J Sci Comput (2017) 73:797–818
DOI 10.1007/s10915-017-0499-3

On a Class of Implicit–Explicit Runge–Kutta Schemes for Stiff Kinetic Equations Preserving the Navier–Stokes Limit

Jingwei Hu[1] · **Xiangxiong Zhang**[1]

Received: 14 February 2017 / Revised: 29 June 2017 / Accepted: 10 July 2017 /
Published online: 17 July 2017

Abstract Implicit–explicit (IMEX) Runge–Kutta (RK) schemes are popular high order time discretization methods for solving stiff kinetic equations. As opposed to the compressible Euler limit (leading order asymptotics of the Boltzmann equation as the Knudsen number ε goes to zero), their asymptotic behavior at the Navier–Stokes (NS) level (next order asymptotics) was rarely studied. In this paper, we analyze a class of existing IMEX RK schemes and show that, under suitable initial conditions, they can capture the NS limit without resolving the small parameter ε, i.e., $\varepsilon = o(\Delta t)$, $\Delta t^m = o(\varepsilon)$, where m is the order of the explicit RK part in the IMEX scheme. Extensive numerical tests for BGK and ES-BGK models are performed to verify our theoretical results.

Keywords Boltzmann equation · BGK/ES-BGK models · IMEX Runge–Kutta schemes · Compressible Euler equations · Navier–Stokes equations

Mathematics Subject Classification 35Q20 · 65L06 · 65L04 · 35Q30 · 35Q31

1 Introduction

The Boltzmann equation is the fundamental equation in kinetic theory. It describes the non-equilibrium dynamics of gas or a system comprised of a large number of particles using a probability distribution function $f(t, x, v)$, where t is time, x is space, and v is (particle)

J. Hu's 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. X. Zhang's research was supported by NSF Grant DMS-1522593.

✉ Jingwei Hu
jingweihu@purdue.edu

Xiangxiong Zhang
zhan1966@purdue.edu

[1] Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA

velocity. After nondimensionalization, the Boltzmann equation reads [9,10,26]:

$$\frac{\partial f}{\partial t} + v \cdot \nabla_x f = \frac{1}{\varepsilon}\mathcal{Q}(f), \quad t \geq 0, \quad x \in \Omega \subset \mathbb{R}^{d_x}, \quad v \in \mathbb{R}^{d_v}. \tag{1.1}$$

Here ε is the Knudsen number, defined as the ratio of the mean free path and the characteristic length scale. $\mathcal{Q}(f)$ is the collision operator—a high-dimensional, nonlinear integral operator modeling the binary collisions among particles. When ε is small (the system is close to continuum regime), one can perform a Chapman-Enskog expansion on (1.1) to derive the compressible Euler equations (Eq. (1.2) without $O(\varepsilon)$ terms) and the Navier–Stokes equations as the leading and the next order asymptotics [3]:

$$\begin{cases} \dfrac{\partial \rho}{\partial t} + \nabla_x \cdot (\rho u) = 0, \\ \dfrac{\partial (\rho u)}{\partial t} + \nabla_x \cdot (\rho u \otimes u + p\mathrm{Id}) = \varepsilon \nabla_x \cdot (\mu \sigma(u)), \\ \dfrac{\partial E}{\partial t} + \nabla_x \cdot ((E+p)u) = \varepsilon \nabla_x \cdot (\mu \sigma(u) u + \kappa \nabla_x T), \end{cases} \tag{1.2}$$

where ρ is density, u is bulk velocity, T is temperature, $p = \rho T$ is pressure, Id is the identity matrix, $E = \frac{d_v}{2}\rho T + \frac{1}{2}\rho u^2$ is total energy, $\sigma(u) = \nabla_x u + (\nabla_x u)^T - \frac{2}{d_v}\nabla_x \cdot u\mathrm{Id}$ ($\nabla_x u$ is a matrix with ij-th component given by $\frac{\partial u_i}{\partial x_j}$), and μ and κ are, respectively, coefficients of viscosity and heat conductivity.

The complexity of the Boltzmann collision operator makes it extremely difficult and expensive for numerical simulation. Therefore, different simpler kinetic models have been proposed to mimic the main properties of the full integral operator. The BGK model [5] assumes a simple relaxation toward the Maxwellian equilibrium:

$$\mathcal{Q}(f) = \frac{\rho T}{\mu}(\mathcal{M}[f] - f), \tag{1.3}$$

where

$$\mathcal{M}[f] = \frac{\rho}{(2\pi T)^{\frac{d_v}{2}}} \exp\left(-\frac{|v-u|^2}{2T}\right), \tag{1.4}$$

with ρ, u, T defined by

$$\rho = \int_{\mathbb{R}^{d_v}} f \, \mathrm{d}v, \quad u = \frac{1}{\rho}\int_{\mathbb{R}^{d_v}} f v \, \mathrm{d}v, \quad T = \frac{1}{d_v \rho}\int_{\mathbb{R}^{d_v}} f|v-u|^2 \, \mathrm{d}v. \tag{1.5}$$

Although it describes the right fluid limit, the BGK model does not give the correct Prandtl number. To correct this defect, the so-called ES-BGK model was introduced by Holway [17], where the Maxwellian is replaced by a Gaussian distribution:

$$\mathcal{Q}(f) = \frac{\rho T}{\mu(1-\nu)}(\mathcal{G}[f] - f), \tag{1.6}$$

where $-\frac{1}{2} \leq \nu < 1$ is a parameter, and

$$\mathcal{G}[f] = \frac{\rho}{\sqrt{\det(2\pi\mathcal{T})}} \exp\left(-\frac{1}{2}(v-u)^T \mathcal{T}^{-1}(v-u)\right), \tag{1.7}$$

with the corrected tensor $\mathcal{T}$ defined by

$$\mathcal{T} = (1-\nu)T\mathrm{Id} + \nu\Theta, \quad \Theta = \frac{1}{\rho}\int_{\mathbb{R}^{d_v}} f(v-u)\otimes(v-u)\, \mathrm{d}v. \tag{1.8}$$

More details about this model can be found in [1].

The BGK/ES-BGK models are greatly simplified compared to the full Boltzmann equation, hence are widely used in various science and engineering applications. Nevertheless, in the presence of small Knudsen number, the numerical simulation of these equations can still be very expensive: the stiff collision term would require an over restrictive time step in a typical explicit scheme. As such, implicit discretization of the collision term is often preferred which allows Δt to be chosen independently of ε (using the conservation property of the collision operator, the implicit $\mathcal{M}$ or $\mathcal{G}$ can be evaluated in an explicit manner without iteration [11,15,24], see Sect. 2). On the other hand, the convection part is non-stiff and can be treated explicitly. In view of these considerations, it is natural to apply implicit–explicit (IMEX) time discretization schemes (c.f. [2]).

The past decades have seen significant development of IMEX schemes for stiff hyperbolic and kinetic equations. Without being exhaustive, we refer to [7,8,12,20,23]. In almost all these works, the main concern is to guarantee the numerical scheme captures the correct macroscopic limit as $\varepsilon \to 0$, i.e., *asymptotic-preserving* [18,21]. In the current context, this means the numerical scheme for (1.1) should become a consistent discretization of the compressible Euler equations (Eq. (1.2) without $O(\varepsilon)$ terms) when $\varepsilon \to 0$ and Δt, Δx being fixed. However, for many science/engineering problems, ε is small but not zero, hence it is very important to also capture the Navier–Stokes (NS) limit (1.2). As commented in [14,18], since the viscous terms are of $O(\varepsilon)$, in general one cannot expect to capture the NS solution with under-resolved mesh sizes and time steps. Yet the situation could be different for high order methods, and this is exactly the motivation of this work. Specifically, we will study the asymptotic behavior of a class of *existing* IMEX Runge–Kutta (RK) schemes for BGK/ES-BGK equations, and prove that they can capture the NS limit without resolving ε, i.e., $\varepsilon = o(\Delta t)$, $\Delta t^m = o(\varepsilon)$, where m is the order of the explicit RK part in the IMEX scheme.

We mention a few related works that have addressed the issue of NS limit to some extent. [4,27] considered a micro-macro decomposition of the BGK equation and then applied the IMEX schemes to the resulting coupled system. These schemes naturally capture the NS limit as the information at the NS level (micro part) is computed directly. However, they are more complicated than solving the BGK equation itself. The very recent work [6] focused on a similar problem as ours: using the linear hyperbolic relaxation system as a prototype, they performed the asymptotic expansion up to $O(\varepsilon)$ for the numerical method, and imposed extra order conditions on the IMEX scheme in order to get a consistent discretization to the diffusion limit. The conditions derived are sufficient but not necessary. In fact, the new IMEX schemes found in [6] require more stages than the commonly used ones [2]. Another recent work [13] considered the IMEX multistep methods for stiff kinetic equations, where the schemes are shown to be able to capture the NS limit under suitable conditions. Although the analysis for multistep methods are easier than Runge–Kutta methods, the former often imposes stronger stability constraints.

The rest of this paper is organized as follows. In Sect. 2, we describe the IMEX RK schemes for the ES-BGK equation along with the characterization of different IMEX schemes. Section 3 proves our main result regarding the NS limit. Extensive numerical examples are presented in Sect. 4 to validate our theoretical finding. The paper is concluded in Sect. 5.

2 IMEX RK Schemes for the BGK/ES-BGK Equations

We first briefly describe the general IMEX RK schemes applied to the stiff kinetic equation (1.1). We will use the ES-BGK model (1.6) as an example (the BGK model is a special case when $\nu = 0$). Define $\tau = \frac{\rho T}{\mu(1-\nu)}$, the scheme reads [12]:

$$f^{(i)} = f^n - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} v \cdot \nabla_x f^{(j)} + \frac{\Delta t}{\varepsilon} \sum_{j=1}^{i} a_{ij} \tau^{(j)} (\mathcal{G}[f^{(j)}] - f^{(j)}), \quad i = 1, \ldots, s, \tag{2.1}$$

$$f^{n+1} = f^n - \Delta t \sum_{i=1}^{s} \tilde{w}_i v \cdot \nabla_x f^{(i)} + \frac{\Delta t}{\varepsilon} \sum_{i=1}^{s} w_i \tau^{(i)} (\mathcal{G}[f^{(i)}] - f^{(i)}). \tag{2.2}$$

Here the matrices $\tilde{A} = (\tilde{a}_{ij})$, $\tilde{a}_{ij} = 0$ for $j \geq i$ and $A = (a_{ij})$, $a_{ij} = 0$ for $j > i$ are $s \times s$ matrices such that the scheme is explicit for the convection part and implicit for the collision part. Along with the coefficient vectors $\tilde{\mathbf{w}} = (\tilde{w}_1, \ldots, \tilde{w}_s)^T$, $\mathbf{w} = (w_1, \ldots, w_s)^T$, they can be represented by a double Butcher tableau:

$$\begin{array}{c|c} \tilde{\mathbf{c}} & \tilde{A} \\ \hline & \tilde{\mathbf{w}}^T \end{array} \qquad \begin{array}{c|c} \mathbf{c} & A \\ \hline & \mathbf{w}^T \end{array} \tag{2.3}$$

where the vectors $\tilde{\mathbf{c}} = (\tilde{c}_1, \ldots, \tilde{c}_s)^T$, $\mathbf{c} = (c_1, \ldots, c_s)^T$ are defined as

$$\tilde{c}_i = \sum_{j=1}^{i-1} \tilde{a}_{ij}, \quad c_i = \sum_{j=1}^{i} a_{ij}. \tag{2.4}$$

