A second-order asymptotic-preserving and positivity-preserving exponential Runge-Kutta method for a class of stiff kinetic equations

Jingwei Hu† and Ruiwen Shu‡

July 11, 2018

Abstract

We introduce a second-order time discretization method for stiff kinetic equations. The method is asymptotic-preserving (AP) – can capture the Euler limit without numerically resolving the small Knudsen number; and positivity-preserving – can preserve the non-negativity of the solution which is a probability density function for arbitrary Knudsen numbers. The method is based on a new formulation of the exponential Runge-Kutta method and can be applied to a large class of stiff kinetic equations including the BGK equation (relaxation type), the Fokker-Planck equation (diffusion type), and even the full Boltzmann equation (nonlinear integral type). Furthermore, we show that when coupled with suitable spatial discretizations the fully discrete scheme satisfies an entropy-decay property. Various numerical results are provided to demonstrate the theoretical properties of the method.

Key words. stiff kinetic equation, exponential Runge-Kutta method, asymptotic-preserving, positivity-preserving, entropy-decay

AMS subject classifications. 82C40, 65L04, 35Q31, 35Q84, 35Q20, 65L06, 65F60

1 Introduction

Kinetic equations describe the non-equilibrium dynamics of a gas or system comprised of a large number of particles. In multiscale modeling hierarchy, they serve as a bridge that connects microscopic Newtonian mechanics and macroscopic continuum mechanics. In this paper, we are concerned with the following class of kinetic equations:

\begin{equation}
\frac{\partial}{\partial t} f + v \cdot \nabla_x f = \frac{1}{\varepsilon} Q(f), \quad t \geq 0, \quad x \in \Omega \subset \mathbb{R}^d, \quad v \in \mathbb{R}^d,
\end{equation}

where \( f = f(t, x, v) \) is the one-particle probability density function (PDF) of time \( t \), position \( x \), and particle velocity \( v \). \( Q \) is the collision operator which acts only in the velocity space.

---

*This research was supported by NSF grant DMS-1620250 and NSF CAREER grant DMS-1654152. Support from DMS-1107291: RNMS KI-Net is also gratefully acknowledged.

†Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA (jingweihu@purdue.edu).

‡Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706, USA (rshu2@wisc.edu).
and models the interactions between particles. Examples of $Q$ include: the Boltzmann collision operator (a nonlinear integral operator) \cite{Q}, the BGK operator (a relaxation type operator) \cite{BGK}, the kinetic Fokker-Planck operator (a diffusion type operator) \cite{Fokker-Planck}, among others. Finally, $\varepsilon$ is the Knudsen number defined as the ratio of the mean free path and typical length scale. The magnitude of $\varepsilon$ indicates the degree of rarefaction of the system. When $\varepsilon$ is small, collisions happen very frequently so that the system is close to the fluid regime. In fact, one can derive the compressible Euler equations from (1.1) as the leading-order asymptotics by sending $\varepsilon \to 0$.

When $\varepsilon$ is small, numerically solving the equation (1.1) is challenging due to the stiff collision term on the right hand side. Any explicit time discretization would suffer from severe stability constraint (time step $\Delta t$ has to be $O(\varepsilon)$). As such, schemes that can remove this constraint are highly desirable. The so-called asymptotic-preserving (AP) scheme \cite{AP} is exactly designed for this kind of problems: it solves the kinetic equation without resolving small scales ($\Delta t$ can be chosen independent of $\varepsilon$), yet when $\varepsilon \to 0$ while keeping $\Delta t$ fixed, it automatically becomes a macroscopic fluid solver, i.e., a consistent discretization to the limiting Euler equations (see \cite{AP-review1, AP-review2} for a comprehensive review of AP schemes).

The AP property is certainly a desired feature for handling multiscale kinetic equations, especially in the near fluid regime. However, most of AP schemes require some implicit treatment or reformulation of the equation such that the positivity of the solution is lost during the construction. This is unphysical since $f$ is a PDF, and sometimes even causes the simulation to break down. The design of high order (at least second order) schemes that are both AP and positivity-preserving turns out to be highly nontrivial and needs to be handled in a problem-dependent basis. Recently, we developed a family of second-order AP and positivity-preserving schemes for the stiff BGK equation \cite{BGK-AP}. The method is based on the implicit-explicit (IMEX) Runge-Kutta framework plus a key correction step utilizing the special structure of the BGK operator. It also works for some hyperbolic systems but is limited to relaxation type operators.

In this paper, we propose a more general time discretization method based on a new exponential Runge-Kutta formulation that can be applied to a large class of stiff kinetic equations including the BGK, the Fokker-Planck, and even the full Boltzmann equations. To summarize, our method possesses the following features:

- The scheme is second-order accurate in the kinetic regime $\varepsilon = O(1)$;
- The scheme is AP: for fixed $\Delta t$, when $\varepsilon \to 0$, it reduces to a second-order scheme for the limiting Euler equations (in fact, the limiting scheme can be made as the optimal second-order strong-stability-preserving (SSP) Runge-Kutta method, i.e., the improved Euler or Heun’s method \cite{SSP-RK, Heun});
- The scheme is positivity-preserving for any $\varepsilon \geq 0$: if $f^n \geq 0$, then $f^{n+1} \geq 0$;
- The time step of the scheme is only constrained by the transport part and can be chosen the same as in the forward Euler method;
- The scheme satisfies an entropy-decay property when coupled with suitable spatial discretizations.

The rest of this paper is organized as follows. In Section 2, we construct the method for the general kinetic equation (1.1) without specifying the collision operator. The emphasis is to make
the method second order and positivity-preserving. In Section 3 we consider the application of the method to specific kinetic equations and discuss its AP property. A comparison with existing similar methods is given as well. In Section 4 we address the issue of solving the homogeneous equation (1.1) without transport term) which is an important building block of the proposed method. In Section 5, we prove the entropy-decay property of the method when coupled with suitable spatial discretizations. Some remarks regarding the spatial and velocity domain discretizations are given in Section 6. Numerical examples are presented in Section 7. The paper is concluded in Section 8.

2 A new exponential Runge-Kutta method for general stiff kinetic equations

We now present the procedure to construct the new exponential Runge-Kutta method. Since the method is quite general and can be applied to a large class of kinetic equations, we will start with the equation (1.1) without specifying the collision operator and derive a scheme that is both second-order accurate and positivity-preserving. Then in Section 3 we will consider specific collision operators and discuss the AP property of the scheme as this latter part is problem dependent.

To begin with, let us introduce the following notation: for the autonomous ODE

$$\frac{d}{dt} f = A(f), \quad f|_{t=t_0} = g, \quad (2.1)$$

where $A$ is an operator, either linear or nonlinear, we use $\exp(sA)g$, $s \geq 0$ to represent its solution at $t = t_0 + s$.

We now consider an ODE resulting from the semi-discretization of the equation (1.1) (only space $x$ is discretized while time $t$ and velocity $v$ are left continuous):

$$\frac{d}{dt} f = T(f) + \frac{1}{\epsilon} Q(f). \quad (2.2)$$

Here $T(f)$ is a discretized operator for the transport term $-v \cdot \nabla_x f$ and $Q(f)$ is the collision operator which may take various forms depending on the application. We assume the operators $T(f)$ and $Q(f)$ are positivity-preserving. To be precise,

- for $T(f)$, we assume

  \[ f \geq 0 \implies f + a\Delta t T(f) \geq 0, \quad \forall \text{ constant } a \text{ s.t. } 0 \leq a\Delta t \leq \Delta t_{FE}, \quad (2.3) \]

  where $\Delta t_{FE}$ is the maximum time step allowance such that the forward Euler method is positivity-preserving;

- for $Q(f)$, we assume the solution to the homogeneous equation

  \[ \frac{d}{dt} f = Q(f) \quad (2.4) \]

  satisfies $f \geq 0$ for all $t \geq t_0$, if the initial data $f|_{t=t_0} = g \geq 0$. In other words,

  \[ g \geq 0 \implies \exp(sQ)g \geq 0, \quad \forall \text{ constant } s \geq 0. \quad (2.5) \]

1Note that $\exp(sA)$ is merely a symbol to denote the map from the solution at time $t_0$ to the solution at time $t_0 + s$, and should not be understood as the matrix exponential except the linear case.
Remark 2.1. The condition (2.3) can be easily satisfied if a positivity-preserving spatial discretization is used, as was done in [13]. The condition (2.5) is a theoretical property that holds for any kinetic equations.

We are ready to construct the numerical method for equation (2.2). For the time being, we assume that the solution to the homogeneous equation (2.4) can be found exactly and will get back to this in Section 4 when discussing specific models. We propose an exponential Runge-Kutta scheme of the following form:

\[
\begin{align*}
    f^{(0)} &= \exp \left( a_0 \Delta t \frac{1}{\varepsilon} Q \right) f^n, \\
    f^{(1)} &= \exp \left( a_1 \Delta t \frac{1}{\varepsilon} Q \right) \left( f^{(0)} + b_1 \Delta t T(f^{(0)}) \right), \\
    f^{(2)} &= f^{(1)} + b_2 \Delta t T(f^{(1)}), \\
    f^{n+1} &= \exp \left( a_2 \Delta t \frac{1}{\varepsilon} Q \right) \left[ w f^{(2)} + (1-w) \exp \left( (1-a_2) \Delta t \frac{1}{\varepsilon} Q \right) f^n \right],
\end{align*}
\]

where the constants \( a_0, a_1, a_2, b_1, b_2, \) and \( w \) are to be determined.

