HIGH-ORDER, ENTROPY-BASED CLOSURES FOR LINEAR TRANSPORT IN SLAB GEOMETRY II: A COMPUTATIONAL STUDY OF THE OPTIMIZATION PROBLEM

GRAHAM ALLDREDGE*, CORY D. HAUCK†, AND ANDRÉ L. TITS‡

Abstract. We present a numerical algorithm to implement entropy-based (M_N) moment models in the context of a simple, linear kinetic equation for particles moving through a material slab. The closure for these models—as is the case for all entropy-based models—is derived through the solution of constrained, convex optimization problem.

The algorithm has two components. The first component is a discretization of the moment equations which preserves the set of realizable moments, thereby ensuring that the optimization problem has a solution (in exact arithmetic). The discretization is a second-order kinetic scheme which uses MUSCL-type limiting in space and a strong-stability-preserving, Runge-Kutta time integrator. The second component of the algorithm is a Newton-based solver for the dual optimization problem, which uses an adaptive quadrature to evaluate integrals in the dual objective and its derivatives. The accuracy of the numerical solution to the dual problem plays a key role in the time step restriction for the kinetic scheme.

We study in detail the difficulties in the dual problem that arise near the boundary of realizable moments, where quadrature formulas are less reliable and the Hessian of the dual objection function is highly ill-conditioned. Extensive numerical experiments are performed to illustrate these difficulties. In cases where the dual problem becomes "too difficult" to solve numerically, we propose a regularization technique to artificially move moments away from the realizable boundary in a way that still preserves local particle concentrations. We present results of numerical simulations for two challenging test problems in order to quantify the characteristics of the optimization solver and to investigate when and how frequently the regularization is needed.

Keywords: *convex optimization, realizability, kinetic theory, transport, entropy-based closures, moment equations.*

AMS classification: 82C70, 35L02, 49M15, 65D32, 65M08

1. Introduction. In transport and kinetic theory, entropy-based closures are used to derive moment models which retain fundamental properties of the underlying kinetic equations such as hyperbolicity, entropy dissipation, and positivity. The resulting models have been studied extensively in the areas of extended thermodynamics [18,48], gas dynamics [24,29,32,33,40,42,57], semiconductors [2–5,26,31,34,41,56], quantum fluids [16, 19], radiative transport [9–12, 20–22, 28, 30, 46, 47, 59, 62], and phonon transport in solids [19]. In spite of their attractive mathematical properties and wide application, entropy methods still suffer from several short-comings. Among these is the issue of whether the closure can generate all physically possible, or *realizable*, moments. Roughly speaking, a vector is realizable if it is the moment of a kinetic distribution. In some applications, there are realizable vectors (on the boundary of the set of realizable moments) for which the defining optimization problem has no solution [29, 32, 33, 57] and so the closure is not well-defined.

*Department of Electrical and Computer Engineering & Institute for Systems Research, University of Maryland College Park, MD 20742 USA, (gwa@umd.edu)

†Computational Mathematics Group, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831 USA, (hauckc@ornl.gov). This author's research was sponsored by the Office of Advanced Scientific Computing Research and performed at the Oak Ridge National Laboratory, which is managed by UT-Battelle, LLC under Contract No. De-AC05-00OR22725.

‡Department of Electrical and Computer Engineering & Institute for Systems Research, University of Maryland College Park, MD 20742 USA, (andre@umd.edu). The work of the first and third authors was supported by the U.S. Department of Energy, under Grant DESC0001862

Even when the closure is well-defined for all realizable moments, as in the application considered here, there are significant computational challenges. Entropy-based closures require the solution to a constrained, convex optimization problem at each point in a space-time mesh which, in most cases, must be solved numerically (and is usually done so via the associated dual problem). Indeed, when compared to conventional closures with simple algebraic forms, the local computational cost of the entropy-based approach can be quite high. However, the approach still has practical merit for large-scale, massively parallel computations that one might see in a complex multi-physics application. This is due to the emerging paradigm in parallel computing in which data transfer—not floating point operations—is the bottleneck to efficient computation [1, 44, 63]. In particular, though entropy minimization requires many expensive function evaluations, the solver that updates the moment equations requires the same amount of data transfer between computational cells as it would for a conventional, algebraic closure.

The bottleneck to implementing a well-defined entropy-based closure lies in the solution of the dual problem for moments near the boundary of realizability, where the condition number of the Hessian of the dual objective function can be arbitrarily large. The situation is further complicated by the fact that the set of realizable moments may not be closed under the action of the numerical solver for the moment equations. Thus realizable moments near the realizable boundary may lead to values at the next time step which are not realizable.

In the current work, we compute numerical solutions for entropy-based moment systems that approximate linear transport in slab geometries. Similar calculations for low-order systems can be found in [9, 10, 20–22, 46, 47, 62]. Here we present a follow-up to the simulations of higher-order systems based on the Maxwell-Boltzmann entropy that were performed in [28]. However in this paper, we focus in much more detail on the optimization problem near the realizable boundary and on the coupling between the optimization problem and the numerical method for solving the moment equations. This includes, on the numerical-solver side, the design of a kinetic scheme that preserves realizability and, on the optimization side, the use of adaptive quadrature and a regularization technique which keeps moments away from the realizable boundary.

The remainder of the paper is organized as follows. In Section 2, we introduce the linear kinetic transport equation, recall the derivation of the entropy-based moment model, and describe the numerical method for solving the moment equations. In Section 3, we discuss difficulties in the solution of the optimization problem. In Section 4, we detail how our implementation addresses these difficulties. In Section 5, we report numerical results, and in Section 6, we state conclusions and discuss further work.

2. Linear Kinetic Equations and Entropy-Based Closures.

2.1. Kinetic Equation. As in [27], we consider the migration of particles with unit speed that are absorbed by or scattered isotropically off of a background material medium with slab geometry. In a kinetic description, the particle system is characterized by non-negative kinetic density $F = F(x, \mu, t)$ that is governed by a kinetic transport equation of the form

$$\partial_t F + \mu \partial_x F + \sigma_{\mathrm{t}} F = \frac{\sigma_{\mathrm{s}}}{2} \langle F \rangle \,. \tag{2.1}$$

The independent variables in (2.1) are the scalar coordinate $x \in (x_{\mathrm{L}}, x_{\mathrm{R}})$ along the direction perpendicular to the slab, the cosine $\mu \in [-1, 1]$ of the angle between the x-axis and the direction of particle travel, and time t. Interactions with the material are characterized by non-negative variables $\sigma_{\mathrm{s}}(x)$, $\sigma_{\mathrm{a}}(x)$, and $\sigma_{\mathrm{t}}(x) := \sigma_{\mathrm{s}}(x)+\sigma_{\mathrm{a}}(x)$ which are the scattering, absorption, and total cross-sections, respectively. For the purposes of this paper, these cross-sections are assumed to be isotropic, i.e., independent of μ. The angle brackets denote integration over μ, i.e., for any integrable function $g = g(\mu)$,

$$\langle g \rangle := \int_{-1}^{1} g(\mu)\, d\mu \,. \tag{2.2}$$

Equation (2.1) is supplemented by boundary and initial conditions

$$F(x_{\mathrm{L}}, \mu, t) = F_{\mathrm{L}}(\mu, t)\,, \quad \mu > 0\,,\ t \geq 0\,, \tag{2.3a}$$
$$F(x_{\mathrm{R}}, \mu, t) = F_{\mathrm{R}}(\mu, t)\,, \quad \mu < 0\,,\ t \geq 0\,, \tag{2.3b}$$
$$F(x, \mu, 0) = F_0(x, \mu)\,, \quad \mu \in [-1, 1]\,,\ x \in [x_{\mathrm{L}}, x_{\mathrm{R}}]\,, \tag{2.3c}$$

where F_0, F_{L}, and F_{R} are given.

2.2. Entropy-Based Closures. Let $\mathbb{P}_N$ be the space of polynomials with degree at most N and let $\mathbf{m} : [-1, 1] \to \mathbb{R}^{N+1}$ be a vector-valued function whose (linearly independent) components form a basis for $\mathbb{P}_N$. Exact equations for the moments

$$\mathbf{u}(x, t) = [u_0, \ldots, u_N]^T := \langle \mathbf{m} F(x, \cdot, t) \rangle \tag{2.4}$$

are found by multiplying the kinetic equation (2.1) by $\mathbf{m}$ and integrating over all angles. This gives the system

$$\partial_t \mathbf{u} + \partial_x \langle \mu \mathbf{m} F \rangle + \sigma_{\mathrm{t}} \mathbf{u} = \sigma_{\mathrm{s}} Q \mathbf{u}\,, \tag{2.5}$$

where the $(N+1)\times(N+1)$ matrix Q is such that $Q\langle \mathbf{m} g \rangle = \langle \mathbf{m} \langle g \rangle \rangle / 2$ for all functions $g \in L^1(d\mu)$.

Entropy-based methods close the system (2.5) by approximating F with an *ansatz* $\mathcal{F}(\mathbf{u}(x, t), \mathbf{m}(\mu))$ which, for given (x, t), solves the constrained, strictly convex optimization problem

$$\underset{g}{\text{minimize}} \ \ \langle \eta(g) \rangle \qquad \text{subject to } \ \langle \mathbf{m} g \rangle = \mathbf{u}\,. \tag{2.6}$$

Here the minimization is with respect to $f \colon [-1, 1] \to \mathbb{R}$ and $\eta \colon \mathbb{R} \to \mathbb{R}$ is strictly convex. If a minimizer exists, it takes the form (see [40])

$$\mathcal{F}(\mathbf{u}, \mathbf{m}) = G_{\hat{\boldsymbol{\alpha}}(\mathbf{u})}\,, \quad G_{\boldsymbol{\alpha}} := \eta_*' \left(\boldsymbol{\alpha}(\mathbf{u})^T \mathbf{m} \right)\,, \tag{2.7}$$

where $\eta_* \colon \mathbb{R} \to \mathbb{R}$ is the Legendre dual of η, η_*' is its derivative, and the vector of Lagrange multipliers (also called "dual variables") $\hat{\boldsymbol{\alpha}}(\mathbf{u}) \in \mathbb{R}^{N+1}$ solves the dual problem

$$\underset{\boldsymbol{\alpha} \in \mathbb{R}^{N+1}}{\text{minimize}} \ \ \left\{ \left\langle \eta_*'(\boldsymbol{\alpha}^T \mathbf{m}) \right\rangle - \boldsymbol{\alpha}^T \mathbf{u} \right\}\,. \tag{2.8}$$

In this paper we focus on the Maxwell-Boltzmann entropy $\eta(z) = z \log(z) - z$. Thus $\eta_*(y) = \eta_*'(y) = e^y$ and

$$G_{\boldsymbol{\alpha}}(\mu) = \exp(\boldsymbol{\alpha}^T \mathbf{m}(\mu)). \tag{2.9}$$

Let $\mathbf{f}$ be the flux defined by the entropy-based ansatz:

$$\mathbf{f}(\mathbf{u}) := \left\langle \mu \mathbf{m} G_{\hat{\boldsymbol{\alpha}}(\mathbf{u})} \right\rangle . \tag{2.10}$$

It is straight-forward to verify (following arguments in [20, 40], for example) that, when written in terms of $\hat{\boldsymbol{\alpha}}$, the closed moment system

$$\partial_t \mathbf{u} + \partial_x \mathbf{f}(\mathbf{u}) + \sigma_{\mathrm{t}} \mathbf{u} = \sigma_{\mathrm{s}} Q \mathbf{u} \tag{2.11}$$

is symmetric hyperbolic and that, when $\sigma_{\mathrm{a}} = 0$ (corresponding to an isolated physical system), it dissipates the strictly convex entropy $h(\mathbf{u}) := \langle \eta(G_{\hat{\boldsymbol{\alpha}}(\mathbf{u})}) \rangle$. For $\mathbf{u} \in \mathbb{R}^{N+1}$, (2.11) is commonly referred to as the *M_N model*.

Correct boundary conditions for the moment system (which involve integration of (2.3) over the entire μ space) are not easily determined because kinetic data is only given for values of μ which correspond to incoming data. Indeed, the issue of proper boundary conditions remains an open question, although some progress has been made for linear systems [36, 37, 39, 55]. We discuss our implementation of the boundaries in Section 2.4.

2.3. Realizability. One of the most challenging aspects of entropy-based closures is the issue of *realizability*.

DEFINITION 2.1. *Let the vector-valued function $\mathbf{m}$ be given and let $L^1_+(d\mu)$ be the set of all non-negative Lebesgue integrable functions g such that $\langle g \rangle > 0$. A vector $\mathbf{v}$ is said to be* realizable *(with respect to $\mathbf{m}$) if there exists a $g \in L^1_+(d\mu)$, such that $\langle \mathbf{m} g \rangle = \mathbf{v}$. The set of all realizable vectors is denoted by $\mathcal{R}_{\mathbf{m}}$.*

The following theorem characterizes the set $\mathcal{R}_{\mathbf{m}}$ when the components of $\mathbf{m}$ are monomials. It is a classical result in the theory of *reduced moments*; see, for example, [58] and references therein.

THEOREM 2.2. *Let $\mathbf{p} = [1, \mu, \ldots, \mu^N]^T$. A necessary and sufficient condition for a vector $\mathbf{v}$ to be realizable with respect to $\mathbf{p}$ is that*

1. *in the case that N is odd, the $(N+1)/2 \times (N+1)/2$ matrices $B^{\pm}$, defined by*

$$B^{\pm}_{kl} := v_{k+l} \pm v_{k+l+1} , \quad k, l \in \{0, \ldots, (N-1)/2\} , \tag{2.12}$$

   are positive definite;
2. *in the case that N is even, the $(N+2)/2 \times (N+2)/2$ matrix B^0 and the $N/2 \times N/2$ matrix B^1, defined by*

$$\begin{aligned} B^0_{kl} &:= v_{k+l} , & k, l &\in \{0, \ldots, N/2\} , \\ B^1_{kl} &:= v_{k+l} - v_{k+l+2} , & k, l &\in \{0, \ldots, (N-2)/2\} , \end{aligned}$$

   are positive definite.

The realizability of moments with respect to any vector-valued function $\mathbf{m}$ whose components form a basis for $\mathbb{P}^N$ can be determined by simply applying a change of basis from $\mathbf{m}$ to $\mathbf{p}$ and then invoking Theorem 2.2. The next theorem characterizes the geometry of $\mathcal{R}_{\mathbf{m}}$.

