An energy stable and conservative multiplicative dynamical low-rank discretization for the Su-Olson problem

Lena Baumann[a], Lukas Einkemmer[b], Christian Klingenberg[a], Jonas Kusch[c]

[a]*Department of Mathematics, University of Würzburg, Emil-Fischer-Strasse 40, Würzburg, DE-97074, Germany*

[b]*Numerical Analysis and Scientific Computing, University of Innsbruck, Technikerstrasse 13, Innsbruck, A-6020, Austria*

[c]*Scientific Computing, Norwegian University of Life Sciences, Drøbakveien 31, Ås, NO-1433, Norway*

Abstract

Computing numerical solutions of the thermal radiative transfer equations on a finely resolved grid can be costly due to high computational and memory requirements. A numerical reduced order method that has recently been applied to a wide variety of kinetic partial differential equations is the concept of dynamical low-rank approximation (DLRA). In this paper, we consider the thermal radiative transfer equations with Su-Olson closure, leading to a linearized kinetic model. For the conducted theoretical and practical considerations we use a multiplicative splitting of the distribution function that poses additional challenges in finding an energy stable discretization and deriving a hyperbolic Courant-Friedrichs-Lewy (CFL) condition. We propose such an energy stable DLRA scheme that makes use of the augmented basis update & Galerkin integrator. This integrator allows for additional basis augmentations, enabling us to give a mathematically rigorous proof of energy stability and local mass conservation. Numerical examples confirm the derived properties and show the computational advantages of the DLRA scheme compared to a numerical solution of the full system of equations.

Keywords: thermal radiative transfer, Su-Olson closure, dynamical low-rank approximation, multiplicative splitting, energy stability, mass

conservation

1. Introduction

Thermal radiative transfer problems are a class of kinetic transport equations modeling the movement of particles that interact with a background medium, for instance by scattering or absorption. By this interaction, the background medium can heat up and itself emit new particles, enforcing the exchange of energy between particles and the background material. This process is described by two coupled equations, one for the particle density $f(t, x, \mu)$ and one for the internal energy $e(t, x)$ of the material. The variable $t \in \mathbb{R}^+$ denotes time, $x \in D \subset \mathbb{R}$ stands for the spatial and $\mu \in [-1, 1]$ for the directional variable. The numerical solution of this set of equations is challenging as its high dimensionality requires enormous computational and memory costs. To overcome these problems, the concept of dynamical low-rank approximation (DLRA) [1] can be used. It is a numerical reduced order method providing accurate and efficient approximations of the solution of kinetic partial differential equations and has already been applied in various fields of research. Recent work for instance has been published on radiation transport [2, 3, 4], radiation therapy [5], plasma physics [6, 7], chemical kinetics [8] and Boltzmann type transport problems [9, 10]. The main idea of DLRA consists in approximating the distribution function as

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

where $\{X_i : i = 1, .., r\}$ are the orthonormal basis functions in space and $\{V_j : j = 1, .., r\}$ the orthonormal basis functions in direction. This splitting can be understood as a continuous analogue to the singular value decomposition of a matrix, explaining why $\mathbf{S} = (S_{ij}) \in \mathbb{R}^{r \times r}$ is called the coefficient or coupling matrix and r the rank of this approximation. However, note that the matrix $\mathbf{S}$ is not required to be diagonal. The time evolution of the low-rank factors is then determined by a projection of the equation onto the tangent space of the low-rank solution manifold. Integrators that ensure that the solution stays on the low-rank manifold while being robust to small singular values (otherwise this may lead to enormous restrictions on the choice of the time

step size [11]) are the *projector-splitting* [12], the *(augmented) basis update & Galerkin* (BUG) [13, 14], and the *parallel integrator* [15].

A challenge for constructing a stable dynamical low-rank scheme is to derive a suitable discretization of the system. To account for the angular dependence of the solution, we consider a modal representation that makes use of a finite expansion of the particle density in terms of spherical harmonics (P_N). This approach is explained in [16, 17] and for instance used in [18, 19, 20]. Also the spatial discretization has to be chosen carefully. In [9] it is shown that for an efficient dynamical low-rank scheme for the isothermal Boltzmann-BGK equation it is useful to consider a multiplicative splitting of the distribution function. To transfer knowledge to such systems, we decide on a multiplicative splitting of the particle density f also for the considered problem. Different from [21], in this case it is per se not clear how to deal with the spatial derivatives. In addition, the multiplicative splitting poses further challenges for the proof of energy stability and the construction of a low-rank scheme to which we account for instance by pursuing a 'first discretize, then low-rank' strategy. Recent work on this topic has been published for the linear Boltzmann-BGK equation in [22]. For the time discretization the potentially stiff opacity term has to be taken into account, leading to a coupled-implicit scheme similar to the one treated in [21], which again is complicated to solve.

In this paper we propose an energy stable multiplicative dynamical low-rank discretization for a linearized internal energy model called the Su-Olson problem. The main novelties are:

- *A multiplicative splitting of the distribution function:* Based on the insights from [9], we consider a multiplicative splitting of the distribution function. Similar to [22], we show that the spatial discretization has to be chosen carefully. Further, the multiplicative splitting requires additional modifications in the low-rank scheme in order to obtain an energy stable numerical discretization of the problem.

- *An energy stable numerical scheme with rigorous mathematical proofs:* We show that the derived DLRA scheme is energy stable and give rigorous mathematical proofs, enabling us to deduce a classic hyperbolic CFL condition. This allows to compute up to a maximal time step size of $\Delta t = \text{CFL} \cdot \Delta x$, which again enhances the performance of the algorithm.

3

- *A mass conservative augmented integrator:* We make use of the augmented BUG integrator from [14], leading to a basis augmentation in two substeps of the numerical scheme. As this integrator allows for further modifications, we include additional basis augmentation steps that ensure the exactness of the projection operators needed for the theoretical proof of energy stability as well as the local conservation of mass.

- *Numerical test examples confirming the derived properties:* We compare the numerical results obtained from the low-rank discretization with the solution of the full system for relevant examples from the literature. The derived properties are validated and the accuracy and the efficiency of the proposed DLRA method is shown.

The structure of the paper is as follows: After the introduction in Section 1, we provide background information on the thermal radiative transfer equations in Section 2. We explain the considered multiplicative structure and derive two possible systems that in the continuous setting are equivalent. In Section 3, we discretize both systems in angle, space and time. Section 4 is devoted to the subject of energy stability. We show that the advection form of the multiplicative Su-Olson problem is generally not numerically stable in the sense of von Neumann, whereas for the conservative form a hyperbolic CFL condition is derived, under which an energy estimate can be given. Section 5 first provides an overview of the method of dynamical low-rank approximation. Then, a provably energy stable DLRA scheme is derived. The provided rigorous stability analysis is the main novelty of this work. In addition, mass conservation can be shown, when using a suitable truncation strategy. The numerical results in Section 6 confirm the derived properties, before in Section 7 a brief conclusion and outlook is given.

2. Thermal radiative transfer

In a one-dimensional setting the thermal radiative transfer equation with absorbing background material is given by

$$\frac{1}{c}\partial_t f(t,x,\mu) + \mu\partial_x f(t,x,\mu) = \sigma(B(t,x) - f(t,x,\mu)),$$

$$\partial_t e(t,x) = \sigma(\langle f(t,x,\cdot) - B(t,x)\rangle_\mu),$$

where an integration over the directional domain $[-1, 1]$ is denoted by $\langle \cdot \rangle_\mu$. The speed of light is denoted by c and the variable σ represents the opacity that encodes the rate at which particles are absorbed by the background medium. The black body radiation $B(T)$ at the material temperature T can be described by the Stefan-Boltzmann law

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

where $a = \frac{4\sigma_{\text{SB}}}{c}$ is the radiation density constant and σ_{SB} the Stefan-Boltzmann constant. Further information on the thermal radiative transfer equations and their relevance in physics can be found in [23, 16, 24]. The above set of equations is not closed. To determine a relation between the temperature T and the internal energy $e(T)$ we follow the ideas of Pomraning [25] and Su and Olson [26] and set $e(T) = \alpha B(T)$. From this point on, we call $\alpha B(T)$ the internal energy of the material. Further, we perform a rescaling $\tau = \frac{t}{c}$ and by an abuse of notation write t instead of τ in the remainder. This leads to the system

$$\partial_t f(t, x, \mu) + \mu \partial_x f(t, x, \mu) = \sigma(B(t, x) - f(t, x, \mu)), \tag{1a}$$
$$\partial_t B(t, x) = \sigma(\langle f(t, x, \cdot) - B(t, x) \rangle_\mu), \tag{1b}$$

where without loss of generality we assume $\alpha = 1$. This system is a closed linearized internal energy model that is analytically solvable and commonly used as a benchmark for numerical examples [27, 18, 19]. In the following, we call equations (1) the *Su-Olson problem*. Note that for the moment we omit initial and boundary conditions.

In [9] it has been shown that for deriving an efficient dynamical low-rank scheme for the isothermal Boltzmann-BGK equation it is crucial to consider a multiplicative splitting of the distribution function. This allows one to separate a generally not low-rank Maxwellian from a remaining low-rank function g, to which the DLRA scheme is subsequently applied. The considered Su-Olson problem is similar in structure to this equation. To transfer knowledge of the construction of stable efficient dynamical low-rank schemes from the Su-Olson problem to more general kinetic equations, we have decided on a multiplicative splitting of the distribution function of the form

$$f(t, x, \mu) = B(t, x)g(t, x, \mu), \tag{2}$$

and apply the low-rank ansatz to g. For this system, we give a mathematically rigorous proof of energy stability and derive a hyperbolic CFL condition,

which allows for determining an optimal time step size, making the proposed scheme even more efficient. In this sense, this paper can be understood as an intermediate step from the Su-Olson problem treated in [21] towards more complicated non-linear problems with multiplicative splitting. The main novelty of this paper consists in providing a rigorous theoretical analysis for the multiplicative Su-Olson problem, distinguishing this work for instance from the multiplicative DLRA scheme proposed in [9], where the time step size of the constructed algorithm is chosen experimentally based on the corresponding numerical experiment. The idea of applying a multiplicative splitting to a linearized model and deriving a concrete CFL condition is also pursued in [22] for the linear Boltzmann-BGK equation. However, there are crucial differences between the multiplicative Su-Olson problem and the multiplicative linear Boltzmann-BGK equation. In the latter, due to the appearance of the Maxwellian equilibrium distribution, the evolution equation couples with the momentum equations. In the Su-Olson problem, the evolution equation for the particle density (1a) couples through (1b) with the background material. This leads to a different notion of stability in the theoretical analysis. In addition, the construction of the low-rank scheme requires new ideas for the basis augmentations as well as an adjusted truncation strategy to ensure mass conservation, which is not available in [22]. Hence, this paper provides novel insights into the construction of conservative and provably stable multiplicative low-rank schemes.

We insert ansatz (2) into (1a) and (1b) and obtain the set of equations

$$\partial_t g(t,x,\mu) = -\mu \partial_x g(t,x,\mu) + \sigma \left(1 - g(t,x,\mu)\right) - \frac{g(t,x,\mu)}{B(t,x)} \partial_t B(t,x) \quad (3a)$$
$$- \mu \frac{g(t,x,\mu)}{B(t,x)} \partial_x B(t,x),$$
$$\partial_t B(t,x) = \sigma B(t,x) \left(\langle g(t,x,\mu) \rangle_\mu - 2\right), \quad (3b)$$

that is called the *advection form* of the multiplicative system. Using the product rule, it splits up the spatial derivatives for B and g in (3a). This corresponds to the form in which the multiplicative splitting in [9] is applied to the Boltzmann-BGK equation. Equation (3a) can be equivalently rewritten into a *conservative form,* leaving the spatial derivative of Bg together and

leading to the system

$$\partial_t g(t, x, \mu) = -\frac{\mu}{B(t, x)} \partial_x \left(B(t, x) g(t, x, \mu) \right) + \sigma \left(1 - g(t, x, \mu) \right) \quad \text{(4a)}$$

$$-\frac{g(t, x, \mu)}{B(t, x)} \partial_t B(t, x),$$

$$\partial_t B(t, x) = \sigma B(t, x) \left(\langle g(t, x, \mu) \rangle_\mu - 2 \right). \quad \text{(4b)}$$

In later considerations, we are interested in the conservation properties of our numerical scheme. For the multiplicative Su-Olson problem, the mass and the momentum of the system shall be defined as follows.

Definition 1 (Macroscopic quantities). *The mass of the multiplicative Su-Olson problem is defined as*

$$\rho(t, x) = \int f(t, x, \mu) \mathrm{d}\mu + B(t, x) = B(t, x) \int g(t, x, \mu) \mathrm{d}\mu + B(t, x).$$

The momentum is given as

$$u(t, x) = \int \mu f(t, x, \mu) \mathrm{d}\mu = B(t, x) \int \mu g(t, x, \mu) \mathrm{d}\mu.$$

In particular, the multiplicative Su-Olson problem satisfies the local conservation law

$$\partial_t \rho(t, x) + \partial_x u(t, x) = 0. \quad \text{(5)}$$

Global conservation of mass is then obtained by integrating over the spatial domain [6].