At every stage of (2.1), since the collision part is implicit, one has to obtain $\tau^{(i)}$ and $\mathcal{G}[f^{(i)}]$ first in order to evaluate $f^{(i)}$. This can be achieved by taking the moments $\langle \cdot \phi \rangle := \int \cdot \phi(v)\, \mathrm{d}v$ with $\phi(v) = (1, v, |v|^2/2)^T$ on both sides of the scheme, which yields [11,24]:

$$\langle \phi f^{(i)} \rangle = \langle \phi f^n \rangle - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla_x \cdot \langle v \phi f^{(j)} \rangle. \tag{2.5}$$

The implicit part is gone since the Gaussian $\mathcal{G}[f]$ defined in (1.7) has the following properties:

$$\int_{\mathbb{R}^{d_v}} \mathcal{G}[f]\, \mathrm{d}v = \int_{\mathbb{R}^{d_v}} f\, \mathrm{d}v = \rho, \tag{2.6}$$

$$\int_{\mathbb{R}^{d_v}} v \mathcal{G}[f]\, \mathrm{d}v = \int_{\mathbb{R}^{d_v}} v f\, \mathrm{d}v = \rho u, \tag{2.7}$$

$$\int_{\mathbb{R}^{d_v}} \frac{|v|^2}{2} \mathcal{G}[f]\, \mathrm{d}v = \int_{\mathbb{R}^{d_v}} \frac{|v|^2}{2} f\, \mathrm{d}v = E, \tag{2.8}$$

$$\int_{\mathbb{R}^{d_v}} (v-u) \otimes (v-u) \mathcal{G}[f]\, \mathrm{d}v = \rho \mathcal{T}. \tag{2.9}$$

Hence one can obtain the macroscopic quantities ρ, u, T at stage i using (2.5), which will define $\tau^{(i)}$ accordingly. To find $\mathcal{G}[f^{(i)}]$, one needs an additional quantity $\Theta^{(i)}$ (1.8). Let $\Sigma = \langle v \otimes v f \rangle = \rho(u \otimes u + \Theta)$, then $\Theta^{(i)}$ can be obtained by finding $\Sigma^{(i)}$ [15]. By taking the moment $\langle \cdot\, v \otimes v \rangle$ on (2.1) and using the facts that

$$\int_{\mathbb{R}^{d_v}} v \otimes v \mathcal{G}[f]\, \mathrm{d}v = \rho(\mathcal{T} + u \otimes u), \tag{2.10}$$

and

$$\rho\mathcal{T} = \rho[(1-\nu)T\mathrm{Id} + \nu\Theta] = \rho(1-\nu)T\mathrm{Id} + \nu\Sigma - \nu\rho u \otimes u, \tag{2.11}$$

we have

$$\begin{aligned}\Sigma^{(i)} &= \Sigma^n - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla_x \cdot \langle v \otimes v v f^{(j)} \rangle \\ &\quad + \frac{\Delta t}{\varepsilon} \sum_{j=1}^{i} a_{ij} \tau^{(j)} (1-\nu) \left[\rho^{(j)} (T^{(j)} \mathrm{Id} + u^{(j)} \otimes u^{(j)}) - \Sigma^{(j)} \right],\end{aligned} \tag{2.12}$$

thus we can find $\Sigma^{(i)}$ as

$$\begin{aligned}\Sigma^{(i)} = c \Bigg[&\Sigma^n - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla_x \cdot \langle v \otimes v v f^{(j)} \rangle \\ &+ \frac{\Delta t}{\varepsilon} \sum_{j=1}^{i-1} a_{ij} \tau^{(j)} (1-\nu) \left[\rho^{(j)} (T^{(j)} \mathrm{Id} + u^{(j)} \otimes u^{(j)}) - \Sigma^{(j)} \right] \Bigg] \\ &+ (1-c) \rho^{(i)} (T^{(i)} \mathrm{Id} + u^{(i)} \otimes u^{(i)}), \quad \text{with } c = \frac{\varepsilon}{\varepsilon + (1-\nu) a_{ii} \tau^{(i)} \Delta t}.\end{aligned} \tag{2.13}$$

Some preliminary notions about the IMEX RK schemes are necessary before we discuss their asymptotic properties. First of all, the double Butcher tableau must satisfy the order conditions (standard order conditions for each tableau and coupling conditions) [16,23]. Then according to the structure of matrix A in the implicit tableau, one can classify the IMEX schemes into following categories [7,12]:

- **Type A** if the matrix A is invertible.
- **Type CK** if the matrix A can be written as

$$\begin{pmatrix} 0 & 0 \\ \mathbf{a} & \hat{A} \end{pmatrix}, \tag{2.14}$$

  and the submatrix $\hat{A} \in \mathbb{R}^{(s-1)\times(s-1)}$ is invertible; in particular, if the vector $\mathbf{a} = 0$, $w_1 = 0$, the scheme is of type ARS.
- If $a_{si} = w_i, \tilde{a}_{si} = \tilde{w}_i, i = 1, \ldots, s$, i.e., $f^{n+1} = f^{(s)}$, the scheme is said to be **globally stiffly accurate (GSA)**.

Next we list a few examples of these schemes (only type CK and GSA schemes are listed here as our following analysis applies to this class). We use (s, σ, p) to denote an IMEX method, where s is the number of stages in the explicit scheme, σ is the number of stages in the implicit scheme, and p is the order of the IMEX scheme.

- ARS(4,4,3) in [2]:

0	0	0	0	0	0
1/2	1/2	0	0	0	0
2/3	11/18	1/18	0	0	0
1/2	5/6	−5/6	1/2	0	0
1	1/4	7/4	3/4	−7/4	0
	1/4	7/4	3/4	−7/4	0

0	0	0	0	0	0
1/2	0	1/2	0	0	0
2/3	0	1/6	1/2	0	0
1/2	0	−1/2	1/2	1/2	0
1	0	3/2	−3/2	1/2	1/2
	0	3/2	−3/2	1/2	1/2

- ARS(2,2,2) in [2]:

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & \gamma & 0 & 0 \\ 1 & \delta & 1-\delta & 0 \\ \hline & \delta & 1-\delta & 0 \end{array} \quad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ \gamma & 0 & \gamma & 0 \\ 1 & 0 & 1-\gamma & \gamma \\ \hline & 0 & 1-\gamma & \gamma \end{array}, \quad \gamma = 1-\frac{\sqrt{2}}{2}, \delta = 1-\frac{1}{2\gamma}$$

- BPR(3,5,3) in [7]:

$$\begin{array}{c|ccccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 2/3 & 4/9 & 2/9 & 0 & 0 & 0 \\ 1 & 1/4 & 0 & 3/4 & 0 & 0 \\ 1 & 1/4 & 0 & 3/4 & 0 & 0 \\ \hline & 1/4 & 0 & 3/4 & 0 & 0 \end{array} \quad \begin{array}{c|ccccc} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1/2 & 1/2 & 0 & 0 & 0 \\ 2/3 & 5/18 & -1/9 & 1/2 & 0 & 0 \\ 1 & 1/2 & 0 & 0 & 1/2 & 0 \\ 1 & 1/4 & 0 & 3/4 & -1/2 & 1/2 \\ \hline & 1/4 & 0 & 3/4 & -1/2 & 1/2 \end{array}$$

- LRR(2,3,2) in [22]:

$$\begin{array}{c|ccccc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 & 0 \\ 1/3 & 1/3 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ \hline & 0 & 1 & 0 & 0 \end{array} \quad \begin{array}{c|ccccc} 0 & 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 & 0 \\ 1/3 & 0 & 0 & 1/3 & 0 \\ 1 & 0 & 0 & 3/4 & 1/4 \\ \hline & 0 & 0 & 3/4 & 1/4 \end{array}$$

- A second order scheme used in [14], we call it IMEX-II-GSA(2,3,2):

$$\begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ \hline & 0 & 1 & 0 \end{array} \quad \begin{array}{c|ccc} 0 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 1 & 1/2 & 0 & 1/2 \\ \hline & 1/2 & 0 & 1/2 \end{array}$$

- IMEX-II-GSA2(4,4,2) in [6]:

$$\begin{array}{c|cccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 & 0 & 0 & 0 \\ 1/3 & 1/6 & 1/6 & 0 & 0 & 0 & 0 \\ 2/3 & -2/3 & 0 & 4/3 & 0 & 0 & 0 \\ 1 & -1/16 & 1/2 & 0 & 9/16 & 0 & 0 \\ \hline & -1/16 & 1/2 & 0 & 9/16 & 0 & 0 \end{array} \quad \begin{array}{c|cccccc} 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/4 & 0 & 1/4 & 0 & 0 & 0 & 0 \\ 1/3 & 0 & 1/12 & 1/4 & 0 & 0 & 0 \\ 2/3 & 0 & -11/12 & 4/3 & 1/4 & 0 & 0 \\ 1 & 0 & 9/31 & 12/31 & 9/124 & 1/4 & 0 \\ \hline & 0 & 9/31 & 12/31 & 9/124 & 1/4 & 0 \end{array}$$

3 Asymptotic Properties of the IMEX RK Schemes

In this section, we discuss in detail the asymptotic properties of the IMEX RK scheme (2.1), (2.2) with respect to the Navier–Stokes limit. For completeness, we first briefly state and prove the results regarding the Euler limit since preserving the leading order asymptotics is prior. More detailed discussion can be found in [12] (IMEX RK applied to the BGK equation) and [15] (first order IMEX applied to the ES-BGK equation).

3.1 Preserving the Euler Limit

For ease of presentation, we rewrite the scheme (2.1), (2.2) using vector notations:

$$\mathbf{F} = f^n \mathbf{e} - \Delta t \tilde{A} v \cdot \nabla_x \mathbf{F} + \frac{\Delta t}{\varepsilon} A \tau (\mathcal{G}[\mathbf{F}] - \mathbf{F}), \tag{3.1}$$

$$f^{n+1} = f^n - \Delta t \tilde{\mathbf{w}}^T v \cdot \nabla_x \mathbf{F} + \frac{\Delta t}{\varepsilon} \mathbf{w}^T \tau (\mathcal{G}[\mathbf{F}] - \mathbf{F}), \tag{3.2}$$

where $\mathbf{F} := (f^{(1)}, \ldots, f^{(s)})^T$, $\mathbf{e} := (1, \ldots, 1)^T$, $\mathcal{G}[\mathbf{F}] := (\mathcal{G}[f^{(1)}], \ldots, \mathcal{G}[f^{(s)}])^T$, and $\tau := \mathrm{diag}(\tau^{(1)}, \ldots, \tau^{(s)})$ is a diagonal matrix. Taking the moments $\langle \cdot \phi \rangle$ on both sides of (3.1), (3.2) yields

$$\langle \phi \mathbf{F} \rangle = \langle \phi f^n \rangle \mathbf{e} - \Delta t \tilde{A} \nabla_x \cdot \langle v \phi \mathbf{F} \rangle, \tag{3.3}$$

$$\langle \phi f^{n+1} \rangle = \langle \phi f^n \rangle - \Delta t \tilde{\mathbf{w}}^T \nabla_x \cdot \langle v \phi \mathbf{F} \rangle. \tag{3.4}$$

For proving the asymptotic properties of schemes solving the ES-BGK equation, we need the following lemma:

Lemma 3.1 $f = \mathcal{G}[f] \Longleftrightarrow f = \mathcal{M}[f]$.