With the previous assumptions on \( T \) and \( Q \), it is easy to see

**Proposition 2.2.** The scheme (2.6) is positivity-preserving, i.e., if \( f^n \geq 0 \), then \( f^{n+1} \geq 0 \) provided

\[
a_0, a_1, b_1, b_2 \geq 0, \quad 0 \leq a_2, w \leq 1,
\]

under the CFL condition

\[
\Delta t \leq \frac{\Delta t_{FE}}{\max(b_1, b_2)},
\]

and the ratio is understood as infinite if the denominator is zero.

We next derive the conditions for (2.6) to be second order in the kinetic regime. Without loss of generality, we assume \( \varepsilon = 1 \).

First of all, given the solution \( f^n = f(t_n) \), if we Taylor expand the exact solution of (2.2) at \( t_{n+1} \) around \( t_n \), we have

\[
\exp(\Delta t(T + Q))f^n = f^n + \Delta t(T(f^n) + Q(f^n)) + \frac{1}{2} \Delta t^2 (T'(f^n)T(f^n) + T'(f^n)Q(f^n) + Q'(f^n)T(f^n) + Q'(f^n)Q(f^n)) + O(\Delta t^3),
\]

where \( Q', T' \) are the Fréchet derivative of \( Q \) and \( T \).

Similarly the exact solution of (2.4) at \( t_{n+1} \) is approximated by

\[
\exp(\Delta tQ)f^n = f^n + \Delta tQ(f^n) + \frac{1}{2} \Delta t^2 Q'(f^n)Q(f^n) + O(\Delta t^3).
\]

Using this in the first equation of (2.6), we have

\[
f^{(0)} = f^n + a_0 \Delta t Q(f^n) + \frac{1}{2} a_0^2 \Delta t^2 Q'(f^n)Q(f^n) + O(\Delta t^3).
\]
Continuing the Taylor expansion of $f^{(1)}$, $f^{(2)}$, and $f^{n+1}$ in (2.6), we have

\begin{align*}
f^{(1)} &= f^{(0)} + b_1 \Delta t T(f^{(0)}) + a_1 \Delta t Q(f^{(0)}) + b_1 \Delta t T(f^{(0)}) \\nonumber
&\quad + \frac{1}{2} a_1^2 \Delta t^2 Q'(f^{(0)}) + b_1 \Delta t T(f^{(0)}) + O(\Delta t^3) \\
&= f^n + \Delta t((a_0 + a_1)Q(f^n) + b_1 T(f^n)) \nonumber
&\quad + \Delta t^2 \left( b_1 a_0 T'(f^n)Q(f^n) + a_1 b_1 Q'(f^n)T(f^n) + \frac{1}{2}(a_0 + a_1)^2 Q'(f^n)Q(f^n) \right) + O(\Delta t^3).
\end{align*}
(2.12)

\begin{align*}
f^{(2)} &= f^{(1)} + b_2 \Delta t T(f^{(1)}) \\
&= f^n + \Delta t((a_0 + a_1)Q(f^n) + b_1 T(f^n)) \nonumber
&\quad + \Delta t^2 \left( b_1 a_0 T'(f^n)Q(f^n) + a_1 b_1 Q'(f^n)T(f^n) + \frac{1}{2}(a_0 + a_1)^2 Q'(f^n)Q(f^n) \right) \\
&\quad + b_2 \Delta t T(f^n) + b_2 \Delta t^2 T'(f^n)((a_0 + a_1)Q(f^n) + b_1 T(f^n)) + O(\Delta t^3) \\
&= f^n + \Delta t((a_0 + a_1)Q(f^n) + (b_1 + b_2) T(f^n)) \nonumber
&\quad + \Delta t^2 \left( b_1 b_2 T'(f^n)T(f^n) + (b_1 a_0 + b_2 a_0) T'(f^n)Q(f^n) \right) \\
&\quad + a_1 b_1 Q'(f^n)T(f^n) + \frac{1}{2}(a_0 + a_1)^2 Q'(f^n)Q(f^n) + O(\Delta t^3).
\end{align*}
(2.13)

\begin{align*}
w f^{(2)} + (1 - w)\exp((1 - a_2)\Delta t Q) f^n &= w f^n + \Delta t((a_0 + a_1)Q(f^n) + (b_1 + b_2) T(f^n)) \nonumber
&\quad + \Delta t^2 \left( b_1 b_2 T'(f^n)T(f^n) + (b_2 a_1 + b_1 a_0 + b_2 a_0) T'(f^n)Q(f^n) \right) \\
&\quad + a_1 b_1 Q'(f^n)T(f^n) + \frac{1}{2}(a_0 + a_1)^2 Q'(f^n)Q(f^n) \nonumber
&\quad + (1 - w) \left[ f^n + (1 - a_2)\Delta t Q(f^n) + \frac{1}{2}(1 - a_2)^2 \Delta t^2 Q'(f^n)Q(f^n) \right] + O(\Delta t^3) \\
&= f^n + \Delta t[(w(a_0 + a_1) + (1 - w)(1 - a_2))Q(f^n) + w(b_1 + b_2) T(f^n)] + \Delta t^2 [w b_1 b_2 T'(f^n)T(f^n) \nonumber
&\quad + w(b_2 a_1 + b_1 a_0 + b_2 a_0) T'(f^n)Q(f^n) + wa_1 b_1 Q'(f^n)T(f^n) \nonumber
&\quad + \frac{1}{2}(w(a_0 + a_1)^2 + (1 - w)(1 - a_2)^2) Q'(f^n)Q(f^n)] + O(\Delta t^3).
\end{align*}
(2.14)

Finally,

\begin{align*}
f^{n+1} &= f^n + \Delta t[(w(a_0 + a_1) + (1 - w)(1 - a_2))Q(f^n) + w(b_1 + b_2) T(f^n)] + \Delta t^2 [w b_1 b_2 T'(f^n)T(f^n) \nonumber
&\quad + w(b_2 a_1 + b_1 a_0 + b_2 a_0) T'(f^n)Q(f^n) + wa_1 b_1 Q'(f^n)T(f^n) \nonumber
&\quad + \frac{1}{2}(w(a_0 + a_1)^2 + (1 - w)(1 - a_2)^2) Q'(f^n)Q(f^n)] \\
&\quad + a_2 \Delta t Q(f^n) + \Delta t Q'(f^n)((w(a_0 + a_1) + (1 - w)(1 - a_2))Q(f^n) + w(b_1 + b_2) T(f^n))] \nonumber
&\quad + \frac{1}{2} a_2^2 \Delta t^2 Q'(f^n)Q(f^n) + O(\Delta t^3) \\
&= f^n + \Delta t[(w(a_0 + a_1 + a_2) + (1 - w))Q(f^n) + w(b_1 + b_2) T(f^n)] + \Delta t^2 [w b_1 b_2 T'(f^n)T(f^n) \nonumber
&\quad + w(b_2 a_1 + b_1 a_0 + b_2 a_0) T'(f^n)Q(f^n) + w(a_1 b_1 + a_2 b_2 + a_2 b_1) Q'(f^n)T(f^n) \nonumber
&\quad + \frac{1}{2}(w(a_0 + a_1 + a_2)^2 + (1 - w)) Q'(f^n)Q(f^n)] + O(\Delta t^3).
\end{align*}
(2.15)
Comparing (2.9) and (2.15), we arrive at the following order conditions:

\[ w(a_0 + a_1 + a_2) + (1 - w) = 1; \quad w(b_1 + b_2) = 1; \quad wb_1b_2 = \frac{1}{2}; \]
\[ w(b_2a_1 + b_1a_0 + b_2a_0) = \frac{1}{2}; \quad w(a_1b_1 + a_2b_2 + a_2b_1) = \frac{1}{2}; \]
\[ w(a_0 + a_1 + a_2)^2 + (1 - w) = 1. \]  

(2.16)

Further simplification yields

**Proposition 2.3.** The scheme (2.6) is second-order accurate for \( \varepsilon = O(1) \) provided

\[ a_0 + a_1 + a_2 = 1, \]  
\[ w(b_1 + b_2) = 1, \]  
\[ wb_1b_2 = \frac{1}{2}, \]  
\[ w(b_2a_1 + (b_1 + b_2)a_0) = \frac{1}{2}. \]  

(2.17) (2.18) (2.19) (2.20)

Combining the positivity conditions and order conditions found in Propositions 2.2 and 2.3, one can obtain a second-order positivity-preserving scheme for equation (2.2). To find a set of parameters satisfying these conditions, first notice that (2.18) and (2.19) imply \( b_1, b_2 \) are the solutions of the quadratic equation

\[ b^2 - \frac{1}{w}b + \frac{1}{2w} = 0, \]  

(2.21)

whose solutions are given by

\[ b_{1,2} = \frac{1}{1 \pm \sqrt{1 - 2w}}, \quad \text{for } 0 < w \leq \frac{1}{2}. \]  

(2.22)

In order to obtain the best CFL condition (minimize \( \max(b_1, b_2) \) in (2.8)), we choose

\[ w = \frac{1}{2}, \quad b_1 = b_2 = 1, \]  

(2.23)

hence the CFL condition (2.8) is the same as the forward Euler method. Then (2.17) and (2.20) reduce to

\[ a_0 + a_1 + a_2 = 1, \quad a_0 = a_2. \]  

(2.24)

To insure positivity, we only need additionally \( a_0, a_1 \geq 0, 0 \leq a_2 \leq 1 \) (see (2.7)). However, to obtain a good AP property, we require

\[ a_0, a_1 > 0, \quad 0 < a_2 < 1. \]  

(2.25)

This will be further elaborated in Section 3. One choice of \( a_0, a_1, a_2 \) is

\[ a_0 = a_1 = a_2 = \frac{1}{3}. \]  

(2.26)

**Remark 2.4.** If one sets \( Q = 0 \), then (2.6) becomes an explicit second-order SSP Runge-Kutta scheme applied to the purely transport equation; moreover, our choice (2.23) just gives the standard optimal one, i.e., the improved Euler or Heun’s method. In what follows, we will refer this scheme as SSP-RK2. If one sets \( T = 0 \), then (2.6) becomes \( f^{n+1} = \exp(\Delta t F_Q) f^n \) which is the exact solution to the homogeneous equation (2.4).
Remark 2.5. For $a_0, a_1 > 0$ and $0 < a_2 < 1$, (2.6) would require 4 times evaluation of the operator $\exp(sQ)$ in each time step. However, similar to the Strang splitting, one can combine the operator $\exp(a_2\Delta t^1 Q)$ in the last stage of the $n$-th step with the operator $\exp(a_0\Delta t^1 Q)$ in the first stage of the $(n+1)$-th step, so that effectively one only needs 3 times of such evaluations in each time step.

