SIAM/ASA J. UNCERTAINTY QUANTIFICATION
Vol. 9, No. 2, pp. 650–680

Uncertainty Quantification for the BGK Model of the Boltzmann Equation Using Multilevel Variance Reduced Monte Carlo Methods*

Jingwei Hu†, Lorenzo Pareschi‡, and Yubo Wang†

Abstract. We propose a control variate multilevel Monte Carlo method for the kinetic Bhatnagar–Gross–Krook model of the Boltzmann equation subject to random inputs. The method combines a multilevel Monte Carlo technique with the computation of the optimal control variate multipliers derived from local or global variance minimization problems. Consistency and convergence analysis for the method equipped with a second-order positivity-preserving and asymptotic-preserving scheme in space and time is also performed. Various numerical examples confirm that the optimized multilevel Monte Carlo method outperforms the classical multilevel Monte Carlo method especially for problems with discontinuities.

Key words. kinetic equation, BGK model, multilevel Monte Carlo method, control variate method, uncertainty quantification, random inputs

AMS subject classifications. 35R60, 35Q20, 65C05

DOI. 10.1137/20M1331846

1. Introduction.

Kinetic theory, from a statistical physics viewpoint [5], represents an essential tool to model the nonequilibrium dynamics in a variety of fields including rarefied gases, semiconductors, plasmas, and even large particle systems in biological and social sciences [4, 27, 34]. The most fundamental kinetic equation, the Boltzmann equation, describes the statistical behavior of a thermodynamic system by taking into account particle transport and binary collisions [3]:

$$\partial_t f + \mathbf{v}\cdot\nabla_{\mathbf{x}} f = \frac{1}{\varepsilon}\mathcal{Q}(f,f), \quad \mathbf{x}\in D\subset\mathbb{R}^3,\ \mathbf{v}\in\mathbb{R}^3,\ t>0, \tag{1.1}$$

where $f = f(\mathbf{x},\mathbf{v},t)$ is the phase space distribution function of position $\mathbf{x}$, velocity $\mathbf{v}$, and time t. The collision term $\mathcal{Q}(f,f)$ is a five-fold, quadratic integral operator, and ε is the Knudsen number, defined as the ratio of the mean free path and the typical length scale. In most applications, ε varies from $O(1)$, the kinetic regime, to $\varepsilon \ll 1$, the fluid regime. Although the Boltzmann equation is widely applicable, the complexity of the collision operator

*Received by the editors April 15, 2020; accepted for publication (in revised form) March 2, 2021; published electronically May 13, 2021.
https://doi.org/10.1137/20M1331846
Funding: The research of the first author was supported in part by NSF grant DMS-1620250 and NSF CAREER grant DMS-1654152. The research of the second author was supported by PRIN Project 2017, 2017KKJP4X, "Innovative numerical methods for evolutionary partial differential equations and applications" with the Italian Ministry of Instruction, University and Research (MIUR).

†Department of Mathematics, Purdue University, West Lafayette, IN 47907 USA (jingweihu@purdue.edu, wang3158@purdue.edu).
‡Department of Mathematics and Computer Science, University of Ferrara, Via Machiavelli 30, 44121-Ferrara, Italy (lorenzo.pareschi@unife.it).

$\mathcal{Q}(f,f)$ makes both analysis and computation of the equation extremely challenging. Hence many simplified collisional models have been introduced to mimic the properties of the full Boltzmann operator. Among these, the Bhatnagar–Gross–Krook (BGK) model [1], which assumes a simple relaxation to equilibrium, has been widely used. The model reads as follows:

$$\partial_t f + \mathbf{v}\cdot\nabla_{\mathbf{x}} f = \frac{1}{\varepsilon}(M[f]-f), \quad \mathbf{x}\in D\subset\mathbb{R}^3,\ \mathbf{v}\in\mathbb{R}^3,\ t>0, \tag{1.2}$$

where $M[f]$ is the so-called Maxwellian equilibrium function given by

$$M[f](\mathbf{x},\mathbf{v},t) = \frac{\rho(\mathbf{x},t)}{(2\pi T(\mathbf{x},t))^{\frac{3}{2}}}\exp\left(-\frac{|\mathbf{v}-\boldsymbol{U}(\mathbf{x},t)|^2}{2T(\mathbf{x},t)}\right), \tag{1.3}$$

where $\rho(\mathbf{x},t)$, $\boldsymbol{U}(\mathbf{x},t)$, $T(\mathbf{x},t)$ are the density, bulk velocity, and temperature defined through the moments of f:

$$\begin{aligned}&\rho(\mathbf{x},t) = \int_{\mathbb{R}^3} f(\mathbf{x},\mathbf{v},t)\ \mathrm{d}\mathbf{v}, \qquad \boldsymbol{U}(\mathbf{x},t) = \frac{1}{\rho(\mathbf{x},t)}\int_{\mathbb{R}^3}\mathbf{v}f(\mathbf{x},\mathbf{v},t)\ \mathrm{d}\mathbf{v},\\ &T(\mathbf{x},t) = \frac{1}{3\rho(\mathbf{x},t)}\int_{\mathbb{R}^3}|\mathbf{v}-\boldsymbol{U}(\mathbf{x},t)|^2 f(\mathbf{x},\mathbf{v},t)\ \mathrm{d}\mathbf{v}.\end{aligned} \tag{1.4}$$

Let $\Phi(\mathbf{v}) = [1,\mathbf{v},\frac{1}{2}|\mathbf{v}|^2]^T$; then one has the following conservation property:

$$\int_{\mathbb{R}^3} M[f](\mathbf{x},\mathbf{v},t)\Phi(\mathbf{v})\mathrm{d}\mathbf{v} = \int_{\mathbb{R}^3} f(\mathbf{x},\mathbf{v},t)\Phi(\mathbf{v})\mathrm{d}\mathbf{v} = \begin{bmatrix}\rho\\ \rho\boldsymbol{U}\\ \frac{3}{2}\rho T+\frac{1}{2}\rho|\boldsymbol{U}|^2\end{bmatrix} =: \begin{bmatrix}\rho\\ \mathbf{m}\\ E\end{bmatrix}, \tag{1.5}$$

where $\mathbf{m}$ is the momentum and E is the total energy. Using (1.5), if we multiply (1.2) by $\Phi(\mathbf{v})$ and integrate over $\mathbf{v}$, we obtain the following local conservation law:

$$\begin{cases}\partial_t\displaystyle\int_{\mathbb{R}^3} f\mathrm{d}\mathbf{v} + \nabla_{\mathbf{x}}\cdot\int_{\mathbb{R}^3}\mathbf{v}f\mathrm{d}\mathbf{v} = 0,\\ \partial_t\displaystyle\int_{\mathbb{R}^3}\mathbf{v}f\mathrm{d}\mathbf{v} + \nabla_{\mathbf{x}}\cdot\int_{\mathbb{R}^3}\mathbf{v}\otimes\mathbf{v}f\mathrm{d}\mathbf{v} = 0,\\ \partial_t\displaystyle\int_{\mathbb{R}^3}\frac{1}{2}|\mathbf{v}|^2 f\mathrm{d}\mathbf{v} + \nabla_{\mathbf{x}}\cdot\int_{\mathbb{R}^3}\frac{1}{2}\mathbf{v}|\mathbf{v}|^2 f\mathrm{d}\mathbf{v} = 0.\end{cases} \tag{1.6}$$

When $\varepsilon\to 0$, formally we have $f\to M[f]$ from (1.2). Replacing f by $M[f]$ in the above local conservation law yields the compressible Euler equations:

$$\begin{cases}\partial_t\rho + \nabla_{\mathbf{x}}\cdot(\rho\boldsymbol{U}) = 0,\\ \partial_t(\rho\boldsymbol{U}) + \nabla_{\mathbf{x}}\cdot(\rho\boldsymbol{U}\otimes\boldsymbol{U}+\rho TI) = 0,\\ \partial_t E + \nabla_{\mathbf{x}}\cdot((E+\rho T)\boldsymbol{U}) = 0.\end{cases} \tag{1.7}$$

In the last decades, research activities in kinetic theory have focused mainly on deterministic kinetic equations, both theoretically and numerically [4, 9, 34], ignoring the presence of uncertain/random inputs. In reality, uncertainties may arise in initial/boundary conditions and

other parameters, like the details of the microscopic interaction, because of incomplete knowledge or imprecise measurement. Recently, there has been a significant interest in studying the impact of these random inputs in kinetic equations; see [17, 26] and the whole collection [19] for an overview. In order to quantify the above uncertainties, the construction of numerical methods for kinetic equations has been mostly oriented on stochastic Galerkin approximation based on generalized polynomial chaos expansion (gPC-sG), already successfully applied to many physical and engineering problems [13, 35]. We mention that recently gPC-sG methods have been successfully applied also to direct simulation Monte Carlo methods for the Boltzmann equation [28]. Despite the fact that gPC-sG methods have been able to show spectral accuracy for smooth solutions, they suffer the drawback of the curse of dimensionality and their highly intrusive nature. On one hand, existing codes for simulating the deterministic kinetic problems need to be completely reconfigured to implement the gPC-sG method. On the other hand, intrusiveness can induce some nonphysical approximations even when the deterministic numerical solvers possess the correct physical properties. For example, due to the gPC expansion, the methods may induce approximations with nonpositive density; furthermore, close to fluid regimes, it is well-known that the gPC-sG system may lose hyperbolicity and lead to spurious solutions [7].

Another class of methods for uncertainty quantification is based on statistical Monte Carlo (MC) sampling, where the random space is sampled and the underlying deterministic PDE is solved for each sample. The nonintrusiveness of the method enables the approximated solutions to inherit properties, like positivity preservation, of the existing deterministic solvers and makes parallel computing feasible for implementation. However, the asymptotic convergence rate is nonimprovable by the central limit theorem, and accelerated algorithms are obtained through variance reduction techniques [2]. In this context, multifidelity methods for kinetic equations have been recently introduced in [10, 11, 22]; see also the recent survey [29] for an introduction to the topic. These methods are capable of providing a significant speedup of the convergence properties of the MC solver using as control variates simplified surrogate models that are cheaper to solve than the full model. A related line of research is based on the use of multilevel Monte Carlo (MLMC) methods (see [14, 16] and [24, 25] for these methods applied to hyperbolic conservation laws), where the approximation of statistical expectation breaks up into telescopic sums of expectations of consecutive mesh sizes. These methods are closely related to multifidelity methods, since they essentially use in a recursive way the solution of the full model with various coarser meshes as surrogate models.

In this manuscript, following the above analogy, we develop MLMC methods in a control variate setting for the multiscale kinetic equations. Therefore, in our MLMC method each level in the telescopic sum depends on an additional parameter which is computed in order to minimize the variance of the solver. We will perform this strategy both locally between two different levels and globally among all levels. As a prototype kinetic equation to design our methodology we consider the BGK model (1.2) of the Boltzmann equation subject to random inputs. Following the well-posedness results in [30, 31], we provide a direct analogue of the former to the BGK equation with random parameters. Due to the nonintrusiveness of MC-type methods, approximations of the statistical moments can preserve properties from the deterministic solvers. We adopt the implicit-explicit Runge–Kutta (IMEX-RK) scheme from [18] to construct a second-order positivity-preserving (the distribution f is positive for all

ε) and asymptotic-preserving (the scheme becomes a solver for the limiting Euler system (1.7) when ε goes to zero) scheme for time and spatial discretizations. Various numerical examples confirm the good performance of MLMC methods compared to standard MC methods and demonstrate that the control variate MLMC method outperforms the classical MLMC method especially for problems with discontinuities.

The rest of this paper is organized as follows. In the next section, we introduce the BGK equation with random inputs and establish the well-posedness of the equation. The MC methods and analysis are presented in section 3, whereas in section 4 we discuss their multilevel extension in a standard and control variate setting. In section 5 we show the numerical results obtained with standard MC, MLMC, and control variate MLMC methods. Finally some conclusions are drawn in section 6. In a separate Appendix A we report the details of the dimension reduction method and the numerical scheme adopted to solve the deterministic BGK equation.

2. The BGK equation with random inputs.

In this section we formulate systematically the BGK equation with random inputs and establish the well-posedness of the equation by extending the results in [30, 31].

2.1. Setup of the problem.

In the BGK equation, due to uncertain initial or boundary conditions, the resulting solution f would be a random variable taking values in the functional space in which the solution of the BGK equation (1.2) lies. In most circumstances, it is the physical observables or macroscopic quantities (such as ρ, $\boldsymbol{U}$, T) at certain time that are of interest; hence we will mainly consider random variables taking values in $L^1(D)$, where D is the physical domain. Following the discussion in [24], we first present some basic concepts from probability theory and functional analysis.

Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space with Ω being the set of elementary events, $\mathscr{F}$ the corresponding σ-algebra, and $\mathbb{P}$ the probability measure mapping Ω into $[0,1]$ such that $\mathbb{P}(\Omega) = 1$. A random variable taking values in $L^1(D)$, a separable Banach space, is defined to be a mapping $X\colon\ \Omega \to L^1(D)$ such that for any $A \in \mathscr{G}$, the preimage $X^{-1}(A) \in \mathscr{F}$, where $X^{-1}(A) = \{w \in \Omega : X(w) \in A\}$ and $(L^1(D), \mathscr{G})$ is a measurable space.

To define the expectation and variance of random variables in $L^1(D)$, we need the concept of Bochner integral by extending the Lebesgue integral theory. The strong measurable mapping $X : \Omega \to L^1(D)$ is *Bochner integrable* if, for any probability measure $\mathbb{P}$ on the measurable space $(\Omega, \mathscr{F})$,

$$\int_\Omega \|X(w)\|_{L^1(D)}\, \mathrm{d}\mathbb{P}(w) < \infty. \tag{2.1}$$

Moreover, any Bochner integrable random variable $X : \Omega \to L^1(D)$ can be approximated by a sequence of simple random variables $\{X_n\}_{n\in\mathbb{N}}$ defined as follows:

$$X_n = \sum_{i=1}^{N} x_{n,i}\chi_{A_{n,i}}, \quad A_{n,i} \in \mathscr{F},\ x_{n,i} \in L^1(D),\ N < \infty. \tag{2.2}$$

To get moments like expectation or central moments like variance, similar to the derivation of the Lebesgue integral, the Bochner integral is defined by taking the limit of sequences of simple random variables $\{X_n(w)\}$; for example, the kth order moments are defined as

$$\mathbb{E}[X^k] := \int_\Omega X^k(w)\ \mathrm{d}\mathbb{P}(w) = \lim_{n\to\infty} \int_\Omega X_n^k(w)\ \mathrm{d}\mathbb{P}(w), \tag{2.3}$$

and the variance is defined as

$$\mathbb{V}[X] := \mathbb{E}[(X - \mathbb{E}[X])^2] = \int_\Omega (X(w) - \mathbb{E}[X])^2\ \mathrm{d}\mathbb{P}(w) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2. \tag{2.4}$$

For the error analysis, we need to introduce the Banach space $L^p(\Omega, \mathscr{F}, \mathbb{P}; L^1(D))$ with the norm

$$\|X\|_{L^p(\Omega;L^1(D))} := (\mathbb{E}[\|X\|^p_{L^1(D)}])^{\frac{1}{p}} < \infty,\ 1 \le p < \infty, \tag{2.5}$$

and $L^\infty(\Omega, \mathscr{F}, \mathbb{P}; L^1(D))$ with the norm

$$\|X\|_{L^\infty(\Omega;L^1(D))} := \operatorname{ess\,sup}_{w\in\Omega} \|X\|_{L^1(D)}. \tag{2.6}$$

