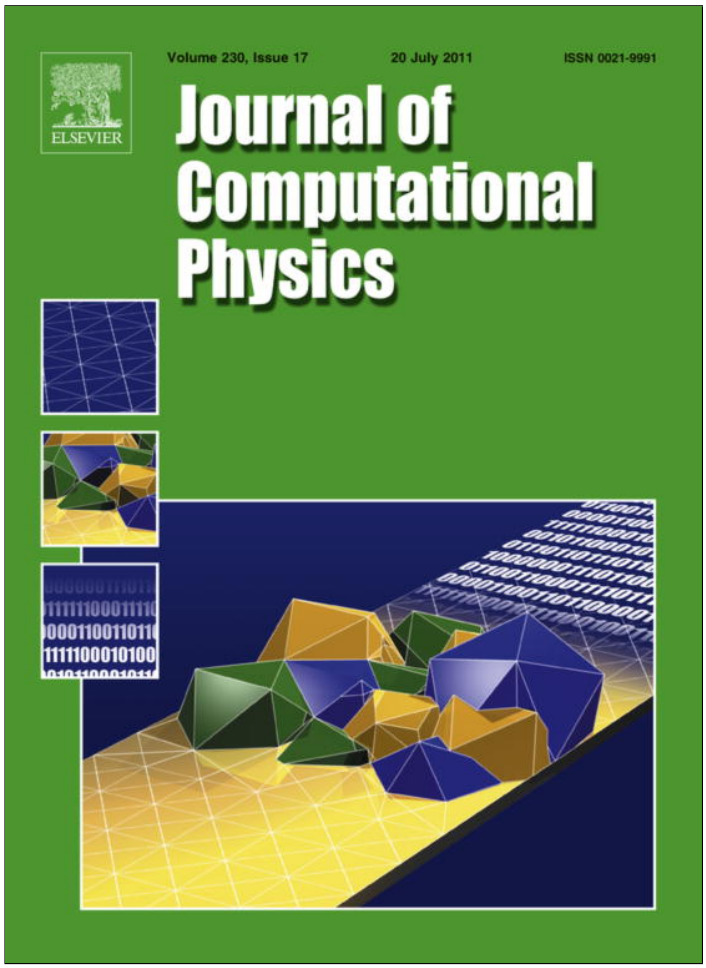


Journal of Computational Physics 230 (2011) 6420–6437

Contents lists available at ScienceDirect

Journal of Computational Physics

journal homepage: www.elsevier.com/locate/jcp

# A class of asymptotic-preserving schemes for the Fokker–Planck–Landau equation ☆

Shi Jin *, Bokai Yan

*Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison, WI 53706, USA*

ARTICLE INFO

*Article history:*
Received 22 October 2010
Received in revised form 1 April 2011
Accepted 3 April 2011
Available online 21 April 2011

*Keywords:*
Fokker–Planck–Landau equation
Fluid dynamic limit
Asymptotic-preserving schemes

ABSTRACT

We present a class of asymptotic-preserving (AP) schemes for the nonhomogeneous Fokker–Planck–Landau (nFPL) equation. Filbet and Jin [16] designed a class of AP schemes for the classical Boltzmann equation, by penalization with the BGK operator, so they become efficient in the fluid dynamic regime. We generalize their idea to the nFPL equation, with a different penalization operator, the Fokker–Planck operator that can be inverted by the conjugate-gradient method. We compare the effects of different penalization operators, and conclude that the Fokker–Planck (FP) operator is a good choice. Such schemes overcome the stiffness of the collision operator in the fluid regime, and can capture the fluid dynamic limit without numerically resolving the small Knudsen number. Numerical experiments demonstrate that the schemes possess the AP property for general initial data, with numerical accuracy uniformly in the Knudsen number.

Published by Elsevier Inc.

## 1. Introduction

The nonlinear Fokker–Planck–Landau (nFPL) equation is widely used in plasma physics. It is a Boltzmann type kinetic equation that describes the dynamics of the phase space density distribution function $f = f(t,x,v)$ of charged particles at position $x$, time $t$ with velocity $v$. The rescaled nFPL equation reads

$$\partial_t f + v\cdot\nabla_x f = \frac{1}{\epsilon}Q(f), \quad x\in\mathbb{R}^{N_x},\ v\in\mathbb{R}^{N_v} \tag{1.1}$$

with the nFPL operator

$$Q(f) = \nabla_v\cdot\int_{\mathbb{R}^{N_v}} A(v-v_*)(f(v_*)\nabla_v f(v) - f(v)\nabla_{v_*}f(v_*))dv_*, \tag{1.2}$$

where the semi-positive definite matrix $A(z)$ is given by

$$A(z) = \Psi(z)\left(I - \frac{z\otimes z}{|z|^2}\right), \quad \Psi(z) = |z|^{\gamma+2}. \tag{1.3}$$

Here $\epsilon$ is the Knudsen number, defined as the ratio of mean free path and the typical length scale. The parameter $\gamma$ is determined by the type of interaction between particles. In the case of inverse power law relationship, that is, when two particles at distance $r$ interact with a force proportional to $1/r^s$, $\gamma = \frac{s-5}{s-1}$. For example, in the cases of the Maxwell molecules $\gamma = 0$ (corresponding to $s = 5$) and for the Coulomb potential $\gamma = -3$ (corresponding to $s = 2$).

☆ This work was partially supported by NSF Grant DMS-0608720 and NSF FRG Grant DMS-0757285. SJ was also supported by a Van Vleck Distinguished Research Prize and a Vilas Associate Award from the University of Wisconsin-Madison.

* Corresponding author. Tel.: +1 608 263 3302; fax: +1 608 263 8891.
*E-mail addresses:* jin@math.wisc.edu (S. Jin), yan@math.wisc.edu (B. Yan).

0021-9991/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jcp.2011.04.002

The nFPL equation is derived as a limit of the Boltzmann equation when all the collisions become grazing. It is more relevant in physics for charged particles, where the Coulomb potential is presented. The first derivation was due to Landau [25,26]. For a mathematical derivation and analysis, we refer to the work of Arsenev and Buryak [1], Degond and Lucquin–Desreux [10], Desvillettes [12], Goudon [19] and also a detailed review by Villani [38]. In this article we will always take $\gamma = -3$, while the scheme itself can be applied to any $\gamma$.

Similar to the classical Boltzmann operator, the nFPL operator (1.2) also preserves mass, momentum and energy. This can be seen from its weak formulation. Noting

$$Q(f) = \nabla_v \cdot \int_{\mathbb{R}^{N_v}} A(v - v_*) f(v_*) f(v) (\nabla_v \log f(v) - \nabla_{v_*} \log f(v_*)) dv_*,$$

one obtains

$$\int_{\mathbb{R}^{N_v}} Q(f) \phi \, dv = -\frac{1}{2} \int \int_{\mathbb{R}^{N_v} \times \mathbb{R}^{N_v}} f(v_*) f(v) (\nabla_v \phi(v) - \nabla_{v_*} \phi(v_*))^T A(v - v_*) (\nabla_v \log f(v) - \nabla_{v_*} \log f(v_*)) dv_* \, dv. \tag{1.4}$$

Here $\nabla \phi$ is a column vector and $(\cdot)^T$ is the matrix transpose operation. The conservations of mass and momentum are straightforward. The conservation of energy follows from the fact that the null space of $A(z)$ is span$\{z\}$, i.e.,

$$A(z) z = 0.$$

Besides if one takes $\phi = \log f$ in (1.4), due to the semi-positivity of $A(z)$, one obtains the entropy dissipation inequality

$$\int_{\mathbb{R}^{N_v}} Q(f) \log f \leqslant 0. \tag{1.5}$$

Here the equality holds only if $f$ is the (local) equilibrium

$$M(x, v) = \frac{\rho(x)}{(2\pi T(x))^{N_v/2}} \exp\left( -\frac{(v - u(x))^2}{2T(x)} \right), \tag{1.6}$$

where the macroscopic quantities are given by

$$\begin{cases} \rho = \int_{\mathbb{R}^{N_v}} f \, dv, \\ u = \frac{1}{\rho} \int_{\mathbb{R}^{N_v}} v f \, dv, \\ T = \frac{1}{N\rho} \int_{\mathbb{R}^{N_v}} (v - u)^2 f \, dv. \end{cases}$$

Finally, as in the classical Boltzmann equation, when $\epsilon \to 0$, the moments of solution to (1.1) are governed asymptotically by the macroscopic Euler equations

$$\begin{cases} \partial_t \rho + \nabla_x \cdot \rho u = 0, \\ \partial_t (\rho u) + \nabla_x \cdot (\rho u \otimes u + p I) = 0, \\ \partial_t E + \nabla_x \cdot ((E + p) u) = 0, \end{cases} \tag{1.7}$$

where the total energy $E = \frac{1}{2} \rho u^2 + \frac{N}{2} \rho T$ and the pressure is given by the equation of state

$$p = \rho T.$$

A lot of efforts have been devoted to the numerical schemes for the nFPL equation recently. In [39,2,11,4–6], conservative finite difference type discretizations for the space homogeneous equation was derived. To reduce the computational cost caused by the high dimensional integral in the collision operator, spectral schemes were derived in [31,32,17]. However all these (explicit) schemes suffer from the stability constraint $\Delta t < C \epsilon \Delta v^2$. Lemou and Mieussens [28] proposed a class of implicit schemes, which invert a linear system, instead of a nonlinear one. However a full matrix needs to be inverted in their scheme.

We would like to develop numerical schemes for Eq. (1.1) that can capture the fluid dynamic limit (1.7) *automatically* when $\epsilon \to 0$. This is the so-called Asymptotic Preserving (AP) scheme, a term first introduced by Jin [22]. An AP scheme is efficient in the fluid dynamic regime ($\epsilon \ll 1$) since it allows one to capture the fluid dynamic limit (1.7) without numerically resolving small scale $\epsilon$. In recent years many AP schemes have been designed for kinetic equations, see for example [16,21] and references therein.

Recently Filbet and Jin [16] proposed a new class of AP schemes for the Boltzmann equation by penalization with BGK operator

$$\frac{f^{n+1} - f^n}{\Delta t} + v \cdot \nabla_x f^n = \frac{1}{\epsilon} \left[ Q(f^n) - \beta (M^n - f^n) + \beta \left( M^{n+1} - f^{n+1} \right) \right]. \tag{1.8}$$

The stiffness in the Boltzmann collision operator $\frac{1}{\epsilon} Q(f)$ can be overcome by the implicitly discretized BGK operator $\frac{\beta}{\epsilon} (M^{n+1} - f^{n+1})$, for large enough constant $\beta$. Since the implicit BGK operator can be solved explicitly, this method avoids the complexation of inverting the $Q(f)$ for small $\epsilon$.