3 Application to specific kinetic equations and AP property

By now, we have obtained a second-order positivity-preserving scheme ((2.6) with coefficients satisfying (2.23) (2.24) (2.25)) for the general stiff kinetic equation (2.2). In this section, we apply the scheme to some specific kinetic equations and discuss its AP property.

We will consider the equation (2.2) with the following collision operators:

- The BGK operator [3], a simple relaxation type operator used to mimic the complicated Boltzmann collision operator:
  \[ Q(f) = \frac{\rho}{(2\pi T)^{\frac{d}{2}}} \exp\left( -\frac{|v-u|^2}{2T} \right), \]  
  \[ \mathcal{M}[f] = \frac{\rho}{(2\pi T)^{\frac{d}{2}}} \exp\left( -\frac{|v-u|^2}{2T} \right), \]  
  where $\rho$, $u$, and $T$ are given by the moments of $f$:
  \[ \rho = \int_{\mathbb{R}^d} f \, dv, \quad u = \frac{1}{\rho} \int_{\mathbb{R}^d} v \, dv, \quad T = \frac{1}{\rho^2} \int_{\mathbb{R}^d} |v-u|^2 \, dv, \]  
  and $\eta$ is some positive function depending only on $\rho$ and $T$.

- The ES-BGK operator [11], a generalized BGK model used to fit realistic values of the transport coefficients:
  \[ Q(f) = \frac{\rho}{\sqrt{\det(2\pi T)}} \exp\left( -\frac{1}{2}(v-u)^T \bar{T}^{-1} (v-u) \right), \]  
  where $\rho$, $u$, $T$ are given in (3.3) and
  \[ \bar{T} = (1 - \nu)TI + \nu \Theta, \quad \Theta = \frac{1}{\rho} \int_{\mathbb{R}^d} f(v-u) \otimes (v-u) \, dv, \]  
  where $-\frac{1}{2} < \nu < 1$ is a parameter and $I$ is the identity matrix. $\eta$ is some positive function of $\rho$ and $T$.

- The Boltzmann collision operator [6], a fundamental equation in kinetic theory describing the binary collisions in a rarefied gas:
  \[ Q(f) = \int_{\mathbb{R}^d} \int_{S^{d-1}} B(v-v_s, \sigma) [f(v')f(v_s') - f(v)f(v_s)] \, d\sigma \, dv_s, \]
where \( v' \) and \( v'_* \) (post-collisional velocities) are defined in terms of \( v \) and \( v_* \) (pre-collisional velocities) as

\[
v' = \frac{v + v_*}{2} + \frac{|v - v_*|}{2} \sigma, \quad v'_* = \frac{v + v_*}{2} - \frac{|v - v_*|}{2} \sigma,
\]

with \( \sigma \) being a vector varying on the unit sphere \( S^{d-1} \). \( B \) is the collision kernel characterizing the scattering rate and is a non-negative function.

- The kinetic Fokker-Planck operator [23], a kinetic model describing the drift and diffusion effects of particles:

\[
Q(f) = \nabla_v \cdot \left( \mathcal{M}[f] \nabla_v \frac{f}{\mathcal{M}[f]} \right),
\]

where \( \mathcal{M}[f] \) is the same as in the BGK model. Using the definition (3.2), (3.9) can be written equivalently as

\[
Q(f) = \nabla_v \cdot \left( \nabla_v f + \frac{(v - u)}{T} f \right),
\]

with \( u \) and \( T \) given by (3.3). This is the more commonly seen drift-diffusion type equation in the literature.

All of the above collision operators \( Q \) satisfy the following properties which can be found in many standard textbooks [6, 23] with perhaps the ES-BGK operator as an exception whose proof is given in [1].

- Conservation of mass, momentum, and energy:

\[
\langle Q(f) \phi \rangle = 0, \quad \langle \cdot \phi \rangle := \int_{\mathbb{R}^d} \cdot d v, \quad \phi(v) = \left(1, v, \frac{|v|^2}{2} \right)^T,
\]

for any function \( f \).

This implies that \( \exp(sQ)g \), the solution to the homogeneous equation (2.4) at \( t = t_0 + s \) with initial data \( f|_{t=t_0} = g \), satisfies the conservation property

\[
\langle (\exp(sQ)g) \phi \rangle = \langle g \phi \rangle, \quad \forall s \geq 0.
\]

- Decay of entropy

\[
\int_{\mathbb{R}^d} Q(f) \log f \ dv \leq 0,
\]

further,

\[
\int_{\mathbb{R}^d} Q(f) \log f \ dv = 0 \iff Q(f) = 0 \iff f = \mathcal{M}[f],
\]

where \( \mathcal{M}[f] \) is the Maxwellian defined in (3.2).

This implies that \( \exp(sQ)g \), the solution to the homogeneous equation (2.4) at \( t = t_0 + s \) with initial data \( f|_{t=t_0} = g \) has the long time behavior

\[
\lim_{s \to \infty} \exp(sQ)g = \mathcal{M}[g],
\]

i.e., \( \exp(sQ)g \) approaches the Maxwellian determined by the moments of the initial condition.
Using these properties, it is easy to show that the spatially inhomogeneous equation (1.1) has the compressible Euler equations as the leading-order asymptotics when $\varepsilon \to 0$. Indeed, taking the moments $\langle \cdot \phi \rangle$ on both sides of (1.1), one obtains

$$\partial_t \langle f \phi \rangle + \nabla_x \cdot \langle fv\phi \rangle = 0 \quad (3.16)$$

by the conservation property of $Q$. On the other hand, when $\varepsilon \to 0$, (1.1) formally implies $Q(f) \to 0$, hence $f \to M[f]$. Substituting $f = M[f] := M[U]$ into (3.16) yields

$$\partial_t U + \nabla_x \cdot \langle M[U]v\phi \rangle = 0, \quad (3.17)$$

where we used the vector $U$ to denote the first $d + 2$ moments of $f$: $U = (\rho, \rho u, E)^T$ with $E = \frac{1}{2} \rho u^2 + \frac{d}{2} \rho T$ being the total energy. The closed system (3.17) is nothing but the compressible Euler equations

$$\begin{cases}
\partial_t \rho + \nabla_x \cdot (\rho u) = 0, \\
\partial_t (\rho u) + \nabla_x \cdot (\rho u \otimes u + p I) = 0, \\
\partial_t E + \nabla_x \cdot ((E + p) u) = 0,
\end{cases} \quad (3.18)$$

where $p = \rho T$ is the pressure.

We now prove the AP property of the proposed scheme.

**Proposition 3.1.** The scheme (2.6) (with coefficients satisfying (2.23), (2.24), (2.25)) applied to the stiff kinetic equation (2.2) with the collision operator $Q$ being the BGK operator (3.1), the ES-BGK operator (3.4), the Boltzmann collision operator (3.7), and the kinetic Fokker-Planck operator (3.9) is asymptotic-preserving, i.e., for any initial data and fixed $\Delta t$, in the limit $\varepsilon \to 0$, (2.6) becomes a second-order scheme SSP-RK2 applied to the limiting Euler system (3.18). Furthermore,

$$\lim_{\varepsilon \to 0} f^{n+1} = M[U^{n+1}], \quad (3.19)$$

i.e., after each time step, $f^{n+1}$ is driven to its corresponding Maxwellian.

**Proof.** First of all, taking the moments $\langle \cdot \phi \rangle$ on (2.6) and using (3.12), one obtains

$$U^{(0)} = U^n, \quad U^{(1)} = U^{(0)} + b_1 \Delta t \langle T(f^{(0)}) \phi \rangle, \quad U^{(2)} = U^{(1)} + b_2 \Delta t \langle T(f^{(1)}) \phi \rangle, \quad U^{n+1} = w U^{(2)} + (1 - w) U^n. \quad (3.20)$$

On the other hand, for $a_0, a_1, a_2 > 0$, using (3.15), it can be seen from (2.6) that as $\varepsilon \to 0$, $f^{(0)}$, $f^{(1)}$, and $f^{n+1}$ are driven to their corresponding Maxwellian:

$$f^{(0)} \to M[U^n] = M[U^{(0)}], \quad f^{(1)} \to M[U^{(0)} + b_1 \Delta t \langle T(f^{(0)}) \phi \rangle] = M[U^{(1)}], \quad f^{n+1} \to M[w U^{(2)} + (1 - w) U^n] = M[U^{n+1}]. \quad (3.21)$$
Finally, substituting $f^{(0)}$ and $f^{(1)}$ into (3.20), one has
\[
U^{(1)} = U^n + b_1 \Delta t \langle T(M[U^n]) \phi \rangle, \\
U^{(2)} = U^{(1)} + b_2 \Delta t \langle T(M[U^{(1)}) \phi \rangle, \\
U^{n+1} = wU^{(2)} + (1 - w)U^n.
\] (3.22)

With the coefficients (2.23) and $T$ a discretized operator for $-v \cdot \nabla_x$, this is just a kinetic scheme for the limiting Euler equations (3.17) using the SSP-RK2 time discretization.

