Karlsruhe Institute of Technology

wave phenomena
analysis and numerics

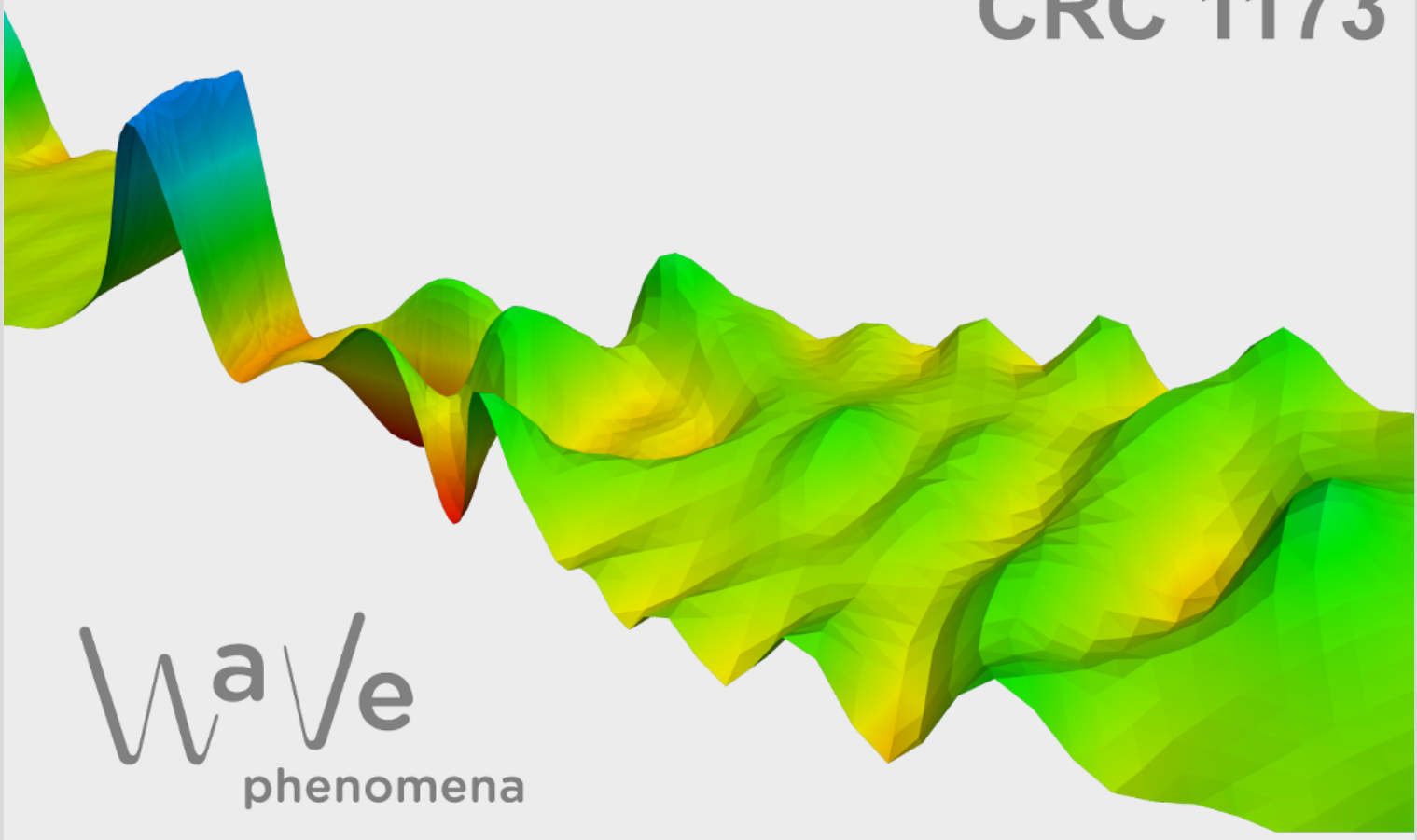
Asymptotic-preserving and energy stable dynamical low-rank approximation for thermal radiative transfer equations

Chinmay Patwardhan, Martin Frank, Jonas Kusch

CRC Preprint 2024/6, February 2024

KARLSRUHE INSTITUTE OF TECHNOLOGY

CRC 1173



Participating universities

Funded by

ISSN 2365-662X

Asymptotic-preserving and energy stable dynamical low-rank approximation for thermal radiative transfer equations

Chinmay Patwardhan[1,*], Martin Frank[1], and Jonas Kusch[2]

[1] *Karlsruhe Institute of Technology, Institute of Applied and Numerical Mathematics, Karlsruhe, Germany*
[2] *Norwegian University of Life Sciences, Scientific Computing, Ås, Norway*

Abstract

The thermal radiative transfer equations model temperature evolution through a background medium as a result of radiation. When a large number of particles are absorbed in a short time scale, the dynamics tend to a non-linear diffusion-type equation called the Rosseland approximation. The main challenges for constructing numerical schemes that exhibit the correct limiting behavior are posed by the solution's high-dimensional phase space and multi-scale effects. In this work, we propose an asymptotic-preserving and rank-adaptive dynamical low-rank approximation scheme based on the macro-micro decomposition of the particle density and a modified augmented basis-update & Galerkin integrator. We show that this scheme, for linear particle emission by the material, dissipates energy over time under a step size restriction that captures the hyperbolic and parabolic CFL conditions. We demonstrate the efficacy of the proposed method in a series of numerical experiments.

Keywords: thermal radiative transfer equations, energy stability, asymptotic-preserving scheme, dynamical low-rank approximation, macro-micro decomposition

1 Introduction

The radiation of particles from a hot source into a cold medium and the corresponding formation of a thermal heat front, known as a Marshak wave, is well modeled by the thermal radiative transfer equations. They consist of a coupled system of partial differential equations governing the transport of particles (represented by the particle density f) and the temperature evolution of the medium (T) [1] given by

$$\frac{1}{c}\partial_t f + \boldsymbol{\Omega}\cdot\nabla_{\mathbf{x}} f = \sigma^a(B(T)-f),$$

$$c_\nu \partial_t T = \int_{\mathbb{S}^2} \sigma^a(f - B(T))\,\mathrm{d}\boldsymbol{\Omega}.$$

The particle density f depends on time t, position $\mathbf{x}$ and direction of flight $\boldsymbol{\Omega} \in \mathbb{S}^2$ and the temperature T on time and position. c and c_ν represent the speed of light and the specific heat of the material, respectively. These two equations are coupled through the absorption and emission of particles by the background material. Numerically simulating the thermal radiative transfer equations poses several challenges. First, to evolve and store the particle density, high memory, and computational resources are required. This is due to its high-dimensional phase space, consisting of temporal, spatial, and angular variables, which can be up to six-dimensional for a three-dimensional

*Corresponding author.
Chinmay Patwardhan: chinmay.patwardhan@kit.edu
Martin Frank: martin.frank@kit.edu
Jonas Kusch: jonas.kusch@nmbu.no

spatial domain. Second, when a large number of these particles are absorbed in small time scales, the dynamics of the system asymptotically converge to a diffusion-type non-linear partial differential equation known as the Rosseland approximation [2] which reads

$$(c_\nu + 4aT^3)\partial_t T = \nabla_{\mathbf{x}} \cdot \left(\frac{4ac}{3\sigma^a} T^3 \nabla_{\mathbf{x}} T \right),$$

where a is the radiation constant. Since the dynamics tend to a diffusion problem, the numerical scheme should also capture this behavior without having to resolve prohibitively small time scales. Numerical methods that do so while efficiently treating the stiffness arising from large absorption terms are called asymptotic-preserving (AP) schemes [3]. Work on asymptotic-preserving schemes for kinetic equations can, for example, be found in [3]–[6].

To address the computational challenges posed by a high-dimensional phase space, we use dynamical low-rank approximation (DLRA) [7], which is a model-order reduction strategy that has recently gained popularity in solving kinetic equations. The fundamental idea behind dynamical low-rank approximation is to evolve the solution on the manifold of rank r functions $\mathcal{M}_r$ by projecting the dynamics to the tangent space of $\mathcal{M}_r$. The evolution equations thus obtained can be interpreted as a Galerkin system, with $2r$ basis functions for the phase space variables, which evolves the coefficients and basis functions according to the dynamics of the problem. Using standard time integrators to evolve the coefficient and basis functions leads to unstable numerical schemes and thus robust integrators, like the projector-splitting integrator (PSI) [8], and the basis-update & Galerkin (BUG) integrators [9]–[12], have been developed.

DLRA has, for example, been shown to reduce computational costs in dose computation in radiation therapy planning [13], in high-scattering problems [14] and neutron criticality problems [15]. In recent works, DLRA was used for the thermal radiative transfer equations [16], [17] where it was shown to significantly reduce computational time. The work in [17] proposes a low-rank scheme based on the augmented BUG integrator of [10] for thermal radiative transfer. Though this scheme is energy-stable and preserves mass locally, it does not include multiscale effects frequently arising in thermal radiative transfer.

Since the thermal radiative transfer equations tend to a diffusion equation, for small time scales, the dynamics are restricted to the manifold of low-rank functions [2] and thus can be accurately represented by a low-rank approximation. Thus, a DLRA scheme combined with techniques to preserve asymptotic behavior can be highly beneficial in tackling both numerical challenges simultaneously. This was investigated for the related problem of radiation transport in [18], where the PSI, along with a macro-micro decomposition to construct an asymptotic-preserving scheme, is used. A key challenge for such schemes is to prove stability and provide a CFL condition that takes into account effects at long time scales (kinetic regime) and small time scales (diffusive regime). In contrast to [18], the scheme constructed in [19] uses the fixed-rank BUG integrator and is energy stable under a CFL restriction, which captures both the kinetic and diffusive regimes. In the kinetic regime, a large number of particles stream in all directions, increasing the rank required to resolve the solution. In contrast, in the diffusive regime, a large number of streaming particles are absorbed and diffused, thus lowering the required rank of the solution [2]. This is well demonstrated in the experiments from [18], [19]. Thus, using a fixed-rank integrator without prior knowledge of the required rank in the regime results in either higher computational costs (due to over-approximation) or a poorly resolved solution (due to under-approximation). One of the ways to address this is by using a DLRA scheme that appropriately chooses and evolves the rank of the solution according to the regime.

This work proposes an asymptotic-preserving and rank-adaptive DLRA scheme for the thermal radiative transfer equations in slab geometry and analyzes its properties. The novelty of this work can be summarized in the following:

- *An asymptotic-preserving, mass conservative and rank-adaptive DLRA integrator:* We propose a new asymptotic-preserving BUG integrator for the thermal radiative transfer equations, based on the macro-micro decomposition [6], [20] of the particle density, the basis

augmentation step from [10] and conservative truncation [21], [22], to capture the underlying dynamics of the system. The proposed algorithm is locally mass conservative.

- *A stability analysis for the proposed asymptotic-preserving DLRA scheme*: We show that the proposed integrator is stable in the energy norm under a CFL restriction that captures both the kinetic and the diffusive regime under linear emission of particles by the background material.

This paper is structured as follows: Following the introduction in Section 1, in Section 2 we review background concepts that are used in this paper and build the modal macro-micro equations for the thermal radiative transfer equations. In Section 3, we present a spatio-temporal discretization for the modal macro-micro equations, which is asymptotic-preserving and stable under a CFL restriction that captures the kinetic and diffusive regime. In Section 4 we present dynamical low-rank integrators for the thermal radiative transfer equations and the stability results. Specifically, in Section 4.1, we present the evolution equations for the modal macro-micro equations using the fixed-rank BUG integrator [9] and prove its stability property. In Section 4.2, we present the asymptotic-preserving BUG integrator and prove the stability of the scheme. Finally, numerical experiments are presented in Section 5.

2 Background

In this section, we review the basic ideas and concepts that are used in this work. The first subsection describes the thermal radiative transfer equations in slab geometry for the gray approximation, its asymptotic limit, and the macro-micro decomposition [6] of the particle density and its angular discretization. In the second subsection, we look at dynamical low-rank approximation [7], the fixed-rank BUG integrator [9] and the augmented BUG integrator [10].

2.1 Thermal radiative transfer equations

In this paper, we consider the dimensionless form of the gray (i.e., frequency-averaged) thermal radiative transfer equations in slab geometry,

$$\frac{\varepsilon^2}{c}\partial_t f + \varepsilon\mu\partial_x f = \sigma^a(B(T) - f), \tag{2.1a}$$

$$\varepsilon^2 c_\nu \partial_t T = \int_{-1}^{1} \sigma^a(f - B(T))\,\mathrm{d}\mu. \tag{2.1b}$$

In the above equations, $f(t, x, \mu)$ represents the particle density (or angular flux) at time $t \in \mathbb{R}^+$, position $x \in D \subset \mathbb{R}$ and direction of flight $\mu \in [-1, 1]$. The temperature of the material is given by $T(t, x)$ and depends on time and position. These are supplemented with initial and boundary conditions, which are later specified according to the problem. $B(T)$ describes the emission of particles by the background material due to blackbody radiation at its current temperature. It is given by the Stefan-Boltzmann law,

$$B(T) = acT^4,$$

where a is the radiation constant and c is the speed of light. The rate of absorption and emission of particles by the background material is specified by the absorption cross-section $\sigma^a(x)$, where we assume that $\sigma^a(x) \geq \sigma_0 > 0$. We denote the integral over μ as $\langle\cdot\rangle_\mu = \int_{-1}^{1} \cdot\,\mathrm{d}\mu$ and thus the scalar flux of the particle density is defined as, $\phi(t, x) = \frac{1}{2}\langle f\rangle_\mu$.

In (2.1) as ε tends to zero, absorption effects dominate the dynamics. A Hilbert expansion of the particle density f yields that, as $\varepsilon \to 0$, the particles are distributed as $B(T)$, i.e., $f = B(T)$,

while the evolution of temperature is given by a diffusion-type non-linear equation known as the Rosseland approximation [2]:

$$c_\nu \partial_t T = \frac{2}{3}\partial_x \left(\frac{1}{\sigma^a} B'(T) \partial_x T \right) - \frac{2}{c} B'(T) \partial_t T, \tag{2.2}$$

where $B'(T) = \frac{d}{dT} B(T)$.

2.1.1 Macro-micro decomposition

The asymptotic analysis of the thermal radiative equations (2.1) shows that multiple time scales are involved in the evolution of temperature and particles. In particular, effects occur at times scales of order $\mathcal{O}(1)$, $\mathcal{O}(\varepsilon)$, and $\mathcal{O}(\varepsilon^2)$ and must be correctly resolved to capture the underlying dynamics of the system. One way to do this is by decomposing the particle density into variables that describe the macroscopic and microscopic effects. This type of decomposition of the particle density is called a macro-micro decomposition and was first proposed in [6]. Since the thermal radiative transfer equations involve three time scales, the particle density is decomposed into macroscopic (B), microscopic (g), and mesoscopic (h) variables. For the thermal radiative transfer equations, this macro-micro ansatz was first proposed in [20]. To be precise, we make the following ansatz for the particle density

$$f(t,x,\mu) = B(T(t,x)) + \varepsilon g(t,x,\mu) + \varepsilon^2 h(t,x), \tag{2.3}$$

where $\langle g \rangle_\mu = 0$. Note that since $\langle g \rangle_\mu = 0$, the total mass of the system is conserved and was used in [23] to construct a numerical scheme that conserves mass in shallow water equations.

Remark 1. The rationale behind calling h the mesoscopic variable instead of the microscopic variable, despite scaling as ε^2 is that it arises as the leading scaled quantity in the decomposition of the macroscopic quantity of radiation transport equation, the scalar flux, and does not depend on the angular variable.

To obtain evolution equations for h, g and T we substitute the macro-micro ansatz (2.3) in the radiative transfer equations (2.1) yielding the following evolution equations

$$\frac{\varepsilon^2}{c}\partial_t h + \frac{\kappa}{c}\sigma^a B'(T) h + \frac{1}{2}\partial_x \langle \mu g \rangle_\mu = -\sigma^a h, \tag{2.4a}$$

$$\frac{\varepsilon^2}{c}\partial_t g + \varepsilon \left(\mathcal{I} - \frac{1}{2}\langle \cdot \rangle_\mu \right)(\mu \partial_x g) + B'(T)\mu \partial_x T + \varepsilon^2 \mu \partial_x h = -\sigma^a g, \tag{2.4b}$$

$$\partial_t T = \kappa \sigma^a h, \tag{2.4c}$$

where we set $\kappa = \frac{2}{c_\nu}$ for ease of presentation. Note that by comparing the $\mathcal{O}(\varepsilon^0)$ terms in all three equations of (2.4) we obtain the Rosseland approximation (2.2) [2], [20].

Initial and boundary conditions It remains to describe the initial and boundary conditions for the macro-micro equations (2.4). Note that we can write the microscopic variable g and mesoscopic variable h as

$$g(t,x,\mu) = \frac{1}{\varepsilon}\left(f(t,x,\mu) - \frac{1}{2}\langle f(t,x,\mu) \rangle_\mu \right), \tag{2.5a}$$

$$h(t,x) = \frac{1}{\varepsilon^2}\left(\frac{1}{2}\langle f(t,x,\mu) \rangle_\mu - B(T)(t,x) \right). \tag{2.5b}$$

Thus, for given initial and boundary conditions of the radiative transfer equations (2.1), we use the above relations to derive the initial and boundary conditions for the macro-micro equations (2.4).

2.1.2 Angular discretization of microscopic variable

The microscopic variable g depends on the direction of flight, μ, and must be discretized in the angular domain. In this work, we use the method of moments or the P_N method [24] to discretize in μ. To obtain the moment equations, let $\{\widetilde{P}_k\}_{k\in\mathbb{N}\cup\{0\}}$ be orthogonal Legendre polynomials with standard $L^2([-1,1])$ norms γ_k, given by $\gamma_k^2 = \frac{2}{2k+1}$. Let $P_k = \widetilde{P}_k/\gamma_k$ denote the k^{th} orthonormal Legendre polynomial satisfying the recurrence relation

$$\mu P_k = a_{k-1}P_{k-1} + a_k P_{k+1}, \qquad a_k = \frac{k+1}{\sqrt{(2k+1)(2k+3)}}.$$

The P_N ansatz for g then reads

$$g(t,x,\mu) \approx g_{\mathrm{P}_N}(t,x,\mu) = \sum_{k=0}^{N} g_k(t,x)P_k(\mu),$$

where g_k is called the k^{th} moment of the system. Since P_k is orthonormal the k^{th} moment is given by $g_k = \langle gP_k\rangle_\mu$. To obtain evolution equations for the moments g_k, $k = 0, 1, \ldots, N$, we multiply (2.4b) by P_k and integrate over $\mu \in [-1,1]$. The evolution equation for the k^{th} moment is then given by

$$\frac{\varepsilon^2}{c}\partial_t g_k + \varepsilon\partial_x\left(a_{k-1}g_{k-1} + a_k g_{k+1}\right) - \varepsilon\frac{\gamma_0\gamma_1}{2}\partial_x g_1\delta_{k0} + \gamma_1\left(B'(T)\partial_x T + \varepsilon^2\partial_x h\right)\delta_{k1} = -\sigma^a g_k.$$

Note that since $\langle g\rangle_\mu = 0$, we get $g_0 = 0$ and we obtain the following system of modal macro-micro equations

$$\frac{\varepsilon^2}{c}\partial_t h + \frac{\kappa}{c}\sigma^a B'(T)h + \frac{\gamma_1}{2}\partial_x g_1 = -\sigma^a h, \tag{2.6a}$$

$$\frac{\varepsilon^2}{c}\partial_t \boldsymbol{g} + \varepsilon\mathbf{A}\partial_x\boldsymbol{g} + \boldsymbol{b}\left(B'(T)\partial_x T + \varepsilon^2\partial_x h\right) = -\sigma^a\boldsymbol{g}, \tag{2.6b}$$

$$\partial_t T = \kappa\sigma^a h, \tag{2.6c}$$

where

$$\boldsymbol{g} = (g_1,\ldots,g_N)^\top \in \mathbb{R}^N, \qquad \mathbf{A} = \begin{bmatrix} 0 & a_1 & & \\ a_1 & 0 & \ddots & \\ & \ddots & \ddots & a_{N-1} \\ & & a_{N-1} & 0 \end{bmatrix} \in \mathbb{R}^{N\times N} \text{ and } \quad \boldsymbol{b} = (\gamma_1, 0, \ldots, 0)^\top \in \mathbb{R}^N.$$

2.2 Dynamical low-rank approximation

In this subsection, we give an overview of the DLRA put forth in [7]. The fundamental motivation behind DLRA is to evolve a solution on a low-rank manifold of a given rank. To make this more concrete, let $g_{ik} = g_k(t,x_i)$ be the evaluation of the k^{th} moment of the microscopic variable g_k at spatial point x_i. The goal is then to evolve $\mathbf{g}$ such that it stays on the manifold of rank r matrices, $\mathcal{M}_r$. In DLRA, a low-rank approximation is computed by projecting the dynamics of the problem onto the tangent space of the manifold [7].

