ENERGY STABLE AND CONSERVATIVE DYNAMICAL LOW-RANK APPROXIMATION FOR THE SU-OLSON PROBLEM*

LENA BAUMANN†, LUKAS EINKEMMER‡, CHRISTIAN KLINGENBERG†, AND JONAS KUSCH§

Abstract. Computational methods for thermal radiative transfer problems exhibit high computational costs and a prohibitive memory footprint when the spatial and directional domains are finely resolved. A strategy to reduce such computational costs is dynamical low-rank approximation (DLRA), which represents and evolves the solution on a low-rank manifold, thereby significantly decreasing computational and memory requirements. Efficient discretizations for the DLRA evolution equations need to be carefully constructed to guarantee stability while enabling mass conservation. In this work, we focus on the Su-Olson closure leading to a linearized internal energy model and derive a stable discretization through an implicit coupling of internal energy and particle density. Moreover, we propose a rank-adaptive strategy to preserve local mass conservation. Numerical results are presented which showcase the accuracy and efficiency of the proposed low-rank method compared to the solution of the full system.

Key words. thermal radiative transfer, Su-Olson closure, dynamical low-rank approximation, energy stability, mass conservation, rank adaptivity

MSC codes. 35L65, 35Q49, 65M12, 65M22

1. Introduction.

Numerically solving the radiative transfer equations is a challenging task, especially due to the high dimensionality of the solution's phase space. A common strategy to tackle this issue is to choose coarse numerical discretizations and mitigate numerical artifacts [23, 27, 32] which arise due to the insufficient resolution, see e.g. [3, 15, 1, 24, 39]. Despite the success of these approaches in a large number of applications, the requirement of picking user-determined and problem dependent tuning parameters can render them impracticable. Another approach to deal with the problem's high dimensionality is the use of model order reduction techniques. A reduced order method which is gaining a considerable amount of attention in the field of radiation transport is dynamical low-rank approximation (DLRA) [20] due to its ability to yield accurate solutions while not requiring an expensive offline training phase. DLRA's core idea is to approximate the solution on a low-rank manifold and evolve it accordingly. Past work in the area of radiative transfer has focused on asymptotic-preserving schemes [10, 9], mass conservation [34], stable discretizations [21], imposing boundary conditions [22, 18] and implicit time discretizations [35]. A discontinuous Galerkin discretization of the DLRA evolution equations for thermal radiative transfer has been proposed in [5].

A key building block of efficient, accurate and stable methods for DLRA is the construction of time integrators which are robust irrespective of small singular values

*Submitted to the editors July, 12, 2023.

Funding: The work of the first author was supported by the Würzburg Mathematics Center for Communication and Interaction (WMCCI) as well as the Stiftung der Deutschen Wirtschaft (Foundation of German Business). The work of the fourth author was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – 491976834.

†Department of Mathematics, University of Wuerzburg, Wuerzburg, DE-97074, Germany (lena.baumann@uni-wuerzburg.de, klingen@mathematik.uni-wuerzburg.de).

‡Numerical Analysis and Scientific Computing, University of Innsbruck, Innsbruck, A-6020, Austria (lukas.einkemmer@uibk.ac.at).

§Scientific Computing, Norwegian University of Life Sciences, Ås, NO-1432, Norway (jonas.kusch@nmbu.no).

in the solution [19]. Three integrators which move on the low-rank manifold while not being restricted by its curvature are the *projector-splitting* (PS) integrator [25], the *basis update & Galerkin* (BUG) integrator [8], and the *parallel* integrator [7]. Since the PS integrator evolves one of the required subflows backward in time, the BUG and parallel integrator are preferable for diffusive problems while facilitating the construction of stable numerical discretization for hyperbolic problems [21]. Moreover, the BUG integrator allows for a basis augmentation step [6] which can be used to construct conservative schemes for the Schrödinger equation [6] and the Vlasov–Poisson equations [14].

In this work we consider the thermal radiative transfer equations using the Su-Olson closure. This leads to a linearized internal energy model for which we propose an energy stable and mass conservative DLRA scheme. The main novelties of this paper are:

- *A stable numerical scheme for thermal radiative transfer*: We show that a naive IMEX scheme fails to guarantee energy stability. To overcome this unphysical behaviour we propose a scheme which advances radiation and internal energy implicitly in a coupled fashion. In addition, our novel analysis gives a classic hyperbolic CFL condition that enables us to operate up to a time step size of $\Delta t = \mathrm{CFL} \cdot \Delta x$.
- *A mass conservative and rank-adaptive integrator*: We employ the basis augmentation step from [6] as well as an adaption of the conservative truncation strategy from [14, 17] to guarantee local mass conservation and rank adaptivity. In contrast to [14, 17] we do not need to impose conservation through a modified L-step equation, but solely use the basis augmentation strategy from [6].

Both these properties are extremely important as they ensure key physical principles and allow us to choose an optimal time step size which reduces the computational effort. Moreover, we demonstrate numerical experiments which underline the derived stability and conservation properties of the proposed low-rank method while showing significantly reduced computational costs and memory requirements compared to the full-order system.

This paper is structured as follows: After the introduction in Section 1, we review the background on thermal radiative transfer and dynamical low-rank approximation in Section 2. In Section 3 we present the evolution equations for the thermal radiative transfer equations when using the rank-adaptive BUG integrator. Section 4 discretizes the resulting equations in angle and space. The main method is presented in Section 5 where a stable time discretization is proposed. We discuss local mass conservation of the scheme in Section 6. Numerical experiments are demonstrated in Section 7.

2. Background.

2.1. Thermal radiative transfer. In this work, we study radiation particles moving through and interacting with a background material. By absorbing particles, the material heats up and emits new particles which can in turn again interact with the background. This process is described by the thermal radiative transport equations

$$\begin{aligned}\frac{1}{c}\partial_t f(t,x,\mu) + \mu\partial_x f(t,x,\mu) &= \sigma(B(t,x) - f(t,x,\mu)),\\ \partial_t e(t,x) &= \sigma(\langle f(t,x,\cdot)\rangle_\mu - B(t,x)),\end{aligned}$$

where we omit boundary and initial conditions for now. This system can be solved for the particle density $f(t,x,\mu)$ and the internal energy $e(t,x)$ of the background medium. Here, $x \in D \subset \mathbb{R}$ is the spatial variable and $\mu \in [-1, 1]$ denotes the directional (or velocity) variable. The opacity σ encodes the rate at which particles are absorbed by the medium and we use brackets $\langle\cdot\rangle_\mu, \langle\cdot\rangle_x$ to indicate an integration over the directional domain and the spatial domain, respectively. Moreover, the speed of light is denoted by c and the black body radiation at the material temperature T is denoted by $B(T)$. It often is described by the Stefan-Boltzmann law

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

where $a = \frac{4\sigma_{\mathrm{SB}}}{c}$ is the radiation density constant and σ_{SB} the Stefan-Boltzmann constant. Different closures exist to determine a relation between the temperature T and the internal energy e. Following the ideas of Pomraning [37] and Su and Olson [38] we assume $e(T) = \alpha B(T)$. Without loss of generality we set $\alpha = 1$ and obtain

$$\partial_t f(t, x, \mu) + \mu \partial_x f(t, x, \mu) = \sigma(B(t, x) - f(t, x, \mu)), \tag{2.1a}$$

$$\partial_t B(t, x) = \sigma(\langle f(t, x, \cdot)\rangle_\mu - B(t, x)). \tag{2.1b}$$

We call this system the Su-Olson problem. It is a linear system for the particle density f and the internal energy B that is analytically solvable and and serves as a common benchmark for numerical considerations [33, 30, 31, 28]. Note that we leave out the speed of light by doing a rescaling of time $\tau = t/c$ and in an abuse of notation use t to denote τ in the remainder. Constructing numerical schemes to solve the above equation is challenging. First, the potentially stiff opacity term has to be treated by an implicit time integration scheme. Second, for three-dimensional spatial domains the computational costs and memory requirements of finely resolved spatial and angular discretizations become prohibitive. To tackle the high dimensionality, we choose a dynamical low-rank approximation which we introduce in the following.

2.2. Dynamical low-rank approximation. The core idea of DLRA is to approximate the solution of a given equation $\partial_t f(t, x, \mu) = F(f(t, x, \mu))$ by a representation of the form

$$f(t, x, \mu) \approx \sum_{i,j=1}^{r} X_i(t, x) S_{ij}(t) V_j(t, \mu), \tag{2.2}$$

where the orthonormal functions $\{X_i : i = 1, ..., r\}$ depend only on t and x and the orthonormal functions $\{V_j : j = 1, ..., r\}$ depend only on t and μ. The number of basis functions is set to r and we call r the rank of this approximation. This terminology stems from the matrix setting for which the concept of DLRA has been introduced [20]. Then, (2.2) can be interpreted as a continuous analogue to the singular value decomposition for matrices. As representation (2.2) is not unique we impose the Gauge conditions $\langle \dot{X}_i, X_j\rangle_x = 0$ and $\langle \dot{V}_i, V_j\rangle_\mu = 0$ from which we can conclude that $\{X_i\}$ and $\{V_j\}$ are uniquely determined for invertible $\mathbf{S} = (S_{ij}) \in \mathbb{R}^{r\times r}$ [20, 10, 13]. That is, we seek for an approximation of f that for each time t lies in the manifold

$$\mathcal{M}_r = \Bigg\{ f \in L^2(D \times [-1, 1]) : f(\cdot, x, \mu) = \sum_{i,j=1}^{r} X_i(\cdot, x) S_{ij}(\cdot) V_j(\cdot, \mu) \text{ with invertible}$$
$$\mathbf{S} = (S_{ij}) \in \mathbb{R}^{r\times r}, X_i \in L^2(D), V_j \in L^2([-1, 1]) \text{ and } \langle X_i, X_j\rangle_x = \delta_{ij},$$
$$\langle V_i, V_j\rangle_\mu = \delta_{ij} \Bigg\}.$$

Note that in the following we denote the full rank and the low-rank solutions as f. Let $f(t,\cdot,\cdot)$ be a path on $\mathcal{M}_r$. A formal differentiation of f with respect to t leads to

$$\dot{f}(t,\cdot,\cdot) = \sum_{i,j=1}^{r} \left(\dot{X}_i(t,\cdot)S_{ij}(t)V_j(t,\cdot) + X_i(t,\cdot)\dot{S}_{ij}(t)V_j(t,\cdot) + X_i(t,\cdot)S_{ij}(t)\dot{V}_j(t,\cdot) \right).$$

These functions restrict the solution dynamics onto the low-rank manifold $\mathcal{M}_r$ and constitute the corresponding tangent space which under the Gauge conditions reads

$$\begin{aligned}\mathcal{T}_f\mathcal{M}_r = \Bigg\{ \dot{f} \in L^2(D\times[-1,1]) : \dot{f}(\cdot,x,\mu) = \sum_{i,j=1}^{r} \Big(&\dot{X}_i(\cdot,x)S_{ij}(\cdot)V_j(\cdot,\mu) \\ &+ X_i(\cdot,x)\dot{S}_{ij}(\cdot)V_j(\cdot,\mu) + X_i(\cdot,x)S_{ij}(\cdot)\dot{V}_j(\cdot,\mu)\Big) \\ &\text{with } \dot{S}_{ij}\in\mathbb{R}, \dot{X}_i \in L^2(D), \dot{V}_j \in L^2([-1,1]) \text{ and } \langle \dot{X}_i, X_j\rangle_x = 0, \\ &\langle \dot{V}_i, V_j\rangle_\mu = 0 \Bigg\}.\end{aligned}$$

Having defined the low-rank manifold and its corresponding tangent space, we now wish to determine $f(t,\cdot,\cdot) \in \mathcal{M}_r$ such that $\partial_t f(t,\cdot,\cdot) \in \mathcal{T}_f\mathcal{M}_r$ and $\|\partial_t f(t,\cdot,\cdot) - F(f(t,\cdot,\cdot))\|_{L^2(D\times[-1,1])}$ is minimized. That is, one wishes to determine f such that

$$\langle \partial_t f(t,\cdot,\cdot) - F(f(t,\cdot,\cdot)), \dot{f}\,\rangle_{x,\mu} = 0 \quad \text{for all } \dot{f} \in \mathcal{T}_f\mathcal{M}_r. \tag{2.3}$$

The orthogonal projector onto the tangent plane $\mathcal{T}_f\mathcal{M}_r$ can be explicitly given as

$$P(f)F(f) = \sum_{j=1}^{r}\langle V_j, F(f)\rangle_\mu V_j - \sum_{i,j=1}^{r} X_i\langle X_iV_j, F(f)\rangle_{x,\mu}V_j + \sum_{i=1}^{r} X_i\langle X_i, F(f)\rangle_x.$$

With this definition at hand, we can reformulate (2.3) as

$$\partial_t f(t,x,\mu) = P(f(t,x,\mu))F(f(t,x,\mu)).$$

To evolve the approximation of the solution in time according to the above equation is not trivial. Indeed standard time integration schemes suffer from the curvature of the low-rank manifold, which is proportional to the smallest singular value of the low-rank solution [20]. Three integrators which move along the manifold without suffering from its high curvature exist: The projector–splitting integrator [25], the BUG integrator [8], and the parallel integrator [7]. In this work, we will use the basis-augmented extension to the BUG integrator [6] which we explain in the following.

