3 \to 0$ for $\mu_4 > 3$. Further, the IMP model predicts, in (4.45), the contribution of the IMP to each moment of the distribution, both constrained and computed moments. The contribution to the first three moments, μ_0, μ_1, and μ_2, is essentially negligible, the contribution to

μ_3 is an appreciable fraction of μ_3, and the contribution to μ_4 is always $\mu_4 - 3$, or the excess in μ_4 above that of a standard normal distribution. This behavior is reminiscent of the symmetric packets proposed in [27], although where Junk used packets with a similar relationship between amplitude and distance from the origin in order to demonstrate the existence of the singular subspace, we have computed analytic forms for the packets which are predictive given μ_3 and μ_4 and applicable throughout Region I, as opposed to only on the singular line itself.

The IMP was found to dominate the value of μ_5, the closing flux in our gas dynamics application. μ_5 grows as μ_3^{-1} in Region I, which is important as it causes multi-valuedness in this flux as the system approaches or departs equilibrium, depending on the exact direction in (μ_3, μ_4)-space it travels. This inverse dependence of the fifth and third moments is in contrast to the interdependence of these two quantities in [42], which has $\mu_5 \propto \mu_3^{-2/3}$. It is believed that this difference has a significant effect on the existence of solutions to the steady normal shock structure problem which approach equilibrium in the upstream or downstream limit.

7.2 The Hybrid Maximum Entropy Reconstruction Method

In Section 2.5 we constructed a new, highly robust and efficient method for computing the MER. Instead of directly solving for the distribution on the infinite domain, we first used our newfound understanding of the IMP to calculate and subtract away its contribution to the constrained moments before passing them to a numerical solution method working on a finite domain. Due to the IMP's effects, this means that the finite domain solution sees much more benign moments, never in Region I, which produce well-behaved and well-conditioned solutions which converge quickly. In order to link this solution in Region I with solutions outside the singular region smoothly, the boundary between the regions was defined by detecting the transition point when the closing flux rapidly changes its μ_3 derivative. As it is not clear how to analytically predict the locus of this boundary, and in the absence of a uniformly valid asymptotic solution around the equilibrium point, an empirical fit was used consisting of a quartic version of a hyperbola (4.40). A weighting function, (4.46), which preserves the sharp transition between the two regions was developed and implemented to ensure continuity in the closing flux and all other moments of the reconstruction across the boundary.

Finding analytical estimates for both the Region I boundary and the weight function are the only remaining problems in the HMER, though the current empirical fits

are good approximations at present. While the HMER does an excellent job of predicting the closing fluxes (see Figure 4.6), it does not accurately predict the Lagrange multipliers, λ_i, for all cases, especially near the transition between the two regions. This is made clear in Figures 4.3, 4.4, and 4.5. It is believed that improving the boundary location and weight functions will improve agreement in these quantities.

7.3 Maximum Entropy Gas Dynamics—Relaxation to Equilibrium

The Hybrid Maximum Entropy Reconstruction was used as a closure to compute collision integrals for the spatially homogeneous, spherically symmetric distribution relaxation to equilibrium of a monatomic gas with a Maxwell molecule collision cross section. Comparison was made with the exact solution computed in [33]. These results show excellent agreement, demonstrating the effectiveness of the reconstruction in representing the key features of the velocity distribution function required to properly compute the collision integral.

7.4 The Velocity-Factorizable Moments Approach

In Section 4.2 we introduced the idea of velocity-factorizable moments for reducing the computational cost of producing a Maximum Entropy Reconstruction solution by eliminating three-dimensional integrals from the reconstruction algorithm. The key step is the elimination of cross-moments from the list of moment constraints that we must match with the Maximum Entropy Reconstruction. The exception are the shear values, which can be incorporated into the reconstruction by rotating the axes of the velocity space into the principal frame of the gas, and then performing the reconstruction in this context where the pressure tensor will be purely diagonal. Higher-order moments are treated as fluxes

7.5 Limited Mach Number Range in the Shock Structure Problem

The analytic expressions for the closing flux in the two regions from Section 2.4 were used to produce linearized versions of the normal shock structure equations near the upstream and downstream equilibrium states. Care had to be taken to ensure that solutions found were consistent, which required that they be real, and that they depart/arrive from/to equilibrium in the proper direction so that they remained in the space where the chosen closing flux model was valid. The results are summarized in Figures 6.1 and 6.2.