THEOREM 2.3. *The set $\mathcal{R}_{\mathbf{m}}$ is a convex cone, i.e., for any moments $\mathbf{v}_1, \mathbf{v}_2 \in \mathcal{R}_{\mathbf{m}}$ and nonnegative constraints c_1 and c_2, with $c_1 + c_2 > 0$, $c_1 \mathbf{v}_1 + c_2 \mathbf{v}_2 \in \mathcal{R}_{\mathbf{m}}$. Furthermore, it is open.*

Proof. The fact that $\mathcal{R}_{\mathbf{m}}$ is a convex cone follows from the fact that $L^1_+(d\mu)$ is also a convex cone. Openness is an corollary of Theorem 2.4 below. □

The issue of realizability raises three questions:

1. Does a solution to (2.6) exist for all $\mathbf{u} \in \mathcal{R}_\mathbf{m}$?
2. If solutions do exist, is the set $\mathcal{R}_\mathbf{m}$ invariant under the dynamics of (2.11)?
3. If such an invariance property holds, can it be preserved at the numerical level?

In general the answer to the first question is "no." The lack of existence is related to the fact that the constraints in (2.6) are not always continuous in the L^1 norm [29, 32, 33, 57]. However, for the case under consideration, the domain of integration is bounded and the components of $\mathbf{m}$ are bounded on that domain. These properties ensure that L^1 continuity holds and a solution exists, leading to the following theorem.

THEOREM 2.4 ([33]). *The function $\hat{\boldsymbol{\alpha}}$ which maps moments $\mathbf{u}$ to dual variables $\boldsymbol{\alpha}$ via the solution of* (2.8) *is a smooth bijection from $\mathcal{R}_\mathbf{m}$ onto $\mathbb{R}^{N+1}$. Its inverse is the moment map $\hat{\mathbf{v}}$ given by*

$$\hat{\mathbf{v}}(\boldsymbol{\alpha}) = \langle \mathbf{m} G_{\boldsymbol{\alpha}} \rangle \, . \tag{2.13}$$

Proof. See [33] and also [7, 32, 45] for similar results. □

Notwithstanding Theorem 2.4, realizability presents significant numerical challenges. Indeed, near the boundary of $\mathcal{R}_\mathbf{m}$, the Hessian of the dual objective (2.8) is ill-conditioned at the solution. This is a consequence of the fact that, on the boundary itself, the constraint equations are uniquely solved by atomic measure—that is, an ansatz made up of delta functions [14].

The question of whether $\mathcal{R}_\mathbf{m}$ is invariant under the dynamics of (2.11) remains open, although partial results can be found in [13]. However, one may at a numerical level enforce the invariance property. We do so below using an appropriate kinetic scheme.

For the remainder of the paper, we will follow the common convention and assume that the components of $\mathbf{m}$ are the first $N+1$ Legendre polynomials, which are orthogonal on $[-1, 1]$ with respect to L^2. With this choice of basis, the entries of Q are given by

$$Q_{lm} = \delta_{lm}\delta_{l0} \, , \tag{2.14}$$

δ_{lm} being the Kronecker δ, so that $Q\mathbf{u} = [u_0, \ldots, 0]^T$.

2.4. Numerical Integration of the Moment System. We implement a numerical solution to (2.11) using a kinetic scheme [17, 19, 25, 53, 54] which is second-order in both space and time. In the context of entropy-based closures, the main benefit of this scheme is that it preserves realizability. In addition, it avoids the direct computation of eigenvalues and (approximate) Riemann solvers [61] which, for most entropy-based moment systems, is expensive due to the complicated relationship between $\mathbf{u}$ and $\mathbf{f}$ in (2.10).

Let $\Delta x = (x_R - x_L)/N_x$ and $\Delta t > 0$ be given mesh parameters, and let $\{x_j\}_{j=-1}^{N_x+2} \times \{t^n\}_{n=0}^{N_t}$ be a uniform space-time mesh defined by $x_j := x_\mathrm{L} + (j - 0.5)\Delta x$ and $t^n := n\Delta t$. The values x_j define the centers of contiguous spatial cells $I_j := (x_{j-1/2}, x_{j+1/2})$, where $x_{j\pm1/2} := x_j \pm \Delta x/2$. The cells with indices $j \in \{-1, 0, N_x + 1, N_x + 2\}$ are "ghost cells", which are not part of the physical domain but are used to implement boundary conditions.

We approximate $\mathbf{u}$ numerically via its the cell averages, letting

$$\frac{1}{\Delta x}\int_{I_j} \mathbf{u}(x,t)\, dx \simeq \mathbf{u}_j(t) \, , \quad j \in \{-1, \ldots, N_x + 2\} \, . \tag{2.15}$$

The semi-discrete, numerical scheme for (2.11) which defines $\mathbf{u}_j$ on the interior of the domain is

$$\partial_t \mathbf{u}_j + \frac{\mathbf{f}_{j+1/2} - \mathbf{f}_{j-1/2}}{\Delta x} + \sigma_\mathrm{t}\mathbf{u}_j = \sigma_\mathrm{s} Q \mathbf{u}_j \,, \quad j \in \{1, \ldots, N_x\} \,, \tag{2.16}$$

the numerical flux $\mathbf{f}_{j+1/2}$ being given by

$$\mathbf{f}_{j+1/2} := \langle \mu \mathbf{m} \hat{G}_{j+1/2} \rangle \,, \quad j \in \{0, \ldots, N_x\} \,, \tag{2.17}$$

and $\hat{G}_{j+1/2}$ is an approximation of the entropy ansatz at the cell edge. These edge values are defined based on the sign of μ, via up-winding:

$$\hat{G}_{j+1/2}(\mu, t) := \begin{cases} \hat{G}_j(\mu, t) + \frac{\Delta x}{2} \hat{s}_j(\mu, t) \,, & \mu > 0 \\ \hat{G}_{j+1}(\mu, t) - \frac{\Delta x}{2} \hat{s}_{j+1}(\mu, t) \,, & \mu < 0 \end{cases} \,, \quad j \in \{0, \ldots, N_x\} \,, \tag{2.18}$$

where $\hat{G}_j$ is the entropy ansatz associated to $\mathbf{u}_j$ via (2.7):

$$\hat{G}_j(\mu, t) := G_{\hat{\boldsymbol{\alpha}}(\mathbf{u}_j(t))}(\mu) \,, \quad j \in \{-1, \ldots, N_x + 2\} \,. \tag{2.19}$$

For $j \in \{0, \ldots, N_x + 1\}$, the quantity $\hat{s}_j$ is an approximation of the spatial derivative of $\hat{G}$ in cell I_j:

$$\hat{s}_j = \mathrm{minmod} \left\{ \theta \frac{\hat{G}_j - \hat{G}_{j-1}}{\Delta x}, \frac{\hat{G}_{j+1} - \hat{G}_{j-1}}{2\Delta x}, \theta \frac{\hat{G}_{j+1} - \hat{G}_j}{\Delta x} \right\} \,, \tag{2.20}$$

where $1 \leq \theta \leq 2$ [38, 49].[(1)] The minmod function selects the real number with smallest absolute value in the convex hull of its arguments. (Note that, in (2.20), $\hat{G}_j$ is needed for $j \in \{-1, 0, \ldots, N_x + 2\}$, which shows the need for four ghost cells, two on each side of $(x_\mathrm{L}, x_\mathrm{R})$.)

As mentioned in Section 2.2, boundary conditions for moment systems remain an open question. In our implementation, we prescribe boundary conditions by specifying moments on the four ghost cells. For periodic boundaries used in some test problems, we simply set

$$\mathbf{u}_{-1}(t) = \mathbf{u}_{N_x - 1}(t) \,, \quad \mathbf{u}_0(t) = \mathbf{u}_{N_x}(t) \,, \quad \mathbf{u}_{N_x+1}(t) = \mathbf{u}_1(t) \,, \quad \mathbf{u}_{N_x+2}(t) = \mathbf{u}_2(t) \,. \tag{2.21}$$

For physical boundary conditions, moments in ghost cells are defined by extending the definitions of $F_\mathrm{L}(\mu, t)$ and $F_\mathrm{R}(\mu, t)$ to all μ and then taking moments:

$$\mathbf{u}_{-1}(t) = \mathbf{u}_0(t) = \langle \mathbf{m} F_\mathrm{L}(\mu, t) \rangle \,, \qquad \mathbf{u}_{N_x+1}(t) = \mathbf{u}_{N_x+2}(t) = \langle \mathbf{m} F_\mathrm{R}(\mu, t) \rangle \,. \tag{2.22}$$

While reasonable, this is clearly not the only option. Further discussion of this issue is given in [27, 60].

To integrate (2.16) in time, we use the optimal, second-order strong-stability-preserving Runge-Kutta (SSP-RK2) method [23], also known as Heun's method or the improved Euler method. We approximate $\mathbf{u}_j(t^n) \simeq \mathbf{u}_j^n$ and let $\mathbf{u}^n$ denote the

[1] Any value of $\theta \in [1, 2]$ will yield a second-order scheme and, roughly speaking, larger values of θ decrease numerical diffusion in the scheme. When $\theta = 1$, monotonic cell averages yield monotonic reconstructions $G_j(\mu, t) + s_j(\mu, t)(x - xj)$. When $\theta = 2$, edge values are bounded by neighboring cell averages.

array containing $\{\mathbf{u}_j^n\}_{j=-1}^{N_x+1}$. For (2.16) with (2.21) or (2.22) in the abstract form $\partial_t \mathbf{u} = L(\mathbf{u})$, the SSP-RK2 method with initial stage $\mathbf{u}^{(0)} := \mathbf{u}^n$ is given by

$$\mathbf{u}^{(1)} := \mathbf{u}^{(0)} + \Delta t L(\mathbf{u}^n), \quad \mathbf{u}^{(2)} := \mathbf{u}^{(1)} + \Delta t L(\mathbf{u}^{(1)}), \quad \mathbf{u}^{n+1} := \frac{1}{2}\left(\mathbf{u}^{(0)} + \mathbf{u}^{(2)}\right) \tag{2.23}$$

for all $n \in \{0, \ldots, N_t - 1\}$.

As discussed in [27], kinetic scheme (2.16)–(2.23) is very inefficient in diffusive regimes, where σ_t is large. In such regimes, accuracy requirements dictate that the spatial and temporal mesh depends inversely on σ_t, even though the solution profile varies on an $O(1)$ scale. However, for the test cases considered later in this paper, σ_t is an $O(1)$ quantity.

2.5. Maintaining realizability. The kinetic scheme (2.16)–(2.23) invokes a solution of the dual problem (2.8) in (2.18) and (2.20), via (2.19). The fact that the dual problem can only be solved approximately must be take into consideration when attempting to maintain realizability of the moments in the numerical solution. We let $\bar{\boldsymbol{\alpha}}(\mathbf{u})$ denote the approximate solution and introduce the associated notation

$$\bar{G}_j(\mu, t) := G_{\bar{\boldsymbol{\alpha}}(\mathbf{u}_j(t))}(\mu, t). \tag{2.24}$$

The use of this approximate solution to the optimization problem means replacing the definition for the numerical flux in (2.17) by

$$\mathbf{f}_{j+1/2} := \left\langle \mu \mathbf{m} \bar{G}_{j+1/2} \right\rangle , \tag{2.25}$$

where $\bar{G}_{j+1/2}$ is computed just as $\hat{G}_{j+1/2}$ in (2.18), but replacing $\hat{G}_j$ by $\bar{G}_j$. The corresponding slopes (see (2.20)) are denoted by $\bar{s}_j$.

We now prove that with an appropriate time-step restriction and appropriate boundary conditions, the resulting kinetic scheme preserves realizable moments. It turns out that with an inexact solution to the dual problem (2.8), the ratios between the ansatzes $G_{\bar{\boldsymbol{\alpha}}}$ and $G_{\hat{\boldsymbol{\alpha}}}$ at each stage of the Runge-Kutta scheme play a key role. We therefore define at each time step t^n

$$\gamma_j^{(m)} := \left(\frac{\bar{G}_j^{(m)}}{\hat{G}_j^{(m)}} \right) , \; m \in \{0, 1\} , \quad \text{and} \quad \gamma_{\max} := \max_{\substack{m \in \{0,1\} \\ j \in \{-1, \ldots, N_x+2\} \\ \mu \in [-1,1]}} \{\gamma_j^{(m)}(\mu)\} , \tag{2.26}$$

where $\hat{G}_j^{(m)} := G_{\hat{\boldsymbol{\alpha}}(\mathbf{u}_j^{(m)})}$ and $\bar{G}_j^{(m)} := G_{\bar{\boldsymbol{\alpha}}(\mathbf{u}_j^{(m)})}$, $\mathbf{u}_j^{(m)}$ being the jth subvector of $\mathbf{u}^{(m)}$, defined in (2.23).[2]

THEOREM 2.5. *Suppose that $\mathbf{u}_j^n \in \mathcal{R}_\mathbf{m}$ for $j \in \{-1, \ldots, N_x + 2\}$. If $\mathbf{u}^{n+1}$ is defined via the kinetic scheme* (2.16),(2.25),(2.18)–(2.23) *with bars replacing hats in (23)-(25) and with time-step restriction*

$$\gamma_{max} \frac{\Delta t}{\Delta x} \frac{2 + \theta}{2} + \sigma_\mathrm{t} \Delta t < 1 \tag{2.27}$$

and if the moments in the ghost cells are realizable at each stage of the Runge-Kutta scheme (2.23)*, then $\mathbf{u}_j^{n+1} \in \mathcal{R}_\mathbf{m}$ for $j \in \{1, \ldots, N_x\}$.*

[2] Here we suppress the dependence on n for clarity.