The BGK equation with random inputs hence reads

$$\begin{aligned} \partial_t f(w;\mathbf{x},\mathbf{v},t) + \mathbf{v}\cdot\nabla_{\mathbf{x}} f(w;\mathbf{x},\mathbf{v},t) &= \frac{1}{\varepsilon}(M[f](w;\mathbf{x},\mathbf{v},t) - f(w;\mathbf{x},\mathbf{v},t)), \\ w\in\Omega,\ \mathbf{x}\in D\subset\mathbb{R}^3,\ \mathbf{v}\in\mathbb{R}^3,\ t>0, \end{aligned} \tag{2.7}$$

where

$$M[f](w;\mathbf{x},\mathbf{v},t) = \frac{\rho(w;\mathbf{x},t)}{(2\pi T(w;\mathbf{x},t))^{\frac{3}{2}}} \exp\left(-\frac{|\mathbf{v} - \boldsymbol{U}(w;\mathbf{x},t)|^2}{2T(w;\mathbf{x},t)}\right) \tag{2.8}$$

with

$$\begin{aligned} &\rho(w;\mathbf{x},t) = \int_{\mathbb{R}^3} f(w;\mathbf{x},\mathbf{v},t)\ \mathrm{d}\mathbf{v}, \qquad \boldsymbol{U}(w;\mathbf{x},t) = \frac{1}{\rho(w;\mathbf{x},t)} \int_{\mathbb{R}^3} \mathbf{v} f(w;\mathbf{x},\mathbf{v},t)\ \mathrm{d}\mathbf{v}, \\ &T(w;\mathbf{x},t) = \frac{1}{3\rho(w;\mathbf{x},t)} \int_{\mathbb{R}^3} |\mathbf{v} - \boldsymbol{U}(w;\mathbf{x},t)|^2 f(w;\mathbf{x},\mathbf{v},t)\ \mathrm{d}\mathbf{v}. \end{aligned} \tag{2.9}$$

The initial condition is given as

$$f(w;\mathbf{x},\mathbf{v},0) = f_0(w;\mathbf{x},\mathbf{v}), \quad w\in\Omega,\ \mathbf{x}\in D\subset\mathbb{R}^3,\ \mathbf{v}\in\mathbb{R}^3. \tag{2.10}$$

For the boundary condition, we consider one of the following:

- periodic boundary: $f(w;\mathbf{x}+\mathbf{a},\mathbf{v},t) = f(w;\mathbf{x},\mathbf{v},t)$ for $\mathbf{x}\in\partial D$ and some $\mathbf{a}\in\mathbb{R}^3$;
- Dirichlet boundary: $f(w;\mathbf{x},\mathbf{v},t) = g(w;\mathbf{x},\mathbf{v},t)$ for $\mathbf{x}\in\partial D$;
- purely diffusive Maxwell boundary: for $\mathbf{x}\in\partial D$,

$$f(w;\mathbf{x},\mathbf{v},t) = M_w(w;\mathbf{x},\mathbf{v},t),\ \mathbf{v}\cdot\mathbf{n} < 0, \tag{2.11}$$

where $\mathbf{n}$ is the outward normal of ∂D and M_w is given by

$$M_w(w;\mathbf{x},\mathbf{v},t) = \frac{\rho_w(w;\mathbf{x},t)}{(2\pi T_w(w;\mathbf{x},t))^{\frac{3}{2}}} \exp\left(-\frac{|\mathbf{v}|^2}{2T_w(w;\mathbf{x},t)}\right), \tag{2.12}$$

where $T_w(w;\mathbf{x},t)$ is the wall temperature and $\rho_w(w;\mathbf{x},t)$ is chosen such that

$$\int_{\mathbf{v}\cdot\mathbf{n}>0} \mathbf{v}\cdot\mathbf{n}\, f(w;\mathbf{x},\mathbf{v},t)\ \mathrm{d}\mathbf{v} = -\int_{\mathbf{v}\cdot\mathbf{n}<0} \mathbf{v}\cdot\mathbf{n}\, M_w(w;\mathbf{x},\mathbf{v},t)\ \mathrm{d}\mathbf{v}. \tag{2.13}$$

2.2. Well-posedness of the equation and some estimates of the macroscopic quantities. In the following, we establish the well-posedness of the BGK equation (2.7) with random inputs. We also obtain some estimates for the macroscopic quantities ρ, $\boldsymbol{U}$, and T. For simplicity, we assume the periodic boundary condition and consider the uncertainty only arising in the initial condition f_0.

First of all, some general estimates on the macroscopic quantities can be obtained pointwise in w following [31] for the deterministic BGK equation.

Proposition 2.1 ([31]). *Suppose that $f(w;\mathbf{x},\mathbf{v},t) \geq 0$. Define $\rho(w;\mathbf{x},t)$, $\boldsymbol{U}(w;\mathbf{x},t)$, $T(w;\mathbf{x},t)$ according to (2.9). Moreover, set*

$$N_q(f)(w;t) := \sup_{\mathbf{x}} \sup_{\mathbf{v}} f(w;\mathbf{x},\mathbf{v},t)|\mathbf{v}|^q, \quad q \geq 0. \tag{2.14}$$

Then the following estimates hold:

$$\frac{\rho(w;\mathbf{x},t)}{T(w;\mathbf{x},t)^{\frac{3}{2}}} \leq C_0 N_0(f), \tag{2.15}$$

$$\rho(w;\mathbf{x},t)(3T(w;\mathbf{x},t) + |\boldsymbol{U}(w;\mathbf{x},t)|^2)^{\frac{q-3}{2}} \leq C_q N_q(f) \quad \text{for } q > 5, \tag{2.16}$$

where C_0, C_q are some positive constants.

Based on the above estimates, one can obtain the existence and uniqueness of the solution to (2.7) also following [31] in a pointwise manner in w.

Theorem 2.2 ([31]). *Set*

$$\mathbb{N}_q(f)(w;t) := \sup_{\mathbf{x}} \sup_{\mathbf{v}} f(w;\mathbf{x},\mathbf{v},t)(1+|\mathbf{v}|^q); \tag{2.17}$$

then, by definition, $N_q(f) \leq \mathbb{N}_q(f)$. Suppose that the initial condition $f_0(w;\mathbf{x},\mathbf{v}) \geq 0$ and that for some $q > 5$,

$$\begin{aligned} &\mathbb{N}_q(f_0)(w) = \sup_{\mathbf{x}} \sup_{\mathbf{v}} f_0(w;\mathbf{x},\mathbf{v})(1+|\mathbf{v}|^q), \\ &\sup_w \mathbb{N}_q(f_0)(w) \leq A_0 < \infty, \end{aligned} \tag{2.18}$$

and

$$\begin{aligned} &\gamma(w;\mathbf{x},t) := \int_{\mathbb{R}^3} f_0(w;\mathbf{x}-\mathbf{v}t,\mathbf{v})\, \mathrm{d}\mathbf{v}, \\ &\inf_w \inf_{\mathbf{x}} \inf_t \gamma(w;\mathbf{x},t) \geq A_1 > 0; \end{aligned} \tag{2.19}$$

then, for fixed Knudsen number $\varepsilon > 0$, there exists a unique mild solution of the initial-value problem (2.7)–(2.10) with periodic boundary condition.

Moreover, for all $t > 0$, the following bounds hold:

$$\mathbb{N}_0(f)(w;t) \leq A_0 \exp\left(\frac{C_0}{\varepsilon}t\right), \quad \mathbb{N}_q(f)(w;t) \leq A_0 \exp\left(\frac{C_q}{\varepsilon}t\right), \tag{2.20}$$

$$\inf_{\mathbf{x}} \rho(w;\mathbf{x},t) \geq A_1 \exp\left(-\frac{t}{\varepsilon}\right), \tag{2.21}$$

where C_0 and C_q are the same constants appearing in Proposition 2.1.

As a direct consequence of Proposition 2.1 and Theorem 2.2, we have the following corollary on the upper bounds of the macroscopic quantities.

Corollary 2.3. *Suppose that the conditions in Theorem 2.2 hold. We also assume the Knudsen number $\varepsilon \geq \varepsilon_0 > 0$. Then for all $t > 0$, the following bounds hold:*

$$\sup_{w}\sup_{\mathbf{x}} \{\rho(w;\mathbf{x},t),\ |\boldsymbol{U}(w;\mathbf{x},t)|,\ T(w;\mathbf{x},t)\} \leq C_1 \exp\left(\frac{C_2}{\varepsilon_0}t\right), \tag{2.22}$$

where C_1 and C_2 are positive constants depending only on A_0, A_1, C_0, and C_q.

Proof. By (2.16), (2.20), and (2.21), we have

$$(3T(w;\mathbf{x},t) + |\boldsymbol{U}(w;\mathbf{x},t)|^2)^{\frac{q-3}{2}} \leq \frac{C_q N_q(f)}{\rho(w;\mathbf{x},t)} \leq \frac{C_q A_0}{A_1} \exp\left(\frac{C_q+1}{\varepsilon}t\right). \tag{2.23}$$

Hence

$$T(w;\mathbf{x},t) \leq \frac{1}{3}\left(\frac{C_q A_0}{A_1}\right)^{\frac{2}{q-3}} \exp\left(\frac{2(C_q+1)}{(q-3)\varepsilon}t\right), \quad |\boldsymbol{U}(w,\mathbf{x},t)| \leq \left(\frac{C_q A_0}{A_1}\right)^{\frac{1}{q-3}} \exp\left(\frac{C_q+1}{(q-3)\varepsilon}t\right). \tag{2.24}$$

By (2.15) and (2.24), we have

$$\begin{aligned}\rho(w;\mathbf{x},t) &\leq C_0 N_0(f) T(w;\mathbf{x},t)^{\frac{3}{2}} \\ &\leq 3^{-\frac{3}{2}} C_0 C_q^{\frac{3}{q-3}} A_0^{\frac{q}{q-3}} A_1^{-\frac{3}{q-3}} \exp\left(\frac{3(C_q+1)+(q-3)C_0}{(q-3)\varepsilon}t\right).\end{aligned} \tag{2.25}$$

■

3. Standard MC method.

In this section, we describe the basic MC sampling method to solve the BGK equation (2.7) and establish some error estimates. For simplicity, we will consider that the uncertainty only comes from the initial condition. The case for the random boundary condition is similar.

3.1. MC method.

Suppose we generate M independent and identically distributed (i.i.d.) random samples f_0^i, $i = 1,\dots,M$, according to the random initial condition $f_0(w;\mathbf{x},\mathbf{v})$. Then each $f_0^i(w;\mathbf{x},\mathbf{v})$ will yield a unique analytical solution to (2.7) at time t, denoted by $f^i(w;\mathbf{x},\mathbf{v},t)$. From $f^i(w;\mathbf{x},\mathbf{v},t)$, we can easily compute

$$\begin{aligned}&\rho^i(w;\mathbf{x},t) = \int_{\mathbb{R}^3} f^i(w;\mathbf{x},\mathbf{v},t)\,d\mathbf{v}, \quad \mathbf{m}^i(w;\mathbf{x},t) = \int_{\mathbb{R}^3} \mathbf{v} f^i(w;\mathbf{x},\mathbf{v},t)\,d\mathbf{v}, \\ &E^i(w;\mathbf{x},t) = \int_{\mathbb{R}^3} \frac{|\mathbf{v}|^2}{2} f^i(w;\mathbf{x},\mathbf{v},t)\,d\mathbf{v};\end{aligned} \tag{3.1}$$

then $\boldsymbol{U}^i$ and T^i are given by

$$\boldsymbol{U}^i(w;\mathbf{x},t) = \frac{\mathbf{m}^i(w;\mathbf{x},t)}{\rho^i(w;\mathbf{x},t)}, \quad T^i(w;\mathbf{x},t) = \frac{2\rho^i(w;\mathbf{x},t)E^i(w;\mathbf{x},t) - |\mathbf{m}^i(w;\mathbf{x},t)|^2}{3(\rho^i(w;\mathbf{x},t))^2}. \tag{3.2}$$

Since it is the macroscopic quantities we are interested in, in the following, without further notice we will use a single variable q to denote ρ, $|\boldsymbol{U}|$, or T.

Given the samples q^i, $i = 1, \dots, M$, the MC estimate of the expectation $\mathbb{E}[q(w; \mathbf{x}, t)]$ is given by

$$\mathbb{E}[q(w; \mathbf{x}, t)] \approx E_M[q(w; \mathbf{x}, t)] := \frac{1}{M} \sum_{i=1}^{M} q^i(w; \mathbf{x}, t). \tag{3.3}$$

To estimate the error between $\mathbb{E}[q(w; \mathbf{x}, t)]$ and $E_M[q(w; \mathbf{x}, t)]$, we need the following lemma.

Lemma 3.1. *For every finite sequence $\{Y_j\}_{j=1}^M$ of independent random variables with zero mean in $L^2(\Omega; L^2(D))$,*

$$\left\| \sum_{j=1}^{M} Y_j \right\|^2_{L^2(\Omega; L^2(D))} = \sum_{j=1}^{M} \|Y_j\|^2_{L^2(\Omega; L^2(D))}. \tag{3.4}$$

Proof. From independence of $\{Y_j\}_{j=1}^M$ and that $\mathbb{E}[Y_j] = 0$,

$$\begin{aligned} \left\| \sum_{j=1}^{M} Y_j \right\|^2_{L^2(\Omega; L^2(D))} &= \int_D \mathbb{E}\left[\left(\sum_{j=1}^{M} Y_j \right)^2 \right] \mathrm{d}\mathbf{x} = \int_D \mathbb{V}\left[\sum_{j=1}^{M} Y_j \right] \mathrm{d}\mathbf{x} \\ &= \int_D \sum_{j=1}^{M} \mathbb{V}[Y_j] \, \mathrm{d}\mathbf{x} = \sum_{j=1}^{M} \int_D \mathbb{E}[Y_j^2] \, \mathrm{d}\mathbf{x} = \sum_{j=1}^{M} \|Y_j\|^2_{L^2(\Omega; L^2(D))}. \end{aligned} \tag{3.5}$$

∎

We have the following consistency theorem.

Theorem 3.2. *For any $M \in \mathbb{N}^+$, at time $t = t_1$,*

$$\|\mathbb{E}[q(w; \mathbf{x}, t_1)] - E_M[q(w; \mathbf{x}, t_1)]\|_{L^2(\Omega; L^1(D))} \le M^{-\frac{1}{2}} |D|^{\frac{1}{2}} \|\mathbb{V}[q(w; \mathbf{x}, t_1)]\|^{\frac{1}{2}}_{L^1(D)}. \tag{3.6}$$

Proof. We interpret the M samples $\{f_0^i\}_{i=1}^M$ as unique realizations of M independent samples of f_0 in the probability space $(\Omega, \mathscr{F}, \mathbb{P})$. In other words, $\{f_0^i\}_{i=1}^M$ are i.i.d. copies of $f_0 \in L^1(D \times \mathbb{R}^3)$. As a result, the corresponding copies of macroscopic quantities $\{q^i(w; \mathbf{x}, t_1)\}_{i=1}^M$ derived from the initial data $\{f_0^i\}_{i=1}^M$ are also independent in $L^2(\Omega; L^1(D))$.