Proof "$\Longrightarrow$": Taking the moment $\frac{1}{\rho}\langle \cdot (v-u) \otimes (v-u) \rangle$ on both sides of $f = \mathcal{G}[f]$ yields

$$\Theta = \mathcal{T} = (1-\nu) T \mathrm{Id} + \nu \Theta \Rightarrow \Theta = T \mathrm{Id},$$

hence $\mathcal{T} = (1-\nu) T \mathrm{Id} + \nu \Theta = T \mathrm{Id}$. When $\mathcal{T} = T \mathrm{Id}$, $\mathcal{G}$ is just the isotropic Maxwellian $\mathcal{M}$, thus $f = \mathcal{G}[f] = \mathcal{M}[f]$.

" $\Longleftarrow$": Taking the moment $\frac{1}{\rho}\langle \cdot (v-u) \otimes (v-u) \rangle$ on both sides of $f = \mathcal{M}[f]$ yields $\Theta = T \mathrm{Id}$ directly. The rest follows the same as above. □

Regarding the Euler limit, we have the following results for IMEX schemes of type A and type CK, respectively.

Proposition 3.2 *If the IMEX scheme* (3.1), (3.2) *is of type A, then for fixed* Δt*, in the limit* $\varepsilon \to 0$*, the scheme becomes the explicit RK scheme characterized by* $(\tilde{A}, \tilde{\mathbf{w}})$ *applied to the limit Euler system (Eq.* (1.2) *without* $O(\varepsilon)$ *terms). If the scheme is additionally GSA, then*

$$\lim_{\varepsilon \to 0} f^{n+1} = \lim_{\varepsilon \to 0} \mathcal{M}[f^{n+1}]. \tag{3.5}$$

Proof Formally passing the limit $\varepsilon \to 0$ in (3.1), one has $\Delta t A \tau (\mathcal{G}[\mathbf{F}] - \mathbf{F}) = 0$. (For convenience we abuse notations by removing $\lim_{\varepsilon \to 0}$: here $\mathbf{F}$ and $\mathcal{G}[\mathbf{F}]$ should be understood as the limiting values for $\varepsilon \to 0$, and similarly for the notations in the following arguments.) This implies $\mathbf{F} = \mathcal{G}[\mathbf{F}]$ since A and τ are invertible. Then by Lemma 3.1, we know $\mathbf{F} = \mathcal{M}[\mathbf{F}]$. Therefore, as $\varepsilon \to 0$, the moment equations (3.3), (3.4) become

$$\langle \phi \mathbf{F} \rangle = \langle \phi f^n \rangle \mathbf{e} - \Delta t \tilde{A} \nabla_x \cdot \langle v \phi \mathcal{M}[\mathbf{F}] \rangle, \tag{3.6}$$

$$\langle \phi f^{n+1} \rangle = \langle \phi f^n \rangle - \Delta t \tilde{\mathbf{w}}^T \nabla_x \cdot \langle v \phi \mathcal{M}[\mathbf{F}] \rangle, \tag{3.7}$$

which is the explicit RK scheme characterized by $(\tilde{A}, \tilde{\mathbf{w}})$ applied to the compressible Euler equations. If the scheme is additionally GSA, then $f^{n+1} = f^{(s)}$. Hence (3.5) is straightforward. □

Proposition 3.3 *If the IMEX scheme* (3.1), (3.2) *is of type CK and GSA, then for fixed* Δt *and consistent initial data:*

$$\lim_{\varepsilon \to 0} f^0(x,v) = \lim_{\varepsilon \to 0} \mathcal{G}[f^0(x,v)] \quad or \quad \lim_{\varepsilon \to 0} f^0(x,v) = \lim_{\varepsilon \to 0} \mathcal{M}[f^0(x,v)], \tag{3.8}$$

in the limit $\varepsilon \to 0$*, the scheme becomes the explicit RK scheme characterized by* $(\tilde{A}, \tilde{\mathbf{w}})$ *applied to the limit Euler system (Eq. (1.2) without* $O(\varepsilon)$ *terms). Furthermore,*

$$\lim_{\varepsilon \to 0} f^{n+1} = \lim_{\varepsilon \to 0} \mathcal{M}[f^{n+1}]. \tag{3.9}$$

Proof If a scheme is of type CK and GSA, then $f^{(1)} = f^n$, $f^{n+1} = f^{(s)}$. Rewrite $\mathbf{F} = (f^{(1)}, \hat{\mathbf{F}})$, $\mathbf{e} = (1, \hat{\mathbf{e}})$, $\mathcal{G}[\mathbf{F}] = (\mathcal{G}[f^{(1)}], \mathcal{G}[\hat{\mathbf{F}}])$, $\hat{\tau} := \mathrm{diag}(\tau^{(2)}, \ldots, \tau^{(s)})$, then (3.1) becomes

$$\hat{\mathbf{F}} = f^n \hat{\mathbf{e}} - \Delta t \tilde{\mathbf{a}} v \cdot \nabla_x f^n - \Delta t \hat{\tilde{A}} v \cdot \nabla_x \hat{\mathbf{F}} + \frac{\Delta t}{\varepsilon} \mathbf{a} \tau^n (\mathcal{G}[f^n] - f^n) + \frac{\Delta t}{\varepsilon} \hat{A} \hat{\tau} (\mathcal{G}[\hat{\mathbf{F}}] - \hat{\mathbf{F}}), \tag{3.10}$$

where we have used a similar notation for matrix $\tilde{A}$ as that in (2.14):

$$\begin{pmatrix} 0 & 0 \\ \tilde{\mathbf{a}} & \hat{\tilde{A}} \end{pmatrix}. \tag{3.11}$$

Taking the moments $\langle \cdot \phi \rangle$ on both sides of (3.10) yields

$$\langle \phi \hat{\mathbf{F}} \rangle = \langle \phi f^n \rangle \hat{\mathbf{e}} - \Delta t \tilde{\mathbf{a}} \nabla_x \cdot \langle v \phi f^n \rangle - \Delta t \hat{\tilde{A}} \nabla_x \cdot \langle v \phi \hat{\mathbf{F}} \rangle. \tag{3.12}$$

Now sending $\varepsilon \to 0$ in (3.10), one has $\Delta t \mathbf{a} \tau^n (\mathcal{G}[f^n] - f^n) + \Delta t \hat{A} \hat{\tau} (\mathcal{G}[\hat{\mathbf{F}}] - \hat{\mathbf{F}}) = 0$, which reduces to $\Delta t \hat{A} \hat{\tau} (\mathcal{G}[\hat{\mathbf{F}}] - \hat{\mathbf{F}}) = 0$ for consistent initial data $f^n = \mathcal{G}[f^n]$ or $f^n = \mathcal{M}[f^n]$ (by Lemma 3.1, these two are equivalent). (Note here again we have abused notations: $\hat{\mathbf{F}}$, $\mathcal{G}[\hat{\mathbf{F}}]$, f^n, $\mathcal{G}[f^n]$ and $\mathcal{M}[f^n]$ should all be understood as the limiting values for $\varepsilon \to 0$, and similarly for notations in the following arguments.) This further implies $\hat{\mathbf{F}} = \mathcal{G}[\hat{\mathbf{F}}]$ since $\hat{A}$ and $\hat{\tau}$ are invertible. Again by Lemma 3.1, we know $\hat{\mathbf{F}} = \mathcal{M}[\hat{\mathbf{F}}]$. Therefore, in the limit, the moment equation (3.12) becomes

$$\langle \phi \hat{\mathbf{F}} \rangle = \langle \phi f^n \rangle \hat{\mathbf{e}} - \Delta t \tilde{\mathbf{a}} \nabla_x \cdot \langle v \phi \mathcal{M}[f^n] \rangle - \Delta t \hat{\tilde{A}} \nabla_x \cdot \langle v \phi \mathcal{M}[\hat{\mathbf{F}}] \rangle, \tag{3.13}$$

i.e., the explicit RK scheme characterized by $(\tilde{A}, \tilde{\mathbf{w}})$ applied to the limit Euler system. Furthermore, since the scheme is GSA, we have $f^{n+1} = \mathcal{M}[f^{n+1}]$, and thus the initial data remains consistent at the next time step. □

3.2 Preserving the Navier–Stokes Limit

To discuss the Navier–Stokes limit, we need the following lemmas.

Lemma 3.4 *$f = \mathcal{G}[f] + O(\varepsilon)$ implies $\mathcal{G}[f] = \mathcal{M}[f] + O(\varepsilon)$.*

Proof The proof is similar to Lemma 3.1. We omit the detail. □

Given a Maxwellian function $\mathcal{M}[f]$, define $\Pi_{\mathcal{M}}$ to be the orthogonal projection to the space

$$L_{\mathcal{M}} = \mathrm{span}\{\mathcal{M}, v\mathcal{M}, |v|^2 \mathcal{M}\}, \tag{3.14}$$

with the inner product defined by $(f, g) := \int f g \frac{1}{\mathcal{M}} \mathrm{d}v$. Then a direct calculation shows that

$$\begin{aligned} (I - \Pi_{\mathcal{M}})(v \cdot \nabla_x \mathcal{M}) = \mathcal{M} \Bigg[& \left(\frac{|v-u|^2}{2T} - \frac{d_v + 2}{2} \right) \frac{(v-u) \cdot \nabla_x T}{T} \\ & + \left(\frac{(v-u) \otimes (v-u)}{T} - \frac{|v-u|^2}{d_v T} \mathrm{Id} \right) : \nabla_x u \Bigg], \end{aligned} \tag{3.15}$$

where the operation : between two matrices is defined as $A : B = \sum_{ij} a_{ij} b_{ij}$. Computing moments of (3.15) yields the following result.

Lemma 3.5

$$\int_{\mathbb{R}^{d_v}} (I - \Pi_{\mathcal{M}})(v \cdot \nabla_x \mathcal{M})(v-u) \otimes (v-u)\,\mathrm{d}v = \rho T \sigma(u), \tag{3.16}$$

$$\int_{\mathbb{R}^{d_v}} (I - \Pi_{\mathcal{M}})(v \cdot \nabla_x \mathcal{M}) \frac{1}{2}(v-u)|v-u|^2\,\mathrm{d}v = \frac{d_v+2}{2}\rho T \nabla_x T, \tag{3.17}$$

where $\sigma(u)$ is the tensor defined in Sect. 1.