Proof. We show for $m \in \{1,2\}$ that $\mathbf{u}_j^{(m-1)} \in \mathcal{R}_\mathbf{m}$ for $j \in \{-1,\ldots,N_x+2\}$ implies $\mathbf{u}_j^{(m)} \in \mathcal{R}_\mathbf{m}$ for $j \in \{1,\ldots,N_x\}$. Realizability for the subvectors of $\mathbf{u}^{n+1}$ then follows from (2.23) and Theorem 2.3 (convexity of $\mathcal{R}_\mathbf{m}$). The key point is to observe that

$$\mathbf{u}_j^{(m)} = \langle \mathbf{m}\Phi_j^{(m)} \rangle \,, \quad j \in \{1,\ldots N_x\} \,,\ m \in \{1,2\} \,, \tag{2.28}$$

where

$$\Phi_j^{(m)} := \hat{G}_j^{(m-1)} - \mu\frac{\Delta t}{\Delta x}\left(\bar{G}_{j+1/2}^{(m-1)} - \bar{G}_{j-1/2}^{(m-1)}\right) + \Delta t\left(-\sigma_\mathrm{t}\hat{G}_j^{(m-1)} + \frac{\sigma_\mathrm{s}}{2}\left\langle \hat{G}_j^{(m-1)} \right\rangle\right) . \tag{2.29}$$

Thus one need only show that $\Phi_j^{(m)} \geq 0$. Stripping away positive terms on the right-hand side of (2.29) gives

$$\Phi_j^{(m)} \geq \hat{G}_j^{(m-1)} - \mu\frac{\Delta t}{\Delta x}\bar{G}_{j+1/2}^{(m-1)} - \Delta t\sigma_\mathrm{t}\hat{G}_j^{(m-1)} \,. \tag{2.30}$$

Assume $\mu \geq 0$. (The case $\mu < 0$ follows from an analogous argument.) If $\bar{s}_j > 0$ (so all arguments of the minmod in (2.20) are non-negative), we have (with bars instead of hats)

$$\bar{s}_j^{(m-1)} \leq \theta\frac{\bar{G}_j^{(m-1)} - \bar{G}_{j-1}^{(m-1)}}{\Delta x} \tag{2.31}$$

so that, using (2.18),

$$\bar{G}_{j+1/2}^{(m-1)} \leq \left(1+\frac{\theta}{2}\right)\bar{G}_j^{(m-1)} - \frac{\theta}{2}\bar{G}_{j-1}^{(m-1)} \leq \frac{2+\theta}{2}\bar{G}_j^{(m-1)} \,. \tag{2.32}$$

Substituting (2.32) into (2.30) and invoking the definition of $\gamma_j^{(m-1)}$ from (2.26) gives

$$\Phi_j^{(m)} \geq \left(1 - \mu\gamma_j^{(m-1)}\frac{\Delta t}{\Delta x}\frac{2+\theta}{2} - \Delta t\sigma_\mathrm{t}\right)\hat{G}_j^{(m-1)}. \tag{2.33}$$

From (2.33), it is clear that (2.27) implies non-negativity of $\Phi_j^{(m)}$. On the other hand, if $\bar{s}_j \leq 0$, we obtain

$$\Phi_j^{(m)} \geq \left(1 - \mu\gamma_j^{(m-1)}\frac{\Delta t}{\Delta x} - \Delta t\sigma_\mathrm{t}\right)\hat{G}_j^{(m-1)}. \tag{2.34}$$

The positivity of the left-hand side of (2.34) is guaranteed by the condition

$$\mu\gamma_j^{(m-1)}\frac{\Delta t}{\Delta x} + \Delta t\sigma_\mathrm{t} < 1 \,, \tag{2.35}$$

which is weaker than (2.27). This concludes the proof. □

REMARK 1. *The reader should note the following:*

1. *The proof of Theorem 2.5 does not depend on the specific form of $\hat{G}_j$ or $\bar{G}_j$, only on the fact that they are positive. Thus the theorem applies to different types of closures and different types of numerical error, so long as positivity of the two approximations is maintained.*

2. *Setting $\gamma_{max} = 1$ recovers the time-step restriction for the case when there is no error in approximating the ansatz. If further $\sigma_t = 0$, then the corresponding time step restriction is exactly what is required to maintain positivity for a single Euler step applied to a semi-discrete MUSCL scheme [51] for a linear advection equation with speed one (the maximum value of $|\mu|$). This is not by accident; in this case, the kinetic scheme* (2.16)–(2.23) *is equivalent to the moments of a semi-discrete MUSCL scheme for the transport equation* (2.1) *with initial condition $\hat{G}$.*
3. *The quantity γ_{max} depends on the solution values at the intermediate Runge-Kutta stage $\mathbf{u}^{(1)}$. This leaves a user with two options: either (i) set an upper bound for γ_{max} to determine a suitable Δt and then require that the optimization error for every cell and every stage is below that bound or (ii) check the error at the intermediate stage and, if it is too high, exit the Runge-Kutta algorithm, go back to the previous time step, and choose a smaller value for Δt. In our implementation, we take the former approach.*
4. *Equation* (2.33) *shows that the less conservative definition of γ_{max} given by $\gamma'_{max} := \max_{m,j,\mu}\{\mu\gamma_j^{(m)}(\mu)\}$ is sufficient to guarantee nonnegativity. This definition was not used in our implementation.*

3. A Study of the Optimization Problem. Solving the dual problem (2.8) is by far the most computationally intensive part of implementing the entropy-based closure. At the outset, we note the following:

1. The optimization routine is always applied to moments which are normalized by dividing by the zeroth-order moment u_0 (the local particle concentration), noting that,
$$\hat{\boldsymbol{\alpha}}(\mathbf{u}) = \hat{\boldsymbol{\alpha}}(\mathbf{u}/u_0) + [\log(u_0), 0, \ldots, 0]^T. \tag{3.1}$$
This normalization makes it simpler to specify tolerances and analyze performance.
2. All calculations are performed in double precision arithmetic, where machine precision—that is, the maximum possible relative error in representing a number as a floating point is $2^{-53} \approx 1.11 \times 10^{-16}$. However, the analysis presented in the remainder of the paper can be easily applied to any floating point system.

3.1. Basics of the optimization. We denote the objective function in (2.8) and its gradient and Hessian, respectively, by

$$f(\boldsymbol{\alpha}) := \langle G_{\boldsymbol{\alpha}} \rangle - \boldsymbol{\alpha}^T \mathbf{u}\,, \quad \mathbf{g}(\boldsymbol{\alpha}) := \langle \mathbf{m} G_{\boldsymbol{\alpha}} \rangle - \mathbf{u}\,, \quad H(\boldsymbol{\alpha}) := \left\langle \mathbf{m}\mathbf{m}^T G_{\boldsymbol{\alpha}} \right\rangle\,. \tag{3.2}$$

Note that f is smooth and strictly convex and H is positive definite for all $\boldsymbol{\alpha}$ and independent of $\mathbf{u}$.

We approach $\hat{\boldsymbol{\alpha}}(\mathbf{u})$ using Newton's method with an Armijo backtracking line search [6] to guarantee global convergence and fast (quadratic) local convergence. Given an initial guess $\boldsymbol{\alpha}_0$, the iterates $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, ... are constructed by

$$\boldsymbol{\alpha}_{k+1} = \boldsymbol{\alpha}_k + \beta^i \mathbf{d}(\boldsymbol{\alpha}_k), \quad k \in \{0, 1, 2, \ldots\}\,, \tag{3.3}$$

where $\mathbf{d}(\boldsymbol{\alpha}_k) := -H^{-1}(\boldsymbol{\alpha}_k)\mathbf{g}(\boldsymbol{\alpha}_k)$ is the Newton direction at $\boldsymbol{\alpha}_k$, $\beta \in (0, 1)$ is the step-size parameter, i is the smallest non-negative integer such that

$$f(\boldsymbol{\alpha}_k + \beta^i \mathbf{d}(\boldsymbol{\alpha}_k)) \leq f(\boldsymbol{\alpha}_k) + \beta^i \xi \mathbf{g}(\boldsymbol{\alpha}_k)^T \mathbf{d}(\boldsymbol{\alpha}_k)\,, \tag{3.4}$$

and $\xi \in (0, 1/2)$. At $t = 0$, we set the initial condition $\boldsymbol{\alpha}_0$ so that $G_{\boldsymbol{\alpha}_0}$ is the isotropic distribution with moment u_0. Assuming the normalization $u_0 = 1$, this means $\boldsymbol{\alpha}_0 = [-\log(2), 0, \dots, 0]^T$. For $t > 0$, the multipliers $\bar{\boldsymbol{\alpha}}$ from the previous time step are the natural choice for the initial condition.

There are two conditions in the stopping criterion. Given parameters $\tau > 0$ and $\varepsilon_\gamma > 0$, we terminate the optimization process at $\boldsymbol{\alpha}_k$ when

$$\|\mathbf{g}(\boldsymbol{\alpha}_k)\| \leq \tau \quad \text{and} \quad \exp\left(5\zeta\|\mathbf{d}(\boldsymbol{\alpha}_k)\|\right) \leq 1 + \varepsilon_\gamma \,. \tag{3.5}$$

Here $\zeta := \max_\mu \|\mathbf{m}(\mu)\|$, and $\|\cdot\|$ is the Euclidean norm. The first condition measures the difference between the moments $\mathbf{u}$ and those of a candidate ansatz $G_{\boldsymbol{\alpha}_k}$ (see (3.2)). The second condition is related to realizability of the moments generated by the kinetic scheme (2.16),(2.25),(2.18)-(2.23). Due to (2.27) in Theorem 2.5, an upper bound on $\gamma_{\max}$ is needed to ensure realizability with a reasonable time step $\Delta t = O(\Delta x)$. We conservatively bound $\gamma_k := G_{\boldsymbol{\alpha}_k}/G_{\hat{\boldsymbol{\alpha}}}$ using the inequality

$$\max_{\mu\in[-1,1]} \gamma_k(\mu) = \max_{\mu\in[-1,1]} \exp((\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}})^T \mathbf{m}(\mu)) \leq \exp\left(\zeta\|\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}\|\right) \,. \tag{3.6}$$

The exact minimizer $\hat{\boldsymbol{\alpha}}$ is unknown, but we can approximate $\|\boldsymbol{\alpha}_k - \hat{\boldsymbol{\alpha}}\| \approx \|\mathbf{d}(\boldsymbol{\alpha}_k)\|$, which is a good asymptotic estimate because Newton's method locally converges quadratically. For our implementation, we further insert a factor of five inside the exponential in the right-hand side of (3.6) to increase our confidence in our upper bound of γ_k. This gives the second condition in (3.5). With this conservative estimate of γ_k, we are typically able to use time steps of at least 90% of the maximum theoretical value for Δt—that is the value of Δt which corresponds to an exact solution of the dual problem and is computed from (2.27) with $\gamma_{\max} = 1$.

3.2. Difficulties Near the Realizable Boundary. Problem (2.8) becomes difficult to solve when the moments $\mathbf{u}$ lie near the boundary $\partial\mathcal{R}_{\mathbf{m}}$ of the set of realizable moments $\mathcal{R}_{\mathbf{m}}$. Such moments are associated with highly anisotropic distributions and/or vacuum states ($F(x, \cdot, t) \equiv 0$) and often occur in the presence of strong sources or when particles enter a void. Refining the spatial mesh in the PDE solver tends to exacerbate the problem since then the sharp dynamics are more fully resolved.

The presence of "near-boundary" moments is challenging for two reasons. First, as mentioned in the introduction, small discretization errors in the PDE solver may generate *unrealizable* moments. This can be overcome with a solver under which $\mathcal{R}_{\mathbf{m}}$ is invariant. The scheme in Section 2.4 is one such solver (assuming the left-hand side of the second condition in (3.5) is indeed an upper bound on $\gamma_{\max}$ in (2.26)). The second issue is that problem (2.8) becomes highly *ill-conditioned* near the boundary of $\mathcal{R}_{\mathbf{m}}$. We next present two examples to illustrate this second point.

Example 1: The M_1 Model. The M_1 model is the simplest example of an entropy-based moment system. It uses only the first two Legendre polynomials: $\mathbf{m} = [m_0, m_1]^T = [1, \mu]^T$. The model was first introduced in [46] in the context of photon radiation and later analyzed in much greater detail in [9]. Unlike the case for most entropy-based models, the relationship between the moments and the multipliers in M_1 can be expressed without the use of integral formulas. This makes M_1 a useful tool for understanding the challenges of solving the dual problem (2.8).

First consider the first-order necessary condition for optimality of the M_1 dual problem. Let $\mathbf{u} = [u_0, u_1]^T$ and $\hat{\boldsymbol{\alpha}}(\mathbf{u}) = [\hat{\alpha}_0(\mathbf{u}), \hat{\alpha}_1(\mathbf{u})]^T$. By solving $\mathbf{g}(\hat{\boldsymbol{\alpha}}(\mathbf{u})) = 0$

with $\mathbf{g}$ given in (3.2), one can show that the optimal multipliers satisfy (see [9])

$$u_0 = \frac{2e^{\hat{\alpha}_0(\mathbf{u})}}{\hat{\alpha}_1(\mathbf{u})}\sinh(\hat{\alpha}_1(\mathbf{u})) \quad \text{and} \quad \frac{u_1}{u_0} = \coth(\hat{\alpha}_1(\mathbf{u})) - \frac{1}{\hat{\alpha}_1(\mathbf{u})}\,. \tag{3.7}$$

A plot of the second relation is given in Figure 3.1(a). The fact that μ is restricted to $[-1,1]$ implies that $|u_1| < u_0$ for any realizable $\mathbf{u}$. Appropriately, the range of $\coth(\alpha_1) - 1/\alpha_1$ is $(-1,1)$. From (3.7), one can show that

$$\hat{\alpha}_0(\mathbf{u}) \to -\infty \quad \text{and} \quad \frac{\hat{\alpha}_1(\mathbf{u})}{\hat{\alpha}_0(\mathbf{u})} \to \text{sign}(u_1) \tag{3.8}$$

as $|u_1|/u_0 \to 1$, while u_0 is held constant. The unbounded growth in the components of $\boldsymbol{\alpha}$ causes numerical overflow and underflow because of the exponential involving $\boldsymbol{\alpha}$ in the objective function and its derivatives (see expression (2.9) for $G_{\boldsymbol{\alpha}}$).

It is also straight-forward to see for M_1 that as $\mathbf{u}$ approaches $\partial\mathcal{R}_{\mathbf{m}}$ the Hessian of the dual problem at $\hat{\boldsymbol{\alpha}}(\mathbf{u})$ becomes singular. Indeed, let $P := \langle \mu^2 G_{\hat{\boldsymbol{\alpha}}(\mathbf{u})}\rangle$. Then a standard calculation shows that the eigenvalues of H are

$$\lambda_\pm = \frac{1}{2}(u_0 + P) \pm |u_1|\sqrt{1 + \left(\frac{u_0 - P}{2u_1}\right)^2} \tag{3.9}$$

and that the bound

$$|u_0 - P| = |\langle (1-\mu^2) G_{\hat{\boldsymbol{\alpha}}(\mathbf{u})}\rangle| \le 2|\langle (1 \pm \mu) G_{\hat{\boldsymbol{\alpha}}(\mathbf{u})}\rangle| = 2|u_0 \pm u_1| \tag{3.10}$$

holds, so that $P \to u_0$ as $u_1 \to \pm u_0$. Thus the ratio λ_+/λ_- tends to ∞ as $u_1 \to \pm u_0$.