**Remark 3.2.** Note that the requirement for nonzero $a_1$, $a_2$, $a_3$ plays an important role here. In order for the scheme to have a nice AP property (works for any initial data, drives $f$ to the corresponding Maxwellian after each time step, the limiting scheme maintains second-order accuracy, etc.), we need all these coefficients to be non-degenerate. See also the discussion in Section 3.2.

### 3.1 A slightly different application

A slightly different example which does not fit exactly into the above framework is the Vlasov-Poisson-Fokker-Planck (VPFP) system, which is a kinetic description of the Brownian motion of a large system of particles in a surrounding bath. Since it can also be treated using the proposed method, we briefly describe it in this subsection.

The system in the high-field regime reads [20]
\[
\partial_t f + v \cdot \nabla_x f - \frac{1}{\varepsilon} \nabla_x \psi \cdot \nabla_v f = \frac{1}{\varepsilon} \nabla_v \cdot (\nabla_v f + v f),
\] (3.23)

with the electric potential $\psi = \psi(t,x)$ solving the Poisson equation
\[
-\Delta_x \psi = \rho - h,
\] (3.24)

where $\rho = \langle f \rangle$ is the density, and $h = h(x)$ is the background charge density satisfying the neutrality condition $\int_{\mathbb{R}^d} \rho(0,x) \, dx = \int_{\mathbb{R}^d} h(x) \, dx$.

One can write (3.23) as
\[
\partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon} \mathcal{Q}(f), \quad \mathcal{Q}(f) = \nabla_v \cdot (\nabla_v f + (v + \nabla_x \psi) f).
\] (3.25)

Here $\mathcal{Q}(f)$ is a Fokker-Planck type operator, and can be written in the form
\[
\mathcal{Q}(f) = \nabla_v \cdot \left( M \nabla_v \frac{f}{M} \right), \quad M(v) = M[\psi](v) = \frac{1}{(2\pi)^{d/2}} \exp \left( -\frac{(v + \nabla_x \psi)^2}{2} \right).
\] (3.26)

This $\mathcal{Q}$ satisfies the mass conservation $\langle \mathcal{Q}(f) \rangle = 0$. Also, $\mathcal{Q}(f) = 0 \iff f = \rho M[\psi]$.

Now taking the moment $\langle \cdot \rangle$ on both sides of (3.25), one has
\[
\partial_t \rho + \nabla_x \cdot \langle vf \rangle = 0.
\] (3.27)

On the other hand, as $\varepsilon \to 0$ in (3.25), formally $\mathcal{Q}(f) \to 0$, hence $f \to \rho M[\psi]$. Substituting this into the above equation, one obtains the limiting equation
\[
\partial_t \rho - \nabla_x \cdot (\rho \nabla_x \psi) = 0.
\] (3.28)

Starting with the form (3.25), it is easy to see that the scheme (2.6) can be applied directly and all the previous discussion regarding the AP property carries over straightforwardly. We omit the detail.
3.2 Comparison with existing methods

Searching the literature, there have been several methods available to solve the stiff kinetic equation (1.1) or equations of a similar structure. Therefore, we devote this subsection to a careful comparison of our method with some of the existing methods.

- If one replaces the solution operator \( \exp(sQ) \) by any second-order approximation, then (2.6) remains second order. In particular, if \( \exp(sQ) \) can be approximated by a second (or higher) order positivity-preserving and AP solver, then one can replace \( \exp(sQ) \) in (2.6) with this solver and still maintains the second-order accuracy, positivity, and AP property. For example, the scheme

\[
g^{(1)} = g + sQ(g^{(1)}),
\]

\[
g^{1} = g^{(1)} - \frac{1}{2}s^2 Q'(g^{(1)}) Q(g^{1}),
\]

produces \( g^{1} \) at \( t = t_0 + s \), which is a second-order positivity-preserving AP approximation to the exact solution \( \exp(sQ)g \), in the case of the BGK operator. Using this in (2.6) would give an IMEX Runge-Kutta method with correction, similar to our previous work [13].

- The following two existing second-order methods for (2.2) are special cases of (2.6):

1. If one considers the second-order Strang splitting

\[
\exp(\Delta t(T + Q)) = \exp\left(\frac{\Delta t}{2}Q\right) \exp(\Delta tT) \exp\left(\frac{\Delta t}{2}Q\right) + O(\Delta t^3),
\]

and discretizes \( \exp(\Delta tT) \) by SSP-RK2, then one arrives at (2.6) with

\[
a_0 = a_2 = \frac{1}{2}, \quad a_1 = 0, \quad b_1 = b_2 = 1, \quad w = \frac{1}{2},
\]

(3.31)

2. For the case \( Q(f) = -\mu f \) with \( \mu > 0 \) a constant, [14] rewrites (2.2) as

\[
\frac{d}{dt} \left( \exp\left( -\frac{t}{\varepsilon} Q \right) f \right) = \exp\left( -\frac{t}{\varepsilon} Q \right) T(f),
\]

(3.32)

and applies SSP-RK2 to (3.32) directly. Then one arrives at (2.6) with

\[
a_0 = a_2 = 0, \quad a_1 = 1, \quad b_1 = b_2 = 1, \quad w = \frac{1}{2},
\]

(3.33)

These two methods are not AP or suffer from order degeneracy in the fluid regime. In fact, in the first method \( a_1 = 0 \) and thus \( f^{(1)} \) is not at local Maxwellian. Therefore, the flux term \( b_2 \Delta t(T(f^{(1)})\phi) \) in (3.20) only approximates the flux in the limiting system up to first-order accuracy, which makes the limiting scheme first order. This order degeneracy of the Strang splitting was discovered already in an early work [15]. Similarly in the second method \( a_0 = a_2 = 0 \) and thus \( f^{(0)} = f^n \) is not at local Maxwellian, which means the flux term \( b_1 \Delta t(T(f^{(0)})\phi) \) in (3.20) is only first-order accurate in the limiting scheme. Moreover, even one starts with a consistent initial data, i.e., \( f^n = M[f^n] \), this method will not drive \( f^{n+1} \) to the local Maxwellian since \( a_2 = 0 \). Hence this error will pollute the solution as well in the next time step.
For the second method, [14] showed that the limiting scheme is second order with consistent initial data, in the case of \( Q(f) = -\mu f \) and \( T \) satisfying a maximum principle. Their proof is based on the following fact: if \( f \) is at local equilibrium (say \( f - f_{eq} = O(\varepsilon) \)), then \( f + \Delta t T(f) \) is also at local equilibrium ((\( f + \Delta t T(f) \)) - (\( f_{eq} + \Delta t T(f_{eq}) \)) = \( O(\varepsilon) \)). This is clearly not the case for equation (1.1), since generally speaking \( f - \Delta tv \cdot \nabla_x f \) is \( O(\Delta t) \) away from its local Maxwellian, even if \( f \) itself is at local Maxwellian.

- In [7], an exponential Runge-Kutta method was proposed for the homogeneous Boltzmann equation. This method is high order, AP, and positivity-preserving. But it is extended to the non-homogeneous equation (1.1) based on the Strang-splitting, hence suffers from the order degeneracy as mentioned above. A non-splitting version was proposed in [19] by a reformulation of the equation. However, the positivity of \( f \) is no longer guaranteed.

To summarize, by a careful choice of the coefficients (2.23) (2.24) (2.25), our scheme (2.6) is different from any existing exponential Runge-Kutta type methods. It is second order, positivity-preserving, and AP (capturing the Euler limit with second-order accuracy for any initial data).

4 Solving the homogeneous equation

A key assumption we made in Section 2 is that the solution to the homogeneous equation (2.4), or equivalently the solution operator \( \exp(sQ) \), can be found exactly. From the previous section we have also seen that this can be relaxed by finding an approximate solution, or an approximate operator \( \tilde{\exp}(sQ) \), such that it is at least second-order accurate in time, i.e.,

\[
\exp(sQ)g \approx \tilde{\exp}(sQ)g + O(s^3);
\]

positivity preserving, i.e.,

\[
g \geq 0 \implies \tilde{\exp}(sQ)g \geq 0, \quad \forall \text{ constant } s \geq 0;
\]

and AP, for which to hold we need \( \tilde{\exp}(sQ) \) satisfy the same long time behavior as \( \exp(sQ) \), i.e.,

\[
\lim_{s \to \infty} \tilde{\exp}(sQ)g = \mathcal{M}[g].
\]

In the following, we will provide the strategy to construct \( \exp(sQ) \) or \( \tilde{\exp}(sQ) \) for all the kinetic equations discussed in Section 3.