The rank-adaptive BUG integrator [6] updates and augments the bases $\{X_i\}, \{V_j\}$ in parallel in the first two steps. In the third step, a Galerkin step is performed for the augmented bases followed by a truncation step to a new rank r_1. In detail, to evolve the approximation of the distribution function from $f(t_0,x,\mu) = \sum_{i,j=1}^{r} X_i^0(x)S_{ij}^0V_j^0(\mu)$ at time t_0 to $f(t_1,x,\mu) = \sum_{i,j=1}^{r_1} X_i^1(x)S_{ij}^1V_j^1(\mu)$ at time $t_1 = t_0 + \Delta t$ the integrator performs the following steps:

***K*-Step**: Write $K_j(t,x) = \sum_{i=1}^{r} X_i(t,x)S_{ij}(t)$. Then we obtain the representation $f(t,x,\mu) = \sum_{j=1}^{r} K_j(t,x)V_j^0(\mu)$ with $\{V_j^0\}$ kept fixed in this step. The basis functions $X_i^0(x)$ with $i = 1,...,r$ are updated by solving the partial differential equation

$$\partial_t K_j(t,x) = \left\langle V_j^0, F\left(\sum_{k=1}^{r} K_k(t,x)V_k^0\right)\right\rangle_\mu, \quad K_j(t_0,x) = \sum_{i=1}^{r} X_i^0(x)S_{ij}^0,$$

and applying Gram Schmidt to $[K_j(t_1,x), X_i^0] = \sum_{i=1}^{2r} \widehat{X}_i^1(x) R_{ij}^1$. Then, the updated and augmented basis in physical space consists of $\widehat{X}_i^1(x)$ with $i = 1, ..., 2r$. Note that R_{ij}^1 is discarded after this step. Compute $\widehat{M}_{ki} = \langle \widehat{X}_k^1, X_i^0 \rangle_x$.

***L*-Step**: Write $L_i(t,\mu) = \sum_{j=1}^r S_{ij}(t) V_j(t,\mu)$. Then we obtain the representation $f(t,x,\mu) = \sum_{i=1}^r X_i^0 L_i(t,\mu)$ with $\{X_i^0\}$ kept fixed in this step. The basis functions $V_j^0(\mu)$ with $j = 1, ..., r$ are updated by solving the partial differential equation

$$\partial_t L_i(t,\mu) = \left\langle X_i^0, F\left(\sum_{\ell=1}^r X_\ell^0 L_\ell(t,\mu)\right)\right\rangle_x, \quad L_i(t_0,\mu) = \sum_{j=1}^r S_{ij}^0 V_j^0(\mu),$$

and applying Gram Schmidt to $[L_i(t_1,\mu), V_j^0(\mu)] = \sum_{j=1}^{2r} \widehat{V}_j^1(\mu) R_{ij}^2$. Then, the updated and augmented basis in velocity space consists of $\widehat{V}_j^1(\mu)$ with $j = 1, ..., 2r$. Note that R_{ij}^2 is discarded after this step. Compute $\widehat{N}_{\ell j} = \langle \widehat{V}_\ell^1, V_j^0 \rangle_\mu$.

***S*-step**: Update S_{ij}^0 with $i,j = 1, ..., r$ to $\widehat{S}_{ij}^1$ with $i,j = 1, ..., 2r$ by solving the ordinary differential equation

$$\dot{\widehat{S}}_{ij}(t) = \left\langle \widehat{X}_i^1 \widehat{V}_j^1, F\left(\sum_{\ell,k=1}^{2r} \widehat{X}_\ell^1 \widehat{S}_{\ell k}(t) \widehat{V}_k^1\right)\right\rangle_{x,\mu}, \quad \widehat{S}_{ij}(t_0) = \sum_{k,\ell=1}^r \widehat{M}_{ik} S_{k\ell}^0 \widehat{N}_{j\ell}.$$

Truncation: Let $\widehat{S}_{ij}^1$ be the entries of the matrix $\widehat{\mathbf{S}}^1$. Compute the singular value decomposition of $\widehat{\mathbf{S}}^1 = \widehat{\mathbf{P}}\widehat{\mathbf{\Sigma}}\widehat{\mathbf{Q}}^\top$ with $\mathbf{\Sigma} = \mathrm{diag}(\sigma_j)$. Given a tolerance ϑ, choose the new rank $r_1 \leq 2r$ as the minimal number such that

$$\left(\sum_{j=r_1+1}^{2r} \sigma_j^2\right)^{1/2} \leq \vartheta.$$

Let $\mathbf{S}^1$ with entries S_{ij}^1 be the $r_1 \times r_1$ diagonal matrix with the r_1 largest singular values and let $\mathbf{P}^1$ with entries P_{ij}^1 and $\mathbf{Q}^1$ with entries Q_{ji}^1 contain the first r_1 columns of $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{Q}}$, respectively. Set $X_i^1(x) = \sum_{i=1}^{2r} \widehat{X}_i^1(x) P_{ij}^1$ for $i = 1, ..., r_1$ and $V_j^1(\mu) = \sum_{j=1}^{2r} \widehat{V}_j^1(\mu) Q_{ji}^1$ for $j = 1, ..., r_1$.

The updated approximation of the solution after one time step is then given by $f(t_1,x,\mu) = \sum_{i,j=1}^{r_1} X_i^1(x) S_{ij}^1 V_j^1(\mu)$. Note that we are not limited to augmenting with the old basis, which we will use to construct our scheme.

3. Dynamical low-rank approximation for Su-Olson. Let us now derive the evolution equations of the rank-adaptive BUG integrator for system (2.1), i.e. the partial differential equations appearing in the K- and L-step and the ordinary differential equation for the S-step. To simplify notation, all derivations are performed for one spatial and one directional variable. However, the derivation trivially extends to higher dimensions. We start with considering the evolution equations for the low-rank approximation of the particle density (2.1a).

***K*-step:** Write $K_j(t,x) = \sum_{i=1}^r X_i(t,x) S_{ij}(t)$. Then we have the representation $f(t,x,\mu) = \sum_{j=1}^r K_j(t,x) V_j^0(\mu)$ for the low-rank approximation of the solution. Again $\{V_j^0\}$ denotes the set of orthonormal basis functions for the velocity space that shall be kept fixed in this step. Inserting this representation of f into (2.1a) and projecting onto $V_k^0(\mu)$ gives the partial differential equation

$$\partial_t K_k(t,x) = -\sum_{j=1}^{r} \partial_x K_j(t,x)\langle V_k^0, \mu V_j^0\rangle_\mu + \sigma\left(B(t,x)\langle V_k^0\rangle_\mu - K_k(t,x)\right). \tag{3.1}$$

***L*-step:** Write $L_i(t,\mu) = \sum_{j=1}^r S_{ij}(t)V_j(t,\mu)$. Then we have the representation $f(t,x,\mu) = \sum_{i=1}^r X_i^0(x)L_i(t,\mu)$ for the low-rank approximation of the solution. Again $\{X_i^0\}$ denotes the set of spatial orthonormal basis functions that shall be kept fixed in this step. Inserting this representation of f into (2.1a) and projecting onto $X_k^0(x)$ yields the partial differential equation

$$\partial_t L_k(t,\mu) = -\mu \sum_{i=1}^{r} \left\langle X_k^0, \frac{\mathrm{d}}{\mathrm{d}x} X_i^0 \right\rangle_x L_i(t,\mu) + \sigma\left(\langle X_k^0, B(t,\cdot)\rangle_x - L_k(t,\mu)\right). \tag{3.2}$$

Lastly, we derive the augmented Galerkin step of the rank-adaptive BUG integrator. We denote the time updated spatial basis augmented with X_i^0 as $\widehat{X}_i^1$. The augmented directional basis $\widehat{V}_i^1$ is constructed in the corresponding way. Then, the augmented Galerkin step is constructed according to:

***S*-step:** We use the initial condition $\widehat{S}_{ij}(t_0) = \sum_{\ell,k=1}^r \langle \widehat{X}_i^1 X_\ell^0\rangle_x S_{\ell k}(t_0)\langle \widehat{V}_j^1 V_k^0\rangle_\mu$ and approximate the solution f as $f(t,x,\mu) = \sum_{i,j=1}^{2r} \widehat{X}_i^1(x)\widehat{S}_{ij}(t)\widehat{V}_j^1(\mu)$. Inserting this representation into (2.1a) and testing against $\widehat{X}_k^1$ and $\widehat{V}_\ell^1$ gives the ordinary differential equation

$$\dot{\widehat{S}}_{k\ell}(t) = -\sum_{i,j=1}^{2r} \left\langle \widehat{X}_k^1, \frac{\mathrm{d}}{\mathrm{d}x}\widehat{X}_i^1 \right\rangle_x \widehat{S}_{ij}(t)\langle \widehat{V}_\ell^1, \mu \widehat{V}_j^1\rangle_\mu + \sigma\left(\langle \widehat{X}_k^1, B(t,\cdot)\rangle_x \langle \widehat{V}_\ell^1\rangle_\mu - \widehat{S}_{k\ell}(t)\right) \tag{3.3}$$

from which we get the augmented quantity $\widehat{S}_{ij}(t)$. Inserting all augmented low-rank factors into (2.1b) leads to the partial differential equation

$$\partial_t B(t,x) = \sigma\left(\sum_{i,j=1}^{2r} \widehat{X}_i^1(x)\widehat{S}_{ij}(t)\langle \widehat{V}_j^1\rangle_\mu - B(t,x)\right). \tag{3.4}$$

Before repeating this process and evolving the subequations further in time we truncate back the augmented quantities to a new rank r_1 using a suitable truncation strategy.

4. Angular and spatial discretization. Having derived the K-, L- and S-step of the rank-adaptive BUG integrator, we can now proceed with discretizing in angle and space. For the angular discretization, we use the modal representations

$$V_j^0(\mu) \simeq \sum_{n=0}^{N-1} V_{nj}^0 P_n(\mu), \quad \widehat{V}_j^1(\mu) \simeq \sum_{n=0}^{N-1} \widehat{V}_{nj}^1 P_n(\mu), \quad L_i(t,\mu) \simeq \sum_{n=0}^{N-1} L_{ni}(t)P_n(\mu),$$

where P_n are the normalized Legendre polynomials. Note that in the following, we use Einstein's sum convention when not stated otherwise to ensure compactness of notation. Let us define the matrix $\mathbf{A} \in \mathbb{R}^{N\times N}$ with entries $A_{mn} := \langle P_m, \mu P_n\rangle_\mu$. Then we can rewrite $\langle V_k^0, \mu V_j^0 \rangle_\mu = V_{km}^0 A_{mn} V_{jn}^0$. The evolution equations with angular discretization then read

(4.1a)
$$\partial_t K_k(t,x) = -\partial_x K_j(t,x) V_{nj}^0 A_{mn} V_{mk}^0 + \sigma \left(B(t,x) V_{0k}^0 - K_k(t,x) \right),$$
(4.1b)
$$\dot{L}_{mk}(t) = -\left\langle X_k^0, \frac{\mathrm{d}}{\mathrm{d}x} X_i^0 \right\rangle_x L_{ni}(t) A_{mn} + \sigma \left(\langle X_k^0, B(t,\cdot)\rangle_x \delta_{m0} - L_{mk}(t) \right),$$
(4.1c)
$$\dot{\widehat{S}}_{k\ell}(t) = -\left\langle \widehat{X}_k^1, \frac{\mathrm{d}}{\mathrm{d}x} \widehat{X}_i^1 \right\rangle_x S_{ij}(t) \widehat{V}_{nj}^1 A_{mn} \widehat{V}_{m\ell}^1 + \sigma \left(\langle \widehat{X}_k^1, B(t,\cdot)\rangle_x \widehat{V}_{0\ell}^1 - \widehat{S}_{k\ell}(t) \right).$$

For the angular discretization of (3.4) we get

$$\partial_t B(t,x) = \sigma \left(\widehat{X}_i^1(x) \widehat{S}_{ij}(t) \widehat{V}_{0j}^1 - B(t,x) \right). \tag{4.1d}$$

To derive a spatial discretization we choose a spatial grid $x_1 < \cdots < x_{n_x}$ with equidistant spacing Δx. The solution in a given cell p is then approximated by

$$X_{pk}(t) \approx \frac{1}{\Delta x} \int_{x_p}^{x_{p+1}} X_k(t,x)\,\mathrm{d}x, \quad K_{pk}(t) \approx \frac{1}{\Delta x} \int_{x_p}^{x_{p+1}} K_k(t,x)\,\mathrm{d}x,$$
$$B_p(t) \approx \frac{1}{\Delta x} \int_{x_p}^{x_{p+1}} B(t,x)\,\mathrm{d}x\,.$$

Spatial derivatives are approximated and stabilized through the tridiagonal stencil matrices $\mathbf{D}^x \approx \partial_x$ and $\mathbf{D}^{xx} \approx \frac{1}{2}\Delta x \partial_{xx}$ with entries

$$D_{p,p\pm 1}^x = \frac{\pm 1}{2\Delta x}\,, \qquad D_{p,p}^{xx} = -\frac{1}{\Delta x}\,, \quad D_{p,p\pm 1}^{xx} = \frac{1}{2\Delta x}\,.$$