For any matrix $\mathbf{g}_r \in \mathcal{M}_r \subset \mathbb{R}^{N_x\times N}$ we have the factorization

$$\mathbf{g}_r(t) = \mathbf{X}(t)\mathbf{S}(t)\mathbf{V}(t)^\top. \tag{2.7}$$

This means that the solution matrix is spanned by the spatial basis $\mathbf{X} : \mathbb{R}^+ \to \mathbb{R}^{N_x\times r}$, the moment basis $\mathbf{V} : \mathbb{R}^+ \to \mathbb{R}^{N\times r}$, and the coefficient matrix $\mathbf{S} : \mathbb{R}^+ \to \mathbb{R}^{r\times r}$. In DLRA, the basis and

coefficient matrices are evolved on the low-rank manifold such that, for $\mathbf{g}_r(t) \in \mathcal{M}_r$ and a given right-hand side $\mathbf{F} : \mathbb{R}^{N_x \times N} \to \mathbb{R}^{N_x \times N}$, the following minimization problem is satisfied at all times t

$$\min_{\dot{\mathbf{g}}_r(t) \in \mathcal{T}_{\mathbf{g}_r(t)}\mathcal{M}_r} \|\dot{\mathbf{g}}_r(t) - \mathbf{F}(\mathbf{g}_r(t))\|_F .$$

Here $\mathcal{T}_{\mathbf{g}_r(t)}\mathcal{M}_r$ denotes the tangent space of $\mathcal{M}_r$ at $\mathbf{g}_r$. A reformulation of this minimization problem [7, Lemma 4.1] projects the right-hand side onto the tangent space and requires solving

$$\dot{\mathbf{g}}_r(t) = \mathbf{P}(\mathbf{g}_r(t))\mathbf{F}(\mathbf{g}_r(t))$$

where

$$\mathbf{P}(\mathbf{g}_r)\mathbf{Z} = \mathbf{X}\mathbf{X}^\top\mathbf{Z} - \mathbf{X}\mathbf{X}^\top\mathbf{Z}\mathbf{V}\mathbf{V}^\top + \mathbf{Z}\mathbf{V}\mathbf{V}^\top$$

is the projection onto the tangent space $\mathcal{T}_{\mathbf{g}_r(t)}\mathcal{M}_r$.

Following [7], evolution equations can be derived for the factorized solution from the above equations as

$$\begin{aligned}
\dot{\mathbf{S}}(t) &= \mathbf{X}(t)^\top\mathbf{F}(\mathbf{g}_r(t))\mathbf{V}(t),\\
\dot{\mathbf{X}}(t) &= (\mathbf{I} - \mathbf{X}(t)\mathbf{X}(t)^\top)\mathbf{F}(\mathbf{g}_r(t))\mathbf{V}(t)\mathbf{S}(t)^{-1},\\
\dot{\mathbf{V}}(t) &= (\mathbf{I} - \mathbf{V}(t)\mathbf{V}(t)^\top)\mathbf{F}(\mathbf{g}_r(t))^\top\mathbf{X}(t)\mathbf{S}(t)^{-\top}.
\end{aligned}$$

In case of over-approximation of the rank, the coefficient matrix $\mathbf{S}$ becomes nearly singular which is a source of instabilities. Thus, robust integrators that do not invert the coefficient matrix have been developed [8]–[11]. In this work, we use the fixed-rank BUG [9] and the augmented BUG integrator [10], which we describe here in brief.

For a given factorized initial solution, $\mathbf{g}_r^0 = \mathbf{X}^0\mathbf{S}^0\mathbf{V}^{0,\top}$, one step of the fixed-rank BUG integrator updates the factors $\mathbf{X}, \mathbf{S}, \mathbf{V}$ from time t_0 to t_1 by the following sub-steps

K-step Update $\mathbf{X}^0$ to $\mathbf{X}^1$ by solving

$$\dot{\mathbf{K}}(t) = \mathbf{F}(\mathbf{K}(t)\mathbf{V}^{0,\top})\mathbf{V}^0, \qquad \mathbf{K}(t_0) = \mathbf{X}^0\mathbf{S}^0.$$

Compute $\mathbf{K}(t_1) = \mathbf{X}^1\mathbf{S}^K$, e.g. by using QR decomposition, and store $\mathbf{M} = \mathbf{X}^{1,\top}\mathbf{X}^0$.

L-step Update $\mathbf{V}^0$ to $\mathbf{V}^1$ by solving

$$\dot{\mathbf{L}}(t) = \mathbf{X}^{0,\top}\mathbf{F}(\mathbf{X}^0\mathbf{L}(t)), \qquad \mathbf{L}(t_0) = \mathbf{S}^0\mathbf{V}^{0,\top}.$$

Compute $\mathbf{L}(t_1)^\top = \mathbf{V}^1\mathbf{S}^{L,\top}$, e.g. by using QR decomposition, and store $\mathbf{N} = \mathbf{V}^{1,\top}\mathbf{V}^0$.

S-step Update the coefficient matrix $\mathbf{S}^0$ to $\mathbf{S}^1$ by performing Galerkin step in the updated basis

$$\dot{\mathbf{S}}(t) = \mathbf{X}^{1,\top}\mathbf{F}(\mathbf{X}^1\mathbf{S}(t)\mathbf{V}^{1,\top})\mathbf{V}^1, \qquad \mathbf{S}(t_0) = \mathbf{M}\mathbf{S}^0\mathbf{N}^\top$$

and set $\mathbf{S}^1 = \mathbf{S}(t_1)$.

Then the approximation at the next time step is set as $\mathbf{g}_r(t_1) = \mathbf{X}^1\mathbf{S}^1\mathbf{V}^{1,\top}$.

Using a fixed-rank integrator comes with several challenges. First, since the rank of the solution is not known beforehand, it is usually over-approximated, which leads to increased computational costs. Second, the rank of the solution may vary over time [10], [13], [17]; thus, a fixed-rank integrator may not capture the solution correctly. Moreover, the fixed-rank BUG integrator does not preserve solution invariances. To overcome these, a rank-adaptive extension of the fixed-rank BUG integrator, known as the augmented BUG integrator, was presented in [10] that appends extra spatial and angular basis vectors by reusing the old basis and truncates the rank to a prescribed tolerance ϑ.

To present the algorithm, we denote all quantities of rank $2r$ with hats and those of rank r without. Then, one step of the augmented BUG integrator updates the solution, $\mathbf{g}_r^0 = \mathbf{X}^0\mathbf{S}^0\mathbf{V}^{0,\top}$ of rank r (note that $\mathbf{g}_r$ represents low-rank approximation and not an approximation of rank r), from time t_0 to t_1 through the following steps

1. Update and expand the spatial and angular basis in parallel.

K-step Solve

$$\dot{\mathbf{K}}(t) = \mathbf{F}(\mathbf{K}(t)\mathbf{V}^{0,\top})\mathbf{V}^0, \qquad \mathbf{K}(t_0) = \mathbf{X}^0\mathbf{S}^0,$$

and compute the updated basis matrix $\widehat{\mathbf{X}} \in \mathbb{R}^{N_x \times 2r}$ as an orthonormal basis of $\left[\mathbf{K}(t_1), \mathbf{X}^0\right]$ and store $\widehat{\mathbf{M}} = \widehat{\mathbf{X}}^\top \mathbf{X}^0 \in \mathbb{R}^{2r\times r}$.

L-step Solve

$$\dot{\mathbf{L}}(t) = \mathbf{X}^{0,\top}\mathbf{F}(\mathbf{X}^0\mathbf{L}(t)), \qquad \mathbf{L}(t_0) = \mathbf{S}^0\mathbf{V}^{0,\top},$$

and compute the updated basis matrix $\widehat{\mathbf{V}} \in \mathbb{R}^{\mathbb{N}\times 2r}$ as an orthonormal basis of $\left[\mathbf{L}(t_1)^\top, \mathbf{V}^0\right]$ and store $\widehat{\mathbf{N}} = \widehat{\mathbf{V}}^\top \mathbf{V}^0 \in \mathbb{R}^{2r\times r}$.

2. Update the coefficient matrix $\mathbf{S}^0$ to $\widehat{\mathbf{S}}$ by performing Galerkin step in the updated and expanded basis

$$\dot{\widehat{\mathbf{S}}}(t) = \widehat{\mathbf{X}}^{1,\top}\mathbf{F}(\widehat{\mathbf{X}}\widehat{\mathbf{S}}(t)\widehat{\mathbf{V}}^\top)\widehat{\mathbf{V}}, \qquad \widehat{\mathbf{S}}(t_0) = \widehat{\mathbf{M}}\mathbf{S}^0\widehat{\mathbf{N}}^\top.$$

3. Truncation to new rank r_1.
Compute the SVD decomposition of $\widehat{\mathbf{S}}$

$$\widehat{\mathbf{S}} = \mathbf{P}\mathbf{\Sigma}\mathbf{Q}^\top,$$

where $\mathbf{P}, \mathbf{Q} \in \mathbb{R}^{2r\times 2r}$ are orthogonal matrices and $\mathbf{\Sigma} \in \mathbb{R}^{2r\times 2r}$ is a diagonal matrix with singular values, $\hat{\sigma}_1, \ldots, \hat{\sigma}_{2r}$. The new rank r_1 is chosen as $1 \le r_1 \le 2r$ such that, for some user-defined ϑ, the following is satisfied:

$$\left(\sum_{i=r_1+1}^{2r} \hat{\sigma}_i^2\right)^{1/2} \le \vartheta.$$

To set the updated factors, we define $\mathbf{P}_{r_1}$ and $\mathbf{Q}_{r_1}$ to be the matrices containing the first r_1 columns of $\mathbf{P}$ and $\mathbf{Q}$, respectively. $\mathbf{\Sigma}_{r_1\times r_1}$ is set as the diagonal matrix containing the first r_1 singular values of $\widehat{\mathbf{S}}$. Then the updated factors are set as $\mathbf{X}^1 = \widehat{\mathbf{X}}\mathbf{P}_{r_1}$, $\mathbf{V}^1 = \widehat{\mathbf{V}}\mathbf{Q}_{r_1}$ and $\mathbf{S}^1 = \mathbf{\Sigma}_{r_1\times r_1}$ and, the approximation at time t_1 is then $\mathbf{g}_r^1 = \mathbf{X}^1\mathbf{S}^1\mathbf{V}^{1,\top}$.

Remark 2. Note that often in practice, to truncate the rank, a relative tolerance of the form $\vartheta \cdot \|\mathbf{\Sigma}\|_2$ is used.

3 Energy stability of modal macro-micro equations

Having discretized the macro-micro equations in angle (2.6), in this section, we present an asymptotic-preserving spatio-temporal discretization and investigate its energy stability.

3.1 Spatio-temporal discretization

We start by noting that for $i, j \in \{1, \ldots, N\}$, we can represent the $(i,j)^{\text{th}}$ term of the flux matrix $\mathbf{A}$ by a quadrature rule. That is,

$$\mathrm{A}_{ij} = \langle \mu P_i P_j \rangle_\mu = \int_{-1}^{1} \mu P_i(\mu) P_j(\mu)\, \mathrm{d}\mu \approx \sum_{k=1}^{N+1} w_k \mu_k P_i(\mu_k) P_j(\mu_k),$$

where $(\mu_k)_{k=1,\ldots,N+1}$ and $(w_k)_{k=1,\ldots,N+1}$ are quadrature points and weights given by the Gauss-Legendre quadrature rule. If we define the matrices $\mathbf{T} \in \mathbb{R}^{N\times(N+1)}$, with $T_{ik} = \sqrt{w_k} P_i(\mu_k)$,

and $\mathbf{M} \in \mathbb{R}^{(N+1)\times(N+1)}$, with $\mathrm{M}_{ij} = \mu_i \delta_{ij}$, then we can write the flux matrix as $\mathbf{A} = \mathbf{T}\mathbf{M}\mathbf{T}^\top$. Given $(|\mathbf{M}|)_{ij} = |\mathrm{M}_{ij}|$ we can define a stabilization matrix for a finite volume discretization as $|\mathbf{A}| = \mathbf{T}\,|\mathbf{M}|\,\mathbf{T}^\top$ and $\mathbf{A}^\pm = \frac{1}{2}\mathbf{T}(\mathbf{M} \pm |\mathbf{M}|)\mathbf{T}^\top$.

Remark 3. The choice of the stabilization matrix used here is not the standard Roe matrix $\widetilde{\mathbf{T}}\left|\widetilde{\mathbf{M}}\right|\widetilde{\mathbf{T}}^\top$ where $\mathbf{A} = \widetilde{\mathbf{T}}\widetilde{\mathbf{M}}\widetilde{\mathbf{T}}^\top$ is the eigendecomposition of the flux matrix. That is, the columns of $\widetilde{\mathbf{T}} \in \mathbb{R}^{N\times N}$ consist of orthogonal eigenvectors of $\mathbf{A}$ and $\widetilde{\mathbf{M}} \in \mathbb{R}^{N\times N}$ has the corresponding eigenvalues on the diagonal. The factorization of the flux matrix used, consisting of transformation matrices in $\mathbb{R}^{N\times(N+1)}$, is needed for the diagonalization of the modal scheme for showing stability in the energy norm. This choice of stabilization matrix was first presented in [19] for the radiative transport equation.

We discretize in space using an equidistant staggered grid, with $\Delta x = 1/N_x$, for a given number of $N_x \in \mathbb{N}$ spatial cells. The cell interfaces are given by $x_{1/2}, \dots, x_{N_x+1/2}$ and the midpoints by x_i for $i \in \{1, \dots, N_x\}$. The temperature (T) and mesoscopic variable (h) are resolved at the full grid points x_i whereas the microscopic variable ($\boldsymbol{g}$) is evaluated at the cell interfaces $x_{i+1/2}$.
Since $B(T) = acT^4$ we get $B'(T) = 4acT^3$ and thus, to simplify the presentation of the algorithm, we define

$$\Psi(t,x) = 4(T(t,x))^3.$$

The values of Ψ at x_i and $x_{i+1/2}$ are given by $\Psi_i^n = \Psi(t^n, x_i)$ and $\Psi_{i+1/2}^n = \frac{\Psi_{i+1}^n + \Psi_i^n}{2}$, respectively. Finally, to discretize in time we employ a forward-backward Euler scheme and obtain the following modal macro-micro scheme

$$\frac{\varepsilon^2}{c}\left(\frac{h_i^{n+1} - h_i^n}{\Delta t}\right) + a\kappa\,\sigma_i^a \Psi_i^n h_i^{n+1} + \frac{\gamma_1}{2}\mathcal{D}^0 g_{1,i}^{n+1} = -\sigma_i^a h_i^{n+1}, \tag{3.1a}$$

$$\frac{\varepsilon^2}{c}\left(\frac{\boldsymbol{g}_{i+1/2}^{n+1} - \boldsymbol{g}_{i+1/2}^n}{\Delta t}\right) + \varepsilon\mathcal{L}\boldsymbol{g}_{i+1/2}^n + \boldsymbol{b}\delta^0\left(ac\,\Psi_{i+1/2}^n T_{i+1/2}^n + \varepsilon^2 h_{i+1/2}^n\right) = -\sigma_{i+1/2}^a \boldsymbol{g}_{i+1/2}^{n+1}, \tag{3.1b}$$

$$\frac{T_i^{n+1} - T_i^n}{\Delta t} = \kappa\,\sigma_i^a h_i^{n+1}, \tag{3.1c}$$

where,

$$\mathcal{D}^-\boldsymbol{g}_{i+1/2} = \frac{\boldsymbol{g}_{i+1/2} - \boldsymbol{g}_{i-1/2}}{\Delta x}, \qquad \mathcal{D}^+\boldsymbol{g}_{i+1/2} = \frac{\boldsymbol{g}_{i+3/2} - \boldsymbol{g}_{i+1/2}}{\Delta x},$$
$$\mathcal{D}^0\boldsymbol{g}_i = \frac{\boldsymbol{g}_{i+1/2} - \boldsymbol{g}_{i-1/2}}{\Delta x} (= \mathcal{D}^-\boldsymbol{g}_{i+1/2}), \qquad \delta^0 T_{i+1/2} = \frac{T_{i+1} - T_i}{\Delta x},$$
$$\mathcal{L}\boldsymbol{g}_{i+1/2} = (\mathbf{A}^+\mathcal{D}^- + \mathbf{A}^-\mathcal{D}^+)\boldsymbol{g}_{i+1/2}.$$

Theorem 1. *In the limit $\varepsilon \to 0$, the modal macro-micro scheme* (3.1) *gives a consistent discretization of the diffusion equation*

$$(1 + \frac{2a\Psi}{c_\nu})\partial_t T = \frac{2ac}{3c_\nu}\partial_x\left(\frac{1}{\sigma^a}\Psi\partial_x T\right).$$

Proof. The $\mathcal{O}(\varepsilon^0)$ term in (3.1b) is given by

$$-\sigma_{i+1/2}^a g_{1,i+1/2}^{n+1} = \gamma_1 ac \Psi_{i+1/2}^n \delta^0 T_{i+1/2}^n. \tag{3.2}$$

Similarly, the $\mathcal{O}(\varepsilon^0)$ term in (3.1a) is

$$-\sigma_i^a h_i^{n+1} = a\kappa\,\sigma_i^a \Psi_i^n h_i^{n+1} + \frac{\gamma_1}{2}\mathcal{D}^0 g_{1,i}^{n+1}$$

which on substituting $g_{1,i+1/2}^{n+1}$ from (3.2) and collecting h_i^{n+1} terms on the left hand side yields

$$(1+a\kappa\Psi_i^n)\sigma_i^a h_i^{n+1} = \frac{\gamma_1^2}{2}ac\left[\frac{\frac{\Psi_{i+1/2}^n}{\sigma_{i+1/2}^a}(T_{i+1}^n - T_i^n) - \frac{\Psi_{i-1/2}^n}{\sigma_{i-1/2}^a}(T_i^n - T_{i-1}^n)}{(\Delta x)^2}\right].$$

Thus, substituting $\sigma_i^a h_i^{n+1}$ in (3.1c)

$$(1+\frac{2a\Psi_i^n}{c_\nu})\left(\frac{T_i^{n+1}-T_i^n}{\Delta t}\right) = \frac{2ac}{3c_\nu}\left[\frac{\frac{\Psi_{i+1/2}^n}{\sigma_{i+1/2}^a}(T_{i+1}^n - T_i^n) - \frac{\Psi_{i-1/2}^n}{\sigma_{i-1/2}^a}(T_i^n - T_{i-1}^n)}{(\Delta x)^2}\right],$$

where we re-substitute $\kappa = \frac{2}{c_\nu}$. This is a discretization of the limiting diffusion equation with an explicit Euler discretization in time and centered differences for spatial derivatives.