Denote $\mathbb{E}[q(w; \mathbf{x}, t_1)] - q^i(w; \mathbf{x}, t_1)$ by $\Delta q^i(w, \mathbf{x}, t_1)$; then

$$\mathbb{E}[\Delta q^i(w, \mathbf{x}, t_1)] = 0 \tag{3.7}$$

and

$$\|\mathbb{E}[q(w; \mathbf{x}, t_1)] - E_M[q(w; \mathbf{x}, t_1)]\|_{L^2(\Omega; L^1(D))} = M^{-1} \left\| \sum_{i=1}^{M} \Delta q^i(w, \mathbf{x}, t_1) \right\|_{L^2(\Omega; L^1(D))}. \tag{3.8}$$

Using the boundedness of domain D,

$$\left\|\sum_{i=1}^{M}\Delta q^i(w,\mathbf{x},t_1)\right\|^2_{L^1(D)} \le |D|\left\|\sum_{i=1}^{M}\Delta q^i(w,\mathbf{x},t_1)\right\|^2_{L^2(D)}. \tag{3.9}$$

Taking the expectation, noting that Δq^i are independent, and using Lemma 3.1, we have

$$\begin{aligned}\left\|\sum_{i=1}^{M}\Delta q^i(w,\mathbf{x},t_1)\right\|_{L^2(\Omega;L^1(D))} &\le |D|^{\frac{1}{2}}\left\|\sum_{i=1}^{M}\Delta q^i(w,x,t_1)\right\|_{L^2(\Omega;L^2(D))}\\ &= |D|^{\frac{1}{2}}\sqrt{\sum_{i=1}^{M}\|\Delta q^i(w,\mathbf{x},t_1)\|^2_{L^2(\Omega;L^2(D))}}\\ &= |D|^{\frac{1}{2}}M^{\frac{1}{2}}\|\Delta q^i(w,\mathbf{x},t_1)\|_{L^2(\Omega;L^2(D))}\\ &= |D|^{\frac{1}{2}}M^{\frac{1}{2}}\|\mathbb{V}[q(w;\mathbf{x},t_1)]\|^{\frac{1}{2}}_{L^1(D)}.\end{aligned} \tag{3.10}$$

As a direct result of Theorem 3.2 and Corollary 2.3, we have the following convergence theorem.

Theorem 3.3. *Under assumptions of Theorem 2.2 and Corollary 2.3, for $0 < t_1 < \infty$, as $M \to \infty$, the MC estimate $E_M[q(w;\mathbf{x},t_1)]$ converges in $L^2(\Omega;L^1(D))$ to $\mathbb{E}[q(w;\mathbf{x},t_1)]$. Furthermore, for any $M \in \mathbb{N}^+$, there holds the error bound*

$$\|\mathbb{E}[q(w;\mathbf{x},t_1)] - E_M[q(w;\mathbf{x},t_1)]\|_{L^2(\Omega;L^1(D))} \le C_1|D|\exp\left(\frac{C_2}{\varepsilon_0}t_1\right)M^{-\frac{1}{2}}. \tag{3.11}$$

Proof. we only need to note that

$$\|\mathbb{V}[q(w;\mathbf{x},t_1)]\|^{\frac{1}{2}}_{L^1(D)} \le \|\mathbb{E}[q^2(w;\mathbf{x},t_1)]\|^{\frac{1}{2}}_{L^1(D)} \le |D|^{\frac{1}{2}}C_1\exp\left(\frac{C_2}{\varepsilon_0}t_1\right). \tag{3.12}$$

■

3.2. MC method with fully discrete scheme. To complete the error analysis, we need to consider the MC method coupled with the fully discrete scheme for the BGK equation, which includes discretization in time, physical space, and velocity space. The details are given in Appendix A. Simply speaking, we are using Gauss quadrature in the velocity space, second-order IMEX-RK scheme for time discretization, and second-order monotonic upstream-centered scheme for conservation laws (MUSCL) [33] finite volume scheme for spatial discretization (under the hyperbolic CFL condition $\Delta t \le C\Delta x$). Overall, this leads to a second-order positivity-preserving and asymptotic-preserving scheme for the deterministic BGK equation. In the following, we assume that the velocity discretization is accurate enough and ignore the work and error in velocity space. It is then reasonable to assume that the numerical solution $q_{\Delta x,\Delta t}(w;\mathbf{x},t_1)$, computed with mesh size Δx and time step Δt corresponding to initial data $f_0(w;\mathbf{x},\mathbf{v})$ up to time t_1, satisfies the following error estimate pointwise in w.

Assumption 3.1. For fixed time $t_1 > 0$, under the hyperbolic CFL condition $\Delta t \le C\Delta x$, we have

$$\|q(w;\mathbf{x},t_1) - q_{\Delta x,\Delta t}(w;\mathbf{x},t_1)\|_{L^1(D)} \le C(w)\left((\Delta x)^2 + (\Delta t)^2\right) \le C_w(\Delta x)^2, \tag{3.13}$$

where C_w is some positive constant.

The MC estimate of the expectation $\mathbb{E}[q(w;\mathbf{x},t)]$ is now given by

$$\mathbb{E}[q(w;\mathbf{x},t)] \approx E_M[q_{\Delta x,\Delta t}(w;\mathbf{x},t)] := \frac{1}{M}\sum_{i=1}^{M} q^i_{\Delta x,\Delta t}(w;\mathbf{x},t). \tag{3.14}$$

We have the following.

Theorem 3.4. *For any $M \in \mathbb{N}^+$, at time $t = t_1$,*

$$\|\mathbb{E}[q(w;\mathbf{x},t_1)] - E_M[q_{\Delta x,\Delta t}(w;\mathbf{x},t_1)]\|_{L^2(\Omega;L^1(D))} \le M^{-\frac{1}{2}}|D|^{\frac{1}{2}}\|\mathbb{V}[q(w;\mathbf{x},t_1)]\|^{\frac{1}{2}}_{L^1(D)} + C_w(\Delta x)^2. \tag{3.15}$$

Proof.

$$\begin{aligned}\|\mathbb{E}[q(w;\mathbf{x},t_1)] - E_M[q_{\Delta x,\Delta t}(w;\mathbf{x},t_1)]\|_{L^2(\Omega;L^1(D))} &\le \|\mathbb{E}[q] - E_M[q]\|_{L^2(\Omega;L^1(D))} \\ &\quad + \|E_M[q] - E_M[q_{\Delta x,\Delta t}]\|_{L^2(\Omega;L^1(D))}.\end{aligned} \tag{3.16}$$

It is enough to apply Theorem 3.2 and Assumption 3.1. ∎

The following corollary is a direct result of Theorem 3.4.

Corollary 3.5. *Under assumptions of Theorem 2.2 and Corollary 2.3, for $0 < t_1 < \infty$, as $M \to \infty$ and Δx, $\Delta t \to 0$, the MC estimate $E_M[q_{\Delta x,\Delta t}(w;\mathbf{x},t_1)]$ converges in $L^2(\Omega;L^1(D))$ to $\mathbb{E}[q(w;\mathbf{x},t_1)]$. Furthermore, for any $M \in \mathbb{N}^+$, there holds the error bound*

$$\|\mathbb{E}[q(w;\mathbf{x},t_1)] - E_M[q_{\Delta x,\Delta t}(w;\mathbf{x},t_1)]\|_{L^2(\Omega;L^1(D))} \le C_1|D|\exp\left(\frac{C_2}{\varepsilon_0}t_1\right)M^{-\frac{1}{2}} + C_w(\Delta x)^2. \tag{3.17}$$

4. Control variate MLMC method.

In this section we first introduce the MLMC method, and then following [11] we discuss the use of control variate techniques to optimize its variance reduction properties locally using two subsequent levels or globally among all levels.

4.1. MLMC method.

The MLMC method is defined as a multilevel discretization in $\mathbf{x}$ and t with a level l dependent number of samples M_l. Suppose we have a nested triangulation $\{\mathcal{T}_l\}_{l=1}^{L}$ of the spatial domain D ($L \in \mathbb{N}^+$ is the number of levels) such that the mesh size Δx_l at level l satisfies

$$\Delta x_l = \sup\{diam(K) : K \in \mathcal{T}_l\} \searrow \text{ as } l \nearrow. \tag{4.1}$$

Set $q^i_{\Delta x_0,\Delta t_0}(w;\mathbf{x},t) := 0$; then given a target level L of spatial resolution, the MLMC estimate of the expectation $\mathbb{E}[q(w;\mathbf{x},t)]$ is given as follows:

$$\begin{aligned}\mathbb{E}[q(w;\mathbf{x},t)] &\approx E^L[q_{\Delta x_L,\Delta t_L}(w;\mathbf{x},t)]\\ &:= \sum_{l=1}^{L} E_{M_l}\left[q_{\Delta x_l,\Delta t_l}(w;\mathbf{x},t) - q_{\Delta x_{l-1},\Delta t_{l-1}}(w;\mathbf{x},t)\right]\\ &= \sum_{l=1}^{L}\sum_{i=1}^{M_l}\frac{1}{M_l}\left[q^i_{\Delta x_l,\Delta t_l}(w;\mathbf{x},t) - q^i_{\Delta x_{l-1},\Delta t_{l-1}}(w;\mathbf{x},t)\right].\end{aligned} \tag{4.2}$$

Hence what we really sample is the difference of solutions at two consecutive levels. At each level l, we separately generate M_l i.i.d. samples f_0^i, $i = 1,\dots,M_l$, of the initial data f_0 on meshes Δx_l and Δx_{l-1}, respectively, and then use the fully discrete scheme for the BGK equation (2.7) to advance solutions $q^i_{\Delta x_l,\Delta t_l}$ and $q^i_{\Delta x_{l-1},\Delta t_{l-1}}$ to a certain time t.

To simplify the notation, we set $q_{\Delta x_0,\Delta t_0}(w;\mathbf{x},t) := 0$ and define the random variable $Y_l := q_{\Delta x_l,\Delta t_l}(w;\mathbf{x},t) - q_{\Delta x_{l-1},\Delta t_{l-1}}(w;\mathbf{x},t)$ and the specific samples $Y_l^i := q^i_{\Delta x_l,\Delta t_l}(w;\mathbf{x},t) - q^i_{\Delta x_{l-1},\Delta t_{l-1}}(w;\mathbf{x},t)$. We have the following consistency and convergence results for the estimator (4.2).

Theorem 4.1. *For any $M_l \in \mathbb{N}^+$, $l = 1,\dots,L$, at time $t = t_1$,*

$$\begin{aligned}\|\mathbb{E}[q(w;\mathbf{x},t_1)] - E^L[q_{\Delta x_L,\Delta t_L}(w;\mathbf{x},t_1)]\|_{L^2(\Omega;L^1(D))} &\le C_w(\Delta x_L)^2\\ &\quad + |D|^{\frac12}\sum_{l=1}^{L} M_l^{-\frac12}\|\mathbb{V}[Y_l]\|^{\frac12}_{L^1(D)}.\end{aligned} \tag{4.3}$$

Proof.

$$\begin{aligned}&\|\mathbb{E}[q] - E^L[q_{\Delta x_L,\Delta t_L}]\|_{L^2(\Omega;L^1(D))}\\ &= \left\|\mathbb{E}[q] - \sum_{l=1}^{L} E_{M_l}[Y_l]\right\|_{L^2(\Omega;L^1(D))}\\ &\le \left\|\mathbb{E}[q] - \sum_{l=1}^{L}\mathbb{E}[Y_l]\right\|_{L^2(\Omega,L^1(D))} + \left\|\sum_{l=1}^{L} E_{M_l}[Y_l] - \sum_{l=1}^{L}\mathbb{E}[Y_l]\right\|_{L^2(\Omega;L^1(D))}\\ &\le \|\mathbb{E}[q] - \mathbb{E}[q_{\Delta x_L,\Delta t_L}]\|_{L^1(D)} + |D|^{\frac12}\sum_{l=1}^{L}\|E_{M_l}[Y_l] - \mathbb{E}[Y_l]\|_{L^2(\Omega;L^2(D))}\\ &= \mathrm{I} + \mathrm{II}.\end{aligned} \tag{4.4}$$

For part I, Assumption 3.1 yields

$$\mathrm{I} = \|q(w;\mathbf{x},t_1) - q_{\Delta x_L,\Delta t_L}(w;\mathbf{x},t_1)\|_{L^1(\Omega;L^1(D))} \le C_w(\Delta x_L)^2. \tag{4.5}$$

For part II, using Lemma 3.1,

$$\mathrm{II} = |D|^{\frac12}\sum_{l=1}^{L} M_l^{-\frac12}\|Y_l^i - \mathbb{E}[Y_l]\|_{L^2(\Omega;L^2(D))} = |D|^{\frac12}\sum_{l=1}^{L} M_l^{-\frac12}\|\mathbb{V}[Y_l]\|^{\frac12}_{L^1(D)}. \tag{4.6}$$

■

Theorem 4.2. *Under the assumptions of Theorem 2.2 and Corollary 2.3, for* $0 < t_1 < \infty$, *as* $M_l \to \infty$ *and* Δx, $\Delta t \to 0$, *the MLMC estimate* $E^L[q_{\Delta x_L, \Delta t_L}(w; \mathbf{x}, t_1)]$ *converges in* $L^2(\Omega; L^1(D))$ *to* $\mathbb{E}[q(w; \mathbf{x}, t_1)]$. *Furthermore, there holds the error bound*

$$\begin{aligned} &\|\mathbb{E}[q(w;\mathbf{x},t_1)] - E^L[q_{\Delta x_L,\Delta t_L}(w;\mathbf{x},t_1)]\|_{L^2(\Omega;L^1(D))} \\ &\quad \le C_w(\Delta x_L)^2 + \left(C_w|D|^{\frac{1}{2}}(\Delta x_1)^2 + C_1|D|\exp\left(\frac{C_2}{\varepsilon_0}t_1\right)\right) M_1^{-\frac{1}{2}} \\ &\quad\quad + \sum_{l=2}^{L} C_w|D|^{\frac{1}{2}}\left((\Delta x_l)^2 + (\Delta x_{l-1})^2\right) M_l^{-\frac{1}{2}}. \end{aligned} \tag{4.7}$$

Proof. From (2.4), we can see that $\mathbb{V}[X] \le \mathbb{E}[X^2]$; then from Theorem 4.1 for $l = 1$,

$$\begin{aligned} \|Y_1^i - \mathbb{E}[Y_1]\|_{L^2(\Omega;L^2(D))} &= \|q^i_{\Delta x_1,\Delta t_1} - \mathbb{E}[q^i_{\Delta x_1,\Delta t_1}]\|_{L^2(\Omega;L^2(D))} \\ &\le \|q^i_{\Delta x_1,\Delta t_1}\|_{L^2(\Omega;L^2(D))} \\ &\le \|q^i_{\Delta x_1,\Delta t_1} - q^i\|_{L^2(\Omega;L^2(D))} + \|q^i\|_{L^2(\Omega;L^2(D))} \\ &\le C_w(\Delta x_1)^2 + |D|^{\frac{1}{2}} C_1 \exp\left(\frac{C_2}{\varepsilon_0}t_1\right), \end{aligned} \tag{4.8}$$

and similarly for $l \ge 2$,

$$\begin{aligned} \|Y_l^i - \mathbb{E}[Y_l]\|_{L^2(\Omega;L^2(D))} &\le \|Y_l^i\|_{L^2(\Omega;L^2(D))} \\ &= \|q^i_{\Delta x_l,\Delta t_l} - q^i_{\Delta x_{l-1},\Delta t_{l-1}}\|_{L^2(\Omega;L^2(D))} \\ &\le \|q^i_{\Delta x_l,\Delta t_l} - q^i\|_{L^2(\Omega;L^2(D))} + \|q^i - q^i_{\Delta x_{l-1},\Delta t_{l-1}}\|_{L^2(\Omega;L^2(D))} \\ &\le C_w((\Delta x_l)^2 + (\Delta x_{l-1})^2). \end{aligned} \tag{4.9}$$

■

Remark 4.3. The summation term on the right-hand side of (4.7) implies that, if the mesh is refined by a factor of 2 as the level increases, then, to balance the errors in different levels, the sample ratio across levels should be chosen as $2^4 = 16$. Furthermore, it should be noted that the above error estimate highly depends on the regularity of the solution and is valid when the solution is smooth (typical when the BGK equation is in the kinetic regime). When the solution contains discontinuities/shocks (typical when the BGK equation is close to the fluid regime), it is well-known that the numerical scheme will not maintain its original order. A second-order scheme as we considered here will generally degenerate to first order or even worse [21]. Therefore, to balance the errors in different levels, the sample ratio can be chosen smaller. Our numerical results in section 5.2 (smooth solutions) and sections 5.3–5.4 (discontinuous solutions) indeed confirmed this prediction (the general trend follows, though the actual ratio value chosen may not be exactly the above predicted number due to the nonnegligible first two terms on the right-hand side of (4.7)).