The goal of this paper is to generalize their idea to the nFPL equation. The diffusive nature in the nFPL operator introduces new stiffness, which requires the penalization term to be also diffusive. The BGK operator is not suitable as a penalization any more. Several candidates are available. Analytical and numerical study in this paper show that the best choice is the following Fokker–Planck (FP) operator,

$$P_{FP}(f) = P_{FP}^M f = \nabla_v \cdot \left( M \nabla_v \left( \frac{f}{M} \right) \right). \tag{1.9}$$

The FP operator is a linear operator when the Maxwellian $M$ is time independent, in the case of the space homogeneous Fokker–Planck equation

$$\partial_t f = P_{FP}^M f.$$

Since $P_{FP}^M f$ preserves the macroscopic variables (density, momentum and energy), the Maxwellian $M$ does not change in time. The study of the FP operator can provide a useful guidance to study the classical Boltzmann operator (see [37]). We refer to [30,14] and the references therein for more detailed study. The numerical methods of FP equation were first introduced by Chang and Cooper [9]. Since then it has been studied in many works, such as [27,15,7]. In this article we also introduce a new discretization for the FP operator, which leads to a symmetric matrix, hence is easy to invert.

Here we summarize our new schemes. The first order scheme for the nFPL Eqs. (1.1) and (1.2) reads

$$\frac{f^{n+1} - f^n}{\Delta t} + v \cdot \nabla_x f^n = \frac{1}{\epsilon} \left( Q(f^n) - \beta P^n f^n + \beta P^{n+1} f^{n+1} \right) \tag{1.10}$$

where $P^n f^n = P_{FP}^{M^n} f^n$ is the FP operator (1.9) and $\beta$ is given by

$$\beta = \beta_0 \max_v \lambda(D_A(f)). \tag{1.11}$$

Here $\beta_0$ is a constant satisfying $\beta_0 > \frac{1}{2}$. A good choice is $\beta_0 = 1$. $\lambda(D_A)$ is the spectral radius of the positive symmetric matrix $D_A$, with $D_A(f)$ defined by

$$D_A(f) = \int A(v - v_*) f_* dv_*. \tag{1.12}$$

A second order implicit–explicit (IMEX) type scheme reads

$$\begin{cases} \frac{f^* - f^n}{\Delta t/2} + v \cdot \nabla_x f^n = \frac{Q(f^n) - \beta P^n f^n}{\epsilon} + \frac{\beta P^* f^*}{\epsilon}, \\ \frac{f^{n+1} - f^n}{\Delta t} + v \cdot \nabla_x f^* = \frac{Q(f^*) - \beta^* P^* f^*}{\epsilon} + \frac{\beta^* P^n f^n + \beta^* P^{n+1} f^{n+1}}{2\epsilon} \end{cases} \tag{1.13}$$

with $P(f)$ the FP operator (1.9). Suggested by numerical experiments, we take

$$\begin{aligned} \beta &= \beta_0 \max_{v,\lambda} \lambda(D_A(f)), \\ \beta^* &= \beta_0 \max_{v,\lambda} \lambda(D_A(f^*)). \end{aligned} \tag{1.14}$$

Again the constant coefficient satisfies $\beta_0 > \frac{1}{2}$. A good choice is $\beta_0 = \left(2 + \sqrt{2}\right)$.

If the initial data is close to equilibrium, i.e., $f^l = M^l + O(\epsilon)$, then the numerical solutions computed by schemes (1.10) and (1.13) always satisfy $f^n - M^n = O(\epsilon)$, due to the implicit discretized FP operator $P^{n+1} f^{n+1}$. Therefore the moments of $f$ solve (1.7) automatically, as $\epsilon \to 0$ with fixed $\Delta x$, $\Delta v$, $\Delta t$. Note that for an explicit scheme one cannot expect $f^n - M^n = O(\epsilon)$ even if initially it is.

If the initial data is far away from equilibrium, i.e., $f^l = M^l + O(1)$, our numerical experiments shows that $f^n - M^n = O(\epsilon)$ for sufficiently large $n$. This is the weakened AP property introduced in [16].

This article is organized as following. In Section 2 we describe the time discretization of the nFPL equation based on penalization. Then we give further details on the implementation of the schemes in Section 3, where we also introduce a symmetric operator to solve the linear system involving $P_{FP}^{n+1}$ efficiently. Finally we perform a series of numerical simulation in Section 4 to demonstrate the desired AP property and the numerical accuracy.

## 2. An AP scheme for the nFPL equation by penalization

An explicit scheme for the classical Boltzmann equation has to use time step $\Delta t \sim \epsilon$, due to the stiffness introduced by $\frac{1}{\epsilon}$ in the collision operator. As $\epsilon \to 0$ this would be too expensive. This is even worse for the nFPL equation since one has to take $\Delta t \sim \epsilon \Delta v^2$. An implicit scheme has no such restriction on the time step. But implicit schemes involve inverting an operator containing $Q(f)$, which cost a lot if one uses Newton's solver.

In [16] a penalization method (1.8) was introduced to overcome this difficulty. The BGK operator is used as the penalization, when $Q(f)$ is the classical Boltzmann operator. In this section we extend Filbet and Jin's idea in [16] in very different way.

The first question is, which operator is suitable as the penalization $P(f)$ for the nFPL operator. Unlike the classical Boltzmann operator, the nFPL operator behaves more like a diffusion operator. The stiffness on the right side of (1.1) comes from two parts: the stiffness due to $\frac{1}{\epsilon}$ when $\epsilon$ is small and the stiffness due to the diffusive nature of (1.2). We first use a toy model to motivate our idea.

### 2.1. A toy nonlinear diffusion equation

Consider the $N$-dimensional diffusion equation for $u(x,t)$, with $x \in \mathbb{R}^N$,

$$\frac{\partial u}{\partial t} = \frac{1}{\epsilon}\nabla_x \cdot (A(u,x)\nabla_x u), \tag{2.1}$$

where $A(u,x)$ is a semi-positive definite $N \times N$ matrix. $A$ can depend on $u$ and $x$.

When $\epsilon$ is small, this equation suffers from the stiffness originated from the diffusive operator and the large coefficient $\frac{1}{\epsilon}$. An explicit scheme requires $\Delta t \sim O(\epsilon(\Delta x)^2)$. We apply the penalization idea to remove this stiffness. The same idea has been used to solve the fourth order surface diffusion equation by Smereka [33]. See also a more recent application in imaging processing [3].

Consider the following scheme with an diffusion term $\frac{\beta}{\epsilon}\nabla_x^2 u$ added and subtracted, but discretized at different time level.

$$\frac{u^{n+1}-u^n}{\Delta t} = \frac{1}{\epsilon}\nabla_x \cdot (A(u^n,x)\nabla_x u^n) - \frac{\beta}{\epsilon}\nabla_x^2 u^n + \frac{\beta}{\epsilon}\nabla_x^2 u^{n+1}. \tag{2.2}$$

For stability one requires

$$\beta \geqslant \frac{1}{2}\max_{x\in\mathbb{R}^N, u\in\mathbb{R}} \lambda(A(u,x)). \tag{2.3}$$

One can show the following result.

**Theorem 2.1.** *The scheme (2.2) is a stable time discretization of (2.1) under the condition (2.3). More precisely, define the energy*

$$E^n = \int \left(|u^n|^2 + \frac{\beta\Delta t}{\epsilon}|\nabla_x u^n|^2\right)dx,$$

*then $E^{n+1} \leqslant E^n$, for any $n \geqslant 0$.*

The proof is similar to that in [3]. For completeness, we give the proof in the Appendix.

Besides, for the isentropic case $A(u,x) = aI$, where $a$ is a constant and $I$ is the identical matrix, one can obtain a positivity preserving scheme under a stronger condition

$$\beta \geqslant \max_{x\in\mathbb{R}^N, u\in\mathbb{R}} \lambda(A(u,x)) = a. \tag{2.4}$$

Without loss of generality, we assume $a = 1$.

**Theorem 2.2.** *Consider the heat Eq. (2.1) in the isentropic case $A(u,x) = I$, with nonnegative initial data $u^I \geqslant 0$. The scheme (2.2) gives nonnegative solutions $u^n$ under the condition (2.4).*

**Proof.** The solution to (2.2) can be written as

$$\begin{aligned} u^{n+1} &= \left(1-\frac{\beta\Delta t}{\epsilon}\nabla_x^2\right)^{-1}\left(1+\frac{(1-\beta)\Delta t}{\epsilon}\nabla_x^2\right)u^n = \left(1-\frac{\beta\Delta t}{\epsilon}\nabla_x^2\right)^{-1}\left\{\left(1-\frac{1}{\beta}\right)\left(1-\frac{\beta\Delta t}{\epsilon}\nabla_x^2\right)+\frac{1}{\beta}\right\}u^n \\ &= \left(1-\frac{1}{\beta}\right)u^n + \frac{1}{\beta}\left(1-\frac{\beta\Delta t}{\epsilon}\nabla_x^2\right)^{-1}u^n. \end{aligned}$$

Notice that the second term

$$u^{n+\beta} = \left(1-\frac{\beta\Delta t}{\epsilon}\nabla_x^2\right)^{-1}u^n$$

is just an approximation of $u((n+\beta)\Delta t)$ solved by the backward Euler scheme

$$\frac{u^{n+\beta}-u^n}{\beta\Delta t} = \frac{1}{\epsilon}\nabla_x^2 u^{n+\beta}.$$

The exact solution of $u^{n+\beta}$ is nonnegative due to the maximum principle. As long as one can find a positive solver for the implicit scheme, one can obtain a nonnegative $u^{n+1}$ by taking the linear combination of $u^n$ and $u^{n+\beta}$, under the condition $\beta \geqslant 1$. □

**Remark 2.3.** To the authors' knowledge, this positivity result is new. The positivity preserving property for the scheme (2.2) to the non-isentropic case $A(u,x) \neq aI$ is not clear yet. Theorem 2.2 suggests that the choice of

$$\beta \geqslant \max_{x\in\mathbb{R}^N, u\in\mathbb{R}} \lambda(A(u,x))$$

is better than $\frac{1}{2} \leqslant \frac{\beta}{\max_{x\in\mathbb{R}^N,u\in\mathbb{R}}\lambda(A(u,x))} < 1$, although in both cases the scheme is stable.

**Remark 2.4.** The proof for positivity is valid as long as, (i) a linear operator $\mathcal{L}$ is used as both the original operator and the penalization operator

$$\frac{u^{n+1}-u^n}{\Delta t} = \frac{1}{\epsilon}\mathcal{L}u^n - \frac{\beta}{\epsilon}\mathcal{L}u^n + \frac{\beta}{\epsilon}\mathcal{L}u^{n+1}.$$

and (ii) the backward Euler gives nonnegative $u^{n+\beta}$. For example, $\mathcal{L}u = -u,\ \mathcal{L}u = \nabla^2 u$ and the linear Fokker–Planck operator we will study later.

**Remark 2.5.** For diffusion Eq. (2.1), one cannot take $P(u) = -\frac{\beta}{\epsilon}u$ as the penalization operator. We give a simple argument here. For simplicity, we consider the one dimensional equation

$$u_t = \frac{1}{\epsilon}u_{xx}$$

with the penalization scheme

$$\frac{u^{n+1}-u^n}{\Delta t} = \frac{1}{\epsilon}u^n_{xx} + \frac{\beta}{\epsilon}u^n - \frac{\beta}{\epsilon}u^{n+1}. \tag{2.5}$$

After the Fourier transform on $x$, one gets

$$\frac{\hat{u}^{n+1}-\hat{u}^n}{\Delta t} = -\frac{k^2}{\epsilon}\hat{u}^n + \frac{\beta}{\epsilon}\hat{u}^n - \frac{\beta}{\epsilon}\hat{u}^{n+1},$$

where $\hat{u}$ is the Fourier transform of $u$, and $k$ the Fourier number. Then

$$\hat{u}^{n+1} = \frac{\epsilon + (\beta - k^2)\Delta t}{\epsilon + \beta\Delta t}\hat{u}^n.$$

For stability uniformly in $\epsilon$, one needs

$$\beta \geqslant \frac{1}{2}\max_k k^2 = O(N_x^2) = O\left(\frac{1}{(\Delta x)^2}\right),$$

where $N_x$ is the number of grid points in the $x$ direction.