□

3.2 Stability analysis

Next, we investigate the stability of the modal macro-micro scheme (3.1) in energy norm for a linearized version of the problem as described in [25]. The linearization assumes that the particles are emitted from the background material proportional to the temperature (instead of the 4^{th} power of temperature as given by the Stefan-Boltzmann law). That is, we set $B(T) = acT$ and thus $\Psi = 1$. Other strategies to linearize the problem include the Su-Olsen closure [26] in which the specific heat, c_ν, is assumed to be proportional to T^3.

Substituting the value of Ψ in (3.1) we get the following modal macro-micro scheme with linear emission of particles:

$$\frac{\varepsilon^2}{c}\left(\frac{h_i^{n+1}-h_i^n}{\Delta t}\right) + a\kappa\,\sigma_i^a h_i^{n+1} + \frac{\gamma_1}{2}\mathcal{D}^0 g_{1,i}^{n+1} = -\sigma_i^a h_i^{n+1}, \tag{3.3a}$$

$$\frac{\varepsilon^2}{c}\left(\frac{\boldsymbol{g}_{i+1/2}^{n+1}-\boldsymbol{g}_{i+1/2}^n}{\Delta t}\right) + \varepsilon\mathcal{L}\boldsymbol{g}_{i+1/2}^n + \boldsymbol{b}\delta^0\left(acT_{i+1/2}^n + \varepsilon^2 h_{i+1/2}^n\right) = -\sigma_{i+1/2}^a \boldsymbol{g}_{i+1/2}^{n+1}, \tag{3.3b}$$

$$\frac{T_i^{n+1}-T_i^n}{\Delta t} = \kappa\,\sigma_i^a h_i^{n+1}. \tag{3.3c}$$

The following norms are defined for the scalar- and vector-valued functions

$$\|u\|^2 = \sum_i u_i^2 \Delta x, \qquad \|\boldsymbol{\phi}\|^2 = \sum_i (\boldsymbol{\phi}_{i+1/2}^\top \boldsymbol{\phi}_{i+1/2})\Delta x.$$

Then, for the linearized modal macro-micro scheme, we have the following stability result:

Theorem 2. *Assume that the time step Δt fulfills the CFL condition for all k, such that $\mu_k \neq 0$,*

$$\Delta t \leq \frac{1}{5c\beta_N}\left(\frac{2\varepsilon\Delta x}{|\mu_k|} + \frac{\sigma_0\Delta x^2}{\mu_k^2}\right), \tag{3.4}$$

where $\beta_N = \max_k\ w_k(N+1)$ and c is the speed of light. Then, the scheme (3.3) is energy stable, that is,

$$e^{n+1} \leq e^n,$$

where the energy is defined as

$$e^n = \left\|aT^n + \frac{\varepsilon^2}{c}h^n\right\|^2 + \left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^n\right\|^2 + \left\|\sqrt{\frac{ac_\nu}{2}}T^n\right\|^2.$$

Remark 4. For the sake of compactness, the proof of this theorem, along with all the required lemmas, are presented in Appendix A. The proof follows the energy stability result in [25] and combines it with the results obtained for the modal macro-micro scheme for radiation transport from [19]. It is roughly divided into three parts; the first part bounds $\left\|aT^{n+1} + \frac{\varepsilon^2}{c}h^{n+1}\right\|^2 + \left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}\right\|^2$ from above using (3.3a) and (3.3b). In the second part we derive an upper bound for $\left\|\sqrt{\frac{ac_\nu}{2}}T^{n+1}\right\|^2$ from (3.3c). Combining the bounds obtained in the first and the second part, we show energy stability subject to step size restriction given by the CFL condition (3.4) in the third part of the proof.

4 Dynamical low-rank approximation for the modal macro-micro equations

The macro-micro decomposition [6], [20] allows us to construct an asymptotic-preserving and energy-stable numerical algorithm for the thermal radiative transfer equations. However, the microscopic variable g is still a high-dimensional quantity since it depends on time, space, and direction of flight. Thus, to reduce computational costs, we use dynamical low-rank approximation [7] to approximate the solution of g by a low-rank factorization. This section is divided into two subsections; in the first subsection, we derive evolution equations for the low-rank factorization of g for the fixed-rank BUG integrator [9]. We present an asymptotic-preserving spatio-temporal discretization and show that the numerical scheme is energy stable for the linearization presented in Section 3. In the second subsection, we extend the scheme to the augmented BUG integrator [10].

Consider the microscopic equation (2.6b) given by

$$\frac{\varepsilon^2}{c}\partial_t \boldsymbol{g} + \varepsilon \mathbf{A}\partial_x \boldsymbol{g} + \boldsymbol{b}\left(ac\Psi\,\partial_x T + \varepsilon^2 \partial_x h\right) = -\sigma^a \boldsymbol{g}. \tag{4.1}$$

The low-rank ansatz for the microscopic variable $\boldsymbol{g}$ reads:

$$\boldsymbol{g}(t,x) \approx \sum_{p,q=1}^{r} X_p(t,x) S_{pq}(t) \mathbf{V}_q(t)^\top,$$

where $r \in \mathbb{N}$ is some given rank and $\mathbf{V}_q = (V_{1,q}, \dots, V_{N,q})^\top \in \mathbb{R}^N$. Thus, we can write the above sum as

$$\boldsymbol{g}(t,x) = \boldsymbol{X}(t,x)^\top \mathbf{S}(t) \mathbf{V}(t)^\top,$$

where

$$\boldsymbol{X} = (X_1, \dots, X_r)^\top \in \mathbb{R}^r, \quad \mathbf{S} = (S_{pq})_{p,q=1}^r \in \mathbb{R}^{r\times r}, \quad \mathbf{V} = \begin{bmatrix}\mathbf{V}_1 & \dots & \mathbf{V}_r\end{bmatrix} \in \mathbb{R}^{N\times r}.$$

4.1 Fixed-rank modal macro-micro BUG integrator

With this low-rank ansatz for $\boldsymbol{g}$ we can now write down the individual steps of the fixed-rank BUG integrator [9] for updating the microscopic variable. To this end let the solution at time t_n be given by $\boldsymbol{g}^n(x) = \boldsymbol{X}^n(x)^\top \mathbf{S}^n \mathbf{V}^{n,\top}$, then the evolution equations for updating $\boldsymbol{X}, \mathbf{S}, \mathbf{V}$ are as follows:

K-step For $\boldsymbol{K}(t,x)^\top = \boldsymbol{X}(t,x)^\top \mathbf{S}(t)$ solve

$$\frac{\varepsilon^2}{c}\partial_t \boldsymbol{K}(t,x) = -\varepsilon\left[\mathbf{V}^{n,\top}\mathbf{A}\mathbf{V}^n\right]\partial_x \boldsymbol{K}(t,x) - \mathbf{V}^{n,\top}\boldsymbol{b}(ac\Psi\partial_x T + \varepsilon^2\partial_x h) - \sigma^a \boldsymbol{K}(t,x),$$

with the initial condition $\boldsymbol{K}(t_n,x) = \boldsymbol{X}^n(x)^\top \mathbf{S}^n$. We denote the updated spatial basis vectors by $\boldsymbol{X}^{n+1}(x)$ which is obtained as the orthonormal basis of $\boldsymbol{K}(t_{n+1},x)$.

L-step For $\mathbf{L}(t) = \mathbf{V}(t)\mathbf{S}(t)^\top$ solve

$$\frac{\varepsilon^2}{c}\dot{\mathbf{L}}(t) = -\varepsilon \mathbf{A}^\top \mathbf{L}(t)\langle \partial_x \boldsymbol{X}^n, \boldsymbol{X}^{n,\top}\rangle_x - \boldsymbol{b}\langle ac\Psi\partial_x T + \varepsilon^2 \partial_x h, \boldsymbol{X}^{n,\top}\rangle_x.$$

where $\langle\cdot,\cdot\rangle_x$ denotes the L^2 - inner product over the spatial domain, and we have the initial condition given by $\mathbf{L}(t_n) = \mathbf{V}^n\mathbf{S}^{n,\top}$. We denote the updated angular basis matrix by $\mathbf{V}^{n+1}$, which is obtained as the orthonormal basis of $\mathbf{L}(t_{n+1})$.

S-step Perform a Galerkin step in the updated spatial and angular basis according to

$$\begin{aligned}\frac{\varepsilon^2}{c}\dot{\mathbf{S}}(t) = &-\varepsilon\langle \boldsymbol{X}^{n+1}, \partial_x \boldsymbol{X}^{n+1,\top}\rangle_x \mathbf{S}(t)\mathbf{V}^{n+1,\top}\mathbf{A}\mathbf{V}^{n+1} - \langle \boldsymbol{X}^{n+1}, ac\Psi\partial_x T + \varepsilon^2\partial_x h\rangle_x \boldsymbol{b}^\top \mathbf{V}^{n+1}\\ &- \langle \sigma^a \boldsymbol{X}^{n+1}, \boldsymbol{X}^{n+1,\top}\rangle_x \mathbf{S}(t),\end{aligned}$$

with the initial condition $\mathbf{S}(t_n) = \langle \boldsymbol{X}^{n+1}, \boldsymbol{X}^{n,\top}\rangle_x \mathbf{S}^n \mathbf{V}^{n,\top}\mathbf{V}^{n+1}$.

4.1.1 Spatio-temporal discretization

The update equations for T and h from Section 2.1.2 along with the evolution equations for the low-rank factors of $\boldsymbol{g}$ give the fixed-rank modal macro-micro BUG equations for the thermal radiative transfer equations. Similar to Section 3, we discretize the fixed-rank modal macro-micro BUG equations in space and time. First, we define

$$\boldsymbol{X}^n_{i+1/2} = \frac{1}{\Delta x}\int_{x_i}^{x_{i+1}} \boldsymbol{X}(t_n, x)\,\mathrm{d}x$$

and $\boldsymbol{K}_{i+1/2}(t) = \boldsymbol{X}_{i+1/2}(t)^\top \mathbf{S}(t) \in \mathbb{R}^r$. Then, for the prescribed data $\boldsymbol{X}^n, \mathbf{V}^n, \mathbf{S}^n, h^n, T^n$ at time t_n the fixed-rank modal macro-micro BUG scheme updates the solution at time t_n through the following steps,

K-step Update

$$\frac{\varepsilon^2}{c}\left[\frac{\boldsymbol{K}^{n+1}_{i+1/2} - \boldsymbol{K}^n_{i+1/2}}{\Delta t}\right] = -\varepsilon\mathcal{L}_K \boldsymbol{K}^n_{i+1/2} - \mathbf{V}^{n,\top}\boldsymbol{b}\delta^0(ac\Psi^n_{i+1/2}T^n_{i+1/2} + \varepsilon^2 h^n_{i+1/2}) - \sigma^a_{i+1/2}\boldsymbol{K}^{n+1}_{i+1/2}, \tag{4.2}$$

where $\boldsymbol{K}^{n,\top}_{i+1/2} = \boldsymbol{X}^{n,\top}_{i+1/2}\mathbf{S}^n$ and

$$\mathcal{L}_K \boldsymbol{K}^n_{i+1/2} = \left[\mathbf{V}^{n,\top}\mathbf{A}^+\mathbf{V}^n\right]\mathcal{D}^-\boldsymbol{K}^n_{i+1/2} + \left[\mathbf{V}^{n,\top}\mathbf{A}^-\mathbf{V}^n\right]\mathcal{D}^+\boldsymbol{K}^n_{i+1/2}.$$

Compute $\boldsymbol{X}^{n+1}_{i+1/2}$ as the orthonormal basis of $\boldsymbol{K}^{n+1}_{i+1/2}$.

L-step Update

$$\begin{aligned}\frac{\varepsilon^2}{c}\left[\frac{\mathbf{L}^{n+1} - \mathbf{L}^n}{\Delta t}\right] = &-\varepsilon\mathcal{L}_L\mathbf{L}^n - \boldsymbol{b}\sum_i \boldsymbol{X}^{n,\top}_{i+1/2}\delta^0(ac\Psi^n_{i+1/2}T^n_{i+1/2} + \varepsilon^2 h^n_{i+1/2})\\ &- \mathbf{L}^{n+1}\sum_i \sigma^a_{i+1/2}\boldsymbol{X}^n_{i+1/2}\boldsymbol{X}^{n,\top}_{i+1/2}\end{aligned} \tag{4.3}$$

where $\mathbf{L}^n = \mathbf{V}^n\mathbf{S}^{n,\top}$ and

$$\mathcal{L}_L\mathbf{L}^n = \mathbf{A}^+\mathbf{L}^n\sum_i \mathcal{D}^-\boldsymbol{X}^n_{i+1/2}\boldsymbol{X}^{n,\top}_{i+1/2} + \mathbf{A}^-\mathbf{L}^n\sum_i \mathcal{D}^+\boldsymbol{X}^n_{i+1/2}\boldsymbol{X}^{n,\top}_{i+1/2}.$$

Compute $\mathbf{V}^{n+1}$ as the orthonormal basis of $\mathbf{L}^{n+1}$.

S-step Update

$$\frac{\varepsilon^2}{c}\left[\frac{\mathbf{S}^{n+1}-\widetilde{\mathbf{S}}^n}{\Delta t}\right] = -\varepsilon\mathcal{L}_S\widetilde{\mathbf{S}}^n - \sum_i \boldsymbol{X}_{i+1/2}^{n+1}\delta^0(ac\Psi_{i+1/2}^n T_{i+1/2}^n + \varepsilon^2 h_{i+1/2}^n)\boldsymbol{b}^\top\mathbf{V}^{n+1} - \sum_i \sigma_{i+1/2}^a \boldsymbol{X}_{i+1/2}^{n+1}\boldsymbol{X}_{i+1/2}^{n+1,\top}\mathbf{S}^{n+1} \tag{4.4}$$

where $\widetilde{\mathbf{S}}^n = \sum_j \boldsymbol{X}_{j+1/2}^{n+1}\boldsymbol{X}_{j+1/2}^{n,\top}\mathbf{S}^n\mathbf{V}^{n,\top}\mathbf{V}^{n+1}$ and

$$\mathcal{L}_S\mathbf{S}^n = \sum_i \boldsymbol{X}_{i+1/2}^{n+1}\mathcal{D}^-\boldsymbol{X}_{i+1/2}^{n+1,\top}\mathbf{S}^n\mathbf{V}^{n+1,\top}\mathbf{A}^+\mathbf{V}^{n+1} + \sum_i \boldsymbol{X}_{i+1/2}^{n+1}\mathcal{D}^+\boldsymbol{X}_{i+1/2}^{n+1,\top}\mathbf{S}^n\mathbf{V}^{n+1,\top}\mathbf{A}^-\mathbf{V}^{n+1},$$

Update T, h:

$$\frac{\varepsilon^2}{c}\left(\frac{h_i^{n+1}-h_i^n}{\Delta t}\right) + a\kappa\,\sigma_i^a\Psi_i^n h_i^{n+1} + \frac{\gamma_1}{2}\mathcal{D}^0\boldsymbol{X}_{i+1/2}^{n+1,\top}\mathbf{S}^{n+1}\mathbf{V}^{n+1,\top}\boldsymbol{e}_1 = -\sigma_i^a h_i^{n+1}, \tag{4.5}$$

$$\frac{T_i^{n+1}-T_i^n}{\Delta t} = \kappa\,\sigma_i^a h_i^{n+1}. \tag{4.6}$$

Theorem 3. *In the limit $\varepsilon \to 0$, the fixed-rank modal macro-micro BUG scheme given by eqs.* (4.2) *to* (4.6) *gives a consistent discretization of the diffusion equation*

$$\left(1+\frac{2a\Psi}{c_\nu}\right)\partial_t T = \frac{2ac}{3c_\nu}\partial_x\left(\frac{1}{\sigma^a}\partial_x T\right).$$

Proof. As $\varepsilon \to 0$, from the K-step (4.2) and L-step (4.3) we obtain

$$\mathbf{V}^{n,\top}\boldsymbol{b}\Psi_{i+1/2}^n\delta^0(acT_{i+1/2}^n) = -\sigma_{i+1/2}^a\boldsymbol{K}_{i+1/2}^{n+1}$$

and

$$\mathbf{L}^{n+1}\sum_i \sigma_{i+1/2}^a\boldsymbol{X}_{i+1/2}^n\boldsymbol{X}_{i+1/2}^{n,\top} = -\boldsymbol{b}\sum_i \boldsymbol{X}_{i+1/2}^{n,\top}\Psi_{i+1/2}^n\delta^0(acT_{i+1/2}^n).$$

If $\boldsymbol{K}_{i+1/2}^{n+1}$ is factorized as $\boldsymbol{K}_{i+1/2}^{n+1,\top} = \boldsymbol{X}_{i+1/2}^{n+1,\top}\mathbf{S}_K$ then $\frac{\Psi_{i+1/2}^n}{\sigma_{i+1/2}^a}\delta^0(acT_{i+1/2}^n)$ lies in the range space of $\boldsymbol{X}_{i+1/2}^{n+1}$. Similarly, if $\mathbf{L}^{n+1} = \mathbf{V}^{n+1}\mathbf{S}_L^\top$ and $\left(\mathbf{S}_L^\top\sum_i \sigma_{i+1/2}^a\boldsymbol{X}_{i+1/2}^n\boldsymbol{X}_{i+1/2}^{n,\top}\right)$ is invertible, then $\boldsymbol{b}$ lies in the range space of $\mathbf{V}^{n+1}$.

