AN ENTROPY SATISFYING CONSERVATIVE METHOD FOR THE FOKKER–PLANCK EQUATION OF THE FINITELY EXTENSIBLE NONLINEAR ELASTIC DUMBBELL MODEL

HAILIANG LIU† AND HUI YU†

Abstract. In this paper, we propose an entropy satisfying conservative method to solve the Fokker–Planck equation of the finitely extensible nonlinear elastic dumbbell model for polymers, subject to homogeneous fluids. Both semidiscrete and fully discrete schemes satisfy all three desired properties—(i) mass conservation, (ii) positivity preserving, and (iii) entropy satisfying—in the sense that these schemes satisfy discrete entropy inequalities for both the physical entropy and the quadratic entropy. These ensure that the computed solution is a probability density and the schemes are entropy stable and preserve the equilibrium solutions. We also prove convergence of the numerical solution to the equilibrium solution as time becomes large. Zero flux at boundary is naturally incorporated, and boundary behavior is resolved sharply. Both one- and two-dimensional numerical results are provided to demonstrate the good qualities of the scheme and the effects of some canonical homogeneous flows.

Key words. Fokker–Planck equations, finitely extensible nonlinear elastic model, relative entropy, positivity preserving

AMS subject classifications. 35K20, 65M08, 76A05, 82C31, 82D60

DOI. 10.1137/110829611

1. Introduction. Dumbbell models with finitely extensible nonlinear elastic (FENE) spring forces are now widely used in numerical flow calculations to capture nonlinear rheological phenomena, both in the classical approach via a closed constitutive equation and in a modern approach in which the polymeric stress tensor is computed via Brownian dynamics simulations [10, 28]. For the dumbbell model the configuration probability density function (pdf) yields information on the probability of finding a dumbbell with a given configuration at a particular material point; hence solving the Fokker–Planck equation directly is desirable, as long as it is feasible [34].

The original empirical FENE spring potential,

\[
\Psi(m) = -\frac{Hb_0}{2} \log \left(1 - \frac{|m|^2}{b_0^2}\right),
\]

was first proposed by Warner [37], where \(H\) is the spring constant and \(m\) is the \(d\)-dimensional connector vector of the beads with \(m \in B := B(0, \sqrt{b_0})\), a ball in \(\mathbb{R}^d\) with radius \(\sqrt{b_0}\) denoting the maximum spring extension. It exhibits, for small extensions, the expected linear behavior and a finite length \(b_0\) in the limit of an infinite force.

This paper is concerned with the numerical solution of the Fokker–Planck equation of the FENE dumbbell model for the pdf \(f = f(x, m, t)\),

\[
\partial_t f + (v \cdot \nabla_x) f + \nabla_m \cdot (\nabla_x v m f) = \frac{2}{\zeta} \nabla_m \cdot (\nabla_m \Psi(m) f) + \frac{2k_B T}{\zeta} \Delta_m f,
\]

∗Received by the editors April 11, 2011; accepted for publication (in revised form) December 12, 2011; published electronically May 22, 2012. This research was supported by the National Science Foundation under Kinetic FRG grant DMS07-57227.

†Mathematics Department, Iowa State University, Ames, IA 50011 (hliu@iastate.edu, legendyu@iastate.edu).
where \( x \in \mathbb{R}^d \) is the macroscopic Eulerian coordinate, \( v(x, t) \), the fluid velocity, is usually governed by the incompressible Navier–Stokes equation, \( \zeta \) is the friction coefficient of the dumbbell beads, \( T \) is the absolute temperature, and \( k_B \) is the Boltzmann constant. We refer to Chapters 11 and 13 of [5] for a comprehensive survey of the physical background and to [17] for some augmented models with inertial forces.

Throughout this paper we consider only homogeneous flows. Therefore the velocity field of the fluid can be written as \( v = Kx \), where \( K = \nabla v \) is independent of the position vector \( x \) in the fluid and has zero trace since we assume the fluid to be incompressible. Let the flow map be defined as

\[
\partial_t X(y; t) = v(X(y; t), t), \quad X(y; 0) = y.
\]

Along the flow map, with a suitable scaling and \( b_0 \to b = \frac{Hb_0}{k_B T} \), we arrive at the following equation for \( f(m, t) := f(X(y; t), m, t) \) for each fixed \( y \):

\[
\begin{align}
\partial_t f &= \frac{1}{2} \nabla_m \cdot \left[ \nabla_m f + \left( \frac{bm}{b - |m|^2} - 2Km \right) f \right], \quad m \in B, \\
f(m, 0) &= f_0(m), \quad m \in B, \\
f = o(b - |m|^2) \quad \text{on } \partial B.
\end{align}
\]

From now on \( K \) is assumed to be a trace-free \( d \times d \) matrix, i.e., \( Tr(K) = 0 \). Boundary requirement (1.3c) is imposed to ensure the existence and uniqueness of the weak solution to (1.3) (see [31]).

The singularity of the Fokker–Planck equation near \( |m| = \sqrt{b} \) makes the boundary issue rather subtle [29] and presents numerous challenges, both analytically and numerically. These issues are particularly important in solving the coupled Navier–Stokes–Fokker–Planck system, in which the behavior of the polymer distribution near boundaries is of significance. Consequently, computing with sharp resolution and stability near boundaries is a major goal. On the other hand the pdf is the practically relevant solution [32] for the underlying Fokker–Planck equation. It is therefore desirable to design a method which preserves three important properties of the pdf: constant integral (mass conservation), positivity preserving, and entropy satisfying in the sense that entropy inequalities are satisfied at the discrete level. In this paper, we develop such a method.

A key concept in the design of our numerical method is the relative entropy. To illustrate the idea, we reformulate the Fokker–Planck equation (1.3a). If \( K \) is normal in the sense that it commutes with its transpose, i.e., \( KK^\top = K^\top K \), it can be verified that the equilibrium solution can be determined explicitly as

\[
M = (b - |m|^2)^{b/2} \exp(m^T K^a m),
\]

where \( K^a \) is the symmetric part of \( K \). Let \( K^a \) be the antisymmetric part of \( K \); then the Fokker–Planck equation can be rewritten as

\[
\begin{align}
\partial_t f &= \frac{1}{2} \nabla_m \cdot (M \nabla_m g - 2K^a m f), \quad f = gM.
\end{align}
\]

Using the zero flux boundary condition (1.6), it can be shown that the relative entropy

\[
E(t) := \int_B \frac{f^2}{M} dm = \int_B g^2 M dm
\]

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
satisfies the inequality
\[ E(t) + \int_0^t \int_B M |\nabla_m g|^2 dmd\tau \leq E(0) \quad \forall t > 0. \]

This entropy dissipation relation ensures that the relative entropy is decreasing in time, and as time evolves the solution is expected to converge toward the equilibrium, i.e.,
\[ \lim_{t \to \infty} f(t, m) = CM(m) \]
for some \( C > 0 \). One may also use the physical entropy defined by
\[ E_p(t) = \int_B f \log \left( \frac{f}{M} \right) dm, \]
which satisfies the following entropy dissipation equation:
\[ E_p(t) + \int_0^t \int_B M |\nabla_m g|^2 g dm \leq E_p(0) \quad \forall t > 0. \]

Note that the physical entropy is bounded as long as \( E \) is bounded since
\[ E_p(t) \leq \int_B f \left( \frac{f}{M} - 1 \right) dm = E(t) - \int_B f dm = E(t) - \int_B f_0(m) dm. \]

For initial density with \( E(0) < \infty \) it suffices to consider the quadratic entropy \( E(t) \), which is particularly convenient to use for higher order methods. As the first step, we shall design a finite volume scheme based on (1.5) and show positivity and stability properties in terms of the relative entropy. More precisely, for both semidiscrete and fully discrete schemes presented in this work, we are able to prove the entropy stability for both quadratic and physical entropy, based on which we also prove the long time convergence. Existence of positive solutions is established as well.

For nonhomogeneous flows, which is the case when considering the coupled problem with the Navier–Stokes equation, we may apply the method developed in this paper using operator splitting. For instance, for each fixed \( m \), one may solve the transport equation
\[ \partial_t f + \nabla \cdot (vf) = 0 \]
with \( v \) obtained from solving the Navier–Stokes equation. With the obtained \( f \) as initial data, one further solves the Fokker–Planck equation with \( K = \nabla_x v(x, t) \). Note that this treatment using operator splitting techniques is a standard tool in fluid simulations; see [15, 16].

1.1. Related work. The regime of physical interest is \( b > 2 \), for which the boundary requirement (1.3c) was shown to be a sharp requirement for the solution to remain a probability density [31]. Moreover, this condition is equivalent to the zero flux boundary condition for \( b > 2 \) as shown in [31],

\[ \nabla_m f + \left( \frac{bm}{b - |m|^2} - 2Km \right) f \cdot m = 0, \quad m \in \partial B. \]

For theoretical results concerning the existence of solutions of the coupled system we refer to [32, 33, 39]; see also the works [11, 12, 13] and the earlier works on this
problem [22, 23]. For rigorous analysis of long-time asymptotics of the FENE model, see [24]; and see [3] for entropy methods to study rate of convergence to equilibrium for Fokker–Planck type equations.

For some special configuration solutions with small flow rates, the use of moment closure approximations has been investigated by several authors; see, e.g., [18, 20, 21, 30, 38]. Most numerical methods developed for the Fokker–Planck equation have been based on the form of (1.3a); see, for example, [1, 2, 19, 37]. Some elaborate numerical algorithms based on spectral methods were recently developed for the Fokker–Planck equation of the FENE model in [15, 16, 26]. A spectral-Galerkin approximation was further introduced in [25] based on a weighted weak formulation for \( f(b - |m|^2)^{-b/4} \). An improved weighted formulation was proposed in [36] in terms of \( f(b - |m|^2)^{-s/2} \) for \( 1 < s \leq b \), leading to a different spectral-Galerkin algorithm. We note that this weighted formulation was also used for specific values of \( s \) in [15, 16, 26] and was analyzed in section 3.2 of [25]. The methods in [25, 36] have provable stability results for certain weighted integrable initial data. However, positivity of the numerical solution is not guaranteed.

Finally we comment on the concept of entropy explored in numerical approximations. There is a vast literature on entropic schemes for related equations including hyperbolic conservation laws and kinetic equations such as Fokker–Planck type equations. For the former, entropy dissipation at the discrete level is often enforced through numerical viscosity so that physical relevant shocks, particularly exact stationary shocks, can be captured; see, e.g., [8, 9, 35]. For the later, information carried by the pdf becomes less and less as time evolves; the probability density is expected to converge to the equilibrium solution in a closed system regardless of how initial data are distributed. The entropy dissipation in time is the underlying mechanism for this phenomenon. To ensure the entropy property at discrete levels, one often uses the logarithmic Landau form

\[
\partial_t f = \frac{1}{2} \nabla \cdot \left( f \nabla \log \frac{f}{M} \right),
\]

see [4]. For a nonlinear Fokker–Planck equation, it was shown in [7] that the scheme based on some entropic averages makes the Landau form equivalent to the underlying equation at the discrete level. Another class of finite difference schemes for Fokker–Planck equations is due to Chang and Cooper; see [14, 27]. This method is based upon the requirement that the discrete Fokker–Planck operator possesses a quasi-equilibrium solution which agrees at the mesh points with a quasi-equilibrium solution of the analytic operator. For a linear Fokker–Planck equation the Chang–Cooper scheme is shown in [6] to make the underlying equation equivalent to the nonlogarithmic Landau form

\[
\partial_t f = \frac{1}{2} \nabla \cdot \left( M \nabla \frac{f}{M} \right)
\]

at the discrete level. In this paper we explore the nonlogarithmic Landau form subject to a nonsymmetric drift term (1.5). The novel features include (1) the equilibrium \( M \) has no positive lower bound, but zero at the boundary, making the Landau formulation singular and numerical computations more difficult; (2) the force due to fluid effects is generally nonconservative; one has to consider a nonsymmetric perturbation upon the usual Landau formulation, which makes the study of long time convergence more interesting; and (3) the natural function space for \( f \) is \( ML^2(Mdm) \), which when \( M \)
vanishes at the boundary is different from the usual weighted space $L^2(Mdm)$ [31], hence the corresponding Galerkin discretization is not standard.

1.2. Contents. This paper is organized as follows. In section 2, we describe the formulation of our scheme for the one-dimensional case. Theoretical analysis for both semidiscrete and fully discrete schemes is provided. In section 3, we generalize the scheme to two space dimensions. Implementation strategies and numerical results of both one and two dimensions are presented in section 4. Finally, in section 5, concluding remarks are given.