The results reveal that there are no consistent solutions which depart from upstream

equilibrium for $M > 1.992$, and that the positive eigenvalue associated with departing solutions below this Mach number diverges here, returning as a strongly negative eigenvalue. It is believed that this is related to the multi-valuedness of the closing flux μ_5 with respect to the departure angle, since solutions at higher Mach numbers still occur under McDonald's approximate closure with a less severe dependence on μ_3 [42]. This certainly represents a failure of the closure scheme and needs remedy.

The question that remains is whether accuracy should be abandoned and the singularity altered to have μ_5 scale with μ_3^{-a} for some $0 < a < 1$, or should something more fundamental be altered. Noting that the high-order polynomial terms in the exponent of the MER cause the tails of the distribution to rapidly vanish, much faster than the near-Guassian tails of test distributions, perhaps a split domain is worthy of consideration. By modifying the constraining moment integrands (the chosen values of ψ in (2.3)) such that they measure only over a finite range, and then adding additional, lower-order moment integrands which span the remainder of the space, one could conceive of a piece-wise MER which falls off at large x more realistically, perhaps modifying the singular behavior of the closure. The possibility is intriguing but it is also possible that this would have the effect of removing the singularity and thus limiting characteristic speeds, again causing subshocks to appear in the shock profile.

7.6 Shock Profiles using the HMERC

The shock profiles produced by using the Hybrid Maximum Entropy Reconstruction to calculate the unclosed moment, μ_{50}, in the case of the velocity-factorizable shock structure problem are qualitatively acceptable, with nearly smooth profiles. The only obviously artificial defects are the kinks produced by derivative discontinuities at an interesting point in the solution, the pseudo-equilibrium point, and those present in the $M = 1.7$ case where the boundary of Region I is crossed.

The presence of the pseudo-equilibrium point is interesting. The solution is drawn back to thermal equilibrium—in the x-direction only, while the energy present in the r-component of particle velocity is still below the downstream equilibrium value. It was speculated that this could be occurring because of the fact that the system of equations (6.45), (6.46), and (6.47) could have taken μ_{02} in place of u as a primary variable, which would result in a multi-valued elimination function, since μ_{02}, along with μ_{30} and μ_{40} has identical values at the upstream and downstream equilibrium states. The pseudo-equilibrium point could be the unique point required to connect

the two values of u. Whether this is true or not, the pseudo-equilibrium point is also noteworthy because it represents a discontinuity in μ_{50} as well as $\frac{du}{dz}$, as well as a slight obstacle to computing complete shock profiles.

In order to compute the shock profiles through the pseudo-equilibrium point, the most practical method found was to detect when the solution has reached the point by detecting when $\mu_{30} > b$, where b is a small, negative value which is set dependent on the precision of the calculation to ensure the pseudo-equilibrium point is detected before it is crossed which could potentially result in undefined behavior. The point at which this state is detected is then reflected across the μ_{30} axis, and μ_{30} and μ_{40} are scaled by a suitable fraction to ensure they are not too close nor too far from equilibrium for the calculation to continue. The required range is usually satisfied by scaling the values back by a factor of 10^{-2}. When the simulation continues from this point, the strong attraction of the downstream equilibrium point overcomes the small errors in the trajectory and the shock profile is completed. By using the same values of u and z as before the pseudo-equilibrium point was reached, this method ensures continuity in the u profile, even if continuity in its derivative is impossible to achieve due to the discontinuity in μ_{50}.

The shock profiles may be continuous and near-smooth, but they are not without flaws. Apart from the discontinuous derivatives mentioned above, the profiles are also considerably thinner than the equivalent profiles computed with DSMC, and slightly thinner even than profiles computed with Navier-Stokes. This is a departure from the behavior encountered when using regularized and modified versions of the MER, in which closures based on the Maximum Entropy Principle produced shocks of comparable thickness to the true BGK solution, and certainly much thicker than Navier-Stokes [41], [42]. Again, this must be due to the effect of the multi-valued μ_{50} closing flux and its particular singular form.

