Journal of Computational Physics 334 (2017) 182–206

Contents lists available at ScienceDirect

Journal of Computational Physics

www.elsevier.com/locate/jcp

An asymptotic-preserving stochastic Galerkin method for the radiative heat transfer equations with random inputs and diffusive scalings ☆

Shi Jin [a,b,*], Hanqing Lu [a]

[a] Department of Mathematics, University of Wisconsin-Madison, Madison, WI 53706, USA
[b] Institute of Natural Sciences, Department of Mathematics, MOE-LSEC and SHL-MAC, Shanghai Jiao Tong University, Shanghai 200240, China

ARTICLE INFO

Article history:
Received 13 June 2016
Received in revised form 16 December 2016
Accepted 18 December 2016
Available online 23 December 2016

Keywords:
Radiative heat transfer equations
Uncertainty quantification
Asymptotic preserving
Diffusion limit
Random inputs
Generalized polynomial chaos
Stochastic Galerkin method
Uniform stability
Spectral accuracy

ABSTRACT

In this paper, we develop an Asymptotic-Preserving (AP) stochastic Galerkin scheme for the radiative heat transfer equations with random inputs and diffusive scalings. In this problem the random inputs arise due to uncertainties in cross section, initial data or boundary data. We use the generalized polynomial chaos based stochastic Galerkin (gPC-SG) method, which is combined with the micro–macro decomposition based deterministic AP framework in order to handle efficiently the diffusive regime. For linearized problem we prove the regularity of the solution in the random space and consequently the spectral accuracy of the gPC-SG method. We also prove the uniform (in the mean free path) linear stability for the space-time discretizations. Several numerical tests are presented to show the efficiency and accuracy of proposed scheme, especially in the diffusive regime.

1. Introduction

The radiative heat transfer equations model the energy transfer in the form of electromagnetic radiation and its interaction with background material temperature. They have applications in a wide variety of subjects, including optics, astrophysics, inertial confinement fusion, and high temperature flow systems. Their mathematical models are well described in many references, see for examples [1,2,18,19].

The behavior of radiation transfer is greatly influenced by the properties of the background material. As a result, intensive interaction between the radiation and material with a small photon mean free path will lead to diffusive radiative behavior. To numerically simulate the system in the diffusive regime, numerical parameters—mesh sizes and time steps in particular—need to be fine enough—compared with the mean free time or path, which is often prohibitively expensive in real applications. In the last two decades, asymptotic-preserving (AP) methods [6] have been shown to be a competitive way to handle small or multiple scales in kinetic equations, see the review [7]. AP schemes were first developed for linear trans-

☆ This research was partially supported by NSF grants DMS-1522184 and DMS-1107291: RNMS KI-Net, by NSFC grant No. 91330203, and by the Office of the Vice Chancellor for Research and Graduate Education at the University of Wisconsin–Madison with funding from the Wisconsin Alumni Research Foundation.

* Corresponding author.
E-mail addresses: sjin@wisc.edu (S. Jin), hanqing@math.wisc.edu (H. Lu).

http://dx.doi.org/10.1016/j.jcp.2016.12.033
0021-9991/

port equations in diffusive regimes, see for stationary transport equations [15,14] and time-dependent transport equations [11,13,16]. AP schemes have also been developed for radiative heat transfer equations, see [12,21]. AP schemes preserve the asymptotic limits from the microscopic models to the macroscopic ones. They *automatically* become good macroscopic solvers when the small scales (mean free path here) are not numerically resolved.

In practical applications, the radiative heat transfer problems almost always involve uncertainties due to modeling and experimental errors. For example, the black body intensity is often modeled empirically and thus may contain uncertainty. Uncertainties could also arise from initial or boundary data, and other coefficients in the equations. The goal of this paper is to develop an efficient numerical method that allows one to conduct uncertainty quantification. Here one has to handle the difficulties due to uncertainties as well as possibly small or multiple scale characterized by the Knudesen number—the dimensionless mean free path.

Only in recent years there started to have works addressing these issues of uncertainty quantification for kinetic equations, in the framework of the stochastic asymptotic-preserving (sAP) schemes. The notion of sAP was first introduced by Jin, Xiu and Zhu in [11] for hyperbolic and kinetic equations with random coefficients and diffusive scalings. It requires a scheme (say the stochastic Galerkin scheme) for the kinetic equations, when the small parameter goes to zero while other numerical parameters are held fixed, becomes a good scheme (say again the stochastic Galerkin) for the limiting equation. See an extension of sAP scheme for the semiconductor Boltzmann equation [10], and a gPC-SG scheme for the Boltzmann equation with uncertainties [4]. The goal of this paper is to develop a sAP scheme for the radiative heat transfer equations with random inputs. Here the complication comes from the nonlinear coupling with a diffusion equation.

We use the generalized polynomial chaos expansion based stochastic Galerkin (gPC-SG) method [22] for the underlying equation. This allows us to use the deterministic solver—we use the micro–macro decomposition based AP scheme introduced by Klar and Schmeiser [12]—which can be easily shown to be sAP, thus the number of the gPC modes, or the degree of orthogonal polynomials used in the gPC approximation—does not depend on the mean free path. To efficiently handle the nonlinearity, we use a linearization procedure so the main part of the implicit time stepping has only the cost and complexity of solving a linear diffusion equation implicitly. Furthermore, for the linearized system we give a regularity result which leads to the proof of spectral convergence of the gPC-SG scheme in the random space. We also establish a linear stability—uniform in the mean free path—for the time and space discretizations. Several numerical tests with different random inputs in cross-section, initial and boundary data are conducted to verify the accuracy and asymptotic properties of the proposed scheme.

The SG method is used here since it allows us to easily extend the deterministic AP framework to the stochastic problems. It is also convenient to study numerical issues such as stability. Another popular method, the stochastic collocation method [23] can also be used for this problem, but its sAP property remains to be explored.

The rest of the paper is organized as follows. In Section 2 we introduce the radiative heat transfer equations with random input and their formal asymptotic limit in the diffusive regime. In Section 3, we derive our sAP scheme, present the fully discrete scheme and analyze its AP property. In Section 4, a uniform linear stability proof for the sAP method is presented. Section 5 focuses on the proof of regularity and spectral convergence in the random space for the linearized problem. Finally, we present several numerical examples with randomness in cross-section, initial and boundary data and compare the resulting quantities with the stochastic collocation method in Section 6.

2. The radiative heat transfer equations with random input and diffusion limit

Let $x \in D \subset \mathbb{R}^3$ be the space variable, $\Omega \in S^2$ be the direction variable, S^2 the unit sphere of $\mathbb{R}^3$, $z \in \mathbb{R}^d (d \geq 1)$ be the random variable and $t \in \mathbb{R}^+$ the time.

We denote by $I = I(x, \Omega, z, t)$ the radiative intensity and by $\theta(x, z, t)$ the material temperature. Introducing the Knudsen number ε, the radiative heat transfer equations in nondimensional form are

$$\varepsilon^2 M \partial_t I + \varepsilon \Omega \cdot \nabla_x I = B(\theta) - I \tag{1a}$$

$$\varepsilon^2 \partial_t \theta = \varepsilon^2 \Delta_x \theta - (B(\theta) - \langle I \rangle) \tag{1b}$$

with the total intensity

$$\langle I \rangle(x, z, t) = \frac{1}{|S|^2} \int_{S^2} I(x, \Omega, z, t) d\Omega, \tag{2}$$

and the black body intensity

$$B(\theta) = \sigma \theta^4, \tag{3}$$

where M is the Mach number ($= 1$ for this paper) and $\sigma = \sigma(x, z) > 0$ is the cross-section depending on the space variable and the random variable.

The initial conditions and reflection transmission boundary conditions are prescribed as following:

$$
\begin{aligned}
&I.C.\ \begin{cases} I(x,\Omega,z,0) = I_I(x,\Omega,z), \\ \theta(x,z,0) = \theta_I(x,z) \end{cases} \\
&B.C.\ \begin{cases} I(\hat{x},\Omega,z,t) = \alpha(n(\hat{x})\cdot\Omega) I(\hat{x},\Omega',z,t) + [1-\alpha(n(\hat{x})\cdot\Omega)] I_B(\hat{x},\Omega,z,t), \quad \text{for } n(\hat{x})\cdot\Omega < 0 \\ \theta(\hat{x},z,t) = \theta_B(\hat{x},z,t) \end{cases}
\end{aligned} \tag{4}
$$

where $\hat{x} \in \partial D$ with outward unit normal $n(\hat{x})$ and $\Omega' = \Omega - 2n(\hat{x})(n(\hat{x})\cdot\Omega)$ is the reflection of Ω on the tangent plane to ∂D. The reflectivity $\alpha,\ 0 \le \alpha \le 1$, depends on the incidence angle.

The random dimensionality d is determined by the number of random variable z used in the input $a(x,z)$ (which may come from cross section, initial or boundary data and etc.), which is typically modeled by a series involving linear combinations of z, i.e.,

$$
a(x,z) \approx \sum_{i=1}^{d} \hat{a}_i(x) z_i \tag{5}
$$

The most widely used such kind of approximation is the Karhunen–Loeve expansion (See [22]).

To approximate the solution, we use the gPC expansion via an orthogonal polynomials series. That is, for random variable $z \in \mathbb{R}^d$, one seeks

$$
\begin{aligned}
\theta(x,z,t) &\approx \theta_N(x,z,t) = \sum_{k=1}^{K} \hat{\theta}_k(x,t)\Phi_k(z), \\
g(x,\mu,z,t) &\approx g_N(x,\mu,z,t) = \sum_{k=1}^{K} \hat{g}_k(x,\mu,t)\Phi_k(z), \\
h(x,z,t) &\approx h_N(x,z,t) = \sum_{k=1}^{K} \hat{h}_k(x,t)\Phi_k(z)
\end{aligned} \tag{6}
$$

where $\left\{\Phi_k(z),\ 1 \le k \le K,\ K = \binom{d+N}{d}\right\}$ are from $\mathbb{P}_N^d$, the d-variate orthogonal polynomials of degree up to $N \ge 1$, and orthonormal

$$
\int \Phi_i(z)\Phi_j(z)\rho(z)dz = \delta_{ij}, \quad 1 \le i,j \le K = \dim(\mathbb{P}_N^d). \tag{7}
$$

Here $\rho(z)$ is the probability density function of z and δ_{ij} the Kronecker delta function (see [24]).

In the diffusion limit $\varepsilon \to 0^+$, for each z, system (1) can be formally approximated by a nonlinear diffusion equation with random input by the following asymptotic procedure: write

$$
\begin{aligned}
I &= I_0 + \varepsilon I_1 + \varepsilon^2 I_2 + \cdots \\
\theta &= \theta_0 + \varepsilon\theta_1 + \varepsilon^2\theta_2 + \cdots
\end{aligned}
$$

Substituting this ansatz into system (1) and collecting the terms of the same order in ε:

$$
O(1): I_0 = \langle I_0 \rangle = B(\theta_0) \tag{8a}
$$

$$
O(\varepsilon): \Omega\cdot\nabla_x I_0 = -I_1,\ \langle I_1 \rangle = 0 \tag{8b}
$$

By taking $\langle\cdot\rangle$ on (1a) and combining with (1b), one obtains the energy conservation equation

$$
\partial_t(\theta + \langle I \rangle) + \nabla_x \cdot \left(\frac{\langle \Omega I \rangle}{\varepsilon} - \nabla_x \theta \right) = 0, \tag{9}
$$

which, using (8a) gives the leading order

$$
\partial_t(\theta_0 + \langle I_0 \rangle) + \nabla_x \cdot (\langle \Omega I_1 \rangle - \nabla_x \theta_0) = 0.
$$

Using (8b) then yields

$$
\partial_t(\theta_0 + B(\theta_0)) = \nabla_x \cdot (\nabla_x \theta_0 + D\nabla_x B(\theta_0)), \tag{10}
$$

with $D = \langle \Omega \otimes \Omega \rangle = \frac{1}{3} I_d$ (where I_d denotes the 3 by 3 identity matrix), or

$$
(1 + B'(\theta_0))\partial_t\theta_0 = \nabla_x \cdot [(B'(\theta_0)/3 + 1)\nabla_x \theta_0]. \tag{11}
$$

When $(1-\rho)(I_B - B(\theta_B)) = 0$ almost everywhere on $\partial D \times S^-$, $\theta_I \in W^{2,\infty}(D)$ and $I_I = B(\theta_I) - \varepsilon\Omega\cdot\nabla_x B(\theta_I)$, there will be no boundary layer and initial layer. Then, when z is viewed as a parameter, for each realization of z, for every $\Delta t > 0$, the solution (I,θ) of system (1) with initial and boundary conditions (4) converges as $\varepsilon \to 0^+$ to $(B(\theta_0),\theta_0)$. The limiting temperature θ_0 is the unique solution of (11) with initial and boundary conditions:

$$\theta_0(x,z,0) = \theta_I(x,z), \qquad \theta_0(\hat{x},z,t) = \theta_B(\hat{x},z,t), \quad \text{for } \hat{x}\in\partial D \tag{12}$$

See [12] for a proof.

3. A stochastic AP scheme

For small values of ε, problem (1) is numerically stiff. We develop a stochastic Galerkin method based on the deterministic AP scheme proposed by Klar and Schmeiser (See [12]).

For simplicity, we consider the one-dimensional case $x\in[0,1]$ and define $\mu = \cos(\Omega\cdot x)$, $\mu\in[-1,1]$. Thus, the angular averaging is defined as:

$$\langle f\rangle = \frac{1}{2}\int_{-1}^{1} f(\mu)d\mu.$$

3.1. A micro–macro decomposition

As in [12], we rewrite the radiative intensity in the form

$$I(x,\mu,z,t) = B(\theta(x,z,t)) + \varepsilon g(x,\mu,z,t) + \varepsilon^2 h(x,z,t), \tag{13}$$

with $\langle g\rangle = 0$. This is a micro–macro decomposition of I into its mean value $\langle I\rangle = B(\theta) + \varepsilon^2 h$ and the remainder εg.

The equations (1a), (1b) are now rewritten as a system for g, h and θ by using (13), taking the angular average of (1a), and subtracting it from (1a):

$$\varepsilon^2\partial_t h + B'(\theta)(\partial_{xx}\theta + h) + \partial_x\langle\mu g\rangle = -h, \tag{14a}$$

$$\varepsilon^2\partial_t g + \mu\partial_x B(\theta) + \varepsilon\partial_x(\mu g - \langle\mu g\rangle) + \varepsilon^2\partial_x(\mu h) = -g, \tag{14b}$$

$$\partial_t\theta = \partial_{xx}\theta + h, \tag{14c}$$

with initial conditions

$$g(x,\mu,z,0) = \frac{1}{\varepsilon}[I_I(x,\mu,z) - \langle I_I\rangle(x,z)], \tag{15a}$$

$$h(x,z,0) = \frac{1}{\varepsilon^2}[\langle I_I\rangle(x,z) - B(\theta_I(x,z))], \tag{15b}$$

$$\theta(x,z,0) = \theta_I(x,z). \tag{15c}$$

Using the boundary conditions (4) with the condition $\langle g\rangle = 0$, evaluated at the boundary $\hat{x} = 0$ and $\hat{x} = 1$, we obtain the equation

$$\int_{-1}^{0}(1+\alpha)g d\mu + \int_{0}^{1}(1-\alpha)\left(\frac{I_B - B(\theta_B)}{\varepsilon} - \varepsilon h\right)d\mu = 0, \text{ at } \hat{x} = 0, \tag{16a}$$

$$\int_{0}^{1}(1+\alpha)g d\mu + \int_{-1}^{0}(1-\alpha)\left(\frac{I_B - B(\theta_B)}{\varepsilon} - \varepsilon h\right)d\mu = 0, \text{ at } \hat{x} = 1. \tag{16b}$$

As long as the boundary is not purely reflective, i.e.,

$$0 \le \alpha < 1$$

the value of $h(\hat{x},z,t)$ can be computed from (16) in terms of the outflow data $g(\hat{x},\mu,z,t)$ as following:

$$h = \frac{1}{\varepsilon}\int_{-1}^{0}\frac{1+\alpha}{1-\alpha}g d\mu + \frac{1}{\varepsilon^2}\int_{0}^{1}(I_B - B(\theta_B))d\mu, \text{ at } \hat{x} = 0, \tag{17a}$$

$$h = \frac{1}{\varepsilon}\int_{0}^{1}\frac{1+\alpha}{1-\alpha}g d\mu + \frac{1}{\varepsilon^2}\int_{-1}^{0}(I_B - B(\theta_B))d\mu, \text{ at } \hat{x} = 1. \tag{17b}$$

Then, boundary conditions for g and θ are given by

$$g(\mu) = \alpha g(-\mu) + (1-\alpha)\left(\frac{I_B - B(\theta_B)}{\varepsilon} - \varepsilon h\right), \ \mu > 0, \ \text{at } \hat{x} = 0, \tag{18a}$$

$$g(\mu) = \alpha g(-\mu) + (1-\alpha)\left(\frac{I_B - B(\theta_B)}{\varepsilon} - \varepsilon h\right), \ \mu < 0, \ \text{at } \hat{x} = 1, \tag{18b}$$

$$\theta(\hat{x}, z, t) = \theta_B(\hat{x}, z, t). \tag{19}$$

When $\alpha = 1$, the purely reflecting case, the reflecting boundary conditions are

$$g(\hat{x}, \mu) = g(\hat{x}, -\mu), \tag{20a}$$

$$\theta(\hat{x}, z, t) = \theta_B(\hat{x}, z, t), \tag{20b}$$

$$h(\hat{x}) \ \text{ solved from (14a).} \tag{20c}$$

3.2. The gPC-SG method for the micro–macro system and the limiting diffusion equation

We now derive a stochastic AP scheme for the micro–macro system (14). One inserts the approximation θ_N, g_N and h_N in (6) into the governing equations and enforces the residue to be orthogonal to the polynomial space spanned by $\{\Phi_1, \cdots, \Phi_K\}$. Thus, we obtain a set of deterministic equations for the expansion coefficients $\{\hat{\theta}_k\}, \{\hat{g}_k\}$ and $\{\hat{h}_k\}$.