2. One-dimensional Fokker–Planck equation. We begin by looking at the Fokker–Planck problem over the interval $B = (-\sqrt{b}, \sqrt{b})$ in one-dimensional space. In such a case $K = 0$ because of the constraint $Tr(K) = 0$; then the problem can be described as

\begin{align}
\partial_t f &= \frac{1}{2} \partial^2_m f + \frac{1}{2} \partial_m \left( \frac{bm}{b - m^2} f \right), \quad m \in B, \quad t > 0, \\
(f(m, 0) &= f_0(m), \quad m \in B, \\
\left. \left( \partial_m f + \frac{bm}{b - m^2} f \right) \right|_{m=\pm\sqrt{b}} &= 0, \quad t > 0.
\end{align}

The associated equilibrium solution reduces to

\[ M(m) = \left( b - m^2 \right)^{\frac{1}{2}}, \quad m \in B, \]

and (2.1a) becomes

\[ \partial_t f = \frac{1}{2} \partial_m (M \partial_m g), \quad \text{where} \ g = \frac{f}{M}. \]

2.1. Semidiscrete scheme. Given a positive integer $N$, we partition the domain $(-\sqrt{b}, \sqrt{b})$ by defining the uniform mesh size $h = \frac{2\sqrt{b}}{N}$ and the cell center at

\[ m_j = -\sqrt{b} + \left( j - \frac{1}{2} \right) h, \quad 1 \leq j \leq N. \]

Notice that at two end points $M(m_{\frac{1}{2}}) = M(m_{N+\frac{1}{2}}) = 0$ and $M(m_{j+\frac{1}{2}}) > 0$ for $1 \leq j \leq N - 1$. On each computational cell $I_j = \left[ m_{j-\frac{1}{2}}, m_{j+\frac{1}{2}} \right]$, we define the cell average of $f$ as

\[ \bar{f}_j(t) = \frac{1}{h} \int_{I_j} f(m, t) \, dm. \]

Integration of (2.2) on $I_j$ yields

\[ \frac{d}{dt} \bar{f}_j = \frac{1}{2h} \int_{I_j} \partial_m (M \partial_m g) \, dm = \frac{1}{2h} M \partial_m g \bigg|_{m_{j-\frac{1}{2}}}^{m_{j+\frac{1}{2}}}. \]

Based on this formulation we derive a finite volume scheme to compute $\{f_j\}$ which approximates $\{\bar{f}_j\}$ by taking the numerical flux

\[ J_{j+\frac{1}{2}} = M \partial_m g = M_{j+\frac{1}{2}} \frac{1}{h} (g_{j+1} - g_j) \quad \text{for} \quad j = 1, \ldots, N - 1 \]

with $M_{j+\frac{1}{2}} := M(m_{j+\frac{1}{2}})$, $g_j(t) = \frac{f_j(t)}{M_j}$, where $M_j = M(m_j)$. We also set

\[ J_{\frac{1}{2}} = J_{N+\frac{1}{2}} = 0 \]

to incorporate the zero flux at the boundary.
Then we obtain a semidiscrete scheme

$$
\frac{d}{dt} f_j = \frac{1}{2h} J_{\frac{1}{2}},
$$

(2.5)

$$
\frac{d}{dt} f_j = \frac{1}{2h} (J_{j+\frac{1}{2}} - J_{j-\frac{1}{2}}), \quad 2 \leq j \leq N - 1,
$$

$$
\frac{d}{dt} f_N = -\frac{1}{2h} J_{N-\frac{1}{2}}
$$

subject to the initial data

$$
f_j(0) = \frac{1}{h} \int_{I_j} f_0(m) \, dm, \quad j = 1, \ldots, N.
$$

**Theorem 2.1.** The semidiscrete scheme (2.5) satisfies the following properties:

1. Conservation of mass: \( \sum_{j=1}^{N} f_j(t) h = \sum_{j=1}^{N} f_j(0) h = \int_B f_0(m) \, dm \forall t > 0 \).

2. Positivity preserving: for any \( t > 0 \), \( f_j(t) \geq 0 \) if \( f_j(0) \geq 0 \).

3. The relative entropy \( E(t) = \sum_{j=1}^{N} \frac{f_j^2}{M_j} h \) is nonincreasing in time with

$$
\frac{d}{dt} E(t) = -\frac{1}{h} \sum_{j=1}^{N-1} (g_{j+1} - g_j)^2 M_{j+\frac{1}{2}} \leq 0.
$$

**Proof.** (1) Summing all equations in (2.5), we have

$$
\frac{d}{dt} \sum_{j=1}^{N} f_j(t) = \sum_{j=1}^{N} \frac{d}{dt} f_j(t) = 0.
$$

So

$$
\sum_{j=1}^{N} f_j(t) h = \sum_{j=1}^{N} f_j(0) h = \int_B f_0(m) \, dm.
$$

(2) Since \( M_j \) is independent of \( t \), we have \( \frac{d}{dt} f_j = M_j \frac{d}{dt} g_j \). The scheme (2.5) can be rewritten as

$$
\frac{d}{dt} g_1 = \frac{1}{2h^2 M_1} M_{\frac{1}{2}}(g_2 - g_1),
$$

(2.6)

$$
\frac{d}{dt} g_j = \frac{1}{2h^2 M_j} [M_{j+\frac{1}{2}}(g_{j+1} - g_j) - M_{j-\frac{1}{2}}(g_j - g_{j-1})], \quad 2 \leq j \leq N - 1,
$$

$$
\frac{d}{dt} g_N = -\frac{1}{2h^2 M_N} M_{N-\frac{1}{2}}(g_N - g_{N-1}).
$$

From (1), we see that

$$
\sum_{j=1}^{N} M_j g_j(t) = \sum_{j=1}^{N} f_j(0) \forall t > 0.
$$

Then all the trajectories of (2.6) remain on this hyperplane. We define a closed set on this hyperplane by

$$
\Sigma = \left\{ \tilde{g} : g_j \geq 0, j = 1, \ldots, N, \text{and} \sum_{j=1}^{N} M_j g_j = \sum_{j=1}^{N} f_j(0) \right\}.
$$

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
Let \( \vec{F}(\vec{g}) \) be the vector field defined by the right-hand side of (2.6); then
\[
\frac{d}{dt} \vec{g} = \frac{1}{2} \vec{F}(\vec{g}).
\]
It suffices to show that \( \Sigma \) is an invariant region of this system. This is indeed the case if the vector field \( \vec{F}(\vec{g}) \) points strictly into \( \Sigma \) on the boundary \( \partial \Sigma \); i.e., for any outward normal vector \( \vec{n} \) on any part of \( \partial \Sigma \),
\[
\vec{F}(\vec{g}) \cdot \vec{n} < 0.
\]
From (2.6), it follows that
\[
\vec{F}(\vec{g}) \cdot \vec{n} = \sum_{j=1}^{N-1} \frac{n_j}{h^2 M_j} \frac{M_j + \frac{1}{2}}{M_j - \frac{1}{2}} (g_{j+1} - g_j) - \sum_{j=2}^{N} \frac{n_j}{h^2 M_j} \frac{M_j - \frac{1}{2}}{M_j + \frac{1}{2}} (g_j - g_{j-1}),
\]
(2.8)
\[
= - \frac{1}{h^2} \sum_{j=1}^{N-1} \left( n_{j+1} - \frac{n_j}{M_j} \right) M_j + \frac{1}{2} (g_{j+1} - g_j).
\]
For each \( \vec{g} \in \partial \Sigma \), we define the set of indices \( S \) such that
\[
S = \{ 1 \leq j \leq N : g_j = 0 \},
\]
which implies that \( S \neq \emptyset \) for any \( \vec{g}_b \in \partial \Sigma \). Then the outward normal vectors \( \vec{n} \) at \( \vec{g}_b \) are of the form
\[
\vec{n} = (n_1, \ldots, n_N)^T \quad \text{with} \quad n_j = \begin{cases} -\alpha_j & \text{if } j \in S, \\ \frac{1}{M_j} & \text{if } j \notin S. \end{cases}
\]
Furthermore, there exists a positive real number \( \gamma \) such that \( \vec{g}_b - \gamma \vec{n} \) is in the interior of \( \Sigma \), which implies that
\[
\alpha_j > 0, \quad j \in S,
\]
and
\[
\sum_{j=1}^{N} M_j n_j = 0, \quad \text{i.e.,} \quad \sum_{j \in S} M_j \alpha_j = \sum_{j \notin S} M_j^2.
\]
Now we look back at (2.8). Note that if \( j, j+1 \in S \), then \( g_j = g_{j+1} = 0 \); if \( j, j+1 \notin S \), then \( \frac{n_{j+1}}{M_{j+1}} - \frac{n_j}{M_j} = 1 - 1 = 0 \). Therefore the nonzero terms in (2.8) are only those with \( j \in S, j+1 \notin S \) or \( j \notin S, j+1 \in S \). Hence
\[
\vec{F}(\vec{g}) \cdot \vec{n} = - \frac{1}{h^2} \left( \sum_{j \in S, j+1 \notin S} \frac{n_j + \frac{1}{2}}{M_j} M_j - \frac{1}{2} (g_{j+1} - g_j) \right)
\]
\[
= - \frac{1}{h^2} \sum_{j \in S, j+1 \notin S} \left( 1 + \frac{\alpha_j}{M_j} \right) M_j + \frac{1}{2} g_{j+1} - \frac{1}{h^2} \sum_{j \notin S, j+1 \in S} \left( 1 + \frac{\alpha_{j+1}}{M_{j+1}} \right) M_{j+1} + \frac{1}{2} g_j < 0.
\]
This leads to the conclusion that \( g_j(t) \geq 0 \) as long as \( g_j(0) \in \Sigma \).
(3) We now show that the relative entropy $E(t)$ is nonincreasing. In fact,

\[
\frac{d}{dt} \sum_{j=1}^{N} \frac{f_j^2}{M_j} h = 2 \sum_{j=1}^{N} \frac{f_j}{M_j} f_j h = \sum_{j=1}^{N} g_j (J_{j+\frac{1}{2}} - J_{j-\frac{1}{2}})
\]

\[
= - \sum_{j=1}^{N-1} (g_{j+1} - g_j) J_{j+\frac{1}{2}}
\]

\[
= - \frac{1}{h} \sum_{j=1}^{N-1} (g_{j+1} - g_j)^2 M_{j+\frac{1}{2}} \leq 0.
\]

**Theorem 2.2.** The physical entropy $E_p(t) = \sum_{j=1}^{N} f_j \log(\frac{f_j}{M_j}) h$ is nonincreasing in time. Moreover,

\[
\frac{d}{dt} E_p(t) = - \frac{1}{2h} \sum_{j=1}^{N-1} M_{j+\frac{1}{2}} \log \left( \frac{g_{j+1}}{g_j} \right) (g_{j+1} - g_j) \leq 0.
\]

**Proof.** A direct calculation using (2.5) and summation by parts gives

\[
\frac{d}{dt} \sum_{j=1}^{N} f_j \log \left( \frac{f_j}{M_j} \right) h = \sum_{j=1}^{N} \frac{d}{dt} f_j \left[ \log \left( \frac{f_j}{M_j} \right) + 1 \right] h
\]

\[
= \frac{1}{2} \sum_{j=1}^{N} (J_{j+\frac{1}{2}} - J_{j-\frac{1}{2}})(\log g_j + 1)
\]

\[
= - \frac{1}{2} \sum_{j=1}^{N-1} J_{j+\frac{1}{2}} (\log g_{j+1} - \log g_j)
\]

\[
= - \frac{1}{2h} \sum_{j=1}^{N-1} M_{j+\frac{1}{2}} (g_{j+1} - g_j) \log \left( \frac{g_{j+1}}{g_j} \right) \leq 0,
\]

where both (2.3) and $(X - Y)(\log X - \log Y) \geq 0$ have been used. \(\square\)

We may also examine the large time behavior of $\hat{g}(t)$. Positivity $\hat{g}(t) > 0$ and the constraint

\[
\sum_{j=1}^{N} g_j(t) M_j h = \int_B f_0(m) dm
\]

together ensure that $\hat{g}(t)$ will remain bounded for all time. Since (2.6) is an autonomous system, what happens as $t \to \infty$ is simple to describe. We summarize this result in the following theorem.