4.2. Quasi-optimal and optimal MLMC method. In this section we generalize the previous MLMC method following [11]. To start with, take the 2 level MLMC method for example.

Suppose we have a low fidelity (coarse mesh) approximation q_1 and a high fidelity (fine mesh) approximation q_2; then the 2 level MLMC method with control variate reads as follows

$$\mathbb{E}[q] \approx E_{M_1}[\lambda q_1] + E_{M_2}[q_2 - \lambda q_1], \tag{4.10}$$

where the multiplier λ has to be determined in order to minimize the overall variance $\mathbb{V}[q] = \lambda^2\mathbb{V}[q_1] + \mathbb{V}[q_2 - \lambda q_1]$. It can be shown that for independent samples the optimal value of λ is given by

$$\lambda = \frac{\mathrm{Cov}[q_1, q_2]}{2\mathbb{V}[q_1]}. \tag{4.11}$$

When $M_2 \ll M_1$, the contribution from $\mathbb{V}[\lambda q_1]$ is negligible compared to $\mathbb{V}[q_2 - \lambda q_1]$. We therefore can only focus on the minimization of the variance $\mathbb{V}[q_1 - \lambda q_2]$. In this case, the optimal value of λ is given by

$$\lambda = \frac{\mathrm{Cov}[q_1, q_2]}{\mathbb{V}[q_1]} \approx \frac{\sum_{i=1}^{M_2}(q_1^i - \bar{q}_1)(q_2^i - \bar{q}_2)}{\sum_{i=1}^{M_2}(q_1^i - \bar{q}_1)^2}, \tag{4.12}$$

where $\bar{q}_1 = E_{M_2}[q_1]$, $\bar{q}_2 = E_{M_2}[q_2]$, and in the above expression the covariance and variance are estimated directly from the MC samples.

Generally, suppose we have L levels of solutions $\{q_{\Delta x_i, \Delta t_i}\}_{i=1,\dots,L}$, from coarsest level $q_{\Delta x_1, \Delta t_1}$ to finest level $q_{\Delta x_L, \Delta t_L}$. Then the MLMC method with control variates is given by

$$\begin{aligned}\mathbb{E}[q(w;\mathbf{x},t)] &\approx E^L_{CV}[q_{\Delta x_L, \Delta t_L}] \\ &:= \prod_{i=1}^{L}\lambda_i E_{M_1}[q_{\Delta x_1, \Delta t_1}] + \sum_{l=2}^{L}\prod_{i=l}^{L}\lambda_i E_{M_l}[q_{\Delta x_l, \Delta t_l} - \lambda_{l-1} q_{\Delta x_{l-1}, \Delta t_{l-1}}].\end{aligned} \tag{4.13}$$

Note that $\{\lambda_l\}_{l=1}^L$ here are the coefficients to be determined and $\lambda_L = 1$. If we only consider the variance reduction for each pair of consecutive levels, then we can easily get the analogy of (4.11) to estimate $\{\lambda_l\}$, which we refer to as the *quasi-optimal MLMC method*:

$$\lambda_{l-1} = \frac{\mathrm{Cov}[q_{\Delta x_l, \Delta t_l}, q_{\Delta x_{l-1}, \Delta t_{l-1}}]}{\mathbb{V}[q_{\Delta x_{l-1}, \Delta t_{l-1}}]} \approx \frac{\sum_{i=1}^{M_l}(q^i_{\Delta x_l, \Delta t_l} - \bar{q}_{\Delta x_l, \Delta t_l})(q^i_{\Delta x_{l-1}, \Delta t_{l-1}} - \bar{q}_{\Delta x_{l-1}, \Delta t_{l-1}})}{\sum_{i=1}^{M_l}(q^i_{\Delta x_{l-1}, \Delta t_{l-1}} - \bar{q}_{\Delta x_{l-1}, \Delta t_{l-1}})^2}, \tag{4.14}$$

where $\bar{q}_{\Delta x_l, \Delta t_l} = E_{M_l}[q_{\Delta x_l, \Delta t_l}]$.

However, if we focus on minimizing the overall variance of the estimator (4.13) and assume that the levels are independent, then denoting

$$\hat{\lambda}_l = \prod_{i=l}^{L}\lambda_i, \quad l = 1, \dots, L, \tag{4.15}$$

the optimality conditions yield a tridiagonal system for $\hat{\lambda}_l$:

$$\begin{aligned}(4.16)\qquad &\hat{\lambda}_l \mathbb{V}[q_{\Delta x_l,\Delta t_l}] - \hat{\lambda}_{l+1}\frac{M_l}{M_l+M_{l+1}}\mathrm{Cov}[q_{\Delta x_{l+1},\Delta t_{l+1}}, q_{\Delta x_l,\Delta t_l}]\\ &- \hat{\lambda}_{l-1}\frac{M_{l+1}}{M_l+M_{l+1}}\mathrm{Cov}[q_{\Delta x_{l-1},\Delta t_{l-1}}, q_{\Delta x_l,\Delta t_l}] = 0, \quad l = 1,\dots,L-1,\end{aligned}$$

where we assumed $\hat{\lambda}_0 = 0$, $\hat{\lambda}_L = 1$, and $q_{\Delta x_0,\Delta t_0} = 0$. A practical way to solve the above tridiagonal system is to rewrite (4.16) in terms of original λ_i. For simplicity, we denote $\mathbb{V}[q_{\Delta x_l,\Delta t_l}]$ by $\mathbb{V}_l$ and $\mathrm{Cov}[q_{\Delta x_{l+1},\Delta t_{l+1}}, q_{\Delta x_l,\Delta t_l}]$ by Cov_l to get

$$\begin{aligned}(4.17)\qquad &\lambda_1\mathbb{V}_1 - \frac{M_1}{M_1+M_2}\mathrm{Cov}_1 = 0,\\ &\lambda_2\mathbb{V}_2 - \frac{M_2}{M_2+M_3}\mathrm{Cov}_2 - \lambda_1\lambda_2\frac{M_3}{M_2+M_3}\mathrm{Cov}_1 = 0,\\ &\lambda_3\mathbb{V}_3 - \frac{M_3}{M_3+M_4}\mathrm{Cov}_3 - \lambda_2\lambda_3\frac{M_4}{M_3+M_4}\mathrm{Cov}_2 = 0,\\ &\dots\\ &\lambda_{L-1}\mathbb{V}_{L-1} - \frac{M_{L-1}}{M_{L-1}+M_L}\mathrm{Cov}_{L-2} - \lambda_{L-2}\lambda_{L-1}\frac{M_L}{M_{L-1}+M_L}\mathrm{Cov}_{L-2} = 0,\end{aligned}$$

which can be easily solved by recursive substitution. This is what we refer to as the *optimal MLMC method*.

Denote the correlation coefficient of $q_{\Delta x_l,\Delta t_l}$ and $q_{\Delta x_{l+1},\Delta t_{l+1}}$ by

$$(4.18)\qquad r_l = \frac{\mathrm{Cov}[q_{\Delta x_l,\Delta t_l}, q_{\Delta x_{l+1},\Delta t_{l+1}}]}{\left(\mathbb{V}[q_{\Delta x_{l+1},\Delta t_{l+1}}]\mathbb{V}[q_{\Delta x_l,\Delta t_l}]\right)^{\frac{1}{2}}};$$

we can prove the following consistency and convergence results for the estimator (4.13).

Theorem 4.4. *For any $M_l \in \mathbb{N}^+$, $l = 1,\dots,L$, if $\{\lambda_l\}$ are quasi-optimal and exact, i.e.,*

$$(4.19)\qquad \lambda_l = \frac{Cov[q_{\Delta x_l,\Delta t_l}, q_{\Delta x_{l+1},\Delta t_{l+1}}]}{\mathbb{V}[q_{\Delta x_l,\Delta t_l}]},$$

then at time $t = t_1$,

$$\begin{aligned}(4.20)\qquad &\|\mathbb{E}[q(w;\mathbf{x},t_1)] - E^L_{CV}[q_{\Delta x_L,\Delta t_L}(w;\mathbf{x},t_1)]\|_{L^2(\Omega;L^1(D))}\\ &\le C_w(\Delta x_L)^2 + |D|^{\frac{1}{2}}M_1{}^{-\frac{1}{2}}\hat{\lambda}_1\|\mathbb{V}[q_{\Delta x_1,\Delta t_1}]\|^{\frac{1}{2}}_{L^1(D)}\\ &\quad + |D|^{\frac{1}{2}}\sum_{l=2}^{L} M_l{}^{-\frac{1}{2}}\hat{\lambda}_l(1-r_{l-1}^2)^{\frac{1}{2}}\|\mathbb{V}[q_{\Delta x_l,\Delta t_l}]\|^{\frac{1}{2}}_{L^1(D)}.\end{aligned}$$

Proof. The proof is similar to Theorem 4.1. All we need is to note that when λ is quasi-optimal, we have for $l \ge 2$,

$$\begin{aligned}(4.21)\qquad \mathbb{V}[q_{\Delta x_l,\Delta t_l} - \lambda_{l-1}q_{\Delta x_{l-1},\Delta t_{l-1}}] &= \mathbb{V}[q_{\Delta x_l,\Delta t_l}] + \lambda_{l-1}^2\mathbb{V}[q_{\Delta x_{l-1},\Delta t_{l-1}}]\\ &\quad - 2\lambda_{l-1}\mathrm{Cov}[q_{\Delta x_l,\Delta t_l}, q_{\Delta x_{l-1},\Delta t_{l-1}}]\\ &= (1-r_{l-1}^2)\mathbb{V}[q_{\Delta x_l,\Delta t_l}].\end{aligned}$$

■

Theorem 4.5. *Under the assumptions of Theorem 2.2 and Corollary 2.3, and if* $\{\lambda_l\}$ *are quasi-optimal and exact, we have for* $0 < t_1 < \infty$, *as* $M_l \to \infty$ *and* Δx, $\Delta t \to 0$, *the quasi-optimal MLMC estimate* $E^L_{CV}[q_{\Delta x_L, \Delta t_L}(w; \mathbf{x}, t_1)]$ *converges in* $L^2(\Omega; L^1(D))$ *to* $\mathbb{E}[q(w; \mathbf{x}, t_1)]$ *with the error bound*

$$
\begin{aligned}
&\|\mathbb{E}[q(w;\mathbf{x},t_1)] - E^L_{CV}[q_{\Delta x_L,\Delta t_L}(w;\mathbf{x},t_1)]\|_{L^2(\Omega;L^1(D))} \\
&\le C_w(\Delta x_L)^2 + \sum_{l=2}^{L} C_w|D|^{\frac{1}{2}}\hat{\lambda}_l M_l^{-\frac{1}{2}}(1-r_{l-1}^2)^{\frac{1}{2}}(\Delta x_l)^2 \\
&\quad + \left(C_w|D|^{\frac{1}{2}}(\Delta x_1)^2 + C_1|D|\exp\left(\frac{C_2}{\varepsilon_0}t_1\right)\right) M_1^{-\frac{1}{2}}\hat{\lambda}_1.
\end{aligned} \tag{4.22}
$$

Remark 4.6. Note that the computational cost for quasi-optimal and optimal MLMC is the same as the standard MLMC method. One can use the data from MLMC to estimate λ_l using (4.14) or (4.17). Finally, we emphasize that in [15] one of the estimators, the weighted recursive difference estimator, in fact coincides with our optimal MLMC strategy (see also [11]). However, the method has never been analyzed in the case of kinetic equations, and additionally, the quasi-MLMC method does not appear in the previous literature. Without solving a tridiagonal system which may suffer from ill-conditioning, the quasi-MLMC method offers an efficient and robust alternative to the optimal MLMC method.

5. Numerical results.

In this section, we present several numerical examples for the BGK equation (2.7) with random initial condition or random boundary condition. The details of the deterministic solver are provided in Appendix A. Simply speaking, we are solving a reduced system (A.6) and (A.7), which is equivalent to the full BGK equation in one spatial dimension. We use the IMEX-RK scheme for time discretization and finite volume scheme for spatial discretization so that the overall method is second order in both time and space. We choose $x \in [0, 1]$ and $v \in [-5, 5]$, where 40 Legendre–Gauss quadrature points are used in the velocity space to ensure that the error in velocity is negligible. The CFL condition is fixed as $\Delta t = 0.1\Delta x$.

5.1. Error evaluation.

In the following, we assume the uncertainties come from either the initial condition or boundary condition. Since the solution is a random field, the numerical error is a random quantity as well. For error analysis, we therefore compute a statistical estimator by averaging numerical errors from several independent experiments.

More precisely, for each method we perform $K = 40$ experiments and get the corresponding approximations $\{q^{(j)}(x,t)\}$, $j = 1, \dots, K$, where q can be ρ, U, or T. We approximate the overall error in norm $\|\cdot\|_{L^2(\Omega;L^1(D))}$ via

$$
E(t) = \sqrt{\frac{1}{K}\sum_{j=1}^{K}\|q^{(j)}(\cdot,t) - q_{\text{ref}}(\cdot,t)\|^2_{L^1(D)}}, \tag{5.1}
$$

where $q_{\text{ref}}(x,t)$ is the reference solution obtained using the stochastic collocation method [35] with 120 Legendre–Gauss collocation points and $N_x = 1280$ spatial points. We are also interested in the error at each spatial point:

$$E_{\Delta x}(x,t) = \sqrt{\frac{1}{K}\sum_{j=1}^{K}(q^{(j)}(x,t) - q_{\text{ref}}(x,t))^2}. \tag{5.2}$$

Sometimes to better evaluate the error from the random domain, we would like to ignore the error induced by spatial discretization. To achieve this, we consider another kind of reference solution, $q_{\text{rel}}(x,t)$, obtained again using the stochastic collocation with 120 collocation points, while in the spatial domain we use the same finest mesh Δx_L as that in the corresponding MLMC method to obtain $q^{(j)}(x,t)$. Therefore, we can assess the error as

$$E_{\text{rel}\Delta x}(x,t) = \sqrt{\frac{1}{K}\sum_{j=1}^{K}(q^{(j)}(x,t) - q_{\text{rel}}(x,t))^2}. \tag{5.3}$$

In each of the following tests, we perform two stages of computations. The experimental stage is to determine the optimal sample allocation parameters (there is some guidance from the theoretical estimates—see Remark 4.3—but we still choose to do a careful testing just as a way to verify the theory). The simulation stage is to perform various methods to estimate the physical quantities of interest.

5.2. Test 1: Smooth random initial condition. We first consider the BGK equation subject to random initial condition:

$$f^0(\mathbf{x},\mathbf{v},z) = 0.5M_{\rho,\boldsymbol{U},T} + 0.5M_{\rho,-\boldsymbol{U},T} \tag{5.4}$$

with

$$M_{\rho,\boldsymbol{U},T}(\mathbf{x},\mathbf{v},z) = \frac{\rho(\mathbf{x},z)}{(2\pi T(\mathbf{x},z))^{\frac{3}{2}}}\exp\left(-\frac{|\mathbf{v}-\boldsymbol{U}(\mathbf{x},z)|^2}{2T(\mathbf{x},z)}\right), \tag{5.5}$$

where

$$\begin{aligned}\rho(\mathbf{x},z) &= \frac{2+\sin(2\pi x)+\frac{1}{2}\sin(4\pi x)z}{3}, \quad \boldsymbol{U}(\mathbf{x}) = (0.2,0,0),\\ T(\mathbf{x},z) &= \frac{3+\cos(2\pi x)+\frac{1}{2}\cos(4\pi x)z}{4},\end{aligned} \tag{5.6}$$

and the random variable z obeys the uniform distribution on $[-1,1]$. The periodic boundary condition is used, and the Knudsen number $\varepsilon = 1$.