Denote

$$\begin{aligned} &\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \cdots, \hat{\theta}_K)^T, \ \ \hat{\mathbf{g}} = (\hat{g}_1, \cdots, \hat{g}_K)^T, \ \ \hat{\mathbf{h}} = (\hat{h}_1, \cdots, \hat{h}_K)^T, \\ &\mathbf{C}(x,t) = (c_{ij}(x,t))_{0\le i,j\le K} \ \text{ with} \\ &c_{ij}(x,t) = \int \theta_N^3 \sigma(x,z)\Phi_i(z)\Phi_j(z)\rho(z)dz. \end{aligned} \tag{21}$$

Then

$$\varepsilon^2 \partial_t \hat{\mathbf{h}} + 4\mathbf{C}(x,t)(\partial_{xx}\hat{\boldsymbol{\theta}} + \hat{\mathbf{h}}) + \partial_x\langle\mu\hat{\mathbf{g}}\rangle = -\hat{\mathbf{h}}, \tag{22a}$$

$$\varepsilon^2 \partial_t \hat{\mathbf{g}} + 4\mu\mathbf{C}(x,t)\partial_x\hat{\boldsymbol{\theta}} + \varepsilon\partial_x(\mu\hat{\mathbf{g}} - \langle\mu\hat{\mathbf{g}}\rangle) + \varepsilon^2\partial_x(\mu\hat{\mathbf{h}}) = -\hat{\mathbf{g}}, \tag{22b}$$

$$\partial_t\hat{\boldsymbol{\theta}} = \partial_{xx}\hat{\boldsymbol{\theta}} + \hat{\mathbf{h}}. \tag{22c}$$

Correspondingly, the initial and boundary data can be projected by the gPC-SG method,

$$\hat{\boldsymbol{\theta}}_I = (\hat{\theta}_{I_1}, \cdots, \hat{\theta}_{I_K})^T, \text{where } \theta_{I_k}(x) = \int \theta_I(x,z)\Phi_k(z)\rho(z)dz \tag{23}$$

and similarly for $\hat{\mathbf{I}}_I$, $\hat{\boldsymbol{\theta}}_B$ and $\hat{\mathbf{I}}_B$.

Thus, (15), (17), (18), (19) or (20) give the initial and boundary conditions for (22) as following:

$$\hat{\mathbf{g}}(x,\mu,0) = \frac{1}{\varepsilon}[\hat{\mathbf{I}}_I(x,\mu) - \langle\hat{\mathbf{I}}_I\rangle(x)], \tag{24a}$$

$$\hat{\mathbf{h}}(x,0) = \frac{1}{\varepsilon^2}[\langle\hat{\mathbf{I}}_I\rangle(x) - B(\hat{\boldsymbol{\theta}}_I(x))], \tag{24b}$$

$$\hat{\boldsymbol{\theta}}(x,0) = \hat{\boldsymbol{\theta}}_I(x). \tag{24c}$$

For $0 \le \alpha < 1$,

$$\hat{\mathbf{h}} = \frac{1}{\varepsilon}\int_{-1}^{0}\frac{1+\alpha}{1-\alpha}\hat{\mathbf{g}}d\mu + \frac{1}{\varepsilon^2}\int_0^1(\hat{\mathbf{I}}_B - B(\hat{\boldsymbol{\theta}}_B))d\mu, \ \text{at } \hat{x} = 0, \tag{25a}$$

$$\hat{\mathbf{h}} = \frac{1}{\varepsilon}\int_{0}^{1}\frac{1+\alpha}{1-\alpha}\hat{\mathbf{g}}d\mu + \frac{1}{\varepsilon^2}\int_{-1}^0(\hat{\mathbf{I}}_B - B(\hat{\boldsymbol{\theta}}_B))d\mu, \ \text{at } \hat{x} = 1, \tag{25b}$$

$$\hat{\mathbf{g}}(\mu) = \alpha\hat{\mathbf{g}}(-\mu) + (1-\alpha)\left(\frac{\hat{\mathbf{I}}_B - B(\hat{\boldsymbol{\theta}}_B)}{\varepsilon} - \varepsilon\hat{\mathbf{h}}\right), \ \mu > 0, \ \text{at } \hat{x} = 0, \tag{25c}$$

$$\hat{\mathbf{g}}(\mu) = \alpha\hat{\mathbf{g}}(-\mu) + (1-\alpha)\left(\frac{\hat{\mathbf{I}}_B - B(\hat{\boldsymbol{\theta}}_B)}{\varepsilon} - \varepsilon\hat{\mathbf{h}}\right),\ \mu < 0,\ \text{at}\ \hat{x} = 1, \tag{25d}$$

$$\hat{\boldsymbol{\theta}}(\hat{x}, t) = \hat{\boldsymbol{\theta}}_B(x, t),\ \hat{x} = 0, 1. \tag{25e}$$

For $\alpha = 1$ at $\hat{x} = 0, 1$,

$$\hat{\mathbf{g}}(\hat{x}, \mu) = \hat{\mathbf{g}}(\hat{x}, -\mu), \tag{26a}$$

$$\hat{\boldsymbol{\theta}}(\hat{x}, t) = \hat{\boldsymbol{\theta}}_B(x, t), \tag{26b}$$

$$\hat{\mathbf{h}}(\hat{x})\ \text{ solved from (22a).} \tag{26c}$$

Similarly as in section 3.2, we can obtain a set of deterministic equations by using K term truncated gPC-SG for θ for the limiting diffusion equation (11):

$$(\mathbf{I} + 4\mathbf{C}(x, t))\partial_t\hat{\boldsymbol{\theta}} = \partial_x\left[\left(\mathbf{I} + \frac{4}{3}\mathbf{C}(x, t)\right)\partial_x\hat{\boldsymbol{\theta}}\right] \tag{27}$$

where $\mathbf{C}(x, t), \mathbf{I}$ and $\hat{\boldsymbol{\theta}}$ are the same as in (21).

3.3. The time discretization

Introduce a time step $\Delta t > 0$ and discrete time $t^n = n\Delta t$. We now employ the semi-implicit time discretization in [12] to the gPC-SG system (22), where backward differences are used for the zeroth order terms (for $\varepsilon \ll 1$) and forward differences for higher order terms:

$$\frac{\varepsilon^2}{\Delta t}(\hat{\mathbf{h}}^{n+1} - \hat{\mathbf{h}}^n) + 4\mathbf{C}^n(\partial_{xx}\hat{\boldsymbol{\theta}}^{n+1} + \hat{\mathbf{h}}^{n+1}) + \partial_x\langle\mu\hat{\mathbf{g}}^{n+1}\rangle = -\hat{\mathbf{h}}^{n+1}, \tag{28a}$$

$$\varepsilon^2\frac{\hat{\mathbf{g}}^{n+1} - \hat{\mathbf{g}}^n}{\Delta t} + 4\mu\mathbf{C}^n\partial_x\hat{\boldsymbol{\theta}}^{n+1} + \varepsilon\partial_x(\mu\hat{\mathbf{g}}^n - \langle\mu\hat{\mathbf{g}}^n\rangle) + \varepsilon^2\partial_x(\mu\hat{\mathbf{h}}^n) = -\hat{\mathbf{g}}^{n+1}, \tag{28b}$$

$$\frac{\hat{\boldsymbol{\theta}}^{n+1} - \hat{\boldsymbol{\theta}}^n}{\Delta t} = \partial_{xx}\hat{\boldsymbol{\theta}}^{n+1} + \hat{\mathbf{h}}^{n+1}. \tag{28c}$$

From equation (28b), $\hat{\mathbf{g}}^{n+1}$ can be expressed in terms of $\hat{\boldsymbol{\theta}}^{n+1}$. Then one inserts $\hat{\mathbf{g}}^{n+1}$ in terms of $\hat{\boldsymbol{\theta}}^{n+1}$ back to equation (28a) and gets $\hat{\mathbf{h}}^{n+1}$ in terms of $\hat{\boldsymbol{\theta}}^{n+1}$. Using the result in (28c), setting

$$\kappa = 1 + \frac{\varepsilon^2}{\Delta t} \tag{29}$$

and letting $\mathbf{I}$ be the $K \times K$ identity matrix, give

$$\hat{\mathbf{g}}^{n+1} = -\frac{1}{\kappa}4\mu\mathbf{C}^n\partial_x\hat{\boldsymbol{\theta}}^{n+1} + \varepsilon\hat{\mathbf{p}}^n, \tag{30a}$$

$$\text{with}\ \hat{\mathbf{p}}^n = \frac{1}{\kappa}\left[\frac{\varepsilon}{\Delta t}\hat{\mathbf{g}}^n - \partial_x(\mu\hat{\mathbf{g}}^n - \langle\mu\hat{\mathbf{g}}^n\rangle) - \varepsilon\mu\partial_x\hat{\mathbf{h}}^n\right], \tag{30b}$$

$$\hat{\mathbf{h}}^{n+1} = [\kappa\mathbf{I} + 4\mathbf{C}^n]^{-1}\left[\frac{1}{\kappa}\partial_x\left(\frac{4}{3}\mathbf{C}^n\partial_x\hat{\boldsymbol{\theta}}^{n+1}\right) - 4\mathbf{C}^n\partial_{xx}\hat{\boldsymbol{\theta}}^{n+1} + \varepsilon\hat{\mathbf{q}}^n\right], \tag{30c}$$

$$\text{with}\ \hat{\mathbf{q}}^n = \frac{\varepsilon}{\Delta t}\hat{\mathbf{h}}^n - \partial_x\langle\mu\hat{\mathbf{p}}^n\rangle, \tag{30d}$$

$$(\kappa\mathbf{I} + 4\mathbf{C}^n)\frac{\hat{\boldsymbol{\theta}}^{n+1} - \hat{\boldsymbol{\theta}}^n}{\Delta t} = \partial_x\left[(\kappa\mathbf{I} + \frac{1}{3\kappa}4\mathbf{C}^n)\partial_x\hat{\boldsymbol{\theta}}^{n+1}\right] + \varepsilon\hat{\mathbf{q}}^n. \tag{30e}$$

Thus, an elliptic equation for $\hat{\boldsymbol{\theta}}^{n+1}$ needs to be first solved, and $\hat{\mathbf{g}}^{n+1}$ and $\hat{\mathbf{h}}^{n+1}$ can be obtained subsequently.

3.4. Positivity of the diffusion coefficient matrices

One problem with the gPC truncation (6) is the loss of positivity for θ_N. Then one may be concerned with the positivity of the diffusion coefficient matrices in (27) and (30e). Below we prove, under suitably mild assumptions, these matrices are positive definite.

Lemma 3.1. *Assume*

$$0 < \sigma_m \le \sigma(x,z) \le \sigma_M, \qquad \forall\, x, z. \tag{31}$$

Suppose $\theta_N^3(x,t,z) \ge -\frac{3}{4\sigma_M}$ *for all* x, t, z, *then the matrix* $\mathbf{I} + \frac{4}{3}\mathbf{C}(x,t)$ *in (27) is positive definite.*

Proof. Let $\mathbf{b} = (\hat{b}_1, \cdots, \hat{b}_K)^T$ be an arbitrary non-zero real vector, and $b(z) = \sum_{j=1}^K \hat{b}_j \Phi_j(z)$ be a random variable constructed by the $\mathbf{b}$ vector. By using the definition of $c_{i,j}(x,t)$ in (21) and the assumption (31), we have for any $x \in D, t > 0$,

$$\begin{aligned}
\mathbf{b}^T \left(\mathbf{I} + \frac{4}{3}\mathbf{C}(x,t)\right)\mathbf{b} &= \sum_{i=1}^K \sum_{j=1}^K \hat{b}_i \left(\delta_{ij} + \frac{4}{3} c_{ij}(x,t)\right) \hat{b}_j \\
&= \sum_{i=1}^K \sum_{j=1}^K \delta_{ij} \hat{b}_i \hat{b}_j + \frac{4}{3} \sum_{i=1}^K \sum_{j=1}^K \hat{b}_i \hat{b}_j \int \theta_N^3 \sigma(x,z) \Phi_i(z) \Phi_j(z) \rho(z) dz \\
&= \sum_{i=1}^K \hat{b}_i^2 + \frac{4}{3} \int \theta_N^3 \sigma(x,z) \left(\sum_{i=1}^K \sum_{j=1}^K \hat{b}_i \hat{b}_j \Phi_i(z) \Phi_j(z) \right) \rho(z) dz \\
&= \sum_{i=1}^K \hat{b}_i^2 + \frac{4}{3} \int \theta_N^3 \sigma(x,z) b(z)^2 \rho(z) dz.
\end{aligned}$$

Now if $\theta_N^3 \ge -\frac{3}{4\sigma_M}$,

$$\begin{aligned}
\mathbf{b}^T (\mathbf{I} + \frac{4}{3}\mathbf{C}(x,t))\mathbf{b} &\ge \sum_{i=1}^K \hat{b}_i^2 - \int b(z)^2 \rho(z) dz \\
&= \sum_{i=1}^K \hat{b}_i^2 - \int \left(\sum_{i=1}^K \sum_{j=1}^K \hat{b}_i \hat{b}_j \Phi_i(z) \Phi_j(z) \right) \rho(z) dz \\
&= \sum_{i=1}^K \hat{b}_i^2 - \sum_{i=1}^K \sum_{j=1}^K \hat{b}_i \hat{b}_j \left(\int \Phi_i(z) \Phi_j(z) \rho(z) dz \right) \\
&= \sum_{i=1}^K \hat{b}_i^2 - \sum_{i=1}^K \sum_{i=1}^K \hat{b}_i \hat{b}_j \delta_{ij} = 0.
\end{aligned}$$

To conclude, $(\mathbf{I} + \frac{4}{3}\mathbf{C}(x,t))$ is positive definite, so is the matrix $\kappa\mathbf{I} + \frac{1}{3\kappa} 4\mathbf{C}^n$ in (30e). □

Remark 3.1. Although we may not know the positivity of θ_N, due to the spectral accuracy, $|\theta_N|$ is close to zero when θ_N is negative. Thus the condition $\theta_N^3 \ge -\frac{3}{4\sigma_M}$ is a very reasonable assumption for given $\sigma_M = O(1)$.