**Theorem 2.3.** Consider the semidiscrete scheme (2.5) subject to the initial data $f_j(0) > 0$ with $\sum_{j=1}^{N} f_j(0) h = \int_B f_0(m) dm$; then

\[
[f_1, f_2, \ldots, f_N]^\top \to C[M_1, \ldots, M_N], \quad t \to \infty,
\]

where

\[
C = \frac{\int_B f_0(m) dm}{\sum_{j=1}^{N} M_j h}.
\]
Proof. Define a functional \( V(g) \) by

\[
V(g) = \sum_{j=1}^{N} (g_j - C)^2 M_j
\]

and \( \bar{g} = C(1, \ldots, 1)^T \). We see that \( \bar{g} \in \Sigma \) as defined in (2.7), satisfying \( F(\bar{g}) = 0 \). A direct verification shows that \( V(\bar{g}) = E - C \int_B f_0(m) dm \), implying \( \frac{d}{dt} V = \frac{d}{dt} E \).

Hence \( V \) satisfies the following:

- \( V(\bar{g}) > 0 \) for any \( g \neq \bar{g} \) (positive definite).
- \( \frac{d}{dt} V \leq 0 \) \( \forall \) \( g \in \Sigma \) (negative semidefinite).
- The set \( \{ \frac{d}{dt} V = 0 \} \cap \Sigma \) does not contain any trajectories of the ODE system (2.5) besides the trajectory \( g(t) = \bar{g} \forall t > 0 \).

With these properties we can apply the Krasovskii–LaSalle principle to conclude that

\[
\lim_{t \to \infty} V(g(t)) = 0
\]

which ensures that \( g_j = \text{const} \), while within \( \Sigma \), \( g = \bar{g} \) must hold. The proof is thus complete. \( \square \)

2.2. Fully discrete scheme. Let the time step be denoted by \( k \), and the mesh ratio \( \lambda = \frac{k}{2h} \). We apply the backward Euler method to the semidiscrete scheme (2.5) to get

\[
\begin{align*}
(f_j^{n+1}) & = f_j^n + \lambda M_j^\frac{1}{2} (g_{j+1}^{n+1} - g_j^{n+1}), \\
(f_j^{n+1}) & = f_j^n + \lambda \left[ M_{j+\frac{1}{2}} (g_{j+1}^{n+1} - g_j^{n+1}) - M_{j-\frac{1}{2}} (g_{j-1}^{n+1} - g_j^{n+1}) \right], 2 \leq j \leq N - 1, \\
(f_N^{n+1}) & = f_N^n + \lambda M_{N-\frac{1}{2}} (g_N^{n+1} - g_{N-1}^{n+1}).
\end{align*}
\]

Given \( \{f_j^n\}, \{f_j^{n+1}\} \) can be obtained from \( f_j^{n+1} = M_j g_j^{n+1} \), where \( \{g_j^{n+1}\} \) solves the following linear system:

\[
\begin{align*}
(M_1 + \lambda M_{\frac{1}{2}}) g_1^{n+1} - \lambda M_{\frac{1}{2}} g_2^{n+1} & = f_1^n, \\
-\lambda M_{-\frac{1}{2}} g_{j-1}^{n+1} + [M_j + \lambda (M_{j+\frac{1}{2}} + M_{j-\frac{1}{2}})] g_j^{n+1} - \lambda M_{j-\frac{1}{2}} g_{j-1}^{n+1} & = f_j^n, 2 \leq j \leq N - 1, \\
-\lambda M_{N-\frac{1}{2}} g_{N-1}^{n+1} + (M_N + \lambda M_{N-\frac{1}{2}}) g_N^{n+1} & = f_N^n.
\end{align*}
\]

Theorem 2.4. The fully discrete scheme (2.11) has a unique solution \( \{f_j^n\} \).

Moreover, the solution satisfies the following properties:

1. Conservation of mass. \( \sum_{j=1}^{N} f_j^{n+1} h = \sum_{j=1}^{N} f_j^n h \).
2. Positivity. If \( f_j^n \geq 0 \), then \( f_j^{n+1} \geq 0 \).
3. The relative entropy

\[
E^n = \sum_{j=1}^{N} \frac{(f_j^n)^2}{M_j} h
\]
is nonincreasing. More precisely,

\begin{equation}
E^{n+1} = E^n - \frac{k}{h} \sum_{j=1}^{N-1} (g_j^{n+1} - g_j^{n+1})^2 M_j + \frac{N}{h} \sum_{j=1}^{N} \frac{(f_j^{n+1} - f_j^n)^2}{M_j}.
\end{equation}

(4) \( f_j^n \) converges as \( n \to \infty \) with

\[ f_j^n \to CM_j, \quad n \to \infty, \]

where \( C \) is defined in (2.10).

Proof. First, we show the existence of a solution to (2.11). Equation (2.12) is a linear system of \( Ag^{n+1} = \vec{f}^n \), where

\[ g^{n+1} = (g_1^{n+1}, \ldots, g_N^{n+1})^T, \quad \vec{f}^n = (f_1^n, \ldots, f_N^n)^T. \]

From the fact that \( A \) is a strictly diagonally dominant matrix, it follows that there is a unique solution \( g^{n+1} = A^{-1} \vec{f}^n \) for any \( \vec{f}^n \).

(1) Summing up the \( N \) equations in (2.11) gives

\[ \sum_{j=1}^{N} f_j^{n+1} - h = \sum_{j=1}^{N} f_j^n h. \]

(2) Since \( M_j > 0, 1 \leq j \leq N \), we only need to prove that \( g_j^{n+1} \geq 0 \forall j \). It suffices to show that \( \min_{1 \leq j \leq N} g_j^{n+1} = g_i^{n+1} \geq 0 \). We only show the case \( 2 \leq i \leq N-1 \), as the cases \( i = 1 \) and \( i = N \) are similar and simpler,

\[ f_i^n = -\lambda M_i - g_i^{n+1} + [M_i + \lambda M_i + \lambda M_i - \lambda M_i + \lambda M_i] g_i^{n+1} - \lambda M_i + \lambda M_i g_i^{n+1} \]

\[ \leq -\lambda M_i - g_i^{n+1} + [M_i + \lambda M_i + \lambda M_i] g_i^{n+1} - \lambda M_i + \lambda M_i g_i^{n+1} \]

\[ = M_i g_i^{n+1}. \]

Hence, \( g_i^{n+1} \geq M_i^{-1} f_i^n \geq 0 \).

(3) As for the relative entropy, we calculate

\[ \sum_{j=1}^{N} \left[ \frac{(f_j^{n+1})^2}{M_j} - \frac{(f_j^n)^2}{M_j} \right] = \sum_{j=1}^{N} \frac{(2f_j^{n+1} + f_j^n - f_j^{n+1})(f_j^{n+1} - f_j^n)}{M_j} \]

\[ = \frac{k}{h} \sum_{j=1}^{N} g_j^{n+1} (J_{j+\frac{1}{2}}^{n+1} - J_{j-\frac{1}{2}}^{n+1}) - \sum_{j=1}^{N} \frac{(f_j^{n+1} - f_j^n)^2}{M_j} \]

\[ = -\frac{k}{h} \sum_{j=1}^{N-1} (g_j^{n+1} - g_j^n)^2 M_j + \frac{N}{h} \sum_{j=1}^{N} \frac{(f_j^{n+1} - f_j^n)^2}{M_j} \leq 0. \]

This yields (2.13), which implies that the relative entropy is nonincreasing.

(4) Since \( E^n \) is nonincreasing and bounded from below, we have

\[ \lim_{n \to \infty} E^n = \inf \{ E^n \}. \]

Observe from (2.13) that \( E^n - E^{n+1} \) is a sum of nonnegative, bounded terms. When passing limit \( n \to \infty \) we conclude that each term must have zero as its limit, that is,

\begin{equation}
\lim_{n \to \infty} (f_j^{n+1} - f_j^n)^2 = 0, \quad \lim_{n \to \infty} (g_j^{n+1} - g_j^n)^2 = 0.
\end{equation}
The first relation in (2.14) tells us that $\bar{\mathbf{g}}^n$ is a Cauchy sequence, which when combined with the completeness of $\Sigma$ (a closed and bounded set in $\mathbb{R}^N$) ensures that $\lim_{n \to \infty} \bar{\mathbf{g}}^n$ exists. The second relation in (2.14) infers that the limit must be $\bar{\mathbf{g}}$. The proof is complete. □

**Theorem 2.5.** The physical entropy

$$E_p^n = \sum_{j=1}^{N} f_j^n \log \left( \frac{f_j^n}{M_j} \right) h$$

is nonincreasing. Moreover,

$$E_p^{n+1} - E_p^n = -\sum_{j=1}^{N-1} \frac{kM_{j+\frac{1}{2}}}{2h} (g_{j+1}^{n+1} - g_j^{n+1})(\log g_{j+1}^{n+1} - \log g_j^{n+1})$$

$$+ \sum_{j=1}^{N} hf_j^n \log \left( \frac{f_j^{n+1}}{f_j^n} \right) \leq 0.$$

**Proof.** For physical entropy $E_p^n = \sum_{j=1}^{N} f_j^n \log \left( \frac{f_j^n}{M_j} \right) h = \sum_{j=1}^{N} g_j^n \log (g_j^n) M_j h$, we calculate

$$E_p^{n+1} - E_p^n = \sum_{j=1}^{N} hf_j^{n+1} \log g_j^{n+1} - \sum_{j=1}^{N} hf_j^n \log g_j^n + \sum_{j=1}^{N} hf_j^n \log g_j^{n+1} - \sum_{j=1}^{N} hf_j^n \log g_j^n$$

$$= \sum_{j=1}^{N} h(f_j^{n+1} - f_j^n) \log g_j^{n+1} + \sum_{j=1}^{N} f_j^n \log \left( \frac{f_j^{n+1}}{f_j^n} \right) h.$$  

Using scheme (2.11) and $\log x \leq x - 1$ for $x > 0$, we estimate

$$E_p^{n+1} - E_p^n \leq \sum_{j=1}^{N} \frac{k}{2} (J_{j+\frac{1}{2}}^{n+1} - J_j^{n+1}) \log g_j^{n+1} + \sum_{j=1}^{N} hf_j^n \left( \frac{f_j^{n+1}}{f_j^n} - 1 \right)$$

$$= -\sum_{j=1}^{N} \frac{kM_{j+\frac{1}{2}}}{2h} (g_{j+1}^{n+1} - g_j^{n+1})(\log g_{j+1}^{n+1} - \log g_j^{n+1})$$

$$+ \sum_{j=1}^{N} (f_j^{n+1} - f_j^n) h \leq 0$$

for the first summation is nonnegative due to monotonicity of $\log x$, and mass conservation implies that the second summation is zero. □

**3. Extension to the multidimensional FENE model.**

**3.1. Reformulation.** Let the matrix $\mathcal{K}$ be decomposed into a sum of the symmetric part and the asymmetric part, i.e.,

$$\mathcal{K} = \mathcal{K}^s + \mathcal{K}^a.$$  

Define $M(m)$ as

$$M(m) = (b - |m|^2)_{\frac{1}{2}} e^{m^T \mathcal{K}^s m}$$

(3.1)
and \( g(m, t) = \frac{f(m, t)}{M(m)} \), then the Fokker–Planck equation (1.3a) can be rewritten as

\[
\partial_t f = \frac{1}{2} \nabla_m \cdot (M \nabla_m g - 2\mathcal{K}^a m f).
\]

**Lemma 3.1.** Let \( f \) be a solution to (3.2). If \( \mathcal{K} \) is normal, then \( M(m) \) is the equilibrium solution to (3.2). Moreover, the relative entropy \( E = \int_B g^2 M dm \) satisfies

\[
\frac{d}{dt} E(t) + \int_B M |\nabla_m g|^2 dm = 0.
\]

**Proof.** Using the zero flux condition in the evolution of \( E \) we find that

\[
\frac{d}{dt} E = \int_B 2g \partial_t f dm
= \int_B g \nabla_m \cdot [M \nabla_m g - 2\mathcal{K}^a m f] dm
= - \int_B M |\nabla_m g|^2 dm + 2 \int_B \nabla_m g \cdot \mathcal{K}^a m f dm.
\]

Let \( B_r \) be a ball with radius \( r < \sqrt{b} \); then using integration by parts we obtain

\[
2 \int_{B_r} \nabla_m g \cdot \mathcal{K}^a m f dm = \int_{B_r} \nabla_m g^2 \cdot \mathcal{K}^a m M dm
= \int_{\partial B_r} g^2 M K^a m \frac{m}{|m|} dS - \int_{B_r} g^2 \nabla_m \cdot (\mathcal{K}^a m M) dm
= \int_{B_r} g^2 \mathcal{K}^a m \cdot \nabla_m M dm,
\]

which, in virtue of \( \nabla_m M = (2\mathcal{K}^a m - \frac{bm}{\sqrt{m^2}}) M \), reduces to

\[
\int_{B_r} Mg^2 m^\top \mathcal{K}^a m dm = \frac{1}{4} \int_{B_r} Mg^2 m^\top (\mathcal{K}^\top \mathcal{K} - \mathcal{K} \mathcal{K}^\top) dm.
\]

This vanishes if \( \mathcal{K} \) is normal. Let \( r \to \sqrt{b} \) and we obtain

\[
\int_B \nabla_m g \cdot \mathcal{K}^a m f dm = 0,
\]

hence the desired estimate (3.3) follows.