Velocity distribution functions, compared between DSMC and the Maximum Entropy Method at equivalent density values in the interior of shocks, show that the Maximum Entropy Method predicts slightly more Gaussian distributions where the distribution is large. This is believed to be the result of the IMP's contribution to the constrained moments making the computed central component significantly nearer to a Gaussian than it would otherwise be, as well as the effect of the pseudo-equilibrium point which drives the solution towards a Gaussian distribution faster than in the more realistic DSMC calculation. It is possible that this has the effect of causing the shock solution in the Maximum Entropy Method case to behave

somewhat more like an Euler or Gaussian solution, which would have an infinitely thin shock. While μ_{40} is significantly affected by the presence of the IMP, μ_{30} and lower-order moments are not very strongly affected, and this could be allowing the lower moments to behave more like they would in the Euler or Gaussian cases.

7.7 A Final Assessment

The five-moment MER is surprisingly and highly singular but this singularity has been well-examined and is thoroughly captured by the IMP model. Nevertheless, this singularity represents a significant stumbling block for applying the MER to gas dynamic problems. In the case of the HMERC, there are stark errors in its predictions where available and restrictive limits on the existence of solutions in at least the shock structure problem. It does, however, perform quite well in the relaxation problem, accurately predicting the rate of relaxation compared with theory.

The Hybrid Maximum Entropy Reconstruction and its enabling analysis of the singular region of the full Maximum Entropy Reconstruction represent the major contribution of this work. No similar analytic results for Maximum Entropy problems of order higher than quadratic have been found, and it is believed that the HMER is the most robust and globally-accurate version of the five-moment Maximum Entropy solution yet published. Improvements could yet be made in analytically predicting the boundary of Region I, and in developing an analytic justification for an appropriate form of the weight function (4.46).

Beyond these improvements, we must look towards extending the analysis of the singularity present in the five-moment MER to higher-order MERs and multi-dimensional MERs. Extension to multiple dimensions is made difficult by the interplay of cross-terms in the expression for the MER, meaning that a simple Gaussian form for the IMP is invalid for the general case. Extension to higher-order is complicated by the fact that the singular line in these higher-order MER problems is still present, but in higher-order cases its actual shape is a function of the lower moments which can no longer be standardized by adjusting coordinate axes. It is possible that some manipulation of the moment constraints in the MER problem can be made to eliminate this difficulty, allowing further progress to be made.

BIBLIOGRAPHY

[1] Rafail V Abramov. "An improved algorithm for the multidimensional moment-constrained maximum entropy problem". In: *Journal of Computational Physics* 226.1 (2007), pp. 621–644.

[2] Naum Iljič Aheizer and N Kemmer. *The classical moment problem and some related questions in analysis*. Oliver & Boyd, 1965.

[3] Petra Aumann et al. "MEGAFLOW: Parallel complete aircraft CFD". In: *Parallel Computing* 27.4 (2001), pp. 415–440.

[4] George Keith Batchelor. *An introduction to fluid dynamics*. Cambridge university press, 2000.

[5] Prabhu Lal Bhatnagar, Eugene P Gross, and Max Krook. "A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems". In: *Physical review* 94.3 (1954), p. 511.

[6] G.A. Bird. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*. Molecular Gas Dynamics and the Direct Simulation of Gas Flows v. 1. Clarendon Press. ISBN: 9780198561958.

[7] GA Bird. "Sophisticated dsmc". In: *Notes prepared for a short course at the DSMC07 meeting, Santa Fe, USA*. 2007.

[8] GA Bird, Timothy J Bartel, and Michael A Gallis. "Forty years of DSMC, and now?" In: *AIP Conference Proceedings*. Vol. 585. 1. AIP. 2001, pp. 372–380.

[9] AV Bobylev. "The Chapman-Enskog and Grad methods for solving the Boltzmann equation". In: *Akademiia Nauk SSSR Doklady*. Vol. 262. 1982, pp. 71–75.

[10] Ludwig Boltzmann. "Further Studies on the Thermal Equilibrium of Gas Molecules". In: *The Kinetic Theory of Gases*. Imperial College Press, 2011, pp. 262–349. DOI: 10.1142/9781848161337_0015. URL: http://www.worldscientific.com/doi/abs/10.1142/9781848161337_0015.

[11] Frederic Bourgault et al. "Information based adaptive robotic exploration". In: *Intelligent Robots and Systems, 2002. IEEE/RSJ International Conference on*. Vol. 1. IEEE. 2002, pp. 540–545.

[12] S. Chapman and T.G. Cowling. *The Mathematical Theory of Non-uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases*. Cambridge Mathematical Library. Cambridge University Press, 1970. ISBN: 9780521408448.