Applying the matrix $\mathbf{D}^x \in \mathbb{R}^{n_x \times n_x}$ corresponds to a first order and the stabilization matrix $\mathbf{D}^{xx} \in \mathbb{R}^{n_x \times n_x}$ to a second order central differencing scheme. Moreover, from now on we assume periodic boundary conditions. Recall the symmetric matrix $\mathbf{A}$. It is diagonalizable in the form $\mathbf{A} = \mathbf{Q}\mathbf{M}\mathbf{Q}^\top$ with $\mathbf{Q}$ orthogonal and $\mathbf{M} = \mathrm{diag}(\sigma_1, ..., \sigma_n)$. We define matrix $|\mathbf{A}|$ as $|\mathbf{A}| = \mathbf{Q}|\mathbf{M}|\mathbf{Q}^\top$. We then obtain the spatially and angular discretized matrix ODEs

$$\begin{aligned} \dot{K}_{pk}(t) = &- D_{qp}^x K_{pj}(t) V_{nj}^0 A_{mn} V_{mk}^0 + D_{qp}^{xx} K_{pj}(t) V_{nj}^0 |A|_{mn} V_{mk}^0 \\ &+ \sigma \left(B_p(t) V_{0k}^0 - K_{pk}(t) \right), \end{aligned} \tag{4.2a}$$
$$\begin{aligned} \dot{L}_{mk}(t) = &- A_{mn} L_{ni}(t) X_{pi}^0 D_{qp}^x X_{qk}^0 + |A|_{mn} L_{ni}(t) X_{pi}^0 D_{qp}^{xx} X_{qk}^0 \\ &+ \sigma \left(\delta_{m0} B_p(t) X_{pk}^0 - L_{mk}(t) \right), \end{aligned} \tag{4.2b}$$
$$\begin{aligned} \dot{\widehat{S}}_{k\ell}(t) = &- \widehat{X}_{pk}^1 D_{pq}^x \widehat{X}_{qi}^1 \widehat{S}_{ij}(t) \widehat{V}_{nj}^1 A_{mn} \widehat{V}_{m\ell}^1 + \widehat{X}_{pk}^1 D_{pq}^{xx} \widehat{X}_{qi}^1 \widehat{S}_{ij}(t) \widehat{V}_{nj}^1 |A|_{mn} \widehat{V}_{m\ell}^1 \\ &+ \sigma \left(\widehat{X}_{pk}^1 B_p(t) \widehat{V}_{0\ell}^1 - \widehat{S}_{k\ell}(t) \right). \end{aligned} \tag{4.2c}$$

Lastly, we obtain from (4.1d) for the internal energy B the spatially discretized equation

$$\dot{B}_p(t) = \sigma \left(\widehat{X}_{pi}^1 \widehat{S}_{ij}(t) \widehat{V}_{0j}^1 - B_p(t) \right) = \sigma \left(u_{p0}^1(t) - B_p(t) \right), \tag{4.2d}$$

where we use the notation $\widehat{X}^1_{pi}\widehat{S}_{ij}(t)\widehat{V}^1_{mj} =: u^1_{pm}(t)$. We can now show that the semi-discrete time-dependent system (4.2) is energy stable. For this, let us first give a definition of the total energy of the system:

DEFINITION 4.1 (Total energy). *Let the matrix $\mathbf{u}^1(t) \in \mathbb{R}^{n_x \times N}$ with low-rank entries $u^1_{pm}(t) = \widehat{X}^1_{pi}\widehat{S}_{ij}(t)\widehat{V}^1_{mj}$ denote the angularly and spatially discretized approximation of the solution of (2.1a) and $\mathbf{B}(t) \in \mathbb{R}^{n_x}$ be the spatially discretized approximation of the solution of (2.1b). Then we call*

$$E(t) := \frac{1}{2}\|\mathbf{u}^1(t)\|_F^2 + \frac{1}{2}\|\mathbf{B}(t)\|_E^2,$$

with $\|\cdot\|_F$ denoting the Frobenius and $\|\cdot\|_E$ denoting the Euclidean norm, the total energy of the system (4.2).

Further, we note the following properties of the chosen spatial stencil matrices which we write down denoting all sums explicitly:

LEMMA 4.2 (Summation by parts). *Let $y, z \in \mathbb{R}^{n_x}$ with indices $p, q = 1, ..., n_x$. In addition, we set $y_0 = y_{n_x}$ and $y_{n+1} = y_1$, for z respectively, due to the periodic boundary conditions. Then the stencil matrices fulfill the following properties:*

$$\sum_{p,q=1}^{n_x} y_p D^x_{pq} z_q = -\sum_{p,q=1}^{n_x} z_p D^x_{pq} y_q, \quad \sum_{p,q=1}^{n_x} z_p D^x_{pq} z_q = 0, \quad \sum_{p,q=1}^{n_x} y_p D^{xx}_{pq} z_q = \sum_{p,q=1}^{n_x} z_p D^{xx}_{pq} y_q.$$

Moreover, let $\mathbf{D}^+ \in \mathbb{R}^{n_x \times n_x}$ be defined as

$$D^+_{p,p} = \frac{-1}{\sqrt{2\Delta x}}\,, \qquad D^+_{p,p+1} = \frac{1}{\sqrt{2\Delta x}}\,.$$

Then, $\sum_{p,q=1}^{n_x} z_p D^{xx}_{pq} z_q = -\sum_{p=1}^{n_x}\left(\sum_{q=1}^{n_x} D^+_{pq} z_q\right)^2$.

Proof. The assertions follow directly by plugging in the definitions of the stencil matrices and rearranging the sums of the products in an adequate way:

$$\begin{aligned}
\sum_{p,q=1}^{n_x} y_p D^x_{pq} z_q &= \frac{1}{2\Delta x}\sum_{p=1}^{n_x} y_p\left(z_{p+1} - z_{p-1}\right) = -\frac{1}{2\Delta x}\sum_{p=1}^{n_x} z_p\left(y_{p+1} - y_{p-1}\right) \\
&= -\sum_{p,q=1}^{n_x} z_p D^x_{pq} y_q, \\
\sum_{p,q=1}^{n_x} z_p D^x_{pq} z_q &= -\sum_{p,q=1}^{n_x} z_p D^x_{pq} z_q = 0,
\end{aligned}$$

$$
\begin{aligned}
\sum_{p,q=1}^{n_x} y_p D^{xx}_{pq} z_q =& -\frac{1}{\Delta x}\sum_{p=1}^{n_x} y_p z_p + \frac{1}{2\Delta x}\sum_{p=1}^{n_x} y_p (z_{p+1}+z_{p-1}) \\
=& -\frac{1}{\Delta x}\sum_{p=1}^{n_x} z_p y_p + \frac{1}{2\Delta x}\sum_{p=1}^{n_x} z_p (y_{p+1}+y_{p-1}) = \sum_{p,q=1}^{n_x} z_p D^{xx}_{pq} y_q, \\
\sum_{p,q=1}^{n_x} z_p D^{xx}_{pq} z_q =& -\frac{1}{\Delta x}\sum_{p=1}^{n_x} z_p^2 + \frac{1}{2\Delta x}\sum_{p=1}^{n_x} z_p (z_{p+1}+z_{p-1}) \\
=& -\frac{1}{2\Delta x}\sum_{p=1}^{n_x}\left(z_p^2 - 2z_p z_{p+1} + z_{p+1}^2\right) = -\frac{1}{2\Delta x}\sum_{p=1}^{n_x}(z_p - z_{p+1})^2 \\
=& -\sum_{p=1}^{n_x}\left(\sum_{q=1}^{n_x} D^{+}_{pq} z_q\right)^2 .
\end{aligned}
$$

∎

With these properties at hand, we can now show dissipation of the total energy:

THEOREM 4.3. *The semi-discrete time-continuous system consisting of (4.2) is energy stable, that is $\dot{E}(t) \leq 0$.*

Proof. Let us start from the S-step in (4.2c)

$$
\begin{aligned}
\dot{\widehat{S}}_{k\ell}(t) =& -\widehat{X}^1_{pk} D^x_{pq}\widehat{X}^1_{qi}\widehat{S}_{ij}(t)\widehat{V}^1_{nj}A_{mn}\widehat{V}^1_{m\ell} + \widehat{X}^1_{pk} D^{xx}_{pq}\widehat{X}^1_{qi}\widehat{S}_{ij}(t)\widehat{V}^1_{nj}|A|_{mn}\widehat{V}^1_{m\ell} \\
&+ \sigma\left(\widehat{X}^1_{pk}(x)B_p(t)\widehat{V}^1_{0\ell} - \widehat{S}_{k\ell}(t)\right).
\end{aligned}
$$

We multiply with $\widehat{X}^1_{\alpha k}\widehat{V}^1_{\beta\ell}$, where $\alpha = 1, ..., n_x$ and $\beta = 0, ..., N-1$, sum over k and ℓ and introduce the projections $P^{X,1}_{\alpha p} = \widehat{X}^1_{\alpha k}\widehat{X}^1_{pk}$ and $P^{V,1}_{m\beta} = \widehat{V}^1_{m\ell}\widehat{V}^1_{\beta\ell}$. With the notation $\widehat{X}^1_{qi}\widehat{S}_{ij}(t)\widehat{V}^1_{nj} = u^1_{qn}(t)$ we get

$$
\begin{aligned}
\dot{u}^1_{\alpha\beta}(t) =& -P^{X,1}_{\alpha p} D^x_{pq} u^1_{qn}(t) A_{mn} P^{V,1}_{m\beta} + P^{X,1}_{\alpha p} D^{xx}_{pq} u^1_{qn}(t)|A|_{mn} P^{V,1}_{m\beta} \\
&+ \sigma\left(P^{X,1}_{\alpha p} B_p(t)\delta_{0m} P^{V,1}_{m\beta} - u^1_{\alpha\beta}(t)\right).
\end{aligned}
$$

Next, we multiply with $u^1_{\alpha\beta}(t)$ and sum over α and β. Note that

$$
P^{X,1}_{\alpha p} u^1_{\alpha\beta}(t) = u^1_{p\beta}(t) \quad \text{and} \quad P^{V,1}_{m\beta} u^1_{p\beta}(t) = u^1_{pm}(t).
$$

This leads to

$$
\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{u}^1(t)\|_F^2 =& -u^1_{pm}(t) D^x_{pq} u^1_{qn}(t) A_{mn} + u^1_{pm}(t) D^{xx}_{pq} u^1_{qn}(t)|A|_{mn} \\
&+ \sigma\left(u^1_{pm}(t) B_p(t)\delta_{0m} - \|\mathbf{u}^1(t)\|^2\right).
\end{aligned}
$$

Recall that we can write $\mathbf{A} = \mathbf{Q}\mathbf{M}\mathbf{Q}^\top$ with $\mathbf{M} = \mathrm{diag}(\sigma_1, ..., \sigma_N)$. Inserting this representation gives

$$
\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{u}^1(t)\|_F^2 =& - u_{pm}^1(t)D_{pq}^x u_{qn}^1(t)Q_{nk}\sigma_k Q_{mk} + u_{pm}^1(t)D_{pq}^{xx}u_{qn}^1(t)Q_{nk}|\sigma_k|Q_{mk} \\
&+ \left(u_{pm}^1(t)B_p(t)\delta_{0m} - \|\mathbf{u}^1(t)\|^2\right) \\
=& - \sigma_k \widetilde{u}_{pk}^1(t)D_{pq}^x \widetilde{u}_{qk}^1(t) + |\sigma_k|\widetilde{u}_{pk}^1(t)D_{pq}^{xx}\widetilde{u}_{qk}^1(t) \\
&+ \left(u_{pm}^1(t)B_p(t)\delta_{0m} - \|\mathbf{u}^1(t)\|^2\right),
\end{aligned}
$$

where $\widetilde{u}_{pk}^1(t) = u_{pm}^1(t)Q_{mk}$. With the properties of the stencil matrices we get

$$
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{u}^1(t)\|_F^2 = -\left(D_{pq}^+ u_{qm}^1(t)|A|_{mn}^{1/2}\right)^2 + \sigma\left(u_{p0}(t)B_p(t) - \|\mathbf{u}^1(t)\|_F^2\right). \tag{4.3}
$$

Next we consider equation (4.2d). Multiplication with $B_p(t)$ and summation over p gives

$$
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|\mathbf{B}(t)\|_E^2 = \sigma\left(u_{p0}(t)B_p(t) - \|\mathbf{B}(t)\|_E^2\right). \tag{4.4}
$$

For the total energy of the system it holds that $E(t) = \frac{1}{2}\|\mathbf{u}^1(t)\|_F^2 + \frac{1}{2}\|\mathbf{B}(t)\|_E^2$. Adding the evolution equations (4.3) and (4.4) we get

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}E(t) =& -\left(\widetilde{D}_{pq}^+ u_{qm}^1(t)|A|_{mn}^{1/2}\right)^2 + \sigma\left(u_{p0}^1(t)B_p(t) - \|\mathbf{u}^1(t)\|_F^2\right) \\
&+ \sigma\left(u_{p0}^1(t)B_p(t) - \|\mathbf{B}(t)\|_E^2\right) \\
=& -\left(\widetilde{D}_{pq}^+ u_{qm}^1(t)|A|_{mn}^{1/2}\right)^2 - \sigma\left((u_{p0}^1(t) - B_p(t))^2 + (u_{pm}^1(t))^2(1-\delta_{m0})\right),
\end{aligned}
$$

where we rewrote $\|\mathbf{B}(t)\|_E^2 = B_p(t)^2$ and $\|\mathbf{u}^1(t)\|_F^2 = (u_{pm}^1(t))^2$. This expression is strictly negative which means that E is dissipated in time. Hence, the system is energy stable. ∎

5. Time discretization.

Our goal is to construct a conservative DLRA scheme which is energy stable under a sharp time step restriction. Constructing time discretization schemes which preserve the energy dissipation shown in Theorem 4.3 while not suffering from the potentially stiff opacity term is not trivial. In fact a naive IMEX time discretization potentially will increase the total energy, which we demonstrate in the following.