**Remark 3.1.** If \( \mathcal{K} \) is not normal, the above estimate can still be obtained if we replace \( M \) by the equilibrium solution. But in such a case, an explicit expression of the equilibrium solution is not available. With \( M \) defined above, we will have

\[
0 \neq 2 \int_B \nabla_m g \cdot \mathcal{K}^a m f dm \leq \frac{1}{2} \int_B M |\nabla_m g|^2 dm + 2a^2 b \int_B Mg^2 dm.
\]

Hence

\[
\frac{d}{dt} E + \frac{1}{2} \int_B M |\nabla_m g|^2 dm \leq 2a^2 b E,
\]
leading to

\[ E(t) \leq e^{2a^2 bt} E(0) \quad \text{for } t > 0. \]

In such a case, \( E \) is no longer decreasing, though still bounded in finite time. In the discretization to follow, we shall focus only on the two-dimensional case, for which \( \mathcal{K} \) has the form

\[
\mathcal{K} = \begin{pmatrix}
  k_{11} & k_{12} \\
  k_{21} & -k_{11}
\end{pmatrix}
\]

with

\[
\mathcal{K}^a = \begin{pmatrix}
  0 & \frac{k_{12} - k_{21}}{2} \\
  -\frac{k_{12} - k_{21}}{2} & 0
\end{pmatrix} = a \begin{pmatrix}
  0 & 1 \\
  -1 & 0
\end{pmatrix}
\]

and

\[
\mathcal{K}^s = \begin{pmatrix}
  \frac{k_{11} + k_{21}}{2} & \frac{k_{12} + k_{21}}{2} \\
  \frac{k_{12} + k_{21}}{2} & -k_{11}
\end{pmatrix}.
\]

### 3.2. Discretization in \( m \in B \).

The domain \( B \) shown in Figure 1 can be represented by \([0, \sqrt{b}) \times [0, 2\pi]\) in the polar coordinate system. Partition \( B \) into uniform rectangles

\[
K_{ij} = \{(r, \theta); r_{i-\frac{1}{2}} \leq r \leq r_{i+\frac{1}{2}}, \theta_{j-\frac{1}{2}} \leq \theta \leq \theta_{j+\frac{1}{2}}\}, \quad 1 \leq i \leq P, 1 \leq j \leq Q,
\]

where

\[
r_{i+\frac{1}{2}} = i \Delta r, \quad \theta_{j+\frac{1}{2}} = j \Delta \theta
\]

with steps of radius and angle

\[
\Delta r = \frac{\sqrt{b}}{P}, \quad \Delta \theta = \frac{2\pi}{Q}.
\]

Let the cell average of \( f \) on \( K_{ij} \) be defined by

\[
\bar{f}_{ij} = \frac{1}{|K_{ij}|} \int_{K_{ij}} f(m, t) \, dm,
\]
where $|K_{ij}| = \Delta \theta \Delta r_i$ is the area of cell $K_{ij}$. Integrate (3.2) over $K_{ij}$ on both sides,
\[
\frac{d}{dt} f_{ij} = \frac{1}{2|K_{ij}|} \int_{K_{ij}} \nabla_m \cdot (M \nabla_m g - 2K^a m f) \, dm 
\]
by the divergence theorem. Here $\nu$ is the outward normal of the cell boundary $\partial K_{ij}$.

In order to derive a finite volume scheme, we use $f_{ij} = g_{i,j} M_{i,j}$ as the numerical solution in $K_{ij}$ to approximate $f_{i,j}$ and represent (3.4) in terms of $\{f_{i,j}\}$.

Because numerical representatives $f$ and $g$ are not defined on $\partial K_{ij}$, we need to define a numerical flux to represent $(M \nabla_m g - 2K^a m f) \cdot \nu$ on $\partial K_{ij}$. To simplify the presentation, we introduce two difference operators,
\[
D_r g_{i,j} = \frac{g_{i+1,j} - g_{i,j}}{\Delta r}, \quad D_\theta g_{i,j} = \frac{g_{i,j+1} - g_{i,j}}{\Delta \theta}.
\]
There are four pieces within $\partial K_{ij}$, denoted by $\gamma_1, \gamma_2, \gamma_3$, and $\gamma_4$. On $\gamma_1 = \{(r, \theta); r = r_{i+\frac{1}{2}}, \theta_{j-\frac{1}{2}} \leq \theta \leq \theta_{j+\frac{1}{2}}\}$, we have
\[
\int_{\gamma_1} M \nabla_m g \cdot \nu \, ds = \int_{\theta_{j-\frac{1}{2}}}^{\theta_{j+\frac{1}{2}}} M(r_{i+\frac{1}{2}}, \theta) \partial_r g(r_{i+\frac{1}{2}}, \theta) r_{i+\frac{1}{2}} \, d\theta
\]
\[
= \int_{\theta_{j-\frac{1}{2}}}^{\theta_{j+\frac{1}{2}}} M(r_{i+\frac{1}{2}}, \theta) D_r g_{i,j} r_{i+\frac{1}{2}} \, d\theta
\]
\[
= \Delta r_{i+\frac{1}{2}} M_{i+\frac{1}{2},j} D_r g_{i,j},
\]
where we use the midpoint rule for the integration in $\theta$, and $\partial_r g(r_{i+\frac{1}{2}}, \theta) = D_r g_{i,j}$.

Similarly, on $\gamma_3 = \{(r, \theta); r = r_{i-\frac{1}{2}}, \theta_{j-\frac{1}{2}} \leq \theta \leq \theta_{j+\frac{1}{2}}\}$,
\[
\int_{\gamma_3} M \nabla_m g \cdot \nu \, ds = -\Delta r_{i-\frac{1}{2}} M_{i-\frac{1}{2},j} D_r g_{i-1,j}.
\]
On $\gamma_2 = \{(r, \theta); r_{i-\frac{1}{2}} \leq r \leq r_{i+\frac{1}{2}}, \theta = \theta_{j+\frac{1}{2}}\}$, we have $\nu = (-\sin \theta, \cos \theta)^T$ and $\nabla_m \cdot \nu = \frac{1}{r} \frac{\partial}{\partial \theta}$, hence
\[
\int_{\gamma_2} M \nabla_m g \cdot \nu \, ds = \int_{r_{i-\frac{1}{2}}}^{r_{i+\frac{1}{2}}} \frac{M(r, \theta_{j+\frac{1}{2}})}{r} \partial_\theta g(r, \theta_{j+\frac{1}{2}}) \, dr
\]
\[
= \int_{r_{i-\frac{1}{2}}}^{r_{i+\frac{1}{2}}} \frac{M(r, \theta_{j+\frac{1}{2}})}{r} D_\theta g_{i,j} \, dr
\]
\[
= \frac{\Delta r_{i+\frac{1}{2}}}{r_i} D_\theta g_{i,j},
\]
where we have taken $\partial_\theta g(r, \theta_{j+\frac{1}{2}}) = D_\theta g_{i,j}$. Similarly, on $\gamma_4 = \{(r, \theta); r_{i-\frac{1}{2}} \leq r \leq r_{i+\frac{1}{2}}, \theta = \theta_{j-\frac{1}{2}}\}$,
\[
\int_{\gamma_4} M \nabla_m g \cdot \nu \, ds = -\frac{\Delta r_{i+\frac{1}{2}}}{r_i} D_\theta g_{i,j-1}.
\]
For the asymmetric part,
\[
2\mathbf{K}^a \cdot \vec{\nu} = 2a \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \end{pmatrix} \cdot \vec{\nu} = \begin{cases} 0 & \text{on } \gamma_1, \\ -2ar & \text{on } \gamma_2, \\ 0 & \text{on } \gamma_3, \\ 2ar & \text{on } \gamma_4. \end{cases}
\]

It follows that
\[
\int_{\partial K_{ij}} (2\mathbf{K}^a m) \cdot \vec{\nu} \, ds = -2a \int_{r_i - \frac{1}{2}}^{r_i + \frac{1}{2}} r f(r, \theta) \, dr + 2a \int_{r_i - \frac{1}{2}}^{r_i + \frac{1}{2}} r f(r, \theta) \, dr.
\]
The numerical flux is chosen to be upwind,
\[
f(r, \theta) = \frac{1}{2}(f_{i,j+1} + f_{i,j}) + \frac{\text{sgn}(a)}{2}(f_{i,j+1} - f_{i,j}).
\]
Hence
\[
\int_{\partial K_{ij}} (2\mathbf{K}^a m) \cdot \vec{\nu} \, ds = -\Delta r r_i [(a + |a|)(f_{i,j+1} - f_{i,j}) + (a - |a|)(f_{i,j} - f_{i,j-1})].
\]

Therefore we obtain the semidiscrete scheme
\[
\frac{d}{dt} f_{i,j} = \frac{r_i + \frac{1}{2}M_{i,j+1}}{2\Delta r r_i} D_r g_{i,j} - \frac{r_i - \frac{1}{2}M_{i,j-1}}{2\Delta r r_i} D_r g_{i-1,j} + \frac{M_{i,j+1}}{2\Delta \theta r_i} D_\theta g_{i,j} - \frac{M_{i,j-1}}{2\Delta \theta r_i} D_\theta g_{i-1,j} + \frac{1}{2}((a + |a|) D_\theta f_{i,j} + (a - |a|) D_\theta f_{i,j-1}).
\]
(3.5)

In regards to (3.5), when \( i = 1, \gamma_3 \) is reduced to a point, so \( \frac{r_i + \frac{1}{2}M_{i+1,j}}{2\Delta r r_i} D_r g_{0,j} \) is understood as 0; when \( i = P, \) the zero flux gives that \( \frac{r_P + \frac{1}{2}M_{P+1,j}}{2\Delta r r_P} D_r g_{P,j} = 0. \) Due to the periodicity of \( f \) and \( M \) with respect to \( \theta, \) we take
\[
f_{i,j} = f_{i,j+Q}, \quad M_{i,j} = M_{i,j+Q} \quad \text{for } 1 \leq j \leq Q.
\]
Thus (3.5) is well defined for \( 1 \leq i \leq P, 1 \leq j \leq Q, \) which can be solved subject to the initial data
\[
f_{i,j}(0) = \frac{1}{|K_{ij}|} \int_{K_{ij}} f_0(m) \, dm.
\]

**Theorem 3.2.** Let \( \{ij\} = \{1 \leq i \leq P, 1 \leq j \leq Q\}. \) The semidiscrete scheme (3.5) has the following properties:
1. \( \sum_{ij} f_{i,j}(t)|K_{ij}| = \sum_{ij} f_{i,j}(0)|K_{ij}| = \int_{B} f_0(m) \, dm. \)
2. Positivity. If \( f_{i,j}(0) \geq 0, \) then \( f_{i,j}(t) \geq 0, \quad t > 0. \)
3. The semidiscrete relative entropy, defined by
\[
E(t) = \sum_{ij} \frac{f_{ij}^2(t)}{M_{ij}} |K_{ij}|
\]
satisfies
\[
E(t) \leq E(0), \quad t > 0,
\]
for normal \( K \) and \( E(t) \leq e^{ct} E(0) \) for general \( K \) with \( c > 0 \) dependent on \( b. \)
Proof. (1) Summation of $\frac{d}{dt} f_{i,j}(t)|K_{ij}|$ over $\{ij\}$ in (3.5) gives
\[
\frac{d}{dt} \sum_{ij} f_{i,j}(t)|K_{ij}| = \sum_{ij} \frac{d}{dt} f_{i,j}(t) \Delta \theta \Delta r_{i} = 0.
\]
So
\[
\sum_{ij} f_{i,j}(t)|K_{ij}| = \sum_{ij} f_{i,j}(0)|K_{ij}| = \int_{B} f_{0}(m) \, dm \quad \forall t > 0.
\]