4.1 The BGK equation

For the homogeneous BGK equation

\[
\partial_t f = Q(f) = \eta(\mathcal{M}[f] - f), \quad f|_{t=t_0} = g,
\]

since \( Q \) conserves mass, momentum and energy, \( \mathcal{M}[f] = \mathcal{M}[g] \) does not change with time, neither does \( \eta \). Hence the solution at \( t = t_0 + s \) can be found analytically:

\[
\exp(sQ)g = e^{-\eta s}g + (1 - e^{-\eta s})\mathcal{M}[g].
\]
4.2 The ES-BGK equation

For the homogeneous ES-BGK equation

$$\partial_t f = Q(f) = \eta (G[f] - f), \quad f|_{t=t_0} = g, \quad (4.6)$$

since $Q$ conserves mass, momentum and energy, $\rho, u, T$ do not change with time, neither does $\eta$. Taking the moment $\langle \frac{1}{\rho} (v - u) \otimes (v - u) \rangle$ on both sides of (4.6) gives

$$\partial_t \Theta = \eta \left( \frac{1}{\rho} \langle (v - u) \otimes (v - u) G[f] \rangle - \Theta \right) = \eta (T - \Theta) = \eta (1 - \nu)(T I - \Theta), \quad (4.7)$$

whose solution is given by

$$\Theta(t_0 + s) = e^{-\eta(1-\nu)s} \Theta(t_0) + (1 - e^{-\eta(1-\nu)s})T I. \quad (4.8)$$

Hence

$$\tilde{T}(t_0 + s) = \nu e^{-\eta(1-\nu)s} \Theta(t_0) + (1 - \nu e^{-\eta(1-\nu)s})T I. \quad (4.9)$$

On the other hand, (4.6) can be integrated to yield

$$\exp(s Q)g = f(t_0 + s) = e^{-\eta s} g + \int_{t_0}^{t_0 + s} e^{-\eta(t_0 + s - \tau)} G[f(\tau)] \, d\tau, \quad (4.10)$$

where $G[f(\tau)]$ only depends on $\rho, u, \tilde{T}(\tau)$. Rather than solving (4.10) exactly, we propose to use a quadrature to approximate the integral part. We adopt the two-point Gauss-Lobatto quadrature, that is,

$$\int_{t_0}^{t_0 + s} e^{-\eta(t_0 + s - \tau)} \psi(\tau) \, d\tau \approx w_1 \psi(t_0) + w_2 \psi(t_0 + s), \quad (4.11)$$

where the weights $w_1, w_2$ are determined by requiring this approximation to be exact for $\psi(\tau) = 1, \tau$. A simple calculation gives

$$w_1 = \frac{1 - e^{-\eta s}}{\eta s} - e^{-\eta s}, \quad w_2 = \frac{1 - e^{-\eta s}}{\eta s}. \quad (4.12)$$

The quadrature in (4.11) has an error $O(s^3)$ for general functions.

Therefore, we approximate the solution in (4.10) as

$$\exp(s Q)g \approx \exp(s Q)g := e^{-\eta s} g + \left(1 - \frac{e^{-\eta s}}{\eta s} - e^{-\eta s}\right) G[\rho, u, \tilde{T}(t_0)] + \left(1 - \frac{1 - e^{-\eta s}}{\eta s}\right) G[\rho, u, \tilde{T}(t_0 + s)], \quad (4.13)$$

with $\tilde{T}(t_0 + s)$ given by (4.9). This approximation is positivity-preserving since (4.13) is a convex combination of positive functions. The resulting scheme is AP because $\exp(s Q)g$ satisfies the long time behavior $\lim_{s \to \infty} \tilde{T}(t_0 + s) \to T I$, and thus $G[\rho, u, \tilde{T}(t_0 + s)] \to M[g]$. Also, three weights in (4.13) converge to 0, 0, 1, respectively, hence $\exp(s Q)g \to M[g]$.

4.3 The Boltzmann equation

For the homogeneous Boltzmann equation

$$\partial_t f = Q(f) = \int_{\mathbb{R}^d} \int_{S^{d-1}} B(v - v_* \sigma)[f(v') f(v_*') - f(v) f(v_*)] \, d\sigma \, dv_*, \quad f|_{t=t_0} = g, \quad (4.14)$$
we adopt the exponential Runge-Kutta method introduced in [7] to find an approximate solution. Since $Q$ conserves mass, momentum and energy, $M[f] = M[g]$ does not change with time. Thus we can rewrite (4.14) as

$$\partial_t((f - M)e^{\mu t}) = (P(f) - \mu M)e^{\mu t},$$

(4.15)

where $P(f) := Q(f) + \mu f$, $\mu > 0$ being a constant, large enough so that $P(f) \geq 0$ (a simple choice is $\mu = \sup_v \int_{\mathbb{R}^d} \int_{S^{d-1}} B(v - v_*, \sigma) f(v_*) \, d\sigma \, dv_*$). Then, by applying the midpoint method to (4.15), one obtains a second-order scheme

$$f^{(1)}(1) = (g - M(t^*) + \lambda e^{\frac{\lambda}{2}} \left( \frac{P(f)}{\mu} - M \right),$$

$$f^1 = (g - M) + \lambda e^{\frac{\lambda}{2}} \left( \frac{P(f)}{\mu} - M \right),$$

(4.16)

with $\lambda = \mu s$, which simplifies to

$$f^{(1)} = e^{-\frac{\lambda}{2}} g + \left( 1 - e^{-\frac{\lambda}{2}} - \frac{\lambda}{2} e^{-\frac{\lambda}{2}} \right) M + \frac{\lambda}{2} e^{-\frac{\lambda}{2}} \frac{P(g)}{\mu},$$

$$f^1 = e^{-\lambda} g + \left( 1 - e^{-\lambda} - \lambda e^{-\frac{\lambda}{2}} \right) M + \lambda e^{-\frac{\lambda}{2}} \frac{P(f^{(1)})}{\mu}.$$  

(4.17)

Therefore, we choose $f^1$ to be the approximate solution at $t = t_0 + s$:

$$\exp(sQ)g = \exp(sQ)g := f^1.$$  

(4.18)

This approximation is positivity-preserving since both $f^{(1)}$ and $f^1$ are convex combinations of positive functions. It is AP since $s \to \infty$ implies $\lambda \to \infty$, thus $f^1 \to M$.

**Remark 4.1.** Here we did not address the issue of velocity domain discretization. To get a fully discrete scheme, one also needs an efficient and positivity-preserving solver for the Boltzmann collision operator (to evaluate the term $P(f)$ in the scheme). Available choices are the direct simulation Monte Carlo (DSMC) method [2] or the recently proposed entropic Fourier method [7].

### 4.4 The kinetic Fokker-Planck equation

For the homogeneous kinetic Fokker-Planck equation

$$\partial_t f = Q(f) = \nabla_v \cdot \left( M[f] \nabla_v \frac{f}{M[f]} \right), \quad f|_{t=t_0} = g,$$

(4.19)

since $Q$ conserves mass, momentum and energy, $M[f] = M[g]$ does not change with time. We adopt the approximation proposed in [18] to discretize $Q$. Define $\tilde{f} = \frac{f}{\sqrt{M}}$, then $\tilde{f}$ solves

$$\partial_t \tilde{f} = \tilde{Q}(\tilde{f}) := \frac{1}{\sqrt{M}} \nabla_v \cdot \left( M \nabla_v \frac{\tilde{f}}{\sqrt{M}} \right), \quad \tilde{f}|_{t=t_0} = \tilde{g} := \frac{g}{\sqrt{M}}.$$  

(4.20)

Hence

$$\exp(sQ)g = \sqrt{M} \exp(sQ)\tilde{g}.$$  

(4.21)

Now it suffices to approximate $\exp(sQ)\tilde{g}$. To do this, we will first discretize the velocity and then use the matrix exponential to solve (4.20). For simplicity we consider $v$ in 1d. We truncate
the velocity domain into a large enough interval $[-|v|_{\text{max}}, |v|_{\text{max}}]$ and discretize it into $N_v$ grid points with $v_i = -|v|_{\text{max}} + (i - 1/2)\Delta v$, $i = 1, \ldots, N_v$, $\Delta v = 2|v|_{\text{max}}/N_v$. Then the operator $\tilde{Q}$ can be approximated by a tridiagonal symmetric matrix $\tilde{Q}^h$ with the entries given by

$$
\tilde{Q}_{i,i}^h = -\frac{1}{\Delta v^2} \frac{\sqrt{M_{i-1}} + \sqrt{M_{i+1}}}{\sqrt{M_i}},
\tilde{Q}_{i,i-1}^h = \tilde{Q}_{i,i+1}^h = \frac{1}{\Delta v^2},
$$

(4.22)

where $M_i = M(v_i)$. Define the vector $\tilde{g}^h = (\tilde{g}_1, \ldots, \tilde{g}_{N_v})^T$ with $\tilde{g}_i = \tilde{g}(v_i)$, then we approximate $\exp(s\tilde{Q})g$ as

$$
(\exp(s\tilde{Q})g)_i \approx \sqrt{M_i} (\exp(s\tilde{Q}^h)\tilde{g}_i),
$$

(4.23)

where $\exp(s\tilde{Q}^h)\tilde{g}^h$ can be computed very accurately by existing matrix exponential algorithms (for simplicity we assume there is no error occurring at this step). This approximation is positivity-preserving since the off-diagonal entries of $\tilde{Q}^h$ are non-negative. It is AP since $s \to \infty$ implies $\sqrt{M_i} (\exp(s\tilde{Q}^h)\tilde{g}_i) \to M_i$. To see this, note that the discretization (4.22) for equation (4.20) is equivalent to the following

$$
\partial_t f_i = \frac{F_{i+1/2} - F_{i-1/2}}{\Delta v}, \quad F_{i+1/2} := \sqrt{M_i M_{i+1}} \left( \frac{f_{i+1}}{M_{i+1}} - \frac{f_i}{M_i} \right).
$$

(4.24)

Define the discrete relative entropy as

$$
H = \sum_i f_i \log \frac{f_i}{M_i} \Delta v,
$$

(4.25)

then

$$
\partial_t H = \sum_i \partial_t f_i \left( \log \frac{f_i}{M_i} + 1 \right) \Delta v = \sum_i (F_{i+1/2} - F_{i-1/2}) \left( \log \frac{f_i}{M_i} + 1 \right)
$$

$$
= -\sum_i F_{i+1/2} \left( \log \frac{f_{i+1}}{M_{i+1}} - \log \frac{f_i}{M_i} \right)
$$

$$
= -\sum_i \sqrt{M_i M_{i+1}} \left( \frac{f_{i+1}}{M_{i+1}} - \frac{f_i}{M_i} \right) \left( \log \frac{f_{i+1}}{M_{i+1}} - \log \frac{f_i}{M_i} \right) \leq 0,
$$

(4.26)

and the equality holds if and only if $f_i/M_i$ is independent of $i$. This implies $f_i = M_i$ by conservation.