In the following, we discretize both sets of equations to compare them in terms of numerical stability. We derive an energy stable dynamical low-rank scheme and give a concrete hyperbolic CFL condition. Note that in contrast to [21, 9], but similar to [22], we first discretize the equations and then apply the low-rank ansatz here.

3. Discretization of the multiplicative system

In this section, we fully discretize the advection form (3) as well as the conservative form (4) of the multiplicative system. We start with the angular and spatial discretization, followed by the time discretization.

7

3.1. Angular discretization

For the angular discretization a modal approach with normalized Legendre polynomials P_ℓ is used [16, 17]. They constitute a complete set of orthogonal functions on the interval $[-1, 1]$ that satisfy $\langle P_k, P_\ell \rangle_\mu = \delta_{k\ell}$. As an approximation, a finite expansion of the distribution function g with N_μ expansion coefficients, called the moments, is used. It writes

$$g(t, x, \mu) \approx g_{N_\mu}(t, x, \mu) = \sum_{\ell=0}^{N_\mu - 1} v_\ell(t, x) P_\ell(\mu).$$

We insert this representation into (3), multiply (3a) with $P_k(\mu)$ and integrate over μ. Further, we introduce the matrix $\mathbf{A} \in \mathbb{R}^{N_\mu \times N_\mu}$ with entries $A_{k\ell} := \langle P_k, \mu P_\ell \rangle_\mu$ and use that $P_0 = \frac{1}{\sqrt{2}}$. This gives

$$\partial_t v_k(t, x) = - \sum_{\ell=0}^{N_\mu - 1} \partial_x v_\ell(t, x) A_{k\ell} + \sigma \left(\sqrt{2} \delta_{k0} - v_k(t, x) \right) \tag{6a}$$

$$- \frac{v_k(t, x)}{B(t, x)} \partial_t B(t, x) - \sum_{\ell=1}^{N_\mu - 1} \frac{v_\ell}{B(t, x)} \partial_x B(t, x) A_{k\ell},$$

$$\partial_t B(t, x) = \sigma B(t, x) \left(\sqrt{2} v_0(t, x) - 2 \right). \tag{6b}$$

Analogously, we obtain for system (4) the angularly discretized equations

$$\partial_t v_k(t, x) = - \sum_{\ell=0}^{N_\mu - 1} \frac{1}{B(t, x)} \partial_x \left(B(t, x) v_\ell(t, x) \right) A_{k\ell} + \sigma \left(\sqrt{2} \delta_{k0} - v_k(t, x) \right)$$

$$\tag{7a}$$

$$- \frac{v_k(t, x)}{B(t, x)} \partial_t B(t, x),$$

$$\partial_t B(t, x) = \sigma B(t, x) \left(\sqrt{2} v_0(t, x) - 2 \right). \tag{7b}$$

Note that the matrix $\mathbf{A}$ is symmetric and diagonalizable in the form $\mathbf{A} = \mathbf{Q}\mathbf{M}\mathbf{Q}^\top$ with $\mathbf{Q}$ orthonormal and $\mathbf{M} = \text{diag}(\sigma_1, ..., \sigma_{N_\mu})$. We then define $|\mathbf{A}| = \mathbf{Q}|\mathbf{M}|\mathbf{Q}^\top$.

3.2. Spatial discretization

For the spatial discretization we prescribe a spatial grid with N_x grid cells and equidistant spacing $\Delta x = \frac{1}{N_x}$. Spatially dependent quantities are approximated at the grid points x_j for $j = 1, ..., N_x$, and denoted by

$$B_j(t) \approx B(t, x_j), \qquad v_{jk}(t) \approx v_k(t, x_j).$$

First-order spatial derivatives ∂_x are approximated using the tridiagonal stencil matrices $\mathbf{D}^x \in \mathbb{R}^{N_x \times N_x}$ and a second-order stabilization term $\mathbf{D}^{xx} \in \mathbb{R}^{N_x \times N_x}$ approximating $\frac{1}{2} \Delta x \partial_{xx}$ is added. The entries of those matrices are defined as

$$D^x_{j,j\pm1} = \frac{\pm1}{2\Delta x} \, , \qquad D^{xx}_{j,j} = -\frac{1}{\Delta x} \, , \qquad D^{xx}_{j,j\pm1} = \frac{1}{2\Delta x} \, .$$

Note that from now on we assume periodic boundary conditions to which we account by setting

$$D^x_{1,N_x} = \frac{-1}{2\Delta x} \, , \qquad D^x_{N_x,1} = \frac{1}{2\Delta x} \, , \qquad D^{xx}_{1,N_x} = D^{xx}_{N_x,1} = \frac{1}{2\Delta x}.$$

The stencil matrices $\mathbf{D}^x$ and $\mathbf{D}^{xx}$ then fulfill the following properties:

Lemma 1. *Let $y, z \in \mathbb{R}^{N_x}$ with indices $i, j = 1, ..., N_x$. It holds*

$$\sum_{i,j=1}^{N_x} y_j D^x_{ji} z_i = -\sum_{i,j=1}^{N_x} z_j D^x_{ji} y_i \, , \quad \sum_{i,j=1}^{N_x} z_j D^x_{ji} z_i = 0 \, , \quad \sum_{i,j=1}^{N_x} y_j D^{xx}_{ji} z_i = \sum_{i,j=1}^{N_x} z_j D^{xx}_{ji} y_i.$$

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

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

Then, $\sum_{i,j=1}^{N_x} z_j D^{xx}_{ji} z_i = -\sum_{j=1}^{N_x} \left(\sum_{i=1}^{N_x} D^+_{ji} z_i \right)^2.$

Proof. See [21, Lemma 4.2]. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

We insert the proposed discretization into the angularly discretized advection form (6) of the equations and add a second-order stabilization term for $B\partial_x v$,

leading to the angularly and spatially discretized set of equations

$$\dot{v}_{jk}(t) = -\sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} D_{ji}^x v_{i\ell}(t) A_{k\ell} + \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} D_{ji}^{xx} v_{i\ell}(t)|A|_{k\ell} \tag{8a}$$

$$+ \sigma\left(\sqrt{2}\delta_{k0} - v_{jk}(t)\right) - \frac{v_{jk}(t)}{B_j(t)}\dot{B}_j(t) - \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} \frac{v_{j\ell}(t)}{B_j(t)} D_{ji}^x B_i(t) A_{k\ell},$$

$$\dot{B}_j(t) = \sigma B_j(t)\left(\sqrt{2}v_{j0}(t) - 2\right). \tag{8b}$$

Inserting the discretization into the angularly discretized conservative form (7) of the equations and adding a second-order stabilization term to $\partial_x(Bv)$ gives

$$\dot{v}_{jk}(t) = -\sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} \frac{1}{B_j(t)} D_{ji}^x B_i(t) v_{i\ell}(t) A_{k\ell}$$

$$+ \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} \frac{1}{B_j(t)} D_{ji}^{xx} B_i(t) v_{i\ell}(t)|A|_{k\ell} \tag{9a}$$

$$+ \sigma\left(\sqrt{2}\delta_{k0} - v_{jk}(t)\right) - \frac{v_{jk}(t)}{B_j(t)}\dot{B}_j(t),$$

$$\dot{B}_j(t) = \sigma B_j(t)\left(\sqrt{2}v_{j0}(t) - 2\right). \tag{9b}$$

Note that due to the different structure of the equations the stabilization term in (8a) is applied to $B\partial_x v$, whereas in (9a) it is added for $\partial_x(Bv)$.

3.3. Time discretization

From [21] we know that constructing an energy stable scheme for the Su-Olson problem is challenging. For the advection form of the equations we start from (8) and apply an explicit Euler step to transport terms. The potentially stiff absorption term is treated implicitly and the time derivative $\partial_t B$ is approximated by its difference quotient. We obtain the following fully

discrete space-time discretization

$$v_{jk}^1 = v_{jk}^0 - \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} D_{ji}^x v_{i\ell}^0 A_{k\ell} + \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} D_{ji}^{xx} v_{i\ell}^0 |A|_{k\ell} \tag{10a}$$

$$+ \sigma \Delta t \left(\sqrt{2} \delta_{k0} - v_{jk}^1 \right) - \Delta t \frac{1}{B_j^0} \frac{B_j^1 - B_j^0}{\Delta t} v_{jk}^1 - \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \frac{v_{j\ell}^0}{B_j^0} D_{ji}^x B_i^0 A_{k\ell},$$

$$B_j^1 = B_j^0 + \sigma \Delta t B_j^1 \left(\sqrt{2} v_{j0}^1 - 2 \right), \tag{10b}$$

that describes one time step from time t_0 to time $t_1 = t_0 + \Delta t$ and holds for all further time steps equivalently. For the conservative form (9) we again apply an explicit Euler step to the transport parts, treat the absorption terms implicitly and approximate $\partial_t B$ by its difference quotient. In addition, we add a factor $\frac{B^1}{B^0}$ in the absorption term of (9a). This gives the fully discrete scheme

$$v_{jk}^1 = v_{jk}^0 - \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \frac{1}{B_j^0} D_{ji}^x B_i^0 v_{i\ell}^0 A_{k\ell} + \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \frac{1}{B_j^0} D_{ji}^{xx} B_i^0 v_{i\ell}^0 |A|_{k\ell}$$

$$\tag{11a}$$

$$+ \sigma \Delta t \frac{B_j^1}{B_j^0} \left(\sqrt{2} \delta_{k0} - v_{jk}^1 \right) - \Delta t \frac{1}{B_j^0} \frac{B_j^1 - B_j^0}{\Delta t} v_{jk}^1,$$

$$B_j^1 = B_j^0 + \sigma \Delta t B_j^1 \left(\sqrt{2} v_{j0}^1 - 2 \right). \tag{11b}$$

Note that the evolution equations (10b) and (11b) for the internal energy B are the same in both schemes. The main difference of (10a) and (11a) consists in the distinct second-order stabilization terms and the additional factor $\frac{B_j^1}{B_j^0}$ in (11a) that we will explain later when showing energy stability.

4. Energy stability

The goal of this section is to investigate energy stability of the derived schemes. Note that this section is closely related to the considerations in [21]. We first introduce the following notations.

Definition 2. *In the following we write $u_{jk}^0 := B_j^0 v_{jk}^0$ and $u_{jk}^1 := B_j^1 v_{jk}^1$ at time t_0 and t_1, respectively. Note that $\mathbf{u}^0 = (u_{jk}^0) \in \mathbb{R}^{N_x \times N_\mu}$ corresponds to $f(t = 0, x, \mu)$ and $\mathbf{v}^0 = (v_{jk}^0) \in \mathbb{R}^{N_x \times N_\mu}$ corresponds to $g(t = 0, x, \mu)$ in (2).*

11

With this notation we can give the definition of the total energy of a fully discretized system.

Definition 3 (Total energy). *Let $\mathbf{u}^0 \in \mathbb{R}^{N_x \times N_\mu}$ be the fully discretized angular solution of the full Su-Olson problem and $\mathbf{B}^0 = (B_j^0) \in \mathbb{R}^{N_x}$ the internal energy at time t_0. Then, the total energy at this time is defined as*

$$E^0 := \frac{1}{2}\|\mathbf{u}^0\|_F^2 + \frac{1}{2}\|\mathbf{B}^0\|_E^2,$$

where $\|\cdot\|_F$ denotes the Frobenius and $\|\cdot\|_E$ the Euclidean norm. For $t_1 = t_0 + \Delta t$ this definition shall hold analogously.

4.1. Advection form

We start with the advection form (10) of the Su-Olson problem which is comparable to the considered low-rank discretization in [9] for the isothermal Boltzmann-BGK equation in the sense that the term $\partial_x(Bv)$ is split up into the sum of $B\partial_x v$ and $v\partial_x B$. We can show that this scheme is not, in general, von Neumann stable.

Theorem 1. *There exist initial values $\mathbf{v}^0 \in \mathbb{R}^{N_x \times N_\mu}$ and $\mathbf{B}^0 \in \mathbb{R}^{N_x}$ such that the advection form (10) of the Su-Olson problem for $\sigma = 0$ is not von Neumann stable.*

Proof. Let us assume a solution v_{jk}^0 that is constant in space and direction, e.g. $v_{jk}^0 = 1$. For this solution all spatial derivatives are zero, i.e. the terms containing $\mathbf{D}^x \mathbf{v}^0$ and $\mathbf{D}^{xx} \mathbf{v}^0$ in (10a) drop out. We further assume that for the opacity it holds $\sigma = 0$, i.e. the Su-Olson problem reduces to a simple advection equation. From (10b) we thus obtain that $B_j^1 = B_j^0 = B_j$, i.e. the internal energy is constant in time. We insert these results into (10a) and get

$$v_{jk}^1 = 1 - \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu - 1} \frac{1}{B_j} D_{ji}^x B_i A_{k\ell}.$$

Multiplication with B_j then leads to

$$u_{jk}^1 = u_{jk}^0 - \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu - 1} D_{ji}^x u_{i\ell}^0 A_{k\ell}.$$

This is a discretization of $\partial_t u + \mu \partial_x u = 0$ with an explicit Euler step forward in time and a centered finite difference scheme in space. From [28] we know that this discretization is not von Neumann stable. $\qquad\square$

This result matches our considerations that the DLRA scheme presented in [9] is not von Neumann stable but stable for relatively small time step sizes. Hence, Theorem 1 is intended to serve as a motivation to seek a generally stable numerical discretization as done in the next section.