5.1. Naive time discretization.

We start from system (4.2) which still depends continuously on the time t. For the time discretization we choose a naive IMEX Euler scheme where we perform a splitting of nternal energy and radiation transport equation. That is, we use an explicit Euler step for the transport part of the evolution equations, treat the internal energy B explicitly and use an implicit Euler step for the radiation absorption term. Note that the scheme describes the evolution from time t_0 to time $t_1 = t_0 + \Delta t$ but holds for all further time steps equivalently. This yields the fully discrete scheme

$$
\begin{aligned}
K_{pk}^1 =& K_{pk}^0 - \Delta t D_{qp}^x K_{pj}^0 V_{nj}^0 A_{mn} V_{mk}^0 + \Delta t D_{qp}^{xx} K_{pj}^0 V_{nj}^0 |A|_{mn} V_{mk}^0 \\
&+ \sigma\left(\Delta t B_p^0 V_{0k}^0 - \Delta t K_{pk}^1\right),
\end{aligned} \tag{5.1a}
$$

$$
\begin{aligned}
L_{mk}^1 =& L_{mk}^0 - \Delta t X_{qk}^0 D_{qp}^x X_{pi}^0 L_{ni}^0 A_{mn} + \Delta t X_{qk}^0 D_{qp}^{xx} X_{pi}^0 L_{ni}^0 |A|_{mn} \\
&+ \sigma\left(\Delta t X_{pk}^0 B_p^0 \delta_{m0} - \Delta t L_{mk}^1\right).
\end{aligned} \tag{5.1b}
$$

We perform a QR-decomposition of the quantities $[K^1_{pk}, X^0_{pk}]$ and $[L^1_{pk}, V^0_{pk}]$ to obtain the augmented and time updated bases $\widehat{X}^1_{pk}$ and $\widehat{V}^1_{pk}$ according to the rank-adaptive BUG integrator [6]. Lastly, we perform a Galerkin step for the augmented bases according to

$$\begin{aligned}\widehat{S}^1_{k\ell} = \widetilde{S}^0_{k\ell} - \Delta t \widehat{X}^1_{pk} D^x_{pq} \widehat{X}^1_{qi} \widetilde{S}^0_{ij} \widehat{V}^1_{nj} A_{mn} \widehat{V}^1_{m\ell} + \Delta t \widehat{X}^1_{pk} D^{xx}_{pq} \widehat{X}^1_{qi} \widetilde{S}^0_{ij} \widehat{V}^1_{nj} |A|_{mn} \widehat{V}^1_{m\ell} \\ + \sigma\left(\Delta t \widehat{X}^1_{pk} B^0_p \widehat{V}^1_{0\ell} - \Delta t \widehat{S}^1_{k\ell}\right),\end{aligned} \tag{5.1c}$$

where $\widetilde{S}^0_{k\ell} := \widehat{X}^1_{pk} X^0_{pi} S^0_{ij} V^0_{nj} \widehat{V}^1_{n\ell}$. The internal energy is then updated via

$$B^1_p = B^0_p + \sigma\Delta t\left(\widehat{X}^1_{pi} \widehat{S}^1_{ij} \widehat{V}^1_{0j} - B^1_p\right). \tag{5.1d}$$

However, this numerical method has the undesirable property that it can increase the total energy during a time step. In Theorem 5.1 we show this analytically. This behavior is, obviously, completely unphysical.

THEOREM 5.1. *Let $\mathbf{u}^0 \in \mathbb{R}^{n_x \times N}$ with entries $u^0_{pm} = X^0_{pk} S^0_{k\ell} V^0_{m\ell}$ denote the angularly and spatially discretized low-rank approximation of the function f at time $t = t_0$, and $\mathbf{u}^1 \in \mathbb{R}^{n_x \times N}$ with entries $u^1_{\alpha\beta} = \widehat{X}^1_{\alpha k} \widehat{S}^1_{k\ell} \widehat{V}^1_{\beta\ell}$ denote the basis augmented angularly and spatially discretized low-rank approximation at time $t = t_1$ using the rank-adaptive BUG integrator. Further, $\mathbf{B}^0 \in \mathbb{R}^{n_x}$ shall denote the spatially discretized low-rank approximation of B at time $t = t_0$, and $\mathbf{B}^1 \in \mathbb{R}^{n_x}$ at time $t = t_1$, respectively. The total energy at time $t = t_0$ is denoted by E^0 and E^1 at time $t = t_1$, respectively. Then, there exist initial value pairs $(\mathbf{u}^0, \mathbf{B}^0)$ and time step sizes Δt such that the naive scheme (5.1) results in $(\mathbf{u}^1, \mathbf{B}^1)$ for which the total energy increases, i.e. for which $E^1 > E^0$.*

Proof. Let us multiply the S-step (5.1c) with $\widehat{X}^1_{\alpha k} \widehat{V}^1_{\beta\ell}$ and sum over k and ℓ. Again we make use of the projections $P^{X,1}_{\alpha p} = \widehat{X}^1_{\alpha k} \widehat{X}^1_{pk}$ and $P^{V,1}_{m\beta} = \widehat{V}^1_{m\ell} \widehat{V}^1_{\beta\ell}$. With the definition of $\widetilde{S}^0_{k\ell}$ we obtain

$$\begin{aligned}u^1_{\alpha\beta} =& u^0_{pm} - P^{X,1}_{\alpha p} \Delta t D^x_{pq} u^0_{qn} A_{mn} P^{V,1}_{m\beta} + P^{X,1}_{\alpha p} \Delta t D^{xx}_{pq} u^0_{qn} |A|_{mn} P^{V,1}_{m\beta} \\ &+ \sigma\left(\Delta t P^{X,1}_{\alpha p} B^0_p \delta_{m0} P^{V,1}_{m\beta} - \Delta t u^1_{\alpha\beta}\right).\end{aligned} \tag{5.2}$$

Let us choose a constant solution in space, i.e., $B^1_p = B^1$ and $u^1_{\alpha\beta} = u^1 \delta_{\beta 0}$ for all spatial indices $p, \alpha = 1, ..., n_x$. The scalar values B^1 and u^1 are chosen such that $B^1 = u^1 + \alpha$ where

$$0 < \alpha < \frac{\sigma\Delta t}{1 + \sigma\Delta t + \sigma^2\Delta t^2 + \frac{1}{2}\sigma^3\Delta t^3} u^1.$$

We can now verify that we obtain our chosen values for B^1_p and $u^1_{\alpha\beta}$ after a single step of (5.2) when using the initial condition

$$B^0_p = B^1 + \sigma\Delta t\alpha = u^1 + \alpha(1 + \sigma\Delta t), \tag{5.3a}$$

$$u^0_{pm} = \left(u^1 + \sigma\Delta t(u^1 - B^0_p)\right)\delta_{m0} = \left(u^1 - \sigma\Delta t\alpha(1 + \sigma\Delta t)\right)\delta_{m0}. \tag{5.3b}$$

To show this, note that since the solution is constant in space, all terms containing the stencil matrices $\mathbf{D}^x$ and $\mathbf{D}^{xx}$ drop out and we are left with

$$u^1_{\alpha\beta} = u^0_{pm} + \sigma\left(\Delta t P^{X,1}_{\alpha p} B^0_p \delta_{m0} P^{V,1}_{m\beta} - \Delta t u^1_{\alpha\beta}\right). \tag{5.4}$$

Since B_p^0 is constant in space and δ_{m0} lies in the span of our basis, we know that all projections in the above equation are exact. Plugging the initial values (5.3) into (5.4) we then directly obtain $u^1_{\alpha\beta} = u^1\delta_{\beta 0}$. Similarly, by plugging (5.3) into (5.1d), we obtain $B_p^1 = B^1$.

Then, we square both of the initial terms (5.3) to get

$$(B_p^0)^2 =(B^1)^2 + 2\sigma\Delta t\alpha B^1 + \sigma^2\Delta t^2\alpha^2 = (B^1)^2 + 2\sigma\Delta t\alpha(u^1+\alpha) + \sigma^2\Delta t^2\alpha^2,$$
$$(u^0_{pm})^2 = \left((u^1)^2 - 2\sigma\Delta t\alpha u^1(1+\sigma\Delta t) + \sigma^2\Delta t^2\alpha^2(1+\sigma\Delta t)^2\right)\delta_{m0}.$$

Summing over p and m, adding these two terms and multiplying with $\frac{1}{2}$ yields

$$E^1 = E^0 + \sigma^2\Delta t^2\alpha u^1 - \sigma\Delta t\alpha^2 - \frac{1}{2}\sigma^2\Delta t^2\alpha^2 - \frac{1}{2}\sigma^2\Delta t^2\alpha^2(1+\sigma\Delta t)^2.$$

Note that $E^1 > E^0$ if

$$\sigma\Delta t u^1 - \alpha - \frac{1}{2}\sigma\Delta t\alpha - \frac{1}{2}\sigma\Delta t\alpha(1+\sigma\Delta t)^2 > 0.$$

Rearranging gives

$$\alpha < \frac{\sigma\Delta t}{1+\sigma\Delta t+\sigma^2\Delta t^2+\frac{1}{2}\sigma^3\Delta t^3}u^1.$$

This is exactly the domain α is chosen from. Hence, we have $E^1 > E^0$, which is the desired result. ∎

5.2. Energy stable space-time discretization. We have seen that the naive scheme presented in (5.1) can increase the total energy in one time step. The main goal of this section is to construct a novel energy stable time integration scheme for which the corresponding analysis leads to a classic hyperbolic CFL condition that enables us to operate up to a time step size of $\Delta t = \mathrm{CFL}\cdot\Delta x$. For constructing this energy stable scheme, we write the original equations in two parts followed by a basis augmentation and correction step.

In detail, we first solve

$$K^\star_{pk} = K^0_{pk} - \Delta t D^x_{qp}K^0_{pj}V^0_{nj}A_{mn}V^0_{mk} + \Delta t D^{xx}_{qp}K^0_{pj}V^0_{nj}|A|_{mn}V^0_{mk}, \tag{5.5a}$$
$$L^\star_{mk} = L^0_{mk} - \Delta t X^0_{qk}D^x_{qp}X^0_{pi}L^0_{ni}A_{mn} + \Delta t X^0_{qk}D^{xx}_{qp}X^0_{pi}L^0_{ni}|A|_{mn}. \tag{5.5b}$$

We perform a QR-decomposition of the augmented quantities $\mathbf{X}^\star\mathbf{R} = [\mathbf{K}^\star, \mathbf{X}^0]$ and $\mathbf{V}^\star\widetilde{\mathbf{R}} = [\mathbf{L}^\star, \mathbf{V}^0]$ to obtain the augmented and time updated bases $\mathbf{X}^\star$ and $\mathbf{V}^\star$. Note that $\mathbf{R}$ and $\widetilde{\mathbf{R}}$ are discarded. With $\widetilde{S}^0_{\alpha\beta} = X^\star_{j\alpha}X^0_{j\ell}S^0_{\ell m}V^0_{km}V^\star_{k\beta}$ we then solve the S-step equation

$$S^\star_{\alpha\beta} = \widetilde{S}^0_{\alpha\beta} - \Delta t X^\star_{p\alpha}D^x_{pq}X^\star_{qi}\widetilde{S}^0_{ij}V^\star_{nj}A_{mn}V^\star_{m\beta} + \Delta t X^\star_{p\alpha}D^{xx}_{pq}X^\star_{qi}\widetilde{S}^0_{ij}V^\star_{nj}|A|_{mn}V^\star_{m\beta}. \tag{5.5c}$$

Second, we solve the coupled equations for the internal energy $\mathbf{B}\in\mathbb{R}^{n_x}$ and the quantity $\widehat{\mathbf{u}}^1_{\mathbf{0}} = (\widehat{u}^1_{j0})_j \in \mathbb{R}^{n_x}$ to which we refer as the zeroth order moment according to

$$\begin{aligned}\widehat{u}^1_{j0} = {} & X^0_{j\ell}S^0_{\ell m}V^0_{0m} - \Delta t D^x_{ji}X^\star_{in}\widetilde{S}^0_{nm}V^\star_{\ell m}A_{0\ell} + \Delta t D^{xx}_{ji}X^\star_{in}\widetilde{S}^0_{nm}V^\star_{\ell m}|A|_{0\ell} \\ & + \sigma\Delta t(B^1_j - \widehat{u}^1_{j0}),\end{aligned} \tag{5.5d}$$
$$B^1_j = B^0_j + \sigma\Delta t(\widehat{u}^1_{j0} - B^1_j). \tag{5.5e}$$