(2) We arrange the solution $\{g_{i,j}\}$ of (3.5) to be a vector $\vec{g} = (g_{1,1}, g_{2,1}, \ldots, g_{p,1}, g_{1,2}, \ldots, g_{p,Q})^T$ and rewrite (3.5) into the vector form
\[
(3.6) \quad \frac{d}{dt} \vec{g} = \vec{F}(\vec{g}).
\]
Due to the mass conservation, we have
\[
\sum_{ij} M_{ij} g_{i,j}|K_{ij}| = \sum_{ij} f_{i,j}(0)|K_{ij}| = \int_{B} f_{0}(m) \, dm \quad \forall t > 0.
\]
Then all the trajectories of (3.6) remain on this hyperplane. We define a closed set $\Sigma$ on this hyperplane by
\[
\Sigma = \left\{ \vec{g} : g_{i,j} \geq 0, (i, j) \in \{ij\}, \text{ and } \sum_{ij} M_{ij} g_{i,j}|K_{ij}| = \int_{B} f_{0}(m) \, dm \right\}.
\]
It suffices to show that $\Sigma$ is an invariant region of the ODE system (3.6). Similar to the argument in the one-dimensional case explored previously, we only need to prove that for any outward normal vector $\vec{n}$ on any part of the boundary of $\Sigma$,
\[
\vec{F}(\vec{g}) \cdot \vec{n} < 0, \quad \vec{g} \in \partial \Sigma.
\]
For each $\vec{g} \in \partial \Sigma$, we define the set of indices $S$ such that
\[
S = \{(i, j) \in \{ij\} : g_{i,j} = 0\},
\]
which implies that $S \neq \emptyset$ for any $\vec{g} \in \partial \Sigma$. Then the outward normal vectors $\vec{n}$ at $\vec{g} \in \partial \Sigma$ are of the form
\[
\vec{n} = (n_{ij}) \quad \text{with} \quad n_{ij} = \begin{cases} -|K_{ij}| \alpha_{ij} & \text{if } (i, j) \in S, \\ |K_{ij}|M_{ij} & \text{if } (i, j) \notin S. \end{cases}
\]
Furthermore, $\alpha_{ij} > 0$ for $(i, j) \in S$.

Here we assume $a \geq 0$. The proof is similar for the case of $a < 0$. Shifting the indices, we have
\[
\vec{F}(\vec{g}) \cdot \vec{n} = -\sum_{ij} \Delta \theta r_{i+\frac{1}{2}} M_{i+\frac{1}{2}}^{r_{i}} D_{r} g_{i,j} \left( \frac{n_{i+1,j}}{|K_{i+1,j}|M_{i+1,j}} - \frac{n_{ij}}{|K_{ij}|M_{ij}} \right)
\]
\[
- \sum_{ij} \frac{\Delta r M_{i+\frac{1}{2}}^{r_{i}}}{2r_{i}} D_{\theta} g_{i,j} \left( \frac{n_{i,j+1}}{|K_{i,j+1}|M_{i,j+1}} - \frac{n_{ij}}{|K_{ij}|M_{ij}} \right)
\]
\[
- \sum_{ij} a \Delta rr_{i} f_{i,j} \left( \frac{n_{ij}}{|K_{ij}|M_{ij}} - \frac{n_{i,j-1}}{|K_{i,j-1}|M_{i,j-1}} \right)
\]
\[
= - I - II - III.
\]
We only analyze I, since the discussion about II and III is analogous. If \((i, j), (i + 1, j) \in S\), then \(g_{i,j} = g_{i+1,j} = 0\), implying \(D_r g_{i,j} = 0\); if \((i, j), (i, j + 1) \notin S\), then \(\frac{|K_{i+1,j}M_{i+1,j}}{|K_{i,j}M_{i,j}}| = 1 - 1 = 0\). Hence the nonzero terms in I are only those with \((i, j) \in S, (i + 1, j) \notin S\) or \((i, j) \notin S, (i + 1, j) \in S\). Therefore

\[
I = \sum_{(i,j) \in S, (i+1,j) \notin S} \frac{\Delta t r_{\frac{i+\frac{k}{2},j}} M_{i+\frac{k}{2},j}}{2} D_r g_{i,j} \left( 1 + \frac{\alpha_{i,j}}{M_{i,j}} \right) \\
+ \sum_{(i,j) \notin S, (i+1,j) \in S} \frac{\Delta t r_{\frac{i+\frac{k}{2},j}} M_{i+\frac{k}{2},j}}{2} D_r g_{i,j} \left( -\frac{\alpha_{i+1,j}}{M_{i+1,j}} - 1 \right).
\]

By the definition of \(S\), \(g_{i,j} = 0\) in the first summation and \(g_{i+1,j} = 0\) in the second summation. So

\[
I = \sum_{(i,j) \in S, (i+1,j) \notin S} \frac{\Delta t r_{\frac{i+\frac{k}{2},j}} M_{i+\frac{k}{2},j}}{2} g_{i+1,j} \left( 1 + \frac{\alpha_{i,j}}{M_{i,j}} \right) \\
+ \sum_{(i,j) \notin S, (i+1,j) \in S} \frac{\Delta t r_{\frac{i+\frac{k}{2},j}} M_{i+\frac{k}{2},j}}{2} (-g_{i,j}) \left( -\frac{\alpha_{i+1,j}}{M_{i+1,j}} - 1 \right) > 0.
\]

With this, \(II > 0\), and \(III > 0\), we arrive at the conclusion that \(\bar{F}(\bar{g}) \cdot \bar{n} < 0\).

(3) Next, we show that \(E(t)\) remains bounded for any \(t > 0\). For definiteness, we assume \(a > 0\).

\[
\frac{d}{dt} E(t) = \sum_{ij} 2 f_{i,j} \frac{df_{i,j}}{dt} |K_{i,j}| = \sum_{ij} 2 g_{i,j} \frac{df_{i,j}}{dt} \Delta \theta \Delta r r_i \\
= \Delta \theta \sum_{ij} g_{i,j} \left( r_{i+\frac{k}{2}} M_{i+\frac{k}{2},j} D_r g_{i,j} - r_{i-\frac{k}{2}} M_{i-\frac{k}{2},j} D_r g_{i-1,j} \right) \\
+ \Delta r \sum_{ij} g_{i,j} \left( M_{i,j+\frac{k}{2}} D_r g_{i,j} - M_{i,j-\frac{k}{2}} D_r g_{i,j-1} \right) \\
+ 2a \Delta r \sum_{ij} g_{i,j} r_i \left( f_{i,j+1} - f_{i,j} \right) \\
= \Delta \theta I + \Delta r II + 2a \Delta r III.
\]

By shifting the indices in \(i\) and using \(r_{\frac{k}{2}} = 0, M_{\frac{k}{2},j} = 0\), we have

\[
I = -\sum_{1 \leq i \leq P-1, 1 \leq j \leq Q} \Delta r r_{i+\frac{k}{2}} M_{i+\frac{k}{2},j} (D_r g_{i,j})^2 = -\Delta r \sum_{ij} r_{i+\frac{k}{2}} M_{i+\frac{k}{2},j} (D_r g_{i,j})^2.
\]
Similarly, shifting the indices in \( j \) gives
\[
\Pi = - \sum_{1 \leq i \leq p} \frac{\Delta \theta M_{i,j+1}}{r_i} (D_{\theta} g_{i,j})^2 + \sum_{1 \leq i \leq p} \frac{M_{i,Q+1}}{r_i} g_{i,Q} D_{\theta} g_{i,Q}
- \sum_{1 \leq i \leq p} \frac{M_{i,j+1}}{r_i} g_{i,1} D_{\theta} g_{i,0}
= - \Delta \theta \sum_{i,j} \frac{M_{i,j+1}}{r_i} (D_{\theta} g_{i,j})^2.
\]
Here we have used \( M_{i,j} = M_{i,Q+1,j} \), \( g_{i,1} = g_{i,Q+1} \), and \( g_{i,0} = g_{i,Q} \).

Summation by parts in \( j \) gives
\[
\Pi = - \sum_{ij} (g_{i,j+1} - g_{i,j}) r_i f_{i,j+1}
= - \sum_{ij} r_i M_{i,j+1} g_{i,j+1} (g_{i,j+1} - g_{i,j})
= - \frac{1}{2} \sum_{ij} r_i M_{i,j+1} (g_{i,j+1} - g_{i,j})^2 + \frac{1}{2} \sum_{ij} r_i g_{ij}^2 (M_{i,j+1} - M_{i,j}).
\]

In the two-dimensional case, \( K \) is normal if and only if \( K \) is either symmetric, i.e., \( a = 0 \), or antisymmetric, i.e., \( M_{i,j+1} = M_{i,j} \). In either case we have
\[
2a \Delta r \Pi \leq 0,
\]
hence \( \frac{d}{dt} E(t) \leq 0 \).

For general matrix \( K \), we have
\[
\frac{d}{dt} E(t) \leq -D(t) + a \Delta r \sum_{ij} r_i g_{ij}^2 (M_{i,j+1} - M_{i,j}),
\]
where
\[
D(t) = \Delta \theta \Delta r \sum_{ij} r_i M_{i,j} + \frac{1}{2} (D_{\theta} g_{i,j})^2 + \Delta \theta \Delta r \sum_{ij} \frac{M_{i,j+1}}{r_i} (D_{\theta} g_{i,j})^2.
\]
For \( A = \max_{ij} \frac{|M_{i,j+1} - M_{i,j}|}{\Delta \theta M_{i,j}} \),
\[
\frac{d}{dt} E(t) \leq -D(t) + \frac{aA}{\beta} E(t).
\]
By Gronwall’s inequality,
\[
E(t) \leq e^{\frac{aA}{\beta}} E(0) - \int_0^t D(\tau) e^{\frac{aA}{\beta}(t-\tau)} d\tau.
\]

**Theorem 3.3.** If \( K \) is normal, then the physical entropy \( E_p(t) = \sum_{ij} f_{i,j} \log(g_{i,j}) \) \(|K_{ij}|\) is nonincreasing in time, satisfying
\[
\frac{d}{dt} E_p(t) \leq - \frac{\Delta \theta}{2} \sum_{ij} r_i M_{i,j} \log \left( \frac{g_{i,j+1}}{g_{i,j}} \right) D_{\theta} g_{i,j}
- \frac{\Delta \theta}{2} \sum_{ij} \frac{M_{i,j+1}}{r_i} \log \left( \frac{g_{i,j+1}}{g_{i,j}} \right) D_{\theta} g_{i,j} \leq 0.
\]
Proof. By mass conservation we obtain

\[
\frac{d}{dt} E_p(t) = \sum_{i,j} \frac{d}{dt} f_{i,j}(\log g_{i,j} + 1)|K_{ij}|
\]

\[
= \sum_{i,j} \frac{d}{dt} f_{i,j} \log g_{i,j}|K_{ij}| =: I + II + III,
\]

where, in virtue of the monotonicity property of the log function,

\[
I = \sum_{i,j} \left( \frac{r_{i+1} M_{i+\frac{1}{2},j}}{2\Delta rr_i} D_r g_{i,j} - \frac{r_{i-\frac{1}{2}} M_{i-\frac{1}{2},j}}{2\Delta rr_i} D_r g_{i-1,j} \right) \log g_{i,j}|K_{ij}|
\]

\[
= -\frac{\Delta \theta}{2} \sum_{i,j} r_{i+\frac{1}{2}} M_{i+\frac{1}{2},j} D_r g_{i,j} (\log g_{i+1,j} - \log g_{i,j}) \leq 0,
\]

\[
II = \sum_{i,j} \left( \frac{M_{i,j+\frac{1}{2}}}{2\Delta \theta r_i^2} D_\theta g_{i,j} - \frac{M_{i,j-\frac{1}{2}}}{2\Delta \theta r_i^2} D_\theta g_{i,j-1} \right) \log g_{i,j}|K_{ij}|
\]

\[
= -\frac{\Delta \theta}{2} \sum_{i,j} M_{i,j+\frac{1}{2}} D_\theta g_{i,j} (\log g_{i,j+1} - \log g_{i,j}) \leq 0
\]

and III = \( \sum_{i,j} a D_\theta f_{i,j} \log g_{i,j}|K_{ij}| \), which corresponds to the case \( a \geq 0 \). The case \( a < 0 \) can be treated in a similar fashion. By summation by parts in \( j \) we have

\[
III = \sum_{i,j} a \Delta rr_i (f_{i,j+1} - f_{i,j}) \log g_{i,j}
\]

\[
= \sum_{i,j} a \Delta rr_i f_{i,j} \log \frac{g_{i,j+1}}{g_{i,j}}.
\]

If \( K \) is symmetric, then III = 0. If \( K \) is antisymmetric, then \( M_{i,j} = M_{i,j-1} \). Therefore,

\[
III = \sum_{i,j} a \Delta rr_i f_{i,j} \log \frac{f_{i,j+1}}{f_{i,j}}
\]

\[
\leq \sum_{i,j} a \Delta rr_i f_{i,j} \left( \frac{f_{i,j+1}}{f_{i,j}} - 1 \right)
\]

\[
= \frac{a}{\Delta \theta} \sum_{i,j} |K_{ij}|(f_{i,j+1} - f_{i,j}) = 0
\]

due to the conservation of mass. \( \square \)

Similar to the one-dimensional case we can show the long time convergence of solutions of the semidiscrete system.