We remark that for the VPFP system, one can solve the homogeneous equation (2.4) by the same method, since for the homogeneous equation the $\psi$ appeared in (3.26) does not change in time.

5 Entropy-decay property

In this section, we discuss the entropy-decay property of our scheme. First of all, we recall the following well-known result in kinetic theory. For the kinetic equation (1.1) with the collision operator being the BGK operator (3.1), the ES-BGK operator (3.4), the Boltzmann collision operator (3.7), or the kinetic Fokker-Planck operator (3.9), one has

$$
\frac{d}{dt} \int f \log f \, dv \, dx \leq 0
$$

(5.1)
under a periodic or compactly supported boundary condition in $x$. This is the famous H-theorem which says that the total entropy of the system is always non-increasing.

We would like to show that our scheme (2.6) coupled with the first-order upwind discretization for the transport term and the homogeneous solvers discussed in Section 4 satisfies a discrete entropy-decay property (a discrete analog of (5.1)). In order to do so, we assume the velocity space is continuous, in particular, this means the Fokker-Planck operator is not discretized and the solution to its homogeneous equation can be found analytically.

For simplicity, we consider the equation (1.1) in 1d:

$$\partial_t f + v \partial_x f = \frac{1}{\varepsilon} Q(f).$$  \hfill (5.2)

We truncate the velocity domain to a large enough interval $[-|v|_{\text{max}}, |v|_{\text{max}}]$ and discretize the transport term by the upwind method ($j$ is the spatial index):

$$(v \partial_x f)_j = \chi_{v \geq 0} \frac{f_j - f_{j-1}}{\Delta x} + \chi_{v < 0} \frac{f_{j+1} - f_j}{\Delta x}.$$  \hfill (5.3)

Define the discrete entropy as

$$S[f] := \Delta x \sum_j S[f_j], \quad S[f_j] := \int f_j \log f_j \, dv,$$  \hfill (5.4)

then we claim that the scheme (2.6) satisfies a discrete entropy-decay property:

$$S[f^{n+1}] \leq S[f^n].$$  \hfill (5.5)

To prove (5.5), we need two building blocks, one is the exponential step decays entropy, i.e., for either the exact $\exp(sQ)$ or approximate $\tilde{\exp}(sQ)$, one has

$$S[\exp(sQ)g] \leq S[g] \quad \text{or} \quad S[\tilde{\exp}(sQ)g] \leq S[g], \quad \forall \text{ constant } s \geq 0,$$  \hfill (5.6)

the other is the transport step decays entropy, i.e., for step of the form $g = f + a \Delta t T(f)$, one has

$$S[g] \leq S[f], \quad \text{under the CFL condition } \Delta t \leq \frac{\Delta x}{a |v|_{\text{max}}}. $$  \hfill (5.7)

We now prove (5.6) and (5.7), respectively.

• For the BGK and Fokker-Planck operators, we have the exact $\exp(sQ)$, hence $S[\exp(sQ)g] \leq S[g]$ follows directly from the analytical result (3.13).

For the ES-BGK operator, note that (4.13) is a convex combination of $g, G[g]$ and $G[\exp(sQ)g]$. One has $S[G[g]] \leq S[g]$ from [1], hence $S[G[\exp(sQ)g]] \leq S[\exp(sQ)g] \leq S[g]$ (the second inequality comes from the analytical result (3.13)). Therefore, $S[\tilde{\exp}(sQ)g] \leq S[g]$ follows from the convexity of $S$.

For the Boltzmann operator, note that in the approximation (4.17), $f^{(1)}$ is a convex combination of $g, \mathcal{M}$ and $\frac{P(g)}{\mu}$, and $f^1$ is a convex combination of $g, \mathcal{M}$ and $\frac{P(f^1)}{\mu}$. In [22], it is proved that $S[\frac{P(g)}{\mu}] \leq S[f]$ (for Maxwell molecules). Therefore, by the convexity of $S$ and $S[\mathcal{M}[g]] \leq S[g]$, one has $S[f^{(1)}] \leq S[g]$, hence $S[f^1] \leq S[g]$. Therefore, $S[\tilde{\exp}(sQ)g] \leq S[g]$. 

16
The transport step \( g = f + a \Delta t T(f) \) with (5.3) plugged in reads
\[
g_j = f_j - a \Delta t \left( \chi_{v \geq 0} \frac{f_j - f_{j-1}}{\Delta x} + \chi_{v < 0} \frac{f_{j+1} - f_j}{\Delta x} \right)
= \left(1 - a \frac{|v| \Delta t}{\Delta x} \right) f_j + a \frac{|v| \Delta t}{\Delta x} \left( \chi_{v \geq 0} f_{j-1} + \chi_{v < 0} f_{j+1} \right).
\tag{5.8}
\]
Hence the right hand side is a convex combination of \( f_j \) and \( \chi_{v \geq 0} f_{j-1} + \chi_{v < 0} f_{j+1} \) under the CFL condition \( \Delta t \leq \frac{\Delta x}{a |v|_{\max}} \). Then using the convexity of function \( f \log f \), one has
\[
S[g] \leq \int \left(1 - a \frac{|v| \Delta t}{\Delta x} \right) f_j \log f_j \, dv
+ \int a \frac{|v| \Delta t}{\Delta x} \left( \chi_{v \geq 0} f_{j-1} + \chi_{v < 0} f_{j+1} \right) \log \left( \chi_{v \geq 0} f_{j-1} + \chi_{v < 0} f_{j+1} \right) \, dv
= S[f] - a \frac{\Delta t}{\Delta x} \left( F_{j+1/2} - F_{j-1/2} \right),
\tag{5.9}
\]
where
\[
F_{j+1/2} := \int |v| \left( \chi_{v \geq 0} f_j \log f_j - \chi_{v < 0} f_{j+1} \log f_{j+1} \right) \, dv
\tag{5.10}
\]
is the discrete entropy flux. Summing over \( j \) in (5.9) and assuming the periodic or compactly supported boundary condition in \( x \), one obtains
\[
S[g] \leq S[f].
\tag{5.11}
\]
Now applying the previous two results in (2.6), we have
\[
S[f^{(2)}] \leq S[f^{(1)}] \leq S[f^{(0)}] \leq S[f^n],
\tag{5.12}
\]
therefore
\[
S[f^{n+1}] \leq w S[f^{(2)}] + (1 - w) S[f^n] \leq S[f^n].
\tag{5.13}
\]
The assertion is proved.

6 A remark on spatial and velocity discretizations

Most of the spatial and velocity discretizations follow our previous paper [13], namely, we use a finite volume method for the \( x \)-variable and finite difference method for the \( v \)-variable.

For the transport term, we adopt the fifth-order finite volume WENO method [21] with a bound-preserving limiter [24, 25] to insure the positivity. Since the treatment of this part is standard and has been described in [13], we omit the detail.

For the collision term, special care needs to be paid when switching between the finite volume and finite difference framework. We briefly describe the procedure in the following. For convenience, we regard \( v \) as continuous and omit it in the discussion.

Let \( I_j = [x_{j-1/2}, x_{j+1/2}] \) be the \( j \)-th spatial cell, and \( \{x_{l,i}\} \ (l = 1, 2, 3) \) denote the three Gauss-Legendre quadrature points in this cell and \( \{w_l\} \) be the corresponding quadrature weights. For a fixed \( v \), suppose we are given the cell average \( f_j \geq 0 \) in \( I_j \), we would like to construct a polynomial \( f_j(x) \) of degree four such that
• $f_j(x)$ is a fifth-order accurate approximation to $f(x)$ in $I_j$ with $f_j$ being its cell average, i.e.,
\[ \frac{1}{\Delta x} \int_{x_{j-1/2}}^{x_{j+1/2}} f_j(x) \, dx = f_j. \] (6.1)

• $f_j(x)$ is non-negative at the Gauss quadrature points, i.e.,
\[ f_{j,l} := f_j(x_{j,l}) \geq 0, \quad l = 1, 2, 3. \] (6.2)

The construction of such a polynomial can be done similarly as described in Section 3.2.2 of our previous paper \[13\]. Provided with $f_j(x)$, it is easy to see (6.1) reduces to
\[ \sum_{l=1}^{3} w_l f_{j,l} = f_j, \] (6.3)since the three-point Gauss-Legendre quadrature is exact for polynomials with degree no more than five.