Following [21, Section 6] we perform the opacity update only on $\mathbf{L} = \mathbf{V}^{\star}\mathbf{S}^{\star}$ according to

$$L^{\star,\mathrm{scat}}_{mk} = \frac{1}{1+\Delta t\sigma} L_{mk} \quad \text{for } k \neq 0 \tag{5.5f}$$

and perform a QR-decomposition $\mathbf{V}^{\star,\mathrm{scat}}\mathbf{S}^{\star,\mathrm{scat},\top} = \mathbf{L}^{\star,\mathrm{scat}}$ to retrieve the factorized basis $\mathbf{V}^{\star,\mathrm{scat}}$ and the coefficients from the matrix $\mathbf{S}^{\star,\mathrm{scat}}$. We then augment the basis matrices according to

$$\widetilde{\mathbf{X}}^1 = \mathrm{qr}([\widehat{\mathbf{u}}_0^1, \mathbf{X}^{\star}]), \quad \widetilde{\mathbf{V}}^1 = \mathrm{qr}([\mathbf{e}_1, \mathbf{V}^{\star,\mathrm{scat}}]). \tag{5.5g}$$

Third, the coefficient matrix is updated via

$$\widetilde{\mathbf{S}}^1 = \widetilde{\mathbf{X}}^{1,\top}\mathbf{X}^{\star}\mathbf{S}^{\star,\mathrm{scat}}\mathbf{V}^{\star,\mathrm{scat},\top}(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^{\top})\widetilde{\mathbf{V}}^1 + \widetilde{\mathbf{X}}^{1,\top}\widehat{\mathbf{u}}_0^1\mathbf{e}_{1,\top}\widetilde{\mathbf{V}}^1 \in \mathbb{R}^{(2r+1)\times(2r+1)}. \tag{5.5h}$$

Then, we obtain the updated solution $\widetilde{\mathbf{X}}^1\widetilde{\mathbf{S}}^1\widetilde{\mathbf{V}}^{1,\top} \in \mathbb{R}^{n_x \times N}$. Lastly, we truncate this rank $2r+1$ solution to a new rank r_1 using a suited truncation strategy such as proposed in [6] or the conservative truncation strategy of [14]. This finally gives the low-rank factors $\mathbf{X}^1, \mathbf{S}^1$ and $\mathbf{V}^1$. We show that the given scheme is energy stable and start with the following Lemma.

Lemma 5.2. *Let us denote $u^1_{jk} := \widetilde{X}^1_{j\alpha}\widetilde{S}^1_{\alpha\beta}\widetilde{V}^1_{k\beta}$. Under the time step restriction $\Delta t \leq \Delta x$ it holds*

$$\frac{\Delta t}{2}(D^x_{ji}u^1_{jk}A_{k\ell} - D^{xx}_{ji}u^1_{jk}|A|_{k\ell})^2 - \left(D^+_{ji}u^1_{ik}|A|^{1/2}_{k\ell}\right)^2 \leq 0. \tag{5.6}$$

Proof. Following [21], we employ a Fourier analysis which allows us to write the stencil matrices $\mathbf{D}^{x,xx,+}$ in diagonal form. Let us define $\mathbf{E} \in \mathbb{C}^{n_x\times n_x}$ with entries

$$E_{k\alpha} = \sqrt{\Delta x}\exp(i\alpha\pi x_k), \quad k,\alpha = 1, ..., n_x$$

with $i \in \mathbb{C}$ being the imaginary unit. Then, the matrix $\mathbf{E}$ is orthonormal, i.e., $\mathbf{E}\mathbf{E}^H = \mathbf{E}^H\mathbf{E} = \mathbf{I}$ (the uppercase H denotes the complex transpose) and it diagonalizes the stencil matrices:

$$\mathbf{D}^{x,xx,+}\mathbf{E} = \mathbf{E}\mathbf{\Lambda}^{x,xx,+}\ . \tag{5.7}$$

The matrices $\mathbf{\Lambda}^{x,xx,+}$ are diagonal with entries

$$\begin{aligned}
\lambda^x_{\alpha,\alpha} =& \frac{1}{2\Delta x}(e^{i\alpha\pi\Delta x} - e^{-i\alpha\pi\Delta x}) = \frac{i}{\Delta x}\sin(\omega_\alpha)\ ,\\
\lambda^{xx}_{\alpha,\alpha} =& \frac{1}{2\Delta x}\left(e^{i\alpha\pi\Delta x} - 2 + e^{-i\alpha\pi\Delta x}\right) = \frac{1}{\Delta x}\left(\cos(\omega_\alpha) - 1\right)\ ,\\
\lambda^{+}_{\alpha,\alpha} =& \frac{1}{\sqrt{2\Delta x}}\left(e^{i\alpha\pi\Delta x} - 1\right) = \frac{1}{\sqrt{2\Delta x}}\left(\cos(\omega_\alpha) + i\sin(\omega_\alpha) - 1\right)\ ,
\end{aligned}$$

where we use $\omega_\alpha := \alpha\pi\Delta x$. Moreover, recall that we can write $\mathbf{A} = \mathbf{Q}\mathbf{M}\mathbf{Q}^{\top}$ where $\mathbf{M} = \mathrm{diag}(\sigma_1, \cdots, \sigma_N)$. We then have with $\widehat{u}_{jk} = E_{j\ell}u_{\ell m}Q_{mk}$

$$\begin{aligned}
&\frac{\Delta t}{2}(D^x_{ji}u^1_{jk}A_{k\ell} - D^{xx}_{ji}u^1_{jk}|A|_{k\ell})^2 - \left(D^+_{ji}u^1_{ik}|A|^{1/2}_{k\ell}\right)^2\\
=&\frac{\Delta t}{2}\left|\lambda^x_{jj}\widehat{u}^1_{jk}\sigma_k - \lambda^{xx}_{jj}\widehat{u}^1_{jk}|\sigma_k|\right|^2 - \left|\lambda^+_{jj}\widehat{u}^1_{jk}|\sigma_k|^{1/2}\right|^2\\
\leq& \left[\Delta t\left(\frac{|\sigma_k|^2}{\Delta x^2}\cdot|1-\cos(\omega_j)|\right) - \frac{|\sigma_k|}{\Delta x}\cdot|1-\cos(\omega_j)|\right](\widehat{u}^1_{jk})^2.
\end{aligned}$$

To ensure negativity, we must have

$$\Delta t \left(\frac{|\sigma_k|^2}{\Delta x^2} \cdot |1 - \cos(\omega_j)| \right) \leq \frac{|\sigma_k|}{\Delta x} \cdot |1 - \cos(\omega_j)| \,.$$

Hence, for $\Delta t \leq \frac{\Delta x}{|\sigma_k|}$ equation (5.6) holds. Since $|\sigma_k| \leq 1$, we have proven the Lemma. ∎

We can now show energy stability of the proposed scheme:

THEOREM 5.3. *Under the time step restriction $\Delta t \leq \Delta x$, the scheme (5.5) is energy stable, i.e.,*

$$\|\mathbf{B}^1\|_E^2 + \|\mathbf{X}^1\mathbf{S}^1\mathbf{V}^{1,\top}\|_F^2 \leq \|\mathbf{B}^0\|_E^2 + \|\mathbf{X}^0\mathbf{S}^0\mathbf{V}^{0,\top}\|_F^2. \tag{5.8}$$

Proof. First, we multiply (5.5e) with B_j^1 and sum over j. Then,

$$\left(B_j^1\right)^2 = B_j^0 B_j^1 + \sigma \Delta t \left(u_{j0}^1 B_j^1 - \left(B_j^1\right)^2 \right).$$

Let us note that

$$B_j^0 B_j^1 = \frac{\left(B_j^1\right)^2}{2} + \frac{\left(B_j^0\right)^2}{2} - \frac{1}{2}(B_j^1 - B_j^0)^2.$$

Hence,

$$\frac{1}{2}\left(B_j^1\right)^2 = \frac{1}{2}\left(B_j^0\right)^2 - \frac{1}{2}(B_j^1 - B_j^0)^2 + \sigma \Delta t \left(u_{j0}^1 B_j^1 - \left(B_j^1\right)^2 \right). \tag{5.9}$$

To obtain a similar expression for $(u_{jk}^1)^2$, we multiply (5.5c) with $X_{j\alpha}^\star V_{k\beta}^\star$ and sum over α and β. For simplicity of notation, let us define $u_{jk}^\star := X_{j\alpha}^\star S_{\alpha\beta}^\star V_{k\beta}^\star$ and $u_{jk}^0 := X_{j\alpha}^\star \widetilde{S}_{\alpha\beta}^0 V_{k\beta}^\star$ as well as the projections $P_{jp}^X := X_{j\alpha}^\star X_{p\alpha}^\star$ and $P_{km}^V := V_{k\beta}^\star V_{m\beta}^\star$. Then, we obtain the system

$$u_{jk}^\star = u_{jk}^0 - \Delta t P_{jp}^X D_{pq}^x u_{qn}^0 A_{mn} P_{km}^V + \Delta t P_{jp}^X D_{pq}^{xx} u_{qn}^0 |A|_{mn} P_{km}^V. \tag{5.10}$$

Next, we define $u_{jk}^1 := \widetilde{X}_{j\alpha}^1 \widetilde{S}_{\alpha\beta}^1 \widetilde{V}_{k\beta}^1$ and note that by construction we have that

$$u_{jk}^1 = \frac{u_{jk}^\star (1 - \delta_{k0})}{1 + \sigma \Delta t} + \widehat{u}_{j0}^1 \delta_{k0}.$$

Hence, plugging in the schemes for $u_{jk}^\star$ and $\widehat{u}_{j0}^1$, that is, (5.10) and (5.5d) we get

$$\begin{aligned}
(1 + \sigma\Delta t) u_{jk}^1 = {} & \left(u_{jk}^0 - \Delta t P_{jp}^X D_{pq}^x u_{qn}^0 A_{mn} P_{km}^V + \Delta t P_{jp}^X D_{pq}^{xx} u_{qn}^0 |A|_{mn} P_{km}^V \right) (1 - \delta_{k0}) \\
& + \Big(X_{j\ell}^0 S_{\ell m}^0 V_{0m}^0 - \Delta t D_{ji}^x X_{in}^\star \widetilde{S}_{nm}^0 V_{\ell m}^\star A_{0\ell} + \Delta t D_{ji}^{xx} X_{in}^\star \widetilde{S}_{nm}^0 V_{\ell m}^\star |A|_{0\ell} \\
& + \sigma \Delta t B_j^1 \Big) \delta_{k0}.
\end{aligned}$$

Let us note that $P_{km}^V P_{jp}^X u_{jk}^1 = u_{jk}^1$ for $k \neq 0$. Hence, multiplying the above equation with u_{jk}^1 and summing over j and k gives

$$\begin{aligned}
\frac{1}{2}\left(u_{jk}^1\right)^2 = \frac{1}{2}\left(u_{jk}^0\right)^2 - \frac{1}{2}(u_{jk}^1 - u_{jk}^0)^2 & - \Delta t u_{jk}^1 D_{ji}^x u_{i\ell}^0 A_{k\ell} + \Delta t u_{jk}^1 D_{ji}^{xx} u_{i\ell}^0 |A|_{k\ell} \\
& + \sigma \Delta t u_{jk}^1 (B_j^1 \delta_{k0} - u_{jk}^1).
\end{aligned}$$

Let us now add the zero term $\Delta t u_{jk}^1 D_{ji}^x u_{i\ell}^1 A_{k\ell}$ and add and subtract the term $\Delta t u_{jk}^1 D_{ji}^{xx} u_{i\ell}^1 |A|_{k\ell}$. Then,

$$\begin{aligned}\frac{1}{2}\left(u_{jk}^1\right)^2 = \frac{1}{2}\left(u_{jk}^0\right)^2 - \frac{1}{2}(u_{jk}^1 - u_{jk}^0)^2 &- \Delta t u_{jk}^1 D_{ji}^x (u_{i\ell}^0 - u_{i\ell}^1) A_{k\ell} \\ &+ \Delta t u_{jk}^1 D_{ji}^{xx}(u_{i\ell}^0 - u_{i\ell}^1)|A|_{k\ell} + \Delta t u_{jk}^1 D_{ji}^{xx} u_{i\ell}^1 |A|_{k\ell} \\ &+ \sigma \Delta t u_{jk}^1 (B_j^1 \delta_{k0} - u_{jk}^1).\end{aligned}$$

In the following, we use Young's inequality which states that for $a, b \in \mathbb{R}$ we have $a \cdot b \leq \frac{a^2}{2} + \frac{b^2}{2}$. We now apply this to the term

$$\begin{aligned}&-\Delta t u_{jk}^1 D_{ji}^x (u_{i\ell}^0 - u_{i\ell}^1) A_{k\ell} + \Delta t u_{jk}^1 D_{ji}^{xx} (u_{i\ell}^0 - u_{i\ell}^1)|A|_{k\ell} \\ &\leq \frac{1}{2}(u_{i\ell}^0 - u_{i\ell}^1)^2 + \frac{\Delta t^2}{2}(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell})^2.\end{aligned}$$