Theorem 3.4. Consider the semidiscrete scheme (3.5) subject to the initial data \( f_{i,j}(0) > 0 \) with \( \sum f_{i,j}(0)|K_{ij}| = \int_B f_0(m) \, dm \). If \( K \) is normal, then

\[
f_{i,j}(t) \to CM_{i,j}, \quad t \to \infty,
\]

where

\[
C = \frac{\int_B f_0(m) \, dm}{\sum_{i,j} M_{i,j} |K_{ij}|}.
\]
Proof. Define a functional $V(g)$ by

$$V(g) = \sum_{ij} (g_{i,j} - C)^2 M_{i,j} |K_{ij}|$$

and $\bar{g}_{i,j} = C$. We see that $\bar{g} \in \Sigma$ with

$$\Sigma := \left\{ g, \quad g_{i,j} \geq 0, \quad \sum_{ij} g_{i,j} M_{i,j} |K_{ij}| = \int_B f_0(m) dm \right\}$$

is the equilibrium solution. A direct verification shows that $V(g) = E(t) - C \int_B f_0(m) dm$, implying that $\frac{d}{dt} V = \frac{d}{dt} E$. Hence $V$ satisfies the following:

- $V(g) > 0$ for any $g \neq \bar{g}$ (positive definite).
- $\frac{d}{dt} V \leq 0$ for all $g$ (negative semidefinite).
- The set $\{ \frac{d}{dt} V = 0 \} \cap \Sigma$ does not contain any trajectories of the ODE system besides the trajectory $g(t) = \bar{g} \forall t > 0$.

With these properties we can apply the Krasovskii–LaSalle principle to conclude that $\lim_{t \to \infty} g(t) = \bar{g}$, which leads to the conclusion. We only verify the third property of $V$: from (3) of Theorem 3.2 it follows that

$$\frac{d}{dt} V = - \sum_{ij} |K_{ij}| \frac{r_{i+1/2} M_{i+1/2,j}}{r_i} (D_r g_{i,j})^2 - \sum_{ij} |K_{ij}| \frac{r_{i+1/2} M_{i,j+1/2}}{r_i} (D_\theta g_{i,j})^2$$

$$- a \Delta r \sum_{ij} r_i M_{i,j+1} (g_{i,j+1} - g_{i,j})^2.$$ 

If $\frac{d}{dt} V = 0$, then each term in the sum on the right side must vanish, that is,

$$D_r g_{i,j} = 0, \quad D_\theta g_{i,j} = 0,$$

which ensures that $g_{i,j} = \text{const}$, while within $\Sigma$, $g = \bar{g}$ must hold. The proof is thus complete. $\square$

3.3. Time discretization. We apply the backward Euler method to (3.5), but treating the asymmetric part explicitly,

$$\frac{f_{i,j}^{n+1} - f_{i,j}^n}{\Delta t} = \frac{r_i + \frac{1}{2} M_{i,j}}{2 \Delta r r_i} D_r g_{i,j}^{n+1} - \frac{r_i - \frac{1}{2} M_{i,j}}{2 \Delta r r_i} D_r g_{i,j}^{n-1}$$

$$+ \frac{M_{i,j}}{2 \Delta \theta r_i} D_\theta g_{i,j}^{n+1} - \frac{M_{i,j}}{2 \Delta \theta r_i} D_\theta g_{i,j}^{n-1}$$

$$+ \frac{1}{2} [a + |a|] D_\theta f_{i,j}^n + (a - |a|) D_\theta f_{i,j}^{n-1},$$

with $f_{i,j}^0 = f_{i,j}(0)$. We assume that $\Delta t$ satisfies the CFL condition

$$\frac{|a| \Delta t}{\Delta \theta} \leq 1.$$

Theorem 3.5. The discrete scheme (3.9) with (3.10) satisfies the following properties:

1. $\sum_{ij} f_{i,j}^n |K_{ij}| = \sum_{ij} f_{i,j}^0 |K_{ij}| \forall n \in \mathbb{N}$. 

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
(2) If the initial data $f_{i,j}^0 \geq 0$, then $f_{i,j}^n \geq 0 \ \forall n \in \mathbb{N}$.
(3) The discrete relative entropy
\[ E^n = \sum_{ij} \left( \frac{(f_{i,j}^n)^2}{M_{i,j}} \right) |K_{ij}| \]
satisfies $E^{n+1} \leq E^n$ for $a = 0$.
(4) $f_{i,j}^n$ converges as $n \to \infty$ with
\[ f_{i,j}^n \to CM_{i,j}, \quad n \to \infty, \]
where $C$ is defined in (3.8).

Proof. (1) Multiply (3.9) by $|K_{ij}|$ and sum over $\{ij\}$ so that
\[ \frac{1}{\Delta t} \left( \sum_{ij} f_{i,j}^{n+1} |K_{ij}| - \sum_{ij} f_{i,j}^n |K_{ij}| \right) = 0. \]
Therefore
\[ \sum_{ij} f_{i,j}^{n+1} |K_{ij}| = \sum_{ij} f_{i,j}^n |K_{ij}| = \cdots = \sum_{ij} f_{i,j}^0 |K_{ij}|. \]

(2) Rewrite the scheme (3.9) in terms of $g_{i,j}^n$ as follows:
\[
\begin{align*}
& - \Delta t \frac{r_i + \frac{1}{2} M_i - \frac{1}{2} r_j}{2(\Delta r)^2 r_i} g_{i,j+1}^{n+1} - \Delta t \frac{r_i + \frac{1}{2} M_i + \frac{1}{2} j - \frac{1}{2} r_j}{2(\Delta r)^2 r_i} g_{i,j+1}^{n+1} - \Delta t \frac{r_i - \frac{1}{2} M_i + \frac{1}{2} j - \frac{1}{2} r_j}{2(\Delta r)^2 r_i} g_{i,j-1}^{n+1} \\
& - \Delta t \frac{M_i + \frac{1}{2} j - \frac{1}{2} r_j}{2(\Delta r)^2 r_i} g_{i,j+1}^{n+1} + (M_{i,j} - (\cdots))g_{i,j+1}^{n+1} = - \frac{(a - |a|) \Delta t M_{i,j-1}}{2 \Delta \theta} g_{i,j-1}^{n+1} \\
& + \left( 1 - \frac{|a| \Delta t}{\Delta \theta} \right) M_{i,j} g_{i,j}^{n+1} + \frac{(a + |a|) \Delta t M_{i,j+1}}{2 \Delta \theta} g_{i,j+1}^{n+1},
\end{align*}
\]
where $(\cdots)$ is the sum of the coefficients of the first four terms on the left-hand side.
The CFL condition (3.10) ensures that the right-hand side of (3.11) is nonnegative.
Note that the coefficient matrix of (3.11) is diagonally dominated. A similar argument
to that in the one-dimensional case can be applied here to prove that $\{g_{i,j}^{n+1}\}$ are
nonnegative. It follows that $\{f_{i,j}^{n+1}\}$ are nonnegative.

(3) We calculate the change of entropy in one time step,
\[
\begin{align*}
E^{n+1} - E^n = & \sum_{ij} \frac{(2f_{i,j}^{n+1} - f_{i,j}^n)(f_{i,j}^{n+1} - f_{i,j}^n)}{M_{i,j}} |K_{ij}| \\
= & 2 \sum_{ij} g_{i,j}^{n+1} (f_{i,j}^{n+1} - f_{i,j}^n) |K_{ij}| - \sum_{ij} \frac{(f_{i,j}^{n+1} - f_{i,j}^n)^2}{M_{i,j}} |K_{ij}| \\
= & 2(I + II) - \sum_{ij} \frac{(f_{i,j}^{n+1} - f_{i,j}^n)^2}{M_{i,j}} |K_{ij}|,
\end{align*}
\]
where

\[
I = \sum_{ij} g_{i,j}^{n+1} \left( \frac{r_{i+\frac{1}{2}} M_{i+\frac{1}{2},j} D_r g_{i,j}^{n+1} - r_{i-\frac{1}{2}} M_{i-\frac{1}{2},j} D_r g_{i,j}^{n+1}}{2\Delta r_{i,j}} \right) |K_{ij}| \Delta t
\]

\[
= \frac{\Delta \theta \Delta t}{2} \left( \sum_{ij} g_{i,j}^{n+1} r_{i+\frac{1}{2}} M_{i+\frac{1}{2},j} D_r g_{i,j}^{n+1} - \sum_{ij} g_{i,j}^{n+1} r_{i-\frac{1}{2}} M_{i-\frac{1}{2},j} D_r g_{i,j}^{n+1} \right)
\]

\[
= -\frac{\Delta \theta \Delta r \Delta t}{2} \sum_{ij} r_{i+\frac{1}{2}} M_{i+\frac{1}{2},j} (D_r g_{i,j}^{n+1})^2 \leq 0,
\]

and similarly, by shifting the index in \( j \), we have

\[
II = \sum_{ij} g_{i,j}^{n+1} \left( \frac{M_{i,j} + \frac{1}{2} D_{\theta} g_{i,j}^{n+1} - M_{i,j} + \frac{1}{2} D_{\theta} g_{i,j}^{n+1}}{2\Delta \theta r_{i,j}} \right) |K_{ij}| \Delta t
\]

\[
= -\frac{\Delta \theta \Delta r \Delta t}{2} \sum_{ij} M_{i,j} + \frac{1}{2} (D_{\theta} g_{i,j}^{n+1})^2 \leq 0.
\]

So \( E^{n+1} \leq E^n \).