4.2. Conservative form

For the conservative form of the discretization of the Su-Olson problem given in (11), we can derive a hyperbolic CFL condition and show that under this time step restriction the total energy of the system decreases over time. We start with the following lemma.

Lemma 2. *Under the time step restriction* $\Delta t \leq \Delta x$ *it holds*

$$\frac{\Delta t}{2} \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \left(\sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell} \right) \right)^2$$
$$- \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \left(\sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} D_{ji}^+ u_{jk}^1 |A|_{k\ell}^{1/2} \right)^2 \leq 0,$$

Proof. See [21, Lemma 5.2]. ☐

With this relation we can now show that the energy of the conservative form (11) of the Su-Olson system dissipates and hence the system is energy stable.

Theorem 2. *Under the time step restriction* $\Delta t \leq \Delta x$ *the fully discrete system* (11) *is energy stable, i.e. it holds* $E^1 \leq E^0$.

Proof. The proof of this theorem is similar to the proof of [21, Theorem 5.3]. We start with equation (11b) and multiply it with B_j^1. This gives

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

Note that it holds

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

13

We insert this relation and sum over j, giving

$$\frac{1}{2}\sum_{j=1}^{N_x}\left(B_j^1\right)^2 = \frac{1}{2}\sum_{j=1}^{N_x}\left(B_j^0\right)^2 - \frac{1}{2}\sum_{j=1}^{N_x}\left(B_j^1 - B_j^0\right)^2 \qquad (12)$$

$$+ \sigma\Delta t\sum_{j=1}^{N_x}\left(B_j^1\right)^2\left(\sqrt{2}v_{j0}^1 - 2\right).$$

Next, we multiply equation (11a) with $B_j^1 B_j^0 v_{jk}^1$ and sum over j and k. This leads to

$$\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1} B_j^1 B_j^0\left(v_{jk}^1\right)^2 = \sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1} B_j^0 v_{jk}^0 B_j^1 v_{jk}^1 - \Delta t\sum_{i,j=1}^{N_x}\sum_{k,\ell=0}^{N_\mu-1} B_j^1 v_{jk}^1 D_{ji}^x B_i^0 v_{i\ell}^0 A_{k\ell}$$

$$+ \Delta t\sum_{i,j=1}^{N_x}\sum_{k,\ell=0}^{N_\mu-1} B_j^1 v_{jk}^1 D_{ji}^{xx} B_i^0 v_{i\ell}^0 |A|_{k\ell} \qquad (13)$$

$$+ \sigma\Delta t\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(B_j^1\right)^2 v_{jk}^1\left(\sqrt{2}\delta_{k0} - v_{jk}^1\right)$$

$$- \sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1} B_j^1\left(v_{jk}^1\right)^2\left(B_j^1 - B_j^0\right)$$

Note that for this step the additional factor $\frac{B_j^1}{B_j^0}$ in the absorption term of (11a) is crucial. As above, it holds for the first term on the right-hand side that

$$\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1} B_j^0 v_{jk}^0 B_j^1 v_{jk}^1$$

$$= \sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(\frac{1}{2}\left(B_j^1 v_{jk}^1\right)^2 + \frac{1}{2}\left(B_j^0 v_{jk}^0\right)^2 - \frac{1}{2}\left(B_j^1 v_{jk}^1 - B_j^0 v_{jk}^0\right)^2\right).$$

We insert the notation $u_{jk}^0 = B_j^0 v_{jk}^0$ and $u_{jk}^1 = B_j^1 v_{jk}^1$, respectively, as well as insert the above relation into (13), bring the last term of (13) to the left-hand

14

side and rearrange the equation. We obtain

$$
\begin{aligned}
\frac{1}{2} \sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(u_{jk}^1\right)^2 = {} & \frac{1}{2} \sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(u_{jk}^0\right)^2 - \frac{1}{2} \sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(u_{jk}^1 - u_{jk}^0\right)^2 \\
& - \Delta t \sum_{i,j=1}^{N_x} \sum_{k,\ell=0}^{N_\mu-1} u_{jk}^1 D_{ji}^x u_{i\ell}^0 A_{k\ell} \\
& + \Delta t \sum_{i,j=1}^{N_x} \sum_{k,\ell=0}^{N_\mu-1} u_{jk}^1 D_{ji}^{xx} u_{i\ell}^0 |A|_{k\ell} \\
& + \sigma \Delta t \sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(B_j^1\right)^2 v_{jk}^1 \left(\sqrt{2}\delta_{k0} - v_{jk}^1\right).
\end{aligned}
\tag{14}
$$

We now add the zero term $\Delta t \sum_{i,j=1}^{N_x} \sum_{k,\ell=0}^{N_\mu-1} u_{jk}^1 D_{ji}^x u_{i\ell}^1 A_{k\ell}$ and add and subtract the second-order term $\Delta t \sum_{i,j=1}^{N_x} \sum_{k,\ell=0}^{N_\mu-1} u_{jk}^1 D_{ji}^{xx} u_{i\ell}^1 |A|_{k\ell}$ giving

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

In the next step, we apply Young's inequality, which states that for $a, b \in \mathbb{R}$

we have $a \cdot b \leq \frac{a^2}{2} + \frac{b^2}{2}$, to the term

$$
- \Delta t \sum_{i,j=1}^{N_x} \sum_{k,\ell=0}^{N_\mu-1} u_{jk}^1 D_{ji}^x \left(u_{i\ell}^0 - u_{i\ell}^1 \right) A_{k\ell} + \Delta t \sum_{i,j=1}^{N_x} \sum_{k,\ell=0}^{N_\mu-1} u_{jk}^1 D_{ji}^{xx} \left(u_{i\ell}^0 - u_{i\ell}^1 \right) |A|_{k\ell}
$$

$$
= - \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \left(u_{i\ell}^0 - u_{i\ell}^1 \right) \left(\sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell} \right) \right)
$$

$$
\leq \frac{1}{2} \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \left(u_{i\ell}^0 - u_{i\ell}^1 \right)^2
$$

$$
+ \frac{(\Delta t)^2}{2} \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \left(\sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell} \right) \right)^2.
$$

In addition, we note that with the properties of the stencil matrices from Lemma 1 we can write

$$
\Delta t \sum_{i,j=1}^{N_x} \sum_{k,\ell=0}^{N_\mu-1} u_{jk}^1 D_{ji}^{xx} u_{i\ell}^1 |A|_{k\ell} = - \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \left(\sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} D_{ji}^+ u_{jk}^1 |A|_{k\ell}^{1/2} \right)^2.
$$

We insert both relations and get

$$
\frac{1}{2} \sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(u_{jk}^1 \right)^2 \leq \frac{1}{2} \sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(u_{jk}^0 \right)^2
$$

$$
+ \frac{(\Delta t)^2}{2} \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \left(\sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(D_{ji}^x u_{jk}^1 A_{k\ell} - D_{ji}^{xx} u_{jk}^1 |A|_{k\ell} \right) \right)^2
$$

$$
- \Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} \left(\sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} D_{ji}^+ u_{jk}^1 |A|_{k\ell}^{1/2} \right)^2
$$

$$
+ \sigma \Delta t \sum_{j=1}^{N_x} \sum_{k=0}^{N_\mu-1} \left(B_j^1 \right)^2 v_{jk}^1 \left(\sqrt{2} \delta_{k0} - v_{jk}^1 \right).
$$

With Lemma 2 we have that under the time step restriction $\Delta t \leq \Delta x$ it

holds

$$\frac{1}{2}\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(u_{jk}^1\right)^2 \leq \frac{1}{2}\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(u_{jk}^0\right)^2 + \sigma\Delta t\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(B_j^1\right)^2 v_{jk}^1\left(\sqrt{2}\delta_{k0} - v_{jk}^1\right).$$

$$(15)$$

To obtain an expression for the total energy of the system, we add (15) and (12). This gives

$$E^1 \leq E^0 - \frac{1}{2}\sum_{j=1}^{N_x}\left(B_j^1 - B_j^0\right)^2 + \sigma\Delta t\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(B_j^1\right)^2 v_{jk}^1\left(\sqrt{2}\delta_{k0} - v_{jk}^1\right)$$

$$+ \sigma\Delta t\sum_{j=1}^{N_x}\left(B_j^1\right)^2\left(\sqrt{2}v_{j0}^1 - 2\right).$$

The term $-\frac{1}{2}\sum_{j=1}^{N_x}\left(B_j^1 - B_j^0\right)^2$ is non-positive. The remaining two terms on the right-hand side can be rewritten and bounded as follows:

$$\sigma\Delta t\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(B_j^1\right)^2 v_{jk}^1\left(\sqrt{2}\delta_{k0} - v_{jk}^1\right) + \sigma\Delta t\sum_{j=1}^{N_x}\left(B_j^1\right)^2\left(\sqrt{2}v_{j0}^1 - 2\right)$$

$$\leq \sigma\Delta t\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(B_j^1\right)^2\left(-\left(v_{jk}^1\right)^2 + 2\sqrt{2}v_{jk}^1\delta_{k0} - 2\delta_{k0}\right)$$

$$= -\sigma\Delta t\sum_{j=1}^{N_x}\sum_{k=0}^{N_\mu-1}\left(B_j^1\right)^2\left(v_{jk}^1 - \sqrt{2}\delta_{k0}\right)^2 \leq 0.$$

Hence, we have shown that under the time step restriction $\Delta t \leq \Delta x$ it holds $E^1 \leq E^0$, and the system is energy stable. $\qquad\square$

5. Dynamical low-rank approximation for the energy stable system

Having attained an energy stable discretization of the multiplicative Su-Olson problem, its practical implementation can still pose numerical challenges such as large memory demands and computational costs, especially in a higher-dimensional setting. To overcome these problems, we introduce the concept of dynamical low-rank approximation.

5.1. Background on DLRA

The method of dynamical low-rank approximation has originally been introduced in a semi-discrete time-continuous matrix setting [1], in which it will also be explained in this section. Let us consider the matrix differential equation

$$\dot{\mathbf{f}}(t) = \mathbf{F}\left(\mathbf{f}(t)\right),$$

where $\mathbf{f}(t) \in \mathbb{R}^{N_x \times N_\mu}$ is the solution of the equation and $\mathbf{F}\left(\mathbf{f}(t)\right) : \mathbb{R}^{N_x \times N_\mu} \to \mathbb{R}^{N_x \times N_\mu}$ denotes its right-hand side. The dynamical low-rank approximation of $\mathbf{f}(t)$ is then given by

$$\mathbf{f}_r(t) = \mathbf{X}(t)\mathbf{S}(t)\mathbf{V}(t)^\top, \tag{16}$$

where $\mathbf{X}(t) \in \mathbf{R}^{N_x \times r}$ is the orthonormal basis in space, $\mathbf{V}(t) \in \mathbf{R}^{N_\mu \times r}$ the orthonormal basis in direction, and $\mathbf{S}(t) \in \mathbb{R}^{r \times r}$ the coupling or coefficient matrix of the rank r approximation. All matrices of the form (16) constitute the manifold of low-rank matrices $\mathcal{M}_r$. The basis matrices $\mathbf{X}(t)$ and $\mathbf{V}(t)$, and the coefficient matrix $\mathbf{S}(t)$ shall then be evolved in time such that the minimization problem

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

is fulfilled at all times t, where $\mathcal{T}_{\mathbf{f}_r(t)}\mathcal{M}_r$ denotes the tangent space of the low-rank manifold $\mathcal{M}_r$ at $\mathbf{f}_r(t)$. In [1], it has been shown that solving this minimization problem is equivalent to projecting the right-hand side $\mathbf{F}\left(\mathbf{f}_r(t)\right)$ by means of an orthogonal projection $\mathbf{P}$ onto $\mathcal{T}_{\mathbf{f}_r(t)}\mathcal{M}_r$, and solving

$$\dot{\mathbf{f}}_r(t) = \mathbf{P}\left(\mathbf{f}_r(t)\right)\mathbf{F}\left(\mathbf{f}_r(t)\right). \tag{17}$$

For $\mathbf{f}_r = \mathbf{X}\mathbf{S}\mathbf{V}^\top$, the orthogonal projection $\mathbf{P}$ onto $\mathcal{T}_{\mathbf{f}_r(t)}\mathcal{M}_r$ is explicitly given [1] as

$$\mathbf{P}(\mathbf{f}_r)\mathbf{F}\left(\mathbf{f}_r\right) = \mathbf{X}\mathbf{X}^\top\mathbf{F}\left(\mathbf{f}_r\right) - \mathbf{X}\mathbf{X}^\top\mathbf{F}\left(\mathbf{f}_r\right)\mathbf{V}\mathbf{V}^\top + \mathbf{F}\left(\mathbf{f}_r\right)\mathbf{V}\mathbf{V}^\top.$$

Different time integrators that make use of the special form of this orthogonal projection such as the projector-splitting [12], the (augmented) basis update & Galerkin (BUG) [13, 14], and the parallel integrator [15] exist. They are able to evolve the solution on the low-rank manifold while not suffering from

18

the stiffness of (17). In this paper, we focus on the augmented basis update & Galerkin (BUG) integrator, that shall be explained in the following.