Since $\beta$ appears in the truncation error for (2.5), this gives the error of (2.5) like $O\left(\frac{\Delta t}{\epsilon(\Delta x)^2}\right)$, which is not good in the regime $\Delta t > O(\epsilon(\Delta x)^2)$.

On the other hand, if one applies the parabolic penalization,

$$\frac{u^{n+1}-u^n}{\Delta t} = \frac{1}{\epsilon}u^n_{xx} - \frac{\beta}{\epsilon}u^n_{xx} + \frac{\beta}{\epsilon}u^{n+1}_{xx},$$

then $\beta \geqslant \frac{1}{2}$ gives a stable scheme. $\beta = \frac{1}{2}$ is the well known Crank–Nicolson scheme while $\beta = 1$ gives the backward Euler scheme.

### *2.2. The choice of penalization operator for the nFPL equation*

As illustrated in the last subsection, the classical BGK operator $P = (M - f)$ used in [16] to penalize the classical Boltzmann equation can not be used here. Instead, we impose the following criteria for the choice of $P$:

(C1) $P(f)$ preserves mass, momentum and energy.
(C2) $P(f)$ is easy to invert, or at least easier than $Q(f)$.
(C3) $P(f)$ contains a diffusion operator.
(C4) $P(f)$ can push $f$ toward the equilibrium $M$.

The condition (C4) implies the operator is well balanced, i.e. $P(f) = 0$ if and only if $f = M$. This is a necessary condition for AP property.

To find a suitable penalization $P(f)$, a key observation is the fact that the diffusion in the nFPL operator (1.2) is on $(f-M)$, not on $f$. In other words, one needs to extract a diffusion operator on $(f-M)$ from (1.2). To do this, note that

$$\nabla f = \nabla\left(M\frac{f}{M}\right) = \frac{f}{M}\nabla M + M\nabla\frac{f}{M} = f\nabla\log M + M\nabla\frac{f}{M},$$

thus the nFPL operator (1.2) can be rewritten as

$$\begin{aligned} Q(f) &= \nabla_v\cdot\int A(v-v_*)(f_*\nabla_v f - f\nabla_{v_*}f_*)dv_* \\ &= \nabla_v\cdot\int A(v-v_*)\left(f_*f(\nabla_v\log M - \nabla_{v_*}\log M_*) + f_*M\nabla\frac{f}{M} - fM_*\nabla_{v_*}\frac{f_*}{M_*}\right)dv_* \\ &= \nabla_v\cdot\left(D_A(f)M\nabla\frac{f}{M}\right) - \nabla_v\cdot(F_A(f)f), \end{aligned} \tag{2.6}$$

where $D_A(f)$ is defined in (1.12), and

$$F_A(f) = \int A(v-v_*)M_*\nabla_{v_*}\frac{f_*}{M_*}dv_*.$$

Here we have used the fact

$$A(v-v_*)(\nabla_v\log M - \nabla_{v_*}\log M_*) = 0.$$

The first term in (2.6) is a diffusion operator on $f-M$ we desire, which can be written as

$$\nabla_v\cdot\left(D_A(f)M\nabla\frac{f-M}{M}\right).$$

Thus a natural choice of the penalization operator is the Fokker–Planck (FP) operator

$$P(f) = P_{FP}^M f = \nabla_v\cdot\left(M\nabla_v\left(\frac{f}{M}\right)\right). \tag{2.7}$$

Motivated by Theorem 2.1, the stability condition on $\beta$ is conjectured as

$$\beta \geqslant \frac{1}{2}\max_v \lambda(D_A(f)). \tag{2.8}$$

The convolution type $2\times 2$ or $3\times 3$ matrix $D_A(f)$ can be computed without difficulty by the Fast Fourier Transform. Actually the Fourier transform of $A(v)$ and $f(v)$ are obtained as a by-product during the computation of $Q(f)$, if one applies a spectral scheme such as in [32]. Then the eigenvalue can be computed easily.

**Remark 2.6.** It is easy to check that this $P(f)$ satisfies the requirements (C1)–(C3) we looked for. As for (C4), let us consider the homogeneous equation

$$\partial_t f = P_{FP}^M f = \nabla_v\cdot\left(M\nabla_v\left(\frac{f}{M}\right)\right).$$

A classical fact is that, the relative entropy

$$H(f|M) = \int f\log\frac{f}{M}\,dv \tag{2.9}$$

decays exponentially along the solution, see [8] for example. In fact one easily derives,

$$-\frac{d}{dt}H(f|M) = I(f|M) := \int f\left|\nabla_v\log\frac{f}{M}\right|^2 dv.$$

With the well known Stam–Gross logarithmic Sobolev inequality [34,20]

$$H(f|M) \leqslant \frac{1}{2}I(f|M),$$

the exponential decay is obtained,

$$H(f|M) \leqslant e^{-2t}H(f^I|M).$$

Now our first order scheme reads,

$$\frac{f^{n+1}-f^n}{\Delta t}+v\cdot\nabla_x f^n=\frac{1}{\epsilon}\left(Q(f^n)-\beta P^n f^n+\beta P^{n+1}f^{n+1}\right) \tag{2.10}$$

with $P^n=P_{FP}^{M^n}$ the FP operator.

First $M^{n+1}$ can be obtained *explicitly* thanks to the fact that the right side of (2.10) preserves density, momentum and energy. Multiply both sides of (2.10) by $\phi=1,\ v,\ \frac{|v|^2}{2}$ and integrate over $v$, one obtains

$$\int\phi\left(\frac{f^{n+1}-f^n}{\Delta t}+v\cdot\nabla_x f^n\right)dv=0.$$

So the moments at $t^{n+1}$ can be derived *explicitly* by,

$$(\rho,\rho u,E)^{n+1}=\int\phi(f^n-\Delta t\,v\cdot\nabla_x f^n)dv \tag{2.11}$$

and $M^{n+1}$ is defined by (1.6). Then one can solve $f^{n+1}$

$$f^{n+1}=\left(1-\frac{\Delta t\beta}{\epsilon}P^{n+1}\right)^{-1}\left(f^n-\Delta t\,v\cdot\nabla_x f^n+\frac{\Delta t}{\epsilon}(Q(f^n)-\beta P^n f^n)\right) \tag{2.12}$$

Section 3.2 describes a detailed algorithm to compute the inverse of $\left(1-\frac{\Delta t\beta}{\epsilon}P^{n+1}\right)$.

### 2.3. The choice of the penalization weight $\beta$

Roughly speaking, the value of $\beta$ determines how much the stiffness in the nFPL operator $Q(f)$ is removed. (2.8) gives a lower bound of $\beta$ for stability. However the equal sign does not give a satisfactory choice of $\beta$. One reason is there is always numerical error in the computation of matrix $D_A(f)$. The choice of $\beta$ on borderline is of high risk. In numerical simulation we take $\beta$ as

$$\beta=\beta_0\max_v\lambda(D_A(f)), \tag{2.13}$$

where $\beta_0>\frac{1}{2}$ is a constant.

To find a suitable $\beta_0$, we reconsider the toy model studied by Filbet and Jin [16],

$$\partial_t f=-\frac{1}{\epsilon}f. \tag{2.14}$$

Apply the first order penalization,

$$\frac{f^{n+1}-f^n}{\Delta t}=-\frac{1}{\epsilon}(f^n-vf^n+vf^{n+1}). \tag{2.15}$$

Then one obtains

$$f^{n+1}=\frac{\epsilon+(v-1)\Delta t}{\epsilon+v\Delta t}f^n.$$

A simple analysis shows that the scheme (2.15) is stable uniformly in $\epsilon$ if $v\geqslant\frac{1}{2}$, analogous to the stability condition $\beta_0\geqslant\frac{1}{2}$ in (2.13). (2.15) with $v=\frac{1}{2}$ gives a second order discretization in time for (2.14). However $v=1$ seems to be a better choice. $v=1$ gives a first order discretization in time, but it gives the fastest decay to equilibrium. Besides, the nonnegativity is guaranteed as long as $v\geqslant 1$. The nonnegativity is a natural requirement since $f$ is the density distribution. The fast decay is important for the AP purpose, when the initial data is not close to the local equilibrium.

For the same reasons, we also take $\beta_0=1$ in (2.13), which is the same as the conclusion we derived in Remark 2.3.

Similarly the second order scheme

$$\begin{cases}\frac{f^*-f^n}{\Delta t/2}=-\frac{1}{\epsilon}(f^n-vf^n+vf^*),\\ \frac{f^{n+1}-f^n}{\Delta t}=-\frac{1}{\epsilon}(f^*-vf^*+v(f^n+f^{n+1})/2),\end{cases}$$

gives

$$f^{n+1}=\frac{\epsilon^2+\epsilon\Delta t(v-1)+\frac{1}{4}\Delta t^2(v^2-4v+2)}{(\epsilon+\Delta tv/2)^2}f^n.$$

Again the scheme is stable if $v\geqslant\frac{1}{2}$. To guarantee the nonnegativity, one needs

$$v-1\geqslant 0,\quad\text{and}\quad v^2-4v+2\geqslant 0.$$

Hence $v\geqslant\left(2+\sqrt{2}\right)$ is a sufficient condition. And $v=\left(2+\sqrt{2}\right)$ gives the fastest decay when $\Delta t\gg\epsilon$.

Therefore the $\beta_0$ is chosen to be $2+\sqrt{2}$ in (1.14) for the second order scheme (1.13).

### 2.4. Other penalizations

Another candidate of the penalization operator is the classical diffusion operator

$$P_D(f) = \nabla_v^2(f - M). \tag{2.16}$$

One can check that this operator satisfies the requirements (C1)–(C3). Besides, (C4) is also satisfied, since for the homogenous equation

$$\frac{\partial}{\partial t} f = \nabla_v^2(f - M),$$

one has the inequality

$$\frac{1}{2}\frac{\partial}{\partial t}\int (f-M)^2 dv = -\int |\nabla_v (f-M)|^2 dv \leqslant 0.$$

However this classical diffusion operator $P_D(f)$ is not qualified as penalization. Numerical simulations in Section 4.5 shows that the penalization scheme (2.10) with $P(f) = P_D(f)$ is not AP in long time. In fact, As shown in Sections 4.5.2 and 4.5.3, this penalization cannot stabilize the scheme (2.10) at all.

Here we give a heuristic explanation [18]. The rigorous analysis is out of the scope of this paper.

Let us consider the homogeneous equation

$$\frac{\partial f}{\partial t} = \mathcal{C}(f).$$

We briefly summarize the trend to equilibrium for the solutions with respect to different operator $\mathcal{C}(f)$.