3.5. The fully discrete scheme

We discretize space using staggered grids with $\Delta x = 1/i_{\max}$:

$$x_i = i\Delta x,\ i = 0,\ \cdots, i_{\max},$$

and

$$x_{i-1/2} = (i - 1/2)\Delta x,\ i = 0,\ \cdots, i_{\max} + 1.$$

The variable $\hat{\boldsymbol{\theta}}$ and $\hat{\mathbf{h}}$ are defined at the grid points x_i, and $\hat{\mathbf{g}}$ is defined at the points $x_{i-1/2}$. The approximations at time t^n are denoted by $\hat{\boldsymbol{\theta}}_i^n$, $\hat{\mathbf{h}}_i^n$ and $\hat{\mathbf{g}}_{i-1/2}^n$ respectively. Let

$$\mathbf{C}_{i+1/2}^n = \frac{1}{2}(\mathbf{C}_{i+1}^n + \mathbf{C}_i^n)$$

the space-discretized version of system (30) reads

$$\hat{\mathbf{g}}_{i-1/2}^{n+1} = -\frac{1}{\kappa} 4\mathbf{C}_{i-1/2}^{n} \mu \frac{\hat{\boldsymbol{\theta}}_i^{n+1} - \hat{\boldsymbol{\theta}}_i^{n+1}}{\Delta x} + \varepsilon \hat{\mathbf{p}}_{i-1/2}^{n},\ i = 1, \cdots, i_{\max}, \tag{32a}$$

$$\text{with } \hat{\mathbf{p}}_{i-1/2}^{n} = \frac{1}{\kappa}\Big[\frac{\varepsilon}{\Delta t}\hat{\mathbf{g}}_{i-1/2}^{n} - \frac{\mu}{\Delta x}(\hat{\mathbf{g}}_i^{n} - \hat{\mathbf{g}}_{i-1}^{n}) + \langle \frac{\mu}{\Delta x}(\hat{\mathbf{g}}_i^{n} - \hat{\mathbf{g}}_{i-1}^{n})\rangle - \frac{\varepsilon\mu}{\Delta x}(\hat{\mathbf{h}}_i^{n} - \hat{\mathbf{h}}_{i-1}^{n})\Big],\ i = 1, \cdots, i_{\max}, \tag{32b}$$

$$\hat{\mathbf{h}}_i^{n+1} = (\kappa \mathbf{I} + 4\mathbf{C}_i^{n})^{-1}\Big[-\frac{4}{(\Delta x)^2}\mathbf{C}_i^{n}(\hat{\boldsymbol{\theta}}_{i+1}^{n+1} - 2\hat{\boldsymbol{\theta}}_i^{n+1} + \hat{\boldsymbol{\theta}}_{i-1}^{n+1}) + \varepsilon \hat{\mathbf{q}}_i^{n}$$
$$+ \frac{4}{3\kappa\,\Delta x}\Big(\mathbf{C}_{i+1/2}^{n}\frac{\hat{\boldsymbol{\theta}}_{i+1}^{n+1} - \hat{\boldsymbol{\theta}}_i^{n+1}}{\Delta x} - \mathbf{C}_{i-1/2}^{n}\frac{\hat{\boldsymbol{\theta}}_i^{n+1} - \hat{\boldsymbol{\theta}}_{i-1}^{n+1}}{\Delta x}\Big)\Big],\ i = 1, \cdots, i_{\max} - 1, \tag{32c}$$

$$\text{with } \hat{\mathbf{q}}_i^{n} = \frac{\varepsilon}{\Delta t}\hat{\mathbf{h}}_i^{n} - \langle \frac{\mu}{\Delta x}(\hat{\mathbf{p}}_{i+1/2}^{n} - \hat{\mathbf{p}}_{i-1/2}^{n})\rangle,\ i = 1, \cdots, i_{\max} - 1, \tag{32d}$$

$$(\kappa \mathbf{I} + 4\mathbf{C}_i^{n})\frac{1}{\Delta t}(\hat{\boldsymbol{\theta}}_i^{n+1} - \hat{\boldsymbol{\theta}}_i^{n}) = \frac{1}{\Delta x}\Big[(\kappa \mathbf{I} + \frac{4}{3\kappa}\mathbf{C}_{i+1/2}^{n})\frac{\hat{\boldsymbol{\theta}}_{i+1}^{n+1} - \hat{\boldsymbol{\theta}}_i^{n+1}}{\Delta x} - (\kappa \mathbf{I} + \frac{4}{3\kappa}\mathbf{C}_{i-1/2}^{n})\frac{\hat{\boldsymbol{\theta}}_i^{n+1} - \hat{\boldsymbol{\theta}}_{i-1}^{n+1}}{\Delta x}\Big] + \varepsilon \hat{\mathbf{q}}_i^{n},$$
$$i = 1, \cdots, i_{\max} - 1. \tag{32e}$$

The free streaming operator in (32b) is discretized in an upwinding fashion:

$$\hat{\mathbf{g}}_i^{n} = \begin{cases} \hat{\mathbf{g}}_{i-1/2}^{n} \text{for } \mu > 0, \\ \hat{\mathbf{g}}_{i+1/2}^{n} \text{for } \mu < 0. \end{cases}$$

A $K(i_{\max}+1) \times K(i_{\max}+1)$ block diagonal system resulting from the implicit discretization of the parabolic equation (32e) (the same as in the diffusion equation) needs to be solved there. There are many fast algorithms for the inversion (see [5]).

It remains to discretize the boundary conditions. $\hat{\mathbf{g}}_{-1/2}^{n+1}$ for $\mu > 0$ ($\hat{\mathbf{g}}_{i_{\max}+1/2}^{n+1}$ for $\mu < 0$) is determined from the boundary conditions (25) and (26) in the following way:

In (25a) and (25b), the outflow data for $\hat{\mathbf{g}}$ are approximated by $\hat{\mathbf{g}}_{1/2}^{n+1}, \mu < 0$ ($\hat{\mathbf{g}}_{i_{\max}-1/2}^{n+1}, \mu > 0$). Then $\hat{\mathbf{h}}_0^{n+1}$ and $\hat{\mathbf{h}}_{i_{\max}}^{n+1}$ are determined by (if the respective boundary point is not purely reflective, i.e. $0 \le \alpha < 1$):

$$\hat{\mathbf{h}}_0^{n+1} = \frac{1}{\varepsilon}\int_{-1}^{0}\frac{1+\alpha}{1-\alpha}\hat{\mathbf{g}}_{1/2}^{n+1}d\mu + \frac{1}{\varepsilon^2}\int_{0}^{1}(\hat{\mathbf{I}}_0^{n+1} - B(\hat{\boldsymbol{\theta}}_0^{n+1}))d\mu, \tag{33a}$$
$$\text{with } \hat{\mathbf{I}}_0^{n+1} = \hat{\mathbf{I}}_B(0),$$

$$\hat{\mathbf{h}}_{i_{\max}}^{n+1} = \frac{1}{\varepsilon}\int_{0}^{1}\frac{1+\alpha}{1-\alpha}\hat{\mathbf{g}}_{i_{\max}-1/2}^{n+1}d\mu + \frac{1}{\varepsilon^2}\int_{-1}^{0}(\hat{\mathbf{I}}_{i_{\max}}^{n+1} - B(\hat{\boldsymbol{\theta}}_{i_{\max}}^{n+1}))d\mu, \tag{33b}$$
$$\text{with } \hat{\mathbf{I}}_{i_{\max}}^{n+1} = \hat{\mathbf{I}}_B(1).$$

This value is then used in the right hand side of (25c) and (25d), where the outflow data on the right hand side are again approximated like above and the left hand side is replaced by $(\hat{\mathbf{g}}_{-1/2}^{n+1} + \hat{\mathbf{g}}_{1/2}^{n+1})/2$ ($(\hat{\mathbf{g}}_{i_{\max}-1/2}^{n+1} + \hat{\mathbf{g}}_{i_{\max}+1/2}^{n+1})/2$):

$$\frac{\hat{\mathbf{g}}_{-1/2}^{n+1}(\mu) + \hat{\mathbf{g}}_{1/2}^{n+1}(\mu)}{2} = \alpha\hat{\mathbf{g}}_{1/2}^{n+1}(-\mu) + \frac{1-\alpha}{\varepsilon}(\hat{\mathbf{I}}_0^{n+1}(\mu) - B(\hat{\boldsymbol{\theta}}_0^{n+1})) - (1-\alpha)\varepsilon\hat{\mathbf{h}}_0^{n+1},\ \mu > 0, \tag{34a}$$

$$\frac{\hat{\mathbf{g}}_{i_{\max}-1/2}^{n+1}(\mu) + \hat{\mathbf{g}}_{i_{\max}+1/2}^{n+1}(\mu)}{2} = \alpha\hat{\mathbf{g}}_{i_{\max}-1/2}^{n+1}(-\mu) + \frac{1-\alpha}{\varepsilon}(\hat{\mathbf{I}}_{i_{\max}}^{n+1}(\mu) - B(\hat{\boldsymbol{\theta}}_{i_{\max}}^{n+1})) - (1-\alpha)\varepsilon\hat{\mathbf{h}}_{i_{\max}}^{n+1},\ \mu < 0. \tag{34b}$$

The boundary for $\hat{\boldsymbol{\theta}}$ is given by (25e):

$$\hat{\boldsymbol{\theta}}_0^{n+1} = \hat{\boldsymbol{\theta}}_B(0, t^{n+1}),\quad \hat{\boldsymbol{\theta}}_{i_{\max}}^{n+1} = \hat{\boldsymbol{\theta}}_B(1, t^{n+1}). \tag{35}$$

For $\alpha = 1$,

$$\hat{\mathbf{g}}_{-\frac{1}{2}}^{n+1}(\mu) = \hat{\mathbf{g}}_{-\frac{1}{2}}^{n+1}(-\mu),\ \mu > 0;\ \ \hat{\mathbf{g}}_{i_{\max}+\frac{1}{2}}^{n+1}(\mu) = \hat{\mathbf{g}}_{i_{\max}+\frac{1}{2}}^{n+1}(-\mu),\ \mu < 0, \tag{36a}$$

$$\hat{\boldsymbol{\theta}}_0^{n+1} = \hat{\boldsymbol{\theta}}_B(0, t^{n+1}),\quad \hat{\boldsymbol{\theta}}_{i_{\max}}^{n+1} = \hat{\boldsymbol{\theta}}_B(1, t^{n+1}), \tag{36b}$$

$$\hat{\mathbf{h}}_0^{n+1} = -(\kappa \mathbf{I} + 4\mathbf{C}_0^{n})^{-1}\left[4\mathbf{C}_0^{n}\frac{-\hat{\boldsymbol{\theta}}_3^{n+1} + 4\hat{\boldsymbol{\theta}}_2^{n+1} - 5\hat{\boldsymbol{\theta}}_1^{n+1} + 2\hat{\boldsymbol{\theta}}_0^{n+1}}{(\Delta x)^2} + \langle\frac{\mu}{\Delta x}(\hat{\mathbf{g}}_{1/2}^{n} - \hat{\mathbf{g}}_{-1/2}^{n})\rangle - \frac{\varepsilon^2}{\Delta t}\hat{\mathbf{h}}_0^{n}\right], \tag{36c}$$

$$\hat{\mathbf{h}}^{n+1}_{\text{imax}} = -(\kappa\mathbf{I} + 4\mathbf{C}^n_{\text{imax}})^{-1}\left[4\mathbf{C}^n_{\text{imax}}\frac{-\hat{\boldsymbol{\theta}}^{n+1}_{\text{imax}-3} + 4\hat{\boldsymbol{\theta}}^{n+1}_{\text{imax}-2} - 5\hat{\boldsymbol{\theta}}^{n+1}_{\text{imax}-1} + 2\hat{\boldsymbol{\theta}}^{n+1}_{\text{imax}}}{(\Delta x)^2}\right.$$
$$\left. + \langle\frac{\mu}{\Delta x}(\hat{\mathbf{g}}^n_{\text{imax}+1/2} - \hat{\mathbf{g}}^n_{\text{imax}-1/2})\rangle - \frac{\varepsilon^2}{\Delta t}\hat{\mathbf{h}}^n_{\text{imax}}\right]. \tag{36d}$$

The values of $\hat{\mathbf{g}}^{n+1}_{-1/2}, \mu > 0$, and $\hat{\mathbf{g}}^{n+1}_{i_{\max}+1/2}, \mu < 0$, computed from these equation (34) or (36), are needed in the next time step.

3.6. The AP property

When $\varepsilon \to 0$, one gets $\kappa = 1$. Thus, (32e) becomes

$$(\mathbf{I} + 4\mathbf{C}^n_i)\frac{1}{\Delta t}(\hat{\boldsymbol{\theta}}^{n+1}_i - \hat{\boldsymbol{\theta}}^n_i) = \frac{1}{\Delta x}\left[(\mathbf{I} + \frac{4}{3}\mathbf{C}^n_{i+1/2})\frac{\hat{\boldsymbol{\theta}}^{n+1}_{i+1} - \hat{\boldsymbol{\theta}}^{n+1}_i}{\Delta x} - (\mathbf{I} + \frac{4}{3}\mathbf{C}^n_{i-1/2})\frac{\hat{\boldsymbol{\theta}}^{n+1}_i - \hat{\boldsymbol{\theta}}^{n+1}_{i-1}}{\Delta x}\right], \tag{37}$$

which is a fully discretized scheme for (27), using implicit time discretization and center difference space discretization. Thus the fully discrete scheme is stochastic-AP (sAP), in the sense that the limiting scheme (37), as the asymptotic limiting solution of the transport scheme (32) when $\varepsilon \to 0$, becomes the stochastic Galerkin approximation of the diffusion equation (11) (see [11] for formal definition of sAP).

3.7. The velocity discretization

For velocity discretization, we employ the discrete-ordinate method. The discrete velocity points are chosen to be the Legendre–Gauss quadrature points. Then the integral in (32b) and (32d) can be computed by using Gauss quadrature rule. This kind of discretization maintains the AP property under a very mild condition. Namely, the quadrature should be exact for quadratic polynomials, see [8].

4. A uniform stability analysis

4.1. Some notations and useful identities

For every grid function $\mathbf{u} = (\mathbf{u}_i)$ we define:

$$\|\mathbf{u}\|^2 = \sum_i \mathbf{u}_i^T\mathbf{u}_i\Delta x. \tag{38}$$

For every velocity dependent grid function $\mu \in [-1, 1] \to \boldsymbol{\phi}(\mu) = (\boldsymbol{\phi}_{i+1/2}(\mu))$, we define:

$$|||\boldsymbol{\phi}||| = \sum_i \langle\boldsymbol{\phi}^T_{i+1/2}\boldsymbol{\phi}_{i+1/2}\rangle\Delta x. \tag{39}$$

Now we give some notations for the finite difference operators:

$$D^-\boldsymbol{\phi}_{i+1/2} = \frac{\boldsymbol{\phi}_{i+1/2} - \boldsymbol{\phi}_{i-1/2}}{\Delta x}, \tag{40a}$$
$$D^+\boldsymbol{\phi}_{i+1/2} = \frac{\boldsymbol{\phi}_{i+3/2} - \boldsymbol{\phi}_{i+1/2}}{\Delta x}, \tag{40b}$$
$$D^c\boldsymbol{\phi}_{i+1/2} = \frac{\boldsymbol{\phi}_{i+3/2} - \boldsymbol{\phi}_{i-1/2}}{2\Delta x}, \tag{40c}$$
$$D^0\boldsymbol{\phi}_i = \frac{\boldsymbol{\phi}_{i+1/2} - \boldsymbol{\phi}_{i-1/2}}{\Delta x}(= D^-\boldsymbol{\phi}_{i+1/2}), \tag{40d}$$
$$\delta^0\mathbf{u}_{i+1/2} = \frac{\mathbf{u}_{i+1} - \mathbf{u}_i}{\Delta x}, \tag{40e}$$
$$\Delta^c\boldsymbol{\phi}_i = \frac{\boldsymbol{\phi}_{i+1} - 2\boldsymbol{\phi}_i + \boldsymbol{\phi}_{i-1}}{(\Delta x)^2}. \tag{40f}$$

We recall some useful formulas derived in [17]:

$$(\mu^+D^- + \mu^-D^+)\boldsymbol{\phi}_{i+1/2} = \mu D^c\boldsymbol{\phi}_{i+1/2} - \frac{\Delta x}{2}|\mu|D^-D^+\boldsymbol{\phi}_{i+1/2}, \tag{41a}$$

$$\sum_i (D^+\boldsymbol{\phi}_{i+1/2})^T(D^+\boldsymbol{\phi}_{i+1/2})\Delta x \le \frac{4}{(\Delta x)^2}\sum_i(\boldsymbol{\phi}_{i+1/2})^T\boldsymbol{\phi}_{i+1/2}\Delta x, \tag{41b}$$

$$|\sum_i \langle(\mu^+D^+ + \mu^-D^-)(\boldsymbol{\psi}_{i+1/2})^T\boldsymbol{\phi}_{i+1/2}\rangle\Delta x| \le \alpha\|\boldsymbol{\phi}\|^2 + \frac{1}{4\alpha}\||\mu|D^+\boldsymbol{\psi}\|^2, \tag{41c}$$

$$\sum_i(\mathbf{u}_i)^T D^0\boldsymbol{\phi}_i\Delta x = -\sum_i(\delta^0\mathbf{u}_{i+1/2})^T\boldsymbol{\phi}_{i+1/2}\Delta x, \tag{41d}$$

$$\sum_i(\boldsymbol{\psi}_{i+1/2})^T D^-\boldsymbol{\phi}_{i+1/2}\Delta x = -\sum_i(D^+\boldsymbol{\psi}_{i+1/2})^T\boldsymbol{\phi}_{i+1/2}\Delta x, \tag{41e}$$

$$\sum_i(\boldsymbol{\phi}_{i+1/2})^T D^c\boldsymbol{\phi}_{i+1/2}\Delta x = 0, \tag{41f}$$

$$\langle\mu\boldsymbol{\phi}\rangle^T\langle\mu\boldsymbol{\phi}\rangle \le \frac{1}{2}\langle|\mu|\boldsymbol{\phi}^T\boldsymbol{\phi}\rangle,\ \boldsymbol{\phi}\in L^2([-1,1]). \tag{41g}$$

4.2. The stability analysis for the gPC-SG for the limiting diffusion equation

First we prove that the gPC-SG scheme (37) for the nonlinear diffusion equation (27) is unconditionally stable. Denote $\hat{\boldsymbol{\phi}}^{n+1}_{i+1/2} = (\mathbf{I} + \frac{4}{3}\mathbf{C}^n_{i+1/2})\delta^0\hat{\boldsymbol{\theta}}^{n+1}_{i+1/2}$, multiply $\hat{\boldsymbol{\theta}}^{n+1}_i$ on both sides of (37) and sum over $i\in\mathbb{Z}$:

$$\begin{aligned}\text{RHS} &= \sum_i(\hat{\boldsymbol{\theta}}^{n+1}_i)^T D^0\hat{\boldsymbol{\phi}}^{n+1}_i = -\sum_i(\delta^0\hat{\boldsymbol{\theta}}^{n+1}_{i+1/2})^T\hat{\boldsymbol{\phi}}^{n+1}_{i+1/2}\\ &= -\sum_i(\delta^0\hat{\boldsymbol{\theta}}^{n+1}_{i+1/2})^T(\mathbf{I} + \frac{4}{3}\mathbf{C}^n_{i+1/2})\delta^0\hat{\boldsymbol{\theta}}^{n+1}_{i+1/2} \le 0.\end{aligned}$$

The second equality used (41d) and the last inequality follows from the fact that $\mathbf{I} + \frac{4}{3}\mathbf{C}^n_{i+1/2}$ is symmetric, positive and definite.

$$\begin{aligned}\text{LHS} &= \sum_i(\hat{\boldsymbol{\theta}}^{n+1}_i)^T(\mathbf{I} + \frac{4}{3}(\mathbf{C})^n_{i+1/2})\frac{1}{\Delta t}(\hat{\boldsymbol{\theta}}^{n+1}_i - \hat{\boldsymbol{\theta}}^n_i)\\ &= \frac{1}{2\Delta t}\left[\sum_i(\hat{\boldsymbol{\theta}}^{n+1}_i)^T\left(\mathbf{I} + \frac{4}{3}(\mathbf{C})^n_{i+1/2}\right)\hat{\boldsymbol{\theta}}^{n+1}_i - \sum_i(\hat{\boldsymbol{\theta}}^n_i)^T\left(\mathbf{I} + \frac{4}{3}(\mathbf{C})^n_{i+1/2}\right)\hat{\boldsymbol{\theta}}^n_i\right.\\ &\quad\left. + \sum_i(\hat{\boldsymbol{\theta}}^{n+1}_i - \hat{\boldsymbol{\theta}}^n_i)^T\left(\mathbf{I} + \frac{4}{3}(\mathbf{C})^n_{i+1/2}\right)(\hat{\boldsymbol{\theta}}^{n+1}_i - \hat{\boldsymbol{\theta}}^n_i)\right]\\ &\ge \frac{1}{2\Delta t}\left[\sum_i(\hat{\boldsymbol{\theta}}^{n+1}_i)^T\left(\mathbf{I} + \frac{4}{3}(\mathbf{C})^n_{i+1/2}\right)\hat{\boldsymbol{\theta}}^{n+1}_i - \sum_i(\hat{\boldsymbol{\theta}}^n_i)^T\left(\mathbf{I} + \frac{4}{3}(\mathbf{C})^n_{i+1/2}\right)\hat{\boldsymbol{\theta}}^n_i\right].\end{aligned}$$

Therefore,

$$\sum_i(\hat{\boldsymbol{\theta}}^{n+1}_i)^T\left(\mathbf{I} + \frac{4}{3}(\mathbf{C})^n_{i+1/2}\right)\hat{\boldsymbol{\theta}}^{n+1}_i \le \sum_i(\hat{\boldsymbol{\theta}}^n_i)^T\left(\mathbf{I} + \frac{4}{3}(\mathbf{C})^n_{i+1/2}\right)\hat{\boldsymbol{\theta}}^n_i,$$

which implies

$$\|\hat{\boldsymbol{\theta}}^{n+1}\|^2 \le \|\hat{\boldsymbol{\theta}}^n\|^2.$$

Thus, scheme (37) is unconditionally stable.