Now as $\varepsilon \to 0$, from the S-step we obtain

$$-\left(\sum_i \Psi_{i+1/2}^n\delta^0(acT_{i+1/2}^n)\boldsymbol{X}_{i+1/2}^{n+1}\right)\left(\boldsymbol{b}^\top\mathbf{V}^{n+1}\right) = \left(\sum_i \sigma_{i+1/2}^a\boldsymbol{X}_{i+1/2}^{n+1}\boldsymbol{X}_{i+1/2}^{n+1,\top}\right)\mathbf{S}^{n+1}. \tag{4.7}$$

Note that since $\frac{\Psi_{i+1/2}^n}{\sigma_{i+1/2}^a}\delta^0(acT_{i+1/2}^n) = \left[\sum_j \frac{\Psi_{j+1/2}^n}{\sigma_{j+1/2}^a}\delta^0(acT_{j+1/2}^n)\boldsymbol{X}_{j+1/2}^{n+1,\top}\right]\boldsymbol{X}_{i+1/2}^{n+1}$ we have

$$\sum_i \delta^0(ac\Psi_{i+1/2}^n T_{i+1/2}^n)\boldsymbol{X}_{i+1/2}^{n+1} = \sum_i \sigma_{i+1/2}^a\left(\frac{\Psi_{i+1/2}^n}{\sigma_{i+1/2}^a}\delta^0(acT_{i+1/2}^n)\boldsymbol{X}_{i+1/2}^{n+1}\right) = \left(\sum_i \sigma_{i+1/2}^a\boldsymbol{X}_{i+1/2}^{n+1}\boldsymbol{X}_{i+1/2}^{n+1,\top}\right)\left(\sum_j \frac{\Psi_{j+1/2}^n}{\sigma_{j+1/2}^a}\delta^0(acT_{j+1/2}^n)\boldsymbol{X}_{j+1/2}^{n+1}\right).$$

Hence, (4.7) becomes

$$\mathbf{S}^{n+1} = -\left(\sum_j \frac{\Psi^n_{j+1/2}}{\sigma^a_{j+1/2}} \delta^0(acT^n_{j+1/2}) \boldsymbol{X}^{n+1}_{j+1/2}\right)\left(\boldsymbol{b}^\top \mathbf{V}^{n+1}\right)$$

and since $\frac{\Psi^n_{i+1/2}}{\sigma^a_{i+1/2}} \delta^0(acT^n_{i+1/2})$ and $\boldsymbol{b}$ lie in the range of the updated spatial and angular basis, scalar multiplication with $\boldsymbol{X}^{n+1,\top}_{i+1/2}$ and $\mathbf{V}^{n+1,\top}$ from the left and right implies

$$\boldsymbol{g}^{n+1}_{i+1/2} = -\frac{\Psi^n_{i+1/2}}{\sigma^a_{i+1/2}} \delta^0(acT^n_{i+1/2}) \boldsymbol{b}^\top. \tag{4.8}$$

The rest of the proof follows along the lines of Theorem 1. □

4.1.2 Energy stability

Next, we investigate the stability of the fixed-rank modal macro-micro BUG scheme in energy norm for the linearized problem (3.3). For the following decomposition of the micro variable

$$\mathbf{g}^n = \begin{bmatrix} \boldsymbol{g}^n_{1/2} \\ \vdots \\ \boldsymbol{g}^n_{N_x+1/2} \end{bmatrix} = \mathbf{X}^n \mathbf{S}^n \mathbf{V}^{n,\top},$$

the norm is defined as

$$\|\mathbf{g}^n\|^2 = \left\|\mathbf{X}^n \mathbf{S}^n \mathbf{V}^{n,\top}\right\|^2_F \Delta x$$

Additionally, we state the following property that we use in the proof of energy stability:

Property 1. *For any $\{c_i\}_{i=1,\dots,N_x} \in \mathbb{R}$ and $\{d_i\}_{i=1,\dots,N_x} \in \mathbb{R}$ we have*

$$\sum_i c_i d_i = \frac{1}{2}\sum_i c_i^2 + \frac{1}{2}\sum_i d_i^2 - \frac{1}{2}\sum_i (c_i - d_i)^2.$$

Theorem 4. *Assume that the time step Δt fulfills the CFL condition* (3.4) *from Theorem 2. Then, the fixed-rank modal macro-micro BUG scheme given by eqs.* (4.2) *to* (4.6) *is energy stable for the linearised problem* (3.3)*, that is,*

$$e^{n+1} \leq e^n,$$

where the energy is defined as

$$e^n = \left\|aT^n + \frac{\varepsilon^2}{c} h^n\right\|^2 + \left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{X}^n \mathbf{S}^n \mathbf{V}^{n,\top}\right\|^2 + \left\|\sqrt{\frac{ac_\nu}{2}} T^n\right\|^2.$$

Proof. Since the proof of the theorem follows along the lines of Theorem 2, to shorten the presentation, we only present the parts of the proof that differ from Theorem 2. That is, we show that the inequalities (A.6) and (A.5) hold for the low-rank scheme. We begin by rewriting the S-step (4.4) of the fixed-rank modal macro-micro BUG scheme as

$$\begin{aligned}\frac{\varepsilon^2}{c\Delta t}\mathbf{S}^{n+1} = \frac{\varepsilon^2}{c\Delta t}\widetilde{\mathbf{S}}^n - \varepsilon \mathcal{L}_S \widetilde{\mathbf{S}}^n - \sum_j \boldsymbol{X}^{n+1}_{j+1/2} \delta^0(acT^n_{j+1/2} + \varepsilon^2 h^n_{j+1/2}) \boldsymbol{b}^\top \mathbf{V}^{n+1} \\ - \sum_j \sigma^a_{j+1/2} \boldsymbol{X}^{n+1}_{j+1/2} \boldsymbol{X}^{n+1,\top}_{j+1/2} \mathbf{S}^{n+1}\end{aligned}$$

and multiply $\boldsymbol{X}_{i+1/2}^{n+1,\top}$ and $\mathbf{V}^{n+1,\top}$ from the left and the right, respectively. If we define $\widetilde{\boldsymbol{g}}_{i+1/2}^{n} = \boldsymbol{X}_{i+1/2}^{n+1,\top}\widetilde{\mathbf{S}}^{n}\mathbf{V}^{n+1,\top}$ and $\boldsymbol{g}_{i+1/2}^{n+1} = \boldsymbol{X}_{i+1/2}^{n+1,\top}\mathbf{S}^{n+1}\mathbf{V}^{n+1,\top}$, we obtain

$$\begin{aligned}\frac{\varepsilon^2}{c\Delta t}\boldsymbol{g}_{i+1/2}^{n+1} = \frac{\varepsilon^2}{c\Delta t}\widetilde{\boldsymbol{g}}_{i+1/2}^{n} &- \varepsilon\boldsymbol{X}_{i+1/2}^{n+1,\top}\mathcal{L}_S\widetilde{\mathbf{S}}^{n}\mathbf{V}^{n+1,\top} \\ &- \sum_j \boldsymbol{X}_{i+1/2}^{n+1,\top}\boldsymbol{X}_{j+1/2}^{n+1}\delta^0(acT_{j+1/2}^{n} + \varepsilon^2 h_{j+1/2}^{n})\boldsymbol{b}^\top\mathbf{V}^{n+1}\mathbf{V}^{n+1,\top} \\ &- \sum_j \sigma_{j+1/2}^{a}\boldsymbol{X}_{i+1/2}^{n+1,\top}\boldsymbol{X}_{j+1/2}^{n+1}\boldsymbol{g}_{j+1/2}^{n+1}\mathbf{V}^{n+1}\mathbf{V}^{n+1,\top}.\end{aligned} \tag{4.9}$$

Defining the projection matrix onto the spatial basis as $\mathbf{P}^X \in \mathbb{R}^{Nx\times N_x}$ with entries

$$P_{ij}^X = \boldsymbol{X}_{i+1/2}^{n+1,\top}\boldsymbol{X}_{j+1/2}^{n+1} = \sum_q X_{i+1/2,q}^{n+1}X_{j+1/2,q}^{n+1}$$

and the projection matrix onto the angular basis as $\mathbf{P}^V = \mathbf{V}^{n+1}\mathbf{V}^{n+1,\top} \in \mathbb{R}^{N\times N}$, (4.9) reads

$$\begin{aligned}\frac{\varepsilon^2}{c\Delta t}\boldsymbol{g}_{i+1/2}^{n+1} = \frac{\varepsilon^2}{c\Delta t}\widetilde{\boldsymbol{g}}_{i+1/2}^{n} &- \varepsilon\boldsymbol{X}_{i+1/2}^{n+1,\top}\mathcal{L}_S\widetilde{\mathbf{S}}^{n}\mathbf{V}^{n+1,\top} \\ &- \sum_j P_{ij}^X\delta^0(acT_{j+1/2}^{n} + \varepsilon^2 h_{j+1/2}^{n})\boldsymbol{b}^\top\mathbf{P}^V - \sum_j \sigma_{j+1/2}^{a}P_{ij}^X\boldsymbol{g}_{j+1/2}^{n+1}\mathbf{P}^V.\end{aligned}$$

Thus, if $1 \le k \le N$, the evolution equation for the k^{th} moment, $g_{i+1/2,k}^{n+1}$, is given by

$$\begin{aligned}\frac{\varepsilon^2}{c\Delta t}g_{i+1/2,k}^{n+1} = \frac{\varepsilon^2}{c\Delta t}\widetilde{g}_{i+1/2,k}^{n} &- \varepsilon\boldsymbol{X}_{i+1/2}^{n+1,\top}\mathcal{L}_S\widetilde{\mathbf{S}}^{n}\mathbf{V}^{n+1,\top}\boldsymbol{e}_k \\ &- \sum_j P_{ij}^X\delta^0(acT_{j+1/2}^{n} + \varepsilon^2 h_{j+1/2}^{n})\boldsymbol{b}^\top\mathbf{P}^V\boldsymbol{e}_k - \sum_j \sigma_{j+1/2}^{a}P_{ij}^X\boldsymbol{g}_{j+1/2}^{n+1}\mathbf{P}^V\boldsymbol{e}_k,\end{aligned} \tag{4.10}$$

where $\boldsymbol{e}_k = (\delta_{ik})_{i=1,\dots,N}$. First, we consider the second term on the right-hand side of (4.10) and split it into two sub-equations

$$\begin{aligned}\boldsymbol{X}_{i+1/2}^{n+1,\top}\mathcal{L}_S\widetilde{\mathbf{S}}^{n}\mathbf{V}^{n+1,\top}\boldsymbol{e}_k &= \boldsymbol{X}_{i+1/2}^{n+1,\top}\left[\sum_j \boldsymbol{X}_{j+1/2}^{n+1}\mathcal{D}^-\boldsymbol{X}_{j+1/2}^{n+1,\top}\widetilde{\mathbf{S}}^{n}\mathbf{V}^{n+1,\top}\mathbf{A}^+\mathbf{V}^{n+1}\right. \\ &\qquad \left. + \sum_j \boldsymbol{X}_{j+1/2}^{n+1}\mathcal{D}^+\boldsymbol{X}_{j+1/2}^{n+1,\top}\widetilde{\mathbf{S}}^{n}\mathbf{V}^{n+1,\top}\mathbf{A}^-\mathbf{V}^{n+1}\right]\mathbf{V}^{n+1,\top}\boldsymbol{e}_k \\ &= \sum_{j,\ell,q} P_{ij}^X\mathcal{D}^-\widetilde{g}_{j+1/2,\ell}^{n}A_{\ell q}^{+}P_{qk}^V + \sum_{j,\ell,q} P_{ij}^X\mathcal{D}^+\widetilde{g}_{j+1/2,\ell}^{n}A_{\ell q}^{-}P_{qk}^V.\end{aligned}$$

Similarly, expanding the third and the fourth term on the right-hand side of (4.10) we get

$$\begin{aligned}\frac{\varepsilon^2}{c\Delta t}g_{i+1/2,k}^{n+1} = \frac{\varepsilon^2}{c\Delta t}\widetilde{g}_{i+1/2,k}^{n} &- \varepsilon\sum_{j,\ell,q} P_{ij}^X\mathcal{D}^-\widetilde{g}_{j+1/2,\ell}^{n}A_{\ell q}^{+}P_{qk}^V - \varepsilon\sum_{j,\ell,q} P_{ij}^X\mathcal{D}^+\widetilde{g}_{j+1/2,\ell}^{n}A_{\ell q}^{-}P_{qk}^V \\ &- \sum_{j,q} P_{ij}^X\delta^0(acT_{j+1/2}^{n} + \varepsilon^2 h_{j+1/2}^{n})b_q P_{qk}^V - \sum_{j,q} P_{ij}^X\sigma_{j+1/2}^{a}g_{j+1/2,q}^{n+1}P_{qk}^V.\end{aligned} \tag{4.11}$$

Multiplying (4.11) by $g^{n+1}_{i+1/2,k}\Delta x$ and summing over i, k we get

$$\begin{aligned}\frac{\varepsilon^2}{c\Delta t}\sum_{i,k}(g^{n+1}_{i+1/2,k})^2\Delta x = \frac{\varepsilon^2}{c\Delta t}\sum_{i,k}\widetilde{g}^n_{i+1/2,k}g^{n+1}_{i+1/2,k}\Delta x - \varepsilon\sum_{j,\ell,q}\mathcal{D}^-\widetilde{g}^n_{j+1/2,\ell}A^+_{\ell q}\sum_{i,k}P^V_{qk}P^X_{ij}g^{n+1}_{i+1/2,k}\Delta x\\ - \varepsilon\sum_{j,\ell,q}\mathcal{D}^+\widetilde{g}^n_{j+1/2,\ell}A^-_{\ell q}\sum_{i,k}P^V_{qk}P^X_{ij}g^{n+1}_{i+1/2,k}\Delta x\\ - \sum_{j,q}\delta^0(acT^n_{j+1/2}+\varepsilon^2h^n_{j+1/2})b_q\sum_{i,k}P^V_{qk}P^X_{ij}g^{n+1}_{i+1/2,k}\Delta x\\ - \sum_{j,q}\sigma^a_{j+1/2}g^{n+1}_{j+1/2,q}\sum_{i,k}P^V_{qk}P^X_{ij}g^{n+1}_{i+1/2,k}\Delta x.\end{aligned} \tag{4.12}$$

Using

$$\sum_i P^X_{ij}g^{n+1}_{i+1/2,k} = g^{n+1}_{j+1/2,k}, \qquad \sum_k P^V_{qk}g^{n+1}_{j+1/2,k} = g^{n+1}_{j+1/2,q}. \tag{4.13}$$

and Property 1, (4.12) reduces to

$$\begin{aligned}\frac{\varepsilon^2}{2c\Delta t}\left(\|\mathbf{g}^{n+1}\|^2 - \|\widetilde{\mathbf{g}}^n\|^2 + \|\mathbf{g}^{n+1}-\widetilde{\mathbf{g}}^n\|^2\right) = -\varepsilon\sum_{j,\ell,q}\mathcal{D}^-\widetilde{g}^n_{j+1/2,\ell}A^+_{\ell q}g^{n+1}_{j+1/2,q}\Delta x\\ - \varepsilon\sum_{j,\ell,q}\mathcal{D}^+\widetilde{g}^n_{j+1/2,\ell}A^-_{\ell q}g^{n+1}_{j+1/2,q}\Delta x\\ - \sum_{j,q}\delta^0(acT^n_{j+1/2}+\varepsilon^2h^n_{j+1/2})b_qg^{n+1}_{j+1/2,q}\Delta x\\ - \sum_{j,q}\sigma^a_{j+1/2}\widetilde{g}^{n+1}_{j+1/2,q}g^{n+1}_{j+1/2,q}\Delta x.\end{aligned}$$

Collecting into a vector, we get

$$\begin{aligned}\frac{\varepsilon^2}{2c\Delta t}\left(\|\mathbf{g}^{n+1}\|^2 - \|\widetilde{\mathbf{g}}^n\|^2 + \|\mathbf{g}^{n+1}-\widetilde{\mathbf{g}}^n\|^2\right) = -\varepsilon\sum_j\left(\mathcal{L}^\top\widetilde{\boldsymbol{g}}^n_{j+1/2}\right)\boldsymbol{g}^{n+1,\top}_{j+1/2}\Delta x\\ - \sum_j\boldsymbol{b}^\top\delta^0(acT^n_{j+1/2}+\varepsilon^2h^n_{j+1/2})\boldsymbol{g}^{n+1,\top}_{j+1/2}\Delta x\\ - \sum_j\sigma^a_{j+1/2}\boldsymbol{g}^{n+1}_{j+1/2}\boldsymbol{g}^{n+1,\top}_{j+1/2}\Delta x.\end{aligned}$$

The above equation is equivalent to (A.6). Similarly, substituting the temperature update (4.6) in (4.5) we get

$$\left(\frac{aT^{n+1}_i+\frac{\varepsilon^2}{c}h^{n+1}_i - aT^n_i - \frac{\varepsilon^2}{c}h^{n+1}_i}{\Delta t}\right) + \frac{\gamma_1}{2}\mathcal{D}^0\boldsymbol{X}^{n+1,\top}_i\mathbf{S}^{n+1}\mathbf{V}^{n+1,\top}\boldsymbol{e}_1 = -\sigma^a_ih^{n+1}_i. \tag{4.14}$$

Multiplying (4.14) by $aT^{n+1}_i + \frac{\varepsilon^2}{c}h^{n+1}_i$, summing over i and using Property 1 yields

$$\begin{aligned}\frac{1}{2\Delta t}\left(\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}\right\| - \left\|aT^n+\frac{\varepsilon^2}{c}h^{n+1}\right\| + \left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1} - aT^n - \frac{\varepsilon^2}{c}h^{n+1}\right\|\right)\\ +\frac{\gamma_1}{2}\sum_i(aT^{n+1}_i+\frac{\varepsilon^2}{c}h^{n+1}_i)\mathcal{D}^0g^{n+1}_{i,1} = -\sum_i\sigma^a_i(aT^{n+1}_i+\frac{\varepsilon^2}{c}h^{n+1}_i)h^{n+1}_i\end{aligned} \tag{4.15}$$

which is the same as (A.5). The rest of the proof follows along the lines of Theorem 2 (see Appendix A).

□

4.1.3 Local mass conservation

Theorem 5. *The fixed-rank modal macro-micro BUG scheme is locally conservative. I.e., if the scalar flux at time t_n is denoted by $\Phi_i^n = acT_i^n + \varepsilon^2 h_i^n$, where $n \in \{0,1\}$ and $g_{i+1/2,k}^{n+1} = \sum_{\ell,m} X_{i+1/2,\ell}^{n+1} S_{\ell m}^{n+1} V_{km}^{n+1}$ the scheme fulfills the discrete conservation law*

$$\frac{\Phi_i^{n+1} - \Phi_i^n}{\Delta t} + c\frac{\gamma_1}{2}\mathcal{D}^0 g_{1,i}^{n+1} = -c\sigma_i^a h_i^{n+1}, \tag{4.16a}$$

$$\frac{c_\nu}{2}\left(\frac{T_i^{n+1} - T_i^n}{\Delta t}\right) = \sigma_i^a h_i^{n+1}. \tag{4.16b}$$

Proof. Since, for zero or periodic boundary conditions, $\sum_i c\frac{\gamma_1}{2}\mathcal{D}^0 g_{1,i}^{n+1} = 0$, this means that the total mass $\sum_i \left(\frac{1}{c}\Phi_i^n + \frac{c_\nu}{2}T_i^n\right)$ is conserved over all time steps n. This result is a direct consequence of the macro-micro strategy as shown in [23] and follows from multiplying (4.6) with ac and adding (4.5) for the linearized problem. □

4.2 Asymptotic-preserving modification to augmented BUG integrator

We see from Theorem 3 that the updated spatial and angular basis span $\frac{\Psi_{i+1/2}^n}{\sigma_{i+1/2}^a}\delta^0(acT_{i+1/2}^n)$ and $\boldsymbol{b}$, respectively. Unlike the fixed-rank BUG integrator [9], naively using the augmented BUG integrator [10] does not guarantee that this property is fulfilled since the truncation step may prune away essential basis vectors. Thus, we propose the following modification to the augmented BUG integrator, based on its basis-augmentation step and conservative truncation [22], to obtain an asymptotic-preserving scheme. To ease the presentation of the integrator, we consider the spatially and angularly discretized problem from Section 4.1 so that $\mathbf{g} \in \mathbb{R}^{N_x \times N}$. Thus, the low-rank ansatz takes the form

$$\mathbf{g}(t) = \mathbf{X}(t)\mathbf{S}(t)\mathbf{V}(t)^\top.$$

Then, one step of the modal macro-micro BUG scheme updates $\mathbf{X}^n, \mathbf{V}^n, \mathbf{S}^n, h^n, T^n$ from time t_n to t_{n+1} by the following steps

1. Spatial and angular basis update

   K-step: Update $\mathbf{K}(t_{n+1}) \in \mathbb{R}^{N_x \times r}$ according to the K-step (4.2) of the fixed-rank modal macro-micro BUG scheme. Then compute $\widehat{\mathbf{X}}^{n+1}$ as an orthonormal basis of $\begin{bmatrix}(\boldsymbol{\sigma^a})^{-1}\Psi^n\boldsymbol{\delta^0}(ac\mathbf{T}^n) & \mathbf{K}(t_{n+1}) & \mathbf{X}^n\end{bmatrix}$ and store $\widehat{\mathbf{M}} = \widehat{\mathbf{X}}^{n+1,\top}\mathbf{X}^n \in \mathbb{R}^{(2r+1)\times r}$.
   L-step: Update $\mathbf{L}(t_{n+1}) \in \mathbb{R}^{N \times r}$ according to the L-step (4.3) of the fixed-rank modal macro-micro BUG scheme. Then compute $\widehat{\mathbf{V}}^{n+1}$ as an orthonormal basis of $\begin{bmatrix}\boldsymbol{b} & \mathbf{L}(t_{n+1}) & \mathbf{V}^n\end{bmatrix}$ and store $\widehat{\mathbf{N}} = \widehat{\mathbf{V}}^{n+1,\top}\mathbf{V}^n \in \mathbb{R}^{(2r+1)\times r}$.