Lemma 3.6 *For the IMEX scheme (3.1), (3.2), one has*

$$\mathcal{M}[\mathbf{F}] = \mathcal{M}[f^n]\mathbf{e} + O(\Delta t), \quad (I - \Pi_{\mathcal{M}[f^n]})(\mathcal{M}[\mathbf{F}]) = O(\Delta t^2). \tag{3.18}$$

Proof Using the differential form of $\mathcal{M}$

$$\mathrm{d}\mathcal{M} = \mathcal{M}\left[\frac{1}{\rho}\mathrm{d}\rho + \frac{(v-u)}{T}\cdot \mathrm{d}u + \left(\frac{(v-u)^2}{2T^2} - \frac{d_v}{2T}\right)\mathrm{d}T\right], \tag{3.19}$$

we have for every $1 \le i \le s$,

$$\begin{aligned}\mathcal{M}[f^{(i)}] = \mathcal{M}[f^n] + \mathcal{M}[f^n]\Bigg[&\frac{1}{\rho^n}(\rho^{(i)} - \rho^n) + \frac{v-u^n}{T^n} \\ &\cdot (u^{(i)} - u^n) + \left(\frac{(v-u^n)^2}{2(T^n)^2} - \frac{d_v}{2T^n}\right)(T^{(i)} - T^n)\Bigg] \\ &+ O((U^{(i)} - U^n)^2),\end{aligned} \tag{3.20}$$

where we used the vector $U := (\rho, \rho u, E)^T$ to represent the macroscopic variables that define the Maxwellian. Therefore,

$$(I - \Pi_{\mathcal{M}[f^n]})\mathcal{M}[f^{(i)}] = O((U^{(i)} - U^n)^2). \tag{3.21}$$

Now from (3.3) we know $\langle \phi \mathbf{F}\rangle - \langle \phi f^n\rangle \mathbf{e} = O(\Delta t)$. This means

$$U^{(i)} - U^n = O(\Delta t), \tag{3.22}$$

which yields the assertion using (3.20) and (3.21). □

We are ready to present the main result.

Theorem 3.7 *If the IMEX scheme (2.1), (2.2) (or its vector form (3.1), (3.2)) is of type CK and GSA and satisfies $\mathbf{c} = \tilde{\mathbf{c}}$, and assume*

$$\mathcal{G}[f^{(i)}] = \mathcal{G}[f^n] + O(\Delta t), \quad 1 \le i \le s, \tag{3.23}$$

then for consistent initial data:

$$f^0(x,v) = \mathcal{G}[f^0] - \frac{\varepsilon}{\tau^0}(I - \Pi_{\mathcal{M}[f^0]})(v \cdot \nabla_x \mathcal{M}[f^0]) + o(\varepsilon), \tag{3.24}$$

and $\varepsilon = o(\Delta t)$, one has

$$f^{(i)} = \mathcal{G}[f^{(i)}] - \frac{\varepsilon}{\tau^n}(I - \Pi_{\mathcal{M}[f^n]})(v \cdot \nabla_x \mathcal{M}[f^n]) + O(\varepsilon \Delta t) + O\left(\frac{\varepsilon^2}{\Delta t}\right), \quad 1 \le i \le s. \tag{3.25}$$

Furthermore, the resulting macroscopic scheme is a consistent discretization to the Navier–Stokes equations (1.2) with the local truncation error:

$$LTE = O(\Delta t^m) + O(\varepsilon \Delta t) + O\left(\frac{\varepsilon^2}{\Delta t}\right), \tag{3.26}$$

where m is the order of the explicit RK scheme. Therefore, in order to capture the NS limit, one needs $LTE = o(\varepsilon)$ which is satisfied if $\Delta t^m = o(\varepsilon)$.

Proof First of all, we write $f^n = \mathcal{G}[f^n] + \varepsilon g^n$, $\hat{\mathbf{F}} = \mathcal{G}[\hat{\mathbf{F}}] + \varepsilon \hat{\mathbf{g}}$, then (3.10) becomes

$$\begin{aligned}\mathcal{G}[\hat{\mathbf{F}}] + \varepsilon\hat{\mathbf{g}} = \mathcal{G}[f^n]\hat{\mathbf{e}} + \varepsilon g^n\hat{\mathbf{e}} - \Delta t\tilde{\mathbf{a}}v\cdot\nabla_x\mathcal{G}[f^n] - \varepsilon\Delta t\tilde{\mathbf{a}}v\cdot\nabla_x g^n\\ - \Delta t\hat{\tilde{A}}v\cdot\nabla_x\mathcal{G}[\hat{\mathbf{F}}] - \varepsilon\Delta t\hat{\tilde{A}}v\cdot\nabla_x\hat{\mathbf{g}} - \Delta t\mathbf{a}\boldsymbol{\tau}^n g^n - \Delta t\hat{A}\hat{\tau}\hat{\mathbf{g}}.\end{aligned} \tag{3.27}$$

Before we prove the main assertion, we have to make sure the scheme (3.27) gives a well-defined $\hat{\mathbf{g}}$. We will show that if $g^n = O(1)$ then $\hat{\mathbf{g}} = O(1)$ (so $g^{n+1} = O(1)$). This can be seen by writing (3.27) as

$$\begin{aligned}\left(\frac{\varepsilon}{\Delta t}I + \hat{A}\hat{\tau}\right)\hat{\mathbf{g}} = \frac{\mathcal{G}[f^n]\hat{\mathbf{e}} - \mathcal{G}[\hat{\mathbf{F}}]}{\Delta t} - \tilde{\mathbf{a}}v\cdot\nabla_x\mathcal{G}[f^n] - \hat{\tilde{A}}v\cdot\nabla_x\mathcal{G}[\hat{\mathbf{F}}]\\ + \left(\frac{\varepsilon}{\Delta t}\hat{\mathbf{e}} - \mathbf{a}\boldsymbol{\tau}^n\right)g^n - \varepsilon\tilde{\mathbf{a}}v\cdot\nabla_x g^n - \varepsilon\hat{\tilde{A}}v\cdot\nabla_x\hat{\mathbf{g}},\end{aligned} \tag{3.28}$$

and noting $\left(\frac{\varepsilon}{\Delta t}I + \hat{A}\hat{\tau}\right)^{-1} = \hat{\tau}^{-1}\hat{A}^{-1} + O\left(\frac{\varepsilon}{\Delta t}\right)$ and the assumption (3.23).

Now by Lemma 3.4, we have $\mathcal{G}[f^n] = \mathcal{M}[f^n] + O(\varepsilon)$, $\mathcal{G}[\hat{\mathbf{F}}] = \mathcal{M}[\hat{\mathbf{F}}] + O(\varepsilon)$. Using these in (3.27) and neglecting $O(\varepsilon)$ terms, one has

$$\mathcal{M}[\hat{\mathbf{F}}] + O(\varepsilon) = \mathcal{M}[f^n]\hat{\mathbf{e}} - \Delta t\tilde{\mathbf{a}}v\cdot\nabla_x\mathcal{M}[f^n] - \Delta t\hat{\tilde{A}}v\cdot\nabla_x\mathcal{M}[\hat{\mathbf{F}}] - \Delta t\mathbf{a}\boldsymbol{\tau}^n g^n - \Delta t\hat{A}\hat{\tau}\hat{\mathbf{g}}. \tag{3.29}$$

Applying the operator $I - \Pi_{\mathcal{M}[f^n]}$ on both sides of (3.29) yields

$$\begin{aligned}(I - \Pi_{\mathcal{M}[f^n]})\mathcal{M}[\hat{\mathbf{F}}] + O(\varepsilon) = -\Delta t\tilde{\mathbf{a}}(I - \Pi_{\mathcal{M}[f^n]})v\cdot\nabla_x\mathcal{M}[f^n]\\ - \Delta t\hat{\tilde{A}}(I - \Pi_{\mathcal{M}[f^n]})v\cdot\nabla_x\mathcal{M}[\hat{\mathbf{F}}] - \Delta t\mathbf{a}\boldsymbol{\tau}^n g^n - \Delta t\hat{A}\hat{\tau}\hat{\mathbf{g}},\end{aligned} \tag{3.30}$$

where we used the fact that $\langle\phi g^n\rangle = \langle\phi\hat{\mathbf{g}}\rangle = 0$, so g^n and $\hat{\mathbf{g}}$ are perpendicular to the space $L_{\mathcal{M}}$ expanded by any $\mathcal{M}$. Next by Lemma 3.6, (3.30) becomes

$$\begin{aligned}O(\Delta t^2) + O(\varepsilon) &= -\Delta t\tilde{\mathbf{a}}(I - \Pi_{\mathcal{M}[f^n]})v\cdot\nabla_x\mathcal{M}[f^n] - \Delta t\hat{\tilde{A}}\hat{\mathbf{e}}(I - \Pi_{\mathcal{M}[f^n]})v\\ &\quad\cdot\nabla_x\mathcal{M}[f^n] - \Delta t\mathbf{a}\boldsymbol{\tau}^n g^n - \Delta t\hat{A}\hat{\tau}\hat{\mathbf{g}}\\ &= -\Delta t\hat{\mathbf{c}}(I - \Pi_{\mathcal{M}[f^n]})v\cdot\nabla_x\mathcal{M}[f^n] - \Delta t\mathbf{a}\boldsymbol{\tau}^n g^n - \Delta t\hat{A}\hat{\tau}\hat{\mathbf{g}},\end{aligned} \tag{3.31}$$

where the second line used the relation (2.4) and the assumption $\mathbf{c} = \tilde{\mathbf{c}}$ (note that $\mathbf{c} = (c_1, \hat{\mathbf{c}})$, $\tilde{\mathbf{c}} = (\tilde{c}_1, \hat{\tilde{\mathbf{c}}})$, thus $\tilde{\mathbf{a}} + \hat{\tilde{A}}\hat{\mathbf{e}} = \hat{\tilde{\mathbf{c}}} = \hat{\mathbf{c}}$). Therefore,

$$\hat{A}\hat{\tau}\hat{\mathbf{g}} = -\hat{\mathbf{c}}(I - \Pi_{\mathcal{M}[f^n]})v\cdot\nabla_x\mathcal{M}[f^n] - \mathbf{a}\boldsymbol{\tau}^n g^n + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right). \tag{3.32}$$

We now assume

$$g^n = -\frac{1}{\tau^n}(I - \Pi_{\mathcal{M}[f^n]})(v\cdot\nabla_x\mathcal{M}[f^n]) + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right). \tag{3.33}$$

Then

$$\hat{A}\hat{\tau}\hat{\mathbf{g}} = -(\hat{\mathbf{c}} - \mathbf{a})(I - \Pi_{\mathcal{M}[f^n]})(v\cdot\nabla_x\mathcal{M}[f^n]) + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right). \tag{3.34}$$

This, written componentwise, is

$$\sum_{j=2}^{i} a_{ij}\tau^{(j)}g^{(j)} = -\sum_{j=2}^{i} a_{ij}(I-\Pi_{\mathcal{M}[f^n]})(v\cdot\nabla_x\mathcal{M}[f^n]) + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right), \quad 2 \le i \le s, \tag{3.35}$$

from which it is easy to show

$$g^{(2)} = -\frac{1}{\tau^{(2)}}(I-\Pi_{\mathcal{M}[f^n]})(v\cdot\nabla_x\mathcal{M}[f^n]) + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right). \tag{3.36}$$

Then using math induction, if one has

$$g^{(j)} = -\frac{1}{\tau^{(j)}}(I-\Pi_{\mathcal{M}[f^n]})(v\cdot\nabla_x\mathcal{M}[f^n]) + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right), \text{ for } j \le i-1, \quad 3 \le i \le s, \tag{3.37}$$

then

$$g^{(i)} = -\frac{1}{\tau^{(i)}}(I-\Pi_{\mathcal{M}[f^n]})(v\cdot\nabla_x\mathcal{M}[f^n]) + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right). \tag{3.38}$$

Since $\tau^{(i)} = \tau^n + O(\Delta t)$,

$$g^{(i)} = -\frac{1}{\tau^{n}}(I-\Pi_{\mathcal{M}[f^n]})(v\cdot\nabla_x\mathcal{M}[f^n]) + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right). \tag{3.39}$$

Hence given the consistent initial data (3.24), we have proved that $f^{(i)} = \mathcal{G}[f^{(i)}] + \varepsilon g^{(i)}$ has the desired form (3.25).