The BUG integrator first updates and augments the spatial basis $\mathbf{X}$ and the directional basis $\mathbf{V}$ in parallel. This leads to an increase from rank r to rank $2r$. Note that we denote augmented quantities of rank $2r$ with hats. Next, a Galerkin step is conducted for the coefficient matrix $\mathbf{S}$ in the augmented setting, before in a last step all augmented quantities are truncated to a new rank $r_1 \leq 2r$. To be more specific, the BUG integrator evolves the low-rank solution $\mathbf{f}_r^0 = \mathbf{X}^0 \mathbf{S}^0 \mathbf{V}^{0,\top}$ at time t_0 to the time-updated low-rank solution $\mathbf{f}_r^1 = \mathbf{X}^1 \mathbf{S}^1 \mathbf{V}^{1,\top}$ at time $t_1 = t_0 + \Delta t$ as follows:

$\boldsymbol{K}$-**Step**: We fix the directional basis $\mathbf{V}^0$ at time t_0 and introduce the notation $\mathbf{K}(t) = \mathbf{X}(t)\mathbf{S}(t)$. Then, we update the spatial basis from $\mathbf{X}^0$ to $\widehat{\mathbf{X}}^1 \in \mathbb{R}^{N_x \times 2r}$ by solving the PDE

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

and then determining $\widehat{\mathbf{X}}^1$ as an orthonormal basis of the augmented matrix $[\mathbf{K}(t_1), \mathbf{X}^0] \in \mathbb{R}^{N_x \times 2r}$, e.g. by QR-decomposition. We store $\widehat{\mathbf{M}} = \widehat{\mathbf{X}}^{1,\top}\mathbf{X}^0 \in \mathbb{R}^{2r \times r}$.

$\boldsymbol{L}$-**Step**: We fix the spatial basis $\mathbf{X}^0$ at time t_0 and introduce the notation $\mathbf{L}(t) = \mathbf{V}(t)\mathbf{S}(t)^\top$. Then, we update the directional basis from $\mathbf{V}^0$ to $\widehat{\mathbf{V}}^1 \in \mathbb{R}^{N_\mu \times 2r}$ by solving the PDE

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

and then determining $\widehat{\mathbf{V}}^1$ as an orthonormal basis of the augmented matrix $[\mathbf{L}(t_1), \mathbf{V}^0] \in \mathbb{R}^{N_\mu \times 2r}$, e.g. by QR-decomposition. We store $\widehat{\mathbf{N}} = \widehat{\mathbf{V}}^{1,\top}\mathbf{V}^0 \in \mathbb{R}^{2r \times r}$.

$\boldsymbol{S}$-**step**: We update the coupling matrix from $\mathbf{S}^0 \in \mathbb{R}^{r \times r}$ to $\widehat{\mathbf{S}}^1 \in \mathbb{R}^{2r \times 2r}$ by solving the ODE

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

Truncation: We compute the singular value decomposition of $\widehat{\mathbf{S}}^1 = \widehat{\mathbf{P}}\boldsymbol{\Sigma}\widehat{\mathbf{Q}}^\top$ with $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_j)$. The new rank $r_1 \leq 2r$ is chosen such that for a prescribed

tolerance parameter ϑ it holds

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

We set $\mathbf{S}^1 \in \mathbb{R}^{r_1 \times r_1}$ to be the matrix containing the r_1 largest singular values. For the update of the spatial and the directional basis we introduce the matrices $\mathbf{P}^1 \in \mathbb{R}^{2r \times r_1}$ and $\mathbf{Q}^1 \in \mathbb{R}^{2r \times r_1}$ containing the first r_1 columns of $\widehat{\mathbf{P}}$ and $\widehat{\mathbf{Q}}$, respectively, and set $\mathbf{X}^1 = \widehat{\mathbf{X}}^1 \mathbf{P}^1 \in \mathbb{R}^{N_x \times r_1}$ and $\mathbf{V}^1 = \widehat{\mathbf{V}}^1 \mathbf{Q}^1 \in \mathbb{R}^{N_\mu \times r_1}$.

Altogether, this gives the time-updated low-rank approximation in the form $\mathbf{f}_r^1 = \mathbf{X}^1 \mathbf{S}^1 \mathbf{V}^{1,\top}$ at time $t_1 = t_0 + \Delta t$. Note that in the following, in an abuse of notation, we will write $\mathbf{f}$ instead of $\mathbf{f}_r$ and call this the low-rank solution of the considered problem.

5.2. DLRA scheme for multiplicative Su-Olson

In this section, the DLRA method is applied to the energy stable conservative form (11) of the Su-Olson problem to evolve $\mathbf{v}^0 = \left(v_{jk}^0\right)$ to $\mathbf{v}^1 = \left(v_{jk}^1\right)$. First note that for the derivation of the low-rank scheme we rewrite equations (11). In (11a), we put all terms containing v_{jk}^1 to the left-hand side and divide by $1 + \sigma \Delta t$. Further, we multiply (11b) with $\frac{1}{B_j^0}$. This establishes the system

$$\frac{B_j^1}{B_j^0} v_{jk}^1 = \frac{1}{1 + \sigma \Delta t} v_{jk}^0 - \frac{\Delta t}{1 + \sigma \Delta t} \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu - 1} \frac{1}{B_j^0} D_{ji}^x B_i^0 v_{i\ell}^0 A_{k\ell} \qquad (18a)$$

$$+ \frac{\Delta t}{1 + \sigma \Delta t} \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu - 1} \frac{1}{B_j^0} D_{ji}^{xx} B_i^0 v_{i\ell}^0 |A|_{k\ell} + \frac{\sqrt{2} \sigma \Delta t}{1 + \sigma \Delta t} \frac{B_j^1}{B_j^0} \delta_{k0},$$

$$\frac{B_j^1}{B_j^0} = 1 + \sigma \Delta t \frac{B_j^1}{B_j^0} \left(\sqrt{2} v_{j0}^1 - 2\right). \qquad (18b)$$

We then apply the DLRA approach to this set of equations. In a first step, we want to update $v_{jk}^0 = \sum_{m,n=1}^{r} X_{jm}^0 S_{mn}^0 V_{kn}^0$ to $\frac{B_j^1}{B_j^0} v_{jk}^* = \sum_{m,n=1}^{4r} \widehat{\widehat{X}}_{jm}^* \widehat{\widehat{S}}_{mn}^* \widehat{\widehat{V}}_{kn}^*$ for $k \neq 0$. This corresponds to an update of the advection part of (18a) without the absorption term coupling to equation (18b). We introduce the

20

notation $K_{jn}^0 = \sum_{m=1}^r X_{jm}^0 S_{mn}^0$ and solve the K-step equation

$$K_{jp}^* = \frac{1}{1+\sigma\Delta t} K_{jp}^0 - \frac{\Delta t}{1+\sigma\Delta t}\frac{1}{B_j^0}\sum_{i=1}^{N_x} D_{ji}^x B_i^0 \sum_{n=1}^r K_{in}^0 \sum_{k,\ell=0}^{N_\mu-1} V_{\ell n}^0 A_{k\ell} V_{kp}^0 \quad (19a)$$

$$+ \frac{\Delta t}{1+\sigma\Delta t}\frac{1}{B_j^0}\sum_{i=1}^{N_x} D_{ji}^{xx} B_i^0 \sum_{n=1}^r K_{in}^0 \sum_{k,\ell=0}^{N_\mu-1} V_{\ell n}^0 |A|_{k\ell} V_{kp}^0.$$

We derive the updated basis $\widehat{\mathbf{X}}^*$ of rank $2r$ from $\widehat{\mathbf{X}}^* = \mathrm{qr}\left([\mathbf{K}^*, \mathbf{X}^0]\right)$. Moreover, we perform an additional basis augmentation step according to

$$\widehat{\widehat{\mathbf{X}}}^* = \mathrm{qr}\left(\left[\widehat{\mathbf{X}}^*, \frac{1}{\mathbf{B}^0}\odot\mathbf{D}^x\left(\mathbf{B}^0\odot\mathbf{X}^0\right), \frac{1}{\mathbf{B}^0}\odot\mathbf{D}^{xx}\left(\mathbf{B}^0\odot\mathbf{X}^0\right)\right]\right). \quad (19b)$$

This basis augmentation step is crucial for the theoretical proof of energy stability of the DLRA scheme as it ensures the exactness of the corresponding projection operators. Its concrete form is motivated from the theoretical stability analysis and will become clear in the proof of energy stability in the next section. Here, the symbol $\odot$ stands for a pointwise multiplication and the vector $\frac{1}{\mathbf{B}^0}\in\mathbb{R}^{N_x}$ is defined to contain the element $\frac{1}{B_j^0}$ for each $j=1,...,N_x$.

In addition, we compute and store $\widehat{\widehat{\mathbf{M}}} = \widehat{\widehat{\mathbf{X}}}^{*,\top}\mathbf{X}^0$. Note that in this paper we perform full rank updates, leading to an increase from rank $2r$ to $4r$. Quantities of rank $2r$ are denoted with one single hat and quantities of rank $4r$ with double hats.

The L-step can be computed in parallel with the K-step. We introduce the notation $L_{mk}^0 = \sum_{n=1}^r S_{nm}^0 V_{nk}^0$ and solve

$$L_{kp}^* = \frac{1}{1+\sigma\Delta t}L_{kp}^0 - \frac{\Delta t}{1+\sigma\Delta t}\sum_{\ell=0}^{N_\mu-1}\sum_{m=1}^r A_{\ell k}L_{\ell m}^0\sum_{i=1}^{N_x} X_{im}^0 B_i^0\sum_{j=1}^{N_x} D_{ij}^x\frac{1}{B_j^0}X_{jp}^0$$

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (19c)$$

$$+ \frac{\Delta t}{1+\sigma\Delta t}\sum_{\ell=0}^{N_\mu-1}\sum_{m=1}^r |A|_{\ell k}L_{\ell m}^0\sum_{i=1}^{N_x} X_{im}^0 B_i^0\sum_{j=1}^{N_x} D_{ij}^{xx}\frac{1}{B_j^0}X_{jp}^0.$$

We derive the updated basis $\widehat{\mathbf{V}}^*$ of rank $2r$ from $\widehat{\mathbf{V}}^* = \mathrm{qr}\left([\mathbf{L}^*, \mathbf{V}^0]\right)$. Moreover, we perform an additional basis augmentation step according to

$$\widehat{\widehat{\mathbf{V}}}^* = \mathrm{qr}\left(\left[\widehat{\mathbf{V}}^*, \mathbf{A}^\top\mathbf{V}^0, |\mathbf{A}|^\top\mathbf{V}^0\right]\right). \quad (19d)$$

21

Again, this additional basis augmentation step is required for the exactness of the corresponding projection operators in the theoretical proof of energy stability. Its concrete form will become clear in the next section. In addition, we compute and store $\widehat{\widehat{\mathbf{N}}} = \widehat{\widehat{\mathbf{V}}}^{*,\top}\mathbf{V}^0$.