4.3. Linear stability uniformly in ε

For $B(\theta) = \sigma(x,z)\theta$, where $\sigma(x,z) > 0$, we prove the linear stability of the scheme. (30) now becomes

$$\frac{\varepsilon^2}{\Delta t}(\hat{\mathbf{h}}^{n+1} - \hat{\mathbf{h}}^n) + \mathbf{C}(\partial_{xx}\hat{\boldsymbol{\theta}}^{n+1} + \hat{\mathbf{h}}^{n+1}) + \partial_x\langle\mu\hat{\mathbf{g}}^{n+1}\rangle = -\hat{\mathbf{h}}^{n+1}, \tag{42a}$$

$$\varepsilon^2\frac{\hat{\mathbf{g}}^{n+1}-\hat{\mathbf{g}}^n}{\Delta t}+\mu\mathbf{C}\partial_x\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon\partial_x(\mu\hat{\mathbf{g}}^n-\langle\mu\hat{\mathbf{g}}^{\mathbf{n}}\rangle)+\varepsilon^2\partial_x(\mu\hat{\mathbf{h}}^n)=-\hat{\mathbf{g}}^{n+1}, \tag{42b}$$

$$\frac{\hat{\boldsymbol{\theta}}^{n+1}-\hat{\boldsymbol{\theta}}^n}{\Delta t}=\partial_{xx}\hat{\boldsymbol{\theta}}^{n+1}+\hat{\mathbf{h}}^{n+1}, \tag{42c}$$

where

$$(\mathbf{C}(x))_{i,j}=\int\sigma(x,z)\Phi_i(z)\Phi_j(z)\rho(z)dz. \tag{43}$$

Since $\sigma(x,z)>0$, $\mathbf{C}(x)$ is a symmetric, positive and definite matrix as proved in [25].

Theorem 4.1. *Denote*

$$\lambda_0=\max_{k,i}\lambda_{i,k},\quad \lambda_{i,k}>0 \text{ are the eigenvalues of } \mathbf{C}_i$$

Let $\mathbf{C}_i=(L_i)^TL_i$ be the Cholesky decomposition with L_i a lower triangular matrix with positive diagonal entries. If Δt satisfies the following CFL condition

$$\Delta t\le\frac{1}{3+\lambda_0}((\Delta x)^2+2\varepsilon\Delta x), \tag{44}$$

then the sequences $\hat{\boldsymbol{\theta}}^n$, $\hat{\mathbf{g}}^n$ and $\hat{\mathbf{h}}^n$ defined by (21) satisfy the energy estimate

$$\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^2\hat{\mathbf{h}}^{n+1}\|^2+\||\varepsilon\hat{\mathbf{g}}^{n+1}\||^2+\|L\hat{\boldsymbol{\theta}}^{n+1}\|^2\le\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n}+\varepsilon^2\hat{\mathbf{h}}^{n}\|^2+\||\varepsilon\hat{\mathbf{g}}^{n}\||^2+\|L\hat{\boldsymbol{\theta}}^{n}\|^2 \tag{45}$$

for every n, and hence the scheme (42) is stable.

The proof follows the deterministic analogy in [9] with the addition of the temperature equation here. It is given in Appendix A.

Remark 4.1. Although the CFL condition (44) still depends on ε, it has a lower bound:

$$\Delta t\le\frac{1}{3+\lambda_0}(\Delta x)^2,$$

as $\varepsilon\to 0$. In this sense it is a uniform in ε stability.

5. Regularity in the random space and spectral accuracy analysis for linear problems

In this section, for simplicity we assume $z\in I_z$ is a one-dimensional random variable, where I_z has finite support (e.g. uniform and Beta distributions). We prove that for linear collision operator $B(\theta)=\sigma(x,z)\theta$, the solutions to the radiative transfer equations with random inputs preserve the regularity in the random space of the initial data. Then based on the regularity, we conduct spectral accuracy analysis and error estimates for the stochastic Galerkin method.

5.1. Regularity in the random space

The l-th order formal differentiation of the radiative transfer equations with respect to z is

$$\varepsilon^2\partial_t(\partial_z^lI)+\varepsilon\mu\partial_x(\partial_z^lI)=\partial_z^l(\sigma(x,z)\theta)-\partial_z^lI, \tag{46a}$$

$$\varepsilon^2\partial_t(\partial_z^l\theta)=\varepsilon^2\partial_{xx}(\partial_z^l\theta)-(\partial_z^l(\sigma(x,z)\theta)-\langle\partial_z^lI\rangle). \tag{46b}$$

Define norm

$$\left(\int_D\int_{I_z}\langle I(t,x,\mu,z)^2\rangle\rho(z)dzdx\right)^{1/2}=\||I\||_\Gamma,$$

$$\left(\int_D\int_{I_z}\theta(t,x,z)^2\rho(z)dzdx\right)^{1/2}=\|\theta\|_\Gamma.$$

Theorem 5.1. *Assume $\sigma(x,z)$ depends on z linearly, and*

$$0<\sigma_m<\sigma<\sigma_M<+\infty,\quad \max_z|\partial_z\sigma|\le\gamma_1,\quad \max_x|\partial_{xx}\sigma|\le\gamma_2.$$

If for integer $m\ge 0$,

$$\||\partial_z^l I_I\|\|_\Gamma^2+\|\sqrt{\sigma}\partial_z^l\theta_I\|_\Gamma^2\le\beta,\ \ \text{for all}\ \ l=0,\cdots,m.$$

Then

$$\||\partial_z^l I(t,\cdot,\cdot,\cdot)\|\|_\Gamma^2+\|\sqrt{\sigma}\partial_z^l\theta(t,\cdot,\cdot,\cdot)\|_\Gamma^2\le C_{\gamma,l}\beta l!e^{\frac{\gamma l}{\varepsilon^2}t},\ \ l=0,\cdots,m,$$

where γ is a constant depending on γ_1, γ_2 and c_m, $C_{\gamma,l}=C_{\gamma,l-1}+1$ is a constant depending on γ and l.

Proof. Multiplying $\partial_z^l I$ to both sides of (46a), taking $\langle\cdot\rangle$ and integrating over $D\times I_z$, one gets

$$\begin{aligned}
\text{LHS}&=\varepsilon^2\int_D\int_{I_z}\langle(\partial_z^l I)\partial_t(\partial_z^l I)\rangle+\varepsilon\int_D\int_{I_z}\langle\mu(\partial_z^l I)\partial_x(\partial_z^l I)\rangle\\
&=\frac{\varepsilon^2}{2}\partial_t\int_D\int_{I_z}\langle(\partial_z^l I)^2\rangle+\frac{\varepsilon}{2}\int_D\int_{I_z}\langle\mu\partial_x(\partial_z^l I)^2\rangle=\frac{\varepsilon^2}{2}\partial_t\int_D\int_{I_z}\langle(\partial_z^l I)^2\rangle,\\
\text{RHS}&=\int_D\int_{I_z}\langle\partial_z^l I\rangle\partial_z^l(\sigma\theta)-\int_D\int_{I_z}\langle(\partial_z^l I)^2\rangle.
\end{aligned}\tag{47}$$

Multiplying $\sigma\partial_z^l\theta$ to both sides of (46b) and integrating over $D\times I_z$, one gets

$$\begin{aligned}
\text{LHS}&=\varepsilon^2\int_D\int_{I_z}(\sigma\partial_z^l\theta)\partial_t(\partial_z^l\theta)=\frac{\varepsilon^2}{2}\partial_t\int_D\int_{I_z}\sigma(\partial_z^l\theta)^2,\\
\text{RHS}&=\varepsilon^2\int_D\int_{I_z}(\sigma\partial_z^l\theta)\partial_{xx}\partial_z^l\theta-\int_D\int_{I_z}\sigma\partial_z^l\theta\partial_z^l(\sigma\theta)+\int_D\int_{I_z}\sigma\partial_z^l\theta\langle\partial_z^l I\rangle\\
&=-\varepsilon^2\int_D\int_{I_z}\partial_x(\sigma\partial_z^l\theta)\partial_x\partial_z^l\theta-\int_D\int_{I_z}\sigma\partial_z^l\theta\partial_z^l(\sigma\theta)+\int_D\int_{I_z}\sigma\partial_z^l\theta\langle\partial_z^l I\rangle\\
&=-\varepsilon^2\int_D\int_{I_z}\sigma(\partial_x\partial_z^l\theta)^2-\varepsilon^2\int_D\int_{I_z}\frac{1}{2}\partial_x\sigma\partial_x(\partial_z^l\theta)^2-\int_D\int_{I_z}\sigma\partial_z^l\theta\partial_z^l(\sigma\theta)+\int_D\int_{I_z}\sigma\partial_z^l\theta\langle\partial_z^l I\rangle\\
&=-\varepsilon^2\int_D\int_{I_z}\sigma(\partial_x\partial_z^l\theta)^2+\varepsilon^2\int_D\int_{I_z}\frac{1}{2}\partial_{xx}\sigma(\partial_z^l\theta)^2-\int_D\int_{I_z}\sigma\partial_z^l\theta\partial_z^l(\sigma\theta)+\int_D\int_{I_z}\sigma\partial_z^l\theta\langle\partial_z^l I\rangle.
\end{aligned}\tag{48}$$

Adding equation (47) and (48) gives

$$\begin{aligned}
\text{LHS}&=\frac{\varepsilon^2}{2}\partial_t(\||\partial_z^l I\|\|_\Gamma^2+\|\sqrt{\sigma}\partial_z^l\theta\|_\Gamma^2),\\
\text{RHS}&=-\varepsilon^2\int_D\int_{I_z}\sigma(\partial_x\partial_z^l\theta)^2+\varepsilon^2\int_D\int_{I_z}\frac{1}{2}\partial_{xx}\sigma(\partial_z^l\theta)^2\\
&\quad+\int_D\int_{I_z}\langle\partial_z^l I\rangle\partial_z^l(\sigma\theta)-\int_D\int_{I_z}\langle(\partial_z^l I)^2\rangle-\int_D\int_{I_z}\sigma\partial_z^l\theta\partial_z^l(\sigma\theta)+\int_D\int_{I_z}\sigma\partial_z^l\theta\langle\partial_z^l I\rangle\\
&=-\varepsilon^2\int_D\int_{I_z}\sigma(\partial_x\partial_z^l\theta)^2+\varepsilon^2\int_D\int_{I_z}\frac{1}{2}\partial_{xx}\sigma(\partial_z^l\theta)^2-\int_D\int_{I_z}\langle(\partial_z^l I)^2\rangle-\langle\partial_z^l I\rangle^2\\
&\quad+\int_D\int_{I_z}\langle\partial_z^l I\rangle\partial_z^l(\sigma\theta)-\int_D\int_{I_z}\langle\partial_z^l I\rangle^2-\int_D\int_{I_z}\sigma\partial_z^l\theta\partial_z^l(\sigma\theta)+\int_D\int_{I_z}\sigma\partial_z^l\theta\langle\partial_z^l I\rangle\\
&\le\frac{\varepsilon^2}{2}\gamma_2\|\partial_z^l\theta\|_\Gamma^2-\int_D\int_{I_z}(\langle\partial_z^l I\rangle-\partial_z^l(\sigma\theta))(\langle\partial_z^l I\rangle-\sigma\partial_z^l\theta)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\varepsilon^2}{2}\gamma_2\|\partial_z^l\theta\|_\Gamma^2 - \int_D\int_{I_z}(\langle\partial_z^l I\rangle - \sigma\partial_z^l\theta - l\partial_z\sigma\,\partial_z^{l-1}\theta)(\langle\partial_z^l I\rangle - \sigma\partial_z^l\theta)\\
&= \frac{\varepsilon^2}{2}\gamma_2\|\partial_z^l\theta\|_\Gamma^2 - \int_D\int_{I_z}(\langle\partial_z^l I\rangle - \sigma\partial_z^l\theta)^2 + l\int_D\int_{I_z}\partial_z\sigma\,\partial_z^{l-1}\theta\langle\partial_z^l I\rangle - l\int_D\int_{I_z}\partial_z\sigma\,\partial_z^{l-1}\theta\sigma\,\partial_z^l\theta\\
&\le \frac{\varepsilon^2}{2}\gamma_2\|\partial_z^l\theta\|_\Gamma^2 + \frac{l}{2}\int_D\int_{I_z}|\partial_z\sigma|\cdot|\partial_z^{l-1}\theta|^2 + |\partial_z\sigma|\cdot\langle\partial_z^l I\rangle^2 + \sigma|\partial_z\sigma|\cdot|\partial_z^{l-1}\theta|^2 + \sigma|\partial_z\sigma|\cdot|\partial_z^l\theta|^2\\
&\le \frac{\varepsilon^2}{2}\gamma_2\|\partial_z^l\theta\|_\Gamma^2 + \frac{\gamma_1 l}{2}\int_D\int_{I_z}\langle\partial_z^l I\rangle^2 + \sigma|\partial_z^l\theta|^2 + \frac{\gamma_1 l}{2}\int_D\int_{I_z}(\sigma+1)|\partial_z^{l-1}\theta|^2\\
&\le \frac{\varepsilon^2}{2}\gamma_2\|\partial_z^l\theta\|_\Gamma^2 + \frac{\gamma_1 l}{2}(\||\partial_z^l I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^l\theta\|_\Gamma^2) + \frac{\gamma_1 l}{2}\int_D\int_{I_z}\frac{\sigma+1}{\sigma}(\langle|\partial_z^{l-1}I|^2\rangle + \sigma|\partial_z^{l-1}\theta|^2)\\
&\le \frac{\varepsilon^2}{2\sigma_m}\gamma_2\|\sqrt{\sigma}\partial_z^l\theta\|_\Gamma^2 + \frac{\gamma_1 l}{2}(\||\partial_z^l I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^l\theta\|_\Gamma^2) + \frac{\gamma_1 l}{2}\frac{\sigma_m+1}{\sigma_m}\int_D\int_{I_z}\langle|\partial_z^{l-1}I|^2\rangle + \sigma|\partial_z^{l-1}\theta|^2\\
&\le \frac{\gamma l}{2}\left[(\||\partial_z^l I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^l\theta\|_\Gamma^2) + (\||\partial_z^{l-1} I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^{l-1}\theta\|_\Gamma^2)\right].
\end{aligned}
\tag{49}
$$

Thus

$$
\partial_t(\||\partial_z^l I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^l\theta\|_\Gamma^2) \le \frac{\gamma l}{\varepsilon^2}\left[(\||\partial_z^l I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^l\theta\|_\Gamma^2) + (\||\partial_z^{l-1} I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^{l-1}\theta\|_\Gamma^2)\right].
\tag{50}
$$