To determine the number of samples needed in MC and MLMC methods as well as the sample ratio across levels in MLMC methods, we proceed as follows.

In the MC method, we consider a series of spatial discretizations: $N = 10, 20, 30, 40$, and for each case, we vary the sample size as $M = 5, 10, 15, \ldots$ The results are shown in Figure 1 (left), where we plot the error (5.1). It can be observed that when the number of samples is small, the statistical error dominates, and when there are enough samples, the spatial error dominates. Therefore, we can roughly determine the minimum number of samples needed so that the statistical error $O(M^{-\frac{1}{2}})$ balances with the spatial/temporal error $O(\Delta x^2)$:

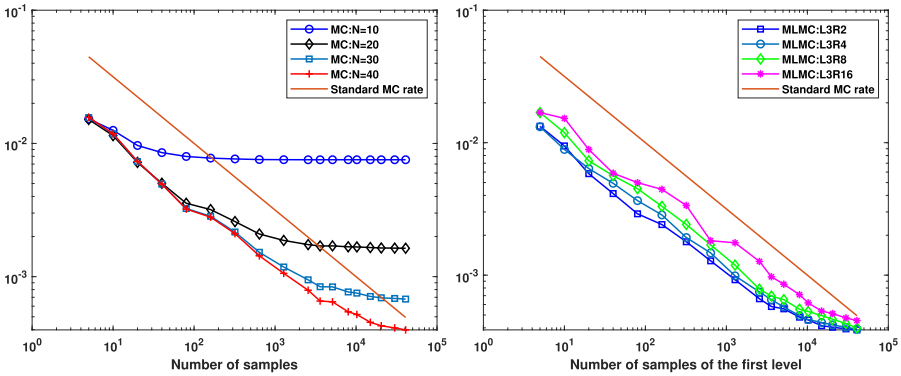

Figure 1. *Test* 1*: Error* (5.1) *(density* ρ*) of MC method (left) and MLMC method (right) versus number of samples (for MLMC, it is the number of samples in the first level).*

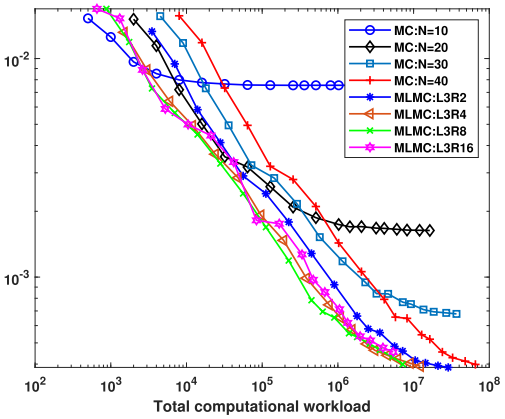

Figure 2. *Test* 1*: Error* (5.1) *(density* ρ*) of MC and MLMC methods versus computational workload.*

- $N = 10$, $M \approx 40$.
- $N = 20$, $M \approx 640$.
- $N = 30$, $M \approx 3300$.
- $N = 40$, $M \approx 10240$.

In the MLMC method, we consider three levels of spatial discretizations: $N_1 = 10$, $N_2 = 20$, $N_3 = 40$, and the corresponding number of samples at each level are chosen as M_1, $M_2 = \frac{M_1}{a}$, and $M_3 = \frac{M_1}{a^2}$, where we test different ratios $a = 2, 4, 8, 16$. We then vary the starting sample size as $M_1 = 16, 32, 48, \ldots$ The results are shown in Figure 1 (right), where we can see that regardless of ratios, the statistical error and spatial/temporal error are roughly balanced when $M_1 \approx 10240$ (the error saturates when the sample size further increases).

In Figure 2 we combine all the previous MC and MLMC results under the scale of workload. Since we are essentially solving a 1D BGK problem, the workload for one deterministic run up to certain time with N spatial points is $O(N^2)$. Then for the MC method with M samples, the total work is $O(MN^2)$. For the MLMC method with ratio a, the amount of work is $O(M_1N_1^2 + M_2(N_1^2 + N_2^2) + M_3(N_2^2 + N_3^2) = \frac{a^2+5a+20}{a^2}M_1N_1^2)$. As we can see clearly from Figure 2, with the same workload, the MLMC methods can achieve better accuracy compared

to various MC methods. Among MLMC methods with different ratios, there is no significant difference except for ratio $a = 2$. Therefore, we empirically set $a = 4$ for this smooth random initial condition test.

Now we fix the mesh sizes $N_1 = 10$, $N_2 = 20$, $N_3 = 40$ and sample sizes $M_1 = 10240$, $M_2 = 2560$, $M_3 = 640$ in the MLMC method. We then find the number of samples in the MC method such that they have the same workload. This means

- $N = 10$, $M = 30720$,
- $N = 20$, $M = 7680$,
- $N = 30$, $M = 3413$,
- $N = 40$, $M = 1920$.

Note that comparing with the numbers we found earlier, for $N = 10$ and 20, the numbers of samples are far beyond the minimum number of samples needed, while for $N = 30$, M is around the minimum number of samples needed. Finally for $N = 40$, the number of samples here is not enough to balance the statistical error and numerical error in the MC method. Using the above parameters, we compare the errors of the standard MC method and three MLMC methods, namely, the standard MLMC, the quasi-optimal MLMC, and optimal MLMC. The results are shown in Figure 3, from which we clearly see the better accuracy of MLMC methods compared to standard MC for fixed workload. On the other hand, the differences of three MLMC methods are not obvious in this example.

Next we examine the errors of the three MLMC methods as defined in (5.2), (5.3). The results are gathered in Figure 4. We can see that the three MLMC methods perform equally well in this test (the differences of the three methods are not significant, though the optimal MLMC has the smallest error overall), largely because the solution is smooth.

To better understand this, we plot the values of λ_1 and λ_2 in the quasi-optimal and optimal MLMC methods in Figure 5. We can see that almost all values are not far from 1, which means the methods are not far from the standard MLMC.

5.3. Test 2: Shock tube problem. In this test, we consider two kinds of shock tube problems with random initial condition. The first one has uncertainty in the interface location:

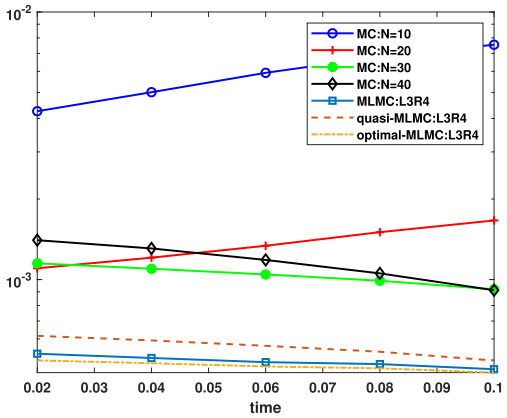

Figure 3. *Test* 1*: Time evolution of the errors* (5.1) *(density* ρ*) using MC and various MLMC methods.*

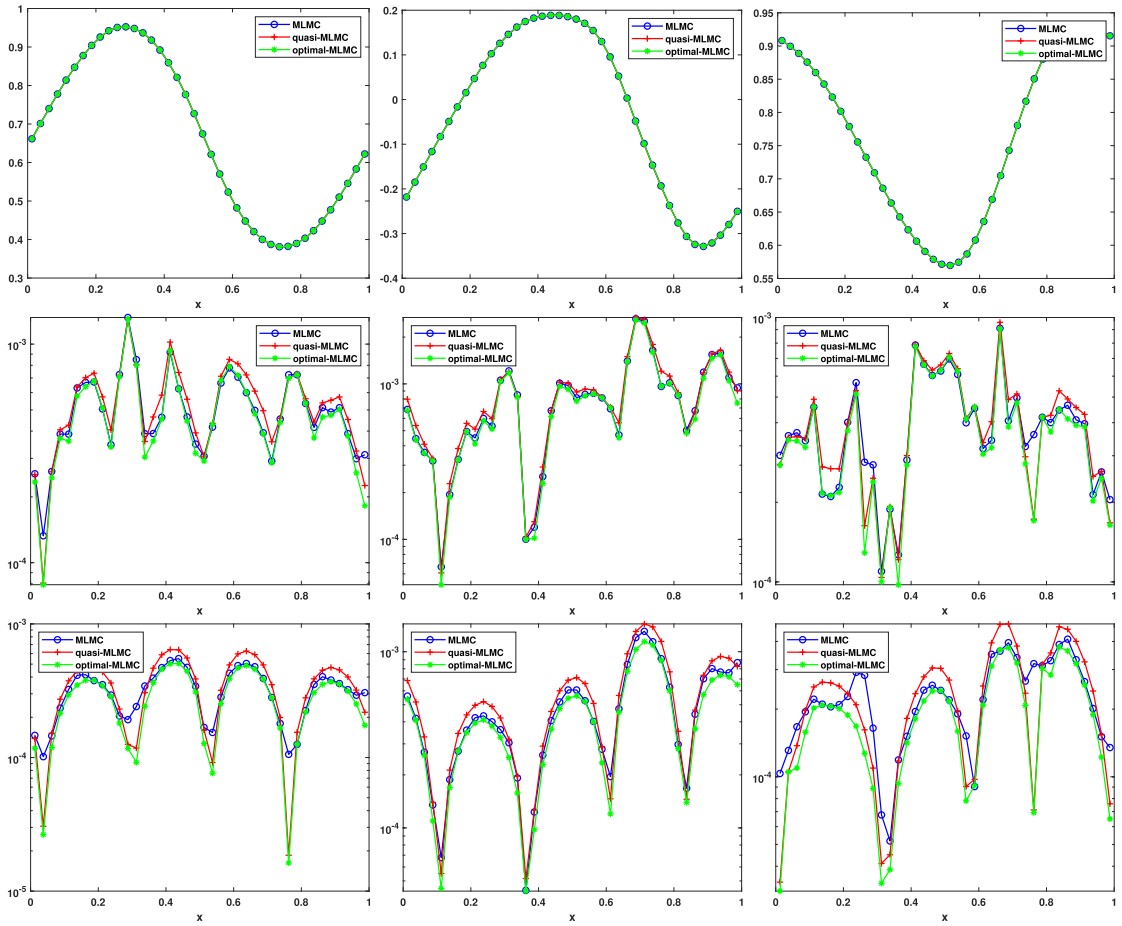

Figure 4. *Test* 1*: Approximated expectation of density* $\mathbb{E}[\rho]$ *(left), velocity* $\mathbb{E}[U]$ *(middle), and temperature* $\mathbb{E}[T]$ *(right) using MLMC, quasi-optimal MLMC, and optimal MLMC methods at time* $t = 0.1$ *(top row). Error* (5.2) *of expectation of density (left), velocity (middle), and temperature (right) using three MLMC methods (middle row). Relative error* (5.3) *of expectation of density (left), velocity (middle), and temperature (right) using three MLMC methods (bottom row).*

$$\text{(5.7)}\qquad \text{I}: \begin{cases} \rho_l = 1, & \boldsymbol{U}_l = (0,0,0), & T_l = 1, & f_0 = M_{\rho_l,\boldsymbol{U}_l,T_l} & x \le 0.5 + 0.05z, \\ \rho_r = 0.125, & \boldsymbol{U}_r = (0,0,0), & T_r = 0.25, & f_0 = M_{\rho_r,\boldsymbol{U}_r,T_r} & x > 0.5 + 0.05z. \end{cases}$$

The second one has uncertainty in the state variables:

$$\text{(5.8)}\qquad \text{II}: \begin{cases} \rho_l = 1 + 0.1(z+1), & \boldsymbol{U}_l = (0,0,0), & T_l = 1, & f_0 = M_{\rho_l,\boldsymbol{U}_l,T_l} & x \le 0.5, \\ \rho_r = 0.125, & \boldsymbol{U}_r = (0,0,0), & T_r = 0.25, & f_0 = M_{\rho_r,\boldsymbol{U}_r,T_r} & x > 0.5. \end{cases}$$

The random variable z obeys the uniform distribution on $[-1, 1]$. We set the Knudsen number $\varepsilon = 10^{-6}$ so that the problem is close to the fluid regime.

For problem I, similarly as the previous example, we perform a series of tests to determine the optimal number of samples needed in MC and MLMC methods as well as the sample

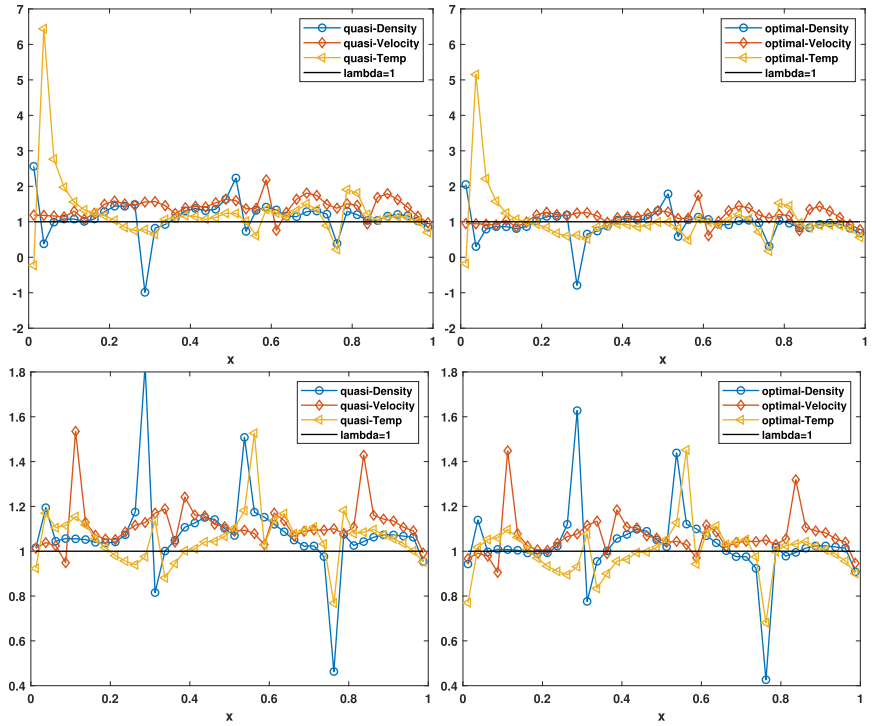

Figure 5. *Test* 1*: Values of* λ_1 *in quasi-optimal (left) and optimal (right) MLMC methods (top row). Values of* λ_2 *in quasi-optimal (left) and optimal (right) MLMC methods (bottom row).*

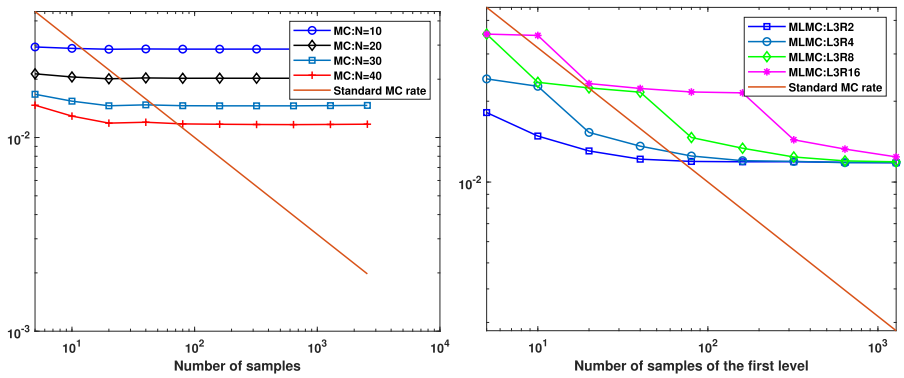

Figure 6. *Test* 2 (I)*: Error* (5.1) *(density* ρ*) of MC method (left) and MLMC method (right) versus number of samples (for MLMC, it is the number of samples in the first level).*

ratio across levels. Figure 6 shows the analogous tests as those in Figure 1. The main difference from the previous example is that the errors saturate much quicker as the number of samples increases. This is due to the low regularity of the solution so that the error from spatial/temporal discretization dominants easily. In Figure 7 we combine both MC and MLMC results under the scale of workload. Similarly as what we observed in Figure 2, with the same workload, the MLMC methods can achieve better accuracy compared to MC. In