The numerical difficulties above are compounded by the fact that, in general, the integrals in the objective, gradient, and Hessian must be approximated by quadrature. Given a set of quadrature points $\mathcal{Q}$, the approximation $f_{\mathcal{Q}}$ has the form

$$f_{\mathcal{Q}}(\boldsymbol{\alpha}) = \sum_{\mu_i \in \mathcal{Q}} w_i G_{\boldsymbol{\alpha}}(\mu_i) - \boldsymbol{\alpha}^T \mathbf{u}, \tag{3.11}$$

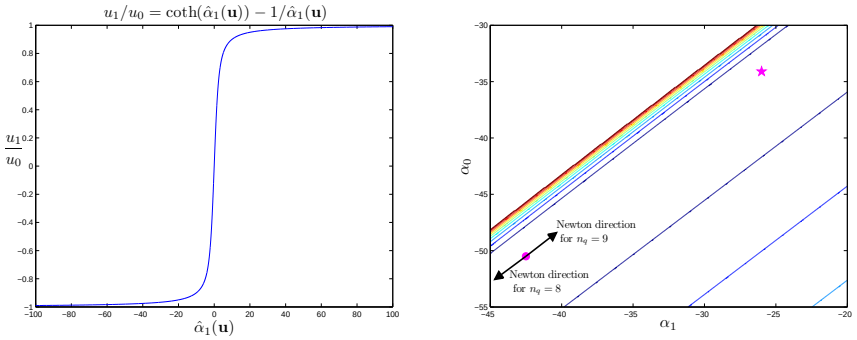
where $w_i > 0$ and $\mu_i \in \mathcal{Q}$ are the quadrature weights and nodes, respectively. For M_1, the first-order necessary conditions for $f_{\mathcal{Q}}$ yield an analog to the second equation of (3.7):

$$\frac{u_1}{u_0} = \frac{\sum w_i \mu_i \exp(\hat{\alpha}_{\mathcal{Q},1}\mu_i)}{\sum w_i \exp(\hat{\alpha}_{\mathcal{Q},1}\mu_i)}, \tag{3.12}$$

where $\hat{\boldsymbol{\alpha}}_{\mathcal{Q}} = [\hat{\alpha}_{\mathcal{Q},0}, \hat{\alpha}_{\mathcal{Q},1}]^T$ denotes the minimizer of $f_{\mathcal{Q}}$. Assuming the quadrature contains at least one node $\mu_i < 0$ and at least one node $\mu_i > 0$, consideration of the range of the right-hand side of (3.12) (with respect to $\hat{\alpha}_{\mathcal{Q},1}$) shows that (3.12) is solvable if and only if

$$\min_{\mu_i \in \mathcal{Q}}\{\mu_i\} < \frac{u_1}{u_0} < \max_{\mu_i \in \mathcal{Q}}\{\mu_i\}. \tag{3.13}$$

Thus the existence of a minimizer of $f_{\mathcal{Q}}$ depends on $\mathcal{Q}$ and on how close $\mathbf{u}$ is to the realizable boundary.



(a) The graph of (3.7), showing the relationship between $\frac{u_1}{u_0}$ and $\hat{\alpha}_1(\mathbf{u})$.

(b) Contours of the objective function for the moments given in Equation (3.15).

FIG. 3.1. *Illustrating difficulties in the optimization.*

The Hessian of the approximate objective function, used in calculating the Newton direction, is given by the following sum of rank-one matrices:

$$H_{\mathcal{Q}}(\boldsymbol{\alpha}) = \sum_{\mu_i \in \mathcal{Q}} w_i G_{\boldsymbol{\alpha}}(\mu_i) \mathbf{m}(\mu_i) \mathbf{m}^T(\mu_i). \tag{3.14}$$

Suppose that $\mathbf{u}$ is near $\partial \mathcal{R}_{\mathbf{m}}$ and $\boldsymbol{\alpha} \to \hat{\boldsymbol{\alpha}}(\mathbf{u})$. Then as a consequence of (3.8), $G_{\boldsymbol{\alpha}}(\mu)$ may vary by (arbitrarily) many orders of magnitude over the interval $[-1, 1]$, in an effort to approximate the ansatz for $\mathbf{u}$, which is nearly a delta-function. In such cases, the limits of finite precision arithmetic mean that many of the terms in (3.14) are effectively zero, making it difficult for $H_{\mathcal{Q}}(\boldsymbol{\alpha})$ to build rank.

In numerical experimentation, we encountered the (non-normalized) moments

$$[u_0, u_1]^T = [1.19788813813286, -1.15179519716325]^T \times 10^{-5} \tag{3.15}$$

which occurred as particles entered the vacuum surrounding an initial impulse. Here $|u_1|/u_0 \approx 0.962$. Figure 3.1(b) shows contours of the objective function for this problem and the effect of the quadrature approximation on the Newton direction. The minimizer of the true objective function, which is marked with a star in the upper right of the figure, is $\hat{\boldsymbol{\alpha}}(\mathbf{u}) \approx [-34.1, -26.0]^T$. A particular iterate $\boldsymbol{\alpha}_k$ is marked with a dot in the lower right of the figure along with the approximate Newton direction computed with an eight-point Gauss-Legendre quadrature. For this particular quadrature $\min_i\{\mu_i\} \approx -0.9603$, and thus according to (3.13), (3.12) is not solvable, i.e., $f_{\mathcal{Q}}$ does not have a minimizer. Not surprisingly, the figure shows that the Newton direction points in the wrong direction. However, as seen in Figure 3.1(b), increasing $n_{\mathcal{Q}}$ by only one suffices to orient the approximation Newton direction correctly. For the nine-point Gauss-Legendre quadrature $\min_i\{\mu_i\} \approx -0.9682$ so that (3.12) is then solvable.

Example 2: A Higher-Order Model. Numerical difficulties also arise in higher-order models. Consider the following M_{15} example with the (normalized) moments

and multipliers

$$\begin{aligned}\mathbf{u} = [\ & 1.0, && 0.837872568, && 0.572819692, && 0.294071376,\\ & 0.079519254, && -0.034894762, && -0.060428124, && -0.037077987,\\ & -0.006145576, && 0.009337451, && 0.007920869, && 0.000075451,\\ & -0.004350212, && -0.002832808, && 0.001074657, && 0.003022835]^T\end{aligned} \tag{3.16a}$$

$$\begin{aligned}\boldsymbol{\alpha} = [\ & -196.5928920, && 230.2769882, && 139.8256880, && -201.8699700,\\ & -183.7928851, && 351.6791766, && -10.2198926, && -278.4913086,\\ & 58.7675482, && 304.0819552, && -258.7624346, && -112.8547686,\\ & 341.8135033, && -269.7545794, && 104.3082746, && -17.0933354]^T\end{aligned} \tag{3.16b}$$

which we encountered in the course of solving the two-beam instability problem discussed in Section 5.3. (Note that $\boldsymbol{\alpha}$ is not $\hat{\boldsymbol{\alpha}}(\mathbf{u})$; rather it is an iterate of the optimization algorithm when attempting to find $\hat{\boldsymbol{\alpha}}(\mathbf{u})$.)

To get a sense of how close a moment is to $\partial\mathcal{R}_{\mathbf{m}}$, one may calculate the minimum eigenvalues of B^+ and B^- from (2.12) in Theorem 2.2. (Recall, one first has to map to the monomial moments to compute these matrices.) For $\mathbf{u}$ in (3.16), the minimum eigenvalues of B^+ and B^- are approximately 2.3×10^{-10} and 2.1×10^{-10}, respectively. For the moment $\hat{\mathbf{v}}(\boldsymbol{\alpha})$, the minimum eigenvalues are approximately 9.3×10^{-10} and 1.3×10^{-10}, respectively. As a reference, the minimum eigenvalues for the normalized "isotropic" moment $[1,0,\ldots,0]^T$ are approximately 1.3×10^{-5}.

From Figure 3.2(a), it is clear that all the structure in the polynomial $p := \boldsymbol{\alpha}^T\mathbf{m}$ is on the left-hand side of the interval. However, because the pointwise values of p are large and negative there, this structure is essentially destroyed when the exponential is applied (Figure 3.2(b)). Even though the function $G_{\boldsymbol{\alpha}}$ appears relatively benign—nothing close to the delta functions which generate the moments on $\partial\mathcal{R}_{\mathbf{m}}$—the condition number of the numerical Hessian (cf. (3.14)) is quite large. Using a very fine 800-point Gauss-Legendre quadrature on each of the subintervals $[-1,0]$ and $[0,1]$ to compute $H_{\mathcal{Q}}$, we find that $\lambda_{\min}(H_{\mathcal{Q}})\approx 4.98\times10^{-12}$ and $\lambda_{\max}(H_{\mathcal{Q}})\approx 2.21$ so that the condition number of $H_{\mathcal{Q}}$ is approximately 4.44×10^{11}.

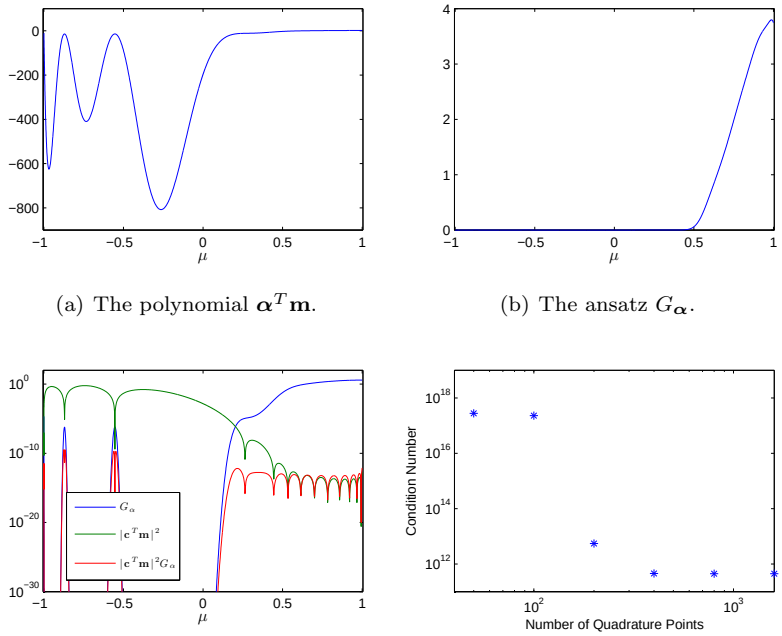
Associated with $\lambda_{\min}(H_{\mathcal{Q}})$ is the unit-length eigenvector $\mathbf{c}_{\mathcal{Q}}$, so $H_{\mathcal{Q}}\mathbf{c}_{\mathcal{Q}} = \lambda_{\min}(H_{\mathcal{Q}})\mathbf{c}_{\mathcal{Q}}$. We evaluate the integrand of the quadratic form

$$U(\mathbf{c},\boldsymbol{\alpha}) := \mathbf{c}^T H(\boldsymbol{\alpha})\mathbf{c} \equiv \langle|\mathbf{c}^T\mathbf{m}|^2 G_{\boldsymbol{\alpha}}\rangle \tag{3.17}$$

at $\mathbf{c}=\mathbf{c}_{\mathcal{Q}}$. (Note that $U(\mathbf{c}_{\mathcal{Q}},\boldsymbol{\alpha}) = \lambda_{\min}(H_{\mathcal{Q}}(\boldsymbol{\alpha}))$ when quadrature $\mathcal{Q}$ is used to evaluate the integral.) The results, given in Figure 3.2(c), show a combination of two effects. First, on the right-hand side of the interval, the polynomial $|\mathbf{c}^T\mathbf{m}|^2$ is very small, but due to the orthogonality relation $\langle m_k m_l\rangle = 2\delta_{kl}/(2k+1)$, this cannot hold everywhere on the interval; indeed, on the left-hand side $|\mathbf{c}^T\mathbf{m}|^2$ becomes $O(1)$. However, on the left-hand side, $G_{\boldsymbol{\alpha}}$ is so small that any contribution to the integral in (3.17) is strongly damped.

Over the entire interval, the most significant contribution to the integral comes from the three peaks in $G_{\boldsymbol{\alpha}}$ on the left-hand hand side. (One of these is at the boundary $\mu=-1.0$.) It is interesting to note that the value of $|\mathbf{c}^T\mathbf{m}|^2$ dips significantly at these peaks so that the product $|\mathbf{c}^T\mathbf{m}|^2 G_{\boldsymbol{\alpha}}$ is $O(10^{-10})$. When $\mathcal{Q}$ is coarsened, the number of quadrature points contained in the support of these peaks decreases, eventually causing the integral (3.17) to decrease and the condition number of $H_{\mathcal{Q}}$ to increase. This effect is displayed in Figure 3.2(d), where we plot the condition number versus the number of quadrature points. In each case, the points are evenly

divided into two Gauss-Legendre quadrature sets on the right and left sides. This result illustrates the need for a highly accurate quadrature set when $\mathbf{u}$ is close to $\partial\mathcal{R}_{\mathbf{m}}$.



(a) The polynomial $\boldsymbol{\alpha}^T\mathbf{m}$.

(b) The ansatz $G_{\boldsymbol{\alpha}}$.

(c) The integrand of the quadratic form.

(d) The condition number for $H_{\mathcal{Q}}(\boldsymbol{\alpha})$.

FIG. 3.2. *Examining the multipliers given in* (3.16).

4. Implementation of the Optimization Algorithm. In this section, we discuss implementation of the optimization algorithm with an adaptive quadrature. We then present a regularization technique to handle moments near the boundary of the realizable set that addresses the difficulties discussed in Section 3.

4.1. Adaptive Quadrature. Most integrals encountered during the optimization process are accurately approximated with just a few Gauss-Legendre quadrature points, but some require many more. Thus a sensible solution is a quadrature adapted to each ansatz $G_{\boldsymbol{\alpha}}$, i.e., a function of $\boldsymbol{\alpha}$. However, as the optimization algorithm iterates through different values of $\boldsymbol{\alpha}$, we must make sure that our changing quadratures do not jeopardize the algorithm's convergence.

4.1.1. Designing the adaptive quadrature. We choose a quadrature separately for $[-1, 0]$ and $[0, 1]$ in order to better capture "one-sided" distributions (as in Example 2 of Section 3.2) and because the integrand in the flux term (2.25) takes different forms for $\mu < 0$ and $\mu > 0$ (see (2.18)). On each half-interval, we use a Gauss-Legendre quadrature and select its order adaptively. To do so, we begin each instance of the optimization process with an initial number of quadrature points $n_{\mathcal{Q}} = N + 5$ on each half interval. Whenever the optimization routine requires an accuracy test of the current quadrature, we actually compute two different quadratures of $G_{\boldsymbol{\alpha}}$ over the half-interval of interest: one with the $n_{\mathcal{Q}}$-point Gauss-Legendre rule and one with the associated $(2n_{\mathcal{Q}} + 1)$-point Gauss-Kronrod rule [15]. The Gauss-Legendre rule is

accepted if the relative difference in the two quadratures is less than the value of a tolerance parameter $\tau_{\mathcal{Q}}$. Otherwise, the value of $n_{\mathcal{Q}}$ is incremented by one and the test is repeated—unless $n_{\mathcal{Q}}$ has already reach a prescribed hard upper bound $n_{\mathcal{Q}}^{\mathrm{MAX}}$. Roughly speaking, this value is chosen to be several orders of magnitude greater than machine precision, while still giving highly accurate approximations for f, $\mathbf{g}$, and H.

4.1.2. Using adaptive quadrature in the optimization algorithm. As mentioned in Section 3.1, we use Newton's method stabilized by a backtracking line search. During the solution of an optimization problem, we never decrease the number of quadrature $n_{\mathcal{Q}}$ points for either half interval, but we do enforce the bound $n_{\mathcal{Q}}^{\mathrm{MAX}}$ on $n_{\mathcal{Q}}$. This ensures that the global convergence properties of Newton's method with backtracking line search (see, e.g., [8]) are not jeopardized: in exact arithmetic (and assuming the stopping criterion is turned off), the constructed sequence will converge to the unique global minimizer of $f_{\mathcal{Q}_{\mathrm{final}}}$, where $\mathcal{Q}_{\mathrm{final}}$ is the final quadrature.