(4) Since \( E^n \) is nonincreasing and bounded from below, we have

\[
\lim_{n \to \infty} E^n = \inf \{ E^n \}.
\]

Observe from analysis of (3) that \( E^n - E^{n+1} \) is a sum of nonnegative, bounded terms. When passing limit \( n \to \infty \) we conclude that each term must have zero as its limit, that is

\[\text{(3.12)} \quad \lim_{n \to \infty} (f_{i,j}^{n+1} - f_{i,j}^n)^2 = 0, \quad \lim_{n \to \infty} [(D_{\theta} g_{i,j}^{n+1})^2 + (D_r g_{i,j}^{n+1})^2] = 0.\]

The first relation in (3.12) says that \( \tilde{g}^n \) is a Cauchy sequence, which when combined with the completeness of \( \Sigma \) (a closed and bounded set in \( \mathbb{R}^{PQ} \)) ensures that \( \lim_{n \to \infty} \tilde{g}^n \) exists. The second relation in (3.12) infers that the limit must be \( \bar{g} \). The proof is complete. \( \Box \)

**Theorem 3.6.** For symmetric \( K \), the physical entropy

\[ E_p^n = \sum_{ij} f_{i,j}^n \log \left( \frac{f_{i,j}^n}{M_{i,j}} \right) |K_{ij}| \]

is nonincreasing. Moreover,

\[ E_p^{n+1} - E_p^n = -\sum_{ij} \log (g_{i,j}^{n+1}) (g_{i,j}^{n+1} - g_{i,j}^n) M_{i,j} |K_{ij}| + \sum_{ij} f_{i,j}^n \log \left( \frac{f_{i,j}^{n+1}}{f_{i,j}^n} \right) |K_{ij}| \leq 0. \]

**Proof.** We calculate

\[ E_p^{n+1} - E_p^n = \sum_{ij} \left( f_{i,j}^{n+1} \log f_{i,j}^{n+1} M_{i,j} - f_{i,j}^n \log f_{i,j}^n M_{i,j} + f_{i,j}^n \log f_{i,j}^{n+1} M_{i,j} - f_{i,j}^n \log f_{i,j}^n M_{i,j} \right) |K_{ij}| \]

\[ = \sum_{ij} \left[ (f_{i,j}^{n+1} - f_{i,j}^n) \log g_{i,j}^{n+1} + f_{i,j}^n \log \frac{f_{i,j}^{n+1}}{f_{i,j}^n} \right] |K_{ij}|. \]
The second sum is nonpositive since
\[
\sum_{ij} f^t_{i,j} \log \left( \frac{f^n_{i,j}}{f^t_{i,j}} \right) |K_{ij}| \leq \sum_{ij} f^n_{i,j} \left( \frac{f^n_{i,j}}{f^t_{i,j}} - 1 \right) |K_{ij}| = 0
\]
for mass is conserved at each time step. The first sum when recalling the fully discrete scheme may be expressed as I + II with
\[
\begin{align*}
I &= \sum_{ij} \Delta t \left( \frac{r_i + \frac{1}{2} M_{i,j} + \frac{1}{2} j}{2 \Delta r r_i} D_r g_{i,j}^{n+1} - \frac{r_i - \frac{1}{2} M_{i,j} - \frac{1}{2} j}{2 \Delta r r_i} D_r g_{i,j}^{n+1} \right) \log g_{i,j}^{n+1} |K_{ij}| \\
&= -\frac{\Delta t \Delta \theta}{2} \sum_{ij} r_i + \frac{1}{2} M_{i,j} + \frac{1}{2} j D_r g_{i,j}^{n+1} \left( \log g_{i,j}^{n+1} - \log g_{i,j}^{n+1} \right) \leq 0
\end{align*}
\]
and
\[
\begin{align*}
II &= \sum_{ij} \Delta t \left( \frac{M_{i,j} + \frac{1}{2} j}{2 \Delta \theta r_i^2} D_{\theta} g_{i,j}^{n+1} - \frac{M_{i,j} - \frac{1}{2} j}{2 \Delta \theta r_i^2} D_{\theta} g_{i,j}^{n+1} \right) \log g_{i,j}^{n+1} |K_{ij}| \\
&= -\frac{\Delta t \Delta \theta}{2} \sum_{ij} M_{i,j} + \frac{1}{2} j D_{\theta} g_{i,j}^{n+1} \left( \log g_{i,j}^{n+1} - \log g_{i,j}^{n+1} \right) \leq 0.
\end{align*}
\]
These together make the proof complete. \( \square \)

Remark 3.2. If the drift term corresponding to the antisymmetric part is made implicit in time discretization, the entropy dissipation relations also hold at the fully discrete level when \( K \) is normal (for both the physical entropy and the quadratic entropy). But such an implicit treatment does not guarantee the positivity preserving property.

4. Numerical implementation and results: Implementation strategies. For the one-dimensional case, we apply the tridiagonal matrix algorithm (also known as the Thomas algorithm) to scheme (2.12). The computation cost is \( O(N) \).

For the two-dimensional case, we use a direct method to solve the linear system \( Ax = b \) with a sparse \( N \times N \) coefficient matrix with \( N = PQ \). If the final time \( t \) is a multiple of the time step \( \Delta t \), the coefficient matrix is the same for each time step. So we only need to compute the LU decomposition once. Furthermore, for large \( N \), the sparsity of the coefficient matrix reduces the complexity significantly, which is about \( O(P^3Q) \). Solving the decomposed system \( LUx = b \) costs \( O(N^2) \). So the total complexity is \( O(N^2) \).

For \( K = 0 \) or antisymmetric, \( M \) is independent of \( \theta \), and we use the Fourier method in \( \theta \) to reduce the computational cost. More precisely, we express the solution as
\[
g_{i,j} = \sum_{l=1}^{Q} \hat{g}_{l} \hat{e}^{-i(j-1)(l-1) \Delta \theta}, \quad \hat{\theta} = \sqrt{-1},
\]
with its inverse
\[
\hat{g}_{l} = \frac{1}{Q} \sum_{j=1}^{Q} g_{i,j} \hat{e}^{i(j-1)(l-1) \Delta \theta}.
\]
For each \( l \), we obtain a linear system of \( (\hat{g}_{1,l}^{n+1}, \ldots, \hat{g}_{P,l}^{n+1})^T \). The Fourier transform and the inverse Fourier transform need \( O(PQ^2) \) operations. And for each time step,
the computational cost of solving $Q$ linear systems is $O(QP)$, since they all have a tri-diagonal coefficient matrix. So the total complexity is $O(P^2)$, which with complexity $O(N^{1.5})$ is clearly faster than the direct solver described above.

**Numerical tests.** We now present our numerical results to demonstrate (i) the accuracy of the schemes, (ii) the capacity to capture equilibrium solutions and the large time behavior of the solution, and (iii) the effects of some typical homogeneous flows.

Denote the initial function without normalization by $\tilde{f}_0(m)$ and the normalized initial data by $f_0(m) = Z^{-1}\tilde{f}_0(m)$, where $Z$ is a normalization factor defined by

$$Z = \int_B \tilde{f}_0(m) \, dm.$$ 

We also denote $Z_M = \int_B M(m) \, dm$.

**4.1. One-dimensional tests.** Denote the numerical solution by $f^n_j$, and the exact solution by $f(m_j, t^n)$.

**Definition 1.** The $L_1$ error is given by

$$\sum_{j=1}^N |f^n_j - f(m_j, t^n)|h,$$

and the $L_\infty$ error is given by

$$\max_{1 \leq j \leq N} |f^n_j - f(m_j, t^n)|.$$

When the exact solution is not available, we replace $f(m_j, t^n)$ by a reference solution to compute the errors.

**4.1.1. Accuracy.** We illustrate the accuracy of scheme (2.12) with several choices of initial data.

**Example 1.** In this example, we consider four kinds of initial data:

(i) $f_0(m) = (b - m^2)^{\alpha b}, \alpha = \frac{1}{4}, \frac{1}{2}, \frac{3}{2}$,

(ii) the distance function $f_0(m) = \sqrt{b} - |m|$, 

(iii) the characteristic function $f_0(m) = \chi_{[-\sqrt{b+\varepsilon}, \sqrt{b-\varepsilon}]}$, $0 < \varepsilon < \sqrt{b}$, and

(iv) a cosine function $f_0(m) = 1 + \cos \left( \frac{2m\pi}{\sqrt{b-\varepsilon}} + \pi \right)$, $0 < \varepsilon < \sqrt{b}$.

We take the numerical solution with $N = 2560$ as the reference solution. Table 1 shows the results from the above initial data when $b = 16$.

**Example 2.** We consider the same initial data as in Example 1 but with $b = 50$. The results are given in Table 2.

**4.1.2. Large time behavior.** The normalized equilibrium solution of the Fokker–Planck equation is

$$f_{eq}(m) = Z_M^{-1} M(m).$$

We define the distance of the solution from the equilibrium as

$$\max_{1 \leq j \leq N} |f^n_j - f_{eq}(m_j)|.$$ 

**Example 3.** Take (iv) in Example 1 as the initial data, and let $b = 16, \varepsilon = 0.01 \sqrt{b}$. The numerical solutions at $t = 0, 1.0, 1.8$ are plotted in Figure 2, which indicate a fast convergence to the equilibrium state. In Table 3 we see that the distance from the equilibrium solution is decreasing. This confirms that the solution converges to the equilibrium solution $f_{eq}$ as time increases.
### Table 1
Error and order of accuracy for Example 1 on a uniform mesh of \( N \) cells: \( b = 16, \Delta t = 0.1 \), final time \( t = 1.8 \).

<table>
<thead>
<tr>
<th>( f_0(m) )</th>
<th>( (b - m^2) \sqrt{\xi} )</th>
<th>( (b - m^2) \sqrt{\eta} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( N )</td>
<td>( L_1 ) error</td>
<td>Order</td>
</tr>
<tr>
<td>20</td>
<td>8.917E-02</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>3.999E-02</td>
<td>1.003</td>
</tr>
<tr>
<td>80</td>
<td>1.998E-02</td>
<td>1.001</td>
</tr>
<tr>
<td>160</td>
<td>9.992E-03</td>
<td>1.000</td>
</tr>
<tr>
<td>320</td>
<td>4.996E-03</td>
<td>1.000</td>
</tr>
<tr>
<td>640</td>
<td>2.498E-03</td>
<td>1.000</td>
</tr>
</tbody>
</table>

### Table 2
Error and order of accuracy for Example 2 on a uniform mesh of \( N \) cells: \( b = 50, \Delta t = 0.1 \), final time \( t = 1.8 \).

| \( f_0(m) \) | \( (b - m^2) \sqrt{\xi} \) | \( \sqrt{\xi} - |m| \) |
|---|---|---|
| \( N \) | \( L_1 \) error | Order | \( L_\infty \) error | Order | \( L_1 \) error | Order | \( L_\infty \) error | Order |
| 20 | 6.805E-02 | | 3.991E-02 | | 6.538E-02 | | 4.011E-02 | |
| 40 | 3.419E-02 | 1.009 | 2.023E-02 | 0.981 | 3.263E-02 | 1.002 | 1.998E-02 | 1.006 |
| 80 | 1.704E-02 | 1.004 | 9.870E-03 | 1.035 | 1.630E-02 | 1.001 | 9.875E-03 | 1.013 |
| 160 | 8.517E-03 | 1.001 | 4.725E-03 | 1.051 | 8.152E-03 | 1.000 | 4.768E-03 | 1.048 |
| 320 | 4.256E-03 | 1.000 | 2.218E-03 | 1.102 | 4.076E-03 | 1.000 | 2.234E-03 | 1.100 |
| 640 | 2.128E-03 | 1.000 | 9.469E-04 | 1.224 | 2.038E-03 | 1.000 | 9.564E-04 | 1.223 |

### Note
\( L_1 \) and \( L_\infty \) errors are computed as

\[
\frac{\| f_0(m) - X_{\eta_1} \|_{L_1}}{\| f_0(m) \|_{L_1}} \quad \text{and} \quad \frac{\| f_0(m) - X_{\eta_1} \|_{L_\infty}}{\| f_0(m) \|_{L_\infty}}
\]

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
Fig. 2. $\tilde{f}_0(m) = 1 + \cos \left( \frac{2m\pi}{\sqrt{b-\varepsilon}} + \pi \right)$ with $b = 16, \varepsilon = 0.01\sqrt{b}$.

Table 3
Numerical convergence to the equilibrium solution measured by distances for Example 3: $b = 16$, $\varepsilon = 0.01\sqrt{b}$, and $N = 160$.

<table>
<thead>
<tr>
<th>$t$</th>
<th>1</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tilde{f}_0(m)$</td>
<td>2.0962E-01</td>
<td>1.6390E-02</td>
<td>4.2450E-03</td>
<td>1.0971E-03</td>
<td>2.8348E-04</td>
</tr>
</tbody>
</table>

Table 4
Relative entropy in Example 4: $b = 16$ and $N = 640$, $\Delta t = 0.1$.

| $t$ | $(b - m^2)^{1/2}$ | $(b - m^2)^{1/2}$ | $\sqrt{b} - |m|$ | $1 + \cos \left( \frac{2m\pi}{\sqrt{b-\varepsilon}} + \pi \right)$ | $\varepsilon = 0.1\sqrt{b}$ | $\varepsilon = 0.01\sqrt{b}$ |
|-----|------------------|------------------|----------------|-------------------------------------------------|--------------------------|--------------------------|
| 0   | 1.8141           | 1.0129E+12       | 1.0129E+12     | 3346.32                                         | 5280.76                  | 5280.76                  |
| 0.2 | 1.1704           | 8.3964           | 35.0358        | 14.8988                                         | 3.0020                   | 3.0020                   |
| 0.6 | 1.0575           | 2.9207           | 1.4865         | 1.6093                                          | 1.2111                   | 1.2111                   |
| 1.0 | 1.0204           | 1.435            | 1.435          | 1.435                                           | 1.435                    | 1.435                    |
| 1.4 | 1.0074           | 1.0488           | 1.0488         | 1.0488                                          | 1.0488                   | 1.0488                   |
| 1.8 | 1.0027           | 1.0174           | 1.0174         | 1.0174                                          | 1.0174                   | 1.0174                   |

4.1.3. Relative entropy. Now we test the relative entropy of the numerical solutions. The scaled discrete entropy is defined as

$$\sum_{j=1}^{N} \frac{(f_{ij}^n)^2}{Z_{M}^{-1}M_{j}} h_i.$$ 

Example 4. We test the time evolution of the relative entropy by using the initial data (i)–(iv) from Example 1. Table 4 shows that the relative entropy is nonincreasing.