Now we use mathematical induction to prove the theorem. It clearly holds for $l = 0$. Assume that

$$
\||\partial_z^{l-1} I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^{l-1}\theta\|_\Gamma^2 \le C_{\gamma,l-1}\beta(l-1)!e^{\frac{\gamma(l-1)}{\varepsilon^2}t},
$$

using Gronwall's inequality,

$$
\begin{aligned}
\||\partial_z^l I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^l\theta\|_\Gamma^2 &\le e^{\frac{\gamma l}{\varepsilon^2}t}(\||\partial_z^l I_I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^l\theta_I\|_\Gamma^2) + \frac{\gamma l}{\varepsilon^2}\int_0^t e^{\frac{\gamma l}{\varepsilon^2}(t-s)}(\||\partial_z^{l-1} I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^{l-1}\theta\|_\Gamma^2)ds\\
&\le e^{\frac{\gamma l}{\varepsilon^2}t}\beta + \frac{\gamma l}{\varepsilon^2}e^{\frac{\gamma l}{\varepsilon^2}t}\int_0^t e^{-\frac{\gamma l}{\varepsilon^2}s}C_{\gamma,l-1}\beta(l-1)!e^{\frac{\gamma(l-1)}{\varepsilon^2}s}ds\\
&= e^{\frac{\gamma l}{\varepsilon^2}t}\beta + \frac{\gamma l}{\varepsilon^2}e^{\frac{\gamma l}{\varepsilon^2}t}C_{\gamma,l-1}\beta(l-1)!\int_0^t e^{-\frac{\gamma}{\varepsilon^2}s}ds\\
&= e^{\frac{\gamma l}{\varepsilon^2}t}\beta + C_{\gamma,l-1}\beta l!e^{\frac{\gamma l}{\varepsilon^2}t}(1-e^{-\frac{\gamma}{\varepsilon^2}t})\\
&\le C_{\gamma,l}\beta l!e^{\frac{\gamma l}{\varepsilon^2}t}
\end{aligned}
\tag{51}
$$

for $C_{\gamma,l} = C_{\gamma,l-1} + 1$. □

5.2. A spectral accuracy analysis

Let I and θ be the solution to the radiation transfer equation (1). We define the operator

$$
P_K I = \sum_{k=1}^K \langle I, \Phi_k\rangle_\rho \Phi_k, \quad P_K\theta = \sum_{k=1}^K \langle\theta, \Phi_k\rangle_\rho\Phi_k.
$$

The error can be split into the projection error R_K^I, R_K^θ and SG error e_K^I, e_K^θ,

$$
I - I_K = I - P_K I + P_K I - I_K := R_K^I + e_K^I,
\tag{52a}
$$

$$
\theta - \theta_K = \theta - P_K\theta + P_K\theta - \theta_K := R_K^\theta + e_K^\theta,
\tag{52b}
$$

where

$$R_K^I = I - P_K I, \quad R_K^\theta = \theta - P_K \theta, \tag{53a}$$

$$e_K^I = P_K I - I_K, \quad e_K^\theta = P_K \theta - \theta_K. \tag{53b}$$

The SG error can be rewritten explicitly as following

$$e_K^I = P_K I - I_K = \sum_{k=1}^{K} (\langle I, \Phi_k \rangle_\rho - \hat{I}_k)\Phi_k = \hat{\mathbf{e}}_I \cdot \mathbf{\Phi},$$

$$e_K^\theta = P_K \theta - \theta_K = \sum_{k=1}^{K} (\langle \theta, \Phi_k \rangle_\rho - \hat{\theta}_k)\Phi_k = \hat{\mathbf{e}}_\theta \cdot \mathbf{\Phi},$$

where

$$\hat{\mathbf{e}}_I = (\langle I, \Phi_1 \rangle_\rho - \hat{I}_1, \cdots, \langle I, \Phi_K \rangle_\rho - \hat{I}_K), \tag{54a}$$

$$\hat{\mathbf{e}}_\theta = (\langle \theta, \Phi_1 \rangle_\rho - \hat{\theta}_1, \cdots, \langle \theta, \Phi_K \rangle_\rho - \hat{\theta}_K), \tag{54b}$$

$$\mathbf{\Phi} = (\Phi_1, \cdots, \Phi_K). \tag{54c}$$

By standard error estimate for orthogonal polynomial approximations [3], and Theorem 5.1, for $0 \le t \le T$,

$$|||R_K^I|||_\Gamma^2 + \|\sqrt{\sigma} R_K^\theta\|_\Gamma^2 \le C_1 K^{-2m} |||\partial_z^m I|||_\Gamma^2 + C_2 K^{-2m} \|\sqrt{\sigma}\partial_z^m \theta\|_\Gamma^2 \le C_{\gamma,m} K^{-2m} \beta m! e^{\frac{\gamma m}{\varepsilon^2} T}. \tag{55}$$

It remains to estimate e_K^I, e_K^θ. Define the operators

$$\mathcal{Q}(I, \theta) = \sigma\theta - I, \quad \tilde{\mathcal{Q}}(I, \theta) = \langle \mathcal{Q}(I, \theta) \rangle, \tag{56a}$$

$$\mathcal{L}_1(I, \theta) = \varepsilon^2 \partial_t I + \varepsilon\mu\partial_x I - \mathcal{Q}(I, \theta), \tag{56b}$$

$$\mathcal{L}_2(I, \theta) = \varepsilon^2 \partial_t \theta - \varepsilon^2 \partial_{xx}\theta + \langle \mathcal{Q}(I, \theta) \rangle. \tag{56c}$$

Define inner product and norm

$$\langle u, v \rangle_\rho = \int_{I_z} u(z)v(z)\rho(z)dz, \ \|u\|_\rho = \left(\int_{I_z} u(z)^2 \rho(z) dz \right)^{1/2}.$$

We first prove two elementary properties on operators $\mathcal{L}_1, \mathcal{L}_2$ and $\mathcal{Q}$.

Lemma 5.1.

$$\langle \mathcal{L}_1(R_K^I, R_K^\theta), \Phi_k \rangle_\rho = -\langle \mathcal{Q}(R_K^I, R_K^\theta), \Phi_k \rangle_\rho, \ k = 1, \cdots, K, \tag{57a}$$

$$\langle \mathcal{L}_2(R_K^I, R_K^\theta), \Phi_k \rangle_\rho = \langle \tilde{\mathcal{Q}}(R_K^I, R_K^\theta), \Phi_k \rangle_\rho, \ k = 1, \cdots, K. \tag{57b}$$

Proof. Since $R_K^I = I - P_K I = I - \sum_{k=1}^K \langle I, \Phi_k \rangle_\rho$, and $\Phi_k \in \text{span}\{\Phi_{K+1}, \Phi_{K+2}, \cdots\}$, due to the orthogonal property of $\{\Phi_k\}$, $\langle R_K^I, \Phi_k \rangle = 0$ for $k = 1, \cdots, K$, thus $\langle \partial_t R_K^I, \Phi_k \rangle = 0$ for $k = 1, \cdots, K$.

The second term in $\langle \mathcal{L}_1(R_K^I, R_K^\theta), \Phi_k \rangle_\rho$ and the first two terms in $\langle \mathcal{L}_2(R_K^I, R_K^\theta), \Phi_k \rangle_\rho$ are zero following similar argument. Therefore, we have, for $k = 1, \cdots, K$,

$$\langle \mathcal{L}_1(R_K^I, R_K^\theta), \Phi_k \rangle_\rho = -\langle \mathcal{Q}(R_K^I, R_K^\theta), \Phi_k \rangle_\rho,$$

$$\langle \mathcal{L}_2(R_K^I, R_K^\theta), \Phi_k \rangle_\rho = \langle \tilde{\mathcal{Q}}(R_K^I, R_K^\theta), \Phi_k \rangle_\rho. \quad \square$$

Lemma 5.2. *Under the assumption of Theorem 5.1,*

$$|||\mathcal{Q}(R_K^I, R_K^\theta)|||_\Gamma^2 \le (1 + \sigma_M) C_{\gamma,m} K^{-2m} \beta m! e^{\frac{\gamma m}{\varepsilon^2} t}.$$

Proof.

$$
\begin{aligned}
\|\|\mathcal{Q}(R_K^I, R_K^\theta)\|\|_\Gamma^2 &= \int_D \int_{I_z} \langle (\sigma R_K^\theta - R_K^I)^2 \rangle \rho(z) dz dx \\
&= \int_D \int_{I_z} \langle (R_K^I)^2 - 2\sigma R_K^I R_K^\theta + \sigma^2 (R_K^\theta)^2 \rangle \rho(z) dz dx \\
&\le \int_D \int_{I_z} \langle (R_K^I)^2 + \sigma (R_K^\theta)^2 + \sigma (R_K^I)^2 + \sigma^2 (R_K^\theta)^2 \rangle \rho(z) dz dx \\
&\le (1+\sigma_M)(\|\|R_K^I\|\|_\Gamma^2 + \|\sqrt{\sigma} R_K^\theta\|_\Gamma^2) \le (1+\sigma_M) C_{\gamma,m} K^{-2m} \beta m! e^{\frac{\gamma m}{\varepsilon^2} t}. \quad \square
\end{aligned}
$$

Since $\mathcal{L}_1(I,\theta) = 0$, $\mathcal{L}_2(I,\theta) = 0$ and $P_K \mathcal{L}_1(I_K, \theta_K) = 0$, $P_K \mathcal{L}_2(I_K, \theta_K) = 0$, from (52) and Lemma 5.1, for $k = 1, \cdots, K$,

$$
\langle \mathcal{L}_1(e_K^I, e_K^\theta), \Phi_k \rangle_\rho = -\langle \mathcal{L}_1(R_K^I, R_K^\theta), \Phi_k \rangle_\rho = \langle \mathcal{Q}(R_K^I, R_K^\theta), \Phi_k \rangle_\rho, \tag{58a}
$$

$$
\langle \mathcal{L}_2(e_K^I, e_K^\theta), \Phi_k \rangle_\rho = -\langle \mathcal{L}_2(R_K^I, R_K^\theta), \Phi_k \rangle_\rho = -\langle \tilde{\mathcal{Q}}(R_K^I, R_K^\theta), \Phi_k \rangle_\rho. \tag{58b}
$$

Now taking the scalar product of $\hat{\mathbf{e}}_I$ in (54a) with

$$
(\langle \mathcal{L}_1(e_K^I, e_K^\theta), \Phi_1 \rangle_\rho, \langle \mathcal{L}_1(e_K^I, e_K^\theta), \Phi_2 \rangle_\rho, \cdots, \langle \mathcal{L}_1(e_K^I, e_K^\theta), \Phi_K \rangle_\rho)
$$

and taking $\langle \cdot \rangle$ and integrating on D, with (56b) and (58a), give

$$
\varepsilon^2 \partial_t \|\|e_K^I\|\|_\Gamma^2 - \int_D \langle \langle \mathcal{Q}(e_K^I, e_K^\theta), e_K^I \rangle_\rho \rangle = \int_D \langle \langle \mathcal{Q}(R_K^I, R_K^\theta), e_K^I \rangle_\rho \rangle. \tag{59}
$$

Then taking the scalar product of $\mathbf{C}\hat{\mathbf{e}}_\theta$ in (54b) for $\mathbf{C}$ defined in (43) with

$$
(\langle \mathcal{L}_2(e_K^I, e_K^\theta), \Phi_1 \rangle_\rho, \langle \mathcal{L}_2(e_K^I, e_K^\theta), \Phi_2 \rangle_\rho, \cdots, \langle \mathcal{L}_2(e_K^I, e_K^\theta), \Phi_K \rangle_\rho)
$$

and integrating on D, with (56c) and (58b), give

$$
\begin{aligned}
&\varepsilon^2 \partial_t \|\sqrt{c} e_K^\theta\|_\Gamma^2 - \varepsilon^2 \int_D \langle \partial_{xx} e_K^\theta, c e_K^\theta \rangle_\rho + \int_D \langle \tilde{\mathcal{Q}}(e_K^I, e_K^\theta), c e_K^\theta \rangle_\rho \\
&= -\int_D \langle \tilde{\mathcal{Q}}(R_K^I, R_K^\theta), c e_K^\theta \rangle_\rho.
\end{aligned} \tag{60}
$$

Adding equation (59) and (60) yields

$$
\begin{aligned}
&\varepsilon^2 \partial_t (\|\|e_K^I\|\|_\Gamma^2 + \|\sqrt{\sigma} e_K^\theta\|_\Gamma^2) \\
&= \int_D \langle \langle \mathcal{Q}(e_K^I, e_K^\theta), e_K^I \rangle_\rho \rangle + \int_D \langle \langle \mathcal{Q}(R_K^I, R_K^\theta), e_K^I \rangle_\rho \rangle + \varepsilon^2 \int_D \langle \partial_{xx} e_K^\theta, \sigma e_K^\theta \rangle_\rho - \int_D \langle \tilde{\mathcal{Q}}(e_K^I, e_K^\theta), \sigma e_K^\theta \rangle_\rho \\
&\quad - \int_D \langle \tilde{\mathcal{Q}}(R_K^I, R_K^\theta), \sigma e_K^\theta \rangle_\rho \\
&= \int_D \langle \langle \mathcal{Q}(e_K^I, e_K^\theta), e_K^I - \sigma e_K^\theta \rangle_\rho \rangle + \int_D \langle \langle \mathcal{Q}(R_K^I, R_K^\theta), e_K^I - \sigma e_K^\theta \rangle_\rho \rangle - \varepsilon^2 \int_D \langle \partial_x e_K^\theta, \partial_x (\sigma e_K^\theta) \rangle_\rho \\
&= \int_D \langle \langle \mathcal{Q}(e_K^I, e_K^\theta), e_K^I - \sigma e_K^\theta \rangle_\rho \rangle + \int_D \langle \langle \mathcal{Q}(R_K^I, R_K^\theta), e_K^I - \sigma e_K^\theta \rangle_\rho \rangle - \frac{\varepsilon^2}{2} \int_D \langle \partial_x (e_K^\theta)^2, \partial_x \sigma \rangle_\rho - \varepsilon^2 \int_D \sigma \|\partial_x e_K^\theta\|_\rho^2 \\
&\le \int_D \langle \langle \mathcal{Q}(e_K^I, e_K^\theta), e_K^I - \sigma e_K^\theta \rangle_\rho \rangle + \int_D \langle \langle \mathcal{Q}(R_K^I, R_K^\theta), e_K^I - \sigma e_K^\theta \rangle_\rho \rangle + \frac{\varepsilon^2}{2} \int_D \langle (e_K^\theta)^2, \partial_{xx} \sigma \rangle_\rho \\
&\le -\|\|e_K^I - \sigma e_K^\theta\|\|_\Gamma^2 + \int_D \langle \|\mathcal{Q}(R_K^I, R_K^\theta)\|_\rho \|e_K^I - \sigma e_K^\theta\|_\rho \rangle + \frac{\varepsilon^2 \gamma_2}{2} \|e_K^\theta\|_\Gamma^2
\end{aligned}
$$

$$\leq -\|\|e_K^I - \sigma e_K^\theta\|\|_\Gamma^2 + \frac{1}{2}\|\|\mathcal{Q}(R_K^I, R_K^\theta)\|\|_\Gamma^2 + \frac{1}{2}\|e_K^I - \sigma e_K^\theta\|_\Gamma^2 + \frac{\varepsilon^2\gamma_2}{2}\|e_K^\theta\|_\Gamma^2$$
$$\leq \frac{1}{2}\|\|\mathcal{Q}(R_K^I, R_K^\theta)\|\|_\Gamma^2 + \frac{\varepsilon^2\gamma_2}{2\sigma_m}\|\sqrt{\sigma}e_K^\theta\|_\Gamma^2$$
$$\leq \frac{1}{2}(1+\sigma_M)C_{\gamma,m}K^{-2m}\beta m! e^{\frac{\gamma m}{\varepsilon^2}t} + \frac{\varepsilon^2\gamma_2}{2\sigma_m}(\|\|e_K^I\|\|_\Gamma^2 + \|\sqrt{\sigma}e_K^\theta\|_\Gamma^2),$$

where the second inequality is by the definition of $\mathcal{Q}$ and Cauchy–Schwartz inequality and the last inequality uses Lemma 5.2 and the assumption of Theorem 5.1.

Thus applying Gronwall's Lemma,

$$\begin{aligned}
&\|\|e_K^I\|\|_\Gamma^2 + \|\sqrt{\sigma}e_K^\theta\|_\Gamma^2 \\
\leq& e^{\frac{\gamma_2}{2\sigma_m}t}(\|\|e_K^{I_I}\|\|_\Gamma^2 + \|\sqrt{\sigma}e_K^{\theta_I}\|_\Gamma^2) + \frac{1}{2\varepsilon^2}(1+\sigma_M)C_{\gamma,m}K^{-2m}\beta m! e^{\frac{\gamma_2}{2\sigma_m}t}\int_0^t e^{(\frac{\gamma m}{\varepsilon^2}-\frac{\gamma_2}{2\sigma_m})s}ds \\
\leq& e^{\frac{\gamma_2}{2\sigma_m}t}(\|\|e_K^{I_I}\|\|_\Gamma^2 + \|\sqrt{\sigma}e_K^{\theta_I}\|_\Gamma^2) + \frac{2\sigma_m(1+\sigma_M)}{\sigma_m\gamma m - \gamma_2\varepsilon^2}C_{\gamma,m}K^{-2m}\beta m! e^{\frac{\gamma m}{\varepsilon^2}t}
\end{aligned} \tag{61}$$

Now we are ready to prove the main spectral convergence theorem:

Theorem 5.2. *Assume c depends on z linearly, and*

$$0 < \sigma_m < \sigma < \sigma_M < +\infty, \quad \max_z |\partial_z\sigma| \leq \gamma_1, \quad \max_x |\partial_{xx}\sigma| \leq \gamma_2$$

Assume $\varepsilon < \sqrt{2\sigma_m\gamma/\gamma_2}$. If for integer $m \geq 0$,

$$\|\|\partial_z^l I_I\|\|_\Gamma^2 + \|\sqrt{\sigma}\partial_z^l\theta_I\|_\Gamma^2 \leq \beta, \quad \text{for all } l = 0, \cdots, m$$

Then

$$\|\|I - I_K\|\|_\Gamma^2 + \|\sqrt{\sigma}(\theta - \theta_K)\|_\Gamma^2 \leq \left(1 + \frac{2\sigma_m(1+\sigma_M)}{\sigma_m\gamma m - \gamma_2\varepsilon^2}\right)C_{\gamma,m}K^{-2m}\beta m! e^{\frac{\gamma m}{\varepsilon^2}t} \tag{62}$$

where γ is a constant depending on γ_1 and γ_2, $C_{\gamma,l}$ is a constant depending on γ and l as in Theorem 5.1.