Hence, using $u_{jk}^1 D_{ji}^{xx} u_{i\ell}^1 |A|_{k\ell} = -\left(D_{ji}^+ u_{ik}^1 |A|_{k\ell}^{1/2}\right)^2$ we get

$$\begin{aligned}\frac{1}{2}\left(u_{jk}^1\right)^2 \leq \frac{1}{2}\left(u_{jk}^0\right)^2 &+ \frac{\Delta t^2}{2}(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell})^2 - \Delta t \left(D_{ji}^+ u_{ik} |A|_{k\ell}^{1/2}\right)^2 \\ &+ \sigma \Delta t u_{jk}^1 (B_j^1 \delta_{k0} - u_{jk}^1).\end{aligned} \tag{5.11}$$

As for the continuous case, we add (5.11) and (5.9) to obtain a time update equation for $E^0 := \frac{1}{2}\left(u_{jk}^0\right)^2 + \frac{1}{2}\left(B_j^0\right)^2$:

$$\begin{aligned}E^1 \leq E^0 &+ \frac{\Delta t^2}{2}(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell})^2 - \Delta t \left(D_{ji}^+ u_{ik}^1 |A|_{k\ell}^{1/2}\right)^2 \\ &+ \sigma \Delta t (u_{j0}^1 B_j^1 - (u_{jk}^1)^2) - \frac{1}{2}(B_j^1 - B_j^0)^2 + \sigma \Delta t \left(u_{j0}^1 B_j^1 - \left(B_j^1\right)^2\right) \\ \leq E^0 &+ \frac{\Delta t^2}{2}(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell})^2 - \Delta t \left(D_{ji}^+ u_{ik}^1 |A|_{k\ell}^{1/2}\right)^2 \\ &- \sigma \Delta t (B_j^1 - u_{jk}^1)^2 - \frac{1}{2}(B_j^1 - B_j^0)^2.\end{aligned} \tag{5.12}$$

With Lemma 5.2 we have that

$$\frac{\Delta t}{2}(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell})^2 - \left(D_{ji}^+ u_{ik}^1 |A|_{k\ell}^{1/2}\right)^2 \leq 0$$

for $\Delta t \leq \Delta x$. Since the truncation step is designed to not alter the zero order moments, we conclude that $E^1 \leq E^0$ and the full scheme is energy stable under the time step restriction $\Delta t \leq \Delta x$. ∎

6. Mass conservation. A drawback of dynamical low-rank approximation using the classical integrators introduced in Section 1 is that the method does not preserve physical invariants. It has been shown in [12] that this problem can be overcome when using a modified L-step equation. On this basis, [14, 17] have presented conservative DLRA algorithms where they additionally introduced a conservative truncation step. In contrast to [14, 17] we do not need to consider a modified L-step equation due to the applied basis augmentation strategy from [6], but use the conservative truncation step. Then we can show that besides being energy stable, our scheme ensures local conservation of mass. The conservative truncation strategy works as follows:

1. Compute $\widetilde{\mathbf{K}} = \widetilde{\mathbf{X}}^1\widetilde{\mathbf{S}}^1$ and split it into two parts $\widetilde{\mathbf{K}} = [\widetilde{\mathbf{K}}^{\text{cons}}, \widetilde{\mathbf{K}}^{\text{rem}}]$ where $\widetilde{\mathbf{K}}^{\text{cons}}$ corresponds to the first and $\widetilde{\mathbf{K}}^{\text{rem}}$ consists of the remaining columns of $\widetilde{\mathbf{K}}$.
   Analogously, distribute $\widetilde{\mathbf{V}}^1 = [\widetilde{\mathbf{V}}^{\text{cons}}, \widetilde{\mathbf{V}}^{\text{rem}}]$ where $\widetilde{\mathbf{V}}^{\text{cons}}$ corresponds to the first and $\widetilde{\mathbf{V}}^{\text{rem}}$ consists of the remaining columns of $\widetilde{\mathbf{V}}$.
2. Derive $\mathbf{X}^{\text{cons}} = \widetilde{\mathbf{K}}^{\text{cons}}/\|\widetilde{\mathbf{K}}^{\text{cons}}\|$ and $\mathbf{S}^{\text{cons}} = \|\widetilde{\mathbf{K}}^{\text{cons}}\|$.
3. Perform a QR-decomposition of $\widetilde{\mathbf{K}}^{\text{rem}}$ to obtain $\widetilde{\mathbf{K}}^{\text{rem}} = \widetilde{\mathbf{X}}^{\text{rem}}\widetilde{\mathbf{S}}^{\text{rem}}$.
4. Compute the singular value decomposition of $\widetilde{\mathbf{S}}^{\text{rem}} = \mathbf{U}\mathbf{\Sigma}\mathbf{W}^\top$ with $\mathbf{\Sigma} = \operatorname{diag}(\sigma_j)$. Given a tolerance ϑ, choose the new rank $r_1 \leq 2r$ as the minimal number such that
$$\left(\sum_{j=r_1+1}^{2r} \sigma_j^2\right)^{1/2} \leq \vartheta.$$
   Let $\mathbf{S}^{\text{rem}}$ be the $r_1 \times r_1$ diagonal matrix with the r_1 largest singular values and let $\mathbf{U}^{\text{rem}}$ and $\mathbf{W}^{\text{rem}}$ contain the first r_1 columns of $\mathbf{U}$ and $\mathbf{W}$, respectively. Set $\mathbf{X}^{\text{rem}} = \widetilde{\mathbf{X}}^{\text{rem}}\mathbf{U}^{\text{rem}}$ and $\mathbf{V}^{\text{rem}} = \widetilde{\mathbf{V}}^{\text{rem}}\mathbf{W}^{\text{rem}}$.
5. Set $\widehat{\mathbf{X}} = [\mathbf{X}^{\text{cons}}, \mathbf{X}^{\text{rem}}]$ and $\widehat{\mathbf{V}} = [\mathbf{e}_1, \mathbf{V}^{\text{rem}}]$. Perform a QR-decomposition of $\widehat{\mathbf{X}} = \mathbf{X}^1\mathbf{R}^1$ and $\widehat{\mathbf{V}} = \mathbf{V}^1\mathbf{R}^2$.
6. Set
$$\mathbf{S}^1 = \mathbf{R}^1 \begin{bmatrix} \mathbf{S}^{\text{cons}} & 0 \\ 0 & \mathbf{S}^{\text{rem}} \end{bmatrix} \mathbf{R}^{2,\top}.$$

The updated solution at time $t_1 = t_0 + \Delta t$ is then given by $\mathbf{u}^1 = \mathbf{X}^1\mathbf{S}^1\mathbf{V}^{1,\top}$. Then, the scheme is conservative:

THEOREM 6.1. *The scheme* (5.5) *is locally conservative. That is, for the scalar flux at time t_n denoted by $\Phi_j^n = X_{j\ell}^n S_{\ell m}^n V_{0m}^n$, where $n \in \{0,1\}$ and $u_{jk}^0 = X_{j\ell}^0 S_{\ell m}^0 V_{km}^0$ it fulfills the conservation law*

$$\Phi_j^1 = \Phi_j^0 - \Delta t D_{ji}^x u_{i\ell}^0 A_{0\ell} + \Delta t D_{ji}^{xx} u_{i\ell}^0 |A|_{0\ell} + \sigma \Delta t (B_j^1 - \Phi_j^1), \tag{6.1a}$$
$$B_j^1 = B_j^0 + \sigma \Delta t (\Phi_j^1 - B_j^1). \tag{6.1b}$$

Proof. The conservatice truncation step is designed such that it does not alter the first column of $\widetilde{\mathbf{X}}^1\widetilde{\mathbf{S}}^1\widetilde{\mathbf{V}}^{1,\top}$. Together with the basis augmentation (5.5g) and correction step (5.5f) we then know that

$$\Phi_j^1 = X_{j\ell}^1 S_{\ell m}^1 V_{0m}^1 = \widetilde{X}_{j\ell}^1 \widetilde{S}_{\ell m}^1 \widetilde{V}_{0m}^1 = \widehat{u}_{j0}^1.$$

Hence, with (5.5d) and (5.5e) we get that

$$\begin{aligned} \Phi_j^1 &= X_{j\ell}^0 S_{\ell m}^0 V_{0m}^0 - \Delta t D_{ji}^x X_{in}^\star \widetilde{S}_{nm}^0 V_{\ell m}^\star A_{0\ell} + \Delta t D_{ji}^{xx} X_{in}^\star \widetilde{S}_{nm}^0 V_{\ell m}^\star |A|_{0\ell} \\ &\quad + \sigma \Delta t (B_j^1 - \Phi_j^1), \\ B_j^1 &= B_j^0 + \sigma \Delta t (\Phi_j^1 - B_j^1). \end{aligned}$$

Since the basis augmentation with $\mathbf{X}^0$ and $\mathbf{V}^0$ ensures $X_{j\ell}^0 S_{\ell m}^0 V_{0m}^0 = X_{in}^\star \widetilde{S}_{nm}^0 V_{\ell m}^\star = u_{i\ell}^0$, the local conservation law (6.1) holds. ∎

Hence, equipped with a conservative truncation step, the energy stable algorithm presented in (5.5) conserves mass locally. To give an overview of the algorithm, we visualize the main steps in Figure 1.

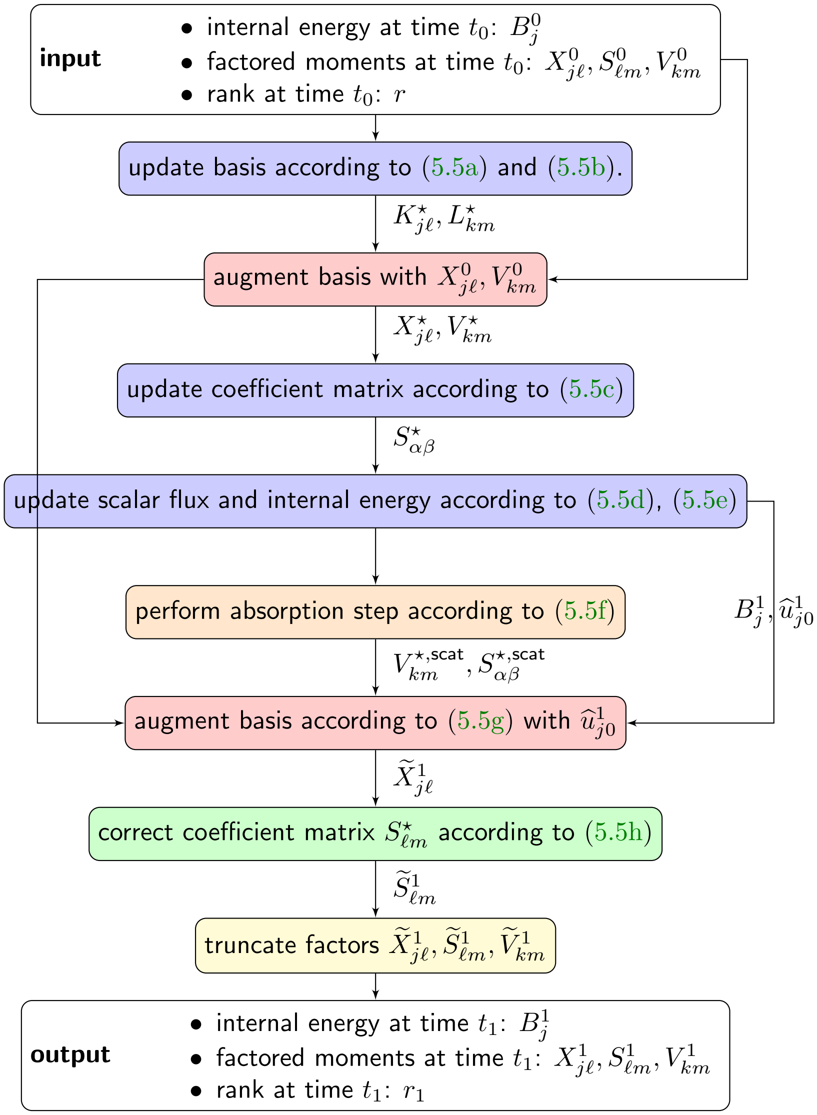

FIG. 1. *Flowchart of the stable and conservative method* (5.5).

7. Numerical results.

In this section we give numerical results to validate the proposed DLRA algorithm. The source code to reproduce the presented numerical results is openly available, see [2].

7.1. 1D Plane source.