- $\mathcal{C}(f) = Q(f)$ is the nFPL operator. It has been proved that, for the hard potential ($\Psi(z) = |z|^{\gamma+2}$, with $\gamma \geqslant 0$) and the mollified soft potential ($\Psi$ is smooth and behaves at infinity like $|z|^{\gamma+2}$, $-3 < \gamma < 0$), the relative entropy (2.9) decays exponentially [13,36]. For the Coulomb case ($\gamma = -3$), a rigorous proof of this exponential decay is not available yet. In a recent work [35], Strain and Guo proved $f$ approaches $M$ in a rate of $e^{-\lambda t^{2/3}}$ when $f$ is close to $M$. We perform some experiments in Section 4.5.1, which numerically verifies that the relative entropy decays exponentially.
- $\mathcal{C}(f) = P_{FP}(f)$ is the linear FP operator. As mentioned before, the relative entropy decays exponentially along solution.
- $\mathcal{C}(f) = P_D(f)$ is the classical diffusion equation. The solution does not enjoy the entropy decaying property. When considering the distance $\|f - M\|_\infty$, the solution $f$ converges to the equilibrium $M$ with a polynomial rate $O(t^{-N_v/2})$

The numerical verifications of these rates are done in Section 4.5.1. The weak decay to equilibrium for classical diffusion operator seems to be the source of unsuitability as the penalization.

From now on, we will always take $P(f)$ to be the FP operator (1.9), except otherwise specified.

## 3. A full discretization of the nFPL equation

We now describe the detailed algorithm for the first order scheme (1.10). The algorithm for the second order scheme (1.13) is similar.

Suppose the numerical solution $f^n$ at time $t^n$ is given, then.

**Step 1** Apply a first order upwind scheme or second order TVD scheme on the transport operator to compute new moments via (2.11) by a quadrature rule, say the trapezoidal rule, then the new Maxwellian $M^{n+1}$ is obtained at each $x$ and $P^{n+1}$ can be defined.

**Step 2** At each $x$, compute the nFPL operator $Q(f^n)$ and the coefficient matrix $D_A f^n$ defined by (1.12). Then the penalization weight $\beta = \beta(x)$ is determined by (1.11).

**Step 3** Discretize the linear FP operator $P^n$ and $P^{n+1}$. One arrives at a linear system in the $v$ direction for each $x$.

**Step 4** Solve the resulting linear system to obtain $f^{n+1}$ in (2.12) at each $x$.

It is very important that one computes $M^{n+1}$ before $Q(f^n)$ and $P^n(f^n)$ are computed. This is equivalent to say that $Q(f^n)$ and $P^n(f^n)$ are assumed to be conservative after numerical approximation. This conservation is not true for many efficient schemes. The spectral scheme on nFPL operator introduced in [32] preserves the mass while conservation of momentum and energy are "spectrally preserved". As for the FP operator, the discretization we are using (see Section 3.2) preserves the mass while the errors in conservation of momentum and energy are $O(\Delta v^2)$. For the first order scheme (2.12), if one computes $Q(f^n)$ and $P^n(f^n)$ first and then computes the moments of $f^{n+1}$ from

$$f^n - \Delta t\, v\cdot\nabla_x f^n + \frac{\Delta t}{\epsilon}(Q(f^n) - \beta P(f^n)),$$

one would get a error of $O(\frac{\Delta t \Delta v^p}{\epsilon})$ in momentum and energy. This could give totally unphysical results. For example one might get negative temperature $T^{n+1}$ and then the new equilibrium $M^{n+1}$ is not a Gaussian at all.

In the following sections we describe how to compute $Q(f)$ and $P(f)$.

### 3.1. Computation of Q(f)

We use the fast spectral method designed by Pareschi et al. [32]. The computational cost is $O(N \log N)$, where $N = N_v^d$ is the grid points in velocity space. The scheme preserves mass exactly, and preserves momentum and energy with the spectral accuracy. Besides, in numerical implementation we will replace $Q(f)$ by $\widetilde{Q}(f) = Q(f) - Q(M)$ to make sure the equilibrium gives well balanced result $\widetilde{Q}(M) = 0$.

Besides, one obtains the Fourier transform of $A(v)$ and $f(v)$ during the implementation of this spectral method. Therefore the matrix $D_A(f)$ can be obtained easily by a simple inverse Fourier transform.

### 3.2. Discretization of P(f)

The discretization of the FP operator (1.9) has been studied in many works. A popular method is initiated by Chang and Cooper [9] and studied later by Larsen et al. [27], Buet and Cordier [15], Buet and Dellacherie [7]. However the Chang–Cooper discretization gives a nonsymmetric matrix, which is not easy to invert. Here we introduce a new discretization based on the symmetrized operator

$$\widetilde{P}^M h = \frac{1}{\sqrt{M}} \nabla_v \cdot \left( M \nabla_v \left( \frac{h}{\sqrt{M}} \right) \right). \tag{3.1}$$

Note

$$P^M f = \sqrt{M} \widetilde{P}^M \frac{f}{\sqrt{M}} \tag{3.2}$$

and we can rewrite (2.12) as

$$\left( \frac{f}{\sqrt{M}} \right)^{n+1} = \left( 1 - \frac{\Delta t \beta}{\epsilon} \widetilde{P}^{n+1} \right)^{-1} \left\{ \frac{1}{\sqrt{M^{n+1}}} \left( f^n - \Delta t v \cdot \nabla_x f^n + \frac{\Delta t}{\epsilon} \left( Q(f^n) - \beta \sqrt{M^n} \widetilde{P}^n \frac{f^n}{\sqrt{M^n}} \right) \right) \right\} \tag{3.3}$$

Now we give the discretization of $\widetilde{P}$ in one dimension. The extension to higher dimension is similar.

$$\begin{aligned} \left( \widetilde{P}^M h \right)_j &= \frac{1}{(\Delta v)^2} \frac{1}{\sqrt{M_j}} \left( \sqrt{M_j M_{j+1}} \left( \left( \frac{h}{\sqrt{M}} \right)_{j+1} - \left( \frac{h}{\sqrt{M}} \right)_j \right) - \sqrt{M_j M_{j-1}} \left( \left( \frac{h}{\sqrt{M}} \right)_j - \left( \frac{h}{\sqrt{M}} \right)_{j-1} \right) \right) \\ &= \frac{1}{(\Delta v)^2} \left( h_{j+1} - \frac{\sqrt{M_{j+1}} + \sqrt{M_{j-1}}}{\sqrt{M_j}} h_j + h_{j-1} \right). \end{aligned} \tag{3.4}$$

Then $\widetilde{P}$ is symmetric (under the normal inner product). Besides, after this discretization, we have the well balanced property

$$P^M M = \sqrt{M} \widetilde{P}^M \sqrt{M} = 0.$$

Therefore, if $\widetilde{P}\left( f/\sqrt{M} \right) = O(\epsilon)$, the inversion gives $f = M + O(\epsilon)$. This is important for the *AP* property.

**Remark 3.1.** This discretization preserves the mass while the errors in conservation of momentum and energy are $O(\Delta v^2)$. One might suggest the discretization of the FP operator based on another equivalent form,

$$Pf = \nabla_v \left( \nabla_v f + \frac{v - u}{T} f \right).$$

The discretized operator can indeed preserve all the moments exactly. However, this discretizaiton does not share the well-balanced property and therefore does not meet the condition (C4). It does not give an AP scheme. Besides, it gives a nonsymmetric matrix, which is not easy to invert.

### 3.3. Inversion of the linear system

We start with a lemma.

**Lemma 3.2.** *Let us write the matrix*

$$A = I - \frac{\Delta t \beta}{\epsilon} \widetilde{P}^{n+1},$$

*where I is an identity matrix and* $\widetilde{P}^{n+1}$ *is discretized as in (3.4). Then A is positive definite.*

**Proof.** Clearly $A$ is symmetric since the discretized $\widetilde{P}$ is symmetric.

For any nonzero vector $h$,

$$h^T A h = \sum_j h_j^2 - \frac{\Delta t\beta}{\epsilon}\sum_j h_j\left(\widetilde{P}^M h\right)_j = \sum_j h_j^2 + \frac{\Delta t\beta}{\epsilon(\Delta v)^2}\sum_j \sqrt{M_j M_{j+1}}\left|\left(\frac{h}{\sqrt{M}}\right)_{j+1} - \left(\frac{h}{\sqrt{M}}\right)_j\right|^2.$$

This is always positive. □

Therefore one can apply the Conjugate Gradient (CG) method on (3.3) to obtain $\left(\frac{f}{\sqrt{M}}\right)^{n+1}$. Then $f^{n+1}$ is obtained. To start the CG algorithm, a good initial guess is

$$f_0^{n+1} = M^{n+1}.$$

Let $f_k^{n+1}$ be the value of $f^{n+1}$ after $k$th iteration.

Then

$$\frac{f_k^{n+1}}{\sqrt{M^{n+1}}} \in \sqrt{M^{n+1}} + \mathrm{span}\{r, Ar, \ldots, A^{k-1}r\},$$

where

$$r = \frac{1}{\sqrt{M^{n+1}}}\left(f^n - \Delta t v\cdot\nabla_x f^n - M^{n+1} + \frac{\Delta t}{\epsilon}\left(Q(f^n) - \beta\sqrt{M^n}\widetilde{P}^n\frac{f^n}{\sqrt{M^n}}\right)\right).$$

Since $Q(f)$ and $P(f)$ preserve mass exactly, $f_k^{n+1}$ shares the same mass with $M^{n+1}$, for all $k \geqslant 0$. As for the momentum and energy, one might question that the vector $r$ could introduce an error of $O(\frac{\Delta t\Delta v^p}{\epsilon})$, where $p$ is the order of accuracy of the velocity discretization for operator $\widetilde{P}$. However our numerical experiments show that the conservations of momentum and energy are quite satisfactory, see Table 1 in Section 4.2.3 for details.

**Remark 3.3.** The use of CG method in an implicit discretization of the collision operator for the Fokker–Planck–Landau equation is not new, see for example [28] where the CG method was used when the linear operator to invert is self-adjoint, and the CG method preserves the exact conservation of mass, momentum and energy when the operator to invert preserves these quantities. The loss of exact conservations of momentum and energy in our method has to do with the discretization of the penalty operator.

## 4. Numerical simulation

### 4.1. The convergence order

First we numerically check that the two schemes (1.10) and (1.13) are indeed first and second order accurate.

To avoid the influence from the boundary, we take the periodic boundary condition in $x$. The initial data are given by $f^I = M^I$, with

$$\rho^I = \frac{2+\sin\pi x}{3}, \quad u^I = 0, \quad T^I = \frac{9+\cos\pi x}{50} \tag{4.1}$$

where $x \in [-1,1]$, $v \in [-\pi,\pi]^2$.