2. Perform a Galerkin update of the coefficient matrix $\widehat{\mathbf{S}}^{n+1}$ similar to (2) with the initial condition $\widetilde{\mathbf{S}}^n = \widehat{\mathbf{M}}^\top\mathbf{S}^n\widehat{\mathbf{N}}^\top$ and the right-hand side as described in (4.4) of the fixed-rank modal macro-micro BUG scheme.

3. Asymptotic-preserving splitting of the basis matrices and truncation

   Set $\widehat{\mathbf{K}} = \widehat{\mathbf{X}}^{n+1}\widehat{\mathbf{S}}^{n+1}$ and split $\widehat{\mathbf{K}} = \begin{bmatrix}\widehat{\mathbf{K}}^{\mathrm{ap}} & \widehat{\mathbf{K}}^{\mathrm{rem}}\end{bmatrix}$ into basis vectors, where $\widehat{\mathbf{K}}^{\mathrm{ap}} = \left[(\boldsymbol{\sigma^a})^{-1}\Psi^n\boldsymbol{\delta^0}(ac\mathbf{T}^n)\right] \in \mathbb{R}^{N_x\times 1}$ and the remaining basis vectors into $\widehat{\mathbf{K}}^{\mathrm{rem}} \in \mathbb{R}^{N_x\times 2r}$. Similarly, split the angular basis $\widehat{\mathbf{V}} = \begin{bmatrix}\widehat{\mathbf{V}}^{\mathrm{ap}} & \widehat{\mathbf{V}}^{\mathrm{rem}}\end{bmatrix}$, where $\widehat{\mathbf{V}}^{\mathrm{ap}} = \left[\boldsymbol{b}\right] \in \mathbb{R}^{N\times 1}$ and $\widehat{\mathbf{V}}^{\mathrm{rem}} \in \mathbb{R}^{N\times 2r}$.

Next compute the QR decomposition of $\widehat{\mathbf{K}}^{\text{rem}}$

$$\widehat{\mathbf{K}}^{\text{rem}} = \widehat{\mathbf{X}}^{\text{rem}}\widehat{\mathbf{S}}^{\text{rem}}.$$

and for truncating the rank compute the SVD of $\widehat{\mathbf{S}}^{\text{rem}}$ as

$$\widehat{\mathbf{S}}^{\text{rem}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{W}^{\top}. \tag{4.17}$$

We truncate the remaining basis vectors to $1 \leq r^* \leq 2r$ such that if $\hat{\sigma}_i$, $i = 1, \ldots, 2r$, are the singular values of $\widehat{\mathbf{S}}^{\text{rem}}$ then for some user defined tolerance ϑ the following is satisfied:

$$\left(\sum_{i=r^*+1}^{2r} \hat{\sigma}_i \right)^{1/2} \leq \vartheta.$$

The new rank is set as $r_1 = r^* + 1$. Then let $\widehat{\mathbf{U}} \in \mathbb{R}^{2r \times r^*}$ and $\widehat{\mathbf{W}} \in \mathbb{R}^{2r \times r^*}$ be the matrices containing the first r^* columns of $\mathbf{U}$ and $\mathbf{W}$, respectively. Similarly, let $\widehat{\boldsymbol{\Sigma}}$ be the first $r^* \times r^*$ block of $\boldsymbol{\Sigma}$; then we set

$$\mathbf{X}^{\text{rem}} = \widehat{\mathbf{X}}^{\text{rem}}\widehat{\mathbf{U}}, \qquad \mathbf{S}^{\text{rem}} = \widehat{\boldsymbol{\Sigma}}, \qquad \mathbf{W}^{n+1} = \widehat{\mathbf{V}}^{\text{rem}}\widehat{\mathbf{W}}.$$

Then we get the updated angular basis $\mathbf{V}^{n+1} \in \mathbb{R}^{N \times r_1}$ by adding columns, i.e.

$$\mathbf{V}^{n+1} = \begin{bmatrix} \widehat{\mathbf{V}}^{\text{ap}} & \mathbf{W}^{n+1} \end{bmatrix}.$$

For the updated spatial basis, we first compute the QR decomposition of $\widehat{\mathbf{K}}^{\text{ap}}$ as

$$\widehat{\mathbf{K}}^{\text{ap}} = \mathbf{X}^{\text{ap}}\mathbf{S}^{\text{ap}}.$$

Then set $\widehat{\mathbf{X}} = \begin{bmatrix} \mathbf{X}^{\text{ap}} & \mathbf{X}^{\text{rem}} \end{bmatrix}$ and subsequently perform a QR decomposition to obtain the updated spatial basis matrix $\mathbf{X}^{n+1} \in \mathbb{R}^{N_x \times r_1}$,

$$\mathbf{X}^{n+1}\mathbf{R}_2 = \widehat{\mathbf{X}}. \tag{4.18}$$

Finally we set the updated coefficient matrix $\mathbf{S}^{n+1} \in \mathbb{R}^{r_1 \times r_1}$ to be

$$\mathbf{S}^{n+1} = \mathbf{R}_2 \begin{bmatrix} \mathbf{S}^{\text{ap}} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{\text{rem}} \end{bmatrix} \tag{4.19}$$

and the approximation at the next time step is set as $\mathbf{g}^{n+1} = \mathbf{X}^{n+1}\mathbf{S}^{n+1}\mathbf{V}^{n+1,\top}$.

4. Update T, h:

$$\frac{\varepsilon^2}{c}\left(\frac{h_i^{n+1} - h_i^n}{\Delta t}\right) + a\kappa\,\sigma_i^a \Psi_i^n h_i^{n+1} + \frac{\gamma_1}{2}\mathcal{D}^0 \boldsymbol{X}_{i+1/2}^{n+1,\top}\mathbf{S}^{n+1}\mathbf{V}^{n+1,\top}\boldsymbol{e}_1 = -\sigma_i^a h_i^{n+1}, \tag{4.20}$$

$$\frac{T_i^{n+1} - T_i^n}{\Delta t} = \kappa\,\sigma_i^a h_i^{n+1}. \tag{4.21}$$

Lemma 1. *For the proposed modal macro-micro BUG scheme, we have*

$$\boldsymbol{R}_2^{-\top}\begin{bmatrix} (\boldsymbol{S}^{ap})^{-\top} & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{U}}^{\top}(\widehat{\boldsymbol{S}}^{rem})^{-\top} \end{bmatrix}\widehat{\boldsymbol{S}}^{n+1,\top}\widehat{\boldsymbol{S}}^{n+1}\begin{bmatrix} \boldsymbol{I}_m & \boldsymbol{0} \\ \boldsymbol{0} & \widehat{\boldsymbol{W}} \end{bmatrix} = \boldsymbol{S}^{n+1},$$

where the matrices are as defined above and $\boldsymbol{I}_m$ is the $m \times m$ identity matrix.

Proof. See Appendix B. □

Theorem 6. *The proposed modal macro-micro BUG scheme is asymptotic-preserving in the sense of Theorem 3.*

Proof. From the K- and L-step of the modal macro-micro BUG scheme we get $(\boldsymbol{\sigma}^{\boldsymbol{a}})^{-1}\Psi^n\boldsymbol{\delta}^{\mathbf{0}}(ac\mathbf{T}^n) \in \text{Range}\,(\widehat{\mathbf{X}}^{n+1})$ and $\boldsymbol{b} \in \text{Range}\,(\widehat{\mathbf{V}}^{n+1})$. Thus, along the lines of Theorem 3 for $\varepsilon \to 0$ we get from the S-step of the modal macro-micro BUG scheme

$$-\left(\widehat{\mathbf{X}}^{n+1,\top}(\boldsymbol{\sigma}^{\boldsymbol{a}})^{-1}\Psi^n\boldsymbol{\delta}^{\mathbf{0}}(ac\mathbf{T}^n)\right)\left(\boldsymbol{b}^\top\widehat{\mathbf{V}}^{n+1}\right) = \widehat{\mathbf{S}}^{n+1}. \tag{4.22}$$

Now, to show that the proposed scheme is asymptotic-preserving, we need to show two things; first, that $(\boldsymbol{\sigma}^{\boldsymbol{a}})^{-1}\Psi^n\boldsymbol{\delta}^{\mathbf{0}}(ac\mathbf{T}^n) \in \text{Range}(\mathbf{X}^{n+1})$ and $\boldsymbol{b} \in \text{Range}(\mathbf{V}^{n+1})$. Second, we need to show that the above relation (4.22) also holds for the truncated factor matrices $\mathbf{X}^{n+1}, \mathbf{V}^{n+1}$ and $\mathbf{S}^{n+1}$. The first property follows directly from the construction of the scheme. For the latter, we can represent $\mathbf{X}^{n+1}$ as

$$\begin{aligned}
\mathbf{X}^{n+1} &= \begin{bmatrix}\mathbf{X}^{\text{ap}} & \mathbf{X}^{\text{rem}}\end{bmatrix}\mathbf{R}_2^{-1} \\
&= \begin{bmatrix}\widehat{\mathbf{K}}^{\text{ap}}(\mathbf{S}^{\text{ap}})^{-1} & \widehat{\mathbf{X}}^{\text{rem}}\widehat{\mathbf{U}}\end{bmatrix}\mathbf{R}_2^{-1} \\
&= \begin{bmatrix}\widehat{\mathbf{K}}^{\text{ap}}(\mathbf{S}^{\text{ap}})^{-1} & \widehat{\mathbf{K}}^{\text{rem}}(\widehat{\mathbf{S}}^{\text{rem}})^{-1}\widehat{\mathbf{U}}\end{bmatrix}\mathbf{R}_2^{-1} \\
&= \begin{bmatrix}\widehat{\mathbf{K}}^{\text{ap}} & \widehat{\mathbf{K}}^{\text{rem}}\end{bmatrix}\begin{bmatrix}(\mathbf{S}^{\text{ap}})^{-1} & \mathbf{0} \\ \mathbf{0} & (\widehat{\mathbf{S}}^{\text{rem}})^{-1}\widehat{\mathbf{U}}\end{bmatrix}\mathbf{R}_2^{-1} \\
&= \widehat{\mathbf{X}}^{n+1}\widehat{\mathbf{S}}^{n+1}\begin{bmatrix}(\mathbf{S}^{\text{ap}})^{-1} & \mathbf{0} \\ \mathbf{0} & (\widehat{\mathbf{S}}^{\text{rem}})^{-1}\widehat{\mathbf{U}}\end{bmatrix}\mathbf{R}_2^{-1}.
\end{aligned}$$

Thus we get

$$\mathbf{X}^{n+1,\top} = \mathbf{R}_2^{-\top}\begin{bmatrix}(\mathbf{S}^{\text{ap}})^{-\top} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{U}}^\top(\widehat{\mathbf{S}}^{\text{rem}})^{-\top}\end{bmatrix}\widehat{\mathbf{S}}^{n+1,\top}\widehat{\mathbf{X}}^{n+1,\top}$$

and similarly, we get the following relation for the updated angular basis

$$\begin{aligned}
\mathbf{V}^{n+1} &= \begin{bmatrix}\widehat{\mathbf{V}}^{\text{ap}} & \mathbf{W}^{n+1}\end{bmatrix} \\
&= \begin{bmatrix}\widehat{\mathbf{V}}^{\text{ap}} & \widehat{\mathbf{V}}^{\text{rem}}\end{bmatrix}\begin{bmatrix}\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix}.
\end{aligned}$$

This yields the relation

$$\mathbf{V}^{n+1} = \widehat{\mathbf{V}}^{n+1}\begin{bmatrix}\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix},$$

where $\mathbf{I}_m$ is the $m\times m$ identity matrix. Now we multiply (4.22) by $\mathbf{R}_2^{-\top}\begin{bmatrix}(\mathbf{S}^{\text{ap}})^{-\top} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{U}}^\top(\widehat{\mathbf{S}}^{\text{rem}})^{-\top}\end{bmatrix}\widehat{\mathbf{S}}^{n+1,\top}$ from the left and by $\begin{bmatrix}\mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix}$ from the right and using Lemma 1 for the right-hand side gives

$$-\left(\mathbf{X}^{n+1,\top}(\boldsymbol{\sigma}^{\boldsymbol{a}})^{-1}\Psi^n\boldsymbol{\delta}^{\mathbf{0}}(ac\mathbf{T}^n)\right)\left(\boldsymbol{b}^\top\mathbf{V}^{n+1}\right) = \mathbf{S}^{n+1}.$$

Thus by multiplying by $\mathbf{X}^{n+1}$ and $\mathbf{V}^{n+1}$ from the left and the right and using the fact that $(\boldsymbol{\sigma}^{\boldsymbol{a}})^{-1}\Psi^n\boldsymbol{\delta}^{\mathbf{0}}(ac\mathbf{T}^n) \in \text{Range}(\mathbf{X}^{n+1})$ and $\boldsymbol{b} \in \text{Range}(\mathbf{V}^{n+1})$ we can show that the proposed scheme is asymptotic-preserving by following the steps from Theorem 3. □

4.2.1 Energy stability

Theorem 7. *Assume that the CFL condition* (3.4) *from Theorem 2 holds. Then the modal macro-micro BUG scheme is energy stable for the linearized problem* (3.3)*, where the energy is the same as defined in Theorem 4.*

Proof. Similar to Theorem 4 we start by considering the S-step of the modal macro-micro BUG scheme where,

$$\begin{aligned}\frac{\varepsilon^2}{c\Delta t}\widehat{\mathbf{S}}^{n+1} = \frac{\varepsilon^2}{c\Delta t}\widetilde{\mathbf{S}}^{n} - \varepsilon\mathcal{L}_S\widetilde{\mathbf{S}}^{n} - \sum_j \widehat{\boldsymbol{X}}^{n+1}_{j+1/2}\delta^0(acT^n_{j+1/2} + \varepsilon^2 h^n_{j+1/2})\boldsymbol{b}^\top\widehat{\mathbf{V}}^{n+1} \\ - \sum_j \sigma^a_{j+1/2}\widehat{\boldsymbol{X}}^{n+1}_{j+1/2}\widehat{\boldsymbol{X}}^{n+1,\top}_{j+1/2}\widehat{\mathbf{S}}^{n+1}.\end{aligned} \tag{4.23}$$

We note that $\sum_i \widehat{\boldsymbol{X}}^{n+1,\top}_{i+1/2}\widetilde{\mathbf{S}}^n\widehat{\mathbf{V}}^{n+1,\top} = \sum_i \boldsymbol{X}^{n,\top}_{i+1/2}\mathbf{S}^n\mathbf{V}^{n,\top}$. Multiplying (4.23) by $\widehat{\boldsymbol{X}}^{n+1,\top}_{i+1/2}$ from the left and $\widehat{\mathbf{V}}^{n+1,\top}$ from the right, then summing over i we get

$$\begin{aligned}\frac{\varepsilon^2}{c\Delta t}\sum_i \widehat{\boldsymbol{X}}^{n+1,\top}_{i+1/2}\widehat{\mathbf{S}}^{n+1}\widehat{\mathbf{V}}^{n+1,\top}\Delta x = \frac{\varepsilon^2}{c\Delta t}\sum_i \boldsymbol{X}^{n,\top}_{i+1/2}\mathbf{S}^n\mathbf{V}^{n,\top}\Delta x - \varepsilon\sum_i \widehat{\boldsymbol{X}}^{n+1,\top}_{i+1/2}\mathcal{L}_S\widetilde{\mathbf{S}}\widehat{\mathbf{V}}^{n+1,\top}\Delta x \\ - \sum_{i,j}\widehat{\boldsymbol{X}}^{n+1,\top}_{i+1/2}\widehat{\boldsymbol{X}}^{n+1}_{j+1/2}\delta^0(acT^n_{j+1/2} + \varepsilon^2 h^n_{j+1/2})\boldsymbol{b}^\top\widehat{\mathbf{V}}^{n+1}\widehat{\mathbf{V}}^{n+1}\Delta x \\ - \sum_{i,j}\sigma^a_{j+1/2}\widehat{\boldsymbol{X}}^{n+1,\top}_{i+1/2}\widehat{\boldsymbol{X}}^{n+1}_{j+1/2}\widehat{\boldsymbol{X}}^{n+1,\top}_{j+1/2}\widehat{\mathbf{S}}^{n+1}\widehat{\mathbf{V}}^{n+1}\Delta x.\end{aligned}$$

Since the truncation step of the integrator does not increase the norm of the solution

$$\left\|\mathbf{X}^{n+1}\mathbf{S}^{n+1}\mathbf{V}^{n+1,\top}\right\| = \left\|\mathbf{S}^{n+1}\right\| \leq \left\|\widehat{\mathbf{S}}^{n+1}\right\| = \left\|\widehat{\mathbf{X}}^{n+1}\widehat{\mathbf{S}}^{n+1}\widehat{\mathbf{V}}^{n+1,\top}\right\|.$$

The rest of the proof follows along the lines of Theorem 4. □

5 Numerical results

The following numerical results can be reproduced with the openly available source code [27].