Then we approximate the $j$-th cell average of $\exp(sQ)f$ by
\[ (\exp(sQ)f)_j = \sum_{l=1}^{3} w_l \exp(sQ)f_{j,l}. \] (6.4)

This approximation is fifth-order accurate in $x$ since the reconstruction $f_{j,l}$ is. It is also conservative, since
\[ \langle (\exp(sQ)f_j) \phi \rangle = \sum_{l=1}^{3} w_l \langle \exp(sQ)f_{j,l} \phi \rangle = \sum_{l=1}^{3} w_l \langle f_{j,l} \phi \rangle = \langle \sum_{l=1}^{3} w_l f_{j,l} \phi \rangle = \langle f_j \phi \rangle, \] (6.5)
where we used (3.12) in the second equality and (6.3) in the last one.

For the mixed regime problem where $\varepsilon = \varepsilon(x)$, one needs to compute $\exp(s(x)Qf)$ with $s(x)$ a given function depending on $x$. To do this, we use the same reconstruction $f_{j,l}$ and approximate $\exp(s(x)Qf)$ by
\[ (\exp(s(x)Qf)_j = \sum_{l=1}^{3} w_l \exp(s(x_{j,l})Q)f_{j,l}, \] (6.6)
which is still fifth-order accurate and conservative.

### 7 Numerical examples

In this section we demonstrate numerically the properties of the proposed scheme (2.6) with coefficients (2.23) (2.26) by solving the 1d BGK equation and Fokker-Planck equation.

We consider the computational domain $x \in [0, 2]$ with periodic boundary condition (except the test in Section 7.2 where the Dirichlet boundary condition is assumed), and a large enough velocity domain $v \in [-|v|_{\text{max}}, |v|_{\text{max}}]$ with $|v|_{\text{max}} = 15$. The $x$-space is discretized into $N_x$ cells with $\Delta x = 2/N_x$ and cell center $x_j = (j-1/2)\Delta x$, $j = 1, \ldots, N_x$. The $v$-space is discretized into $N_v$ grid points with $\Delta v = 2|v|_{\text{max}}/N_v$ and $v_i = -|v|_{\text{max}} + (i-1/2)\Delta v$, $i = 1, \ldots, N_v$. Unless specified, $N_v = 150$ is used in all tests so that the discretization error in $v$ is much smaller than that in $x$ and $t$.

To compute the matrix exponential (4.23) resulting from the discretization of the Fokker Planck operator, we used the code by S. Guttel \[10\] for the test in Section 7.1 and the MATLAB function ‘expm’ for other tests.
\[
\begin{array}{|c|c|c|c|c|c|c|c|}
\hline
\varepsilon & \varepsilon = 1e + 00 & \varepsilon = 1e - 02 & \varepsilon = 1e - 04 & \varepsilon = 1e - 06 & \varepsilon = 1e - 08 & \varepsilon = 1e - 10 \\
\hline
\text{Nx}=10 & 5.60e-04 & 4.64e-04 & 4.67e-04 & 4.67e-04 & 4.67e-04 & 4.67e-04 \\
\text{Order} & 3.25 & 3.56 & 3.68 & 3.68 & 3.68 & 3.68 \\
\hline
\text{Nx}=20 & 5.91e-05 & 3.93e-05 & 3.65e-05 & 3.65e-05 & 3.65e-05 & 3.65e-05 \\
\text{Order} & 3.77 & 3.80 & 3.03 & 3.89 & 3.89 & 3.89 \\
\hline
\text{Nx}=40 & 4.33e-06 & 2.83e-06 & 4.46e-06 & 2.46e-06 & 2.46e-06 & 2.46e-06 \\
\text{Order} & 4.36 & 3.31 & -0.23 & 4.49 & 4.49 & 4.49 \\
\hline
\text{Nx}=80 & 2.11e-07 & 2.86e-07 & 5.24e-06 & 1.10e-07 & 1.10e-07 & 1.10e-07 \\
\text{Order} & 4.95 & 6.24e-08 & 3.25e-06 & 6.29e-09 & 6.29e-09 & 6.29e-09 \\
\hline
\text{Nx}=160 & 1.27e-08 & 6.24e-08 & 3.25e-06 & 6.29e-09 & 6.29e-09 & 6.29e-09 \\
\text{Order} & 2.14 & 1.55e-08 & 1.23e-06 & 1.45e-09 & 1.45e-09 & 1.45e-09 \\
\hline
\text{Nx}=320 & 7.30e-10 & 3.88e-09 & 3.74e-07 & 3.68e-10 & 3.68e-10 & 3.68e-10 \\
\text{Order} & 1.99 & 2.01 & 1.40 & 2.11 & 2.11 & 2.11 \\
\hline
\text{Nx}=640 & 1.83e-10 & 9.71e-10 & 1.03e-07 & 2.82e-10 & 9.20e-11 & 9.20e-11 \\
\text{Order} & 2.00 & 2.00 & 1.86 & 0.38 & 2.00 & 2.00 \\
\hline
\end{array}
\]

Table 1: Accuracy test of the scheme for the BGK equation.

### 7.1 Accuracy test

We first verify the second-order accuracy of the scheme. We consider inconsistent initial data

\[ f(0, x, v) = 0.5M_{\rho,u,T} + 0.3M_{\rho,0.5u,T}, \]

with

\[ \rho = 1 + 0.2 \sin(\pi x), \quad u = 1, \quad T = \frac{1}{1 + 0.2 \sin(\pi x)^2}, \]

and compute the solution to time \( t = 0.1 \). We choose different values of \( \varepsilon \), ranging from the kinetic regime (\( \varepsilon = 1 \)) to the fluid regime (\( \varepsilon = 10^{-10} \)). We choose different \( \Delta x \) and set \( \Delta t = 0.5 \Delta x / |v|_{\text{max}} \). This CFL number is not small enough to guarantee the positivity which is pretty restrictive due to the spatial discretization. We will consider the positivity-preserving property in the following test. For the same reason, the positivity-preserving limiter is turned off here. Since the exact solution is not available, the numerical solution on a finer mesh \( \Delta x/2 \) is used as a reference solution to compute the error for the solution on the mesh of size \( \Delta x \):

\[ \text{error}_{\Delta t, \Delta x} := \| f_{\Delta t, \Delta x} - f_{\Delta t/2, \Delta x/2} \|_{L^2_x}. \]

The results are shown in Tables 7.1 and 7.2. For the Fokker-Planck equation, due to the second-order discretization error in the velocity space, one has to choose a larger \( N_v \) in order to see the temporal error. In all these results, the spatial error dominates for small \( N_x \), and the temporal error dominates for large \( N_x \). One can clearly see that in both the kinetic regime \( \varepsilon = O(1) \) and the fluid regime \( \varepsilon \ll 1 \), the scheme is second order. Note that there is some extent of order reduction in the intermediate regime \( \varepsilon = O(\Delta t) \). The uniform accuracy of the AP scheme is an open problem and we do not attempt to address this issue in the current work.

### 7.2 Positivity-preserving property

We now illustrate the positivity-preserving property of the scheme. Consider the initial data

\[ f(0, x, v) = M_{\rho,u,T}, \]

19
with
\[(\rho, u, T) = \begin{cases} (1, 0, 1), & 0 \leq x \leq 1, \\ (0.125, 0, 0.25), & 1 \leq x \leq 2. \end{cases} \quad (7.5)\]

With the positivity-preserving limiter, the CFL condition of our scheme is \(\Delta t \leq \frac{1}{12} \frac{\Delta x}{|v|_{\text{max}}}\) (note that 1/12 comes from the spatial discretization and the forward Euler method also has the same constraint). We choose \(\Delta t = \frac{1}{24} \frac{\Delta x}{|v|_{\text{max}}}\) and \(N_x = 80\).

For the BGK equation, no negative cells are detected in the simulation. For the Fokker-Planck equation, one technical issue is that we are not aware of any algorithms that can guarantee the numerically computed matrix exponential is positive if the exact matrix exponential is. To demonstrate that no negative values are caused by our time discretization, we use ‘expm’ function in MATLAB to compute the matrix exponential and set the negative entries of the resulting matrix to zero. With this modification, no negative cells are detected in the simulation.

As a comparison, we solve the same equations with the same initial data and spatial/velocity discretization, but using the ARS(2,2,2) scheme in time [2], which is a standard second-order accurate IMEX scheme with no positivity-preserving property. The number of negative cells during the simulation is tracked. The result for the BGK equation is already included in the previous paper [13] and is omitted here. The result for the Fokker-Planck equation is shown in Figure 1. Here to make the comparison fair, when we compute \((I - sQh^{-1}g)^{-1}g_h\) (an operator needs to be evaluated in the IMEX scheme), we first compute the matrix \((I - sQh^{-1})^{-1}\) which is not necessarily positive at the numerical level, and then set the negative entries to zero in this matrix. This is to make sure that no negative values are generated due to the failure of positivity-preserving in the matrix inversion. In Figure 1 one can still see a lot of negative cells in the fluid regime.