For the S-step we use the information from the K- and L-step, set $\widetilde{S}^0_{mn} = \sum_{j,k=1}^r \widehat{\widehat{M}}_{mj}S^0_{jk}\widehat{\widehat{N}}_{nk}$, and solve

$$
\widehat{\widehat{S}}^*_{qp} = \frac{1}{1+\sigma\Delta t}\widetilde{S}^0_{qp}
$$

$$
- \frac{\Delta t}{1+\sigma\Delta t}\sum_{j=1}^{N_x}\widehat{\widehat{X}}^*_{jq}\frac{1}{B^0_j}\sum_{i=1}^{N_x}D^x_{ji}B^0_i\sum_{m,n=1}^{4r}\widehat{\widehat{X}}^*_{im}\widetilde{S}^0_{mn}\sum_{k,\ell=0}^{N_\mu-1}\widehat{\widehat{V}}^*_{\ell n}A_{k\ell}\widehat{\widehat{V}}^*_{kp} \quad (19\mathrm{e})
$$

$$
+ \frac{\Delta t}{1+\sigma\Delta t}\sum_{j=1}^{N_x}\widehat{\widehat{X}}^*_{jq}\frac{1}{B^0_j}\sum_{i=1}^{N_x}D^{xx}_{ji}B^0_i\sum_{m,n=1}^{4r}\widehat{\widehat{X}}^*_{im}\widetilde{S}^0_{mn}\sum_{k,\ell=0}^{N_\mu-1}\widehat{\widehat{V}}^*_{\ell n}|A|_{k\ell}\widehat{\widehat{V}}^*_{kp}.
$$

In the next step, we consider the equations for $k = 0$. In this case, the equations for $\frac{B^1_j}{B^0_j}v^1_{j0}$ and $\frac{B^1_j}{B^0_j}$ couple and we solve the system

$$
\frac{B^1_j}{B^0_j}v^1_{j0} = \frac{1}{1+\sigma\Delta t}\sum_{m,n=1}^r X^0_{jm}S^0_{mn}V^0_{0n}
$$

$$
- \frac{\Delta t}{1+\sigma\Delta t}\sum_{i=1}^{N_x}\frac{1}{B^0_j}D^x_{ji}B^0_i\sum_{m,n=1}^{4r}\widehat{\widehat{X}}^*_{im}\widetilde{S}^0_{mn}\sum_{\ell=0}^{N_\mu-1}\widehat{\widehat{V}}^*_{\ell n}A_{0\ell} \quad (19\mathrm{f})
$$

$$
+ \frac{\Delta t}{1+\sigma\Delta t}\sum_{i=1}^{N_x}\frac{1}{B^0_j}D^{xx}_{ji}B^0_i\sum_{m,n=1}^{4r}\widehat{\widehat{X}}^*_{im}\widetilde{S}^0_{mn}\sum_{\ell=0}^{N_\mu-1}\widehat{\widehat{V}}^*_{\ell n}|A|_{0\ell} + \frac{\sqrt{2}\sigma\Delta t}{1+\sigma\Delta t}\frac{B^1_j}{B^0_j},
$$

$$
\frac{B^1_j}{B^0_j} = 1 + \sigma\Delta t\frac{B^1_j}{B^0_j}\left(\sqrt{2}v^1_{j0} - 2\right). \quad (19\mathrm{g})
$$

From (19f) and (19g) we can then retrieve $\mathbf{v}^1_0 = \left(v^1_{j0}\right)$ and $\mathbf{B}^1 = \left(B^1_j\right)$. We use the latter to divide out the factor $\frac{B^1_j}{B^0_j}$ in the low-rank representation of $\frac{B^1_j}{B^0_j}v^*_{jk} = \sum_{m,n=1}^{4r}\widehat{\widehat{X}}^*_{jm}\widehat{\widehat{S}}^*_{mn}\widehat{\widehat{V}}^*_{kn}$. We perform the transformation step

$$
K^{*,\mathrm{trans}}_{jp} = \frac{B^0_j}{B^1_j}K^*_{jp}. \quad (19\mathrm{h})
$$

From a QR-decomposition we obtain $\widehat{\widehat{\mathbf{X}}}^{*,\text{trans}}\widehat{\widehat{\mathbf{S}}}^{*,\text{trans}} = \text{qr}\left(\mathbf{K}^{*,\text{trans}}\right)$. Then we perform an additional basis augmentation step according to

$$\widehat{\widehat{\mathbf{X}}}^1 = \text{qr}\left(\left[\mathbf{v}_0^1, \widehat{\widehat{\mathbf{X}}}^{*,\text{trans}}\right]\right), \quad \widehat{\widehat{\mathbf{V}}}^1 = \text{qr}\left(\left[\mathbf{e}_1, \widehat{\widehat{\mathbf{V}}}^*\right]\right), \tag{19i}$$

where we add $\mathbf{v}_0^1$ to the updated spatial low-rank basis since the mass of the system is given by the zeroth order moment $\mathbf{v}_0^1$. In the directional basis, we add $\mathbf{e}_1 \in \mathbb{R}^{N_\mu}$, denoting the first unit vector in $\mathbb{R}^{N_\mu}$. This ensures mass conservation of the proposed low-rank scheme. Then, we have to adjust the coefficient matrix $\widehat{\widehat{\mathbf{S}}}^{*,\text{trans}}$ correspondingly as

$$\widehat{\widehat{\mathbf{S}}}^1 = \widehat{\widehat{\mathbf{X}}}^{1,\top}\widehat{\widehat{\mathbf{X}}}^{*,\text{trans}}\widehat{\widehat{\mathbf{S}}}^{*,\text{trans}}\widehat{\widehat{\mathbf{V}}}^*\left(\mathbf{I} - \mathbf{e}_1\mathbf{e}_1^\top\right)\widehat{\widehat{\mathbf{V}}}^1 + \widehat{\widehat{\mathbf{X}}}^{1,\top}\mathbf{v}_0^1\mathbf{e}_1^\top\widehat{\widehat{\mathbf{V}}}^1 \in \mathbb{R}^{(4r+1)\times(4r+1)}. \tag{19j}$$

In a last step, we truncate the augmented quantities $\widehat{\widehat{\mathbf{X}}}^1, \widehat{\widehat{\mathbf{S}}}^1$ and $\widehat{\widehat{\mathbf{V}}}^1$ back to a new rank r_1, using the truncation strategy described in [14] or an adjusted truncation method inspired by [7] that ensures conservation of mass and will be given in Section 5.4. Altogether, we obtain the updated low-rank factors $\mathbf{X}^1, \mathbf{S}^1$ and $\mathbf{V}^1$ such that $v_{jk}^1 = \sum_{m,n=1}^{r_1} X_{jm}^1 S_{mn}^1 V_{kn}^1$. The structure of the DLRA scheme is visualized in Figure 1.

Note that this scheme is significantly different from the one presented in [21] as the multiplicative structure leads to additional basis augmentations in (19b) and (19d) as well as a different solution of the coupled equations (19f) and (19g).

5.3. Energy stability of the proposed low-rank scheme
It can be shown that the proposed DLRA scheme preserves the energy stability of the full system. The rewriting of equations (11) into (18) as well as the basis augmentations in (19b) and (19d) differentiate the proposed DLRA method from existing schemes and are crucial for the proof.

Theorem 3. *Under the time step restriction $\Delta t \leq \Delta x$, the fully discrete DLRA scheme* (19) *is energy stable, i.e. it holds $E^1 \leq E^0$.*

Proof. We start with the internal energy $\mathbf{B}$ and multiply (19g) with $B_j^0 B_j^1$. This leads to

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

23

input
- internal energy at time t_0: B_j^0
- low-rank factors at time t_0: $X_{jm}^0, S_{mn}^0, V_{kn}^0$
- rank at time t_0: r

update bases according to (19a) and (19c)

K_{jp}^*, L_{kp}^*

augment bases with X_{jm}^0, V_{kn}^0

$\widehat{X}_{jm}^*, \widehat{V}_{kn}^*$

augment bases with $\frac{1}{B_j^0} \sum_{i=1}^{N_x} D_{ji}^x B_i^0 X_{im}^0$, $\frac{1}{B_j^0} \sum_{i=1}^{N_x} D_{ji}^{xx} B_i^0 X_{im}^0$ and $\sum_{\ell=0}^{N_\mu} A_{k\ell} V_{\ell n}^0$, $\sum_{\ell=0}^{N_\mu} |A|_{k\ell} V_{\ell n}^0$ according to (19b) and (19d)

$\widehat{\widehat{X}}_{jm}^*, \widehat{\widehat{V}}_{kn}^*$

update coefficient matrix according to (19e)

$\widehat{\widehat{S}}_{mn}^*$

update zeroth order moment and internal energy according to (19f) and (19g)

perform transformation step according to (19h)

B_j^1, v_{j0}^1

$X_{jm}^{*,\text{trans}}, S_{mn}^{*,\text{trans}}$

augment bases with $\mathbf{v}_0^1$ and $\mathbf{e}_1$ according to (19i)

$\widehat{\widehat{X}}_{jm}^1, \widehat{\widehat{V}}_{kn}^1$

adjust coefficient matrix $\widehat{\widehat{S}}_{mn}^{*,\text{trans}}$ according to (19j)

$\widehat{\widehat{S}}_{mn}^1$

truncate factors $\widehat{\widehat{X}}_{jm}^1, \widehat{\widehat{S}}_{mn}^1, \widehat{\widehat{V}}_{kn}^1$

output
- internal energy at time t_1: B_j^1
- low-rank factors at time t_1: $X_{jm}^1, S_{mn}^1, V_{kn}^1$
- rank at time t_1: r_1

Abbildung 1: Flowchart of the stable and conservative method (19).

Analogously to the proof of Theorem 2, we rewrite the product $B_j^0 B_j^1$ and sum over j, rendering the expression

$$\frac{1}{2}\sum_{j=1}^{N_x}\left(B_j^1\right)^2 = \frac{1}{2}\sum_{j=1}^{N_x}\left(B_j^0\right)^2 - \frac{1}{2}\sum_{j=1}^{N_x}\left(B_j^1 - B_j^0\right)^2 \tag{20}$$

$$+ \sigma\Delta t\sum_{j=1}^{N_x}\left(B_j^1\right)^2\left(\sqrt{2}v_{j0}^1 - 2\right).$$

In the next step, we multiply (19e) with $\widehat{\widehat{X}}_{\alpha q}^* \widehat{\widehat{V}}_{\beta p}^*$ and sum over q and p. We introduce the projection operators $P_{\alpha j}^{X*} = \sum_{q=1}^{4r}\widehat{\widehat{X}}_{\alpha q}^*\widehat{\widehat{X}}_{jq}^*$ and $P_{k\beta}^{V*} = \sum_{p=1}^{4r}\widehat{\widehat{V}}_{kp}^*\widehat{\widehat{V}}_{\beta p}^*$ as well as the notations $v_{\alpha\beta}^* := \sum_{p,q=1}^{4r}\widehat{\widehat{X}}_{\alpha q}^*\widehat{\widehat{S}}_{qp}^*\widehat{\widehat{V}}_{\beta p}^*$ and $v_{\alpha\beta}^0 := \sum_{p,q=1}^{4r}\widehat{\widehat{X}}_{\alpha q}^*\widetilde{S}_{qp}^0\widehat{\widehat{V}}_{\beta p}^*$. We obtain

$$v_{\alpha\beta}^* = \frac{1}{1+\sigma\Delta t}v_{\alpha\beta}^0 - \frac{\Delta t}{1+\sigma\Delta t}\sum_{i,j=1}^{N_x}\sum_{k,\ell=0}^{N_\mu-1}P_{\alpha j}^{X*}\frac{1}{B_j^0}D_{ji}^x B_i^0 v_{i\ell}^0 A_{k\ell}P_{k\beta}^{V*} \tag{21}$$

$$+ \frac{\Delta t}{1+\sigma\Delta t}\sum_{i,j=1}^{N_x}\sum_{k,\ell=0}^{N_\mu-1}P_{\alpha j}^{X*}\frac{1}{B_j^0}D_{ji}^{xx}B_i^0 v_{i\ell}^0|A|_{k\ell}P_{k\beta}^{V*}.$$

Further, we denote $\widehat{\widehat{v}}_{\alpha\beta}^1 := \sum_{p,q=1}^{4r+1}\widehat{\widehat{X}}_{\alpha q}^1\widehat{\widehat{S}}_{qp}^1\widehat{\widehat{V}}_{\beta p}^1$. From equation (19j), we can derive the equation

$$\frac{B_\alpha^1}{B_\alpha^0}\widehat{\widehat{v}}_{\alpha\beta}^1 = v_{\alpha\beta}^*\left(1-\delta_{\beta 0}\right) + \frac{B_\alpha^1}{B_\alpha^0}v_{\alpha 0}^1\delta_{\beta 0}.$$

We insert (21) and (19f), establishing the expression

$$
\frac{B_\alpha^1}{B_\alpha^0}\widehat{\widetilde{v}}_{\alpha\beta}^1 (1+\sigma\Delta t) = \left(v_{\alpha\beta}^0 - \Delta t \sum_{i,j=1}^{N_x}\sum_{k,\ell=0}^{N_\mu-1} P_{\alpha j}^{X^*}\frac{1}{B_j^0}D_{ji}^x B_i^0 v_{i\ell}^0 A_{k\ell}P_{k\beta}^{V^*} \right.
$$

$$
\left. + \Delta t \sum_{i,j=1}^{N_x}\sum_{k,\ell=0}^{N_\mu-1} P_{\alpha j}^{X^*}\frac{1}{B_j^0}D_{ji}^{xx} B_i^0 v_{i\ell}^0 |A|_{k\ell}P_{k\beta}^{V^*} \right)(1-\delta_{\beta 0})
$$

$$
+ \left(v_{\alpha\beta}^0 - \Delta t \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} \frac{1}{B_\alpha^0}D_{\alpha i}^x B_i^0 v_{i\ell}^0 A_{0\ell} \right.
$$

$$
\left. + \Delta t \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} \frac{1}{B_\alpha^0}D_{\alpha i}^{xx} B_i^0 v_{i\ell}^0 |A|_{0\ell} + \sqrt{2}\sigma\Delta t\frac{B_\alpha^1}{B_\alpha^0} \right)\delta_{\beta 0}.
$$