**Table 1**
The errors in moments when inverting the linear system (3.3).

| | | $\Delta v = 0.4$ | $\Delta v = 0.2$ | $\Delta v = 0.1$ |
|---|---|---|---|---|
| $Err(1)$ | $\epsilon = 1$ | $2.027\times10^{-9}$ | $1.572\times10^{-9}$ | $1.454\times10^{-9}$ |
| | $\epsilon = 10^{-2}$ | $2.017\times10^{-9}$ | $1.565\times10^{-9}$ | $1.449\times10^{-9}$ |
| | $\epsilon = 10^{-4}$ | $4.380\times10^{-10}$ | $3.219\times10^{-10}$ | $2.810\times10^{-10}$ |
| | $\epsilon = 10^{-6}$ | $3.654\times10^{-11}$ | $1.941\times10^{-11}$ | $1.145\times10^{-11}$ |
| $Err(v)$ | $\epsilon = 1$ | $1.502\times10^{-7}$ | $3.820\times10^{-8}$ | $9.601\times10^{-9}$ |
| | $\epsilon = 10^{-2}$ | $6.218\times10^{-6}$ | $1.529\times10^{-6}$ | $3.796\times10^{-7}$ |
| | $\epsilon = 10^{-4}$ | $1.135\times10^{-6}$ | $2.821\times10^{-7}$ | $7.020\times10^{-8}$ |
| | $\epsilon = 10^{-6}$ | $1.229\times10^{-6}$ | $3.114\times10^{-7}$ | $7.821\times10^{-8}$ |
| $Err(\lvert v\rvert^2)$ | $\epsilon = 1$ | $8.346\times10^{-8}$ | $6.226\times10^{-8}$ | $5.691\times10^{-8}$ |
| | $\epsilon = 10^{-2}$ | $2.691\times10^{-7}$ | $1.118\times10^{-7}$ | $7.584\times10^{-8}$ |
| | $\epsilon = 10^{-4}$ | $1.690\times10^{-6}$ | $4.471\times10^{-7}$ | $1.760\times10^{-7}$ |
| | $\epsilon = 10^{-6}$ | $1.749\times10^{-6}$ | $4.602\times10^{-7}$ | $1.797\times10^{-7}$ |

The spectral scheme described in [32] allows us to compute the nFPL operator (1.2) efficiently. Numerical experiments shows that $N_v = 32$ can give satisfactory results.

We compute the solutions with the number of grid points $N_x = 32, 64, 128, 256, 512, 1024$ respectively. The time step is given by $\Delta t = \Delta x/8$. After time $t_{\max} = 0.125$ we check the following error

$$e_{\Delta x} = \max_{t\in(0,t_{\max})} \frac{\|f_{\Delta x}(t) - f_{2\Delta x}(t)\|_p}{\|f^I\|_p} \tag{4.2}$$

This can be considered as an estimation of the relative error in $L^p$ norm, where $f_h$ is the numerical solution computed from a grid of size $\Delta x = \frac{x_{\max}-x_{\min}}{N_x}$. The numerical scheme is said to be $k$th order if $e_{\Delta x} \leqslant C\Delta x^k$, for $\Delta x$ small enough.

For (1.10) the first order upwind scheme is applied to the transport operator. As for (1.13), the transport operator is solved by a second order TVD scheme using the van Leer slope limiter (see [29] for details).

Fig. 1 gives the convergence order in $L^1$ norm, showing that the two schemes are first order and second order in $x$, respectively (hence in time) uniformly in $\epsilon$, as expected.

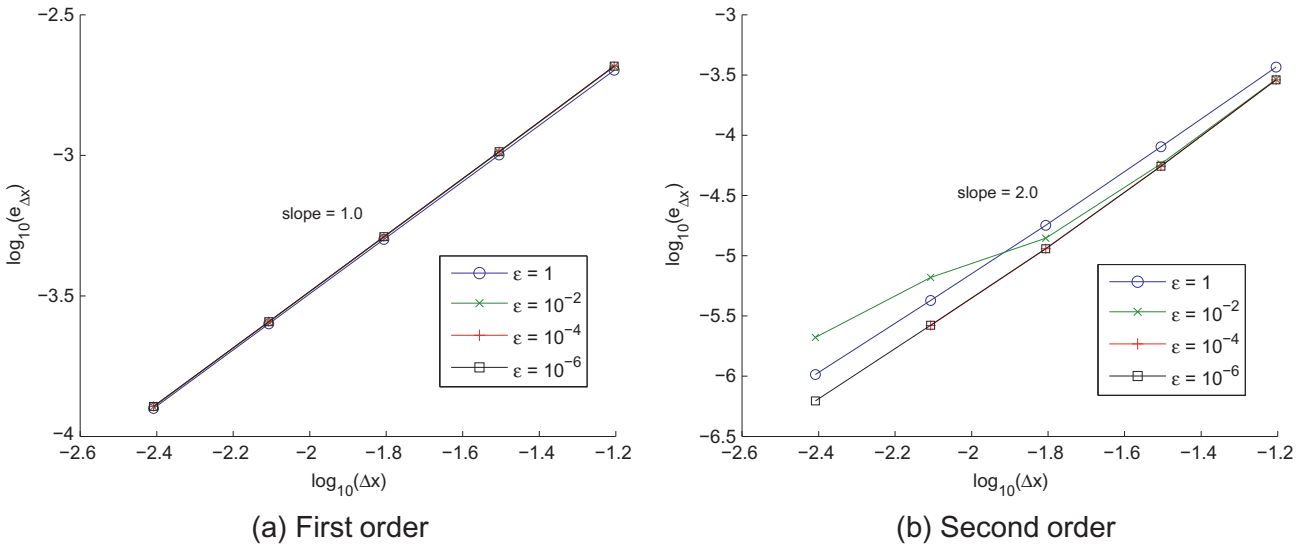



**Fig. 1.** The $l^1$ errors (4.2) of the first order scheme (1.10) (left) and the second order scheme (1.13) (right) with different Knudsen number $\epsilon$.

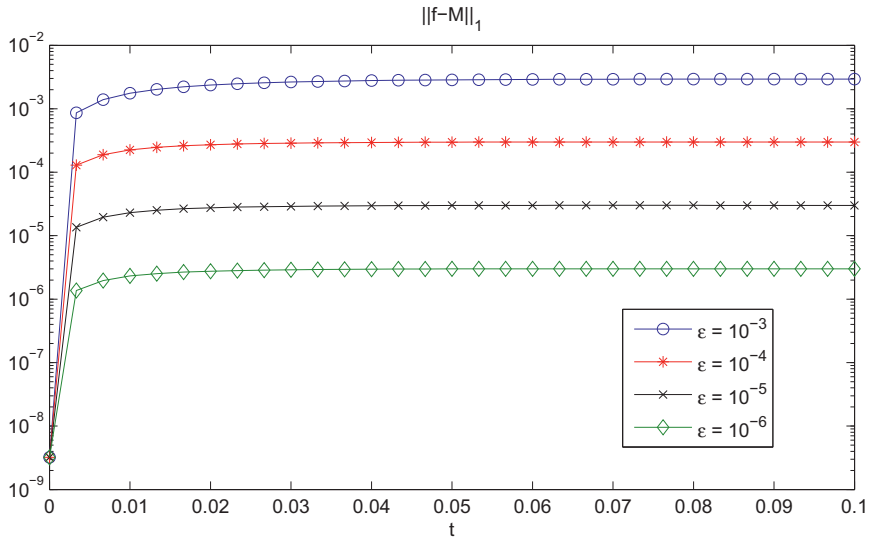



**Fig. 2.** The time evolution of $\|f - M\|_1$ for different $\epsilon$ with equilibrium initial data. The solutions are computed by the first order scheme. The mesh sizes are the same. $v \in [-6,6]^2$, $N_v = 64$, $x \in [-1,1]$, $N_x = 100$, $\Delta t = \Delta x/v_{\max}$.

### 4.2. The AP property

#### 4.2.1. The AP property for equilibrium initial data

We first demonstrate that the distribution $f$ would stay close to the equilibrium $M$, if initially it does. We apply the first order schemes (1.10) and (1.12) on the equilibrium initial data $f^l = M^l$, with the macroscopic variables given by

$$\rho^l = \frac{2+\sin \pi x}{3}, \quad u^l = 0, \quad T^l = \frac{3+\cos \pi x}{4} \tag{4.3}$$

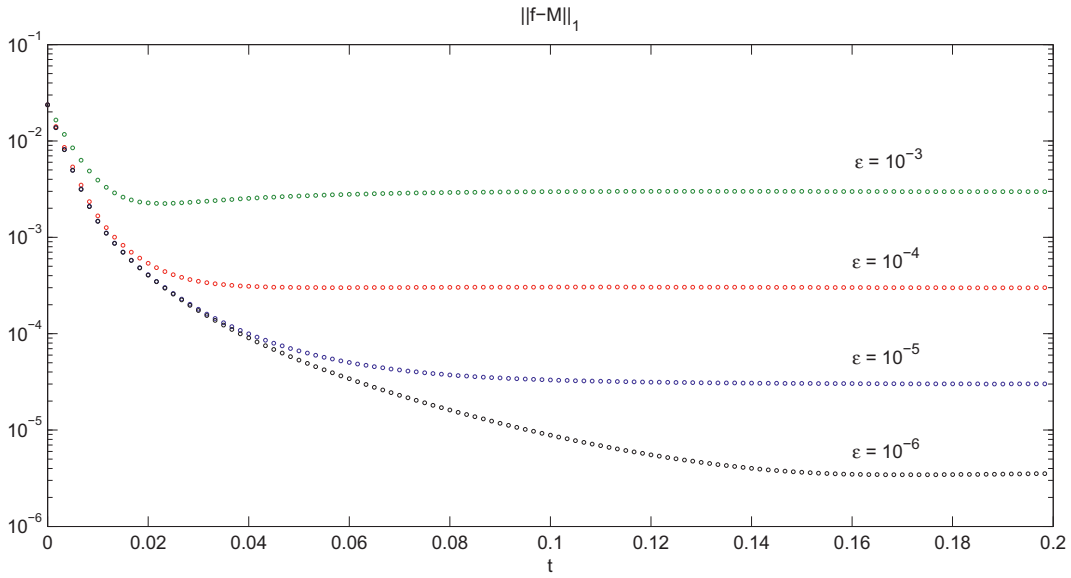
where $x \in [-1,1]$, $v \in [-6,6]^2$.