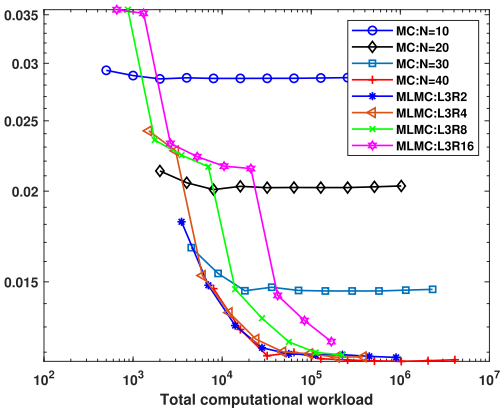

Figure 7. *Test* 2 (I)*: Error* (5.1) *(density* ρ*) of MC and MLMC methods versus computational workload.*

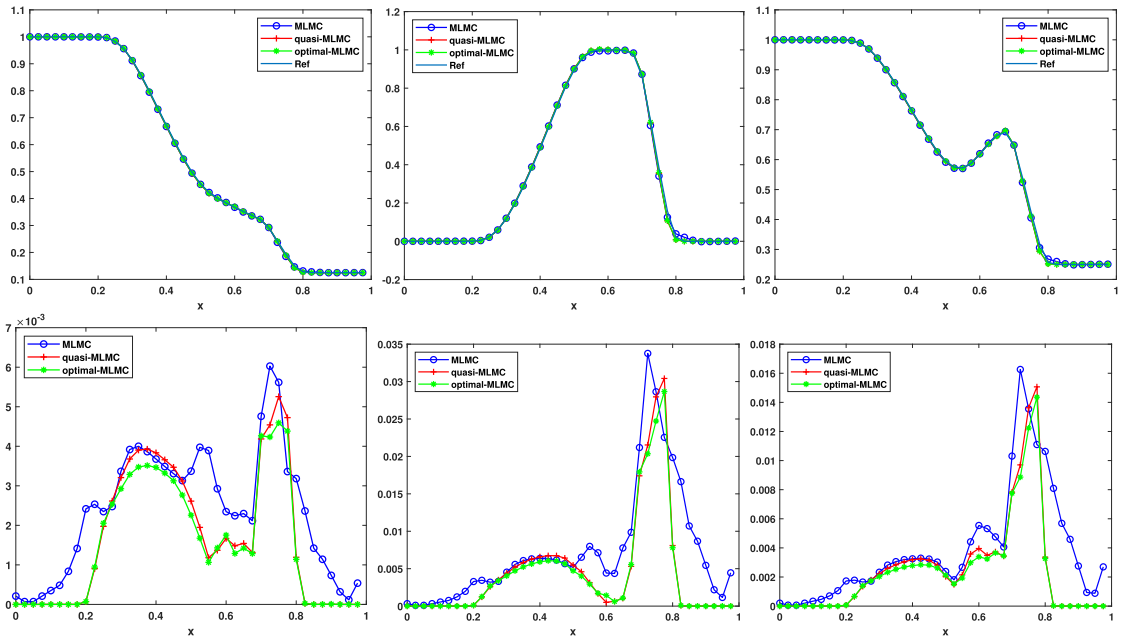

Figure 8. *Test* 2 (I)*: Approximated expectation of density* $\mathbb{E}[\rho]$ *(left), velocity* $\mathbb{E}[U]$ *(middle), and temperature* $\mathbb{E}[T]$ *(right) using MLMC, quasi-optimal MLMC, and optimal MLMC methods at time* $t = 0.15$ *(top row). Relative error* (5.3) *of expectation of density (left), velocity (middle), and temperature (right) using three MLMC methods (bottom row).*

addition, the MLMC methods with ratios $a = 2, 4$ are more accurate than $a = 8, 16$. This is consistent with our earlier theoretical prediction; see Remark 4.3. From the right plot in Figure 6, we also see that $M_1 \approx 320$ is the minimum number of samples needed for the MLMC method to balance the statistical error and spatial/temporal error. Therefore, we choose the following parameters in the MLMC methods: mesh sizes $N_1 = 10$, $N_2 = 20$, $N_3 = 40$ and sample sizes $M_1 = 320$, $M_2 = 80$, $M_3 = 20$. In Figures 8–9, we report the results obtained using the standard MLMC, quasi-optimal MLMC, and optimal MLMC

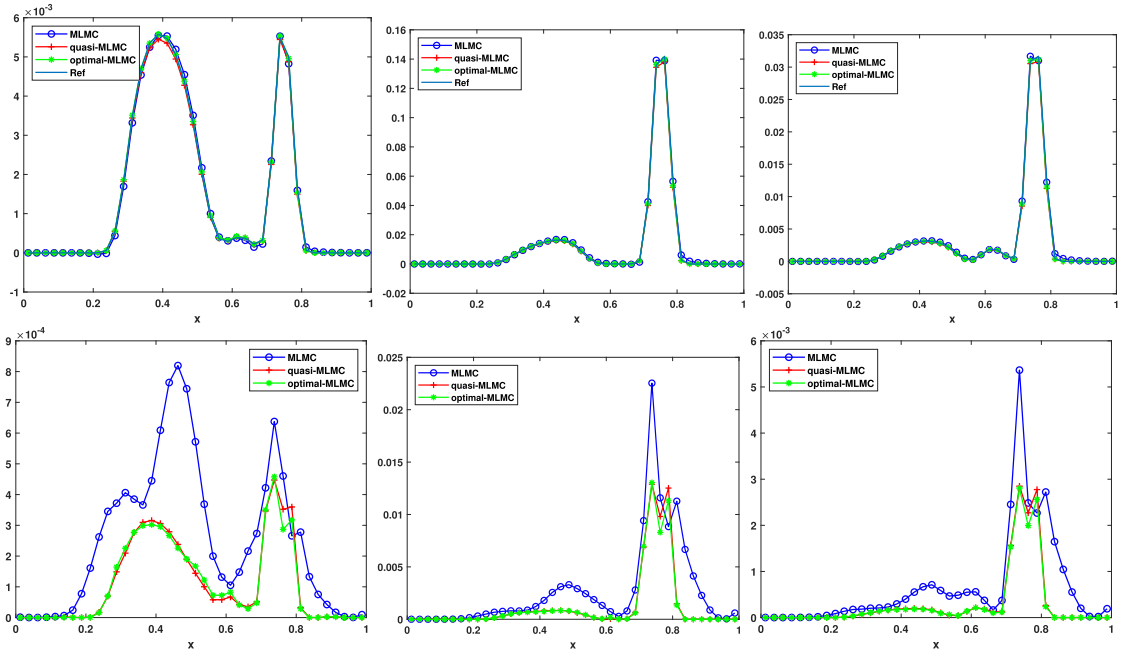

Figure 9. *Test* 2 (I)*: Approximated variance of density* $\mathbb{V}[\rho]$ *(left), velocity* $\mathbb{V}[U]$ *(middle), and temperature* $\mathbb{V}[T]$ *(right) using MLMC, quasi-optimal MLMC, and optimal MLMC methods at time* $t = 0.15$ *(top row). Relative error* (5.3) *of variance of density (left), velocity (middle), and temperature (right) using three methods (bottom row).*

methods. We mainly examine the approximation to the expectation $\mathbb{E}[q]$ as the proposed quasi-MLMC and optimal MLMC methods are especially designed to minimize the variance in the estimation of $\mathbb{E}[q]$. As a by-product, we also plot the approximation to the variance $\mathbb{V}[q]$ using the samples generated for expectation. Note that the MLMC methods are based on the linearity of the expectation operator, not the variance operator. Hence to approximate the variance, we approximate separately two different expectations $\mathbb{E}[q^2]$ and $\mathbb{E}[q]$ and use them to obtain $\mathbb{V}[q] = \mathbb{E}[q^2] - (\mathbb{E}[q])^2$. We refer to [20] for other approaches to variance approximation including error control. The results clearly show that both control variate MLMC methods outperform the standard MLMC in regions where the solution presents strong variations, namely, close to the shock position. Although the results are very close, as expected, the optimal MLMC method performs slightly better than the quasi-optimal MLMC.

For problem II, we choose the following parameters: mesh sizes $N_1 = 10, N_2 = 20, N_3 = 40$ and number of samples $M_1 = 640, M_2 = 160, M_3 = 40$ (these parameters are chosen based on a similar test as problem I, and we omit the detail). The results are shown in Figures 10 and 11, where the same observation as problem I is obtained.

To better see the difference of the three MLMC methods, we plot the values of λ_1 and λ_2 in the quasi-optimal and optimal MLMC methods for both problems I and II in Figures 12 and 13. It is clear that for these problems with shocks/discontinuities the values are far from one in various regions of the computational domain. This is particularly true for the temperature and velocity in agreement with the corresponding errors observed in the previous figures.

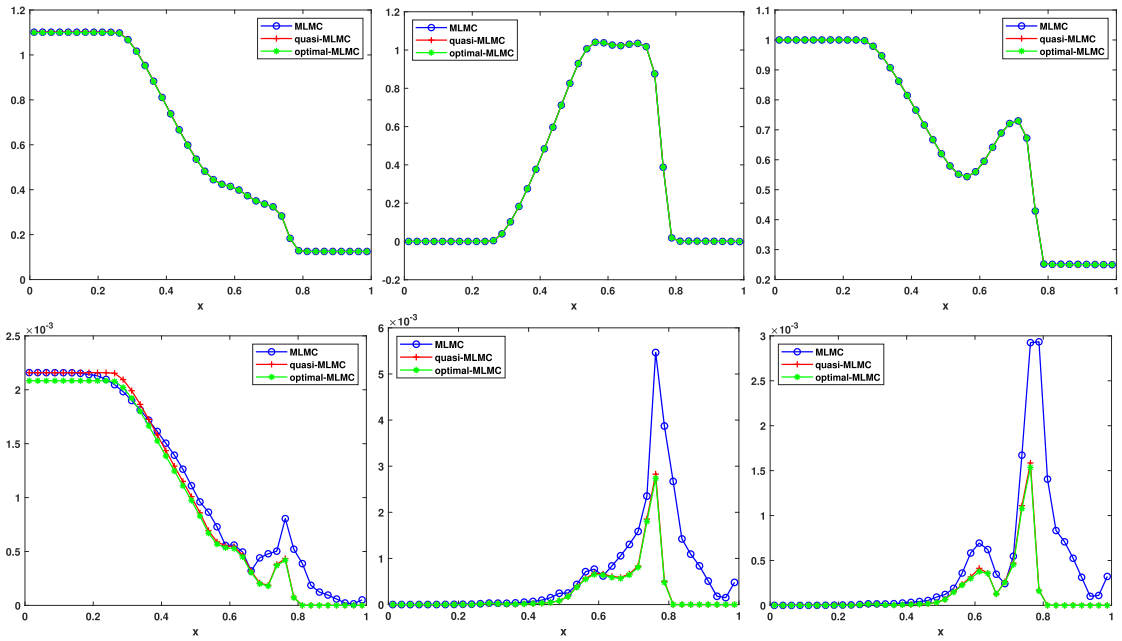

Figure 10. *Test* 2 (II)*: Approximated expectation of density* $\mathbb{E}[\rho]$ *(left), velocity* $\mathbb{E}[U]$ *(middle), and temperature* $\mathbb{E}[T]$ *(right) using MLMC, quasi-optimal MLMC, and optimal MLMC methods at time* $t = 0.15$ *(top row). Relative error* (5.3) *of expectation of density (left), velocity (middle), and temperature (right) using three MLMC methods (bottom row).*

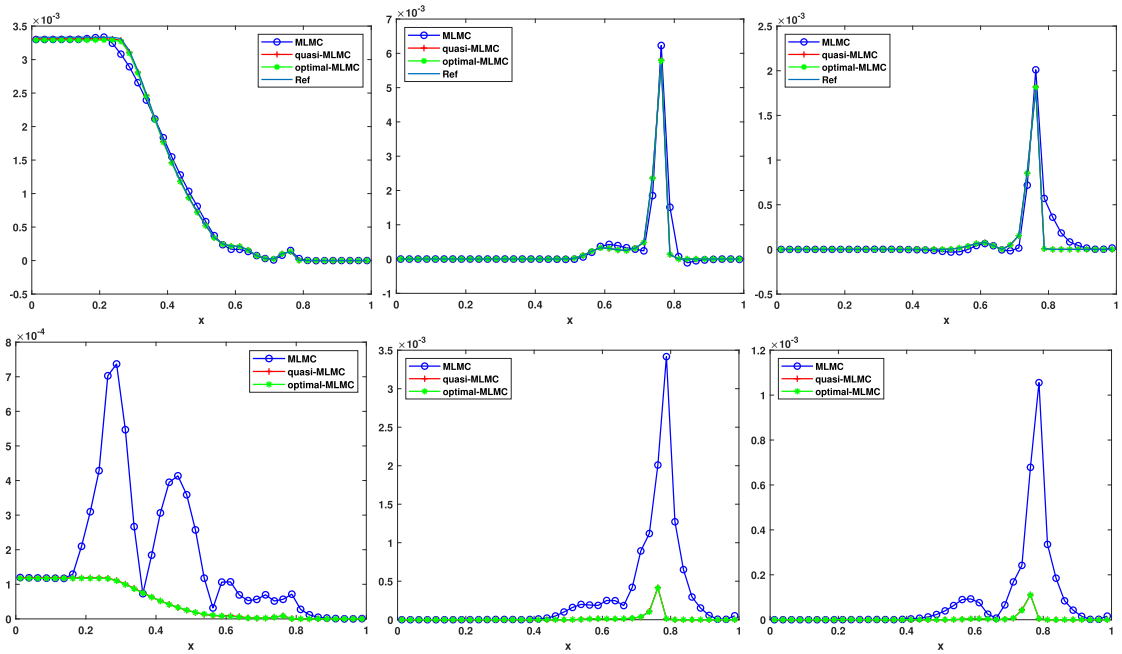

Figure 11. *Test* 2 (II)*: Approximated variance of density* $\mathbb{V}[\rho]$ *(left), velocity* $\mathbb{V}[U]$ *(middle), and temperature* $\mathbb{V}[T]$ *(right) using MLMC, quasi-optimal MLMC, and optimal MLMC methods at time* $t = 0.15$ *(top row). Relative error* (5.3) *of variance of density (left), velocity (middle), and temperature (right) using three methods (bottom row).*