We now use the fact that we have augmented the spatial basis according to (19b) and (19d). This allows us to write any function $h_i^0 \in \mathrm{span}\,(X_i^0)$ and $\tilde{h}_\ell^0 \in \mathrm{span}\,(V_\ell^0)$ as

$$
\sum_{i,j=1}^{N_x} P_{\alpha j}^{X^*}\frac{1}{B_j^0}D_{ji}^x B_i^0 h_i^0 = \frac{1}{B_\alpha^0}\sum_{i=1}^{N_x} D_{\alpha i}^x B_i^0 h_i^0 \ \ \text{and} \ \ \sum_{k,\ell=0}^{N_\mu-1} \tilde{h}_\ell^0 A_{k\ell}P_{k\beta}^{V^*} = \sum_{\ell=0}^{N_\mu-1}\tilde{h}_\ell^0 A_{\beta\ell},
$$

$$
\sum_{i,j=1}^{N_x} P_{\alpha j}^{X^*}\frac{1}{B_j^0}D_{ji}^{xx} B_i^0 h_i^0 = \frac{1}{B_\alpha^0}\sum_{i=1}^{N_x} D_{\alpha i}^{xx} B_i^0 h_i^0 \ \ \text{and} \ \ \sum_{k,\ell=0}^{N_\mu-1} \tilde{h}_\ell^0 |A|_{k\ell} P_{k\beta}^{V^*} = \sum_{\ell=0}^{N_\mu-1}\tilde{h}_\ell^0 |A|_{\beta\ell}.
$$

The basis augmentations as well as the properties of the projection operators enable us to obtain a representation of the form

$$
\frac{B_\alpha^1}{B_\alpha^0}\widehat{\widetilde{v}}_{\alpha\beta}^1 (1+\sigma\Delta t) = F(1-\delta_{\beta 0}) + F\delta_{\beta 0} + \sqrt{2}\sigma\Delta t\frac{B_\alpha^1}{B_\alpha^0}\delta_{\beta 0}
$$

with

$$
F = v_{\alpha\beta}^0 - \Delta t \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} \frac{1}{B_\alpha^0}D_{\alpha i}^x B_i^0 v_{i\ell}^0 A_{\beta\ell} + \Delta t \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} \frac{1}{B_\alpha^0}D_{\alpha i}^{xx} B_i^0 v_{i\ell}^0 |A|_{\beta\ell}.
$$

We use that on the right-hand side $F\delta_{\beta 0}$ cancels out. This yields the equation

$$\frac{B_\alpha^1}{B_\alpha^0}\widehat{\widehat{v}}_{\alpha\beta}^1\left(1+\sigma\Delta t\right) = v_{\alpha\beta}^0 - \Delta t \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1}\frac{1}{B_\alpha^0}D_{\alpha i}^x B_i^0 v_{i\ell}^0 A_{\beta\ell}$$

$$+ \Delta t \sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1}\frac{1}{B_\alpha^0}D_{\alpha i}^{xx} B_i^0 v_{i\ell}^0 |A|_{\beta\ell} + \sqrt{2}\sigma\Delta t\frac{B_\alpha^1}{B_\alpha^0}\delta_{\beta 0}.$$

In the next step, we multiply with $B_\alpha^1 B_\alpha^0 \widehat{\widehat{v}}_{\alpha\beta}^1$, sum over α and β, rearrange, and introduce the notation $\widehat{\widehat{u}}_{\alpha\beta}^1 = B_\alpha^1 \widehat{\widehat{v}}_{\alpha\beta}^1$. This leads to

$$\sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1}\left(\widehat{\widehat{u}}_{\alpha\beta}^1\right)^2 = \sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1} u_{\alpha\beta}^0\widehat{\widehat{u}}_{\alpha\beta}^1 - \Delta t \sum_{i,\alpha=1}^{N_x}\sum_{\ell,\beta=0}^{N_\mu-1}\widehat{\widehat{u}}_{\alpha\beta}^1 D_{\alpha i}^x u_{i\ell}^0 A_{\beta\ell}$$

$$+ \Delta t \sum_{i,\alpha=1}^{N_x}\sum_{\ell,\beta=0}^{N_\mu-1}\widehat{\widehat{u}}_{\alpha\beta}^1 D_{\alpha i}^{xx} v_{i\ell}^0 |A|_{\beta\ell}$$

$$+ \sigma\Delta t \sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1}\left(B_\alpha^1\right)^2\widehat{\widehat{v}}_{\alpha\beta}^1\left(\sqrt{2}\delta_{\beta 0} - \widehat{\widehat{v}}_{\alpha\beta}^1\right).$$

Inserting the relation

$$\sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1} u_{\alpha\beta}^0\widehat{\widehat{u}}_{\alpha\beta}^1 = \sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1}\left(\frac{1}{2}\left(\widehat{\widehat{u}}_{\alpha\beta}^1\right)^2 + \frac{1}{2}\left(u_{\alpha\beta}^0\right)^2 - \frac{1}{2}\left(\widehat{\widehat{u}}_{jk}^1 - u_{jk}^0\right)^2\right)$$

establishes the expression

$$\frac{1}{2}\sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1}\left(\widehat{\widehat{u}}_{\alpha\beta}^1\right)^2 = \frac{1}{2}\sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1}\left(u_{\alpha\beta}^0\right)^2 - \frac{1}{2}\sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1}\left(\widehat{\widehat{u}}_{jk}^1 - u_{jk}^0\right)^2$$

$$- \Delta t \sum_{i,\alpha=1}^{N_x}\sum_{\ell,\beta=0}^{N_\mu-1}\widehat{\widehat{u}}_{\alpha\beta}^1 D_{\alpha i}^x u_{i\ell}^0 A_{\beta\ell}$$

$$+ \Delta t \sum_{i,\alpha=1}^{N_x}\sum_{\ell,\beta=0}^{N_\mu-1}\widehat{\widehat{u}}_{\alpha\beta}^1 D_{\alpha i}^{xx} v_{i\ell}^0 |A|_{\beta\ell}$$

$$+ \sigma\Delta t \sum_{\alpha=1}^{N_x}\sum_{\beta=0}^{N_\mu-1}\left(B_\alpha^1\right)^2\widehat{\widehat{v}}_{\alpha\beta}^1\left(\sqrt{2}\delta_{\beta 0} - \widehat{\widehat{v}}_{\alpha\beta}^1\right).$$

27

This is the same expression as in equation (14) in the proof of Theorem 2. We apply the same estimates as in the proof of Theorem 2 and add the resulting equation with equation (20). Analogously to the proof of Theorem 2 and due to the fact that the truncation step does not alter the zeroth order moment, we obtain energy stability of the multiplicative DLRA scheme under the time step restriction $\Delta t \leq \Delta x$. $\qquad\square$

5.4. Mass conservation

In addition, the DLRA scheme (19) can be shown to be mass conservative when using a suitable truncation strategy. We follow the ideas in [7, 29, 30] and adjust the truncation step such that it conserves the zeroth order moment. Different from [7] and as explained in [30], we do not need to adjust the L-step equation due to the usage of the augmented BUG integrator from [14]. Starting from the augmented quantities $\widehat{\widehat{\mathbf{X}}}^1$, $\widehat{\widehat{\mathbf{S}}}^1$ and $\widehat{\widehat{\mathbf{V}}}^1$, the conservative truncation strategy then works as follows:

1. We set $\widehat{\widehat{\mathbf{K}}}^1 = \widehat{\widehat{\mathbf{X}}}^1 \widehat{\widehat{\mathbf{S}}}^1$ and split it into two parts $\widehat{\widehat{\mathbf{K}}}^1 = [\widehat{\widehat{\mathbf{K}}}^{1,\mathrm{cons}}, \widehat{\widehat{\mathbf{K}}}^{1,\mathrm{rem}}]$, where $\widehat{\widehat{\mathbf{K}}}^{1,\mathrm{cons}}$ corresponds to the first and $\widehat{\widehat{\mathbf{K}}}^{1,\mathrm{rem}}$ to the remaining columns of $\widehat{\widehat{\mathbf{K}}}^1$. Analogously, we split $\widehat{\widehat{\mathbf{V}}}^1$ into $\widehat{\widehat{\mathbf{V}}}^1 = [\widehat{\widehat{\mathbf{V}}}^{1,\mathrm{cons}}, \widehat{\widehat{\mathbf{V}}}^{1,\mathrm{rem}}]$, where $\widehat{\widehat{\mathbf{V}}}^{1,\mathrm{cons}}$ corresponds to the first and $\widehat{\widehat{\mathbf{V}}}^{1,\mathrm{rem}}$ to the remaining columns of $\widehat{\widehat{\mathbf{V}}}^1$.

2. We compute $\widehat{\widehat{\mathbf{X}}}^{1,\mathrm{cons}} = \widehat{\widehat{\mathbf{K}}}^{1,\mathrm{cons}}/\|\widehat{\widehat{\mathbf{K}}}^{1,\mathrm{cons}}\|$ and $\widehat{\widehat{\mathbf{S}}}^{1,\mathrm{cons}} = \|\widehat{\widehat{\mathbf{K}}}^{1,\mathrm{cons}}\|$.

3. We perform a QR-decomposition of $\widehat{\widehat{\mathbf{K}}}^{1,\mathrm{rem}}$ to get $\widehat{\widehat{\mathbf{K}}}^{1,\mathrm{rem}} = \widehat{\widehat{\mathbf{X}}}^{1,\mathrm{rem}}\widehat{\widehat{\mathbf{S}}}^{1,\mathrm{rem}}$.

4. We compute the singular value decomposition of $\widehat{\widehat{\mathbf{S}}}^{1,\mathrm{rem}} = \widehat{\widehat{\mathbf{P}}}\mathbf{\Sigma}\widehat{\widehat{\mathbf{Q}}}^{\top}$ with $\mathbf{\Sigma} = \mathrm{diag}(\sigma_j)$. The new rank $r_1 \leq 4r$ is chosen such that for a prescribed tolerance parameter ϑ it holds

$$\left(\sum_{j=r_1+1}^{4r} \sigma_j^2\right)^{1/2} \leq \vartheta.$$

We set $\mathbf{S}^{1,\mathrm{rem}} \in \mathbb{R}^{r_1 \times r_1}$ to be the matrix containing the r_1 largest singular values. For the update of the spatial and the directional basis we introduce the matrices $\widehat{\widehat{\mathbf{P}}}^{\mathrm{rem}} \in \mathbb{R}^{4r \times r_1}$ and $\widehat{\widehat{\mathbf{Q}}}^{\mathrm{rem}} \in \mathbb{R}^{4r \times r_1}$ containing the first r_1 columns of $\widehat{\widehat{\mathbf{P}}}$ and $\widehat{\widehat{\mathbf{Q}}}$, respectively, and set $\mathbf{X}^{1,\mathrm{rem}} = \widehat{\widehat{\mathbf{X}}}^{1,\mathrm{rem}}\widehat{\widehat{\mathbf{P}}}^{\mathrm{rem}} \in \mathbb{R}^{N_x \times r_1}$ and $\mathbf{V}^{1,\mathrm{rem}} = \widehat{\widehat{\mathbf{V}}}^{1,\mathrm{rem}}\widehat{\widehat{\mathbf{Q}}}^{\mathrm{rem}} \in \mathbb{R}^{N_\mu \times r_1}$.

5. We set $\widetilde{\mathbf{X}}^1 = [\widehat{\widehat{\mathbf{X}}}^{1,\text{cons}}, \mathbf{X}^{1,\text{rem}}]$ and $\widetilde{\mathbf{V}}^1 = [\mathbf{e}_1, \mathbf{V}^{1,\text{rem}}]$ and perform a QR-decomposition to obtain $\widetilde{\mathbf{X}}^1 = \mathbf{X}^1\mathbf{R}^1$ and $\widetilde{\mathbf{V}}^1 = \mathbf{V}^1\mathbf{R}^2$, respectively.

6. We compute

$$\mathbf{S}^1 = \mathbf{R}^1 \begin{bmatrix} \widehat{\widehat{\mathbf{S}}}^{1,\text{cons}} & 0 \\ 0 & \mathbf{S}^{1,\text{rem}} \end{bmatrix} \mathbf{R}^{2,\top}.$$

This leads to the updated solution $\mathbf{v}^1 = \mathbf{X}^1\mathbf{S}^1\mathbf{V}^{1,\top}$ after one time step at time $t_1 = t_0 + \Delta t$.

In order to show local mass conservation for the proposed DLRA scheme, we translate the macroscopic quantities given in Definition 1 to the fully discretized setting.

Definition 4 (Fully discretized macroscopic quantities)**.** *The mass of the fully discretized multiplicative Su-Olson problem at time t_0 is defined as*

$$\rho_j^0 := \sqrt{2}B_j^0 v_{j0}^0 + B_j^0.$$

The momentum is given as

$$u_j^0 := \sqrt{2}B_j^0 \sum_{\ell=0}^{N_\mu-1} v_{j\ell}^0 A_{0\ell}.$$

For $t_1 = t_0 + \Delta t$ the definitions shall hold analogously.

We can then show that the DLRA algorithm together with the conservative truncation strategy fulfills the following local conservation law.

Theorem 4. *The DLRA scheme* (19) *together with the conservative truncation strategy is locally mass conservative, i.e. it fulfills the local conservation law*

$$\frac{1}{\Delta t}\left(\sqrt{2}B_j^1\Phi_j^1 + B_j^1 - \left(\sqrt{2}B_j^0\Phi_j^0 + B_j^0\right)\right) \tag{22}$$

$$= -\sqrt{2}\sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} D_{ji}^x B_i^0 v_{i\ell}^0 A_{0\ell} + \sqrt{2}\sum_{i=1}^{N_x}\sum_{\ell=0}^{N_\mu-1} D_{ji}^{xx} B_i^0 v_{i\ell}^0 |A|_{0\ell},$$

where $\Phi_j^0 = \sum_{m,n=1}^r X_{jm}^0 S_{mn}^0 V_{0n}^0$, $\Phi_j^1 = \sum_{m,n=1}^{r_1} X_{jm}^1 S_{mn}^1 V_{0n}^1$ as well as $v_{jk}^0 = \sum_{m,n=1}^r X_{jm}^0 S_{mn}^0 V_{kn}^0$. This is a discretization of the continuous local conservation law given in (5).