Substituting (3.25) into (2.1), and taking the moments $\langle\cdot\,\phi\rangle$, we have

$$\begin{aligned}\langle\phi f^{(i)}\rangle = \langle\phi f^n\rangle - \Delta t\sum_{j=1}^{i-1}\tilde{a}_{ij}\nabla_x\cdot\left\langle v\phi\left(\mathcal{G}[f^{(j)}] - \frac{\varepsilon}{\tau^n}(I-\Pi_{\mathcal{M}[f^n]})(v\cdot\nabla_x\mathcal{M}[f^n])\right)\right\rangle \\ + O(\varepsilon\Delta t^2) + O(\varepsilon^2).\end{aligned} \tag{3.40}$$

Using Lemma 3.5 and the notations

$$\begin{aligned}U = (\rho, \rho u, E)^T, \quad F(U) = (\rho u, \rho u\otimes u + p\mathrm{Id}, (E+p)u)^T, \\ S(U) = (0, \mu\sigma(u), \mu\sigma(u)u + \frac{d_v+2}{2}\mu(1-\nu)\nabla_x T)^T,\end{aligned} \tag{3.41}$$

the scheme (3.40) can be written as

$$\begin{aligned}U^{(1)} &= U^n, \\ U^{(i)} &= U^n - \Delta t\sum_{j=1}^{i-1}\tilde{a}_{ij}\nabla_x\cdot F(U^{(j)}) + \varepsilon\Delta t\sum_{j=1}^{i-1}\tilde{a}_{ij}\nabla_x\cdot S(U^n) + O(\varepsilon\Delta t^2) \\ &\quad + O(\varepsilon^2), \quad 2 \le i \le s, \\ U^{n+1} &= U^{(s)}.\end{aligned} \tag{3.42}$$

We want to show it is a consistent discretization to the NS equation (same as (1.2) with $\kappa = \frac{d_v+2}{2}\mu(1-\nu)$):

$$\partial_t U + \nabla_x\cdot F(U) = \varepsilon\nabla_x\cdot S(U). \tag{3.43}$$

Note that a standard explicit RK scheme applied to (3.43) should be

$$\begin{aligned} U^{(1)} &= U^n, \\ U^{(i)} &= U^n - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla_x \cdot F(U^{(j)}) + \varepsilon \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} \nabla_x \cdot S(U^{(j)}), \quad 2 \le i \le s, \\ U^{n+1} &= U^{(s)}. \end{aligned} \tag{3.44}$$

For this scheme, the local truncation error is $O(\Delta t^m)$, where m is the order of the method. Assume u is the true solution to (3.43), this means

$$\begin{aligned} O(\Delta t^m) &= \frac{u^{n+1} - u^n}{\Delta t} + \sum_{j=1}^{s-1} \tilde{a}_{sj} \nabla_x \cdot F(u^{(j)}) - \varepsilon \sum_{j=1}^{s-1} \tilde{a}_{sj} \nabla_x \cdot S(u^{(j)}) \\ &= \frac{u^{n+1} - u^n}{\Delta t} + \tilde{a}_{s1} \nabla_x \cdot F(u^n) + \sum_{j=2}^{s-1} \tilde{a}_{sj} \nabla_x \cdot F\left(u^n - \Delta t \sum_{j_1=1}^{j-1} \tilde{a}_{jj_1} \nabla_x \cdot F(u^{(j_1)}) \right) \\ &\quad - \varepsilon \sum_{j=1}^{s-1} \tilde{a}_{sj} \nabla_x \cdot S(u^n) + O(\varepsilon \Delta t), \end{aligned} \tag{3.45}$$

where a Taylor expansion was performed in the last step to extract the $O(\varepsilon \Delta t)$ term.

Now for the scheme (3.42), its local truncation error is

$$\begin{aligned} \mathrm{LTE} &= \frac{u^{n+1} - u^n}{\Delta t} + \sum_{j=1}^{s-1} \tilde{a}_{sj} \nabla_x \cdot F(u^{(j)}) - \varepsilon \sum_{j=1}^{s-1} \tilde{a}_{sj} \nabla_x \cdot S(u^n) + O(\varepsilon \Delta t) + O\left(\frac{\varepsilon^2}{\Delta t} \right) \\ &= \frac{u^{n+1} - u^n}{\Delta t} + \tilde{a}_{s1} \nabla_x \cdot F(u^n) + \sum_{j=2}^{s-1} \tilde{a}_{sj} \nabla_x \cdot F\left(u^n - \Delta t \sum_{j_1=1}^{j-1} \tilde{a}_{jj_1} \nabla_x \cdot F(u^{(j_1)}) \right) \\ &\quad - \varepsilon \sum_{j=1}^{s-1} \tilde{a}_{sj} \nabla_x \cdot S(u^n) + O(\varepsilon \Delta t) + O\left(\frac{\varepsilon^2}{\Delta t} \right) \\ &= O(\Delta t^m) + O(\varepsilon \Delta t) + O\left(\frac{\varepsilon^2}{\Delta t} \right), \end{aligned} \tag{3.46}$$

where a Taylor expansion was performed in the second step and the estimate in (3.45) was used in the last step. □

Remark 3.8 The assumption $\mathbf{c} = \tilde{\mathbf{c}}$ is satisfied by quite a few CK and GSA schemes in the literature, e.g., the ones listed in Sect. 2. In fact, it is the condition often assumed to simplify the order conditions in designing a standard IMEX scheme [2].

Remark 3.9 The assumption (3.23) can be removed if we consider the BGK model since it is easy to show $\mathcal{M}[f^{(i)}] = \mathcal{M}[f^n] + O(\Delta t)$, see Lemma 3.6. For the ES-BGK case, it is not easy to show since Σ, hence $\mathcal{G}$, depends on ε. Yet in our numerical tests, we do observe that $\hat{\mathbf{g}} = O(1)$, to prove which (3.23) is needed.

Remark 3.10 The $O\left(\frac{\varepsilon^2}{\Delta t}\right)$ term in (3.25), and correspondingly in (3.26), can be improved to $O(\varepsilon^2)$ if we assume $g^{(i)} - g^n = O(\Delta t)$. However, this is very hard to prove in general

as also pointed out in [14]. Since our main conclusion remains the same (the scheme can capture the NS limit without resolving ε), we choose not to make this assumption.

Remark 3.11 We are not able to prove a similar result for IMEX schemes of type A following the same argument without imposing extra conditions. In fact, if a type A scheme is not GSA, we expect an error or blow-up of the order at least $O\left(\frac{\Delta t^m}{\varepsilon}\right)$ for g^{n+1} even if $g^n = O(1)$, which is verified in numerical tests (see Example 3). On the other hand, even a type A scheme is GSA, the analysis here still cannot carry over unless we impose extra conditions on the scheme.

Remark 3.12 Two second-order type A schemes were recently proposed in [6] with the aim to capture the NS limit (in fact, they used the relaxation system as a prototype but the analysis is expected to hold also for kinetic equations). We emphasize that the goal in this paper is different: in [6], extra order conditions were derived to match $O(\varepsilon)$ terms, hence new IMEX schemes with more stages need to be constructed; while the aim in this paper is to investigate the existing, widely used IMEX schemes, such as those listed in Sect. 2 (IMEX-II-GSA2(4,4,2) from [6] is included as well but all we need is: it is an IMEX scheme of type CK and GSA and satisfies $\mathbf{c} = \tilde{\mathbf{c}}$).

4 Numerical Results

In this section, we test the IMEX RK schemes on two different models: the $1+1$ BGK model and the $1+3$ ES-BGK model. The fifth order finite difference WENO method [19] is used for approximating spatial derivatives. For velocity domain discretization, we use uniform grid points in a large enough interval or domain.

For the BGK model with $d_x = d_v = 1$, we set $\mu = \rho T$, so $\tau = \frac{\rho T}{\mu} = 1$. Since $\sigma(u) \equiv 0$, there is no viscosity term in the NS equations (1.2), and $\kappa = \frac{d_v+2}{2}\mu = \frac{3}{2}\rho T$.

For the ES-BGK model with $d_x = 1$ and $d_v = 3$, we set $\nu = -\frac{1}{2}$, $\mu = \sqrt{T}$, so $\tau = \frac{\rho T}{(1-\nu)\mu} = \frac{2}{3}\rho\sqrt{T}$ and $\kappa = \frac{d_v+2}{2}\mu(1-\nu) = \frac{15}{4}\sqrt{T}$. Hence the Prandtl number is $\mathrm{Pr} = \frac{d_v+2}{2}\frac{\mu}{\kappa} = \frac{1}{1-\nu} = \frac{2}{3}$.

4.1 Accuracy Tests

Example 1 We test the accuracy of the numerical schemes solving the BGK model for a smooth solution. The consistent initial data is taken as

$$f^0(x, v) = \mathcal{M} - \varepsilon(I - \Pi_{\mathcal{M}})(v\partial_x \mathcal{M}),$$

where

$$\mathcal{M}(x, v) = \frac{\rho(x)}{\sqrt{2\pi T(x)}} \exp\left(-\frac{(v-u(x))^2}{2T(x)}\right),$$

with $\rho(x) = 1 + 0.2\sin(\pi x)$, $T = \frac{1}{\rho(x)}$, $u = 1$.

We use 100 uniform points for velocity in the interval $v \in [-10, 10]$, and Nx uniform points for space in the interval $x \in [0, 2]$ with periodic boundary conditions. The mesh size is $\Delta x = \frac{2}{Nx}$ and we set $\Delta t = 0.1\Delta x$. Since the exact solution is not available, we use the numerical solution on a finer mesh with mesh size $\Delta x/2$ as the reference solution to compute the error for solutions on the mesh size of Δx.

Table 1 shows the result for the third order accurate ARS(4,4,3) scheme at time $= 1$. We can observe that the order of accuracy is no less than or around three when $\Delta t \ll \varepsilon$ or $\Delta t \gg \varepsilon$. On the other hand, we can also see obvious order reduction in the intermediate regime $\Delta t \sim O(\varepsilon)$. In general, the order of accuracy of IMEX schemes in the intermediate regime is highly nontrivial. In a recent work [6], uniform accuracy was observed for second order schemes constructed therein for a linear hyperbolic relaxation system. We also tested the two second order schemes in [6] IMEX-I-GSA2(3,4,2) and IMEX-II-GSA2(4,4,2) for this example. The second order accuracy is indeed achieved for IMEX-I-GSA2(3,4,2) for the same ε and the same meshes as listed in Table 1 even in the intermediate regime $\Delta t \sim O(\varepsilon)$. Nonetheless, it is an open problem to justify why this scheme can maintain uniform accuracy.