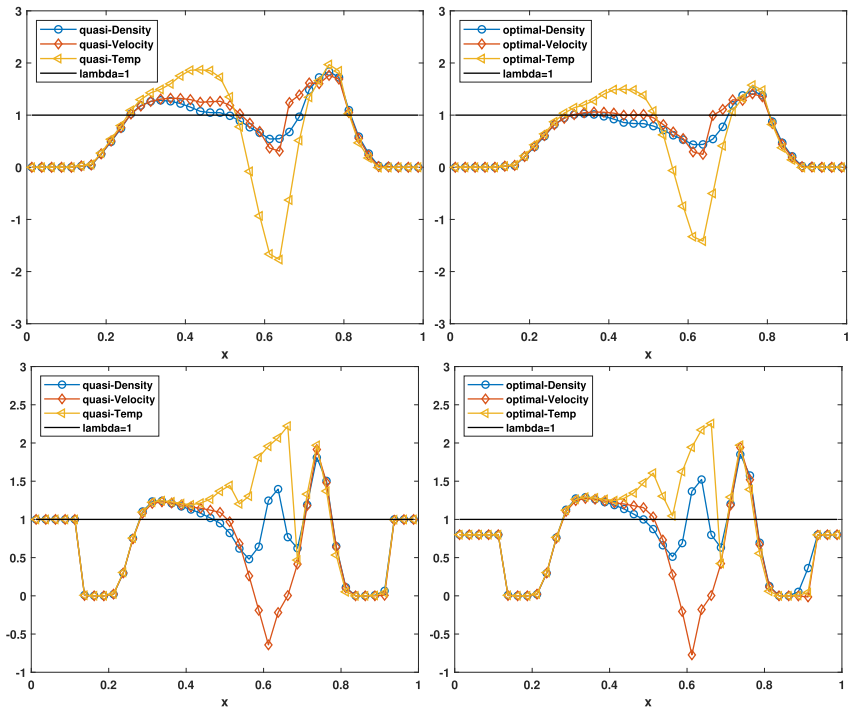

Figure 12. *Test* 2 (I)*: Values of* λ_1 *in quasi-optimal (left) and optimal (right) MLMC methods (top row). Values of* λ_2 *in quasi-optimal (left) and optimal (right) MLMC methods (bottom row).*

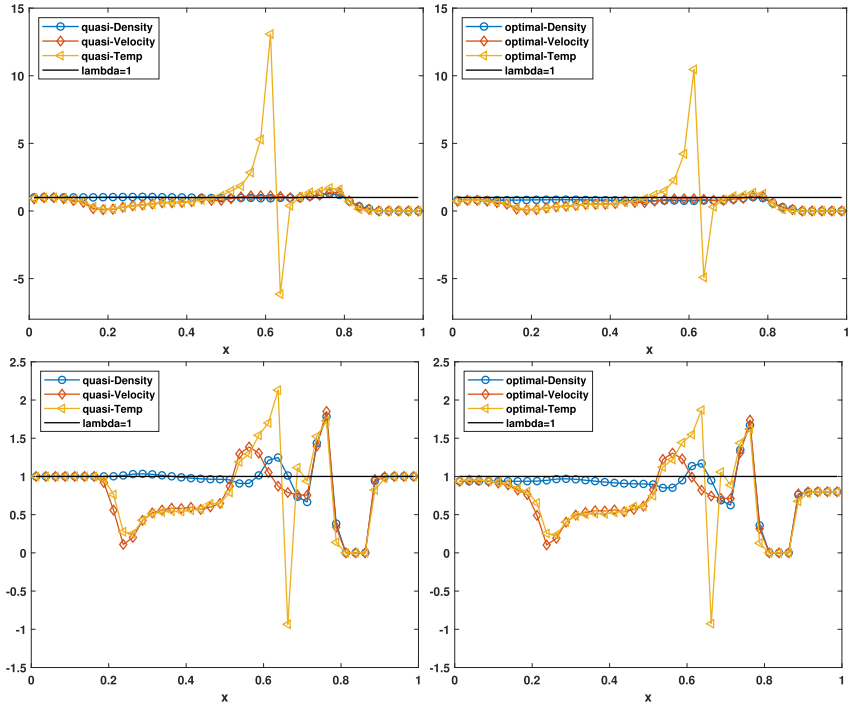

Figure 13. *Test* 2 (II)*: Values of* λ_1 *in quasi-optimal (left) and optimal (right) MLMC methods (top row). Values of* λ_2 *in quasi-optimal (left) and optimal (right) MLMC methods (bottom row).*

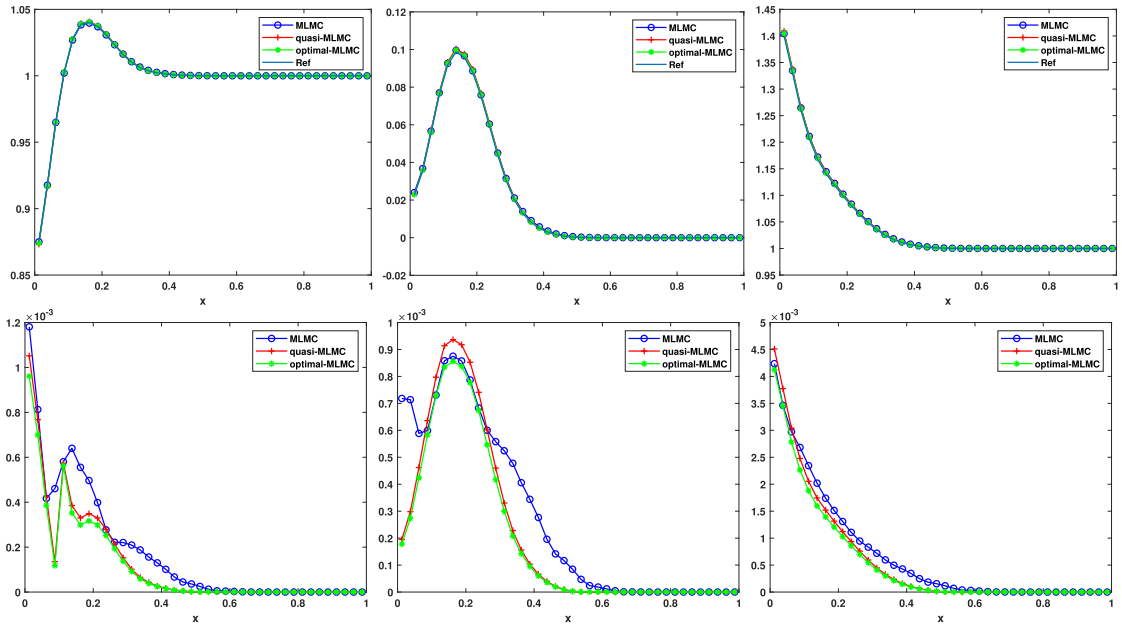

Figure 14. *Test* 3*: Approximated expectation of density* $\mathbb{E}[\rho]$ *(left), velocity* $\mathbb{E}[U]$ *(middle), and temperature* $\mathbb{E}[T]$ *(right) using MLMC, quasi-optimal MLMC, and optimal MLMC methods at time* $t = 0.1$ *(top row). Relative error* (5.3) *of expectation of density (left), velocity (middle), and temperature (right) using three methods (bottom row).*

5.4. Test 3: Sudden heating problem. In the last test, we consider a problem with random boundary condition. The gas is initially in a constant state with $\rho_0 = 1$, $\boldsymbol{U}_0 = (0, 0, 0)$, $T_0 = 1$, and $f_0(\mathbf{x}, \mathbf{v}) = M_{\rho_0, \boldsymbol{U}_0, T_0}$. At time $t = 0$, we suddenly change the wall temperature at left boundary $x = 0$ to

$$T_w(z) = 3(T_0 + sz), \ s = 0.2, \tag{5.9}$$

where the random variable z obeys the uniform distribution on $[-1, 1]$. We assume purely diffusive Maxwell boundary condition at $x = 0$ and homogeneous Neumann boundary condition at $x = 1$. The Knudsen number is set as $\varepsilon = 0.1$. This is a classical benchmark test in kinetic theory. With the sudden rise of the wall temperature, the gas close to the wall is heated, and accordingly the pressure rises sharply and pushes the gas away, forming a shock propagating into the domain.

We compare the three MLMC methods using the following parameters: mesh sizes $N_1 = 10, N_2 = 20, N_3 = 40$ and number of samples $M_1 = 1280, M_2 = 320, M_3 = 80$ (these parameters are chosen based on a similar test as in previous examples). The results are shown in Figure 14 and Figure 15. Again the control variate MLMC methods outperform the standard MLMC in all simulations, and the optimal MLMC method yields slightly better results than the quasi-optimal MLMC.

6. Conclusions. We have introduced a control variate MLMC method for the BGK model of the Boltzmann equation with uncertainty. Well-posedness of the BGK equation with random parameters, consistency, and convergence analysis for various MC-type methods are

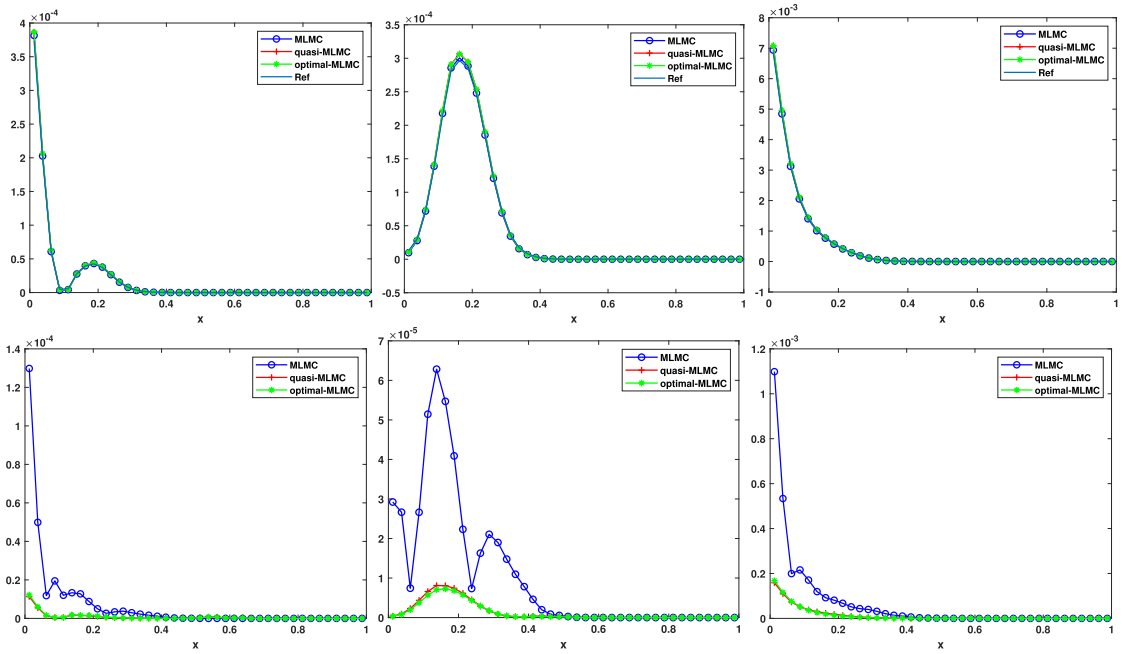

Figure 15. *Test* 3*: Approximated variance of density* $\mathbb{V}[\rho]$ *(left), velocity* $\mathbb{V}[U]$ *(middle) and temperature* $\mathbb{V}[T]$ *(right) using MLMC, quasi-optimal MLMC and optimal MLMC methods at time* $t = 0.1$ *(top row). Relative error* (5.3) *of variance of density (left), velocity (middle) and temperature (right) using three methods (bottom row).*

established. Extensive numerical results confirm that the MLMC methods perform much better than the standard MC, and the control variate MLMC is capable of providing further improvement over the conventional MLMC, in particular for problems close to fluid regimes and in presence of discontinuities, where the fidelity degree of the various levels is reduced and traditional gPC-SG based methods may fail (see [7]). In addition to an optimal strategy, we have introduced a simplified quasi-optimal approach that does not require solving a tridiagonal system of linear equations. In the numerical examples, this simplified approach provided only slightly less accurate results than those obtained with the optimal strategy. The control variate MLMC methods here developed naturally extend to other kinetic equations of Boltzmann type which combines deterministic discretizations in the phase space with MC sampling in the random space. In particular, even if our study were limited to one space dimension, we expect the gains of MLMC methods over standard MC to be even more significant in higher dimensions.

Appendix A. Dimension reduction method and deterministic solver for the BGK equation.

In this appendix, we briefly describe the dimension reduction method adopted to reduce the computational complexity of the BGK equation and the details of the numerical methods used to discretize time, physical space, and velocity space. Since the MC methods are nonintrusive, our discussion will be based on the deterministic equation (1.2) for simplicity.

A.1. The Chu reduction method. The BGK equation (1.2) is formulated in a six-dimensional phase space where computations can be extremely expensive. Under certain homogeneity assumptions, one can reduce the dimension using the so-called Chu reduction [6].

Let $\mathbf{x} = (x_1, x_2, x_3)$, $\mathbf{v} = (v_1, v_2, v_3)$, and $\boldsymbol{U} = (U_1, U_2, U_3)$. If the solution f only varies in one spatial dimension, then effectively we are solving a one-dimensional problem, and it is reasonable to assume the following:

$$\partial_{x_2} f = \partial_{x_3} f = 0, \quad U_2 = U_3 = 0. \tag{A.1}$$

Then (1.2) becomes

$$\partial_t f(x_1, v_1, v_2, v_3, t) + v_1 \partial_{x_1} f(x_1, v_1, v_2, v_3, t) = \frac{1}{\varepsilon} \left(M[f] - f(x_1, v_1, v_2, v_3, t) \right), \tag{A.2}$$

where

$$M[f](x_1, v_1, v_2, v_3, t) = \frac{\rho(x_1, t)}{(2\pi T(x_1, t))^{\frac{3}{2}}} \exp\left(-\frac{(v_1 - U_1(x_1, t))^2 + v_2^2 + v_3^2}{2T(x_1, t)} \right). \tag{A.3}$$

The Chu reduction proceeds by introducing two distribution functions:

$$\phi(x_1, v_1, t) := \iint_{\mathbb{R}^2} f(x_1, v_1, v_2, v_3, t) \, \mathrm{d}v_2 \mathrm{d}v_3, \tag{A.4}$$

$$\psi(x_1, v_1, t) := \iint_{\mathbb{R}^2} \left(\frac{1}{2} v_2^2 + \frac{1}{2} v_3^2 \right) f(x_1, v_1, v_2, v_3, t) \, \mathrm{d}v_2 \mathrm{d}v_3. \tag{A.5}$$

It is then easy to derive that ϕ and ψ satisfy the following system:

$$\partial_t \phi(x_1, v_1, t) + v_1 \partial_{x_1} \phi(x_1, v_1, t) = \frac{1}{\varepsilon} \left(M_\phi(x_1, v_1, t) - \phi(x_1, v_1, t) \right), \tag{A.6}$$

$$\partial_t \psi(x_1, v_1, t) + v_1 \partial_{x_1} \psi(x_1, v_1, t) = \frac{1}{\varepsilon} \left(M_\psi(x_1, v_1, t) - \psi(x_1, v_1, t) \right), \tag{A.7}$$

where

$$M_\phi(x_1, v_1, t) := \iint_{\mathbb{R}^2} M[f] \, \mathrm{d}v_2 \mathrm{d}v_3 = \frac{\rho(x_1, t)}{\sqrt{2\pi T(x_1, t)}} \exp\left(-\frac{(v_1 - U_1(x_1, t))^2}{2T(x_1, t)} \right), \tag{A.8}$$

$$M_\psi(x_1, v_1, t) := \iint_{\mathbb{R}^2} \left(\frac{1}{2} v_2^2 + \frac{1}{2} v_3^2 \right) M[f] \, \mathrm{d}v_2 \mathrm{d}v_3 = T(x_1, t) M_\phi. \tag{A.9}$$

Denoting $\int_{\mathbb{R}} \cdot \, \mathrm{d}v_1 = \langle \cdot \rangle$, it is easy to see the following relation holds:

$$\begin{aligned} \rho &= \int_{\mathbb{R}} \phi \, \mathrm{d}v_1 = \int_{\mathbb{R}} M_\phi \, \mathrm{d}v_1, \\ m &= \rho U_1 = \int_{\mathbb{R}} v_1 \phi \, \mathrm{d}v_1 = \int_{\mathbb{R}} v_1 M_\phi \, \mathrm{d}v_1, \\ E &= \frac{1}{2} \rho U_1^2 + \frac{3}{2} \rho T = \int_{\mathbb{R}} \left(\frac{1}{2} v_1^2 \phi + \psi \right) \mathrm{d}v_1 = \int_{\mathbb{R}} \left(\frac{1}{2} v_1^2 M_\phi + M_\psi \right) \mathrm{d}v_1. \end{aligned} \tag{A.10}$$

Now our task is to solve the reduced 1D BGK system (A.6)–(A.7).