### 7.3 AP property

Finally, to illustrate the AP property, we use the proposed scheme to solve the BGK and Fokker-Planck equations in a mixed regime (\(\varepsilon\) is a function of \(x\) so that in part of the domain the problem is in kinetic regime and while in other part it is in fluid regime). We take the same

Table 2: Accuracy test of the scheme for the Fokker-Planck equation. Here \(N_v = 600\).

| \(N_x\)  | \(\varepsilon = 1e + 00\) | \(\varepsilon = 1e - 01\) | \(\varepsilon = 1e - 02\) | \(\varepsilon = 1e - 03\) | \(\varepsilon = 1e - 04\) | \(\varepsilon = 1e - 05\) | \(\varepsilon = 1e - 06\) | \(\varepsilon = 1e - 07\) |
|--------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| 10     | 5.30e-04         | 4.64e-04         | 4.62e-04         | 4.66e-04         | 4.66e-04         | 4.66e-04         | 4.66e-04         | 4.66e-04         |
| 20     | 5.50e-05         | 4.32e-05         | 3.93e-05         | 4.63e-05         | 3.65e-05         | 3.65e-05         | 3.65e-05         | 3.65e-05         |
| Order  | 3.27             | 3.42             | 3.56             | 3.33             | 3.68             | 3.68             | 3.68             | 3.68             |
| 40     | 3.89e-06         | 2.82e-06         | 3.42e-06         | 1.29e-05         | 2.54e-06         | 2.46e-06         | 2.46e-06         | 2.46e-06         |
| Order  | 3.82             | 3.94             | 3.52             | 1.85             | 3.85             | 3.89             | 3.89             | 3.89             |
| 80     | 1.80e-07         | 1.29e-07         | 5.47e-07         | 4.23e-06         | 2.16e-06         | 1.10e-07         | 1.10e-07         | 1.10e-07         |
| Order  | 4.43             | 4.45             | 2.64             | 1.61             | 0.23             | 4.49             | 4.49             | 4.49             |
| 160    | 1.13e-08         | 9.34e-09         | 3.56e-08         | 1.35e-07         | 1.25e-06         | 2.97e-06         | 1.16e-08         | 6.30e-09         |
| Order  | 3.99             | 3.79             | 2.02             | 1.76             | -0.46            | 3.25             | 4.12             | 4.12             |
| 320    | 2.64e-09         | 2.07e-09         | 3.56e-08         | 3.53e-07         | 1.80e-06         | 1.08e-07         | 1.50e-09         | 1.45e-09         |
| Order  | 2.10             | 2.17             | 1.92             | 1.82             | 0.72             | -3.22            | 2.07             | 2.11             |
| 640    | 6.66e-10         | 5.77e-10         | 9.93e-09         | 1.01e-07         | 7.15e-07         | 3.91e-07         | 3.62e-10         | 3.68e-10         |
| Order  | 1.98             | 1.85             | 1.84             | 1.80             | 1.33             | -1.86            | 2.05             | 1.98             |
Figure 1: Total number of negative cells for the ARS(2,2,2) scheme applied to the Fokker-Planck equation during time evolution. Blue line: $\varepsilon = 10^{-6}$. Red line: $\varepsilon = 10^{-8}$.

initial data as in (7.1)-(7.2) and $N_x = 40$.

For the BGK equation, we consider $\varepsilon = \varepsilon(x)$ as follows:

$$
\varepsilon(x) = \varepsilon_0 + (\tanh(1 - 11(x - 1)) + \tanh(1 + 11(x - 1))), \quad \varepsilon_0 = 10^{-5}.
$$

(7.6)

We compare the macroscopic quantities at time $t = 0.5$ with a reference solution computed by SSP-RK2 with $N_x = 80$. Note that for our scheme, $\Delta t = \frac{1}{24} \frac{\Delta x}{|v|_{\text{max}}} \approx 7 \times 10^{-5}$; while for the explicit SSP-RK2 scheme, $\Delta t = \frac{1}{240} \frac{\Delta x}{|v|_{\text{max}}} \approx 7 \times 10^{-6}$ which needs to resolve $\varepsilon$. One can see a good agreement with the reference solution in Figure 2.

For the Fokker-Planck equation, we consider the following $\varepsilon(x)$:

$$
\varepsilon(x) = \varepsilon_0 + (\tanh(1 - 11(x - 1)) + \tanh(1 + 11(x - 1))), \quad \varepsilon_0 = 5 \times 10^{-4}.
$$

(7.7)

The numerical parameters are chosen the same as the BGK case, except in the reference solution, $\Delta t = \frac{1}{540} \frac{\Delta x}{|v|_{\text{max}}} \approx 3 \times 10^{-6}$ in order to satisfy the explicit parabolic CFL condition. The result is shown in Figure 3, and again with good agreement.

8 Conclusion

We introduced a new exponential Runge-Kutta time discretization method for a class of stiff kinetic equations. The method is second order, AP, and positivity-preserving. We applied the method to the relaxation type equations (BGK and ES-BGK equations), the diffusion type equations (kinetic Fokker-Planck and Vlasov-Poisson-Fokker-Planck equations), and even the full
Figure 2: The BGK equation in a mixed regime. Left to right: density $\rho$, velocity $u$, and temperature $T$. Solid line: reference solution computed by the explicit SSP-RK2 scheme. Dots: solution computed by the proposed scheme.

Figure 3: The Fokker-Planck equation in a mixed regime. Left to right: density $\rho$, velocity $u$, and temperature $T$. Solid line: reference solution computed by the explicit SSP-RK2 scheme. Dots: solution computed by the proposed scheme.
Boltzmann equation. Further, we showed that the method satisfies an entropy-decay property when coupled with upwind discretization for the transport term. Numerical examples for the BGK and Fokker-Planck equations were presented to demonstrate the properties of the proposed method.

Acknowledgement

The first author would like to thank Prof. Lili Ju for helpful discussion on exponential methods.

References

[1] P. Andries, P. Le Tallec, J.-P. Perlat, and B. Perthame. The Gaussian-BGK model of Boltzmann equation with small Prandtl number. *Eur. J. Mech. B/Fluids*, 19:813–830, 2000.

[2] U. Ascher, S. Ruuth, and R. Spiteri. Implicit-explicit Runge-Kutta methods for time-dependent partial differential equations. *Appl. Numer. Math.*, 25:151–167, 1997.

[3] P. L. Bhatnagar, E. P. Gross, and M. Krook. A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. *Phys. Rev.*, 94:511–525, 1954.

[4] G. A. Bird. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*. Clarendon Press, Oxford, 1994.

[5] Z. Cai, Y. Fan, and L. Ying. An entropic Fourier method for the Boltzmann equation. *preprint*.

[6] C. Cercignani. *The Boltzmann Equation and Its Applications*. Springer-Verlag, New York, 1988.

[7] G. Dimarco and L. Pareschi. Exponential Runge-Kutta methods for stiff kinetic equations. *SIAM J. Numer. Anal.*, 49:2057–2077, 2011.

[8] S. Gottlieb, D. Ketcheson, and C.-W. Shu. *Strong Stability Preserving Runge-Kutta and Multistep Time Discretizations*. World Scientific, 2011.

[9] S. Gottlieb, C.-W. Shu, and E. Tadmor. Strong stability-preserving high-order time discretization methods. *SIAM Rev.*, 43:89–112, 2001.

[10] S. Güttel. *Rational Krylov methods for operator functions*. PhD thesis, Institut für Numerische Mathematik und Optimierung, Technische Universität Bergakademie Freiberg, 2010.

[11] L. Holway. Kinetic theory of shock structure using an ellipsoidal distribution function. In *Proceedings of the 4th International Symposium on Rarefied Gas Dynamics*, volume I, pages 193–215, New York, 1966. Academic Press.
[12] J. Hu, S. Jin, and Q. Li. Asymptotic-preserving schemes for multiscale hyperbolic and kinetic equations. In R. Abgrall and C.-W. Shu, editors, Handbook of Numerical Methods for Hyperbolic Problems: Applied and Modern Issues, chapter 5, pages 103–129. North-Holland, 2017.

[13] J. Hu, R. Shu, and X. Zhang. Asymptotic-preserving and positivity-preserving implicit-explicit schemes for the stiff BGK equation. SIAM J. Numer. Anal., 56:942–973, 2018.

[14] J. Huang and C.-W. Shu. Bound-preserving modified exponential Runge-Kutta discontinuous Galerkin methods for scalar hyperbolic equations with stiff source terms. J. Comput. Phys., 361:111–135, 2018.

[15] S. Jin. Runge-Kutta methods for hyperbolic conservation laws with stiff relaxation terms. J. Comput. Phys., 122:51–67, 1995.

[16] S. Jin. Efficient asymptotic-preserving (AP) schemes for some multiscale kinetic equations. SIAM J. Sci. Comput., 21:441–454, 1999.

[17] S. Jin. Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. Riv. Mat. Univ. Parma, 3:177–216, 2012.

[18] S. Jin and B. Yan. A class of asymptotic-preserving schemes for the Fokker-Planck-Landau equation. J. Comput. Phys., 230:6420–6437, 2011.

[19] Q. Li and L. Pareschi. Exponential Runge-Kutta for the inhomogeneous Boltzmann equations with high order of accuracy. J. Comput. Phys., 259:402–420, 2014.

[20] F. Poupaud. Runaway phenomena and fluid approximation under high fields in semiconductor kinetic theory. Z. Angew. Math. Mech., 72:359–372, 1992.

[21] C.-W. Shu. Essentially non-oscillatory and weighted essentially non-oscillatory schemes for hyperbolic conservation laws. In Advanced numerical approximation of nonlinear hyperbolic equations, pages 325–432. Springer, 1998.

[22] C. Villani. Fisher information estimates for Boltzmann’s collision operator. J. Math. Pures Appl., 77:821–837, 1998.

[23] C. Villani. A review of mathematical topics in collisional kinetic theory. In S. Friedlander and D. Serre, editors, Handbook of Mathematical Fluid Mechanics, volume I, pages 71–305. North-Holland, 2002.

[24] X. Zhang and C.-W. Shu. On maximum-principle-satisfying high order schemes for scalar conservation laws. J. Comput. Phys., 229:3091–3120, 2010.

[25] X. Zhang and C.-W. Shu. Maximum-principle-satisfying and positivity-preserving high-order schemes for conservation laws: survey and new developments. In Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, page rspa20110153. The Royal Society, 2011.