Proof. The conservative truncation strategy is designed to leave the zeroth order moment unchanged, i.e. it holds $\sum_{m,n=1}^{4r} \widehat{\widehat{X}}_{jm}^1 \widehat{\widehat{S}}_{mn}^1 \widehat{\widehat{V}}_{0n}^1 = v_{j0}^1$. In addition, we know from the basis augmentation (19i) and the adjustment step (19j) that it holds $\sum_{m,n=1}^{4r} \widehat{\widehat{X}}_{jm}^1 \widehat{\widehat{S}}_{mn}^1 \widehat{\widehat{V}}_{0m}^1 = \sum_{m,n=1}^{r_1} X_{jm}^1 S_{mn}^1 V_{0n}^1$. Combining both equalities, we obtain

$$\Phi_j^1 = \sum_{m,n=1}^{r_1} X_{jm}^1 S_{mn}^1 V_{0n}^1 = \sum_{m,n=1}^{4r} \widehat{\widehat{X}}_{jm}^1 \widehat{\widehat{S}}_{mn}^1 \widehat{\widehat{V}}_{0n}^1 = v_{j0}^1.$$

We insert this relation into the coupled equations (19f) and (19g). We multiply (19f) with $\sqrt{2}\,(1 + \sigma\Delta t)$, rearrange it, and multiply both equations with B_j^0. This leads to

$$\sqrt{2}B_j^1\Phi_j^1 = \sqrt{2}B_j^0\Phi_j^0 - \sqrt{2}\Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} D_{ji}^x B_i^0 \sum_{m,n=1}^{4r} \widehat{\widehat{X}}_{im}^* \widetilde{S}_{mn}^0 \widehat{\widehat{V}}_{\ell n}^* A_{0\ell}$$

$$+ \sqrt{2}\Delta t \sum_{i=1}^{N_x} \sum_{\ell=0}^{N_\mu-1} D_{ji}^{xx} B_i^0 \sum_{m,n=1}^{4r} \widehat{\widehat{X}}_{im}^* \widetilde{S}_{mn}^0 \widehat{\widehat{V}}_{\ell n}^* |A|_{0\ell} \qquad (23a)$$

$$+ \sigma\Delta t B_j^1 \left(2 - \sqrt{2}\Phi_j^1\right),$$

$$B_j^1 = B_j^0 + \sigma\Delta t B_j^1 \left(\sqrt{2}\Phi_j^1 - 2\right). \qquad (23b)$$

Due to the basis augmentations with $\mathbf{X}^0$ and $\mathbf{V}^0$ from the augmented BUG integrator it can be concluded that

$$\sum_{m,n=1}^{4r} \widehat{\widehat{X}}_{im}^* \widetilde{S}_{mn}^0 \widehat{\widehat{V}}_{\ell n}^* = \sum_{m,n=1}^{r} X_{im}^0 S_{mn}^0 V_{\ell n}^0 = v_{i\ell}^0.$$

We insert this relation into (23a), add equations (23a) and (23b), and rearrange the result. This leads to the local conservation law (22), ensuring the local conservation of mass. $\qquad\square$

6. Numerical results

In this section, we compare the solution of the DLRA scheme (19) to the solution of the full equations (18) to underline the efficiency and accuracy of the proposed method. We give different test examples in 1D and 2D that validate our theoretical results.

6.1. 1D plane source

We first examine the 1D plane source test case which is a common test example for thermal radiative transfer [31, 32, 3]. We consider the spatial domain $D = [-10, 10]$ and the directional domain $[-1, 1]$. The initial condition is chosen to be the cutoff Gaussian

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

with constant deviation $\sigma_{\mathrm{IC}} = 0.03$. The traveling particles are initially centered around $x = 1$ and move into all directions $\mu \in [-1, 1]$. The initial internal energy is set to $B^0 = 1$ and the opacity to the constant value $\sigma = 1$. For the low-rank computations we use an initial rank of $r = 10$. The total mass m^n at time t_n is defined as $m^n = \Delta x \sum_{j=1}^{N_x} \left(\sqrt{2}B_j^n v_{j0}^n + B_j^n\right)$. As computational parameters we use $N_x = 1000$ cells in the spatial domain and $N_\mu = 500$ moments for the approximation in the directional domain. The time step size is determined by $\Delta t = \mathrm{CFL} \cdot \Delta x$ with $\mathrm{CFL} = 0.99$, according to the corresponding CFL condition.

For the theoretical proof of energy stability of the multiplicative DLRA scheme (19) the basis augmentations to rank $4r+1$ performed in (19b) and (19d) are essential. However, in practical implementations it can be observed that the standard basis augmentations to rank $2r+1$ provide similar solutions while being significantly faster.For this reason, we propose to leave out the basis augmentations presented in (19b) and (19d) in practical applications.

We compare the solution of the full system (full) with the solution of the reduced DLRA scheme with rank $2r + 1$ (DLRA) and the solution of the augmented DLRA scheme with rank $4r + 1$ (DLRA BasisAug). Figure 2 displays the solution $f(x, \mu)$ computed with the three different solvers. In Figure 3 the numerical results for the scalar flux $\Phi = \frac{1}{\sqrt{2}}\langle f \rangle_\mu$ and the dimensionless temperature $T = \sqrt[4]{B}$ are shown. All quantities are captured well by both DLRA schemes. In addition, Figure 3 shows the evolution of the rank r. For a chosen tolerance parameter of $\vartheta = 10^{-1}\|\Sigma\|_F$ the rank of both DLRA schemes increases up to $r = 23$ before it significantly reduces again. Note that the ranks of the reduced as well as the augmented DLRA scheme show a similar behavior as they are displayed after the corresponding truncation step. The relative mass error $\frac{|m^0 - m^n|}{|m^0|}$ is of order $\mathcal{O}(10^{-13})$, i.e. the DLRA
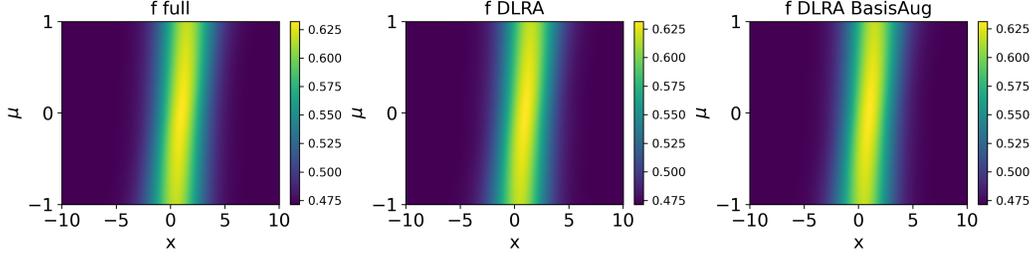
31

Abbildung 2: Numerical results for the solution $f(x, \mu)$ of the 1D plane source problem at time $t_{\text{end}} = 8$ computed with the multiplicative full solver (left), the reduced multiplicative DLRA scheme (middle), and the augmented multiplicative DLRA scheme (right).

scheme is mass conservative up to machine precision. These results confirm our theoretical findings. The memory demand reduce from $\mathcal{O}\left(N_x N_\mu\right)$ for the full solver to $\mathcal{O}\left(r \max\left(N_x, N_\mu\right)\right)$ for the DLRA schemes.

6.2. 1D external source

In a second test example an external source term $Q(x)$ is added to the conservative form (4) of the Su-Olson system, leading to the equations

$$\partial_t g(t, x, \mu) = - \frac{\mu}{B(t, x)}\partial_x \left(B(t, x)g(t, x, \mu)\right) + \sigma\left(1 - g(t, x, \mu)\right)$$
$$- \frac{g(t, x, \mu)}{B(t, x)}\partial_t B(t, x) + \frac{Q(x)}{B(t, x)},$$
$$\partial_t B(t, x) = \sigma B(t, x)\left(\langle g(t, x, \mu)\rangle_\mu - 2\right).$$

The source term again generates radiation that moves through and interacts with the background material which in turn heats up and itself emits particles. The resulting travelling temperature wave is called a *Marshak wave* [33]. Again, we compare the solution of the full equations (full) with the solution of the reduced DLRA scheme with rank $2r + 1$ (DLRA) and the solution of the augmented DLRA scheme with rank $4r + 1$ (DLRA BasisAug). All schemes are adjusted to take the additional source term into account. For our numerical example we choose the function $Q(x) = \chi_{[-0.5, 0.5]}(x)/a$ with $a = \frac{4\sigma_{\text{SB}}}{c}$ being the radiation constant. The initial internal energy is set to $B^0 = 50$. All other initial settings and computational parameters remain unchanged from the previous test example given in Section 6.1.

Figures 4 and 5 display the numerical results for the solution $f(x, \mu)$, for the scalar flux $\Phi = \frac{1}{\sqrt{2}}\langle f\rangle_\mu$, and for the dimensionless temperature $T = \sqrt[4]{B}$,
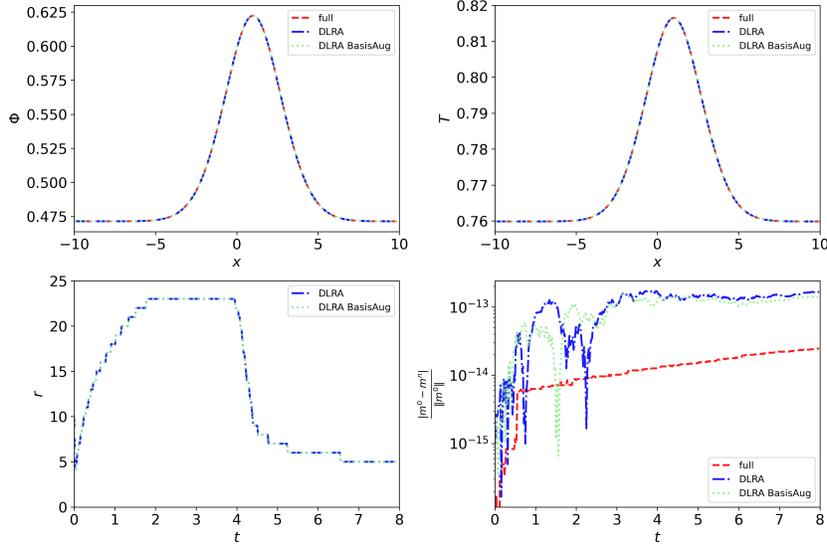
32

Abbildung 3: Top row: Numerical results for the scalar flux Φ (left) and temperature T (right) of the 1D plane source problem at time $t_{\mathrm{end}} = 8$ computed with the multiplicative full solver, the reduced multiplicative DLRA scheme, and the augmented multiplicative DLRA scheme. Bottom row: Evolution of the rank in time for both multiplicative DLRA methods (left) and evolution of the relative mass error compared for all methods (right).

computed with the full and both DLRA solvers. We again observe that both DLRA schemes capture the solution of the full system. The rank r increases up to a value of $r = 23$ for a chosen tolerance parameter of $\vartheta = 10^{-3}\|\mathbf{\Sigma}\|_F$. Due to the additional source term, there is no conservation of mass in this test example.