Given an iterate $\boldsymbol{\alpha}_k$, we compute quadrature weights and nodes according to the criteria outlined in Section 4.1.1. The computed quadrature $\mathcal{Q}$ is then used to determine the Newton step and the step size. Upon exiting the line search, we compute a new quadrature $\mathcal{Q}_{\mathrm{new}}$ and check that the line search criterion (3.4) still holds when using $\mathcal{Q}_{\mathrm{new}}$ to evaluate all the requisite integrals. If it does, we update the iterate according to (3.3), using $\mathcal{Q}_{\mathrm{new}}$ to compute the Newton step, and perform the linesearch for the next iteration. If (3.4) is *not* satisfied with the quadrature $\mathcal{Q}_{\mathrm{new}}$, we continue the line search using $\mathcal{Q}_{\mathrm{new}}$ until (3.4) is satisfied again, and so on until an integer i is arrived at for which (3.4) holds.

We have observed cases where, after switching to a finer quadrature during the line search, the previous estimate of the value of the objective function was evidently very poor. This happened, for example, when new quadrature nodes revealed structure in the ansatz that was not visible to the previous quadrature nodes. If the two estimates of the initial value of the objective function differ greatly, the line search may backtrack all the way back to the previous iterate (that is, i in (3.4) is large enough that β^i underflows). In this case, we recompute the Newton direction using the new quadrature.

Further complicating matters, it may happen that with the new quadrature, the Hessian is so poorly conditioned that the search direction cannot be computed with sufficient accuracy. This suggests that the Newton iteration has strayed off course because of a previous quadrature that was too coarse. Accordingly, given a prescribed large value $\kappa_{\max} > 0$, if $\mathrm{cond}\,(H_{\mathcal{Q}_{\mathrm{new}}}(\boldsymbol{\alpha}_k)) > \kappa_{\max}$ at some iteration k, we successively consider previous iterates $\boldsymbol{\alpha}_{k-1}, \boldsymbol{\alpha}_{k-2}, \ldots$ to determine the largest positive non-negative integer $i < k$ such that $\mathrm{cond}\,(H_{\mathcal{Q}_{\mathrm{new}}}(\boldsymbol{\alpha}_i)) \leq \kappa_{\max}$. We then set $\boldsymbol{\alpha}_{k+1} := \boldsymbol{\alpha}_i$ and continue the optimization, still with the new quadrature.

4.1.3. Coupling to the Flux Terms in the PDE Solver. As mentioned in Section 4.1.2, the number of quadrature points $n_{\mathcal{Q}}$ used to evaluate integrals is never decreased during the execution of the optimization algorithm. Consequently, upon exiting the optimization process, the final quadrature may be much finer than needed to accurately approximate the flux $\mathbf{f}_{j+1/2}$ in (2.25). Therefore, upon exiting the optimization algorithm, we generate (according to the criteria in Section 4.1.1) new quadrature sets for $\bar{\boldsymbol{\alpha}}(\mathbf{u}_j)$; that is, a new quadrature is generated "from scratch," starting with $n_{\mathcal{Q}} = N + 5$ points on each half interval so that the tolerance $\tau_{\mathcal{Q}}$ is satisfied and also $\mathrm{cond}(H_{\mathcal{Q}}(\bar{\boldsymbol{\alpha}})) \leq \kappa_{\max}$. For this quadrature, let $\mathcal{Q}_j^+$ and $\mathcal{Q}_j^-$ denote positive and negative values of μ respectively in each cell I_j. Then the flux is computed

by

$$\mathbf{f}_{j+1/2} \equiv \left\langle \mu \mathbf{m} \bar{G}_{j+1/2} \right\rangle \simeq \sum_{\mu \in \mathcal{Q}_j^+} \mu \mathbf{m}(\mu) \bar{G}_{j+1/2,l} + \sum_{\mu \in \mathcal{Q}_{j+1}^-} \mu \mathbf{m}(\mu) \bar{G}_{j+1/2,r} \tag{4.1}$$

The evaluation of the edge values requires the computation of slopes $\bar{s}_j$ for each cell I_j, as given in (2.20) (with hats replaced by bars). Computation of the slopes is implemented by passing the values $\bar{\boldsymbol{\alpha}}_{j\pm1}$ (the computed multipliers from cells $I_{j\pm1}$) to the cell I_j, evaluating $\bar{G}_{j\pm1}$ for every $\mu \in \mathcal{Q}_j := \mathcal{Q}_j^- \cup \mathcal{Q}_j^+$, and then using these ansatz values to compute $\bar{s}_j$ for every $\mu \in \mathcal{Q}_j$. There are two reasons for this approach: First, while the quadrature sets may vary from cell to cell, evaluation of the slopes must occur on a common set of values for μ. Second, passing multipliers rather than kinetic data significantly reduces the communication between cells.

4.1.4. Computational Limits. With an adaptive quadrature and its careful implementation in the optimization algorithm, we have been able to solve challenging problems on fine spatial meshes. (See Section 5.) However, for any given quadrature set $\mathcal{Q}$, there are almost always moments for which the approximate gradient $\mathbf{g}_\mathcal{Q}$ remains bounded away from zero, i.e., there is a constant C which depends on the quadrature, such that $||\mathbf{g}_\mathcal{Q}(\boldsymbol{\alpha})|| \geq C > 0$ for all $\boldsymbol{\alpha} \in \mathbb{R}^{N+1}$ (cf. (3.13) for the case of the M_1 model). This is a consequence of the fact that the moments on the boundary of $\mathcal{R}_\mathbf{m}$ can only be generated by a unique atomic distribution—that is, a linear combination of delta functions [14].

If $\{\mathbf{u}^{(l)}\}$ is a sequence of moments such that $\mathbf{u}^{(l)} \to \mathbf{u}_{\text{bndry}} \in \partial\mathcal{R}_\mathbf{m}$, then as l increases, $G_{\hat{\boldsymbol{\alpha}}(\mathbf{u}^{(l)})}$ more closely approximates the unique atomic distribution which generates $\mathbf{u}_{\text{bndry}}$. In particular, the effective support of $G_{\hat{\boldsymbol{\alpha}}(\mathbf{u}^{(l)})}$ (i.e., the set of μ on which $G_{\hat{\boldsymbol{\alpha}}(\mathbf{u}^{(l)})}$ is large enough to affect the numerical evaluation of the moment integral) begins to shrink. If $\mathcal{Q}$ is not co-located with the limiting atomic distribution, then for l sufficiently large, the effective support of $G_{\hat{\boldsymbol{\alpha}}(\mathbf{u}^{(l)})}$ will eventually fail to contain any points of $\mathcal{Q}$. For the M_1 case, where the boundary moments are generated by a single delta function at $\mu = \pm1$, a quadrature set such as Gauss-Lobatto can be chosen to contain these endpoints, and for such a quadrature, the approximate M_1 problem is always solvable (cf. (3.13)). However, when $N > 1$, the moments on the boundary may be generated by delta functions located anywhere on the interval $[-1, 1]$, making it is impossible to co-locate a single quadrature set with all these possibilities. Hence, near the boundary, the quadrature must depend on $\mathbf{u}$ and, in addition, it must be locally adaptive so that it can track the support of $G_{\boldsymbol{\alpha}_k}$ during the course of the optimization process.[3] In doing so, the quadrature routine must carefully select and deselect points in order to maintain a practical bound on $n_\mathcal{Q}$ and still guarantee convergence of the optimization algorithm. We have yet to design such a quadrature.

Further, the true Hessian $H(\hat{\boldsymbol{\alpha}}(\mathbf{u}))$ at the dual solution approaches singularity as $\mathbf{u}$ approaches the realizable boundary, and thus an accurate approximation $H_\mathcal{Q}(\boldsymbol{\alpha})$ will also be poorly conditioned for $\boldsymbol{\alpha}$ sufficiently close to $\hat{\boldsymbol{\alpha}}(\mathbf{u})$. Consequently, even if the norm of the gradient is very small, the norm of the Newton step $\mathbf{d}_\mathcal{Q} := -H_\mathcal{Q}^{-1}\mathbf{g}_\mathcal{Q}$ may be too big to satisfy the second condition in (3.5), which is used to guarantee realizability. Indeed, numerical experiments show that this stopping criteria is more difficult to satisfy than the small gradient condition and that, for higher-order

[3] An entirely different closure approach based on choosing a quadrature set from the moments is the *Quadrature Method of Moments* (QMOM). See [52] and references therein.

$n_{\mathcal{Q}}^{\text{MAX}}$ \ $\mathbf{u}^{(l)}$	M_1	M_2		
	$\begin{pmatrix} 1 \\ 1-2^{-l} \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1/\sqrt{3}-2^{-l} \\ 0 \end{pmatrix}$	$\begin{pmatrix} 1 \\ 1-2^{-l} \\ 1-2^{-l} \end{pmatrix}$	$\begin{pmatrix} 1 \\ 0 \\ 1-2^{-l} \end{pmatrix}$
50	11	19	10	10
100	13	15	12	12
200	15	17	14	14
400	17	19	16	16
600	18	21	17	17
800	19	23	18	18
1000	20	24	18	16

TABLE 4.1

The maximum value of l for which we were able to close $\mathbf{u}^{(l)}$ using no more than the specified number of quadrature points $n_{\mathcal{Q}}^{\text{MAX}}$ over either $\mu \in [-1, 0]$ or $\mu \in [0, 1]$ to achieve the tolerance $\tau_{\mathcal{Q}} = 0.5 \times 10^{-12}$, as explained in Section 4.1.1. For reference, $2^{-10} \approx 1 \times 10^{-3}$, $2^{-15} \approx 3 \times 10^{-5}$, and $2^{-20} \approx 1 \times 10^{-6}$.

models, the singularity of the Hessian usually causes the optimization to fail before inaccuracies in the quadrature come into play.

Example 1: The M_1 and M_2 Models. To get a sense of how close the moments $\mathbf{u}$ can get to $\partial\mathcal{R}_{\mathbf{m}}$ before our optimization algorithm fails to produce a point that satisfies the stopping criterion, we experimented with several static M_1 and M_2 problems (where the realizable sets are easy to calculate explicitly). In the case of M_1, $\mathcal{R}_{\mathbf{m}} = \{(u_0, u_1) : |u_1| < u_0\}$ while for M_2, $\mathcal{R}_{\mathbf{m}} = \{(u_0, u_1, u_2) \in \mathbb{R}^3 : 0 < u_0\,,\, 3u_1^2 < 2u_0u_2 + u_0^2\,,\, u_2 < u_0\}$ (see Figure 4.1).

We consider four different sequences of moments $\{\mathbf{u}^{(l)}\}$ that approach the boundary of realizability as $l \to \infty$. For each sequence, we find the largest value of l for which the optimization algorithm can close the moment $\mathbf{u}^{(l)}$ without exceeding a given value of $n_{\mathcal{Q}}^{\text{MAX}}$ when attempting to satisfy the quadrature tolerance $\tau_{\mathcal{Q}} = 0.5 \times 10^{-12}$ (as explained in Section 4.1.1). Table 4.1 shows the results for these experiments. The results show that even with 1000 quadrature points, it is difficult to get within 10^{-6} units of the boundary with this particular value of $\tau_{\mathcal{Q}}$. Further, we can see that if we limit our algorithm to using at most 200 quadrature points on each side ($\mu \in [-1, 0]$ and $\mu \in [0, 1]$), we can still get about 5×10^{-5} units away from the boundary.

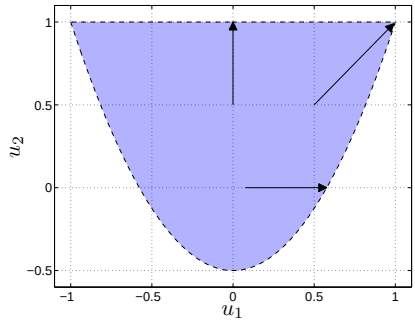


FIG. 4.1. *The set of normalized realizable moments $\mathcal{R}_{\mathbf{m}}|_{u_0=1}$ in M_2 and the paths we take to the boundary in Table 4.1.*

Example 2: A Higher-Order Model. We revisit the M_{15} model with a small perturbation of the multiplier from the example in (3.16):[4]

$$\begin{aligned}\boldsymbol{\alpha} \;=\; [&\text{-1.9930449}, \quad 2.3573737, \quad 1.3695484, \quad \text{-2.0424320},\\ &\text{-1.7877058}, \quad 3.4897586, \quad \text{-0.1169555}, \quad \text{-2.7645420},\\ &0.6175618, \quad 2.9482952, \quad \text{-2.4912231}, \quad \text{-1.1399532},\\ &3.3044482, \quad \text{-2.5119482}, \quad 0.8977758, \quad \text{-0.1487460}]^T \times 10^2\end{aligned} \tag{4.2}$$

For this value of $\boldsymbol{\alpha}$, the minimum eigenvalues of B^+ and B^- associated with the moment $\hat{\mathbf{v}}(\boldsymbol{\alpha})$ are approximately 1.3×10^{-14} and 3.3×10^{-15}, respectively. By this measure, the $\hat{\mathbf{v}}(\boldsymbol{\alpha})$ is significantly closer to the $\partial\mathcal{R}_{\mathbf{m}}$ than is the moment generated by the multiplier in (3.16).

Figure 4.2 contains the same results as Figure 3.2, except that multipliers in (3.16) are replaced by those in (4.2). It is interesting to note that the profile of $G_{\boldsymbol{\alpha}}$ in Figure 4.2(b) is beginning to look like an atomic distribution, although it is still fairly smooth.

We again compute $H_{\mathcal{Q}}$ using a very fine 800-point Gauss-Legendre quadrature on each half interval, which is accurate enough to resolve the structure in $G_{\boldsymbol{\alpha}}$, and find that $\lambda_{\min}(H_{\mathcal{Q}}) \simeq -1.30 \times 10^{-16}$ and $\lambda_{\max}(H_{\mathcal{Q}}) \simeq 2.89$. (The fact that the computed value of $\lambda_{\min}(H_{\mathcal{Q}})$ is negative is a result of roundoff error from double precision arithmetic.) Thus the condition number of $H_{\mathcal{Q}}$ is at least $O(10^{16})$. It may in fact be larger, but no further conclusions can be drawn without increasing the working precision. Moreover, the large condition number means that the relative error in the computed Newton step may be $O(1)$ or greater. Figure 4.2(d) shows that refining the quadrature over two orders of magnitude has little effect on the calculated condition number. (This is not the case for the multiplier in (3.16), where increasing the number of quadrature points from $O(10)$ to $O(10^3)$ eventually improves the condition number (see Figure 3.2(d)). Because $G_{\boldsymbol{\alpha}}$ is still relatively smooth, we again conclude that the limitations in the optimization algorithm are not due to the quadrature in the case, but rather to the conditioning of the Hessian.