4.2. Two-dimensional tests. Denote the numerical solution by $f_{ij}^n$ and the exact solution by $f(r_i, \theta_j, t_n)$.

**Definition 2.** $L_1$ error is given by

$$\sum_{i,j} |f_{ij}^n - f(r_i, \theta_j, t_n)||K_{ij}|,$$


\begin{table}[h]
\centering
\caption{Error and order of accuracy for Example 5: $b = 40$, $K = 0$, final time $t = 4$, $\Delta t = 0.05$, and the reference solution is given by $P = Q = 320$.}
\begin{tabular}{|c|c|c|c||c|c|c|c|}
\hline
\multicolumn{1}{|c|}{\text{$\hat{f}_0(m)$}} & \multicolumn{3}{|c|}{$(b - m^2)^{\frac{3}{2}}$} & \multicolumn{3}{|c|}{$(b - m^2)^{\frac{3}{2}}$} \\
\hline
\text{$P = Q$} & \text{$L_1$ error} & \text{Order} & \text{$L_\infty$ error} & \text{Order} & \text{$L_1$ error} & \text{Order} & \text{$L_\infty$ error} & \text{Order} \\
\hline
20 & 1.0127E-01 & 1.5472E-02 & 1.0171E-01 & 1.5674E-02 & 5.1013E-02 & 0.999 & 3.0840E-03 & 1.232 \\
40 & 5.0798E-02 & 7.1504E-03 & 5.1032E-02 & 7.2466E-03 & 3.0924E-03 & 0.999 & 3.0840E-03 & 1.232 \\
80 & 2.5420E-02 & 3.0428E-03 & 2.5527E-02 & 3.0840E-03 & 1.232 \\
\hline
\end{tabular}
\end{table}

\begin{table}[h]
\centering
\caption{Error and order of accuracy for Example 5: $k_{11} = 0.5$, $k_{12} = k_{21} = 0.15$, final time $t = 4$, $\Delta t = 0.05$.}
\begin{tabular}{|c|c|c|c||c|c|c|c|}
\hline
\multicolumn{1}{|c|}{\text{$\hat{f}_0(m)$}} & \multicolumn{3}{|c|}{\text{$M(m)$ with $b = 40$}} & \multicolumn{3}{|c|}{\text{$M(m)$ with $b = 100$}} \\
\hline
\text{$P = Q$} & \text{$L_1$ error} & \text{Order} & \text{$L_\infty$ error} & \text{Order} & \text{$L_1$ error} & \text{Order} & \text{$L_\infty$ error} & \text{Order} \\
\hline
20 & 1.6176E-01 & 2.6971E-02 & 2.1369E-01 & 2.0079E-01 & 1.099 & 2.0079E-01 & 1.060 \\
40 & 8.1719E-02 & 1.1744E-01 & 1.1278E-01 & 1.1278E-01 & 1.099 & 2.0079E-01 & 1.060 \\
\hline
\end{tabular}
\end{table}

and the $L_\infty$ error is given by

$$
\max_{i,j} |f_{i,j}^n - f(r_i, \theta_j, t_n)|.
$$

Again when the exact solution is not available, we replace $f(r_i, \theta_j, t_n)$ by a reference solution to compute the errors.

The scaled discrete relative entropy is defined by

$$
\sum_{ij} \frac{(f_{ij}^n)^2}{Z_M M_{ij}} |K_{ij}|
$$

and the distance from the equilibrium solution by

$$
\max_{i,j} |f_{i,j}^n - f_{eq}(r_i, \theta_j)|.
$$

4.2.1. Accuracy test. We test the two-dimensional accuracy also with several choices of initial data.

Example 5. In this test, we consider the two-dimensional problem with $K = 0, b = 40$ and two types of initial data:

(i) $\hat{f}_0(m) = (b - |m|^2)^{\alpha}$, $\alpha = \frac{1}{4}, \frac{1}{2}, \frac{3}{2}$,

(ii) $\hat{f}_0(m) = \cos(3\pi |m|^2) + 1$, and

(iii) $\hat{f}_0(m) = M(m)$.

The results are given in Table 5.

In Table 6, we choose a symmetric $K$ with different values of $b$ and let $\hat{f}_0(m) = M(m)$. In this particular case, we know that the exact solution is independent of $t$, which is given by $f_{eq}(m) = Z_M^{-1} M(m)$. 

---

Copyright © by SIAM. Unauthorized reproduction of this article is prohibited.
1.00448
18
7.44458E-05
1
1.00005
2.96284E-05
−4.64376E-02
3.12548E-04
15
1.00596
1.30858E-03
7.56072E-03
1.00042
(−1.37345E-04
1.00004
1.04837
9
1.58325E-04
1
−3.75317E-04
−1
1.3167
−12
1.00267
9.25247
(2.27856E-03
1.06563
1
1.42082E-02
6
1234
HAILIANG LIU AND HUI YU

Table 7
Relative entropy in Example 6: \( P = Q = 40, b = 40, \mathcal{K} = 0 \).

| \( f_0(m) \) | \( (b - m^2)^\frac{3}{2} \) | \( (b - m^2)^\frac{5}{2} \) | \( \cos (3\pi \frac{|m|^2}{b}) + 1 \) |
|-------------|----------------|----------------|----------------|
| 1           | 1.000056      | 1              | 1.00487        |
| 2           | 1.00056       | 1              | 1.00448        |
| 3           | 1.000056      | 1              | 1.00042        |
| 4           | 1.00005      | 1              | 1.00004        |

Table 8
Numerical convergence to the equilibrium solution measured by distances for Example 7: \( b = 16, k_{11} = 1.1, k_{12} = 0.15, k_{21} = 0.15, \Delta t = 0.05 \).

| \( f_0(m) \) | \( (b - |m|^2)^\frac{3}{2} \) | \( (b - |m|^2)^\frac{5}{2} \) | \( \cos (3\pi \frac{|m|^2}{b}) + 1 \) |
|-------------|----------------|----------------|----------------|
| 3           | 8.80349E-02   | 1.42082E-02   | 2.27856E-03    |
| 6           | 2.75317E-04   | 3.12548E-04   | 1.58325E-04    |
| 9           | 7.44458E-05   | 2.96284E-05   | 1.37345E-04    |
| 12          | 1.04837       | 9.25247       | 3.12548        |
| 15          | 1.58325       | 4.6376E-02    | 3.12548        |
| 18          | 2.96284E-05   | 7.56072E-03   | 3.12548        |

4.2.2. Entropy decreasing and large time behavior.

Example 6. Consider the initial data (i) and (ii) in Example 5 with symmetric \( \mathcal{K} \), i.e., \( a = 0 \). The Fokker–Planck equation has an equilibrium solution \( f_{eq}(m) = Z_M^{-1}M(m) \) whose relative entropy is 1. Table 7 shows that the relative entropy is nonincreasing and converges to 1. Especially in the second column where we take the equilibrium solution as the initial data, the relative entropy stays the same.

Example 7. Let \( a = 0 \), i.e., \( k_{12} = k_{21} \). A comparison of solution behavior for two different initial data but with same \( b = 16 \) is plotted in Figures 3 and 4. Moreover, Table 8 shows that solutions in these two tests converge to the equilibrium solution.

4.2.3. Positivity preserving. Another feature of our scheme is positivity preserving. Let

\[
f_{\min}^n = \min_{i,j} f_{i,j}^n
\]
AN ENTROPY SATISFYING METHOD FOR THE FENE MODEL

4.2. Flow effects. Let \((x, y)\) be the macroscopic Eulerian coordinate and \(\nabla = (\frac{\partial}{\partial x}, \frac{\partial}{\partial y})^T\), associated with a fluid velocity field \(\vec{v}(x, y)\).

**Example 8** (simple extensional flow). We consider a homogeneous planar strain flow with the velocity field

\[
\vec{v} = (\alpha x, -\alpha y),
\]

where \(\alpha\) is the extensional rate. Then the velocity gradient tensor is

\[
\mathcal{K} = \nabla \vec{v} = \begin{pmatrix} \alpha & 0 \\ 0 & -\alpha \end{pmatrix}.
\]

This flow is irrotational and forms a strain flow. With this extensional flow, we consider the initial data with four separate peaks, defined by \(\tilde{f}_0(m) = \delta_\varepsilon(m)\), where

\[
\delta_\varepsilon(m) = \begin{cases} 
\cos\left(\frac{\pi (m_1 - m_{10})}{\varepsilon}\right) + 1 & \text{if } |m_1 - m_{10}| \leq \varepsilon \text{ and } |m_2 - m_{20}| \leq \varepsilon, \\
0 & \text{otherwise},
\end{cases}
\]

where \((m_{10}, m_{20}) \in \{(\pm \beta, 0), (0, \pm \beta)\}\) and \(\varepsilon < \beta < \sqrt{b} - \varepsilon\).

Note that in such a case the normalized equilibrium solution is

\[
f(m) = Z_M^{-1} M(m), \quad M(m) = (b - |m|^2)^{b/2} e^{\alpha(x^2 - y^2)}.
\]

The solutions at different times are plotted in Figure 6. In these tests we can see that the proposed method can well capture the equilibrium solutions for extensional flows.

The contours in Figure 7 show how the equilibrium solution \(f_{eq}(m) = Z_M^{-1} M(m)\) changes with respect to \(\alpha\). Observe that the two peaks of the equilibrium solution move away from each other as \(\alpha\) gets larger. For large \(\alpha\) one expects to see sharp peaks near boundary, with an amplification factor \(e^{\alpha x}\) of the profile \((b - |m|^2)^{b/2}\). Due
to the low order of our scheme, its performance when \( \alpha \beta \) becomes large tends to be less satisfactory. A higher order extension of the present method constitutes a future publication.

**Example 9** (steady state shear flow). The steady state shear flow has the velocity field

\[
\vec{v} = (\gamma y, 0),
\]

where \( \gamma \) is a constant shear rate, and the velocity gradient tensor is

\[
\mathcal{K} = \nabla \vec{v} = \begin{pmatrix} 0 & \gamma \\ 0 & 0 \end{pmatrix}.
\]

Let \( \gamma = 0.1, 0.3, 0.5, 1.0, 2.0 \). Figure 8 gives the contour plots of \( f^n_{i,j} \) at \( t = 4 \), from which the shear effects are clearly seen. Note that since for shear flow, \( \mathcal{K} \) is not normal, we do not expect the scheme to capture the large time behavior of the solution.

**Example 10** (a vortex). A typical vortex has the velocity field

\[
\vec{v} = (-\gamma y, \gamma x),
\]

with velocity gradient tensor

\[
\mathcal{K} = \nabla \vec{v} = \begin{pmatrix} 0 & \gamma \\ -\gamma & 0 \end{pmatrix}.
\]

Note that \( \mathcal{K} \) is not symmetric but it is normal, i.e., \( \mathcal{K}^T \mathcal{K} = \mathcal{K} \mathcal{K}^T \), hence \( f_{eq}(m) = Z_M^{-1} M(m) \) is still an equilibrium solution. In addition, \( \mathcal{K}^* = 0 \) in this case, so \( M(m) = (b - |m|^2)^{\beta} \). Table 9 shows the convergence to \( f_{eq} \) as \( t \) increases.
5. Conclusion. In this paper, we have investigated the Fokker–Planck equation which is of bead-spring type FENE dumbbell model for polymers, with our focus on the development of an entropy satisfying numerical method for the Fokker–Planck equation subject to zero flux on the boundary. We constructed simple, easy-to-implement conservative schemes which preserve equilibrium solutions and proved that they satisfy all three desired properties of the pdf: constant integral (mass conservation), positivity preserving, and entropy satisfying for $K$ normal. We also proved the long time convergence to the equilibrium solution at discrete levels. The goal of our future work is to extend the numerical method and analytical results herein to a higher order discontinuous Galerkin method.

Acknowledgments. The authors thank the referees who provided valuable comments, resulting in improvements in this paper.

REFERENCES