5.1 Rectangular pulse test case

We consider gray thermal radiative transfer equations in slab geometry (2.1) on the spatial domain $D = [-10, 10]$. The initial distribution of the temperature is given by the rectangular pulse

$$T(t = 0, x) = \frac{100}{\sigma^a(x)} \cdot \chi_{[-0.5,0.5]}(x).$$

The particle density is initially at an equilibrium with the temperature and is given by

$$f(t = 0, x, \mu) = acT(t = 0, x).$$

Subsequently, as time progresses, the particles move into all directions $\mu \in [-1, 1]$ while undergoing isotropic absorption at the rate $\sigma^a(x) = 0.5$. In this test case, we assume that no particles are present at the boundary during the entire simulation, and the temperature remains zero at the boundary as well. The initial and boundary conditions for g and h can be derived from those for temperature and particle density by using the relations (2.5). We assume that constants are scaled

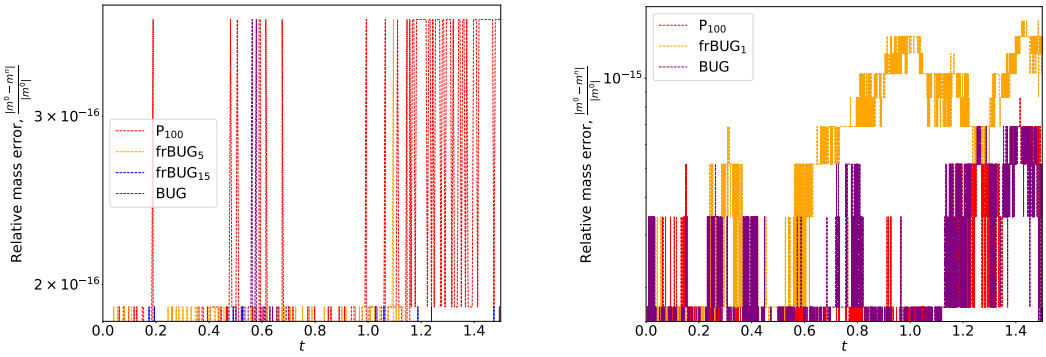
to 1, i.e., we set the radiation constant $a = 1$, speed of light $c = 1$, and the specific heat is $c_\nu = 1$. The mass at time t_n is defined as

$$m^n = \sum_i \left(aT_i^n + \frac{\varepsilon^2}{c} h_i^n + \frac{c_\nu}{2} T_i^n \right) \Delta x. \tag{5.1}$$

The spatial domain is divided into $N_x = 501$ spatial cells, and we use $N = 100$ Legendre polynomials to represent the angular variable. The moment solutions (P_N) are computed using the modal macro-micro scheme (3.3). For $\varepsilon = 1$, we choose a rank of $r = 5, 15$ for the fixed-rank modal macro-micro BUG integrator (frBUG) and an initial rank of $r = 1$ for the modal macro-micro BUG integrator (BUG). For rank truncation we set the tolerance parameter to $\vartheta = 5 \cdot 10^{-2} \|\mathbf{\Sigma}\|_2$ and the end time is set to $t_{\text{end}} = 1.5$. The step size is chosen according to (3.4)

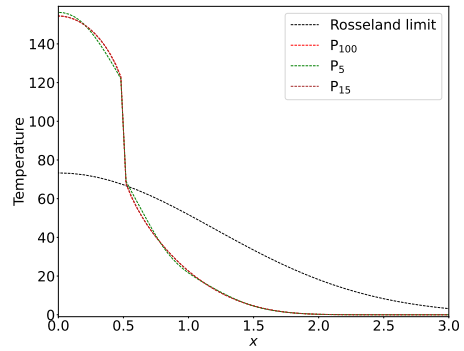
$$\Delta t = \min_k \left\{ \frac{1}{5c\beta_N} \left(\frac{2\varepsilon \Delta x}{|\mu_k|} + \frac{\sigma_0 \Delta x^2}{\mu_k^2} \right) \right\}$$

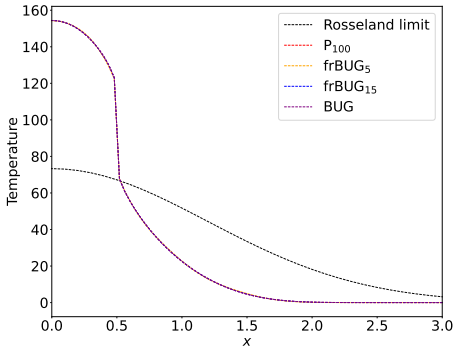
where, for $N = 100$ the step size is minimal for $\mu_k = -0.999719$ which gives the step size $\Delta t \approx 0.005$. We compare the low-rank approximations with the moment solutions P_5, P_{15} and P_{100}, and the results are given in Figure 2. The relative mass error of all the low-rank solutions as well as the P_{100} solution is given in Figure 1a. Note that the chosen step size for the P_5 and P_{15} solutions differs from that of the rest of the solutions and is minimal for a different quadrature point, which we do not specify here. From the plots of temperature and scalar flux, we see that the solution of the fixed-rank modal macro-micro BUG integrator with $r = 15$ (BUG_{15}) and the modal macro-micro BUG integrator agree well with the full moment solution P_{100} and are different from the Rosseland diffusion limit. Additionally, the BUG_5 approximation performs much better than the P_5 solution. All the integrators dissipate energy over time, as we see from Figure 3.

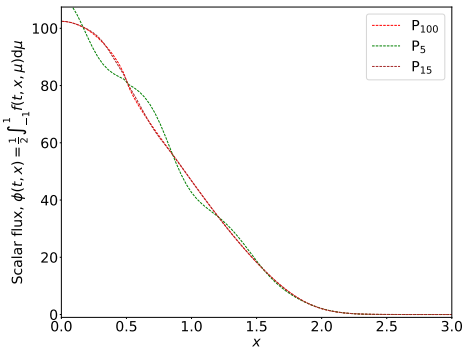


(a) $\varepsilon = 1.0$ (b) $\varepsilon = 10^{-5}$

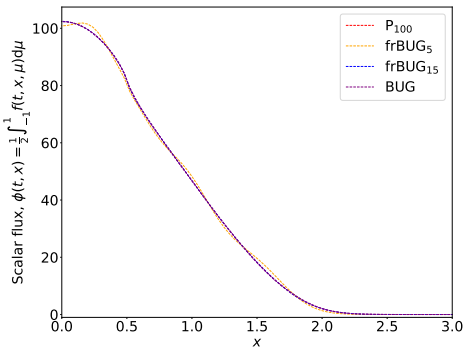
Figure 1: Relative mass error for the rectangular pulse test case in the kinetic and diffusive regime.



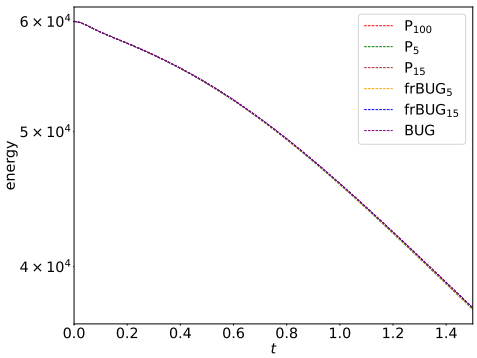




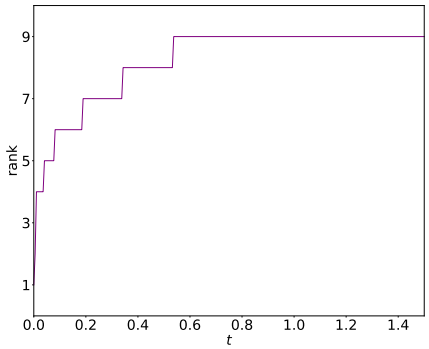
(a) moment methods



(b) low-rank methods



(c) energy over time



(d) rank over time BUG

Figure 2: Numerical results of the rectangular pulse test case in the kinetic regime, i.e., $\varepsilon = 1$ at $t = 1.5$. In the first row, we present the temperature profile at end-time for the moment and low-rank methods; in the second row, we have the corresponding scalar flux. In the last row, we have the energy of the system over time for all the methods and the rank evolution of the BUG integrator.

For $\varepsilon = 10^{-5}$, we use a coarser spatial grid with $N_x = 201$ cells. The rank used for the fixed-rank modal macro-micro BUG integrator is $r = 1$, and the modal macro-micro BUG integrator starts with the same initial rank $r = 1$. The tolerance parameter and end time are the same as

in the kinetic regime. We see from Figures 1 and 3 that the solutions from the full modal macro-micro integrator, fixed-rank modal macro-micro BUG integrator, and modal macro-micro BUG integrator agree well with the limiting Rosseland approximation. Additionally, all the methods dissipate energy over time.

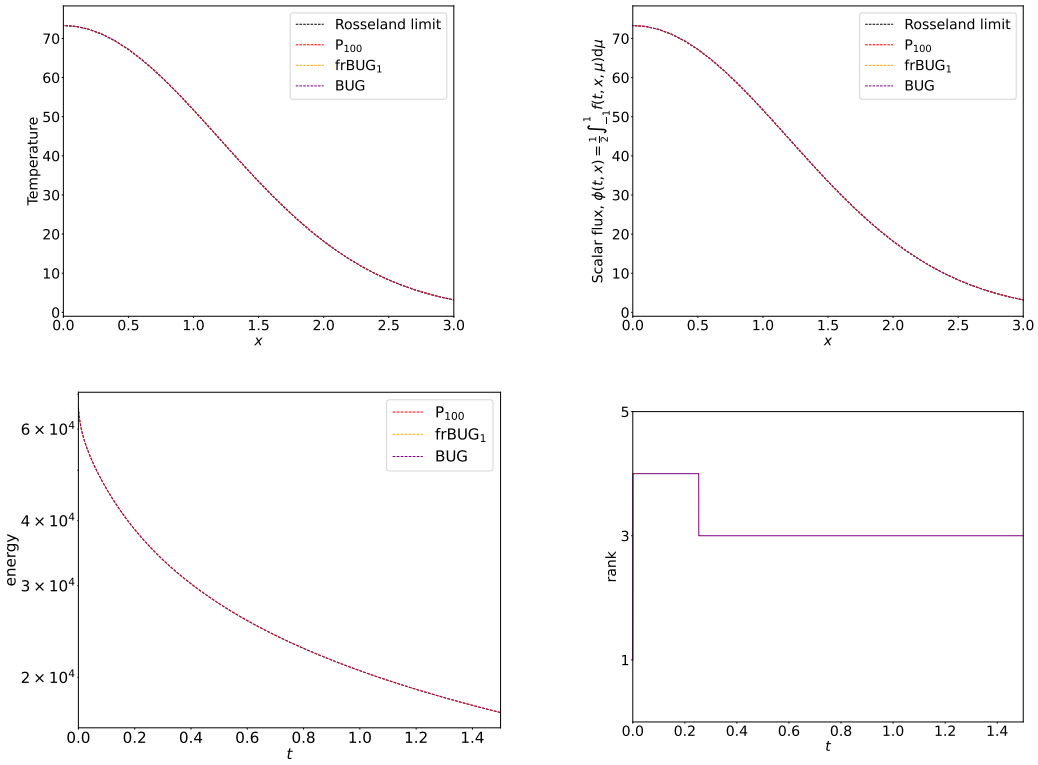


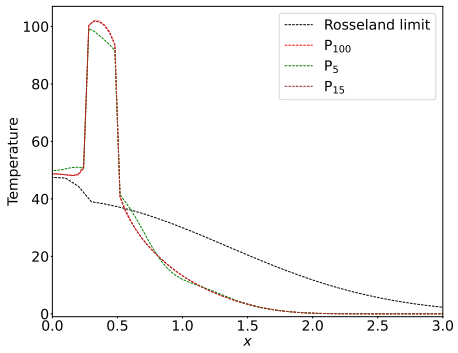
Figure 3: Numerical results of the rectangular pulse test case in the diffusive regime, i.e. $\varepsilon = 10^{-5}$ at $t = 1.5$. Top left: Temperature profile, Top right: Scalar flux, Bottom left: Energy of the system over time for all the methods, Bottom right: Rank evolution of rank-adaptive integrator over time.

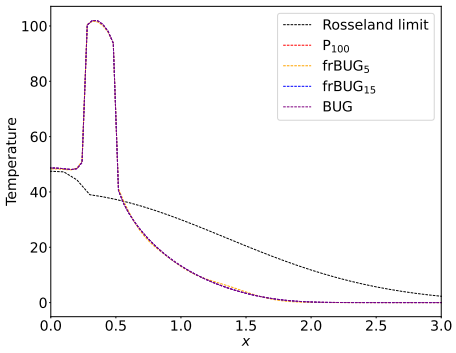
5.2 Absorber test case

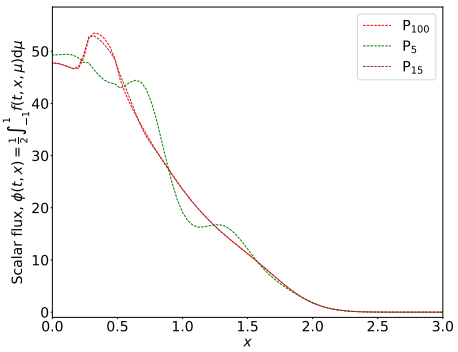
To study the behavior of the methods in an inhomogeneous medium, we place an absorber in the middle of the domain. That is, we set the absorption coefficient to

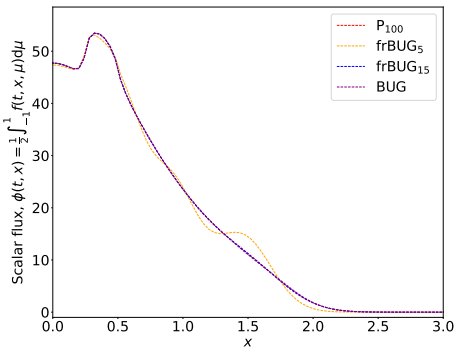
$$\sigma^a(x) = \begin{cases} 5, & \text{if } -0.25 \leq x \leq 0.25, \\ 0.5 & \text{else} \end{cases}.$$

The remaining parameters, along with the end time, are the same as in the rectangular pulse test case. The temperature and scalar flux, along with other parameters, are depicted in Figure 4 for $\varepsilon = 1$ and in Figure 5 for $\varepsilon = 10^{-5}$.



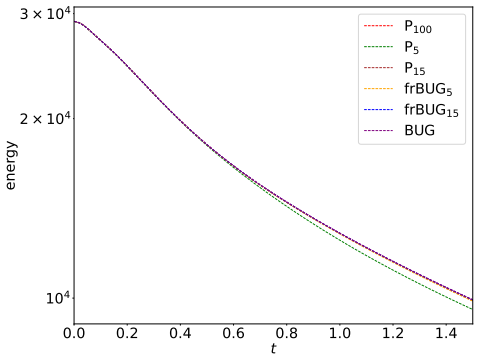


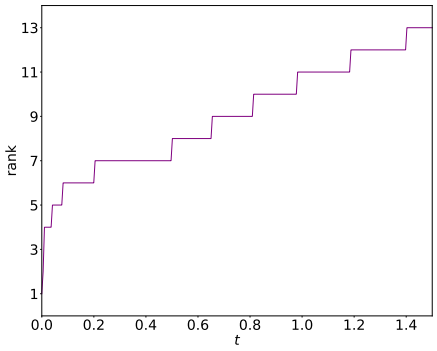




(a) moment methods

(b) low-rank methods





(c) energy over time

(d) rank over time BUG

Figure 4: Numerical results of the absorber test case in the kinetic regime, i.e., $\varepsilon = 1$ at $t = 1.5$. In the first row, we present the temperature profile at end-time for the moment and low-rank methods; in the second row, we have the corresponding scalar flux. In the last row, we have the energy of the system over time for all the methods and the rank evolution of the BUG integrator.

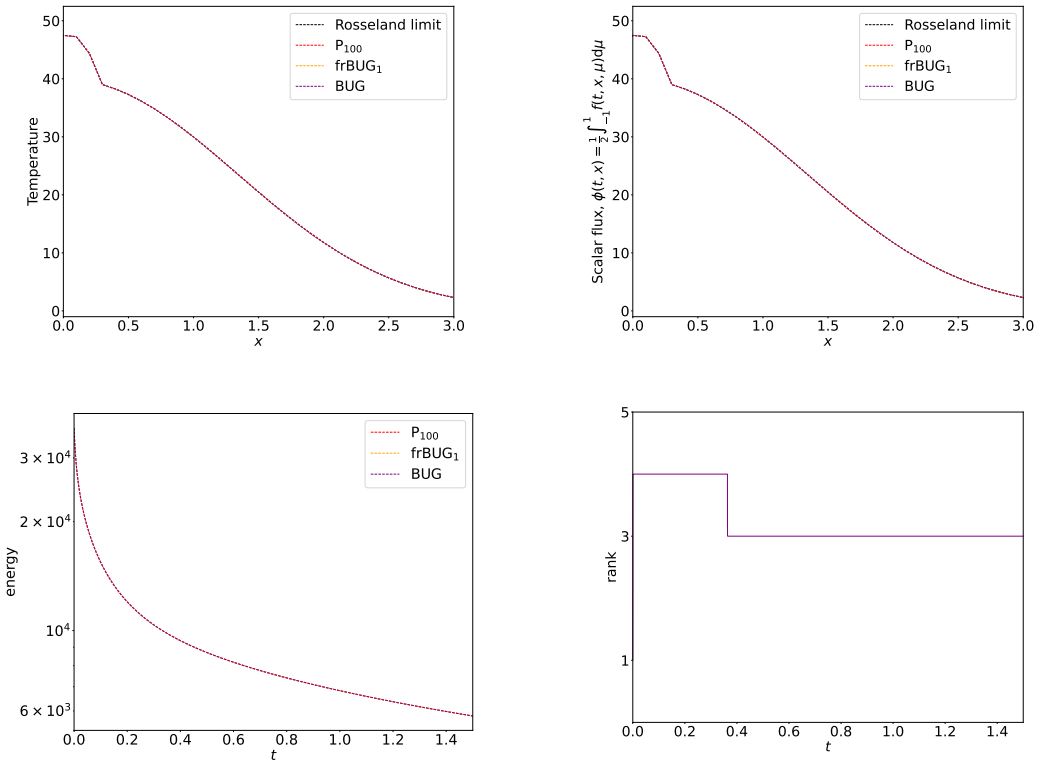


Figure 5: Numerical results of the absorber test case in the diffusive regime, i.e., $\varepsilon = 10^{-5}$ at $t = 1.5$. Top left: Temperature profile, Top right: Scalar flux, Bottom left: Energy of the system over time for all the methods, Bottom right: Rank evolution of rank-adaptive integrator over time.

6 Conclusion

In this work, we propose a modal macro-micro BUG integrator for the radiative heat transfer equations. We show that this integrator is energy stable for the linearized problem under a CFL condition that captures the kinetic regime and diffusive regime of the thermal radiative transfer equations. Additionally, the full modal macro-micro scheme's stability and the fixed-rank macro-micro BUG scheme have been investigated.

Declaration of competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

The work of Chinmay Patwardhan and Martin Frank was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Project-ID 258734477 – SFB 1173.

References

[1] J. R. Howell, M. P. Menguc, and R. Siegel, *Thermal Radiation Heat Transfer (5th ed.)* CRC Press, 2010. DOI: https://doi.org/10.1201/9781439894552.

[2] S. Rosseland, *Astrophysik auf atomtheoretischer Grundlage* (Struktur der Materie in Einzeldarstellungen ; 11). Berlin: Springer, 1931.

[3] S. Jin, "Asymptotic preserving (ap) schemes for multiscale kinetic and hyperbolic equations: A review.," *Lecture Notes for Summer School on "Methods and Models of Kinetic Theory" (M&MKT), Porto Ercole (Grosseto, Italy)*, pp. 177–216, 2010.

[4] J. Hu, S. Jin, and Q. Li, "Chapter 5 - asymptotic-preserving schemes for multiscale hyperbolic and kinetic equations," in *Handbook of Numerical Methods for Hyperbolic Problems*, ser. Handbook of Numerical Analysis, R. Abgrall and C.-W. Shu, Eds., vol. 18, Elsevier, 2017, pp. 103–129. DOI: https://doi.org/10.1016/bs.hna.2016.09.001.

[5] A. Klar, "An asymptotic preserving numerical scheme for kinetic equations in the low Mach number limit," *SIAM J. Numer. Anal.*, vol. 36, no. 5, pp. 1507–1527, 1999, ISSN: 0036-1429,1095-7170.

[6] M. Lemou and L. Mieussens, "A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit," *SIAM J. Sci. Comput.*, vol. 31, no. 1, pp. 334–368, 2008, ISSN: 1064-8275,1095-7197. DOI: 10.1137/07069479X.

[7] O. Koch and C. Lubich, "Dynamical low-rank approximation," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 2, pp. 434–454, 2007. DOI: 10.1137/050639703.