Proof. From (52a) and (52b), one has

$$\|\|I - I_K\|\|_\Gamma^2 + \|\sqrt{\sigma}(\theta - \theta_K)\|_\Gamma^2 \leq \left(\|\|R_K^I\|\|_\Gamma^2 + \|\sqrt{\sigma}R_K^\theta)\|_\Gamma^2\right) + \left(\|\|e_K^I\|\|_\Gamma^2 + \|\sqrt{\sigma}e_K^\theta)\|_\Gamma^2\right)$$

Note that

$$e_K^{I_I} = P_K I - I_K|_{t=0} = 0, \quad e_K^{\theta_I} = P_K\theta - \theta_K|_{t=0} = 0.$$

Then combining (55) with (61) gives (62). □

Remark 5.1. The error in (62) shows that, for $\varepsilon \to 0$, one needs $K >> e^{\gamma t/\varepsilon^2}$. This justifies the need to the notion of AP such that practically, one can take K (and other numerical parameters) independent of ε.

6. Numerical tests

We consider the one-dimensional slab geometry introduced in the previous section. Furthermore, we assume a nonreflecting boundary, i.e., $\alpha = 0$. For the velocity discretization, 16 Gauss quadrature points are used. The spatial grid spacing is $h = 0.01$. We assume the same CFL condition as the deterministic case (See [20]) and thus take $\Delta t = 0.001$. We use the 4th order gPC-SG method and compare it with stochastic collocation method with 20 points sampling z, both for the transfer equation and its limiting diffusion equation. In stochastic collocation, one applies the deterministic AP solver to a set of selected sample points and then approximates the solution via an interpolation procedure. (See [22] for an overview of stochastic collocation methods.) The numerical results are examined by two quantities, the mean value and the standard deviation of θ. Given the gPC coefficients $\hat{\theta}_k$ of θ, the mean value and standard deviation are calculated as

$$E[\theta] \approx \hat{\theta}_1, \quad Sd[\theta] \approx \sqrt{\sum_{k=2}^{K}\hat{\theta}_k^2} \tag{63}$$

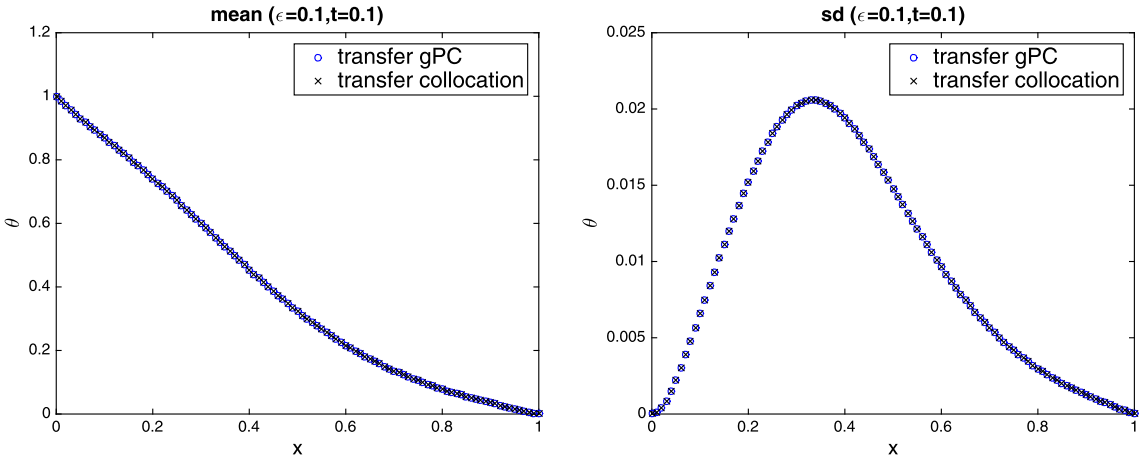

Fig. 1. Test 1. The mean (left) and standard deviation (right) of θ at time $t = 0.1$, obtained by the 4th-order gPC-SG (circles) and the 20-point stochastic collocation (crosses) with $\varepsilon = 0.1$.

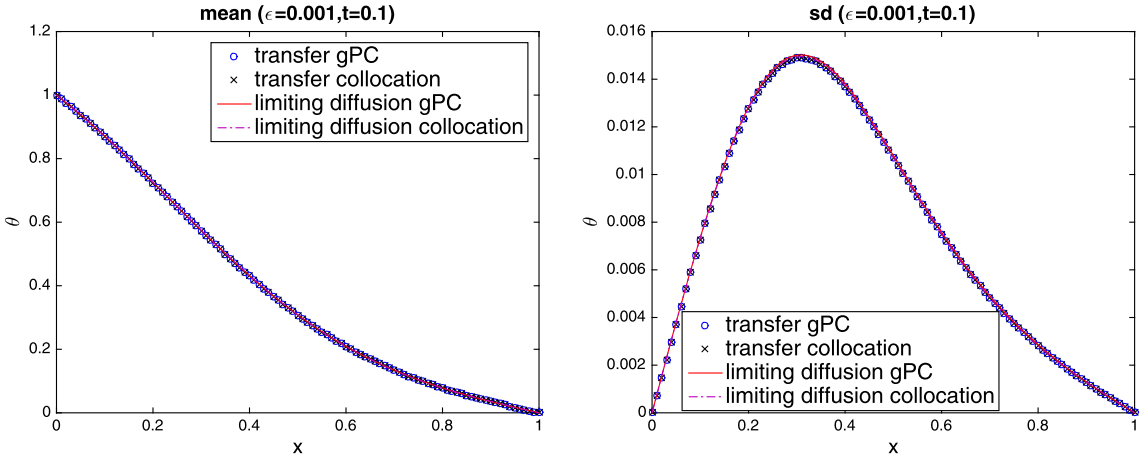

Fig. 2. Test 1. The mean (left) and standard deviation (right) of θ at time $t = 0.1$, obtained by the 4th-order gPC-SG (circles) and the 20-point stochastic collocation (crosses), along with the solutions of the random diffusion limit obtained by the 4th-order gPC-SG (solid line) and the 20-point stochastic collocation (dashed line) with $\varepsilon = 0.001$.

6.1. Test 1: 1D randomness in cross-section

We first consider the randomness in the cross-section with the following initial and boundary conditions:

$$I_I(x, \mu, z, 0) = 0, \quad \theta_I(x, z, 0) = 0, \quad x \in [0, 1]$$

$$\theta_B(0, z, t) = 1, \ \theta_B(1, z, t) = 0;$$

$$I_B(0, \mu, z, t) = 1 + 0.5z, \ \mu > 0, \quad I_B(1, \mu, z, t) = 0, \ \mu < 0,$$

and random coefficient

$$\sigma(z) = 1 + 0.5z, \ z \sim U[-1, 1].$$

The random space is just one dimension and z has a uniform distribution.

We first set $\varepsilon = 0.1$ to be a relatively large number. Compared with the reference solution obtained by stochastic collocation method, one can see a good agreement on both mean and standard deviation in Fig. 1.

Then we consider the case of a very small $\varepsilon = 0.001$. The efficiency of AP method is notable as the same mesh is used with much smaller ε. And again one can observe good agreement between gPC solutions and stochastic collocations. Furthermore, we plot the "semi-exact" solutions which are obtained by solving the limiting nonlinear diffusion equation (9) by the gPC-SG method and stochastic collocation respectively. Good agreements can be observed among these four in Fig. 2.

To compare the gPC-SG solution with the reference solution, we define the error in mean and standard deviation with L^2 norm in x as following:

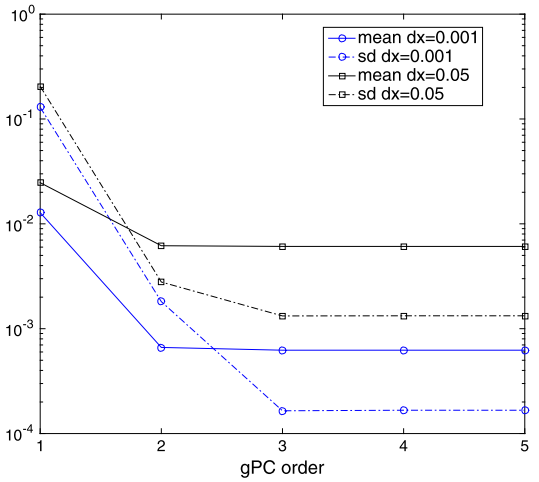

Fig. 3. Test 1. Error in mean (solid lines) and standard deviation (dashed lines), with respect to the gPC order for $\varepsilon = 0.001$. (Circles: $\Delta x = 0.001$ and squares: $\Delta x = 0.05$.)

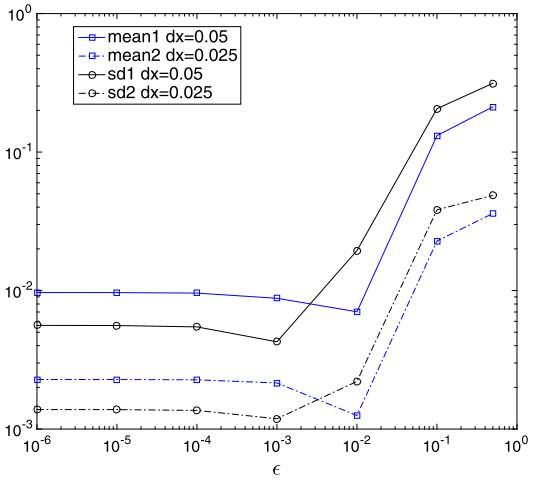

Fig. 4. Test 1. Differences in the mean (solid lines) and standard deviation (dashed line) between the 4th-order gPC-SG solution and the limiting diffusion solution with respect to different values of ε. (Squares: $\Delta x = 0.05$ and circles: $\Delta x = 0.025$.)

$$\mathcal{E}_{\text{mean}}(t) = \|\mathrm{E}[\theta^{\text{gPC}}] - \mathrm{E}[\theta^{\text{ref}}]\|_{L^2},$$
$$\mathcal{E}_{\text{sd}}(t) = \|\mathrm{Sd}[\theta^{\text{gPC}}] - \mathrm{Sd}[\theta^{\text{ref}}]\|_{L^2}.$$

Fig. 3 shows the errors at time $t = 0.1$ with respect to increasing gPC order with different meshes $\Delta x = 0.001$ (circles) and $\Delta x = 0.05$ (squares). We employ the 20-point stochastic collocation method as reference solution. The time step is $\Delta t = 0.1\Delta x$. One can observe fast exponential convergence with respect to the gPC order no matter one uses a coarse mesh (Δx is much larger than ε) or an intermediate mesh (Δx is of order ε).

In Fig. 4, we show the differences of the mean and standard deviation between the solution to the limiting diffusion equation and the 4th-order gPC-SG solution with respect to various values of ε up to time $t = 0.1$. The differences are measured with L^2 norm in x as following:

$$\epsilon_{\text{mean}}(t) = \|\mathrm{E}[\theta^{\varepsilon}] - \mathrm{E}[\theta^{0}]\|_{L^2},$$
$$\epsilon_{\text{sd}}(t) = \|\mathrm{Sd}[\theta^{\varepsilon}] - \mathrm{Sd}[\theta^{0}]\|_{L^2}.$$

Obviously, the differences decrease as ε gets smaller, thus showing the scheme captures the diffusion limit very well. When ε is small enough, the difference saturates as the numerical errors from the spatial and temporal discretizations become dominating. The solution to the limiting diffusion equation is obtained by the 20-point stochastic collocation method as reference.

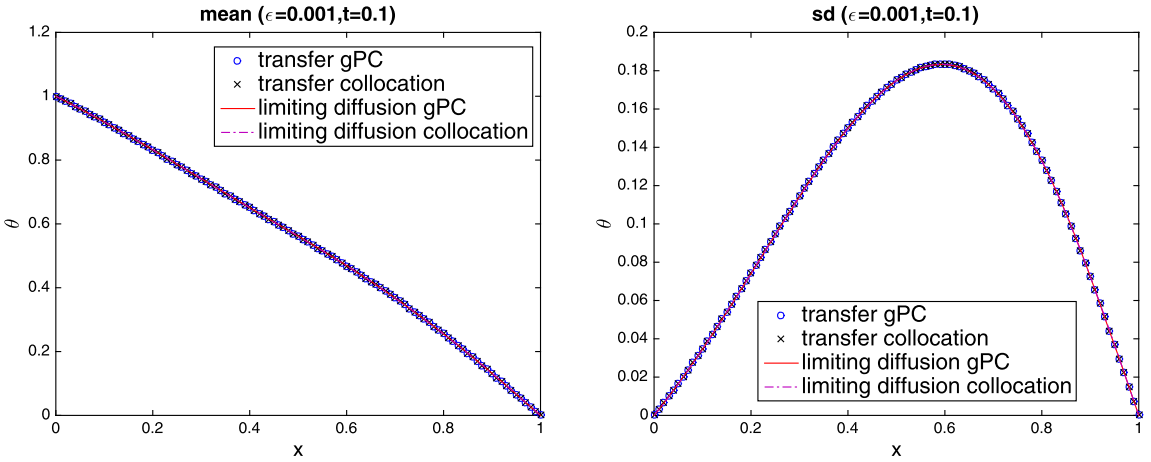

Fig. 5. Test 2. The mean (left) and standard deviation (right) of θ at time $t = 0.1$, $\varepsilon = 0.001$ obtained by the 4th-order gPC-SG (circles) and the 20-point stochastic collocation (crosses), along with the solutions of the random diffusion limit obtained by the 4th-order gPC-SG (solid line) and the 20-point stochastic collocation (dashed line).

6.2. Test 2: 1D randomness in initial data for $\varepsilon = 0.001$

We now consider the randomness in the initial data:

$$\theta_I(x, z, 0) = 0.5 + 0.5z,\ I_I(x, \mu, z, 0) = \sigma(0.5 + 0.5z)^4,\ z \sim U(-1, 1),$$

with a constant cross-section term and the same boundary data,

$$\sigma = 1$$

$$\theta_B(0, z, t) = 1,\ \theta_B(1, z, t) = 0;$$

$$I_B(0, \mu, z, t) = 1,\ \mu > 0,\ \ I_B(1, \mu, z, t) = 0,\ \mu < 0.$$

We only examine the case of a very small $\varepsilon = 0.001$ and observe good agreements of the two quantities obtained by the gPC-SG method, the stochastic collocation and the "semi-exact" solutions, as can be seen from Fig. 5.

6.3. Test 3: 1D randomness in boundary for $\varepsilon = 0.001$

We now consider the randomness in the boundary data:

$$\theta_B(0, z, t) = 1 + 0.5z,\ z \sim U(-1, 1),\ \theta_B(1, z, t) = 0;$$

$$I_B(0, \mu, z, t) = \sigma(1 + 0.5z)^4,\ \mu > 0,\ \ I_B(1, \mu, z, t) = 0,\ \mu < 0,$$

with the same constant cross-section term and the initial data

$$\sigma = 1,\ \theta_I(x, z, 0) = I_I(x, \mu, z, 0) = 0.$$

The mean and standard deviation obtained by the gPC-SG method, the stochastic collocation and the "semi-exact" solution match well, as shown in Fig. 6.

6.4. Test 4: 2D randomness in cross-section for $\varepsilon = 0.001$

We then model the random input in cross-section but as a random field of two dimension.

$$\sigma(x, z_1, z_2) = \frac{3}{4} + \frac{1}{4}\cos(2\pi x) + \frac{1}{8}\cos(4\pi x)z_1 + \frac{1}{12}\cos(6\pi x)z_2,$$

$$z_1 \sim U(-1, 1),\ z_2 \sim U(-1, 1),\ z_1, z_2 \text{ are independent from each other.}$$

This resembles the form of the well known Karhunen–Loeve expansion, which is widely used for modeling random fields [22].