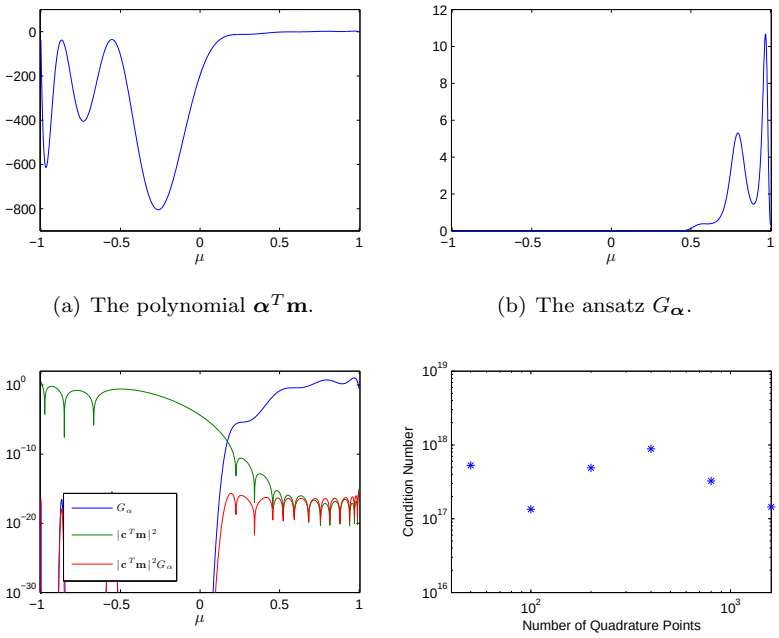
4.2. Regularization at the Boundary of Realizability. Given the limitations in the optimization algorithm discussed in Section 4.1, we propose here a regularization procedure to modify moments $\mathbf{u}$ for which $\hat{\boldsymbol{\alpha}}(\mathbf{u})$ is "too difficult" to compute. Based on the experiments in Section 4.1.4, the closure problem for $\mathbf{u}$ is deemed too difficult to solve when the number of quadrature points needed during the optimization process exceeds a soft upper bound $n_{\mathcal{Q}}^{\max}$ (typically chosen to be a few hundred). In such cases we replace the normalized moment $\mathbf{u}$ by a regularized moment

$$\mathbf{v}(r) := (1-r)\mathbf{u} + rQ\mathbf{u}, \tag{4.3}$$

where Q is defined in (2.14) and $r \in (0,1)$ is a regularization factor that should be chosen as small as possible (see below). The regularization (4.3) exploits the convexity of $\mathcal{R}_{\mathbf{m}}$ (Theorem 2.3): It produces a new moment $\mathbf{v}(r) \in \mathcal{R}_{\mathbf{m}}$ that is closer to the "isotropic" moment $Q\mathbf{u}$ with multiplier $\hat{\boldsymbol{\alpha}}(Q\mathbf{u}) = [\log(u_0/2), 0, \ldots, 0]^T$ *without changing the number of particles* u_0. Note that $G_{\hat{\boldsymbol{\alpha}}(Q\mathbf{u})}(\mu) \equiv u_0/2$ and the condition number of the Hessian $H(\hat{\boldsymbol{\alpha}}(Q\mathbf{u}))$ is $O(N)$.

In order to maintain realizability in the PDE algorithm (i.e, to apply Theorem 2.5), we must also replace each subvector of $\mathbf{u}^{(m)}$, $m \in \{0,1\}$, by its regularized version

[4] The relative difference between the multipliers in (3.16) and (4.2) is roughly 5.3% when measured in the ℓ_∞ norm



(a) The polynomial $\boldsymbol{\alpha}^T\mathbf{m}$.

(b) The ansatz $G_{\boldsymbol{\alpha}}$.

(c) The integrand of the quadratic form.

(d) The condition number for $H_{\mathcal{Q}}(\boldsymbol{\alpha})$.

FIG. 4.2. *Examining the multipliers given in* (4.2).

at each Runge-Kutta stage (cf. (2.23)) of the kinetic scheme for (2.11). The result is a modified set of moment equations. In a single forward Euler step this modification takes the form

$$\frac{\mathbf{u}_j^{n+1}-\mathbf{u}_j^n}{\Delta t}+\left[\mathbf{f}((1-r)\mathbf{u}_j^n+rQ\mathbf{u})\right]_x+\tilde{\sigma}_{\mathrm{t}}\mathbf{u}_j^n=\left[\tilde{\sigma}_{\mathrm{s}}^{(0)}Q+\tilde{\sigma}_{\mathrm{s}}^{(1)}(I-Q)\right]\mathbf{u}_j^n \tag{4.4}$$

where, for the purpose of exposition, the x variable is held continuous and

$$\tilde{\sigma}_{\mathrm{t}}:=\sigma_{\mathrm{t}}+\frac{r}{\Delta t}\,,\qquad \tilde{\sigma}_{\mathrm{s}}^{(0)}:=\sigma_{\mathrm{s}}+\frac{r}{\Delta t}\,,\qquad \tilde{\sigma}_{\mathrm{s}}^{(1)}:=r\sigma_{\mathrm{t}}\,. \tag{4.5}$$

Thus the effect of the regularization is (i) to introduce a modification of the flux; (ii) to increase isotropic scattering from σ_{s} to $\tilde{\sigma}_{\mathrm{s}}^{(0)}$; and (iii) to introduce non-isotropic scattering via $\tilde{\sigma}_{\mathrm{s}}^{(1)}$. However, since $\tilde{\sigma}_{\mathrm{t}}-\tilde{\sigma}_{\mathrm{s}}^{(0)}=\sigma_{\mathrm{t}}-\sigma_{\mathrm{s}}=\sigma_{\mathrm{a}}$, the number of particles absorbed during the time step is not affected by the regularization. Note that all of these modifications are spatially and temporally dependent since r varies between spatial cells and time steps.

The use of "numerical" scattering to regularize moment problems has been reported before in [43,50]. The regularization acts as barrier that prevents the entropy-ansatz from getting too close to the delta-type distributions which characterize the boundary $\partial\mathcal{R}_{\mathbf{m}}$. In this sense, it can be viewed as numerical dissipation in μ-space. More detailed analysis of the modified moment system that is generated by the regularization will be the subject of future work.

To select a value for r in the regularization, we first choose an initial small value $r_0>0$ and then generate a sequence $\{r_\ell\}$ with the recursion formula $r_{\ell+1}:=$

$\min(2r_\ell, r_{\max})$. We choose the smallest value of ℓ for which the optimization stopping criterion can be satisfied without using more than $n_{\mathcal{Q}}^{\max}$ quadrature points. If we reach the maximum value $r_{\max}$, we solve the dual problem with $r = r_{\max}$ no matter how many quadrature points it takes (up to $n_{\mathcal{Q}}^{\mathrm{MAX}}$).

The selection of the parameter $r_{\max}$ must strike a balance between the number of quadrature points and the error introduced by the regularization. To ensure a completely robust scheme, the value of $r_{\max}$ should be set to 1. Indeed if $r_{\max}$ is smaller, then there is no simple way to ensure that the number of quadrature points or the condition number of the Hessian (exact or approximate) remains bounded. However, we have found that for $n_{\mathcal{Q}}^{\max} = 200$, a value of $r_{\max}$ between 10^{-4} and 10^{-5} gives satisfactory results in almost all cases. In simulations presented in the next section, the maximum number of quadrature points used is 255, and the number 200 is exceeded in less than 1.2% of the optimization problems solved.

REMARK 2. *In some cases, the regularization can be used to solve the dual problem for a* non-regularized *moment* $\mathbf{u}$ *when a direct application of Newton's method fails. This is done as follows. Define a decreasing sequence* $r_\ell \searrow 0$ *and successively solve the dual problem to find* $\hat{\boldsymbol{\alpha}}(\mathbf{v}(r_\ell))$*, using* $\hat{\boldsymbol{\alpha}}(\mathbf{v}(r_{\ell-1}))$ *as an initial condition. This defines a new path in* $\boldsymbol{\alpha}$*-space to the minimizer* $\hat{\boldsymbol{\alpha}}(\mathbf{u})$ *of the original problem. In practice, we were indeed able to solve some* M_{15} *problems for which Newton's method either failed or needed thousands of quadrature points. However, the fraction of moments for which this method worked when the Newton method did not was relatively small, and hence we did not include it in our implementation.*

5. Results. We first test the convergence of the kinetic scheme from Section 2.4 in space and time, and then, closely following [27], we test our algorithm on two benchmark problems for which the closure is challenging to compute. We consider the steady-state solution for the two-beam instability problem and transient solutions for a plane source problem. In all cases, we use the following parameter values.

τ	=	10^{-8},	upper bound for $\|\mathbf{g}(\bar{\boldsymbol{\alpha}})\|$
ε_γ	=	0.01,	upper bound on $\gamma_{\max} - 1$ to maintain realizability
$\tau_{\mathcal{Q}}$	=	0.5×10^{-12},	quadrature tolerance
r_0	=	10^{-8},	initial regularization factor
$r_{\max}$	=	10^{-4},	maximum regularization factor
$n_{\mathcal{Q}}^{\max}$	=	200,	'soft' maximum $n_{\mathcal{Q}}$ for regularization
$n_{\mathcal{Q}}^{\mathrm{MAX}}$	=	1000,	'hard' maximum $n_{\mathcal{Q}}$ for regularization
$\kappa_{\max}$	=	$1/\mathtt{eps}$,	the largest allowable value of $\mathrm{cond}\,(H_{\mathcal{Q}}(\boldsymbol{\alpha}))$
β	=	1/2,	line search stepsize decrease parameter
ξ	=	0.001,	line search sufficient decrease parameter
θ	=	2.0,	slope limiting parameter

Here $\mathtt{eps} = 2^{-52}$ is twice the machine precision in double precision arithmetic. Note also that the value $n_{\mathcal{Q}}^{\mathrm{MAX}}$ was never attained during any of the simulations.

5.1. Convergence Test. To test the convergence in space and time of the kinetic scheme from Section 2.4, we devised a simple test problem with an initial condition that is sinusoidal in space and isotropic in angle:

$$F_0(x, \mu) = \frac{1}{2}\left(2 + \cos(4\pi x)\right), \qquad x \in [0, 1]\,. \tag{5.1}$$

We use periodic boundary conditions as in (2.21).

We run the tests on the M_1 and M_{15} models, and since we are interested primarily in the closure of the fluxes, we set $\sigma_t = 0$. We reach a final time $t^{N_t} = \sqrt{3}/5$ with a fixed time step $\Delta t = 0.45\Delta x$[5], which satisfies (2.27). Since an analytic solution is unavailable for either model, we perform a high-resolution simulation with $N_x = 2560$ spatial cells to compute a reference solution $\mathbf{u}_{\text{ref}}$.

We define L_1 and L_∞ errors at time t_{final} for each moment using the piecewise-linear spatial reconstruction

$$\mathbf{u}_{\Delta x}(x) = \mathbf{u}_j^{N_t} + (x - x_j)\frac{\mathbf{u}_{j+1}^{N_t} - \mathbf{u}_{j-1}^{N_t}}{2\Delta x} \qquad x \in I_j, \qquad j \in \{1, \ldots, N_x\}. \tag{5.2}$$

The L_1 and L_∞ errors for simulations of cell-size Δx are given by

$$\mathbf{e}_{\Delta x}^1 := \int_0^1 |\mathbf{u}_{\text{ref}} - \mathbf{u}_{\Delta x}| dx \quad \text{and} \quad \mathbf{e}_{\Delta x}^\infty := \max_{x\in[0,1]} |\mathbf{u}_{\text{ref}} - \mathbf{u}_{\Delta x}|\,, \tag{5.3}$$

respectively, where the absolute value is defined component-wise. Since the integrand in (5.3) is just the difference of two piecewise-linear functions, we compute $\mathbf{e}_{\Delta x}^1$ exactly. For $q \in \{1, \infty\}$, the order of convergence $\boldsymbol{\nu}$ between two successive meshes of size Δx_1 and Δx_2 is defined by the equality

$$\mathbf{e}_{\Delta x_2}^q / \mathbf{e}_{\Delta x_1}^q = (\Delta x_1/\Delta x_2)^{\boldsymbol{\nu}}\,, \tag{5.4}$$

where all operations are performed component-wise.

The results are shown in Tables 5.1 and 5.2. We only report errors in the zeroth- and first-order moments. For both M_1 and M_{15} models, these moments exhibit second-order convergence.

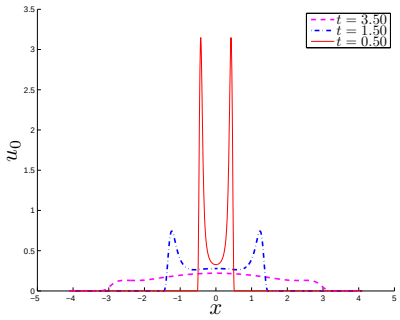
TABLE 5.1
M_1 convergence study with $\theta = 2$.

Number of cells	u_0				u_1			
	L_1 error	ν	L_∞ error	ν	L_1 error	ν	L_∞ error	ν
10	3.34e-01		5.94e-01		9.74e-02		2.07e-01	
20	6.41e-02	2.38	1.28e-01	2.21	3.37e-02	1.52	8.02e-02	1.36
40	1.20e-02	2.41	2.87e-02	2.15	8.40e-03	2.00	2.12e-02	1.91
80	3.45e-03	1.79	1.63e-02	0.81	2.35e-03	1.83	5.70e-03	1.89
160	9.08e-04	1.92	5.77e-03	1.49	5.95e-04	1.98	2.11e-03	1.42
320	2.02e-04	2.16	1.32e-03	2.12	1.29e-04	2.20	4.90e-04	2.11
640	4.70e-05	2.10	2.55e-04	2.37	2.74e-05	2.23	9.00e-05	2.44

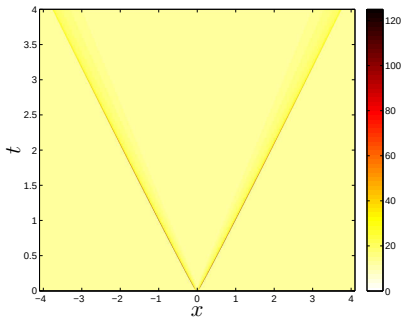
TABLE 5.2
M_{15} convergence study with $\theta = 2$.