[8] C. Lubich and I. V. Oseledets, "A projector-splitting integrator for dynamical low-rank approximation," *Bit Numer Math*, vol. 54, pp. 171–188, 2014. DOI: 10.1007/s10543-013-0454-0.

[9] G. Ceruti and C. Lubich, "An unconventional robust integrator for dynamical low-rank approximation," *Bit Numer Math*, vol. 62, pp. 23–44, 2022. DOI: 10.1007/s10543-021-00873-0.

[10] G. Ceruti, J. Kusch, and C. Lubich, "A rank-adaptive robust integrator for dynamical low-rank approximation," *Bit Numer Math*, vol. 62, pp. 1149–1174, 2022. DOI: 10.1007/s10543-021-00907-7.

[11] G. Ceruti, J. Kusch, and C. Lubich, *A parallel rank-adaptive integrator for dynamical low-rank approximation*, 2023. arXiv: 2304.05660 [math.NA].

[12] G. Ceruti, L. Einkemmer, J. Kusch, and C. Lubich, "A robust second-order low-rank bug integrator based on the midpoint rule," *arXiv preprint arXiv:2402.08607*, 2024.

[13] J. Kusch and P. Stammer, "A robust collision source method for rank adaptive dynamical low-rank approximation in radiation therapy," *ESAIM: M2AN*, vol. 57, no. 2, pp. 865–891, 2023. DOI: 10.1051/m2an/2022090.

[14] L. Einkemmer, J. Hu, and Y. Wang, "An asymptotic-preserving dynamical low-rank method for the multi-scale multi-dimensional linear transport equation," *Journal of Computational Physics*, vol. 439, p. 110 353, 2021, ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2021.110353.

[15] J. Kusch, B. Whewell, R. McClarren, and M. Frank, "A low-rank power iteration scheme for neutron transport criticality problems," *Journal of Computational Physics*, vol. 470, p. 111 587, Dec. 2022. DOI: 10.1016/j.jcp.2022.111587.

[16] G. Ceruti, M. Frank, and J. Kusch, "Dynamical low-rank approximation for Marshak waves," Karlsruhe Institute of Technology, CRC 1173 Preprint 2022/76, Dec. 2022. DOI: 10.5445/IR/1000154134.

[17] L. Baumann, L. Einkemmer, C. Klingenberg, and J. Kusch, *Energy stable and conservative dynamical low-rank approximation for the su-olson problem*, 2023. arXiv: 2307.07538 [math.NA].

[18] L. Einkemmer, J. Hu, and Y. Wang, "An asymptotic-preserving dynamical low-rank method for the multi-scale multi-dimensional linear transport equation," *J. Comput. Phys.*, vol. 439, Paper No. 110353, 21, 2021, ISSN: 0021-9991,1090-2716. DOI: 10.1016/j.jcp.2021.110353.

[19] L. Einkemmer, J. Hu, and J. Kusch, "Asymptotic-preserving and energy stable dynamical low-rank approximation," *SIAM Journal on Numerical Analysis*, vol. 62, no. 1, pp. 73–92, 2024. DOI: 10.1137/23M1547603.

[20] A. Klar and C. Schmeiser, "Numerical passage from radiative heat transfer to nonlinear diffusion models," *Math. Models Methods Appl. Sci.*, vol. 11, no. 5, pp. 749–767, 2001, ISSN: 0218-2025,1793-6314. DOI: 10.1142/S0218202501001082.

[21] L. Einkemmer, A. Ostermann, and C. Scalone, "A robust and conservative dynamical low-rank algorithm," *Journal of Computational Physics*, vol. 484, p. 112 060, 2023, ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2023.112060.

[22] L. Einkemmer, J. Kusch, and S. Schotthöfer, *Conservation properties of the augmented basis update & galerkin integrator for kinetic problems*, 2023. arXiv: 2311.06399 [math.NA].

[23] J. Koellermeier, P. Krah, and J. Kusch, *Macro-micro decomposition for consistent and conservative model order reduction of hyperbolic shallow water moment equations: A study using pod-galerkin and dynamical low rank approximation*, 2023. arXiv: 2302.01391 [math.NA].

[24] K. M. Case and P. F. Zweifel, *Linear transport theory*. Addison-Wesley, 1967.

[25] S. Jin and H. Lu, "An asymptotic-preserving stochastic galerkin method for the radiative heat transfer equations with random inputs and diffusive scalings," *Journal of Computational Physics*, vol. 334, pp. 182–206, 2017, ISSN: 0021-9991. DOI: https://doi.org/10.1016/j.jcp.2016.12.033.

[26] B. Su and G. L. Olson, "An analytical benchmark for non-equilibrium radiative transfer in an isotropically scattering medium," *Annals of Nuclear Energy*, vol. 24, no. 13, pp. 1035–1055, 1997, ISSN: 0306-4549. DOI: https://doi.org/10.1016/S0306-4549(96)00100-4.

[27] C. Patwardhan, J. Kusch, and M. Frank, *Numerical testcases for "asymptotic-preserving and energy stable dynamical low-rank approximation for thermal radiative transfer equations"*, 2024. [Online]. Available: https://github.com/chinsp/publication-Asymptotic-preserving-and-energy-stable-DLRA-for-thermal-radiative-transfer-equations.git.

A Proof of Theorem 2

We start by stating some lemmas and properties used to prove stability in energy norm (Theorem 2) for the linearized modal macro-micro scheme (3.3).

Lemma 2 (Lemma 3.3 [19](Summation by parts))**.** *For vectors $\boldsymbol{\phi}_{i+1/2}, \boldsymbol{\zeta}_{i+1/2} \in \mathbb{R}^N$ where $i = 0, \dots, N_x$, the equality*

$$\sum_i \boldsymbol{\zeta}_{i+1/2}^\top \mathcal{D}^\pm \boldsymbol{\phi}_{i+1/2} = -\sum_i (\mathcal{D}^\mp \boldsymbol{\zeta}_{i+1/2})^\top \boldsymbol{\phi}_{i+1/2} \tag{A.1}$$

holds for periodic or zero values at the boundary.

Lemma 3. *Let $\boldsymbol{\phi}_{i+1/2} \in \mathbb{R}^{N+1}$, for $i = 0, \dots, Nx$, then the following inequality holds*

$$\sum_i \left(\mathcal{D}^+ \boldsymbol{\phi}_{i+1/2}\right)^2 \leq \frac{4}{\Delta x^2} \sum_i (\boldsymbol{\phi}_{i+1/2})^2. \tag{A.2}$$

Proof. Expanding the left-hand side using the definition of $\mathcal{D}^+$

$$\begin{aligned}\sum_i \left(\mathcal{D}^+ \boldsymbol{\phi}_{i+1/2}\right)^2 &= \frac{1}{\Delta x^2} \sum_i (\boldsymbol{\phi}_{i+3/2} - \boldsymbol{\phi}_{i+1/2})^2 \\ &= \frac{2}{\Delta x^2} \sum_i (\boldsymbol{\phi}_{i+1/2})^2 - \frac{2}{\Delta x^2} \sum_i \boldsymbol{\phi}_{i+3/2}^\top \boldsymbol{\phi}_{i+1/2}.\end{aligned}$$

Using Young's inequality for the last term on the right-hand side in the above equation we get

$$\left| \frac{2}{\Delta x^2} \sum_i \boldsymbol{\phi}_{i+3/2}^\top \boldsymbol{\phi}_{i+1/2} \right| \leq \frac{1}{\Delta x^2} \sum_i (\boldsymbol{\phi}_{i+1/2})^2 + \frac{1}{\Delta x^2} \sum_i (\boldsymbol{\phi}_{i+3/2})^2. \tag{A.3}$$

A change in the index in the last term gives the desired result. □

The constructed macro-micro system (3.3) with the stabilization matrix $|\mathbf{A}|$ is related to the full P_N system through the flux matrix $\mathbf{A}_f = \left(\left[P_{i-1} P_{j-1} \mu \right]_\mu \right)_{i,j=1}^{N+1}$ and Roe matrix $|\mathbf{A}_f| = \mathbf{T}_f \, |\mathbf{M}| \, \mathbf{T}_f^\top$, where $\mathbf{T}_f = \left(\sqrt{w_k} P_{i-1}(\mu_k) \right)_{i,j=1}^{N+1}$ such that $\mathbf{A}_f = \mathbf{T}_f \mathbf{M} \mathbf{T}_f^\top$. We additionally define

$$\frac{1}{\gamma_0} \boldsymbol{b} = \mathbf{a} = (a_0, 0, \ldots, 0)^\top \in \mathbb{R}^N, \qquad \mathbf{a}_f = (0, a_0, 0, \ldots, 0)^\top \in \mathbb{R}^{N+1}.$$

Lemma 4 (Lemma 3.4 [19] (P_N preservation))**.** *For a given vector $\boldsymbol{g} \in \mathbb{R}^N$ define its extension $\boldsymbol{v} \coloneqq (0, g_1, \ldots, g_N)^\top \in \mathbb{R}^{N+1}$ as well as $\hat{\boldsymbol{v}}_{i+1/2} \coloneqq \boldsymbol{T}_f^\top \boldsymbol{v}_{i+1/2} \in \mathbb{R}^{N+1}$. Then,*

$$\boldsymbol{g}^\top \boldsymbol{A}^2 \boldsymbol{g} = \hat{\boldsymbol{v}}^\top \boldsymbol{M}^2 \hat{\boldsymbol{v}}, \quad \boldsymbol{g}^\top |\boldsymbol{A}| \, \boldsymbol{g} = \hat{\boldsymbol{v}}^\top |\boldsymbol{M}| \, \hat{\boldsymbol{v}}, \quad \boldsymbol{g}^\top \boldsymbol{a}\boldsymbol{a}^\top \boldsymbol{g} = \hat{\boldsymbol{v}}^\top \boldsymbol{T}_f^\top \boldsymbol{a}_f \boldsymbol{a}_f^\top \boldsymbol{T}_f \hat{\boldsymbol{v}}.$$

Two main properties of the advection operator, $\mathcal{L}$, that are used in proving energy stability are

Lemma 5 (Lemma 3.5 [19] (Positivity))**.** *For a given discrete function $\boldsymbol{g}_{i+1/2}^n$, the advection operator fulfills the properties*

$$\sum_j \boldsymbol{g}_{i+1/2}^{n+1,\top} \mathcal{L} \boldsymbol{g}_{i+1/2}^{n+1} = \sum_i \frac{\Delta x}{2} \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \, |\boldsymbol{A}| \, \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} \geq 0$$

and

$$\begin{aligned} \sum_j \boldsymbol{g}_{i+1/2}^{n+1,\top} \mathcal{L} \boldsymbol{g}_{i+1/2}^{n} = \sum_i \frac{\Delta x}{2} \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \, |\boldsymbol{A}| \, \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} \\ + \sum_i (\boldsymbol{g}_{i+1/2}^{n} - \boldsymbol{g}_{i+1/2}^{n+1})^\top (\boldsymbol{A}^+ \mathcal{D}^+ + \boldsymbol{A}^- \mathcal{D}^-) \boldsymbol{g}_{i+1/2}^{n+1}. \end{aligned}$$

Lemma 6 (Lemma 3.6 [19] (Boundedness))**.** *For a given discrete function $\boldsymbol{g}_{i+1/2}^n$, the advection operator fulfills the property*

$$\sum_i \left[(\boldsymbol{A}^+ \mathcal{D}^+ + \boldsymbol{A}^- \mathcal{D}^-) \boldsymbol{g}_{i+1/2}^{n+1} \right]^2 \leq 2 \beta_N \sum_i \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \boldsymbol{A}^2 \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1},$$

where $\beta_N = \max_k \; w_k (N+1)$ is bounded for all N.

With these lemma we now present the proof of theorem 2:

Proof. We start by plugging in the update equation of the macro variable, T, (3.3c) into the update equation of the mesoscopic variable, h, (3.3a). This gives

$$\frac{(aT_i^{n+1} + \frac{\varepsilon^2}{c} h_i^{n+1}) - (aT_i^n + \frac{\varepsilon^2}{c} h_i^n)}{\Delta t} + \frac{\gamma_1}{2} \mathcal{D}^0 g_{1,i}^{n+1} = -\sigma_i^a h_i^{n+1}, \tag{A.4a}$$

$$\frac{1}{c} \left(\frac{\boldsymbol{g}_{i+1/2}^{n+1} - \boldsymbol{g}_{i+1/2}^n}{\Delta t} \right) + \frac{1}{\varepsilon} \mathcal{L} \boldsymbol{g}_{i+1/2}^n = -\frac{\sigma_{i+1/2}^a}{\varepsilon^2} \boldsymbol{g}_{i+1/2}^{n+1} - \frac{1}{\varepsilon^2} \boldsymbol{b} \, \delta^0 (acT_{i+1/2}^n + \varepsilon^2 h_{i+1/2}^n). \tag{A.4b}$$

Next we multiply the update equation for the micro equation (A.4a) by $(aT_i^{n+1} + \frac{\varepsilon^2}{c} h_i^{n+1}) \Delta x$, and sum over i,

$$\begin{aligned} \frac{1}{\Delta t} \sum_i (aT_i^{n+1} + \frac{\varepsilon^2}{c} h_i^{n+1})^2 \Delta x - \frac{1}{\Delta t} \sum_i (aT_i^{n+1} + \frac{\varepsilon^2}{c} h_i^{n+1})(aT_i^n + \frac{\varepsilon^2}{c} h_i^n) \Delta x \\ + \frac{\gamma_1}{2} \sum_i (aT_i^{n+1} + \frac{\varepsilon^2}{c} h_i^{n+1}) \mathcal{D}^0 g_{1,i}^{n+1} \Delta x = - \sum_i (aT_i^{n+1} + \frac{\varepsilon^2}{c} h_i^{n+1}) \sigma_i^a h_i^{n+1} \Delta x. \end{aligned}$$

Using the summation of products property 1 we get

$$\frac{1}{2\Delta t}\left(\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}\right\|^2-\left\|aT^{n}+\frac{\varepsilon^2}{c}h^{n}\right\|^2+\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}-aT^{n}-\frac{\varepsilon^2}{c}h^{n}\right\|^2\right)+ \\ \frac{\gamma_1}{2}\sum_i(aT_i^{n+1}+\frac{\varepsilon^2}{c}h_i^{n+1})\mathcal{D}^0 g_{1,i}^{n+1}\Delta x=-\sum_i(aT_i^{n+1}+\frac{\varepsilon^2}{c}h_i^{n+1})\sigma_i^a h_i^{n+1}\Delta x \tag{A.5}$$

Multiply (A.4b) by $\boldsymbol{g}_{i+1/2}^{n+1,\top}\Delta x$, and sum over i, and using the summation property 1

$$\frac{1}{2\Delta t}\left(\frac{1}{c}\left\|\mathbf{g}^{n+1}\right\|^2-\frac{1}{c}\left\|\mathbf{g}^{n}\right\|^2+\frac{1}{c}\left\|\mathbf{g}^{n+1}-\mathbf{g}^{n}\right\|^2\right)+\frac{1}{\varepsilon}\sum_i \boldsymbol{g}_{i+1/2}^{n+1,\top}\mathcal{L}\boldsymbol{g}_{i+1/2}^{n}\Delta x \\ \leq-\frac{\sigma_0}{\varepsilon^2}\left\|\mathbf{g}^{n+1}\right\|^2-\frac{\gamma_1}{\varepsilon^2}\sum_i g_{1,i+1/2}^{n+1}\,\delta^0(acT_{i+1/2}^{n}+\varepsilon^2 h_{i+1/2}^{n})\Delta x, \tag{A.6}$$

where we use $\sigma_0\leq\sigma_{i+1/2}^a,\forall i$, and $\boldsymbol{g}_{i+1/2}^{n+1,\top}\boldsymbol{b}=\gamma_1 g_{1,i+1/2}^{n+1}$.
Then (A.5) $+\frac{\varepsilon^2}{\gamma_0^2 c}\times$ (A.6), where $\gamma_0^2=2$ is the normalization factor of the zeroth order Legendre polynomial, gives

$$\begin{aligned}&\frac{1}{2\Delta t}\left(\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}\right\|^2+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}\right\|^2-\left\|aT^{n}+\frac{\varepsilon^2}{c}h^{n}\right\|^2-\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n}\right\|^2\right.\\&\left.+\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}-aT^{n}-\frac{\varepsilon^2}{c}h^{n}\right\|^2+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}-\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n}\right\|^2\right)\\&+\frac{\gamma_1}{2}\sum_i(aT_i^{n+1}+\frac{\varepsilon^2}{c}h_i^{n+1})\mathcal{D}^0 g_{1,i}^{n+1}\Delta x+\frac{\varepsilon}{2c}\sum_i \boldsymbol{g}_{i+1/2}^{n+1,\top}\mathcal{L}\boldsymbol{g}_{i+1/2}^{n}\Delta x\\&\leq-\sum_i(aT_i^{n+1}+\frac{\varepsilon^2}{c}h_i^{n+1})\sigma_i^a h_i^{n+1}\Delta x-\frac{\sigma_0}{2c}\left\|\mathbf{g}^{n+1}\right\|^2-\frac{\gamma_1}{\gamma_0^2}\sum_i g_{1,i+1/2}^{n+1}\,\delta^0(aT_{i+1/2}^{n}+\frac{\varepsilon^2}{c}h_{i+1/2}^{n})\Delta x\,.\end{aligned} \tag{A.7}$$

Using discrete integration by parts from Lemma 2 we get

$$\sum_i g_{1,i+1/2}^{n+1}\,\delta^0(aT_{i+1/2}^{n}+\frac{\varepsilon^2}{c}h_{i+1/2}^{n})\Delta x=-\sum_i \mathcal{D}^0 g_{1,i}^{n+1}(aT_i^{n}+\frac{\varepsilon^2}{c}h_i^{n})\Delta x. \tag{A.8}$$

With the use of (A.8) we rewrite (A.7) as

$$\begin{aligned}&\frac{1}{2\Delta t}\left(\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}\right\|^2+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}\right\|^2-\left\|aT^{n}+\frac{\varepsilon^2}{c}h^{n}\right\|^2-\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n}\right\|^2\right.\\&\left.+\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}-aT^{n}-\frac{\varepsilon^2}{c}h^{n}\right\|^2+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}-\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n}\right\|^2\right)\\&+\frac{\varepsilon}{2c}\sum_i \boldsymbol{g}_{i+1/2}^{n+1,\top}\mathcal{L}\boldsymbol{g}_{i+1/2}^{n}\Delta x\leq-\frac{\sigma_0}{2c}\left\|\mathbf{g}^{n+1}\right\|^2-\sum_i(aT_i^{n+1}+\frac{\varepsilon^2}{c}h_i^{n+1})\sigma_i^a h_i^{n+1}\Delta x\\&+\frac{\gamma_1}{2}\sum_i \mathcal{D}^0 g_{1,i}^{n+1}(-aT_i^{n+1}-\frac{\varepsilon^2}{c}h_i^{n+1}+aT_i^{n}+\frac{\varepsilon^2}{c}h_i^{n})\Delta x.\end{aligned}$$

Using Young's inequality,

$$\frac{\gamma_1}{2}\sum_i \mathcal{D}^0 g_{1,i}^{n+1}(-aT_i^{n+1}-\frac{\varepsilon^2}{c}h_i^{n+1}+aT_i^{n}+\frac{\varepsilon^2}{c}h_i^{n})\Delta x\leq\alpha\left\|-aT^{n+1}-\frac{\varepsilon^2}{c}h^{n+1}+aT^{n}+\frac{\varepsilon^2}{c}h^{n}\right\| \\ +\frac{1}{4\alpha}\sum_i\frac{\gamma_1^2}{4}(\mathcal{D}^0 g_{1,i}^{n+1})^2\Delta x.$$