[13] Timothée Ewart et al. "Mass flow rate measurements in a microchannel, from hydrodynamic to near free molecular regimes". In: *Journal of fluid mechanics* 584 (2007), pp. 337–356.

[14] Mohammad Faghri Hongwei Sun. “Effects of rarefaction and compressibility of gaseous flow in microchannel using DSMC”. In: *Numerical Heat Transfer: Part A: Applications* 38.2 (2000), pp. 153–168.

[15] Joel H Ferziger and Hans G Kaper. “Mathematical theory of transport processes in gases”. In: (1972).

[16] Iain Fraser. “An application of maximum entropy estimation: the demand for meat in the United Kingdom”. In: *Applied Economics* 32.1 (2000), pp. 45–59.

[17] Mohamed Gad-el-Hak. *The MEMS handbook*. CRC press, 2001.

[18] Dj Gilbarg and D Paolucci. “The structure of shock waves in the continuum theory of fluids”. In: *journal of Rational Mechanics and Analysis* 2.5 (1953), pp. 617–642.

[19] Harold Grad. “On the kinetic theory of rarefied gases”. In: *Communications on pure and applied mathematics* 2.4 (1949), pp. 331–407.

[20] Stephen F Gull and Geoff J Daniell. “Image reconstruction from incomplete and noisy data”. In: (1978).

[21] Nicolas G Hadjiconstantinou. “The limits of Navier-Stokes theory and kinetic extensions for describing small-scale gaseous hydrodynamics”. In: *Physics of Fluids* 18.11 (2006), p. 111301.

[22] Francis Begnaud Hildebrand. *Introduction to numerical analysis*. Courier Corporation, 1987.

[23] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 2012.

[24] Edwin T Jaynes and G Larry Bretthorst. *Probability theory: the logic of science*. Vol. 200. 3. Cambridge University Press Cambridge, 1996.

[25] ET Jaynes. “The relation of Bayesian and maximum entropy methods”. In: *Maximum-entropy and Bayesian methods in science and engineering*. Springer, 1988, pp. 25–29.

[26] Michael Junk. “Domain of definition of Levermore’s five-moment system”. In: *Journal of Statistical Physics* 93.5 (1998), pp. 1143–1167.

[27] Michael Junk. “Maximum entropy for reduced moment problems”. In: *Mathematical Models and Methods in Applied Sciences* 10.07 (2000), pp. 1001–1025.

[28] Michael Junk. “Maximum entropy moment problems and extended Euler equations”. In: *Transport in Transition Regimes*. Springer, 2004, pp. 189–198.

[29] Michael Junk and Andreas Unterreiter. “Maximum entropy moment systems and Galilean invariance”. In: *Continuum Mechanics and Thermodynamics* 14.6 (2002), pp. 563–576.

[30] A Karchani and RS Myong. “Convergence analysis of the direct simulation Monte Carlo based on the physical laws of conservation”. In: *Computers & Fluids* 115 (2015), pp. 98–114.

[31] Jeffrey T Kiehl et al. *Description of the NCAR Community Climate Model (CCM3). Technical Note*. Tech. rep. National Center for Atmospheric Research, Boulder, CO (United States). Climate and Global Dynamics Div., 1996.

[32] Song-Charng Kong et al. *Modeling and experiments of HCCI engine combustion using detailed chemical kinetics with multidimensional CFD*. Tech. rep. SAE Technical paper, 2001.

[33] Max Krook and Tai Tsun Wu. “Exact solutions of the Boltzmann equation”. In: *The Physics of Fluids* 20.10 (1977), pp. 1589–1595.

[34] GJ LeBeau and FE Lumpkin Iii. “Application highlights of the DSMC Analysis Code (DAC) software for simulating rarefied flows”. In: *Computer Methods in Applied Mechanics and Engineering* 191.6 (2001), pp. 595–609.

[35] C David Levermore. “Moment closure hierarchies for kinetic theories”. In: *Journal of Statistical Physics* 83.5 (1996), pp. 1021–1065.

[36] C David Levermore and William J Morokoff. “The Gaussian moment closure for gas dynamics”. In: *SIAM Journal on Applied Mathematics* 59.1 (1998), pp. 72–96.

[37] C David Levermore, William J Morokoff, and BT Nadiga. “Moment realizability and the validity of the Navier–Stokes equations for rarefied gas dynamics”. In: *Physics of Fluids* 10.12 (1998), pp. 3214–3226.