Number of cells	u_0				u_1			
	L_1 error	ν	L_∞ error	ν	L_1 error	ν	L_∞ error	ν
10	1.58e-01		2.56e-01		2.48e-02		4.43e-02	
20	3.43e-02	2.20	5.08e-02	2.33	1.45e-03	4.09	3.36e-03	3.72
40	4.50e-03	2.92	8.25e-03	2.62	3.01e-03	-1.04	5.62e-03	-0.74
80	6.84e-04	2.71	1.96e-03	2.07	1.35e-03	1.15	2.15e-03	1.38
160	1.48e-04	2.21	4.53e-04	2.11	3.81e-04	1.83	6.71e-04	1.68
320	3.32e-05	2.15	1.21e-04	1.89	9.81e-05	1.95	1.98e-04	1.75
640	6.23e-06	2.41	2.49e-05	2.28	2.38e-05	2.03	4.29e-05	2.20

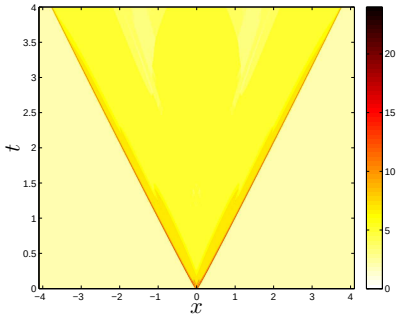
[5]Except for the final time step which is chosen to meet the final time exactly.



(a) Snapshots of the solution, $u_0(x,t)$ at $t = 0.5, 1.5$ and 3.5.



(b) The average number of quadrature points used (over both Runge-Kutta stages) for each point in space and time.



(c) The total number of iterations (over both Runge-Kutta stages) needed to solve the optimization problem at each point in space and time.

FIG. 5.1. *A simulation of the M_1 model of the plane-source problem with 1000 spatial cells. We did not need to regularize the moments anywhere in the course of solving the problem.*

5.2. Plane Source. In this problem, we model an infinite domain with an isotropic impulse in μ as initial condition. We choose a purely scattering medium, $\sigma_t = \sigma_s = 1$, and discretize the initial condition

$$F_0(x,\mu) = 0.5\delta(x) + F_{\text{floor}}\,, \tag{5.5}$$

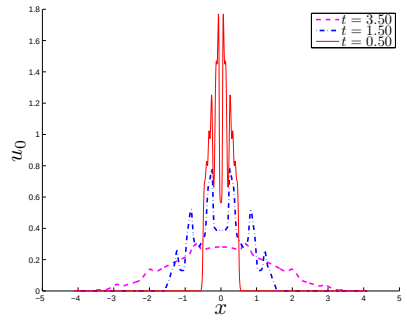
where $F_{\text{floor}} = 0.5 \times 10^{-8}$ is used to keep moments away from the realizable boundary. Although the problem is posed on an infinite domain, a finite domain is required for practical computation and boundary conditions must be specified. As in [27], we approximate the infinite domain by the interval $[-L/2, L/2]$, where $L := 2t_{\text{final}} + 0.2$ is chosen so that the boundary has negligible effects on the solution. At the right and left ends of the boundary, we enforce the boundary conditions

$$F_L(\mu,t) = F_{\text{floor}}\,, \quad \text{and} \quad F_R(\mu,t) = F_{\text{floor}} \tag{5.6}$$

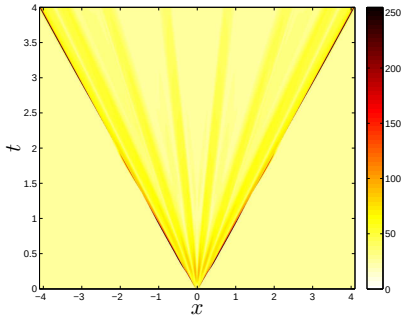
for $t \geq 0$ and, following (2.22), set values of $\mathbf{u}$ in the ghost cells by taking moments of (5.6).

The results illustrate characteristics of the optimization, but do not reveal any new qualitative behavior beyond what was already noted in [27]. The results are presented in Figures 5.1, 5.2, and 5.3 for $N = 1, 7$, and 15 respectively. We chose $t_{\text{final}} = 4$ with $N_x = 1000$ spatial cells and $\Delta t = 0.45\Delta x$.

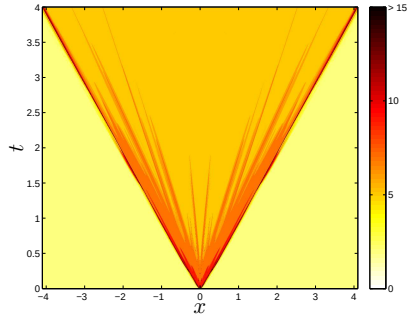
5.3. Two-beam Instability. In this problem, particles constantly stream into the domain from the left at $x_L = -0.5$ and the right at $x_R = 0.5$ into the initially



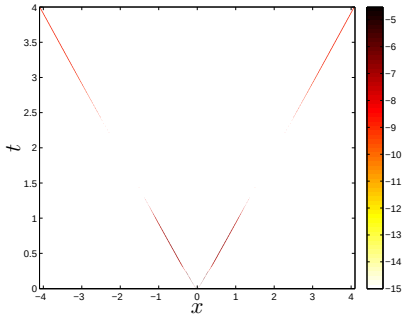
(a) Snapshots of the solution, $u_0(x,t)$ at $t = 0.5, 1.5$ and 3.5.



(b) The average number of quadrature points used (over both Runge-Kutta stages) for each point in space and time.



(c) The total number of iterations (over both Runge-Kutta stages) needed to solve the optimization problem at each point in space and time. The maximum number of iterations needed for one time step was 269 (off scale).



(d) The quantity $\log_{10} \|\mathbf{u} - \mathbf{v}(r)\|$, which measures the difference between the original moments $\mathbf{u}$ and the regularized moments $\mathbf{v}(r)$ closed by the optimization. (The white space indicates where moments were not regularized.)

FIG. 5.2. *A simulation of the M_7 model of the plane-source problem with 1000 spatial cells.*

(almost) vacuous interior. There is no scattering: $\sigma_t = \sigma_a = 2$. The boundary conditions are 'forward-peaked,'

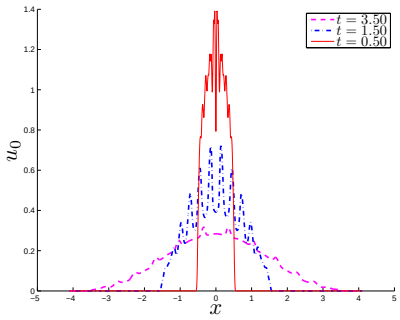
$$F_L(\mu,t) = \max(\exp(-10(\mu-1)^2), F_{\text{floor}}), \quad F_R(\mu,t) = \max(\exp(-10(\mu+1)^2), F_{\text{floor}}), \tag{5.7}$$

with $F_{\text{floor}} := 0.5 \times 10^{-8}$, and the initial condition is $F_0(x,\mu) \equiv F_{\text{floor}}$.

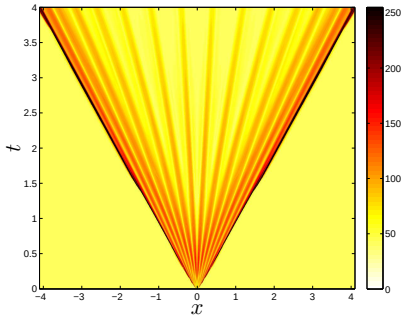
Results for $N = 1, 7$, and 15 are presented in Figures 5.4, 5.5 and 5.6 respectively, again with $N_x = 1000$ and $\Delta t = 0.45\Delta x$. In each case, the most difficult optimization problems occur when particles first enter the interior and when particles originating from right and left boundaries meet in the interior. The lack of scattering in this problem means that for $t < 0.5$, almost all of the particles on the left (resp. right) half of the domain have positive (resp. negative) velocities. As particles starting from the left boundary cross those starting from the right boundary, the support of the ansatz eventually grows to the full interval $[-1, 1]$ and the optimization becomes easier.

6. Conclusions and Discussion. We have considered the entropy-based moment closure models M_N. We presented a kinetic scheme that is formally second-order in space and time and discussed in detail the challenges in solving the associated optimization problem.

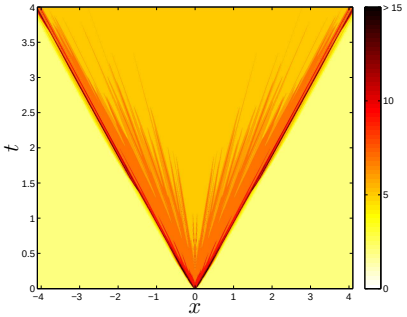
First we gave sufficient conditions for an approximation of the true solution to maintain realizability of the kinetic scheme. We then devised a complete solution to



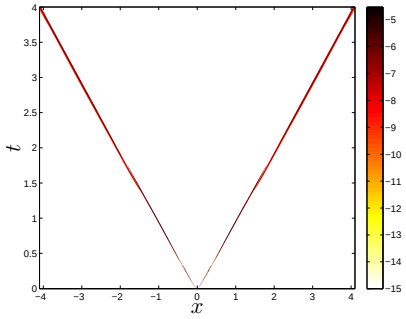
(a) Snapshots of the solution, $u_0(x,t)$ at $t = 0.5, 1.5$ and 3.5.



(b) The average number of quadrature points used (over both Runge-Kutta stages) for each point in space and time.



(c) The total number of iterations (over both Runge-Kutta stages) needed to solve the optimization problem at each point in space and time. The maximum number of iterations needed for one time step was 315 (off scale).



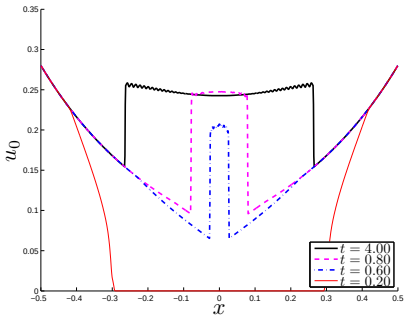
(d) The quantity $\log_{10} \|\mathbf{u} - \mathbf{v}(r)\|$, which measures the difference between the original moments $\mathbf{u}$ and the regularized moments $\mathbf{v}(r)$ closed by the optimization. (The white space indicates where moments were not regularized.)

FIG. 5.3. *A simulation of the M_{15} model of the plane-source problem with 1000 spatial cells.*

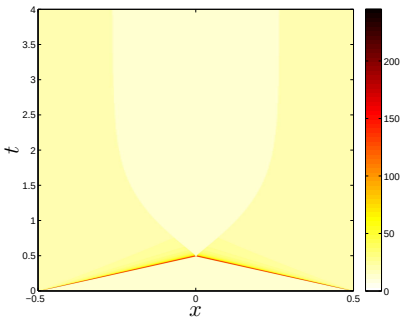
the optimization problem up to the limits of finite-precision arithmetic. The quadrature problem turns out to be a key hurdle to finding sufficiently precise multipliers. With an appropriate adaptive quadrature, we were able to solve a large class of problems.

For moments that are so close to the boundary of realizability that we could not compute their closure, we showed how they can be regularized to nearby moments while preserving the number of particles. For problems which require regularization, only a small perturbation was required to find solvable moments, and the resulting solution did not appear to suffer.

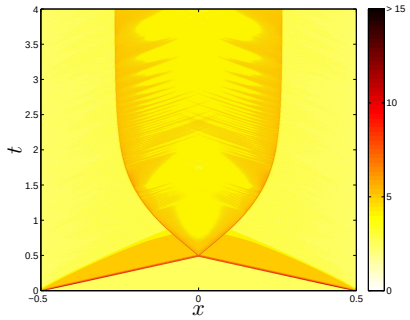
Future work must first take advantage of the parallelizability of the problem, which is one of the primary motivations for our work on the M_N models. The methods presented here also need to be tested in two- and three-dimensional models, where addressing the quadrature problem will likely take a great deal of effort. In addition, more advanced numerical methods are required to efficiently simulate problems for which σ_t is very large. Such methods are needed to maintain realizability, to allow for $O(\Delta x)$ time steps, and to capture the well-known diffusion limit [35]. From an applications point of view, we plan to examine other physically motivated entropies, such as Fermi-Dirac and Bose-Einstein. In addition coupling to hydrodynamic equations will be considered.



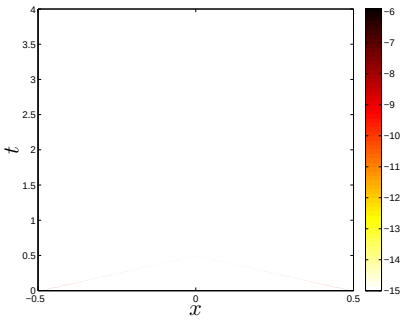
(a) Snapshots of the solution, $u_0(x,t)$ at $t = 0.2, 0.6, 0.8$ and 4.0.



(b) The average number of quadrature points (over both Runge-Kutta stages) used for each point in space and time.



(c) The total number of iterations (over both Runge-Kutta stages) needed to solve the optimization problem at each point in space and time. The maximum number of iterations needed for one time step was 128 (off scale).



(d) The quantity $\log_{10} \|\mathbf{u} - \mathbf{v}(r)\|$, which measures the difference between the original moments $\mathbf{u}$ and the regularized moments $\mathbf{v}(r)$ closed by the optimization. (The white space indicates where moments were not regularized.)

FIG. 5.4. *A simulation of the M_1 model of the two-beam instability with 1000 spatial cells.*

REFERENCES

[1] *International exascale software project roadmap*, 2009.

[2] A.M. ANILE, W. ALLEGRETTO, AND C. RINGHOFER, *Mathematical Problems in Semiconductor Physics*, vol. 1823 of Lecture Notes in Mathematics, Springer-Verlag, Berlin, 2003. Lectures given at the C.I.M.E. Summer School held in Cetraro, Italy on July 15-22, 1998.

[3] A. M. ANILE AND O. MUSCATO, *Improved hydrodynamical model for carrier transport in semiconductors*, Phys. Rev. B, 51 (1995), pp. 16728–16740.

[4] A. M. ANILE AND S. PENNISI, *Thermodynamic derivation of the hydrodynamical model for charge transport in semiconductors*, Phys. Rev. B, 46 (1992), pp. 187–193.

[5] A. M. ANILE AND V. ROMANO, *Hydrodynamical modeling of charge carrier transport in semiconductors*, Meccanica, 35 (2000), pp. 249–296.

[6] L. ARMIJO, *Minimization of functions having lipschitz continuous first partial derivatives*, (1966).

[7] J. M. BORWEIN AND A. S. LEWIS, *Duality relationships for entropy-like minimization problems*, SIAM J. Control Optim., 1 (1991), pp. 191–205.