The initial and boundary data are the same as test 1:

$$I_I(x, \mu, z, 0) = 0,\ \ \theta_I(x, z, 0) = 0,\ \ x \in [0, 1],$$

$$\theta_B(0, z, t) = 1,\ \theta_B(1, z, t) = 0;$$

$$I_B(0, \mu, z, t) = \sigma(x, z_1, z_2),\ \mu > 0,\ \ I_B(1, \mu, z, t) = 0,\ \mu < 0.$$

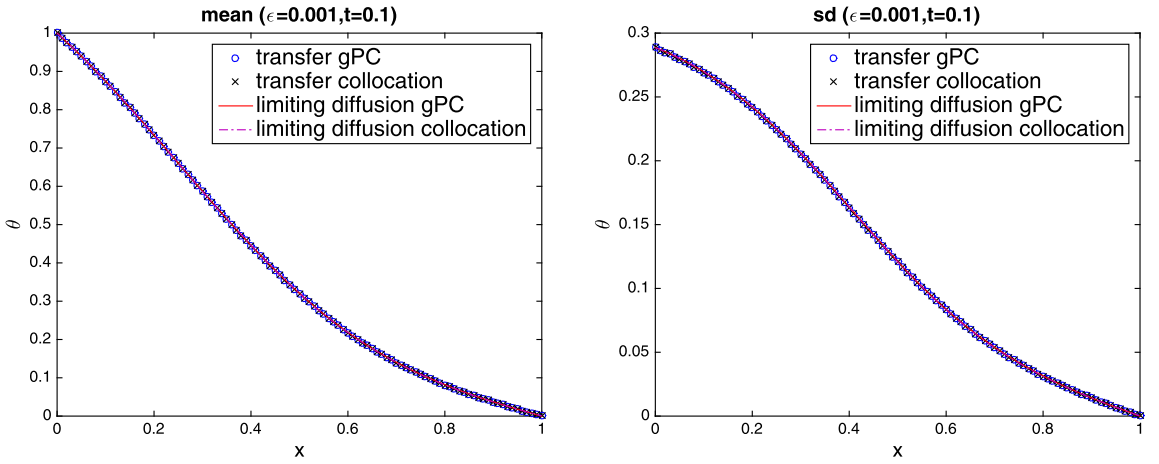

Fig. 6. Test 3. The mean (left) and standard deviation (right) of θ at time $t = 0.1$, $\varepsilon = 0.001$ obtained by the 4th-order gPC-SG (circles) and the 20-point stochastic collocation (crosses), along with the solutions of the random diffusion limit obtained by the 4th-order gPC-SG (solid line) and the 20-point stochastic collocation (dashed line).

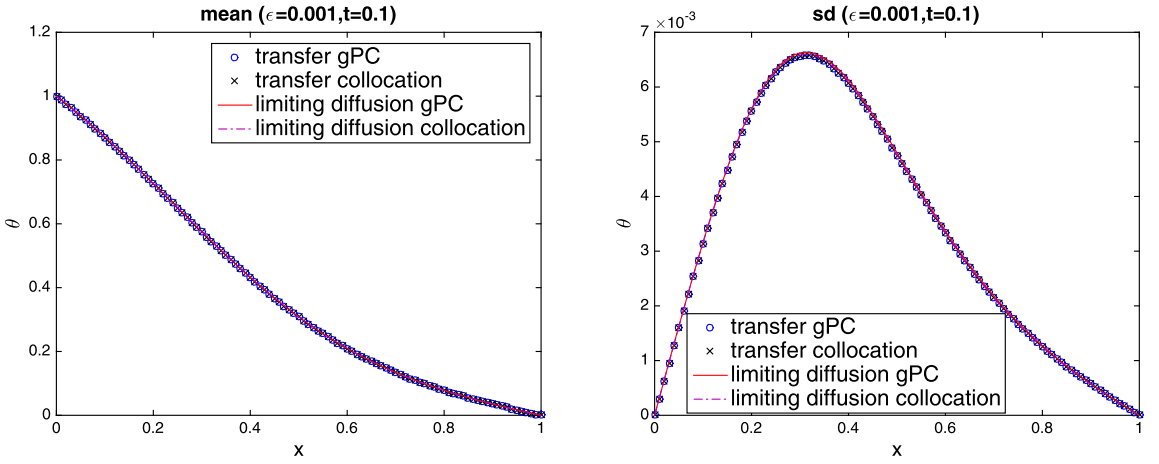

Fig. 7. Test 4. The mean (left) and standard deviation (right) of θ at time $t = 0.1$, $\varepsilon = 0.001$ obtained by the 9th-order gPC-SG (circles) and the 20-point stochastic collocation (crosses), along with the solution of the random diffusion limit obtained by the 9th-order gPC-SG (solid line) and the 20-point stochastic collocation (dashed line).

The mean and standard deviation of the solution are shown in Fig. 7, where a good agreement can be observed between the gPC-SG method of order 9 and the stochastic collocation over 20^2 Legendre–Gauss quadrature points. The computing efficiency of the gPC-SG method in high dimensional space becomes notable even in this 2D case. The number of sample points needed grows exponentially with dimension but the order of gPC-SG does not. And similar to previous cases, with small $\varepsilon = 0.001$, the results match well with the "semi-exact" solution.

Appendix A. Proof of Theorem 4.1

Proof. Plug (42c) into (42a):

$$\varepsilon^2 \frac{\hat{\mathbf{h}}^{n+1} - \hat{\mathbf{h}}^n}{\Delta t} + \mathbf{C} \frac{\hat{\boldsymbol{\theta}}^{n+1} - \hat{\boldsymbol{\theta}}^n}{\Delta t} + \partial_x \langle \mu \hat{\mathbf{g}} \rangle^{n+1} = -\hat{\mathbf{h}}^{n+1}, \tag{64a}$$

$$\varepsilon^2 \frac{\hat{\mathbf{g}}^{n+1} - \hat{\mathbf{g}}^n}{\Delta t} + \mu \mathbf{C} \partial_x \hat{\boldsymbol{\theta}}^{n+1} + \varepsilon \partial_x (\mu \hat{\mathbf{g}}^n - \langle \mu \hat{\mathbf{g}} \rangle^n) + \varepsilon^2 \partial_x (\mu \hat{\mathbf{h}}^n) = -\hat{\mathbf{g}}^{n+1}. \tag{64b}$$

Then with space discretization introduced in section 3.5

$$\frac{(\mathbf{C}_i \hat{\boldsymbol{\theta}}_i^{n+1} + \varepsilon^2 \hat{\mathbf{h}}_i^{n+1}) - (\mathbf{C}_i \hat{\boldsymbol{\theta}}_i^{n} + \varepsilon^2 \hat{\mathbf{h}}_i^{n})}{\Delta t} + D^0 \langle \mu \hat{\mathbf{g}}_i^{n+1} \rangle = -\hat{\mathbf{h}}_i^{n+1}, \tag{65a}$$

$$\frac{\hat{\mathbf{g}}_{i+1/2}^{n+1}-\hat{\mathbf{g}}_{i+1/2}^{n}}{\Delta t}+\frac{1}{\varepsilon}(Id-\langle\cdot\rangle)(\mu^{+}D^{-}+\mu^{-}D^{+})\hat{\mathbf{g}}_{i+1/2}^{n}=-\frac{1}{\varepsilon^{2}}\hat{\mathbf{g}}_{i+1/2}^{n+1}-\frac{1}{\varepsilon^{2}}\mu\delta^{0}(\mathbf{C}_{i+1/2}\hat{\boldsymbol{\theta}}_{i+1/2}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_{i+1/2}^{n}). \tag{65b}$$

Step 1

Multiply (65a) by $(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1}+\varepsilon^2\hat{\mathbf{h}}_i^{n+1})^T$, sum over $i\in\mathbb{Z}$,

$$\begin{aligned}&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}^{n+1}\|^{2}-\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n}+\varepsilon^{2}\hat{\mathbf{h}}^{n}\|^{2}+\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}^{n+1}-\mathbf{C}\hat{\boldsymbol{\theta}}^{n}-\varepsilon^{2}\hat{\mathbf{h}}^{n}\|^{2})\\&+\sum_{i}(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_i^{n+1})^{T}D^{0}\langle\mu\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x=-\sum_{i}(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_i^{n+1})^{T}\hat{\mathbf{h}}_i^{n+1}\Delta x.\end{aligned} \tag{66}$$

Multiply (65b) by $(\hat{\mathbf{g}}_{i+1/2}^{n+1})^T$ and take $\langle\cdot\rangle$, sum over i:

$$\begin{aligned}&\frac{1}{2\Delta t}(\|\|\hat{\mathbf{g}}^{n+1}\|\|^{2}-\|\|\hat{\mathbf{g}}^{n}\|\|^{2}+\|\|\hat{\mathbf{g}}^{n+1}-\hat{\mathbf{g}}^{n}\|\|^{2})+\frac{1}{\varepsilon}\sum_{i}\langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^{T}[(Id-\langle\cdot\rangle)(\mu^{+}D^{-}+\mu^{-}D^{+})\hat{\mathbf{g}}_{i+1/2}^{n}]\rangle\Delta x\\&=-\frac{1}{\varepsilon^{2}}\|\|\hat{\mathbf{g}}^{n+1}\|\|^{2}-\frac{1}{\varepsilon^{2}}\sum_{i}\langle\mu\hat{\mathbf{g}}_{i+1/2}^{n+1}\rangle^{T}\delta^{0}(\mathbf{C}_{i+1/2}\hat{\boldsymbol{\theta}}_{i+1/2}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_{i+1/2}^{n})\Delta x.\end{aligned} \tag{67}$$

Use $\langle\hat{\mathbf{g}}_{i+1/2}^{n+1}\rangle=0$ for every i,

$$\sum_{i}\langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^{T}[(Id-\langle\cdot\rangle)(\mu^{+}D^{-}+\mu^{-}D^{+})\hat{\mathbf{g}}_{i+1/2}^{n}]\rangle\Delta x=\sum_{i}\langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^{T}((\mu^{+}D^{-}+\mu^{-}D^{+})\hat{\mathbf{g}}_{i+1/2}^{n})\rangle\Delta x. \tag{68}$$

Then (66) + $\varepsilon^2\times$ (67) gives

$$\begin{aligned}&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}^{n+1}\|^{2}-\|\mathbf{C}^{n}\hat{\boldsymbol{\theta}}^{n}+\varepsilon^{2}\hat{\mathbf{h}}^{n}\|^{2}+\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}^{n+1}-\mathbf{C}\hat{\boldsymbol{\theta}}^{n}-\varepsilon^{2}\hat{\mathbf{h}}^{n}\|^{2})\\&+\sum_{i}(\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}^{n+1})^{T}D^{0}\langle\mu\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x+\frac{\varepsilon^{2}}{2\Delta t}(\|\|\hat{\mathbf{g}}^{n+1}\|\|^{2}-\|\|\hat{\mathbf{g}}^{n}\|\|^{2}+\|\|\hat{\mathbf{g}}^{n+1}-\hat{\mathbf{g}}^{n}\|\|^{2})\\&+\varepsilon\sum_{i}\langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^{T}((\mu^{+}D^{-}+\mu^{-}D^{+})\hat{\mathbf{g}}_{i+1/2}^{n})\rangle\Delta x=-\sum_{i}(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_i^{n+1})^{T}\hat{\mathbf{h}}_i^{n+1}\Delta x\\&-\|\|\hat{\mathbf{g}}^{n+1}\|\|^{2}-\sum_{i}\langle\mu\hat{\mathbf{g}}_{i+1/2}^{n+1}\rangle^{T}\delta^{0}(\mathbf{C}_{i+1/2}\hat{\boldsymbol{\theta}}_{i+1/2}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_{i+1/2}^{n})\Delta x\end{aligned} \tag{69}$$

Use discrete integration by parts:

$$\sum_{i}\langle\mu\hat{\mathbf{g}}_{i+1/2}^{n+1}\rangle^{T}\delta^{0}(\mathbf{C}_{i+1/2}\hat{\boldsymbol{\theta}}_{i+1/2}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_{i+1/2}^{n})\Delta x=-\sum_{i}\langle\mu D^{0}\hat{\mathbf{g}}_i^{n+1}\rangle^{T}(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_i^{n})\Delta x$$

Step 2

Rewrite (69) as following:

$$\begin{aligned}&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}^{n+1}\|^{2}+\|\|\varepsilon\hat{\mathbf{g}}^{n+1}\|\|^{2}-\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n}+\varepsilon^{2}\hat{\mathbf{h}}^{n}\|^{2}-\|\|\varepsilon\hat{\mathbf{g}}^{n}\|\|^{2}\\&+\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}^{n+1}-\mathbf{C}\hat{\boldsymbol{\theta}}^{n}-\varepsilon^{2}\hat{\mathbf{h}}^{n}\|^{2}+\|\|\varepsilon\hat{\mathbf{g}}^{n+1}-\varepsilon\hat{\mathbf{g}}^{n}\|\|^{2})\\&+\varepsilon\sum_{i}\langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^{T}((\mu^{+}D^{-}+\mu^{-}D^{+})\hat{\mathbf{g}}_{i+1/2}^{n})\rangle\Delta x\\=&-\sum_{i}(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1}+\varepsilon^{2}\hat{\mathbf{h}}_i^{n+1})^{T}\hat{\mathbf{h}}_i^{n+1}\Delta x-\|\|\hat{\mathbf{g}}^{n+1}\|\|^{2}+\sum_{i}\langle\mu D^{0}\hat{\mathbf{g}}_i^{n+1}\rangle^{T}(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1}-\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n})\Delta x\\&+\sum_{i}\langle\mu D^{0}\hat{\mathbf{g}}_i^{n+1}\rangle^{T}(-\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1}-\varepsilon^{2}\hat{\mathbf{h}}_i^{n+1}+\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n}+\varepsilon^{2}\hat{\mathbf{h}}_i^{n})\Delta x.\end{aligned} \tag{70}$$

Use Young's inequality:

$$\begin{aligned}
&\sum_i \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T (-\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1} - \varepsilon^2\hat{\mathbf{h}}_i^{n+1} + \mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n} + \varepsilon^2\hat{\mathbf{h}}_i^{n})\Delta x \\
\le& \alpha \| -\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1} - \varepsilon^2\hat{\mathbf{h}}^{n+1} + \mathbf{C}\hat{\boldsymbol{\theta}}^{n} + \varepsilon^2\hat{\mathbf{h}}^{n}\|^2 + \frac{1}{4\alpha}\sum_i \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle \Delta x.
\end{aligned}$$

Thus, let $\alpha = \frac{1}{2\Delta t}$,we have

$$\begin{aligned}
&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1} + \varepsilon^2\hat{\mathbf{h}}^{n+1}\|^2 + \||\varepsilon\hat{\mathbf{g}}^{n+1}\||^2 - \|\mathbf{C}\hat{\boldsymbol{\theta}}^{n} + \varepsilon^2\hat{\mathbf{h}}^{n}\|^2 - \||\varepsilon\hat{\mathbf{g}}^{n}\||^2 \\
&+ \||\varepsilon\hat{\mathbf{g}}^{n+1} - \varepsilon\hat{\mathbf{g}}^{n}\||^2) + \varepsilon\sum_i \langle (\hat{\mathbf{g}}_{i+1/2}^{n+1})^T((\mu^+D^- + \mu^-D^+)\hat{\mathbf{g}}_{i+1/2}^{n})\rangle \Delta x \\
\le& -\sum_i (\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1} + \varepsilon^2\hat{\mathbf{h}}_i^{n+1})^T\hat{\mathbf{h}}_i^{n+1}\Delta x - \||\hat{\mathbf{g}}^{n+1}\||^2 \\
&+ \sum_i \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1} - \mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n})\Delta x + \frac{\Delta t}{2}\sum_i \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle \Delta x.
\end{aligned} \tag{71}$$

Multiply (42c) by $(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1})^T$, sum over $i \in \mathbb{Z}$,

$$\frac{1}{\Delta t}\sum_i (\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1})^T(\hat{\boldsymbol{\theta}}_i^{n+1} - \hat{\boldsymbol{\theta}}_i^{n})\Delta x = \sum_i (\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1})^T(\Delta^c\hat{\boldsymbol{\theta}}_i^{n+1} + \hat{\mathbf{h}}_i^{n+1})\Delta x. \tag{72}$$

Since $\mathbf{C}_i$ is symmetric, positive and definite, one can use the Cholesky decomposition

$$\mathbf{C}_i = L_i(L_i)^T,$$

where L_i is a lower triangular matrix with positive diagonal entries. Then (72) becomes

$$\frac{1}{2\Delta t}(\|L\hat{\boldsymbol{\theta}}^{n+1}\|^2 - \|L\hat{\boldsymbol{\theta}}^{n}\|^2 + \|L\hat{\boldsymbol{\theta}}^{n+1} - L\hat{\boldsymbol{\theta}}^{n}\|^2) = \sum_i (\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1})^T(\Delta^c\hat{\boldsymbol{\theta}}_i^{n+1} + \hat{\mathbf{h}}_i^{n+1})\Delta x. \tag{73}$$

Adding equation (71) and (73),

$$\begin{aligned}
&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1} + \varepsilon^2\hat{\mathbf{h}}^{n+1}\|^2 + \||\varepsilon\hat{\mathbf{g}}^{n+1}\||^2 + \|L\hat{\boldsymbol{\theta}}^{n+1}\|^2 - \|\mathbf{C}\hat{\boldsymbol{\theta}}^{n} + \varepsilon^2\hat{\mathbf{h}}^{n}\|^2 \\
&- \||\varepsilon\hat{\mathbf{g}}^{n}\||^2 - \|L\hat{\boldsymbol{\theta}}^{n}\|^2 + \||\varepsilon\hat{\mathbf{g}}^{n+1} - \varepsilon\hat{\mathbf{g}}^{n}\||^2 + \|L\hat{\boldsymbol{\theta}}^{n+1} - L\hat{\boldsymbol{\theta}}^{n}\|^2) \\
&+ \varepsilon\sum_i \langle (\hat{\mathbf{g}}_{i+1/2}^{n+1})^T((\mu^+D^- + \mu^-D^+)\hat{\mathbf{g}}_{i+1/2}^{n})\rangle \Delta x \\
\le& -\||\hat{\mathbf{g}}^{n+1}\||^2 - \|\varepsilon\hat{\mathbf{h}}^{n+1}\|^2 + \sum_i \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1} - \mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n})\Delta x \\
&+ \frac{\Delta t}{2}\sum_i \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle \Delta x + \sum_i (\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1})^T\Delta^c\hat{\boldsymbol{\theta}}^{n+1}\Delta x.
\end{aligned} \tag{74}$$