Example 2 We test the accuracy of the numerical schemes solving the ES-BGK model for a smooth solution. Denote $v = (v_1, v_2, v_3)^T$, the consistent initial data is taken as

$$f^0(x, v) = \mathcal{M} - \frac{\varepsilon}{\tau}(I - \Pi_{\mathcal{M}})(v_1 \partial_x \mathcal{M}),$$

where

$$\mathcal{M}(x, v) = \frac{\rho(x)}{(\sqrt{2\pi T(x)})^3} \exp\left(-\frac{(v_1 - u(x))^2 + v_2^2 + v_3^2}{2T(x)}\right),$$

with $\rho(x) = 1 + 0.2\sin(\pi x)$, $T = \frac{1}{\rho(x)}$, $u(x) = (1, 0, 0)^T$.

For velocity discretization, we use $80 \times 80 \times 80$ uniform points in the domain $v \in [-10, 10] \times [-10, 10] \times [-10, 10]$. We use Nx uniform points for space in the interval $x \in [0, 2]$ with periodic boundary conditions. The time step is taken as $\Delta t = 0.1\Delta x$. The numerical solution on a finer mesh with mesh size $\Delta x/2$ is used as the reference solution to compute the error for solutions on the mesh of size Δx.

Table 2 shows the result for the ARS(4,4,3) method. Similar to Example 1, we can observe that the order of accuracy is is no less than or around three when $\Delta t \ll \varepsilon$ or $\Delta t \gg \varepsilon$, and we can also see obvious order reduction in the intermediate regime $\Delta t \sim O(\varepsilon)$.

4.2 Accurate Approximations to the Shear Stress and Heat Flux

As discussed in Theorem 3.7, if a numerical scheme solving the ES-BGK equation satisfies the required condition, it can preserve the Navier–Stokes limit using under-resolved time step Δt.

We verify the property (3.25) by comparing the two quantities $\frac{f - \mathcal{G}[f]}{\varepsilon}$ and $-\frac{1}{\tau}(I - \Pi_{\mathcal{M}})(v \cdot \nabla_x \mathcal{M})$ in the numerical solution. In particular, by Lemma 3.5, a desired numerical solution should also satisfy

$$\int_{\mathbb{R}^{d_v}} \frac{f - \mathcal{G}[f]}{\varepsilon}(v - u) \otimes (v - u)\, \mathrm{d}v = -(1 - \nu)\mu\sigma(u) + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right), \quad (4.1)$$

$$\int_{\mathbb{R}^{d_v}} \frac{f - \mathcal{G}[f]}{\varepsilon}\frac{1}{2}(v - u)|v - u|^2\, \mathrm{d}v = -\kappa \nabla_x T + O(\Delta t) + O\left(\frac{\varepsilon}{\Delta t}\right). \quad (4.2)$$

Therefore, we also verify these two moments of $\frac{f - \mathcal{G}[f]}{\varepsilon}$ can produce the correct shear stress $\mu\sigma(u)$ and heat flux $\kappa \nabla_x T$ for smooth solutions.

Example 3 We consider the BGK model with the same consistent initial data as that in Example 1. We use 100 uniform points for $v \in [-10, 10]$, 100 uniform points for $x \in [0, 2]$,

Table 1 Example 1: The third order ARS(4,4,3) scheme. L^∞ error in $f(t, x, v)$ at time $= 1$

	Nx $= 10$	Nx $= 20$	Order	Nx $= 40$	Order	Nx $= 80$	Order	Nx $= 160$	Order	Nx $= 320$	Order	Nx $= 640$	Order
$\varepsilon = 1$	1.42E−2	2.18E−3	2.70	1.57E−4	3.79	6.56E−6	4.58	2.92E−7	4.49	2.97E−8	3.30	3.69E−9	3.01
$\varepsilon = 0.01$	3.37E−3	1.61E−4	3.39	4.43E−6	5.12	2.58E−7	4.10	3.44E−8	2.91	4.99E−9	2.78	6.63E−10	2.91
$\varepsilon = 10^{-4}$	3.89E−3	1.89E−4	4.37	6.05E−6	4.96	1.35E−7	5.48	3.11E−8	*2.12*	1.45E−8	*1.10*	6.37E−9	*1.19*
$\varepsilon = 10^{-6}$	3.90E−3	1.89E−4	4.36	6.21E−6	4.93	1.92E−7	5.02	5.74E−9	5.06	1.82E−10	4.98	1.13E−10	*0.69*
$\varepsilon = 10^{-8}$	3.90E−3	1.89E−4	4.36	6.21E−6	4.93	1.92E−7	5.01	6.06E−9	4.99	2.80E−10	4.44	2.23E−11	3.65

$\Delta t = 0.1\Delta x = 0.1\frac{2}{Nx}$. Order reduction in the intermediate regime $\Delta t \sim O(\varepsilon)$ is marked in italics

Table 2 Example 2: The third order ARS(4,4,3) scheme. L^∞ error in $f(t, x, v)$ at time $= 0.1$

	$N_x = 8$	$N_x = 16$	Order	$N_x = 32$	Order	$N_x = 64$	Order	$N_x = 128$	Order
$\varepsilon = 1$	1.28E−3	1.58E−4	3.02	6.40E−6	4.63	1.13E−7	5.82	1.05E−8	3.43
$\varepsilon = 10^{-2}$	4.27E−4	9.31E−6	5.52	2.89E−7	5.01	8.19E−8	*1.82*	4.13E−8	*0.99*
$\varepsilon = 10^{-4}$	3.64E−4	5.96E−6	5.93	2.08E−7	4.84	2.00E−8	4.81	5.73E−9	*1.80*
$\varepsilon = 10^{-6}$	3.63E−4	5.86E−6	5.95	1.74E−7	5.08	4.68E−9	5.22	1.58E−10	4.89
$\varepsilon = 10^{-8}$	3.63E−4	5.86E−6	5.95	1.73E−7	5.08	4.50E−9	5.27	7.60E−11	5.89

$\Delta t = 0.1\Delta x = 0.1\frac{2}{N_x}$. Order reduction in the intermediate regime $\Delta t \sim O(\varepsilon)$ is marked in italics

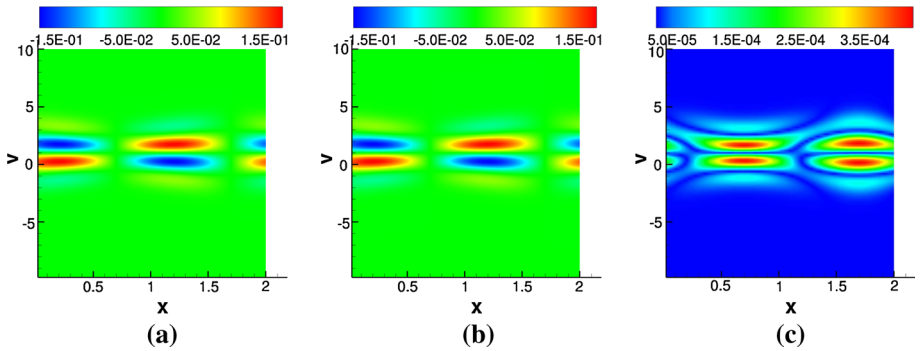

Fig. 1 Example 3: ARS(4,4,3) scheme with a consistent initial condition. The *contour plots* in $x - v$ plane. time $= 0.2$. $\varepsilon = 10^{-8}$. $\Delta x = \frac{2}{100}$. $\Delta t = 0.002$. The maximum pointwise error $\| \frac{f-\mathcal{M}}{\varepsilon} + (I - \Pi_{\mathcal{M}}) v \partial_x \mathcal{M} \|_\infty$ is $4.22E-4$. **a** $\frac{f-\mathcal{M}}{\varepsilon}$. **b** $-(I - \Pi_{\mathcal{M}}) v \partial_x \mathcal{M}$. **c** $\left| \frac{f-\mathcal{M}}{\varepsilon} + (I - \Pi_{\mathcal{M}}) v \partial_x \mathcal{M} \right|$

and $\Delta t = 0.1 \Delta x = 0.002$, and compute the solution up to time $= 0.2$ with $\varepsilon = 10^{-8}$. We test different IMEX RK schemes listed in Sect. 2 which all satisfy the required condition in Theorem 3.7.

Figure 1 shows the contour plots of two quantities $\frac{f-\mathcal{M}}{\varepsilon}$ and $-(I - \Pi_{\mathcal{M}})(v \partial_x \mathcal{M})$ and their difference for the ARS(4,4,3) scheme (recall in the BGK case, $\mathcal{G} = \mathcal{M}$ and $\tau = 1$). In fact, all schemes in Sect. 2 produce similar results and we omit the detail.

For the simple $1+1$ BGK model considered here, the shear stress is zero. Thus we compare the moment $\int_{\mathbb{R}} \frac{f-\mathcal{M}[f]}{\varepsilon} \frac{1}{2}(v-u)|v-u|^2 \, dv$ to the heat flux $-\kappa T_x = -\frac{3}{2} \rho T T_x$. In Fig. 2a, we can see that all type CK and GSA schemes we've tested can capture the correct heat flux.

We have also tested an inconsistent initial condition by choosing $f^0(x, v) = \mathcal{M}$. It seems that except the IMEX-II-GSA(2,3,2) scheme, all other schemes can still produce similar results as those in Figs. 1 and 2a. IMEX-II-GSA(2,3,2), however, is quite sensitive to the initial condition, which suggests the necessity of a consistent initial condition (3.24) in Theorem 3.7.

Figure 2b shows the result of a type A non-GSA scheme IMEX-SSP2(2,2,2) in [23]. It produces huge errors for the heat flux. The blow-up rate for different Δt suggests an error of order $O(\frac{\Delta t^2}{\varepsilon})$. We emphasize that this does not necessarily imply IMEX-SSP2(2,2,2) cannot capture the NS limit. It simply means that IMEX-SSP2(2,2,2) does not satisfy the sufficient conditions derived in this paper. We have also tested several other non-GSA schemes: IMEX-SSP3(3,3,2) and IMEX-SSP3(4,3,3) in [23], which all produce blow-ups in $\frac{f-\mathcal{M}}{\varepsilon}$ as $\varepsilon \to 0$.

Example 4 We consider the ES-BGK model with the same initial data as that in Example 2 except the initial temperature and velocity are set as $T = \frac{2+0.2\cos(\pi x)}{\rho(x)}$, $u(x) = 1 + 0.2\cos(\pi x)$. For velocity discretization, we use $40 \times 40 \times 40$ uniform points in the domain $v \in [-10, 10] \times [-10, 10] \times [-10, 10]$. We use 100 uniform points for space in the interval $x \in [0, 2]$ with periodic boundary conditions. The time step is taken as $\Delta t = 0.1 \Delta x = 0.002$.