[38] HW Liepmann, R Narasimha, and M Chahine. “Theoretical and experimental aspects of the shock structure problem”. In: *Applied Mechanics*. Springer, 1966, pp. 973–979.

[39] Joseph Loschmidt. *Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft: I [-IV].* aus der KK Hof-und Staatsdruckerei, 1876.

[40] EA Malkov et al. “High-accuracy deterministic solution of the Boltzmann equation for the shock wave structure”. In: *Shock Waves* 25.4 (2015), pp. 387–397.

[41] James G. McDonald and Clinton P.T. Groth. “Numerical Solution of Maximum-Entropy-Based Hyperbolic Moment Closures for the Prediction of Viscous Heat-Conducting Gaseous Flows”. In: *Computational Fluid Dynamics 2010: Proceedings of the Sixth International Conference on Computational Fluid Dynamics, ICCFD6, St Petersburg, Russia, on July 12-16, 2010*. Ed. by Alexander Kuzmin. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 653–659. ISBN: 978-3-642-17884-9. DOI: 10.1007/978-3-642-17884-9_83. URL: http://dx.doi.org/10.1007/978-3-642-17884-9_83.

[42] James McDonald and Manuel Torrilhon. “Affordable robust moment closures for CFD based on the maximum-entropy hierarchy”. In: *Journal of Computational Physics* 251 (2013), pp. 500–523.

[43] Lawrence R Mead and Nikos Papanicolaou. “Maximum entropy in the problem of moments”. In: *Journal of Mathematical Physics* 25.8 (1984), pp. 2404–2417.

[44] Eckart Meiburg. “Comparison of the molecular dynamics method and the direct simulation Monte Carlo technique for flows around simple geometries”. In: *The Physics of fluids* 29.10 (1986), pp. 3107–3113.

[45] HM Mott-Smith. “The solution of the Boltzmann equation for a shock wave”. In: *Physical Review* 82.6 (1951), p. 885.

[46] Taku Ohwada. “Structure of normal shock waves: Direct numerical analysis of the Boltzmann equation for hard-sphere molecules”. In: *Physics of Fluids A: Fluid Dynamics* 5.1 (1993), pp. 217–234.

[47] Marc A Rieffel. “A method for estimating the computational requirements of DSMC simulations”. In: *Journal of Computational Physics* 149.1 (1999), pp. 95–113.

[48] Thomas E Schwartzentruber, Leonardo C Scalabrin, and Iain D Boyd. “Hybrid particle-continuum simulations of nonequilibrium hypersonic blunt-body flowfields”. In: *Journal of Thermophysics and Heat Transfer* 22.1 (2008), pp. 29–37.

[49] Claude Elwood Shannon. “A mathematical theory of communication”. In: *ACM SIGMOBILE Mobile Computing and Communications Review* 5.1 (2001), pp. 3–55.

[50] James Alexander Shohat and Jacob David Tamarkin. *The problem of moments*. 1. American Mathematical Soc., 1943.

[51] Stefan K Stefanov. “On DSMC calculations of rarefied gas flows with small number of particles in cells”. In: *SIAM Journal on Scientific Computing* 33.2 (2011), pp. 677–702.

[52] Myron Tribus. “Chapter Five - The Principle of Maximum Entropy”. In: *Rational Descriptions, Decisions and Designs*. Ed. by Myron Tribus. Pergamon Unified Engineering Series. Pergamon, 1969, pp. 119–179.

[53] M Usami et al. “DSMC Calculation of Karman Vortex Flow and Taylor Vortex Flow by an Improved New Collision Scheme (U-system 3)”. In: *AIP Conference Proceedings*. Vol. 1084. 1. AIP. 2008, pp. 365–370.

[54] Wolfgang Wagner. “A convergence proof for Bird’s direct simulation Monte Carlo method for the Boltzmann equation”. In: *Journal of Statistical Physics* 66.3 (1992), pp. 1011–1044.

[55] W. Weiss. “Continuous shock structure in extended thermodynamics”. In: *Phys. Rev. E* 52 (6 Dec. 1995), R5760–R5763. DOI: 10.1103/PhysRevE.52.R5760. URL: http://link.aps.org/doi/10.1103/PhysRevE.52.R5760.

[56] David M Young and Robert Todd Gregory. *A survey of numerical mathematics*. Vol. 1. Courier Corporation, 1988.