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

$$\begin{aligned}\frac{1}{2\Delta t}\Bigg(&\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}\right\|^2+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}\right\|^2-\left\|aT^{n}+\frac{\varepsilon^2}{c}h^{n}\right\|^2-\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n}\right\|^2\\&+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}-\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n}\right\|^2\Bigg)+\frac{\varepsilon}{2c}\sum_i \boldsymbol{g}_{i+1/2}^{n+1,\top}\mathcal{L}\boldsymbol{g}_{i+1/2}^{n}\Delta x\\&\le-\frac{\sigma_0}{2c}\left\|\mathbf{g}^{n+1}\right\|^2-\sum_i(aT_i^{n+1}+\frac{\varepsilon^2}{c}h_i^{n+1})\sigma_i^a h_i^{n+1}\Delta x+\frac{\Delta t}{2}\frac{\gamma_1^2}{4}\sum_i(\mathcal{D}^0 g_{1,i}^{n+1})^2\Delta x.\end{aligned} \tag{A.9}$$

This gives an upper bound for the term $\left\|aT^{n+1}+\frac{\varepsilon^2}{c}h^{n+1}\right\|^2+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}\right\|$. Next, we derive an upper bound for $\left\|\sqrt{\frac{a}{\kappa}}T^{n+1}\right\|^2$. For that we multiply the temperature update (3.3c) by $aT_i^{n+1}\Delta x$ and sum over i

$$\frac{1}{\Delta t}\sum_i aT_i^{n+1}(T_i^{n+1}-T_i^n)\Delta x=\kappa\sum_i aT_i^{n+1}\sigma_i^a h_i^{n+1}\Delta x.$$

Thus, we get using the summation of products property 1

$$\frac{1}{2\Delta t}\left(\left\|\sqrt{\frac{a}{\kappa}}T^{n+1}\right\|^2-\left\|\sqrt{\frac{a}{\kappa}}T^{n}\right\|^2+\left\|\sqrt{\frac{a}{\kappa}}T^{n+1}-\sqrt{\frac{a}{\kappa}}T^{n}\right\|^2\right)=\sum_i aT_i^{n+1}\sigma_i^a h_i^{n+1}\Delta x. \tag{A.10}$$

Adding (A.9) and (A.10) gives

$$\begin{aligned}\frac{1}{2\Delta t}\Bigg(&e^{n+1}-e^n+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}-\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n}\right\|^2\Bigg)+\frac{\varepsilon}{2c}\sum_i \boldsymbol{g}_{i+1/2}^{n+1,\top}\mathcal{L}\boldsymbol{g}_{i+1/2}^{n}\Delta x\\&\le-\frac{\sigma_0}{2c}\left\|\mathbf{g}^{n+1}\right\|^2-\frac{\sigma_0}{c}\left\|\varepsilon h^{n+1}\right\|^2+\frac{\Delta t}{2}\frac{\gamma_1^2}{4}\sum_i(\mathcal{D}^0 g_{1,i}^{n+1})^2\Delta x-\left\|\sqrt{\frac{a}{\kappa}}T^{n+1}-\sqrt{\frac{a}{\kappa}}T^{n}\right\|^2.\end{aligned} \tag{A.11}$$

Since $-\frac{\sigma_0}{c}\left\|\varepsilon h^{n+1}\right\|^2\le 0$ and $-\left\|\sqrt{\frac{a}{\kappa}}T^{n+1}-\sqrt{\frac{a}{\kappa}}T^{n}\right\|^2\le 0$ we bound them from above by 0.

Then (A.11) becomes

$$\begin{aligned}\frac{1}{2\Delta t}\Bigg(&e^{n+1}-e^n+\left\|\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1}-\frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n}\right\|^2\Bigg)+\frac{\varepsilon}{2c}\sum_i \boldsymbol{g}_{i+1/2}^{n+1,\top}\mathcal{L}\boldsymbol{g}_{i+1/2}^{n}\Delta x\\&\le-\frac{\sigma_0}{2c}\left\|\mathbf{g}^{n+1}\right\|^2+\frac{\Delta t}{2}\frac{\gamma_1^2}{4}\sum_i(\mathcal{D}^0 g_{1,i}^{n+1})^2\Delta x.\end{aligned} \tag{A.12}$$

Next, we split the last term on the left-hand side of (A.12) as

$$\sum_i \boldsymbol{g}_{i+1/2}^{n+1}\mathcal{L}\boldsymbol{g}_{i+1/2}^{n}\Delta x=P+Q$$

where

$$P=\sum_i \boldsymbol{g}_{i+1/2}^{n+1}\mathcal{L}\boldsymbol{g}_{i+1/2}^{n+1}\Delta x,\quad Q=\sum_i \boldsymbol{g}_{i+1/2}^{n+1}\mathcal{L}(\boldsymbol{g}_{i+1/2}^{n}-\boldsymbol{g}_{i+1/2}^{n+1})\Delta x.$$

Then, using Lemma 5 we have,

$$\begin{aligned}P&=\frac{\Delta x}{2}\sum_i \mathcal{D}^+\boldsymbol{g}_{i+1/2}^{n+1,\top}\left|\mathbf{A}\right|\mathcal{D}^+\boldsymbol{g}_{i+1/2}^{n+1}\Delta x,\\Q&=-\sum_i(\mathbf{A}^+\mathcal{D}^+ +\mathbf{A}^-\mathcal{D}^-)\boldsymbol{g}_{i+1/2}^{n+1,\top}(\boldsymbol{g}_{i+1/2}^{n}-\boldsymbol{g}_{i+1/2}^{n+1}).\end{aligned} \tag{A.13}$$

Thus, using Young's inequality

$$|Q| \leq \alpha \left\|\mathbf{g}^{n+1} - \mathbf{g}^{n}\right\| + \frac{1}{4\alpha} \sum_i \left[(\mathbf{A}^+\mathcal{D}^+ + \mathbf{A}^-\mathcal{D}^-)\boldsymbol{g}_{i+1/2}^{n+1}\right]^2 \Delta x$$

and Lemma 5 we get

$$|Q| \leq \alpha \left\|\mathbf{g}^{n+1} - \mathbf{g}^{n}\right\| + \frac{2\beta_N}{4\alpha} \sum_i \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \mathbf{A}^2 \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} \Delta x. \tag{A.14}$$

We set $\alpha = \frac{\varepsilon}{2c\Delta t}$, then $\left\| \frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n+1} - \frac{\varepsilon}{\gamma_0 c}\mathbf{g}^{n} \right\|$ gets canceled out and we get

$$\begin{aligned}\frac{1}{2\Delta t}\left(e^{n+1} - e^{n}\right) + \frac{\varepsilon\Delta x}{4c} \sum_i \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} |\mathbf{A}| \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} \Delta x - \beta_N \frac{\Delta t}{2} \sum_i \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \mathbf{A}^2 \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} \Delta x \\ \leq -\frac{\sigma_0}{2c}\left\|\mathbf{g}^{n+1}\right\|^2 + \frac{\Delta t}{2}\frac{\gamma_1^2}{4} \sum_i (\mathcal{D}^0 g_{1,i}^{n+1})^2 \Delta x\end{aligned} \tag{A.15}$$

We rewrite the last term on the right-hand side of (A.15) as

$$\frac{\gamma_1^2}{4} \sum_i (\mathcal{D}^0 g_{1,i}^{n+1})^2 \Delta x = \frac{\gamma_1^2}{4} \sum_i (\mathcal{D}^+ g_{1,i}^{n+1})^2 \Delta x = \frac{1}{2} \sum_i \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \mathbf{a}\mathbf{a}^\top \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} \Delta x$$

and we get

$$\begin{aligned}\frac{1}{2\Delta t}\left(e^{n+1} - e^{n}\right) \leq -\frac{\sigma_0}{2c}\left\|\mathbf{g}^{n+1}\right\|^2 \\ + \frac{1}{2} \sum_i \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \left[\Delta t \left(\beta_N \mathbf{A}^2 + \frac{1}{2}\mathbf{a}\mathbf{a}^\top\right) - \frac{\varepsilon\Delta x}{2c}|\mathbf{A}|\right] \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} \Delta x.\end{aligned} \tag{A.16}$$

If we define $\mathbf{v} = (0, g_1, \ldots, g_N)^\top \in \mathbb{R}^{N+1}$ and $\hat{\mathbf{v}} = \mathbf{T}_f^\top \mathbf{v} \in \mathbb{R}^{N+1}$, then we have from Lemma 4

$$\mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \mathbf{A}^2 \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} = \mathcal{D}^+ \hat{\mathbf{v}}_{i+1/2}^{n+1,\top} \mathbf{M}^2 \mathcal{D}^+ \hat{\mathbf{v}}_{i+1/2}^{n+1},$$

$$\mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} |\mathbf{A}| \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} = \mathcal{D}^+ \hat{\mathbf{v}}_{i+1/2}^{n+1,\top} |\mathbf{M}| \mathcal{D}^+ \hat{\mathbf{v}}_{i+1/2}^{n+1},$$

and

$$\mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1,\top} \mathbf{a}\mathbf{a}^\top \mathcal{D}^+ \boldsymbol{g}_{i+1/2}^{n+1} = \mathcal{D}^+ \hat{\mathbf{v}}_{i+1/2}^{n+1,\top} \mathbf{T}_f^\top \mathbf{a}_f \mathbf{a}_f^\top \mathbf{T}_f \mathcal{D}^+ \hat{\mathbf{v}}_{i+1/2}^{n+1}.$$

As given in [19, Thoerem 3.2], since $\mathbf{T}_f^\top \mathbf{a}_f = \frac{1}{\sqrt{3}} \mathbf{T}_f^\top \mathbf{e}_2 = \sqrt{\frac{w_k}{3}} P_1(\mu_k) = \sqrt{\frac{w_k}{2}}\mu_k$, we have

$$\begin{aligned}\mathcal{D}^+ \hat{\mathbf{v}}_{i+1/2}^{n+1,\top} \mathbf{T}_f^\top \mathbf{a}_f \mathbf{a}_f^\top \mathbf{T}_f \mathcal{D}^+ \hat{\mathbf{v}}_{i+1/2}^{n+1} &= \left(\sum_{k=1}^{N+1} \mathcal{D}^+ \hat{v}_{i+1/2,k}^{n+1} \sqrt{\frac{w_k}{2}} \mu_k\right)^2 \\ &= \frac{1}{2} \sum_{k,\ell}^{N+1} \mathcal{D}^+ \hat{v}_{i+1/2,k}^{n+1} \sqrt{w_k}\mu_k \mathcal{D}^+ \hat{v}_{\ell,i+1/2}^{n+1} \sqrt{w_\ell}\mu_\ell \\ &\overset{\text{Young}}{\leq} \frac{1}{4} \sum_{k,\ell}^{N+1} \left(\mathcal{D}^+ \hat{v}_{i+1/2,k}^{n+1}\right)^2 w_k \mu_k^2 \\ &\qquad + \frac{1}{4} \sum_{k,\ell}^{N+1} \left(\mathcal{D}^+ \hat{v}_{\ell,i+1/2}^{n+1}\right)^2 w_\ell \mu_\ell^2 \\ &= \frac{N+1}{2} \sum_{k}^{N+1} \left(\mathcal{D}^+ \hat{v}_{i+1/2,k}^{n+1}\right)^2 w_k \mu_k^2.\end{aligned}$$

Since we can write $\left\|\mathbf{g}^{n+1}\right\|^2 = \left\|\hat{\mathbf{v}}^{n+1}\right\|^2 = \sum_{k,i}(\hat{v}^{n+1}_{i+1/2,k})^2\Delta x$ we have that (A.16) becomes

$$\frac{1}{2\Delta t}\left(e^{n+1}-e^n\right) \leq -\frac{\sigma_0}{2c}\sum_{k,i}(\hat{v}^{n+1}_{i+1/2,k})^2\Delta x + \frac{1}{2}\sum_{k,i}(\mathcal{D}^+\hat{v}^{n+1}_{i+1/2,k})^2\left[\Delta t\left(\beta_N\mu_k^2+\frac{N+1}{4}w_k\mu_k^2\right)-\frac{\varepsilon\Delta x}{2c}\left|\mu_k\right|\right]\Delta x.$$

Using Lemma 3 we get $\sum_i\left(\mathcal{D}^+\hat{\mathbf{v}}^{n+1}_{i+1/2}\right)^2 \leq \frac{4}{\Delta x^2}\sum_i(\hat{\mathbf{v}}^{n+1}_{i+1/2})^2$, thus

$$\frac{1}{2}\left(e^{n+1}-e^n\right) \leq \frac{1}{2}\frac{\Delta t}{\Delta x}\sum_{k,i}(\hat{v}^{n+1}_{i+1/2,k})^2\left[\frac{4\Delta t}{\Delta x^2}\left(\beta_N\mu_k^2+\frac{1}{4}\beta_N\mu_k^2\right)-\frac{4\varepsilon}{2c\Delta x}\left|\mu_k\right|-\frac{\sigma_0}{c}\right].$$

Hence for ensuring stability we must have for all k, where $\mu_k \neq 0$,

$$\Delta t \leq \frac{1}{5c\beta_N}\left(\frac{\sigma_0\Delta x^2}{\mu_k^2}+\frac{2\varepsilon\Delta x}{|\mu_k|}\right).$$

We note that β_N remains bounded for all N. □

B Proof of Lemma 1

Proof. As the SVD of $\widehat{\mathbf{S}}^{\mathrm{rem}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{W}^\top$ (4.17) where $\mathbf{U}$ and $\mathbf{W}$ are orthogonal, we have

$$\widehat{\mathbf{U}}^\top\left(\widehat{\mathbf{S}}^{\mathrm{rem}}\right)^{-\top} = (\widehat{\mathbf{U}}^\top\mathbf{U})\boldsymbol{\Sigma}^{-\top}\mathbf{W}^\top = (\mathbf{S}^{\mathrm{rem}})^{-\top}\widehat{\mathbf{W}}^\top.$$

Consider

$$\begin{aligned}\widehat{\mathbf{X}}^{n+1}\widehat{\mathbf{S}}^{n+1}\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix} &= \widehat{\mathbf{K}}\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{X}^{\mathrm{ap}}\mathbf{S}^{\mathrm{ap}} & \widehat{\mathbf{X}}^{\mathrm{rem}}\widehat{\mathbf{S}}^{\mathrm{rem}}\end{bmatrix}\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{X}^{\mathrm{ap}} & \widehat{\mathbf{X}}^{\mathrm{rem}}\end{bmatrix}\begin{bmatrix}\mathbf{S}^{\mathrm{ap}} & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{S}}^{\mathrm{rem}}\widehat{\mathbf{W}}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{X}^{\mathrm{ap}} & \widehat{\mathbf{X}}^{\mathrm{rem}}\end{bmatrix}\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{U}}\end{bmatrix}\begin{bmatrix}\mathbf{S}^{\mathrm{ap}} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}^{\mathrm{rem}}\end{bmatrix}\end{aligned}$$

where we use that $\widehat{\mathbf{S}}^{\mathrm{rem}}\widehat{\mathbf{W}} = \widehat{\mathbf{U}}\mathbf{S}^{\mathrm{rem}}$. We have from (4.18) that the updated spatial basis matrix has the form $\mathbf{X}^{n+1}\mathbf{R}_2 = \begin{bmatrix}\mathbf{X}^{\mathrm{ap}} & \mathbf{X}^{\mathrm{rem}}\end{bmatrix} = \begin{bmatrix}\mathbf{X}^{\mathrm{ap}} & \widehat{\mathbf{X}}^{\mathrm{rem}}\widehat{\mathbf{U}}\end{bmatrix}$ and thus

$$\widehat{\mathbf{X}}^{n+1}\widehat{\mathbf{S}}^{n+1}\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix} = \mathbf{X}^{n+1}\mathbf{R}_2\begin{bmatrix}\mathbf{S}^{\mathrm{ap}} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}^{\mathrm{rem}}\end{bmatrix}.$$

Hence we get the following relation

$$\begin{aligned}\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{W}}^\top\end{bmatrix}\widehat{\mathbf{S}}^{n+1,\top}\widehat{\mathbf{S}}^{n+1}\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix} &= \begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{W}}^\top\end{bmatrix}\widehat{\mathbf{S}}^{n+1,\top}\widehat{\mathbf{X}}^{n+1,\top}\widehat{\mathbf{X}}^{n+1}\widehat{\mathbf{S}}^{n+1}\begin{bmatrix}\mathbf{I}_m & \mathbf{0}\\ \mathbf{0} & \widehat{\mathbf{W}}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{S}^{\mathrm{ap},\top} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}^{\mathrm{rem},\top}\end{bmatrix}\mathbf{R}_2^\top\mathbf{X}^{n+1,\top}\mathbf{X}^{n+1}\mathbf{R}_2\begin{bmatrix}\mathbf{S}^{\mathrm{ap}} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}^{\mathrm{rem}}\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{S}^{\mathrm{ap},\top} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}^{\mathrm{rem},\top}\end{bmatrix}\mathbf{R}_2^\top\mathbf{R}_2\begin{bmatrix}\mathbf{S}^{\mathrm{ap}} & \mathbf{0}\\ \mathbf{0} & \mathbf{S}^{\mathrm{rem}}\end{bmatrix}.\end{aligned}$$

Putting it all together we have

$$
\begin{aligned}
\mathbf{R}_2^{-\top} \begin{bmatrix} (\mathbf{S}^{\mathrm{ap}})^{-\top} & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{U}}^{\top} (\widehat{\mathbf{S}}^{\mathrm{rem}})^{-\top} \end{bmatrix} & \widehat{\mathbf{S}}^{n+1,\top} \widehat{\mathbf{S}}^{n+1} \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{W}} \end{bmatrix} \\
&= \mathbf{R}_2^{-\top} \begin{bmatrix} (\mathbf{S}^{\mathrm{ap}})^{-\top} & \mathbf{0} \\ \mathbf{0} & (\mathbf{S}^{\mathrm{rem}})^{-\top} \widehat{\mathbf{W}}^{\top} \end{bmatrix} \widehat{\mathbf{S}}^{n+1,\top} \widehat{\mathbf{S}}^{n+1} \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{W}} \end{bmatrix} \\
&= \mathbf{R}_2^{-\top} \begin{bmatrix} \mathbf{S}^{\mathrm{ap}} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{\mathrm{rem}} \end{bmatrix}^{-\top} \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{W}}^{\top} \end{bmatrix} \widehat{\mathbf{S}}^{n+1,\top} \widehat{\mathbf{S}}^{n+1} \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ \mathbf{0} & \widehat{\mathbf{W}} \end{bmatrix} \\
&= \mathbf{R}_2^{-\top} \begin{bmatrix} \mathbf{S}^{\mathrm{ap}} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{\mathrm{rem}} \end{bmatrix}^{-\top} \begin{bmatrix} \mathbf{S}^{\mathrm{ap},\top} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{\mathrm{rem},\top} \end{bmatrix} \mathbf{R}_2^{\top} \mathbf{R}_2 \begin{bmatrix} \mathbf{S}^{\mathrm{ap}} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{\mathrm{rem}} \end{bmatrix} \\
&= \mathbf{R}_2 \begin{bmatrix} \mathbf{S}^{\mathrm{ap}} & \mathbf{0} \\ \mathbf{0} & \mathbf{S}^{\mathrm{rem}} \end{bmatrix} \\
&= \mathbf{S}^{n+1}.
\end{aligned}
$$

□