Three type CK and GSA schemes are tested: ARS(4,4,3), BPR(3,5,3), and IMEX-II-GSA(2,3,2). The error $\| \frac{f-\mathcal{G}[f]}{\varepsilon} + \frac{1}{\tau}(I - \Pi_{\mathcal{M}})(v_1 \partial_x \mathcal{M}) \|_\infty$ at time $= 0.1$ for these three schemes are 5.76E−5, 4.17E−5 and 1.21E−6 respectively for $\varepsilon = 10^{-8}$. Next, we compare

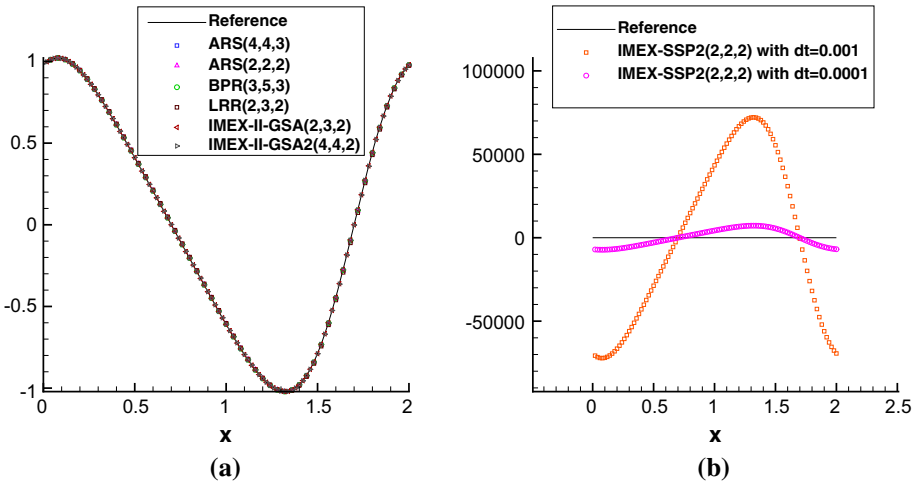

Fig. 2 Example 3: The *symbols* are the moment $\int_{\mathbb{R}} \frac{f-\mathcal{M}[f]}{\varepsilon} \frac{1}{2}(v-u)|v-u|^2 \, dv$ in numerical solutions at time $= 0.2$. The *solid line* reference is the heat flux $-\kappa T_x$. $\varepsilon = 10^{-8}$. $\Delta x = \frac{2}{100}$. **a** Type CK and GSA schemes with a consistent initial condition. $\Delta t = 0.002$. **b** IMEX-SSP2(2,2,2) scheme with a consistent initial condition with two different time steps. The blow up rate suggests an error of order $O\left(\frac{\Delta t^2}{\varepsilon}\right)$

the two moments (4.1) and (4.2) with the shear stress and heat flux computed from the macroscopic quantities. See Fig. 3.

We remark that as in the previous example, with an inconsistent initial condition $f^0(x, v) = \mathcal{M}$, ARS(4,4,3) and BPR(3,5,3) can still produce similar results but IMEX-II-GSA(2,3,2) cannot.

4.3 The Lax Shock Tube Problem

Example 5 We finally test the ARS(4,4,3) scheme solving the ES-BGK model for the Lax shock tube problem at time $= 1.3$ with the initial states

$$\begin{pmatrix} \rho \\ u \\ p \end{pmatrix} = \begin{cases} (0.445, 0.698, 3.528)^T, & -5 \le x \le 0, \\ (0.5, 0, 0.571)^T, & 0 < x \le 5. \end{cases}$$

The initial condition is taken as $f^0(x, v) = \mathcal{M} - \frac{\varepsilon}{\tau}(I - \Pi_{\mathcal{M}})(v_1 \partial_x \mathcal{M})$.

We first use $80 \times 80 \times 80$ uniform points for the velocity domain $[-20, 20] \times [-20, 20] \times [-20, 20]$, and $Nx = 200$ uniform points for the spatial discretization, i.e., $\Delta x = \frac{10}{200}$. The CFL number is taken as 0.2, i.e., $\Delta t = 0.2 \frac{\Delta x}{\|v_1\|_\infty}$ where $\|v_1\|_\infty = 20$. In Fig. 4, we compare the numerical solution of the ES-BGK model with those of the compressible Navier–Stokes equations and the compressible Euler equations. The reference solution of the Euler equations was generated by the exact Riemann solution [25]. The reference solution of the Navier–Stokes equations was generated by using the fifth order WENO method for the convection and fourth order finite difference for the diffusion on a grid of 20,000, 40,000 and 100,000 points for $\varepsilon = 10^{-2}, 10^{-3}$ and 10^{-4} respectively. See the appendix in [28] for a similar scheme. We can see that the numerical solution of the ES-BGK model is very close to that of the NS equations. For better visualization, we again compare the two

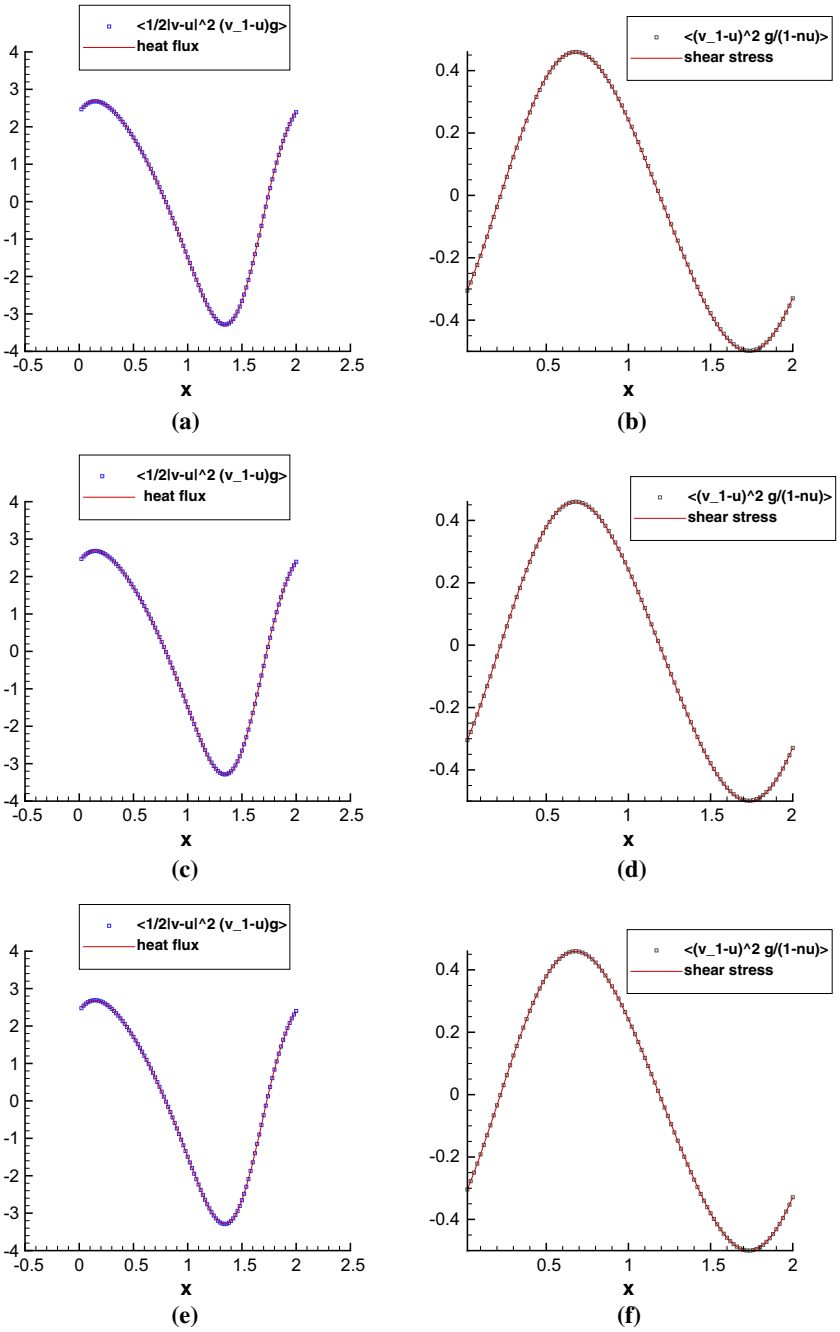

Fig. 3 Example 4: time $= 0.2$. $\varepsilon = 10^{-8}$. $\Delta x = \frac{2}{100}$. $\Delta t = 0.002$. Here $g = \frac{f - \mathcal{G}[f]}{\varepsilon}$. The *symbols* are the moments $\int_{\mathbb{R}^3} g \frac{1}{2}(v_1 - u)|v - u|^2 \, \mathrm{d}v$ (*left*) and $\frac{1}{1-\nu} \int_{\mathbb{R}^3} g(v_1 - u)^2 \, \mathrm{d}v$ (*right*) in numerical solutions. The *solid line* reference is the heat flux $-\kappa T_x$ (*left*) and the shear stress is $-\mu\sigma(u)$ (*right*). **a** ARS(4,4,3). **b** ARS(4,4,3). **c** BPR(3,5,3). **d** BPR(3,5,3). **e** IMEX-II-GSA(2, 3, 2). **f** IMEX-II-GSA(2,3,2)

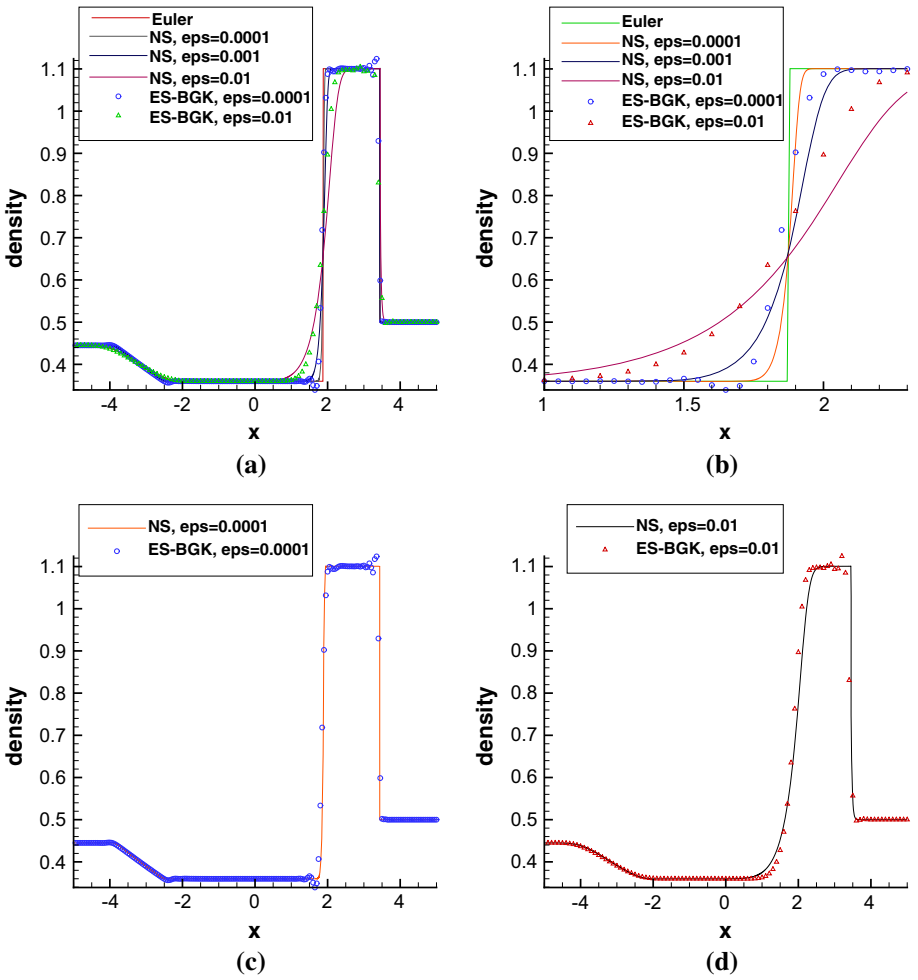

Fig. 4 Example 5: Comparison between reference solutions of Navier–Stokes equations and the numerical solutions of ARS(4,4,3) for the ES-BGK model using $\Delta x = \frac{10}{200}$ and $\Delta t = 0.2\frac{\Delta x}{\|v_1\|_\infty} = 0.2\frac{\Delta x}{20}$ at time $= 1.3$. Uniform $80 \times 80 \times 80$ points for the velocity domain $[-20, 20] \times [-20, 20] \times [-20, 20]$

moments (4.1) and (4.2) with the shear stress and heat flux computed from the macroscopic quantities. See Fig. 5 for the case of $\varepsilon = 10^{-4}$. Figure 6 shows the case of $\varepsilon = 10^{-8}$, where $\Delta x = \frac{10}{80}$, $\Delta t = 0.2\frac{\Delta x}{\|v_1\|_\infty} = 0.2\frac{\Delta x}{20}$ and uniform $40 \times 40 \times 40$ points for the velocity domain $[-20, 20] \times [-20, 20] \times [-20, 20]$ are used.