For different $\epsilon$, we show the time evolution of

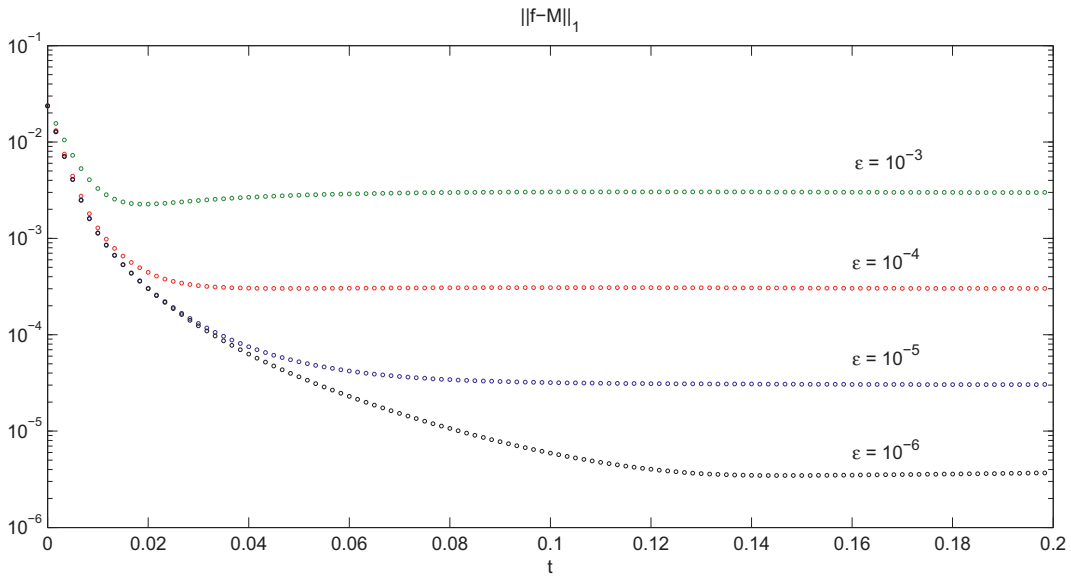
$$\|f - M\|_1 = \int \int |f - M| dx\, dv$$

(approximated by the trapezoidal rule). The results are shown in Fig. 2. As expected, $f^n - M^n = O(\epsilon)$ for all $n \geqslant 1$.

The second order IMEX scheme (1.13) gives similar results.




(a) First order scheme.




(b) Second order scheme.

**Fig. 3.** The time evolution of $\|f - M\|_1$ for different $\epsilon$ with non-equilibrium initial data. The solutions are computed by the first order scheme (a) and second order scheme (b), respectively. $v \in [-6,6]^2$, $N_v = 64$, $x \in [-1,1]$, $N_x = 100$, $\Delta t = \Delta x / v_{\max}$.

*4.2.2. The weakened AP property for non-equilibrium initial data*

Next we start with the "double peak" non-equilibrium initial data

$$f^I = \frac{\rho^I}{2\pi T^I}\cdot\frac{1}{2}\left(\exp\left(-\frac{(v-u^I)^2}{2T^I}\right)+\exp\left(-\frac{(v+u^I)^2}{2T^I}\right)\right), \tag{4.4}$$

where

$$\rho^I = \frac{2+\sin\pi x}{3}, \quad u^I = (0.2,0), \quad T^I = \frac{3+\cos\pi x}{4}. \tag{4.5}$$

The time evolutions of $\|f-M\|_1$ for different $\epsilon$ are shown in Fig. 3, with first order (circle) and second order (solid line) schemes. We have numerically shown that, for general initial data, the scheme is "weak" AP after transient steps, namely, $f^n - M^n = O(\epsilon)$ for $n$ sufficiently large. This is the weakened AP property. This behavior is similar to that in [16], where the classical Boltzmann equation is penalized by the BGK operator. Fig. 3 shows that the two schemes, with $\beta_0 = 1$ and $\beta_0 = \left(2+\sqrt{2}\right)$, need almost the same transient steps.

*4.2.3. The conservation of moments in solving the linear system (3.3)*

Next we show the CG method can preserve the moments well.

We use the "double-peak" non-equilibrium initial condition (4.4) with the macroscopic variables (4.5), where $x\in[-1,1]$, $v\in[-6,6]^2$. We take $N_x = 100$, while $N_v = 32, 64, 128$, respectively. Correspondingly $\Delta v = \frac{v_{\max}-vmin}{N_v} \approx 0.4, 0.2, 0.1$.

We use the first order scheme (1.10) with (1.11) for one step and compute the $l^1$ error in moments

$$Err(\phi) = \sum_x \left|\sum_v \left((f^1 - M^1)\phi\right)\right|\Delta v^2 \Delta x,$$

where $\phi = 1$, $v$, $|v|^2$, $M^1$ is computed from (2.11) while $f^1$ is obtained by solving (3.3) with a CG scheme.

The results are shown in Table 1. The moments are preserved very well when the CG scheme is applied to solve the linear system. The errors in moments are uniformly small in $\epsilon$. Besides, the conservations get improved on a finer grid in $v$.

*4.3. The Riemann problem*

Now we simulate the Sod shock tube problem, where the initial condition is $f^I = M^I$ with

$$\begin{cases} (\rho,u_1,T) = (1,0,1), & \text{if } -0.5 \leqslant x < 0, \\ (\rho,u_1,T) = (1/8,0,1/4), & \text{if } 0 \leqslant x \leqslant 0.5. \end{cases} \tag{4.6}$$

The Neumann boundary condition in the $x$-direction is applied.

In this test we take $x\in[0,1]$, $v\in[-6,6]^2$, $\epsilon = 0.001$. Numerical experiments show that $N_v = 32$ is enough for our simulation. We choose $N_x = 100$ and $\Delta t = \frac{\Delta x}{v_{\max}} \approx 5\times 10^{-3}$. We compare this under-resolved solution to a fully resolved solution by

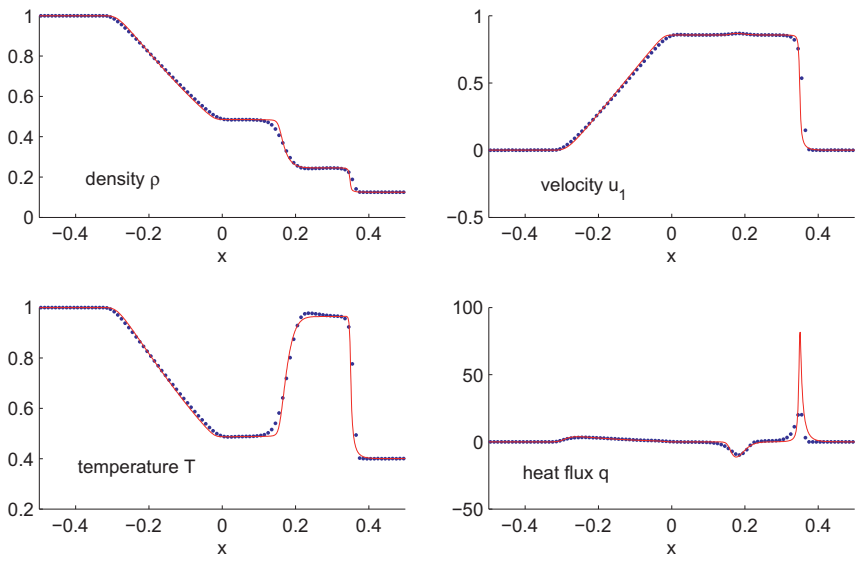



**Fig. 4.** The comparison of density, velocity, temperature and heat flux at $t = 0.2$ between the resolved computation by the explicit second order Runger–Kutta scheme (solid line) and the under-resolved solutions by the second order IMEX type AP scheme (dots).

the explicit second order Runger–Kutta scheme, where we take $N_x = 2000$ and $\Delta t = \min\left\{\frac{\Delta x}{v_{\max}}, \epsilon\Delta v^2\right\} \approx 10^{-5}$. We compute the macroscopic variable $\rho$, $u_1$, $T$ and $q$, where the heat flux $q$ is given by

$$q(t,x) = \frac{1}{\epsilon}\int_{\mathbb{R}^{N_v}} (v_1 - u_1)|v-u|^2 f(t,x,v)dv.$$

The results are compared at $t_{\max} = 0.2$ and shown in Fig. 4. One can see the macroscopic quantities are well approximated although the mesh size and time steps are much bigger than $\epsilon$, thus the computational cost has been reduced significantly than a fully resolved computation.

### 4.4. Mixing regimes

Now we consider the case where the Knudsen number $\epsilon$ increases smoothly from $\epsilon_0$ to $O(1)$, then jumps back to $\epsilon_0$,

$$\epsilon(x) = \begin{cases} \epsilon_0 + \frac{1}{2}(\tanh(5-10x) + \tanh(5+10x)), & x \leqslant 0.3, \\ \epsilon_0, & x > 0.3 \end{cases}$$

with $\epsilon_0 = 0.001$. The picture of $\epsilon$ is shown in Fig. 5. This problem involves mixed kinetic and fluid regimes.

To avoid the influence from the boundary, we take periodic boundary condition in $x$. The initial data are given by $f^I = M^I$, with the macroscopic quantities given by (4.3). Again we take $x \in [-1,1]$, $v \in [-6,6]^2$.

In this test we compare the macroscopic variable obtained by our new second order scheme (1.13) and the explicit Runger–Kutta scheme. For the explicit Runger–Kutta scheme, we take $N_x = 1000$, $\Delta t = \min\left\{\frac{\Delta x}{v_{\max}}, \frac{\epsilon_0}{\Delta v^2}\right\} \approx 10^{-5}$. For our scheme (1.13), we take $N_x = 100$, $\Delta t = \frac{\Delta x}{v_{\max}} = 5 \times 10^{-3}$. The results are compared up to $t_{\max} = 0.2$ in Fig. 6. Our scheme can capture the macroscopic behavior efficiently, with much larger mesh size and time steps.

### 4.5. The comparison on different penalization operators

This section is devoted to the comparison of the two different penalizing operators (1.9) and (2.16). We will show numerically that the classical diffusion operator (2.16) is not suitable to be the penalization.

#### 4.5.1. Trend to the equilibrium

First we show the convergence rate to equilibrium mentioned in Section 2.4. We start with the homogeneous equation

$$\frac{\partial}{\partial t} f = \beta Q(f)$$

with $Q(f)$ the nFPL operator, FP operator and classical diffusion operator, respectively. Here $\beta = 1$ for nFPL operator, and $\beta$ given by

$$\beta = \max_v \lambda(D_A(M))$$

for the FP and diffusion operators. We solve this equation by a second order midpoint scheme, with $\Delta t$ constrained by the CFL condition $\Delta t \sim \Delta v^2$.

The double peak shape initial data are used

$$f^I = \frac{\rho}{2\pi T}\frac{1}{2}\left(\exp\left(-\frac{(v-u)^2}{2T}\right) + \exp\left(-\frac{(v+u)^2}{2T}\right)\right)$$

with $\rho = 1$, $u = (1,0)$, $T = 0.2$. We take $v_{\max} = 16$, $N_v = 128$.