We consider the thermal radiative transfer equations as described in (2.1a) on the spatial domain $D = [-10, 10]$. As initial distribution we choose a cutoff Gausian

$$u(t = 0, x) = \max\left(10^{-4}, \frac{1}{\sqrt{2\pi\sigma_{\mathrm{IC}}^2}} \exp\left(-\frac{(x-1)^2}{2\sigma_{\mathrm{IC}}^2}\right)\right),$$

with constant deviation $\sigma_{\mathrm{IC}} = 0.03$. Particles are initially centered around $x = 1$ and move into all directions $\mu \in [-1, 1]$. The initial value for the internal energy is set to

$B^0 = 1$ and we start computations with a rank of $r = 20$. The opacity σ is set to the constant value of 1. Note that this setting is an extension of the so-called *plane source* problem, which is a common test case for the radiative transfer equation [16]. In the context of dynamical low-rank approximation it has been studied in [6, 21, 34, 36]. We compare the solution of the full coupled-implicit system without DLRA which reads

$$u^1_{jk} = u^0_{jk} - \Delta t D^x_{ji} u^0_{i\ell} A_{k\ell} + \Delta t D^{xx}_{ji} u^0_{i\ell} |A|_{k\ell} + \sigma \Delta t (B^1_j \delta_{k0} - u^1_{jk}) \tag{7.1a}$$

$$B^1_j = B^0_j + \sigma \Delta t (u^1_{j0} - B^1_j) \tag{7.1b}$$

to the presented energy stable mass conservative DLRA solution from (5.5). We refer to (7.1) as the full system. The total mass at any time t_n shall be defined as $m^n = \Delta x \sum_j \left(u^n_{j0} + B^n_j \right)$. As computational parameters we use $n_x = 1000$ cells in the spatial domain and $N = 500$ moments to represent the directional variable. The time step size is chosen as $\Delta t = \mathrm{CFL} \cdot \Delta x$ with a CFL number of $\mathrm{CFL} = 0.99$. In Figure 2 we present computational results for the solution $f(x, \mu)$, the scalar flux $\Phi = \langle f \rangle_\mu$ and the temperature T at the end time $t_{\text{end}} = 8$. Further, the evolution of the rank r in time, and the relative mass error $\frac{|m^0 - m^n|}{\|m^0\|}$ are shown. One can observe that the DLRA scheme captures well the behaviour of the full system. For a chosen tolerance of $\vartheta = 10^{-1} \|\mathbf{\Sigma}\|_2$ the rank increases up to $r = 24$ before it reduces again. The relative mass error is of order $\mathcal{O}(10^{-14})$. Hence, our proposed scheme is mass conservative up to machine precision.

7.2. 1D Su-Olson problem. For the next test problem we add a source term $Q(x)$ to the previously investigated equations leading to

$$\begin{aligned} \partial_t f(t, x, \mu) + \mu \partial_x f(t, x, \mu) &= \sigma (B(t, x) - f(t, x, \mu)) + Q(x), \\ \partial_t B(t, x) &= \sigma (\langle f(t, x, \cdot) \rangle_\mu - B(t, x)). \end{aligned}$$

In our example we use the source function $Q(x) = \chi_{[-0.5, 0.5]}(x)/a$ with $a = \frac{4\sigma_{\text{SB}}}{c}$ being the radiation constant. Again we consider the spatial domain $D = [-10, 10]$ and choose the initial condition

$$u(t = 0, x) = \max \left(10^{-4}, \frac{1}{\sqrt{2\pi\sigma^2_{\text{IC}}}} \exp \left(-\frac{(x-1)^2}{2\sigma^2_{\text{IC}}} \right) \right),$$

with constant deviation $\sigma_{\text{IC}} = 0.03$ and particles moving into all directions $\mu \in [-1, 1]$. The initial value for the internal energy is set to $B_0 = 50$, the initial value for the rank to $r = 20$. The opacity σ is again chosen to have the constant value of 1. As computational parameters we use $n_x = 1000$ cells in the spatial domain and $N = 500$ moments to represent the directional variable. The time step size is chosen as $\Delta t = \mathrm{CFL} \cdot \Delta x$ with a CFL number of $\mathrm{CFL} = 0.99$. The isotropic source term generates radiation particles flying through and interacting with a background material. The interaction is driven by the opacity σ. In turn, particles heat up the material leading to a travelling temperature front, also called a *Marshak wave* [26]. Again this travelling heat wave can lead to the emission of new particles from the background material generating a particle wave. At a given time point $t_{\text{end}} = 3.16$ this waves can be seen in Figure 3 where we display numerical results for the solution $f(x, \mu)$, the scalar flux $\Phi = \langle f \rangle_\mu$ and the temperature T. We compare the solution of the full coupled-implicit system differing from (7.1) by an additional source term to the presented energy stable mass conservative DLRA solution from (5.5) where we have

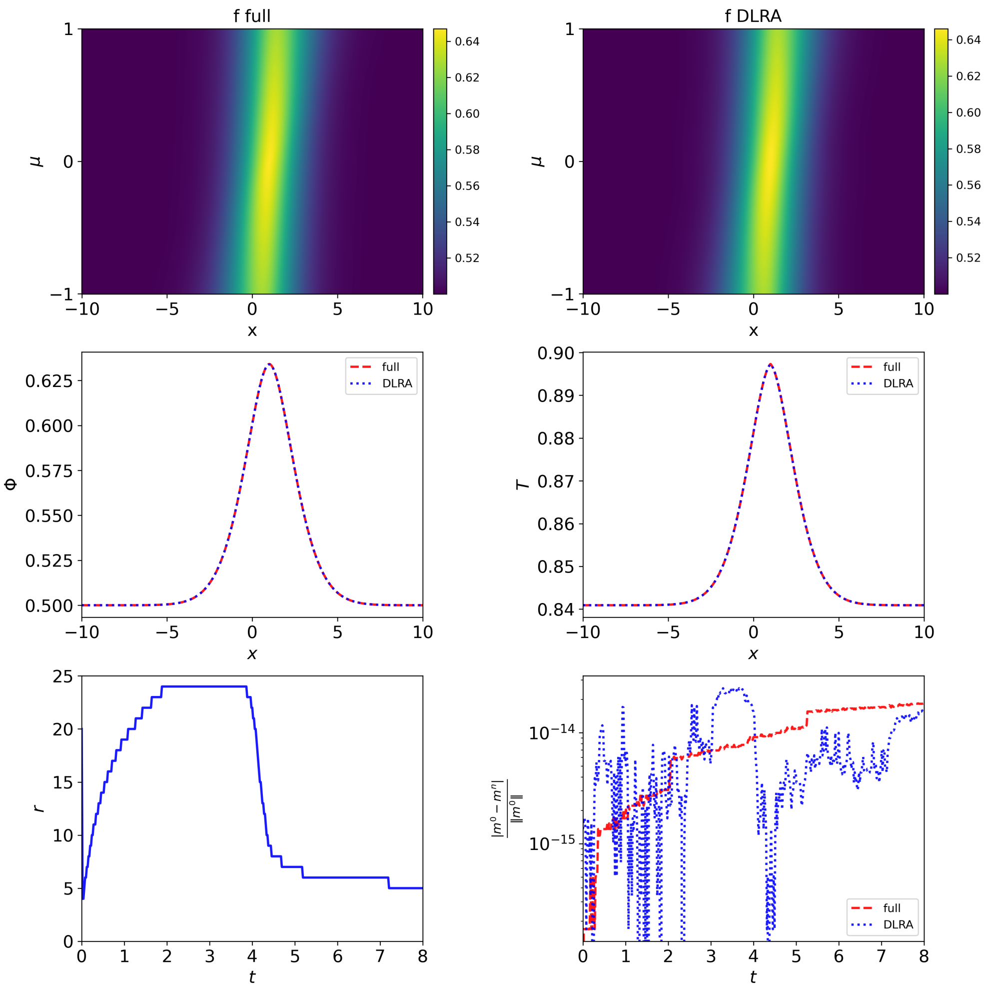

FIG. 2. *Top row: Numerical results for the solution $f(x, \mu)$ of the plane source problem at time $t_{end} = 8$ computed with the full coupled-implicit system (left) and the DLRA system (right). Middle row: Travelling particle (left) and heat wave (right) for both the full system and the DLRA system. Bottom row: Evolution of the rank in time for the DLRA method (left) and relative mass error compared for both methods (right).*

also added this source term. Further, the evolution of the rank in time is presented for a tolerance parameter of $\vartheta = 10^{-2}\|\mathbf{\Sigma}\|_2$. Again we observe that the proposed DLRA scheme approximates well the behaviour of the full system. In addition, a very low rank is sufficient to obtain accurate results. Note that due to the source term there is no mass conservation in this example.

7.3. 2D Beam. To approve computational benefits of the presented method we extend it to a two-dimensional setting. The set of equations becomes:

$$\begin{aligned} \partial_t f(t,\mathbf{x},\mathbf{\Omega}) + \mathbf{\Omega}\cdot\nabla_{\mathbf{x}} f(t,\mathbf{x},\mathbf{\Omega}) &= \sigma(B(t,\mathbf{x}) - f(t,\mathbf{x},\mathbf{\Omega})), \\ \partial_t B(t,\mathbf{x}) &= \sigma(\langle f(t,\mathbf{x},\cdot)\rangle_{\mathbf{\Omega}} - B(t,\mathbf{x})). \end{aligned}$$

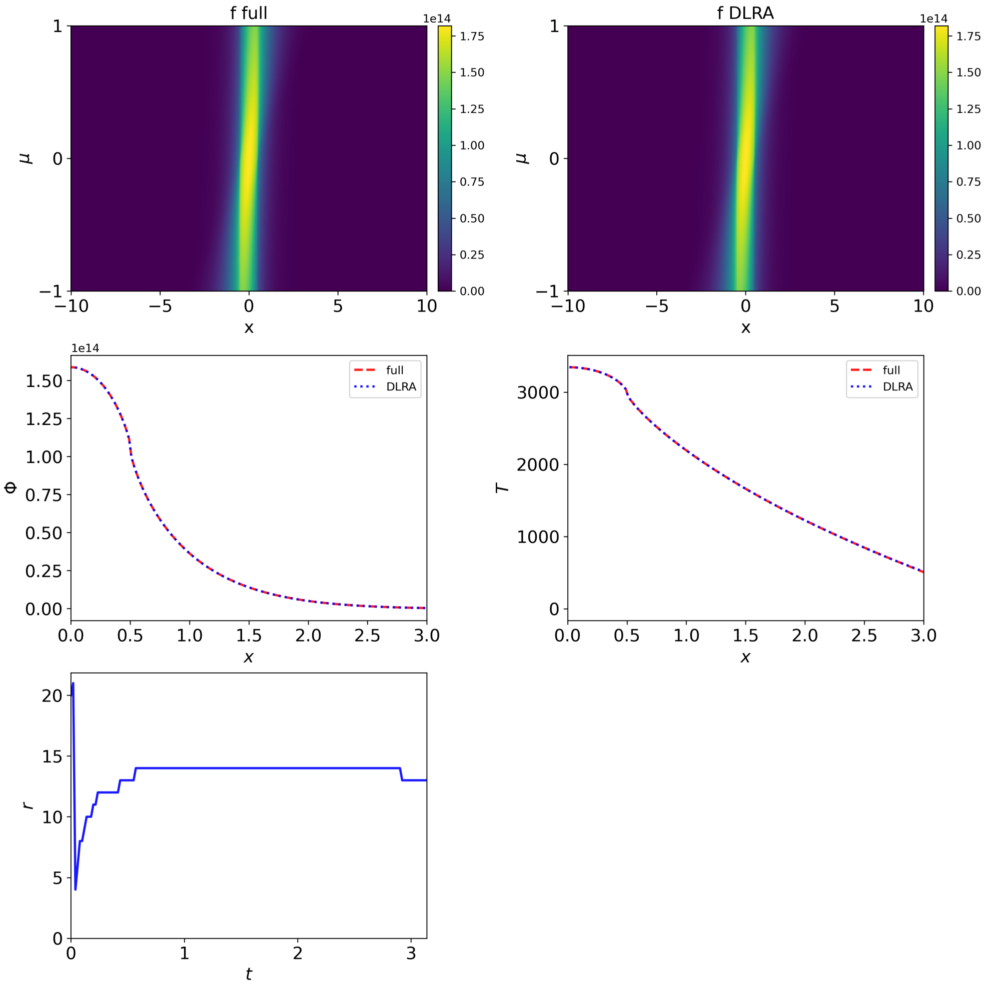

FIG. 3. *Top row: Numerical results for the solution $f(x, \mu)$ of the Su-Olson problem at time $t_{end} = 3.16$ computed with the full coupled-implicit system (left) and the DLRA system (right). Middle row: Travelling particle (left) and heat wave (right) for both the full system and the DLRA system. Bottom row: Evolution of the rank in time for the DLRA method.*

For the numerical experiments let $\mathbf{x} = (x_1, x_2) \in [-1, 1] \times [-1, 1]$, $\mathbf{\Omega} = (\Omega_1, \Omega_2, \Omega_3) \in \mathcal{S}^2$ and $\sigma = 0.5$. The initial condition of the two-dimensional beam is given by

$$f(t = 0, \mathbf{x}, \mathbf{\Omega}) = 10^6 \cdot \frac{1}{2\pi\sigma_x^2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma_x^2}\right) \cdot \frac{1}{2\pi\sigma_\Omega^2} \exp\left(-\frac{(\Omega_1 - \Omega^\star)^2 + (\Omega_3 - \Omega^\star)^2}{2\sigma_\Omega^2}\right),$$

with $\Omega^\star = \frac{1}{\sqrt{2}}$, $\sigma_x = \sigma_\Omega = 0.1$. The initial value for the internal energy is set to $B^0 = 1$, the initial value for the rank to $r = 100$. The total mass at any time t_n shall be defined as $m^n = \Delta x_1 \Delta x_2 \sum_j \left(u_{j0}^n + B_j^n\right)$. We perform our computations on a spatial grid with $N_{\text{CellsX}} = 500$ points in x_1 and $N_{\text{CellsY}} = 500$ points in x_2. For the angular basis we use again a modal approach, namely the spherical harmonics (P_N) method. Technical details can be found in [4, 31, 29], whereas [36, 22] relates the method to dynamical low-rank approximation. The polynomial degree shall be chosen large enough such that the behaviour is captured correctly but small enough to stay in a reasonable computational regime. An increasing order of unknowns usually leads to an increasing complexity and therefore to the need of a higher polynomial degree. For our example we use a polynomial degree of $n_{\mathrm{PN}} = 29$ corresponding to 900 expansion coefficients in angle. The time step size is chosen as $\Delta t = \mathrm{CFL} \cdot \Delta x$ with a CFL number of $\mathrm{CFL} = 0.7$. We compare the solution of the two-dimensional full system corresponding to (7.1) to the two-dimensional DLRA solution corresponding to (5.5). The extension to two dimensions is straightforward. In Figure 4 we show numerical results for the scalar flux $\Phi = \int_{\mathcal{S}^2} f(t, \mathbf{x}, \cdot)\, \mathrm{d}\mathbf{\Omega}$ and the temperature T at the time $t = 0.5$. We again observe the accuracy of the proposed DLRA scheme. For this setup the computational benefit of the DLRA method is significant as the run time compared to the solution of the full problem is reduced by a factor of approximately 8 from 20023 seconds to 2509 seconds. For the evolution of the rank r in time and the relative mass error $\frac{|m^0 - m^n|}{\|m^0\|}$ we consider a time interval up to $t = 1.5$. In Figure 5 one can observe that for a chosen tolerance parameter of $\vartheta = 5 \cdot 10^{-4} \|\mathbf{\Sigma}\|_2$ the rank increases but does not approach its allowed maximal value of 100. Further, the relative mass error stagnates and the DLRA method shows its mass conservation property.