Remark 4.1 For discontinuous problems, the thickness of a shock layer is about $O(\varepsilon)$ so one has to use resolved mesh size Δx in the numerical scheme. This, as a result, requires Δt to be chosen at least of order $O(\varepsilon)$ due to the CFL condition. Our focus here is to show the semi-discrete high order IMEX scheme does not need the time step to resolve $O(\varepsilon)$ in order to capture the NS limit for smooth solutions. Hence we do not attempt to address the issue of spatial discretization.

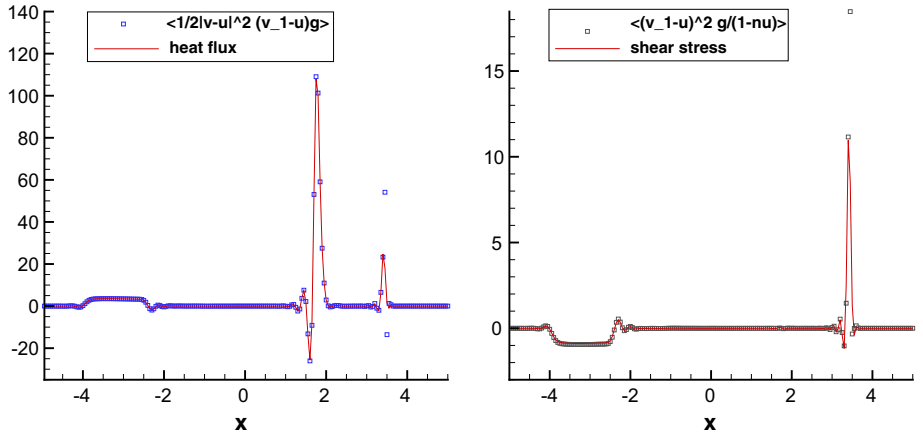

Fig. 5 Example 5: The numerical solutions of ARS(4,4,3) at time $= 1.3$ for $\varepsilon = 10^{-4}$. $\Delta t = 0.2\frac{\Delta x}{\|v_1\|_\infty} = 0.2\frac{\Delta x}{20}$ and $\Delta x = \frac{10}{200}$. Here $g = \frac{f-\mathcal{G}[f]}{\varepsilon}$. The heat flux is $-\kappa T_x$ and the shear stress is $-\mu\sigma(u)$. Uniform $80 \times 80 \times 80$ points for the velocity domain $[-20, 20] \times [-20, 20] \times [-20, 20]$

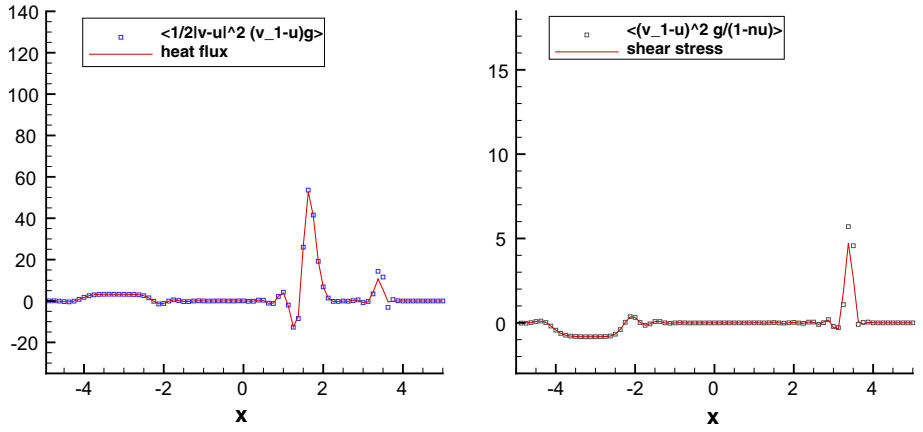

Fig. 6 Example 5: The numerical solutions of ARS(4,4,3) at time $= 1.3$ for $\varepsilon = 10^{-8}$. $\Delta t = 0.2\frac{\Delta x}{\|v_1\|_\infty} = 0.2\frac{\Delta x}{20}$ and $\Delta x = \frac{10}{80}$. Here $g = \frac{f-\mathcal{G}[f]}{\varepsilon}$. The heat flux is $-\kappa T_x$ and the shear stress is $-\mu\sigma(u)$. Uniform $40 \times 40 \times 40$ points for the velocity domain $[-20, 20] \times [-20, 20] \times [-20, 20]$

5 Conclusion

IMEX RK schemes are popular methods to solve the stiff kinetic equations. Their asymptotic behavior with respect to the leading Euler limit has been studied extensively in the literature. In this work, we investigate their behavior at the Navier–Stokes level and prove that for a class of existing IMEX schemes (type CK and GSA), under consistent initial condition, they can capture the NS limit without resolving ε. That is, for $\varepsilon = o(\Delta t)$, we only need $\Delta t^m = o(\varepsilon)$, where m is the order of the explicit RK scheme in an IMEX method. For simplicity, we only considered the BGK/ES-BGK models, for which the implicit collision operators can be solved easily without iteration. In the future, we will study the application of IMEX schemes for the full Boltzmann equation.

References

1. Andries, P., Le Tallec, P., Perlat, J.-P., Perthame, B.: The Gaussian-BGK model of Boltzmann equation with small Prandtl number. Eur. J. Mech. B Fluids **19**, 813–830 (2000)
2. Ascher, U., Ruuth, S., Spiteri, R.: Implicit-explicit Runge–Kutta methods for time-dependent partial differential equations. Appl. Numer. Math. **25**, 151–167 (1997)
3. Bardos, C., Golse, F., Levermore, D.: Fluid dynamic limits of kinetic equations. I. Formal derivations. J. Stat. Phys. **63**, 323–344 (1991)
4. Bennoune, M., Lemou, M., Mieussens, L.: Uniformly stable numerical schemes for the Boltzmann equation preserving the compressible Navier–Stokes asymptotics. J. Comput. Phys. **227**, 3781–3803 (2008)
5. Bhatnagar, P.L., Gross, E.P., Krook, M.: A model for collision processes in gases. I. Small amplitude processes in charged and neutral one-component systems. Phys. Rev. **94**, 511–525 (1954)
6. Boscarino, S., Pareschi, L.: On the asymptotic properties of IMEX Runge–Kutta schemes for hyperbolic balance laws. J. Comput. Appl. Math. **316**, 60–73 (2017)
7. Boscarino, S., Pareschi, L., Russo, G.: Implicit–explicit Runge–Kutta schemes for hyperbolic systems and kinetic equations in the diffusion limit. SIAM J. Sci. Comput. **35**, A22–A51 (2013)
8. Caflisch, R.E., Jin, S., Russo, G.: Uniformly accurate schemes for hyperbolic systems with relaxation. SIAM J. Numer. Anal. **34**, 246–281 (1997)
9. Cercignani, C.: The Boltzmann Equation and Its Applications. Springer, New York (1988)
10. Chapman, S., Cowling, T.G.: The Mathematical Theory of Non-uniform Gases, 3rd edn. Cambridge University Press, Cambridge (1991)
11. Coron, F., Perthame, B.: Numerical passage from kinetic to fluid equations. SIAM J. Numer. Anal. **28**, 26–42 (1991)
12. Dimarco, G., Pareschi, L.: Asymptotic preserving implicit–explicit Runge–Kutta methods for nonlinear kinetic equations. SIAM J. Numer. Anal. **51**, 1064–1087 (2013)
13. Dimarco, G., Pareschi, L.: Implicit–explicit linear multistep methods for stiff kinetic equations. SIAM J. Numer. Anal. **55**, 664–690 (2017)
14. Filbet, F., Jin, S.: A class of asymptotic-preserving schemes for kinetic equations and related problems with stiff sources. J. Comput. Phys. **229**, 7625–7648 (2010)
15. Filbet, F., Jin, S.: An asymptotic preserving scheme for the ES-BGK model of the Boltzmann equation. J. Sci. Comput. **46**, 204–224 (2011)
16. Hairer, E., Wanner, G.: Solving Ordinary Differential Equations. II: Stiff and Differential-Algebraic Problems. Springer, New York (1987)
17. Holway, L.: Kinetic theory of shock structure using an ellipsoidal distribution function. In: Proceedings of the 4th International Symposium on Rarefied Gas Dynamics, vol. I, pp. 193–215. Academic Press, New York (1966)
18. Hu, J., Jin, S., Li, Q.: Asymptotic-preserving schemes for multiscale hyperbolic and kinetic equations, chapter 5. In: Abgrall, R., Shu, C.-W. (eds.) Handbook of Numerical Methods for Hyperbolic Problems, pp. 103–129. North-Holland, Amsterdam (2017)
19. Jiang, G.-S., Shu, C.-W.: Efficient implementation of weighted ENO schemes. J. Comput. Phys. **126**(1), 202–228 (1996)
20. Jin, S.: Runge–Kutta methods for hyperbolic conservation laws with stiff relaxation terms. J. Comput. Phys. **122**, 51–67 (1995)
21. Jin, S.: Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. Riv. Mat. Univ. Parma **3**, 177–216 (2012)
22. Liotta, S.F., Romano, V., Russo, G.: Central schemes for balance laws of relaxation type. SIAM J. Numer. Anal. **38**, 1337–1356 (2000)
23. Pareschi, L., Russo, G.: Implicit–explicit Runge–Kutta methods and applications to hyperbolic systems with relaxation. J. Sci. Comput. **25**, 129–155 (2005)
24. Pieraccini, S., Puppo, G.: Implicit–explicit schemes for BGK kinetic equations. J. Sci. Comput. **1**, 1–28 (2007)
25. Toro, E.F.: Riemann Solvers and Numerical Methods for Fluid Dynamics: A Practical Introduction. Springer, Berlin (2013)
26. Villani, C.: A review of mathematical topics in collisional kinetic theory. In: Friedlander, S., Serre, D. (eds.) Handbook of Mathematical Fluid Mechanics, vol. I, pp. 71–305. North-Holland, Amsterdam (2002)
27. Xiong, T., Jang, J., Li, F., Qiu, J.-M.: High order asymptotic preserving nodal discontinuous Galerkin IMEX schemes for the BGK equation. J. Comput. Phys. **284**, 70–94 (2015)
28. Zhang, X.: On positivity-preserving high order discontinuous Galerkin schemes for compressible Navier–Stokes equations. J. Comput. Phys. **328**, 301–343 (2017)