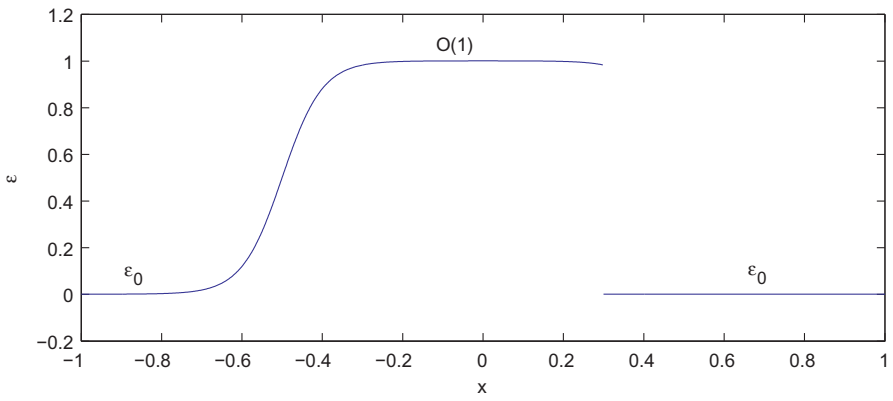



**Fig. 5.** An $x$-dependent $\epsilon(x)$.

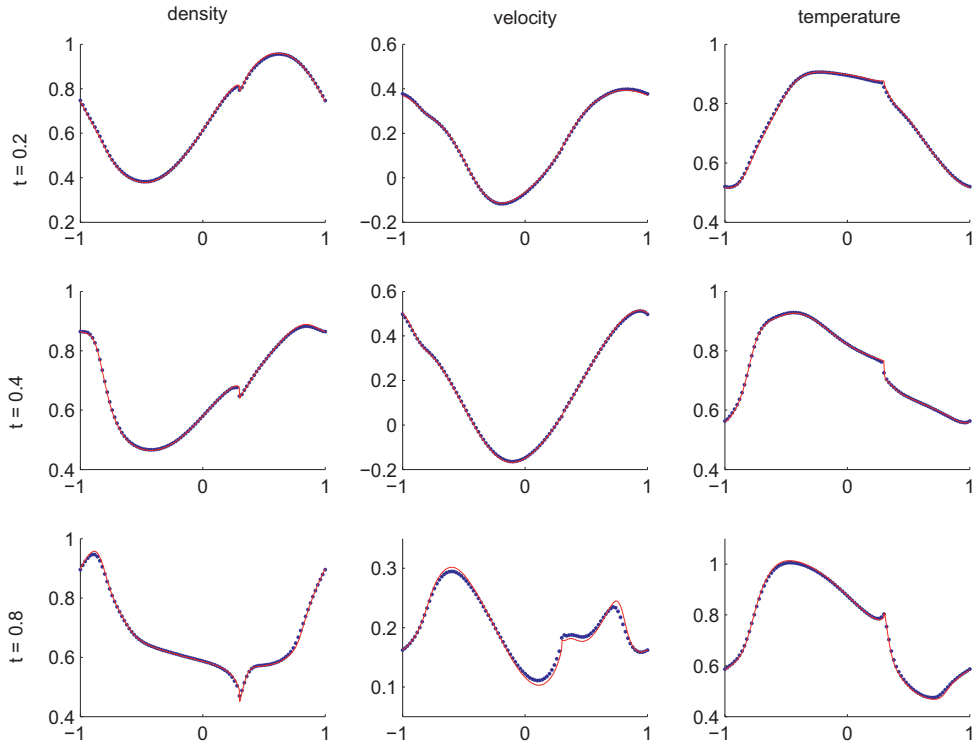



**Fig. 6.** For mixing regime, the comparison between the resolved solutions (solid line) given by the explicit Runger–Kutta scheme and the solutions (dots) obtained by our new scheme (1.10) with coarse grid and large time step.

As we mentioned in Section 2.4, the relative entropy decays exponentially for the FP operator and nFPL operator along the solutions. We show the relative entropy $H(f|M)$ in Fig. 7. The solution corresponding to FP operator has an exponential decay to the equilibrium. The solution with nFPL operator approaches equilibrium with a lower, but still exponential rate. The solution of classical diffusion gets to equilibrium with a polynomial rate ($t^{-1}$). It loses the control over the solution of nFPL as time evolves, even a large $\beta$ is used.

*4.5.2. Penalization on the homogeneous equation*

Next we check the behavior of numerical solutions when the two operators, FP and classical diffusion, are used as penalization.

We still work on the (rescaled) homogeneous equation and apply the first order scheme,

$$\frac{f^{n+1}-f^n}{\Delta t}=\frac{1}{\epsilon}\left(Q(f^n)-\beta P(f^n)+\beta P(f^{n+1})\right), \tag{4.7}$$

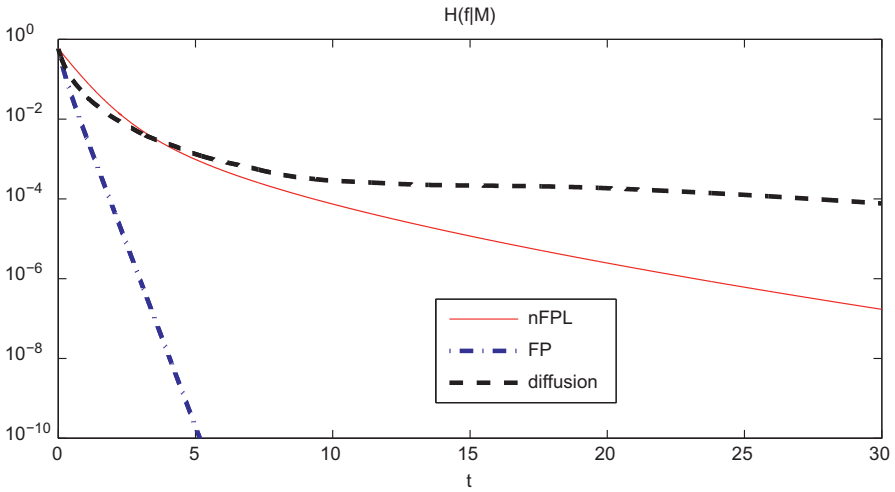



**Fig. 7.** The trend to equilibrium for the homogeneous equation with different operators. The evolution of relative entropy $H(f|M)=\int f\log\frac{f}{M}\,dv$ is plotted.

where $Q(f)$ is the nFPL operator and $P(f)$ is either the FP or the classical diffusion operator, and we take

$$\beta = \beta_0 \max_v \lambda\left(\int A(v - v_*) f_* dv_*\right).$$

The equilibrium initial data $f^I = M^I$ is used, with $v \in [-6,6]^2$, $N_v = 64$, and $\rho = 1$, $u = 0$, $T = 0.8$. We take $\epsilon = 10^{-6}$ and $\Delta t = 0.01$.

For $P(f)$ to be the FP operator, we take $\beta_0 = 1$. For $P(f)$ to be the classical diffusion operator, we take $\beta_0 = 2, 4, 6$. We compute the time evolution of $\|f - f_{true}\|_\infty$. Note that the true solution is just the steady state $f_{true} = M$ for all the time. The results are shown in Fig. 8. The solution derived when penalized by FP stays at equilibrium $f = M$, while the solution penalized by the classical diffusion deviates from the equilibrium very soon, whatever the choice of $\beta$.

This gives a direct numerical evidence that the classical diffusion cannot be used as penalization for the nFPL operator.

*4.5.3. Nonhomogeneous case*

Finally we move to the fully nonhomogeneous equation. We have numerically checked the AP property in Section 4.2, when the FP operator is used as penalization. Here we show the IMEX schemes (1.10) and (1.13) are not AP if penalized by the classical diffusion.

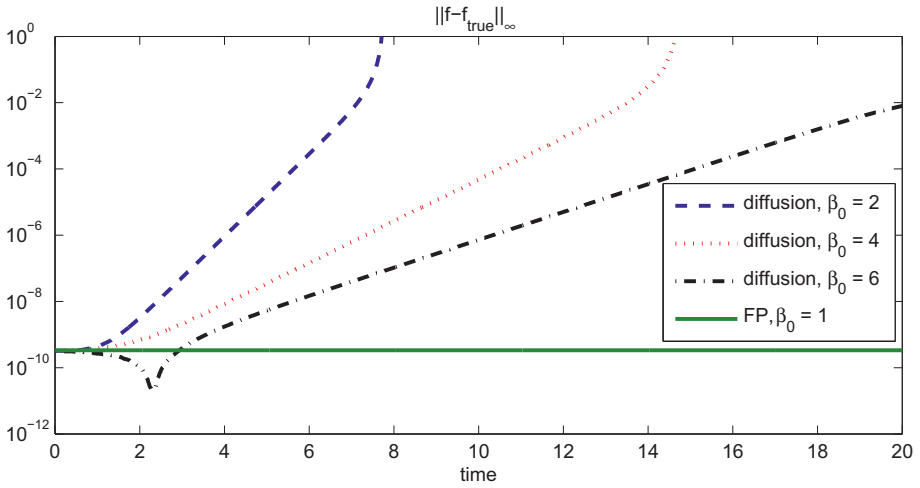



**Fig. 8.** The time evolution of the error $\|f - f_{true}\|_\infty$ when the homogeneous nFPL equation is penalized by the classical diffusion with different $\beta_0$ and the FP operator. $\epsilon = 10^{-6}$.

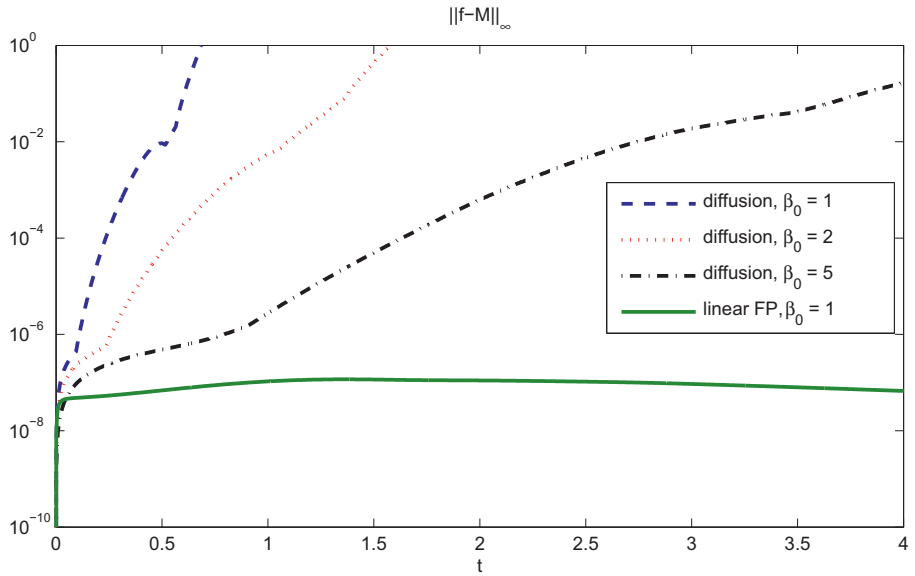



**Fig. 9.** The time evolution of $\|f - M\|_\infty$ when the non-homogenous nFPL equation is penalized by the classical diffusion with different $\beta_0$ and FP operator. $\epsilon = 10^{-6}$.

The equilibrium initial data $f^I = M^I$ is considered, with the macroscopic quantities given by (4.3). We take $x \in [-1,1]$, $N_x = 100$, $v \in [-6,6]^2$, $N_v = 64$, $\Delta t = \Delta x/v_{\max}$.

We take $P(f)$ in (1.10) to be the classical diffusion operator. We compare the results with the one obtained when penalized by the FP operator. The comparison is shown in Fig. 9. We give the time evolution of $\|f - M\|_\infty$ for the scheme with the FP ($\beta_0 = 1$ in solid lines) and with the classical diffusion with different $\beta_0$. The simulation shows, if the classical diffusion operator is used as penalization, $f$ would get away from $M$ even if initially they are close. A larger $\beta$ can decelerate this departing. But after long time we always get $f - M \sim O(1)$. Therefore the scheme is not AP anymore.