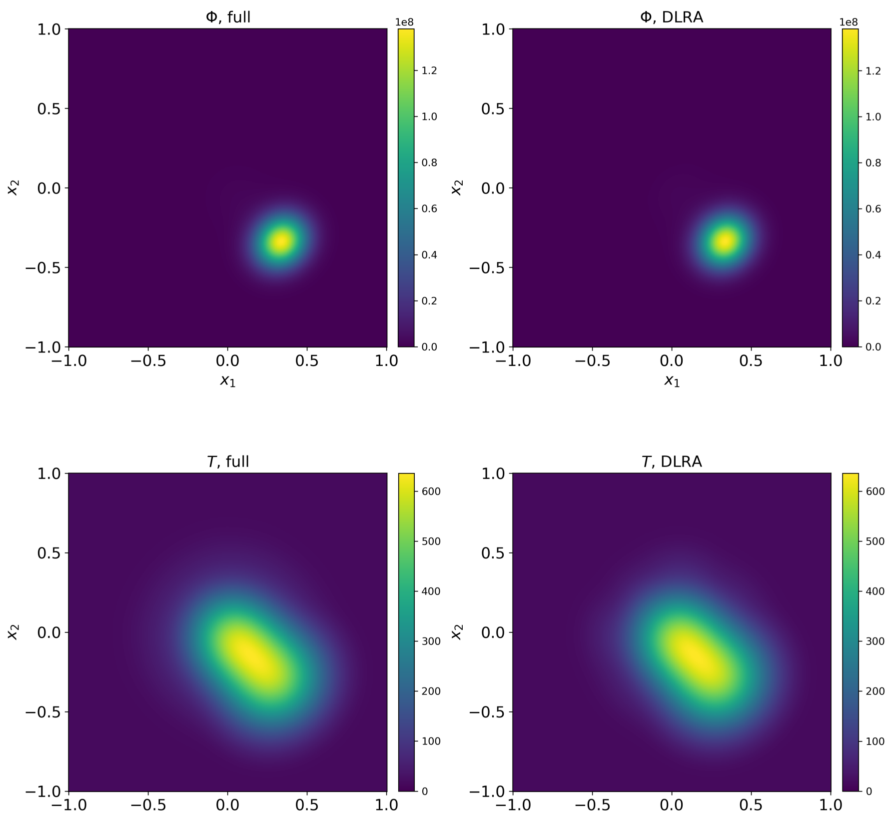

Fig. 4. *Numerical results of the scalar flux and the temperature for the 2D beam example for the full coupled-implicit system (left) and the DLRA system (right) at the time $t = 0.5$.*

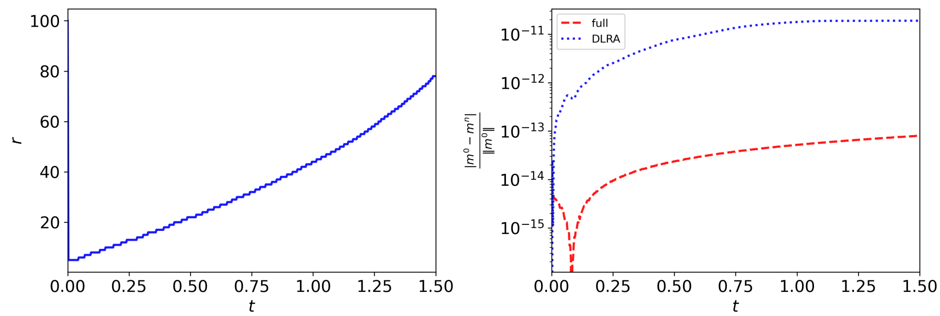

FIG. 5. *Evolution of the rank in time for the 2D beam example for the DLRA method (left) and relative mass error compared for both methods (right) until a time of* $t = 1.5$.

8. Conclusion and outlook.

We have introduced an energy stable and mass conservative dynamical low-rank algorithm for the Su-Olson problem. The key points leading to these properties consist in treating both equations in a coupled-implicit way and using a mass conservatice truncation strategy. Numerical examples both in 1D and 2D validate the accuracy of the DLRA method. Its efficiency compared to the solution of the full system can especially be seen in the two-dimensional setting. For future work, we propose to implement the parallel integrator of [7] for further enhancing the efficiency of the DLRA method. Moreover, we expect to draw conclusions from this Su-Olson system to the Boltzmann-BGK system and the DLRA algorithm presented in [11] regarding stability and an appropriate choice of the size of the time step.

REFERENCES

[1] I. Abu-Shumays, *Angular quadratures for improved transport computations*, Transport Theory and Statistical Physics, 30 (2001), pp. 169–204.

[2] L. Baumann, L. Einkemmer, C. Klingenberg, and J. Kusch, *Numerical testcases for "Energy stable and conservative dynamical low-rank approximation for the Su-Olson problem"*, (2023), https://github.com/JonasKu/publication-Energy-stable-and-conservative-dynamical-low-rank-approximation-for-the-Su-Olson-problem.git.

[3] T. Camminady, M. Frank, K. Küpper, and J. Kusch, *Ray effect mitigation for the discrete ordinates method through quadrature rotation*, Journal of Computational Physics, 382 (2019), pp. 105–123.

[4] K. M. Case and P. F. Zweifel, *Linear Transport Theory*, Addison-Wesley, Reading, Massachusetts, 1967.

[5] G. Ceruti, M. Frank, and J. Kusch, *Dynamical low-rank approximation for Marshak waves*, CRC 1173 Preprint 2022/76, Karlsruhe Institute of Technology, (2022).

[6] G. Ceruti, J. Kusch, and C. Lubich, *A rank-adaptive robust integrator for dynamical low-rank approximation*, BIT Numerical Mathematics, 62 (2022), pp. 1149–1174.

[7] G. Ceruti, J. Kusch, and C. Lubich, *A parallel rank-adaptive integrator for dynamical low-rank approximation*, arXiv preprint arXiv:2304.05660, (2023).

[8] G. Ceruti and C. Lubich, *An unconventional robust integrator for dynamical low-rank approximation*, BIT Numerical Mathematics, 62 (2022), pp. 23–44.

[9] L. Einkemmer, J. Hu, and J. Kusch, *Asymptotic–preserving and energy stable dynamical low-rank approximation*, arXiv preprint arXiv:2212.12012, (2022).

[10] 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, 439 (2021), p. 110353.

[11] L. Einkemmer, J. Hu, and L. Ying, *An efficient dynamical low-rank algorithm for the Boltzmann-BGK equation close to the compressible viscous flow regime*, SIAM Journal on Scientific Computing, 43 (2021), pp. B1057–B1080.

[12] L. Einkemmer and J. Ilon, *A mass, momentum, and energy conservative dynamical low-rank scheme for the Vlasov equation*, Journal of Computational Physics, 443 (2021), p. 110493.

[13] L. Einkemmer and C. Lubich, *A low-rank projector-splitting integrator for the Vlasov-Poisson equation*, SIAM Journal on Scientific Computing, 40 (2018), pp. B1330–B1360.

[14] L. Einkemmer, A. Ostermann, and C. Scalone, *A robust and conservative dynamical low-rank algorithm*, arXiv preprint arXiv:2206.09374, (2022).

[15] M. Frank, J. Kusch, T. Camminady, and C. D. Hauck, *Ray effect mitigation for the discrete ordinates method using artificial scattering*, Nuclear Science and Engineering, 194 (2020), pp. 971–988.

[16] B. D. Ganapol, *Analytical benchmarks for nuclear engineering applications*, Case Studies in Neutron Transport Theory, (2008).

[17] W. Guo and J.-M. Qiu, *A conservative low rank tensor method for the Vlasov dynamics*, arXiv preprint arXiv:2201.10397, (2022).

[18] J. Hu and Y. Wang, *An adaptive dynamical low rank method for the nonlinear Boltzmann equation*, Journal of Scientific Computing, 92 (2022), p. 75.

[19] E. Kieri, C. Lubich, and H. Walach, *Discretized dynamical low-rank approximation in the presence of small singular values*, SIAM Journal on Numerical Analysis, 54 (2016), pp. 1020–1038.

[20] O. Koch and C. Lubich, *Dynamical low-rank approximation*, SIAM Journal on Matrix Analysis and Applications, 29 (2007), pp. 434–454.

[21] J. Kusch, L. Einkemmer, and G. Ceruti, *On the stability of robust dynamical low-rank approximations for hyperbolic problems*, SIAM Journal on Scientific Computing, 45 (2023), pp. A1–A24.

[22] J. Kusch and P. Stammer, *A robust collision source method for rank adaptive dynamical low-rank approximation in radiation therapy*, ESAIM: M2AN, (2022).

[23] K. D. Lathrop, *Ray effects in discrete ordinates equations*, Nuclear Science and Engineering, 32 (1968), pp. 357–369.

[24] K. D. Lathrop, *Remedies for ray effects*, Nuclear Science and Engineering, 45 (1971), pp. 255–268.

[25] C. Lubich and I. V. Oseledets, *A projector-splitting integrator for dynamical low-rank approximation*, BIT Numerical Mathematics, 54 (2014), pp. 171–188.

[26] R. E. Marshak, *Effect of radiation on shock wave behavior*, Physics of Fluids, 1 (1958), pp. 24–29.

[27] K. A. Mathews, *On the propagation of rays in discrete ordinates*, Nuclear science and engineering, 132 (1999), pp. 155–180.

[28] R. G. McClarren, T. M. Evans, R. B. Lowrie, and J. D. Densmore, *Semi-implicit time integration for P_n thermal radiative transfer*, Journal of Computational Physics, 227 (2008), pp. 7561–7586.

[29] R. G. McClarren and C. D. Hauck, *Robust and accurate filtered spherical harmonics expansions for radiative transfer*, Journal of Computational Physics, 229 (2010), pp. 5597–5614.

[30] R. G. McClarren, J. P. Holloway, and T. A. Brunner, *Analytic P_1, solutions for time-dependant, thermal radiative transfer in several geometries*, Journal of Quantitative Spectroscopy and Radiative Transfer, 109 (2008), pp. 389–403.

[31] R. G. McClarren, J. P. Holloway, and T. A. Brunner, *On solutions to the P_n equations for thermal radiative transfer*, Journal of Computational Physics, 227 (2008), pp. 2864–2885.

[32] J. Morel, T. Wareing, R. Lowrie, and D. Parsons, *Analysis of ray-effect mitigation techniques*, Nuclear science and engineering, 144 (2003), pp. 1–22.

[33] G. L. Olson, L. H. Auer, and M. L. Hall, *Diffusion, P_1, and other approximate forms of radiation transport*, Journal of Quantitative Spectroscopy and Radiative Transfer, 62 (2000), pp. 619–634.

[34] Z. Peng and R. G. McClarren, *A high-order/low-order (HOLO) algorithm for preserving conservation in time-dependent low-rank transport calculations*, Journal of Computational Physics, 447 (2021), p. 110672.

[35] Z. Peng and R. G. McClarren, *A sweep-based low-rank method for the discrete ordinate transport equation*, Journal of Computational Physics, 473 (2023), p. 111748.

[36] Z. Peng, R. G. McClarren, and M. Frank, *A low-rank method for two-dimensional time-dependent radiation transport calculations*, Journal of Computational Physics, 421 (2020), p. 109735.

[37] G. C. Pomraning, *The non-equilibrium marshak wave problem*, Journal of Quantitative Spectroscopy and Radiative Transfer, 21 (1979), pp. 249–261.

[38] B. Su and G. L. Olson, *An analytical benchmark for non-equilibrium radiative transfer in an isotropically scattering medium*, Annals of Nuclear Energy, 24 (1997), pp. 1035–1055.

[39] J. Tencer, *Ray effect mitigation through reference frame rotation*, Journal of Heat Transfer, 138 (2016).