6.3. 2D beam

Computational benefits of the DLRA method can especially be observed in higher-dimensional settings. We extend the Su-Olson problem (1) to a 2D spatial and a 2D angular setting. The corresponding equations is given by

$$\partial_t f(t, \mathbf{x}, \mathbf{\Omega}) + \mathbf{\Omega} \cdot \nabla_{\mathbf{x}} f(t, \mathbf{x}, \mathbf{\Omega}) = \sigma(B(t, \mathbf{x}) - f(t, \mathbf{x}, \mathbf{\Omega})),$$
$$\partial_t B(t, \mathbf{x}) = \sigma(\langle f(t, \mathbf{x}, \cdot)\rangle_{\mathbf{\Omega}} - B(t, \mathbf{x})),$$
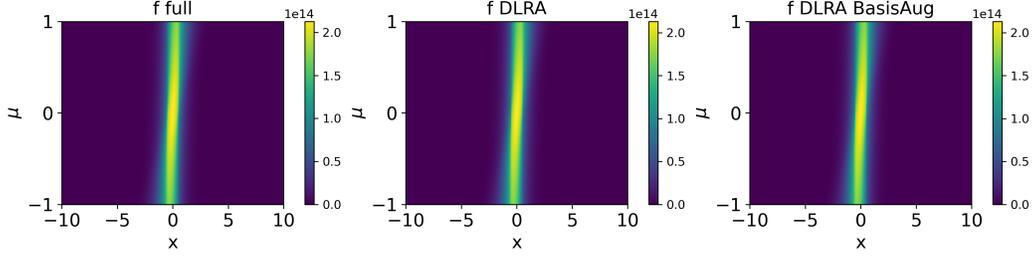
33

Abbildung 4: Numerical results for the solution $f(x, \mu)$ of the 1D external source problem at time $t_{\text{end}} = 3.16$ computed with the multiplicative full solver (left), the reduced multiplicative DLRA scheme (middle) and the augmented multiplicative DLRA scheme (right).
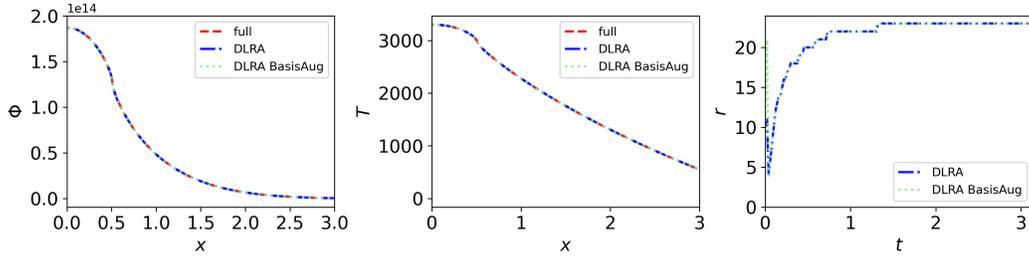

Abbildung 5: Numerical results for the scalar flux Φ (left) and temperature T (middle) of the 1D external source problem at time $t_{\text{end}} = 3.16$ computed with the multiplicative full solver, the reduced multiplicative DLRA scheme, and the augmented multiplicative DLRA scheme. Right: Evolution of the rank in time for both multiplicative DLRA methods.

where $\mathbf{x} = (x, y)$ and $\mathbf{\Omega} = (\Omega_x, \Omega_y, \Omega_z)$. For the numerical computations let $\mathbf{x} \in [-1, 1] \times [-1, 1]$ and $\mathbf{\Omega} \in \mathcal{S}^2$. The initial condition is chosen to be

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

with $\Omega^\star = \frac{1}{\sqrt{2}}$, $\sigma_{\mathbf{x}} = \sigma_{\mathbf{\Omega}} = 0.1$. The initial internal energy is set to $B^0 = 1$ and the opacity to the constant value $\sigma = 0.5$. The low-rank computations are performed with an initial rank of $r = 100$. The total mass m^n at time t_n is defined as $m^n = \Delta x \Delta y \sum_{j=1}^{N_x \cdot N_y} \left(B_j^n v_{j0}^n + B_j^n\right)$. As computational parameters we use $N_x = N_y = 500$ cells in each spatial dimension. For the 2D angular discretization we consider a spherical harmonics expansion with a polynomial degree of $N_{\mathbf{\Omega}} = 29$, corresponding to 900 expansion coefficients in angle. The time step size is determined by $\Delta t = \text{CFL} \cdot \Delta x$ with CFL $= 0.7$.

34

In Figure 6 the numerical results for the scalar flux $\Phi = \int_{\mathcal{S}^2} f(t, \mathbf{x}, \mathbf{\Omega}) \, \mathrm{d}\mathbf{\Omega}$ and the temperature $T = 4\pi\sqrt{2}\sqrt[4]{B}$ at time $t = 0.5$, computed with full multiplicative solver and the reduced multiplicative DLRA scheme, are displayed. Note that the extension of the full as well as the DLRA scheme to 2D is straightforward. It can be observed that the solution of the full multiplicative solver is captured accurately by the multiplicative DLRA scheme. Figure 7 presents the evolution of the rank r of the reduced multiplicative DLRA scheme. For a chosen tolerance parameter of $\vartheta = 5 \cdot 10^{-4} \|\mathbf{\Sigma}\|_F$ the rank increases continuously over time but does not reach its allowed maximal value of $r_{\max} = 100$. Both the full and the DLRA scheme are mass conservative as the relative mass errors are of order $\mathcal{O}\left(10^{-14}\right)$ and $\mathcal{O}\left(10^{-11}\right)$, respectively. For this setup, the computational benefit of the DLRA method is significant. The scheme is implemented in Julia v1.7 and performed on a MacBook Pro with M1 chip, resulting in a decrease of run time by a factor of approximately 35.5 from 61385 seconds to 1731 seconds. Due to the efficient implementation of the coupling of the equations in the multiplicative setting, the proposed multiplicative DLRA scheme is approximately 1.5 times faster than the non-multiplicative DLRA scheme presented in [21].

7. Conclusion and outlook

We have presented a DLRA discretization for the multiplicative Su-Olson problem that is energy stable and mass conservative. To achieve both of these features, additional basis augmentations in the augmented BUG integrator combined with an adjusted truncation step are performed. This enables us to give a mathematically rigorous stability analysis. Numerical test examples confirm the theoretical properties and validate the accuracy and computational advantages of the DLRA scheme. However, the extension of the considered stability analysis from a linear to a non-linear problem, for example the isothermal Boltzmann-BGK equation treated in [9], poses additional challenges as the general theoretical setting is significantly more difficult. For example, to our knowledge, the theoretical framework we have used here is not available in the non-linear case. Nevertheless, the analysis performed in this paper provides valuable insights into the choice of a suitable space discretization and stabilization when considering a multiplicative splitting of the distribution function, also for more complicated, potentially non-linear problems. In this sense, the theoretical considerations on the multiplicative splitting approach provided in this work can be extremely useful
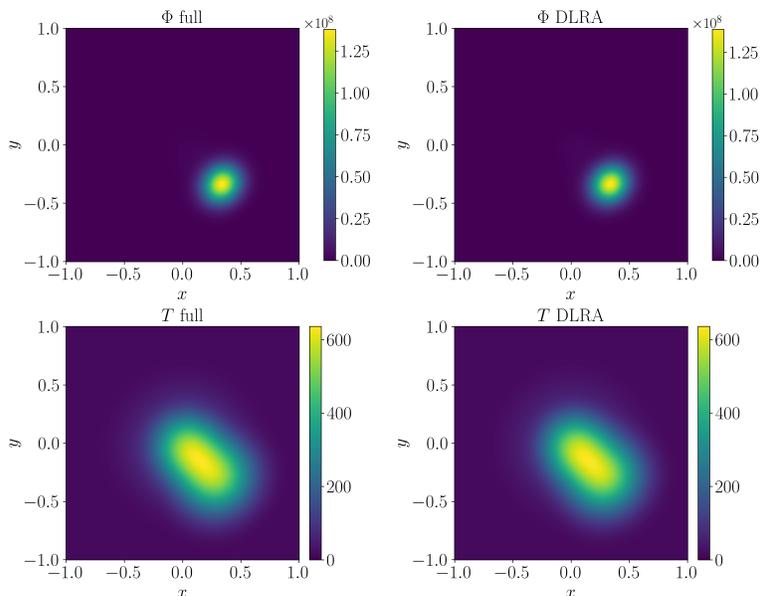
Abbildung 6: Numerical results for the scalar flux Φ (top row) and temperature T (bottom row) of the 2D beam problem at time $t_{\text{end}} = 0.5$ computed with the multiplicative full solver (left) and the reduced multiplicative DLRA scheme (right).

for the construction of stable and efficient DLRA schemes for more general equations, which we consider future work.

Acknowledgements

Declarations

The authors have nothing to declare.

Literatur

[1] O. Koch, C. Lubich, Dynamical low-rank approximation, SIAM Journal on Matrix Analysis and Applications 29 (2) (2007) 434–454.
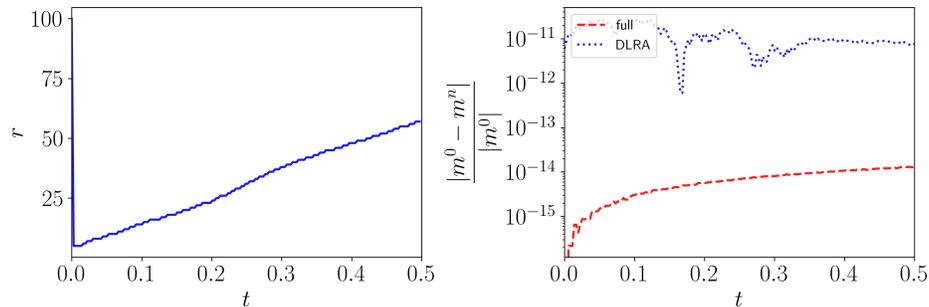
Abbildung 7: Evolution of the rank in time for the 2D beam problem for the reduced multiplicative DLRA method (left) and evolution of the relative mass error in time compared for both methods (right).

[2] C. Patwardhan, M. Frank, J. Kusch, Asymptotic-preserving and energy stable dynamical low-rank approximation for thermal radiative transfer equations, Multiscale Modeling & Simulation 23 (1) (2025) 278–312.

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

[4] P. Yin, E. Endeve, C. D. Hauck, S. R. Schnake, Towards dynamical low-rank approximation for neutrino kinetic equations. Part I: Analysis of an idealized relaxation model, Mathematics of Computation (2024).

[5] J. Kusch, P. Stammer, A robust collision source method for rank adaptive dynamical low-rank approximation in radiation therapy, ESAIM: M2AN 57 (2) (2023) 865–891.

[6] L. Einkemmer, C. Lubich, A quasi-conservative dynamical low-rank algorithm for the Vlasov equation, SIAM Journal on Scientific Computing 41 (5) (2019) B1061–B1081.

[7] L. Einkemmer, A. Ostermann, C. Scalone, A robust and conservative dynamical low-rank algorithm, Journal of Computational Physics 484 (2023) 112060.

[8] M. Prugger, L. Einkemmer, C. F. Lopez, A dynamical low-rank approach to solve the chemical master equation for biological reaction networks, Journal of Computational Physics 489 (2023) 112250.

[9] L. Einkemmer, J. Hu, 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 (5) (2021) B1057–B1080.

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

[11] E. Kieri, C. Lubich, H. Walach, Discretized dynamical low-rank approximation in the presence of small singular values, SIAM Journal on Numerical Analysis 54 (2) (2016) 1020–1038.

[12] C. Lubich, I. V. Oseledets, A projector-splitting integrator for dynamical low-rank approximation, BIT Numerical Mathematics 54 (1) (2014) 171–188.

[13] G. Ceruti, C. Lubich, An unconventional robust integrator for dynamical low-rank approximation, BIT Numerical Mathematics 62 (1) (2022) 23–44.

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

[15] G. Ceruti, J. Kusch, C. Lubich, A parallel rank-adaptive integrator for dynamical low-rank approximation, SIAM Journal on Scientific Computing 46 (3) (2024) B205–B228.

[16] G. I. Bell, S. Glasstone, Nuclear Reactor Theory, Van Nostrand Reinhold Company, New York, 1970.

[17] K. M. Case, P. F. Zweifel, Linear Transport Theory, Addison-Wesley Series in Nuclear Engineering, Addison-Wesley, Reading, MA, 1967.

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

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

[20] S. Dargaville, A. Buchan, R. Smedley-Stevenson, P. Smith, C. Pain, Angular adaptivity with spherical harmonics for Boltzmann transport, Journal of Computational Physics 397 (2019) 108846.

[21] L. Baumann, L. Einkemmer, C. Klingenberg, J. Kusch, Energy stable and conservative dynamical low-rank approximation for the Su-Olson problem, SIAM Journal on Scientific Computing 46 (2) (2024) B137–B158.

[22] L. Baumann, L. Einkemmer, C. Klingenberg, J. Kusch, A stable multiplicative dynamical low-rank discretization for the linear Boltzmann-BGK equation, arXiv preprint arXiv:2411.06844 (2024).

[23] G. C. Pomraning, The Equations of Radiation Hydrodynamics, Vol. 54 of International series of monographs in natural philosophy, Pergamon Press, Oxford, 1973.

[24] J. R. Howell, M. P. Mengüc, K. Daun, R. Siegel, Thermal Radiation Heat Transfer, seventh Edition, CRC Press, Boca Raton, 2020. doi:https://doi.org/10.1201/9780429327308.

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

[26] B. Su, G. L. Olson, An analytical benchmark for non-equilibrium radiative transfer in an isotropically scattering medium, Annals of Nuclear Energy 24 (13) (1997) 1035–1055.

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

[28] R. J. LeVeque, Finite Difference Methods for Ordinary and Partial Differential Equations: Steady-State and Time-Dependent Problems, SIAM, Philadelphia, 2007. doi:https://doi.org/10.1137/1.9780898717839.

[29] W. Guo, J.-M. Qiu, A conservative low rank tensor method for the Vlasov dynamics, SIAM Journal on Scientific Computing 46 (1) (2024) A232–A263.

[30] L. Einkemmer, J. Kusch, S. Schotthöfer, Conservation properties of the augmented basis update & Galerkin integrator for kinetic problems, ar-Xiv preprint arXiv:2311.06399 (2023).

[31] B. Ganapol, R. Baker, J. Dahl, Homogeneous infinite media time-dependent analytical benchmarks, Los Alamos National Laboratory (2001).

[32] B. D. Ganapol, Analytical benchmarks for nuclear engineering applications. Case studies in neutron transport theory, Nuclear energy agency, Organisation for economic co-operation and development (2008).

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