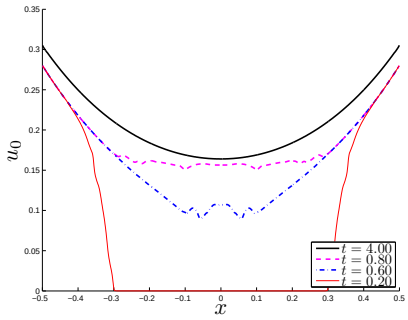
[8] S.P. BOYD AND L.VANDENBERGHE, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[9] THOMAS A. BRUNNER AND JAMES PAUL HOLLOWAY, *One-dimensional Riemann solvers and the maximum entropy closure*, J. Quant Spect. and Radiative Trans, 69 (2001), pp. 543 – 566.

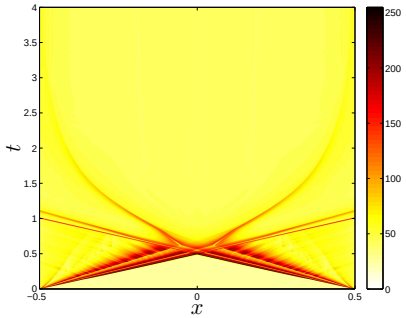
[10] T. A. BRUNNER AND J. P. HOLLOWAY, *Two-dimensional time-dependent Riemann solvers for neutron transport*, J. Comp Phys., 210 (2005), pp. 386–399.

[11] J. CERNOHORSKY AND S. A. BLUDMAN, *Stationary neutrino radiation transport by maximum entropy closure*, tech. report, Lawrence Berkely National Laboratory, 1994.

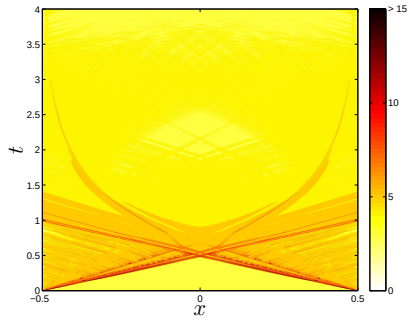
[12] J. CERNOHORSKY, L. J. VAN DEN HORN, AND J. COOPERSTEIN, *Maximum entropy eddington*



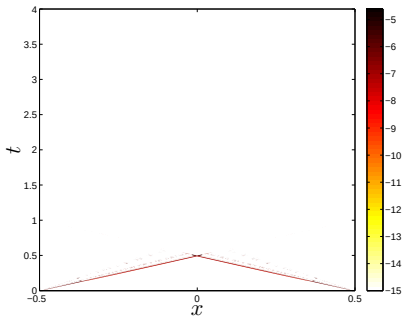
(a) Snapshots of the solution, $u_0(x,t)$ at $t = 0.2, 0.6, 0.8$ and 4.0.



(b) The average number of quadrature points (over both Runge-Kutta stages) used for each point in space and time.



(c) The total number of iterations (over both Runge-Kutta stages) needed to solve the optimization problem at each point in space and time. The maximum number of iterations needed for one time step was 284 (off scale).



(d) The quantity $\log_{10} \|\mathbf{u} - \mathbf{v}(r)\|$, which measures the difference between the original moments $\mathbf{u}$ and the regularized moments $\mathbf{v}(r)$ closed by the optimization. (The white space indicates where moments were not regularized.)

FIG. 5.5. *A simulation of the M_7 model of the two-beam instability with 1000 spatial cells.*

factors in flux-limited neutrino diffusion, Journal of Quantitative Spectroscopy and Radiative Transfer, 42 (1989), pp. 603 – 613.

[13] JEAN-FRANOIS COULOMBEL AND THIERRY GOUDON, *Entropy-based moment closure for kinetic equations: Riemann problem and invariant regions*, Journal of Hyperbolic Differential Equations, 6 (2006), pp. 649–671.

[14] RAÚL E. CURTO AND LAWRENCE A. FIALKOW, *Recursivness, positivity and truncated moment problems*, Houston Journal of Mathematics, 4 (1991), pp. 603–635.

[15] P.J. DAVIS AND P. RABINOWITZ, *Methods of Numerical Integration*, Academic Press, Orlando, second ed., 1984.

[16] P. DEGOND AND C. RINGHOFER, *Quantum moment hydrodynamics and the entropy principle*, J. Stat. Phys., 112 (2003), pp. 587–627.

[17] S. M. DESHPANDE, *Kinetic theory based new upwind methods for inviscid compressible flows*, in American Institute of Aeronautics and Astronautics, New York, 1986. Paper 86-0275.

[18] W. DREYER, *Maximisation of the entropy in non-equilibrium*, Journal of Physics A Mathematical General, 20 (1987), pp. 6505–6517.

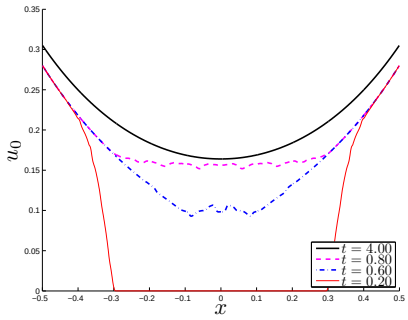
[19] W. DREYER, M. HERRMANN, AND M. KUNIK, *Kinetic solutions of the Boltzmann-Peierls equation and its moment systems*, Continuum Mechanics and Thermodynamics, 16 (2004), pp. 453–469.

[20] B. DUBROCA AND J.-L. FUEGAS, *Étude théorique et numérique d'une hiérarchie de modèles aus moments pour le transfert radiatif*, C.R. Acad. Sci. Paris, I. 329 (1999), pp. 915–920.

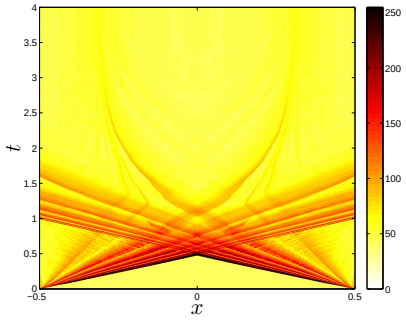
[21] B. DUBROCA AND A. KLAR, *Half-moment closure for radiative transfer equations*, J. Comput. Phys., 180 (2002), pp. 584–596.

[22] MARTIN FRANK, BRUNO DUBROCA, AND AXEL KLAR, *Partial moment entropy approximation to radiative heat transfer*, J. Comput. Phys., 218 (2006), pp. 1–18.

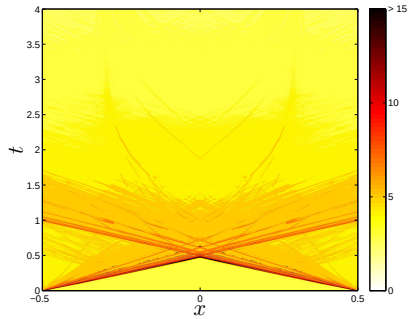
[23] SIGAL GOTTLIEB, CHI-WANG SHU, AND EITAN TADMOR, *Strong stability-preserving high-order time discretization methods*, SIAM Review, 43 (2001), pp. pp. 89–112.



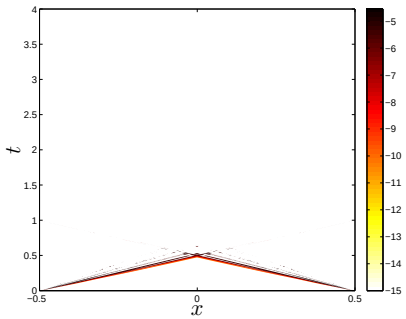
(a) Snapshots of the solution, $u_0(x,t)$ at $t = 0.2, 0.6, 0.8$ and 4.0.



(b) The average number of quadrature points (over both Runge-Kutta stages) used for each point in space and time.



(c) The total number of iterations (over both Runge-Kutta stages) needed to solve the optimization problem at each point in space and time. The maximum number of iterations needed for one time step was 364 (off scale).



(d) The quantity $\log_{10} \|\mathbf{u} - \mathbf{v}(r)\|$, which measures the difference between the original moments $\mathbf{u}$ and the regularized moments $\mathbf{v}(r)$ closed by the optimization. (The white space indicates where moments were not regularized.)

FIG. 5.6. *A simulation of the M_{15} model of the two-beam instability with 1000 spatial cells.*

[24] CLINTON GROTH AND JAMES MCDONALD, *Towards physically realizable and hyperbolic moment closures for kinetic theory*, Continuum Mechanics and Thermodynamics, (2009).

[25] A. HARTEN, P. D. LAX, AND VAN LEER, *On upstream differencing and Godunov-type schemes for hyperbolic conservation laws*, SIAM Rev., 25 (1983), pp. 35–61.

[26] C. D. HAUCK, *Entropy-Based Moment Closures in Semiconductor Models*, PhD thesis, University of Maryland, College Park, 2006.

[27] ———, *High-order entropy-based closures for linear transport in slab geometry*, Comm. Math. Sci., 9 (2011).

[28] ———, *High-order entropy-based closures for linear transport in slab geometry*, Communications in Mathematical Sciences, (to appear).

[29] CORY D. HAUCK, C. DAVID LEVERMORE, AND ANDRÉ L. TITS, *Convex duality and entropy-based moment closures: Characterizing degenerate densities*, SIAM J. Control Optim., 47 (2008), pp. 1977–2015.

[30] CORY D. HAUCK AND RYAN G. MCCLARREN, *Positive P_N closures*, SIAM Journal on Scientific Computing, (to appear).

[31] ANSGAR JÜNGEL, STEFAN KRAUSE, AND PAOLA PIETRA, *A hierarchy of diffusive higher-order moment equations for semiconductors*, SIAM Journal on Applied Mathematics, 68 (2007), pp. 171–198.

[32] M. JUNK, *Domain of definition of Levermore's five moment system*, J. Stat. Phys., 93 (1998), pp. 1143–1167.

[33] ———, *Maximum entropy for reduced moment problems*, Math. Mod. Meth. Appl. S., 10 (2000), pp. 1001–1025.

[34] M. JUNK AND V. ROMANO, *Maximum entropy moment system of the semiconductor Boltzmann equation using Kane's dispersion relation*, Continuum Mechanics and Thermodynamics, 17 (2005), pp. 247–267.

[35] E.W. LARSEN, *The asymptotic diffusion limit of discretized transport problems*, Nucl. Sci. Eng.,

112 (1992), pp. 336–346.
[36] E. W. Larsen and C. G. Pomraning, *The P_N theory as an asymptotic limit of transport theory in planar geometry—I: Analysis*, Nucl. Sci. Eng, 109 (1991), pp. 49–75.
[37] ———, *The P_N theory as an asymptotic limit of transport theory in planar geometry—II: Numerical results*, Nucl. Sci. Eng., 109 (1991), pp. 76–85.
[38] Bram Van Leer, *Towards the ultimate conservative difference scheme. iv. a new approach to numerical convection*, Journal of Computational Physics, 23 (1977), pp. 276 – 299.
[39] C. D. Levermore, *Boundary conditions for moment closures*. Presented at Institute for Pure and Applied Mathematics University of California, Los Angeles, CA on May 27, 2009.
[40] ———, *Moment closure hierarchies for kinetic theory*, J. Stat. Phys., 83 (1996), pp. 1021–1065.
[41] ———, *Moment closure hierarchies for the Boltzmann-Poisson equation*, VLSI Design, 6 (1998), pp. 97–101.
[42] C. D. Levermore, W. J. Morokoff, and B. T. Nadiga, *Moment realizability and the validity of the Navier-Stokes equations for rarefied gas dynamics*, Phys. Fluids, 10 (1998), pp. 3214–3226.
[43] Ryan G. McClarren and Cory D. Hauck, *Robust and accurate filtered spherical harmonics expansions for radiative transfer*, Journal of Computational Physics, 229 (2010), pp. 5597 – 5614.
[44] Sally A. McKee, William A. Wulf, and Trevor C. Landon, *Bounds on memory bandwidth in streamed computations*, in Euro-Par '95: Proceedings of the First International Euro-Par Conference on Parallel Processing, London, UK, 1995, Springer-Verlag, pp. 83–99.
[45] L. R. Mead and N. Papanicolaou, *Maximum entropy in the problem of moments*, Journal of Mathematical Physics, 25 (1984), pp. 2404–2417.
[46] G. N. Minerbo, *Maximum entropy Eddington factors*, J. Quant. Spectrosc. Radiat. Transfer, 20 (1978), pp. 541–545.
[47] P. Monreal and M. Frank, *Higher order minimum entropy approximations in radiative transfer*. preprint.
[48] I. Müller and T. Ruggeri, *Rational Extended Thermodynamics*, vol. 37 of Springer Tracts in Natural Philosophy, Springer-Verlag, New York, second ed., 1993.
[49] Haim Nessyahu and Eitan Tadmor, *Non-oscillatory central differencing for hyperbolic conservation laws*, Journal of Computational Physics, 87 (1990), pp. 408 – 463.
[50] Gordon L. Olson, *Second-order time evolution of P_N equations for radiation transport*, 228 (2009), pp. 3072–3083.
[51] S. Osher, *Convergence of Generalized MUSCL Schemes*, SIAM Journal on Numerical Analysis, 22 (1985), pp. 947–961.
[52] A. Passalacqua and R.O. Fox, *Advanced continuum modelling of gas-particle flows beyond the hydrodynamic limit*, Applied Mathematical Modelling, 35 (2011), pp. 1616 – 1627.
[53] B. Perthame, *Boltzmann type schemes for gas dynamics and the entropy property*, SIAM J. on Numer. Anal., 27 (1990), pp. 1405–1421.
[54] B. Perthame, *Second-order Boltzmann schemes for compressible euler equations in one and two space dimensions*, SIAM J. Numer. Anal., 29 (1992), pp. 1–19.
[55] G. C. Pomraning, *Variational boundary conditions for the spherical harmonics approximation to the neutron transport equation*, Ann. Phys., 27 (1964), pp. 193–215.
[56] Salvatore La Rosa, Giovanni Mascali, and Vittorio Romano, *Exact maximum entropy closure of the hydrodynamical model for Si semiconductors: The 8-moment case*, SIAM Journal on Applied Mathematics, 70 (2009), pp. 710–734.
[57] J. Schneider, *Entropic approximation in kinetic theory*, Math. Model. Numer. Anal., 38 (2004), pp. 541–561.
[58] James Alexander Shohat and Jacob David Tamarkin, *The Problem of Moments*, American Mathematical Society, New York, 1943.
[59] J. M. Smit, L. J. van den Horn, and S. A. Bludman, *Closure in flux-limited neutrino diffusion and two-moment transport*, Astron. Astrophys., 356 (2000), pp. 559–569.
[60] Henning Struchtrup, *Kinetic schemes and boundary conditions for moment equations*, Z. Angew. Math. Phys., 51 (2000), pp. 346–365.
[61] Eleuterio F. Toro, *Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction*, Springer, New York, 2009.
[62] D. Wright, M. Frank, and A. Klar, *The minimum entropy approximation to the radiative transfer equation*. preprint.
[63] Wm. A. Wulf and Sally A. McKee, *Hitting the memory wall: implications of the obvious*, SIGARCH Comput. Archit. News, 23 (1995), pp. 20–24.