Denote $\boldsymbol{\phi}_{i+1/2}^{n+1} = \frac{\hat{\boldsymbol{\theta}}_{i+1}^{n+1} - \hat{\boldsymbol{\theta}}_i^{n+1}}{\Delta x}$, then $\Delta^c\hat{\boldsymbol{\theta}}_i^{n+1} = D^0\boldsymbol{\phi}_i^{n+1}$, and

$$\begin{aligned}
&\sum_i (\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1})^T\Delta^c\hat{\boldsymbol{\theta}}_i^{n+1}\Delta x \\
=& \sum_i (\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1})^T D^0\boldsymbol{\phi}_i^{n+1}\Delta x \\
=& -\sum_i (\boldsymbol{\phi}_i^{n+1})^T\mathbf{C}_i\boldsymbol{\phi}_i^{n+1}\Delta x = -\|L\boldsymbol{\phi}^{n+1}\|^2 \le 0.
\end{aligned} \tag{75}$$

Using Young's Inequality,

$$\begin{aligned}
&\sum_i \langle \mu D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T(\mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n+1} - \mathbf{C}_i\hat{\boldsymbol{\theta}}_i^{n})\Delta x = \sum_i \langle \mu L_i D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T(L_i\hat{\boldsymbol{\theta}}_i^{n+1} - L_i\hat{\boldsymbol{\theta}}_i^{n})\Delta x \\
\le& \alpha\sum_i \langle \mu L_i D^0 \hat{\mathbf{g}}_i^{n+1}\rangle^T \langle \mu L_i D^0 \hat{\mathbf{g}}_i^{n+1}\rangle \Delta x + \frac{1}{4\alpha}\|L\hat{\boldsymbol{\theta}}^{n+1} - L\hat{\boldsymbol{\theta}}^{n}\|^2.
\end{aligned} \tag{76}$$

Let $\alpha = \frac{\Delta t}{2}$, then (76) becomes

$$\begin{aligned}&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1} + \varepsilon^2\hat{\mathbf{h}}^{n+1}\|^2 + \||\varepsilon\hat{\mathbf{g}}^{n+1}\||^2 + \|L\hat{\boldsymbol{\theta}}^{n+1}\|^2 - \|\mathbf{C}\hat{\boldsymbol{\theta}}^{n} + \varepsilon^2\hat{\mathbf{h}}^{n}\|^2 - \||\varepsilon\hat{\mathbf{g}}^{n}\||^2\\ &- \|L\hat{\boldsymbol{\theta}}^{n}\|^2 + \||\varepsilon\hat{\mathbf{g}}^{n+1} - \varepsilon\hat{\mathbf{g}}^{n}\||^2) + \varepsilon\sum_i \langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^T((\mu^+D^- + \mu^-D^+)\hat{\mathbf{g}}_{i+1/2}^{n})\rangle\Delta x\\ \le &-\||\hat{\mathbf{g}}^{n+1}\||^2 + \frac{\Delta t}{2}\sum_i \langle\mu L_i D^0\hat{\mathbf{g}}_i^{n+1}\rangle^T\langle\mu L_i D^0\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x + \frac{\Delta t}{2}\sum_i\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle^T\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x.\end{aligned} \tag{77}$$

Denote $\lambda_i > 0$ the largest eigenvalue of $\mathbf{C}_i$. Then

$$\begin{aligned}\sum_i\langle\mu L_i D^0\hat{\mathbf{g}}_i^{n+1}\rangle^T\langle\mu L_i D^0\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x &\le \sum_i \lambda_i\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle^T\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x\\ &\le \lambda_0\sum_i\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle^T\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x.\end{aligned}$$

Thus

$$\begin{aligned}&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1} + \varepsilon^2\hat{\mathbf{h}}^{n+1}\|^2 + \||\varepsilon\hat{\mathbf{g}}^{n+1}\||^2 + \|L\hat{\boldsymbol{\theta}}^{n+1}\|^2 - \|\mathbf{C}\hat{\boldsymbol{\theta}}^{n} + \varepsilon^2\hat{\mathbf{h}}^{n}\|^2 - \||\varepsilon\hat{\mathbf{g}}^{n}\||^2 - \|L\hat{\boldsymbol{\theta}}^{n}\|^2\\ &+ \||\varepsilon\hat{\mathbf{g}}^{n+1} - \varepsilon\hat{\mathbf{g}}^{n}\||^2) + \varepsilon\sum_i \langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^T((\mu^+D^- + \mu^-D^+)\hat{\mathbf{g}}_{i+1/2}^{n})\rangle\Delta x\\ \le &-\||\hat{\mathbf{g}}^{n+1}\||^2 + \frac{\Delta t}{2}(\lambda_0+1)\sum_i\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle^T\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x.\end{aligned} \tag{78}$$

Step 3

Now we separate the last term on the left hand side of (78) into two parts:

$$\begin{aligned}&\sum_i\langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^T((\mu^+D^- + \mu^-D^+)\hat{\mathbf{g}}_{i+1/2}^{n})\rangle\Delta x\\ =&\sum_i\langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^T((\mu^+D^- + \mu^-D^+)\hat{\mathbf{g}}_{i+1/2}^{n+1})\rangle\Delta x + \sum_i\langle(\hat{\mathbf{g}}_{i+1/2}^{n+1})^T(\mu^+D^- + \mu^-D^+)(\hat{\mathbf{g}}_{i+1/2}^{n} - \hat{\mathbf{g}}_{i+1/2}^{n+1})\rangle\Delta x\\ =&A + B\end{aligned} \tag{79}$$

where

$$\begin{aligned}A &= \sum_i\langle\mu(\hat{\mathbf{g}}_{i+1/2}^{n+1})^T D^c\hat{\mathbf{g}}_{i+1/2}^{n+1}\rangle\Delta x - \frac{\Delta x}{2}\sum_i\langle(|\mu|\hat{\mathbf{g}}_{i+1/2}^{n+1})^T D^-D^+\hat{\mathbf{g}}_{i+1/2}^{n+1}\rangle\Delta x\\ &= \frac{\Delta x}{2}\sum_i\langle|\mu|(D^+\hat{\mathbf{g}}_{i+1/2}^{n+1})^T D^+\hat{\mathbf{g}}_{i+1/2}^{n+1}\rangle\Delta x,\end{aligned} \tag{80}$$

$$B = -\sum_i\langle[(\mu^+D^+ + \mu^-D^-)\hat{\mathbf{g}}_{i+1/2}^{n+1}]^T(\hat{\mathbf{g}}_{i+1/2}^{n} - \hat{\mathbf{g}}_{i+1/2}^{n+1})\rangle\Delta x, \tag{81}$$

and

$$|B| \le \alpha\|\hat{\mathbf{g}}^{n+1} - \hat{\mathbf{g}}^{n}\|^2 + \frac{1}{4\alpha}\||\mu|D^+\hat{\mathbf{g}}^{n+1}\|^2. \tag{82}$$

If $\alpha = \frac{\varepsilon}{2\Delta t}$, $\|\hat{\mathbf{g}}^{n+1} - \hat{\mathbf{g}}^{n}\|^2$ can be canceled out and one gets

$$\begin{aligned}&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1} + \varepsilon^2\hat{\mathbf{h}}^{n+1}\|^2 + \||\varepsilon\hat{\mathbf{g}}^{n+1}\||^2 + \|L\hat{\boldsymbol{\theta}}^{n+1}\|^2 - \|\mathbf{C}\hat{\boldsymbol{\theta}}^{n} + \varepsilon^2\hat{\mathbf{h}}^{n}\|^2 - \||\varepsilon\hat{\mathbf{g}}^{n}\||^2 - \|L\hat{\boldsymbol{\theta}}^{n}\|^2)\\ &+ \varepsilon\frac{\Delta x}{2}\sum_i\langle|\mu|(D^+\hat{\mathbf{g}}_{i+1/2}^{n+1})^T D^+\hat{\mathbf{g}}_{i+1/2}^{n+1}\rangle\Delta x - \frac{\Delta t}{2}\||\mu|D^+\hat{\mathbf{g}}^{n+1}\|^2\\ \le &-\||\hat{\mathbf{g}}^{n+1}\||^2 + \frac{\Delta t}{2}(\lambda_0+1)\sum_i\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle^T\langle\mu D^0\hat{\mathbf{g}}_i^{n+1}\rangle\Delta x.\end{aligned} \tag{83}$$

Step 4

The last term on the left hand side of (83) is

$$\begin{aligned}\frac{\Delta t}{2}\||\mu|D^+\hat{\mathbf{g}}^{n+1}\|^2 &= \frac{\Delta t}{2}\sum_i \langle|\mu|^2(D^+\hat{\mathbf{g}}^{n+1}_{i+1/2})^T D^+\hat{\mathbf{g}}^{n+1}_{i+1/2}\rangle\Delta x \\ &\le \frac{\Delta t}{2}\sum_i \langle|\mu|(D^+\hat{\mathbf{g}}^{n+1}_{i+1/2})^T D^+\hat{\mathbf{g}}^{n+1}_{i+1/2}\rangle\Delta x\end{aligned} \tag{84}$$

for $|\mu| \le 1$.

Then the last term on the right hand side of (83)

$$\frac{\Delta t}{2}(\lambda_0+1)\sum_i \langle\mu D^0\hat{\mathbf{g}}^{n+1}_i\rangle^T\langle\mu D^0\hat{\mathbf{g}}^{n+1}_i\rangle\Delta x \le \frac{\Delta t}{4}(\lambda_0+1)\sum_i\langle|\mu|(D^+\hat{\mathbf{g}}^{n+1}_{i+1/2})^T D^+\hat{\mathbf{g}}^{n+1}_{i+1/2}\rangle\Delta x. \tag{85}$$

Step 5

Using the results of (84) and (85) with (83), one gets

$$\begin{aligned}&\frac{1}{2\Delta t}(\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^2\hat{\mathbf{h}}^{n+1}\|^2+\||\varepsilon\hat{\mathbf{g}}^{n+1}\||^2+\|L\hat{\boldsymbol{\theta}}^{n+1}\|^2-\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n}+\varepsilon^2\hat{\mathbf{h}}^{n}\|^2-\||\varepsilon\hat{\mathbf{g}}^{n}\||^2-\|L\hat{\boldsymbol{\theta}}^{n}\|^2)\\ &\le -\||\hat{\mathbf{g}}^{n+1}\||^2+\left[\frac{\Delta t}{4}(\lambda_0+3)-\varepsilon\frac{\Delta x}{2}\right]\sum_i\langle|\mu|(D^+\hat{\mathbf{g}}^{n+1}_i)^T D^+\hat{\mathbf{g}}^{n+1}_i\rangle\Delta x\\ &\le -\||\hat{\mathbf{g}}^{n+1}\||^2+\left[\frac{\Delta t}{4}(\lambda_0+3)-\varepsilon\frac{\Delta x}{2}\right]^+\sum_i\langle(D^+\hat{\mathbf{g}}^{n+1}_i)^T D^+\hat{\mathbf{g}}^{n+1}_i\rangle\Delta x\\ &\le -\||\hat{\mathbf{g}}^{n+1}\||^2+\left[\frac{\Delta t}{4}(\lambda_0+3)-\varepsilon\frac{\Delta x}{2}\right]^+\frac{4}{\Delta x^2}\||\hat{\mathbf{g}}^{n+1}\||^2.\end{aligned} \tag{86}$$

This means that we have the final energy estimate

$$\|\mathbf{C}\hat{\boldsymbol{\theta}}^{n+1}+\varepsilon^2\hat{\mathbf{h}}^{n+1}\|^2+\||\varepsilon\hat{\mathbf{g}}^{n+1}\||^2+\|L\hat{\boldsymbol{\theta}}^{n+1}\|^2 \le \|\mathbf{C}\hat{\boldsymbol{\theta}}^{n}+\varepsilon^2\hat{\mathbf{h}}^{n}\|^2+\||\varepsilon\hat{\mathbf{g}}^{n}\||^2+\|L\hat{\boldsymbol{\theta}}^{n}\|^2, \tag{87}$$

if Δt is such that

$$\left[\frac{\Delta t}{4}(\lambda_0+3)-\varepsilon\frac{\Delta x}{2}\right]\frac{4}{(\Delta x)^2}\le 1.$$

This implies

$$\Delta t \le \frac{1}{3+\lambda_0}((\Delta x)^2+2\varepsilon\Delta x).$$

(87) clearly implies that $\hat{I}^{n+1}$, given by (13), is bounded by the initial data. □

References

[1] E.E. Anderson, Heat transfer in semitransparent solids, Adv. Heat Transf. 11 (1975) 317.
[2] Subrahmanyan Chandrasekhar, Radiative Transfer, Courier Corporation, 2013.
[3] Gautschi Walter, Orthogonal Polynomials: Computation and Approximation, Oxford University Press, 2004.
[4] Jingwei Hu, Shi Jin, A stochastic Galerkin method for the Boltzmann equation with uncertainty, J. Comput. Phys. 315 (2016) 150–168.
[5] Jitesh Jain, Hong Li, Stephen Cauley, Cheng-Kok Koh, Venkataramanan Balakrishnan, Numerically stable algorithms for inversion of block tridiagonal and banded matrices, Purdue ECE technical report (1358), 2007.
[6] Shi Jin, Efficient asymptotic-preserving (ap) schemes for some multiscale kinetic equations, SIAM J. Sci. Comput. 21 (2) (1999) 441–454.
[7] Shi Jin, Asymptotic preserving (ap) schemes for multiscale kinetic and hyperbolic equations: a review, in: Lecture Notes for Summer School on "Methods and Models of Kinetic Theory" (M&MKT), Porto Ercole, Grosseto, Italy, 2010, pp. 177–216.
[8] Jin Shi, David Levermore, The discrete-ordinate method in diffusive regimes, Transp. Theory Stat. Phys. 20 (5–6) (1991) 413–439.
[9] Shi Jin, Jian-Guo Liu, Zheng Ma, A micro–macro decomposition based stochastic asymptotic-preserving scheme for linear transport equations in diffusive regimes with random inputs. Preprint, 2016.
[10] Shi Jin, Liu Liu, An asymptotic-preserving stochastic Galerkin method for the semiconductor Boltzmann equation with random inputs and diffusive scalings, Multiscale Model. Simul. (2016), in press.
[11] Shi Jin, Dongbin Xiu, Xueyu Zhu, Asymptotic-preserving methods for hyperbolic and transport equations with random inputs and diffusive scalings, J. Comput. Phys. 289 (2015) 35–52.
[12] A. Klar, C. Schmeiser, Numerical passage from radiative heat transfer to nonlinear diffusion models, Math. Models Methods Appl. Sci. 11 (05) (2001) 749–767.

[13] Axel Klar, An asymptotic-induced scheme for nonstationary transport equations in the diffusive limit, SIAM J. Numer. Anal. 35 (3) (1998) 1073–1094 (electronic).
[14] Edward W. Larsen, J.E. Morel, Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes. II, J. Comput. Phys. 83 (1) (1989) 212–236.
[15] Edward W. Larsen, J.E. Morel, Warren F. Miller, Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes, J. Comput. Phys. 69 (2) (1987) 283–324.
[16] Mohammed Lemou, Luc Mieussens, A new asymptotic preserving scheme based on micro–macro formulation for linear kinetic equations in the diffusion limit, SIAM J. Sci. Comput. 31 (1) (2008) 334–368.
[17] Jian-Guo Liu, Luc Mieussens, Analysis of an asymptotic preserving scheme for linear kinetic equations in the diffusion limit, SIAM J. Numer. Anal. 48 (4) (2010) 1474–1491.
[18] Michael F. Modest, Radiative Heat Transfer, Academic Press, 2013.
[19] G.C. Pomraning, Initial and boundary conditions for equilibrium diffusion theory, J. Quant. Spectrosc. Radiat. Transf. 36 (1) (1986) 69–84.
[20] Christian Schmeiser, Alexander Zwirchmayr, Convergence of moment methods for linear kinetic equations, SIAM J. Numer. Anal. 36 (1) (1998) 74–88.
[21] Wenjun Sun, Song Jiang, Kun Xu, An asymptotic preserving unified gas kinetic scheme for gray radiative transfer equations, J. Comput. Phys. 285 (2015) 265–279.
[22] Dongbin Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.
[23] Dongbin Xiu, Jan S. Hesthaven, High-order collocation methods for differential equations with random inputs, SIAM J. Sci. Comput. 27 (3) (2005) 1118–1139.
[24] Dongbin Xiu, George Em Karniadakis, The Wiener–Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2) (2002) 619–644.
[25] Dongbin Xiu, Jie Shen, Efficient stochastic Galerkin methods for random diffusion equations, J. Comput. Phys. 228 (2) (2009) 266–281.