## 5. Conclusion

A penalization based asymptotic-preserving scheme for the nonlinear Fokker–Planck–Landau (nFPL) equation has been introduced in this article. The basic idea comes from the BGK-penalization for the classical Boltzmann equation studied by Filbet and Jin [16]. However the diffusive nature of nFPL operator makes the BGK operator not suitable as the penalization term. We use the (linear) Fokker–Planck (FP) operator as the penalization instead. The FP operator possesses the good properties of collision operator, such as the conservation of moments and entropy dissipation. Besides, the FP operator, which also contains a diffusive term, can overcome the stiffness in the nFPL operator. To solve the linear system involving FP operator implicitly, we introduce a central discretization and derive a symmetric matrix, therefore a Conjugate Gradient scheme can be applied easily. Several numerical experiments are also carried out to verify the performance of the new scheme for different regimes and its AP property.

The case of particles interacting through Coulomb potential is studied. However the scheme can apply to other cases (e.g. the Maxwell potential) without any difficulties.

The boundary conditions are beyond the scope of this paper. There are very few studies on AP schemes in this direction except [23,24]. It is an important subject for future research.

We numerically verified our scheme is AP beyond the initial transient layer. However the theoretical analysis for our scheme is still lacking and is a subject of future research.

## Acknowledgment

The authors thank Professor Francis Filbet for the fruitful discussions about this work.

## Appendix A. Proof of Theorem 2.1

We take $\nabla u$ as a column vector. Its transpose is written as $(\nabla u)^T$.

Multiply (2.2) by $u^{n+1}$ on both sides, then integrate over $x$. For the left side, we use $u^{n+1} = \frac{1}{2}\left((u^{n+1} + u^n) + (u^{n+1} - u^n)\right)$. For the right side, we apply the integration by parts. Then

$$\frac{\|u^{n+1}\|_2^2 - \|u^n\|_2^2}{2\Delta t} + \frac{\|u^{n+1} - u^n\|_2^2}{2\Delta t} = \frac{1}{\epsilon}\int \left[(\nabla u^{n+1})^T(\beta I - A(u^n, x))\nabla u^n\right]dx - \frac{\beta}{\epsilon}\int \left[\nabla u^{n+1} \cdot \nabla u^{n+1}\right]dx, \tag{5.1}$$

where $I$ is $N \times N$ identity matrix and $\|\cdot\|_2$ is the regular $L^2$ norm.

While for a symmetric matrix $P$, we have the following inequality holds,

$$x^T P y \leqslant \frac{1}{2}\lambda(x^T x + y^T y)$$

with $\lambda$ the spectral radius of $P$. One can easily show this by first diagonalizing $P$ and then applying the Cauchy–Schwarz inequality.

Then

$$(\nabla u^{n+1})^T(\beta I - A(u^n, x))\nabla u^n \leqslant \frac{1}{2}\max|\beta - \lambda(A)|\left(|\nabla u^{n+1}|^2 + |\nabla u^n|^2\right) \leqslant \frac{1}{2}\beta\left(|\nabla u^{n+1}|^2 + |\nabla u^n|^2\right).$$

The last inequality follows from the condition (2.3).

Hence

$$\frac{\|u^{n+1}\|_2^2 - \|u^n\|_2^2}{\Delta t} + \frac{\|u^{n+1} - u^n\|_2^2}{\Delta t} \leqslant \frac{\beta}{\epsilon}\int \left(|\nabla u^n|^2 + |\nabla u^{n+1}|^2\right)dx - \frac{2\beta}{\epsilon}\int |\nabla u^{n+1}|^2 dx = -\frac{\beta}{\epsilon}\int \left(|\nabla u^{n+1}|^2 - |\nabla u^n|^2\right)dx.$$

Therefore, the total energy of $u$ by

$$E(u) = \int \left(u^2 + \Delta t\frac{\beta}{\epsilon}|\nabla u|^2\right)dx, \tag{5.2}$$

satisfies the energy dissipation

$$E(u^{n+1}) - E(u^n) \leqslant -\|u^{n+1} - u^n\|_2^2 \leqslant 0.7. \qquad \square$$

## References


[1] O.E. Buryak, A.A. Arsen'ev, On the connection between a solution of the Boltzmann equation and a solution of the Fokker–Planck–Landau equation, Mathematics of the USSR-Sbornik 69 (1991) 465–478.
[2] Yu. A. Berezin, V.N. Khudick, M.S. Pekker, Conservative finite-difference schemes for the Fokker–Planck equation not violating the law of an increasing entropy, Journal of Computational Physics 69 (1) (1987) 163–174.
[3] A.L. Bertozzi, N. Ju, H.W. Lu, A biharmonic-modified forward time stepping method for fourth order nonlinear diffusion equations, Discrete and Continuous Dynamical Systems 29 (4) (2011) 1367–1391.
[4] C. Buet, S. Cordier, Conservative and entropy decaying numerical scheme for the isotropic Fokker–Planck–Landau equation, Journal of Computational Physics 145 (1) (1998) 228–245.
[5] C. Buet, S. Cordier, Numerical analysis of conservative and entropy schemes for the Fokker–Planck–Landau equation, SIAM Journal on Numerical Analysis 36 (3) (1999) 953–973.
[6] C. Buet, S. Cordier, P. Degond, M. Lemou, Fast algorithms for numerical, conservative, and entropy approximations of the Fokker–Planck–Landau equation, Journal of Computational Physics 133 (2) (1997) 310–322.
[7] C. Buet, S. Dellacherie, On the Chang and Cooper scheme applied to a linear Fokker–Planck equation, Communications in Mathematical Sciences 8 (4) (2010) 1079–1090.
[8] J.A. Carrillo, G. Toscani, Exponential convergence toward equilibrium for homogeneous Fokkerc–Planck-type equations, Mathematical Methods in the Applied Sciences 21 (1998) 1269–1286.
[9] J.S. Chang, G. Cooper, A practical difference scheme for Fokker–Planck equations, Journal of Computational Physics 6 (1) (1970) 1–16.
[10] P. Degond, B. Lucquin-Desreux, The Fokker–Planck asymptotics of the Boltzmann collision operator in the Coulomb case, Mathematical Models and Methods in Applied Sciences (M3AS) 2 (2) (1992) 167–182.
[11] P. Degond, B. Lucquin-desreux, An entropy scheme for the Fokker–Planck collision operator of plasma kinetic theory, Numerische Mathematik 68 (1994) 239–262.
[12] L. Desvillettes, On asymptotics of the Boltzmann equation when the collisions become grazing, Transport Theory and Statistical Physics 21 (3) (1992) 259–276.
[13] L. Desvillettes, C. Villani, On the spatially homogeneous Landau equation for hard potentials. Part II: Htheorem and applications, Communications in Partial Differential Equations 25 (1) (2000) 261–298.
[14] L. Desvillettes, C. Villani, On the trend to global equilibrium in spatially inhomogeneous entropy-dissipating systems. Part I: The linear Fokker–Planck equation, Communications on Pure and Applied Mathematics 54 (2001) 1–42.
[15] E.M. Epperlein, Implicit and conservative difference scheme for the Fokker–Planck equation, Journal of Computational Physics 112 (2) (1994) 291–297.
[16] F. Filbet, S. Jin, A class of asymptotic-preserving schemes for kinetic equations and related problems with stiff sources, Journal of Computational Physics 229 (20) (2010) 7625–7648.
[17] F. Filbet, L. Pareschi, A numerical method for the accurate solution of the Fokker–Planck–Landau equation in the nonhomogeneous case, Journal of Computational Physics 179 (1) (2002) 1–26.
[18] T. Goudon, Private communication.
[19] T. Goudon, On boltzmann equations and Fokker–Planck asymptotics: Influence of grazing collisions, Journal of Statistical Physics 89 (3–4) (1997) 751–776.
[20] L. Gross, Logarithmic Sobolev inequalities, American Journal of Mathematics 97 (1975) 1061–1083.
[21] S. Jin, Asymptotic preserving (ap) schemes for multiscale kinetic and hyperbolic equations: a review, Lecture Notes for Summer School on "Methods and Models of Kinetic Theory" (M&MKT), Porto Ercole (Grosseto, Italy), June 2010. Rivista di Matematica della UniversitÃ di Parma, submitted for publication.
[22] S. Jin, Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations, SIAM Journal on Scientific Computing 21 (2) (1999) 441–454.
[23] S. Jin, D. Levermore, The discrete-ordinate method in diffusive regimes, Transport Theory and Statistical Physics 20 (5) (1991) 413–439.
[24] S. Jin, D. Levermore, Fully-discrete numerical transfer in diffusive regimes, Transport Theory and Statistical Physics 22 (6) (1993) 739–791.
[25] L.D. Landau, Die kinetische gleichung für den fall Coulombscher vechselwirkung, Physik Zeitsch der Sowjetunion 154 (1963).
[26] L.D. Landau, The transport equation in the case of the Coulomb interaction, in: D. Ter Haar (Ed.), Collected Papers of L.D. Landau, Pergamon press, Oxford, 1981, pp. 163–170.
[27] E.W. Larsen, C.D. Levermore, G.C. Pomraning, J.G. Sanderson, Discretization methods for one-dimensional Fokker–Planck operators, Journal of Computational Physics 61 (3) (1985) 359–390.
[28] M. Lemou, L. Mieussens, Implicit schemes for the Fokker–Planck–Landau equation, SIAM Journal on Scientific Computing 27 (3) (2005) 809–830.
[29] R.J. LeVeque, Numerical Methods for Conservation Laws, Birkhauser-Verlag, Basel, 1990.
[30] P.A. Markowich, C. Villani, On the trend to equilibrium for the Fokker–Planck equation: An interplay between physics and functional analysis, Matematica Contemporanea (SBM) 19 (2000) 1–31.
[31] L. Pareschi, B. Perthame, A Fourier spectral method for homogeneous Boltzmann equations, Transport Theory and Statistical Physics 25 (3) (1996) 369–382.
[32] L. Pareschi, G. Russo, G. Toscani, Fast spectral methods for the Fokker–Planck–Landau collision operator, Journal of Computational Physics 165 (1) (2000) 216–236.
[33] P. Smereka, Semi-implicit level set methods for curvature and surface diffusion motion, Journal of Scientific Computing 19 (1–3) (2003) 439–456.
[34] A.J. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon, Information and Control 2 (2) (1959) 101–112.
[35] R. Strain, Y. Guo, Exponential decay for soft potentials near Maxwellian, Archive for Rational Mechanics and Analysis 187 (2008) 287–339, doi:10.1007/s00205-007-0067-3.
[36] G. Toscani, C. Villani, On the trend to equilibrium for some dissipative systems with slowly increasing a priori bounds, Jornal of Statistical Physics 98 (5) (2000) 1279–1309.
[37] G. Toscani, C. Villani, Sharp entropy dissipation bounds and explicit rate of trend to equilibrium for the spatially homogeneous Boltzmann equation, Communications in Mathematical Physics 203 (1999) 667–706, doi:10.1007/s002200050631.
[38] Cédric Villani, A Review of Mathematical Topics in Collisional Kinetic Theory, Handbook of Mathematical Fluid Dynamics, vol. 1, North-Holland, 2002. pp. 71–74.
[39] J.C. Whitney, Finite difference methods for the Fokker–Planck equation, Journal of Computational Physics 6 (3) (1970) 483–509.