A.2. The fully discrete scheme. The fully discrete scheme used to solve (A.6)–(A.7) consists of three components: velocity discretization, time discretization, and spatial discretization.

Velocity discretization. In the velocity space, we follow the discrete velocity method (see section 4.1.1 in [12] or [23] for example), which satisfies a discrete entropy decay property.

We first truncate the infinite velocity domain into a bounded interval $[-R, R]$ and then discretize it using N_v-point Gauss quadrature with (ξ_k, w_k), $k = 1, 2, \ldots, N_v$, as abscissae and weights. To obtain M_ϕ, M_ψ from ϕ and ψ, normally one could use the relation in (A.10), where the continuous integral is replaced by the Gauss quadrature. However, due to the domain truncation error, the resulting moments are not sufficiently accurate. To remove this error, we assume

$$M_\phi = \exp(\alpha_1 + \alpha_2 v_1 + \alpha_3 v_1^2), \quad M_\psi = -\frac{1}{2\alpha_3} M_\phi \tag{A.11}$$

and determine α_1, α_2, α_3 such that

$$\begin{bmatrix} \langle M_\phi \rangle \\ \langle v_1 M_\phi \rangle \\ \langle \frac{1}{2} v_1^2 M_\phi + M_\psi \rangle \end{bmatrix} = \begin{bmatrix} \langle \phi \rangle \\ \langle v_1 \phi \rangle \\ \langle \frac{1}{2} v_1^2 \phi + \psi \rangle \end{bmatrix} := \begin{bmatrix} \rho \\ m \\ E \end{bmatrix}, \tag{A.12}$$

where $\langle u(v_1) \rangle := \sum_{k=1}^{N_v} u(\xi_k) w_k$ denotes the quadrature sum in the interval $[-R, R]$. The above nonlinear system is solved by the Newton–Raphson algorithm.

Time discretization. Due to the possibly stiff collision term, we use the IMEX-RK scheme [8, 32] for the time discretization. In particular, we employ the second-order IMEX-RK scheme proposed in [18], which is positivity preserving and asymptotic preserving (preserving the Euler limit without Δt resolving ε).

Specifically, we discretize (A.6) and (A.7) as

$$\begin{aligned}
\phi^{(i)} &= \phi^n - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} v_1 \partial_{x_1} \phi^{(j)} + \Delta t \sum_{j=1}^{i} a_{ij} \frac{1}{\varepsilon} \left(M_\phi^{(j)} - \phi^{(j)} \right), \quad i = 1, \ldots, \nu, \\
\psi^{(i)} &= \psi^n - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} v_1 \partial_{x_1} \psi^{(j)} + \Delta t \sum_{j=1}^{i} a_{ij} \frac{1}{\varepsilon} \left(M_\psi^{(j)} - \psi^{(j)} \right), \quad i = 1, \ldots, \nu, \\
\phi^{n+1} &= \phi^{(\nu)} + \alpha \Delta t^2 \frac{1}{\varepsilon^2} \left(M_\phi^{n+1} - \phi^{n+1} \right), \\
\psi^{n+1} &= \psi^{(\nu)} + \alpha \Delta t^2 \frac{1}{\varepsilon^2} \left(M_\psi^{n+1} - \psi^{n+1} \right),
\end{aligned} \tag{A.13}$$

where the values of the coefficients $\tilde{a}_{ij}, a_{ij}, \alpha$ are given in section 2.6.1 of [18]. To implement the above scheme explicitly, we first solve the moment system for $i = 1, \ldots, \nu$:

$$\begin{bmatrix} \langle \phi^{(i)} \rangle \\ \langle v_1 \phi^{(i)} \rangle \\ \langle \frac{1}{2} v_1^2 \phi^{(i)} + \psi^{(i)} \rangle \end{bmatrix} = \begin{bmatrix} \langle \phi^{n} \rangle \\ \langle v_1 \phi^{n} \rangle \\ \langle \frac{1}{2} v_1^2 \phi^{n} + \psi^{n} \rangle \end{bmatrix} - \Delta t \sum_{j=1}^{i-1} \tilde{a}_{ij} \begin{bmatrix} \langle v_1 \partial_{x_1} \phi^{(j)} \rangle \\ \langle v_1^2 \partial_{x_1} \phi^{(j)} \rangle \\ \langle \frac{1}{2} v_1^3 \partial_{x_1} \phi^{(j)} + v_1 \partial_{x_1} \psi^{(j)} \rangle \end{bmatrix}, \tag{A.14}$$
$$\begin{bmatrix} \langle \phi^{n+1} \rangle \\ \langle v_1 \phi^{n+1} \rangle \\ \langle \frac{1}{2} v_1^2 \phi^{n+1} + \psi^{n+1} \rangle \end{bmatrix} = \begin{bmatrix} \langle \phi^{(\nu)} \rangle \\ \langle v_1 \phi^{(\nu)} \rangle \\ \langle \frac{1}{2} v_1^2 \phi^{(\nu)} + \psi^{(\nu)} \rangle \end{bmatrix},$$

which is obtained by taking the moments of (A.13) and using (A.12). Hence we can obtain $\rho^{(i)}$, $m^{(i)}$ and $E^{(i)}$ first and use them to define $M_\phi^{(i)}$ and $M_\psi^{(i)}$. Finally we solve (A.13) to get $\phi^{(i)}$ and $\psi^{(i)}$.

Spatial discretization. In the physical space, we use the second-order MUSCL finite volume scheme [33].

Here we take the following first-order in time scheme for ϕ as an illustration (suppose it is evaluated at velocity point $v_1 = \xi_k$):

$$\frac{\phi_k^{n+1}(x_1) - \phi_k^n(x_1)}{\Delta t} + \xi_k \partial_{x_1} \phi_k^n(x_1) = \frac{1}{\varepsilon} \left((M_\phi)_k^{n+1}(x_1) - \phi_k^{n+1}(x_1) \right). \tag{A.15}$$

Suppose $x_1 \in [a, b]$ and $[a, b]$ is divided into N_x uniform cells with size $\Delta x = (b-a)/N_x$, where $a = x_{\frac{1}{2}}$, $b = x_{N_x+\frac{1}{2}}$. In the cell $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$, define the cell average as

$$\phi_{j,k}^n := \frac{1}{\Delta x} \int_{x_{j-\frac{1}{2}}}^{x_{j+\frac{1}{2}}} \phi_k^n(x_1)\, \mathrm{d}x_1. \tag{A.16}$$

Then integrating (A.15) over $[x_{j-\frac{1}{2}}, x_{j+\frac{1}{2}}]$ yields

$$\frac{\phi_{j,k}^{n+1} - \phi_{j,k}^n}{\Delta t} + \frac{F_{j+\frac{1}{2},k}^n - F_{j-\frac{1}{2},k}^n}{\Delta x} = \frac{1}{\varepsilon} \left((M_\phi)_{j,k}^{n+1} - \phi_{j,k}^{n+1} \right), \tag{A.17}$$

where $(M_\phi)_{j,k}^{n+1} := (M_\phi)_k^{n+1}(x_j)$. Note that we have replaced the cell average of $(M_\phi)_k^{n+1}$ by its point value at cell center x_j (the error introduced by this is $O(\Delta x^2)$ which does not destroy the overall order of the method). $F_{j+\frac{1}{2},k}^n$ is the flux at interface $x_{j+\frac{1}{2}}$ and is defined as

$$F_{j+\frac{1}{2},k}^n = \max(0, \xi_k) \phi_{l,j,k}^n + \min(0, \xi_k) \phi_{r,j+1,k}^n, \tag{A.18}$$

with the left interface and right interface values $\phi_{l,j,k}^n, \phi_{r,j,k}^n$ given by

$$\begin{cases} \phi_{l,j,k}^n = \phi_{j,k}^n + \dfrac{1}{2} \Delta x \sigma_{j,k}^n, \\ \phi_{r,j,k}^n = \phi_{j,k}^n - \dfrac{1}{2} \Delta x \sigma_{j,k}^n, \end{cases} \tag{A.19}$$

where $\sigma_{j,k}^n$ is the slope of the linear reconstruction and is chosen to be the MC limiter ($\theta = 2$):

$$\sigma_{j,k}^n = \mathrm{minmod} \left(\frac{\phi_{j+1,k}^n - \phi_{j-1,k}^n}{2\Delta x}, \theta \left(\frac{\phi_{j,k}^n - \phi_{j-1,k}^n}{\Delta x} \right), \theta \left(\frac{\phi_{j+1,k}^n - \phi_{j,k}^n}{\Delta x} \right) \right). \tag{A.20}$$

REFERENCES

[1] 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 (1954), p. 511.

[2] R. E. Caflisch, *Monte Carlo and quasi-Monte Carlo methods*, Acta Numer., 7 (1998), pp. 1–49.

[3] C. Cercignani, *The Boltzmann equation*, in The Boltzmann Equation and Its Applications, Springer, Cham, 1988, pp. 40–103.

[4] C. Cercignani, R. Illner, and M. Pulvirenti, *The Mathematical Theory of Dilute Gases*, Springer, Cham, 1994.

[5] S. Chapman, T. G. Cowling, and D. Burnett, *The Mathematical Theory of Non-Uniform Gases: An Account of the Kinetic Theory of Viscosity, Thermal Conduction and Diffusion in Gases*, Cambridge University Press, Cambridge, UK, 1990.

[6] C. Chu, *Kinetic-theoretic description of the formation of a shock wave*, Phys. Fluids, 8 (1965), pp. 12–22.

[7] B. Després, G. Poëtte, and D. Lucor, *Robust uncertainty propagation in systems of conservation laws with the entropy closure method*, in Uncertainty Quantification in Computational Fluid Dynamics, Springer, Cham, 2013, pp. 105–149.

[8] G. Dimarco and L. Pareschi, *Asymptotic preserving Implicit-Explicit Runge–Kutta methods for nonlinear kinetic equations*, SIAM J. Numer. Anal., 51 (2013), pp. 1064–1087.

[9] G. Dimarco and L. Pareschi, *Numerical methods for kinetic equations*, Acta Numer., 23 (2014), pp. 369–520.

[10] G. Dimarco and L. Pareschi, *Multiscale control variate methods for uncertainty quantification in kinetic equations*, J. Comput. Phys., 388 (2019), pp. 63–89.

[11] G. Dimarco and L. Pareschi, *Multiscale variance reduction methods based on multiple control variates for kinetic equations with uncertainties*, Multiscale Model. Simul., 18 (2020), pp. 351–382.

[12] E. Gabetta, L. Pareschi, and G. Toscani, *Relaxation schemes for nonlinear kinetic equations*, SIAM J. Numer. Anal., 34 (1997), pp. 2168–2194.

[13] R. G. Ghanem and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Springer-Verlag, New York, 1991.

[14] M. B. Giles, *Multilevel Monte Carlo path simulation*, Oper. Res., 56 (2008), pp. 607–617.

[15] A. Gorodetsky, G. Geraci, M. Eldred, and J. Jakeman, *A generalized approximate control variate framework for multifidelity uncertainty quantification*, J. Comput. Phys., 408 (2020), 109257.

[16] S. Heinrich, *Multilevel Monte Carlo methods*, in International Conference on Large-Scale Scientific Computing, Springer, Cham, 2001, pp. 58–67.

[17] J. Hu and S. Jin, *Uncertainty quantification for kinetic equations*, in Uncertainty Quantification for Hyperbolic and Kinetic Equations, S. Jin and L. Pareschi, eds., SEMA SIMAI Springer Ser. 14, Springer Cham, 2017, pp. 193–229.

[18] 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 (2018), pp. 942–973.

[19] S. Jin and L. Pareschi, eds., *Uncertainty Quantification for Hyperbolic and Kinetic Equations*, SEMA SIMAI Springer Ser., Springer, Cham, 2017.

[20] S. Krumscheid, F. Nobile, and M. Pisaroni, *Quantifying uncertain system outputs via the multilevel Monte Carlo method Part* I*: Central moment estimation*, J. Comput. Phys., 414 (2020), p. 109466.

[21] R. LeVeque, *Numerical Methods for Conservation Laws*, Birkhäuser, Basel, 1992.

[22] L. Liu and X. Zhu, *A bi-fidelity method for the multiscale Boltzmann equation with random parameters*, J. Comput. Phys., 402 (2020), 108914.

[23] L. Mieussens, *Discrete-velocity models and numerical schemes for the Boltzmann-BGK equation in plane and axisymmetric geometries*, J. Comput. Phys., 162 (2000), pp. 429–466.

[24] S. Mishra, N. H. Risebro, C. Schwab, and S. Tokareva, *Numerical solution of scalar conservation laws with random flux functions*, SIAM/ASA J. Uncertain. Quantifi., 4 (2016), pp. 552–591.

[25] S. Mishra, C. Schwab, and J. Sukys, *Multi-level Monte Carlo finite volumen methods for uncertainty quantification in nonlinear systems of balance laws*, in Uncertainty Quantification in Computational Fluid Dynamics, H. Bijl, D. Lucor, S. Mishra, and C. Schwab, eds., Lect. Notes Comput. Sci. Eng. 92, Springer, Cham, 2013, pp. 225–294.

[26] L. Pareschi, *An introduction to uncertainty quantification for kinetic equations and related problems*, in Trails in Kinetic Theory. Foundational Aspects and Numerical Methods, G. Albi, S. Merino-Aceituno, A. Nota, and M. Zanella, eds., SEMA SIMAI Springer Ser. 25, Springer, Cham, 2021.
[27] L. Pareschi and G. Toscani, *Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods*, Oxford University Press, Oxford, UK, 2013.
[28] L. Pareschi and M. Zanella, *Monte Carlo stochastic Galerkin methods for the Boltzmann equation with uncertainties: Space-homogeneous case*, J. Comput. Phys., 423 (2020), 109822.
[29] B. Peherstorfer, K. Willcox, and M. Gunzburger, *Survey of multifidelity methods in uncertainty propagation, inference, and optimization*, SIAM Rev., 60 (2018), pp. 550–591.
[30] B. Perthame, *Global existence to the BGK model of Boltzmann equation*, J. Differential Equations, 82 (1989), pp. 191–205.
[31] B. Perthame and M. Pulvirenti, *Weighted L^∞ bounds and uniqueness for the Boltzmann BGK model*, Arch. Ration. Mech. Anal., 125 (1993), pp. 289–295.
[32] G. Puppo and S. Pieraccini, *Implicit-explicit schemes for BGK kinetic equations*, J. Sci. Comput., 32 (2007), pp. 1–28.
[33] B. van Leer, *Towards the ultimate conservative difference scheme,* V. *A second order sequel to Godunov's method*, J. Comput. Phys., 32 (1979), pp. 101–136.
[34] C. Villani, *A review of mathematical topics in collisional kinetic theory*, in Handbook of Mathematical Fluid Dynamics, vol. 1, Elseiver, New York, 2002, pp. 3–8.
[35] D. Xiu, *Numerical Methods for Stochastic Computations*, Princeton University Press, Princeton, NJ, 2010.