MATHEMATICS OF COMPUTATION
Volume 94, Number 353, May 2025, Pages 1199–1233
https://doi.org/10.1090/mcom/3997
Article electronically published on August 8, 2024

TOWARDS DYNAMICAL LOW-RANK APPROXIMATION FOR NEUTRINO KINETIC EQUATIONS. PART I: ANALYSIS OF AN IDEALIZED RELAXATION MODEL

PEIMENG YIN, EIRIK ENDEVE, CORY D. HAUCK, AND STEFAN R. SCHNAKE

Abstract. Dynamical low-rank approximation (DLRA) is an emerging tool for reducing computational costs and provides memory savings when solving high-dimensional problems. In this work, we propose and analyze a semi-implicit dynamical low-rank discontinuous Galerkin (DLR-DG) method for the space homogeneous kinetic equation with a relaxation operator, modeling the emission and absorption of particles by a background medium. Both DLRA and the discontinuous Galerkin (DG) scheme can be formulated as Galerkin equations. To ensure their consistency, a weighted DLRA is introduced so that the resulting DLR-DG solution is a solution to the fully discrete DG scheme in a subspace of the standard DG solution space. Similar to the standard DG method, we show that the proposed DLR-DG method is well-posed. We also identify conditions such that the DLR-DG solution converges to the equilibrium. Numerical results are presented to demonstrate the theoretical findings.

1. Introduction

In this paper, we consider high-order approximation methods for solving kinetic equations using low-dimensional surrogates that capture their essential features. These methods have been demonstrated to be computationally cheaper for many high-dimensional dynamical systems (see, e.g., [20]), and dynamical low-rank approximation (DLRA) is one well-known method used for this purpose. Specifically, we analyze a dynamical low-rank discontinuous Galerkin (DLR-DG) method used to approximate a space homogeneous kinetic equation that models the emission and absorption of particles by a background medium. This background medium is

Received by the editor October 28, 2023, and, in revised form, May 22, 2024.
2020 *Mathematics Subject Classification.* Primary 65N12, 65N30, 65F55.
Key words and phrases. Kinetic equations, radiation transport, dynamical low-rank approximation, discontinuous Galerkin method, semi-implicit time integration, unconventional integrator.
Research at Oak Ridge National Laboratory was supported under contract DE-AC05-00OR22725 from the U.S. Department of Energy to UT-Battelle, LLC. This work was supported by the U.S. Department of Energy, Office of Science, Office of Advanced Scientific Computing Research via the Applied Mathematics Program and the Scientific Discovery through Advanced Computing (SciDAC) program. This research was supported by Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

represented as an external source which determines the equilibrium of the model for long time simulations.

Kinetic models of particle systems involve the evolution of the particle distribution function $f(\boldsymbol{p}, \boldsymbol{x}, t)$, a phase-space density depending on the particle momentum $\boldsymbol{p} \in \mathbb{R}^3$, position $\boldsymbol{x} \in \mathbb{R}^3$, and time t. Kinetic equations, governing the evolution of f, are expressed as a balance between phase-space advection (e.g., due to inertia and external forces) and collisions (e.g., due to interparticle interactions or interactions with a background). In the absence of collisions, the distribution function can develop complex phase-space structures, while collisions tend to drive f towards an equilibrium, characterized by (spatially) local conditions, in which the dynamics can be accurately described by fluid models (where variables depend only on $\boldsymbol{x}$ and t). As such, kinetic models are high-dimensional models that can exhibit low-dimensional structure under certain conditions (e.g., particle systems undergoing frequent collisions).

DLRA methods can be traced back to the Dirac–Frenkel–McLachlan variational principle developed in the 1930s [13, 19] and function by evolving a dynamical system on the Riemannian manifold of fixed rank matrices. This evolution is achieved by projecting the right-hand side of a matrix differential equation onto the tangent space of the manifold, which yields a set of differential equations that govern the factors of an SVD-like decomposition. As such, they can be suitable for modeling high-dimensional systems that exhibit dynamics in a lower-dimensional manifold (e.g., kinetic equations). Recently, they have been applied to simulate high-dimensional quantum systems, biological cellular systems [6, 23, 28], kinetic/transport equations [12, 14–17, 32–34], hyperbolic problems with uncertainty [26], and neural network training [36]. Several integrators have been developed to overcome the stiffness induced by the high curvature of the manifold [10, 24, 29]. In this paper, we analyze a *semi-implicit basis-update & Galerkin (SIBUG) integrator* which is the unconventional integrator of [10] with implicit time stepping in each of the substeps of the integrator. For collision operators, including the relaxation operator studied in this work, implicit time discretization is desired because the short time scales induced by collisions can render explicit methods inefficient.

The discontinuous Galerkin (DG) method is a finite element method that uses a discontinuous piecewise polynomial space to approximate the numerical solution. The method offers several advantages, such as high-order accuracy on a compact stencil, compatibility with *hp*-adaptivity, and the ability to handle domains with complex geometry [22, 35, 37]. Its mathematical formulation makes it amenable to rigorous analysis. Moreover, DG methods are attractive for solving kinetic equations because of their ability to maintain structural properties (e.g., asymptotic limits [2, 21, 27] and conservation [3, 11]) of the continuum model formulation, in part, because of flexibility in the approximation spaces. However, the use of the DG methods to solve kinetic equations in full dimensionality, without any form of adaptivity to reduce the total number of degrees of freedom, can be computationally expensive.

The DLR-DG method studied in this paper applies DLRA to the matrix differential equation resulting from the semi-discretization of the kinetic equation using the DG method. The combination of DLRA and DG methods aims to leverage the benefits of both approaches by lowering the computational complexity relative to standard DG methods while retaining high-order accuracy. In this work, we

consider a model kinetic equation of relaxation-type in reduced dimensionality (by assuming homogeneity in physical space and imposing axial symmetry in momentum space) and focus on establishing conditions for which the DLR-DG formulation possesses the same properties as the standard DG scheme. The equation is, e.g., used to model the emission and absorption of neutrinos by stellar matter during the explosion of massive stars [8, 9, 30]. We use spherical-polar momentum space coordinates; as a result, a volume Jacobian appears in the inner product of the DG scheme, giving a matrix weight in the matrix differential equation. This weighting must be respected in the DLRA formulation in order to obtain similar stability and well-posedness results as in the standard DG method.

Additionally, many systems for which DLRA is applied are autonomous with equilibria that arise from the invariants embedded in the dynamics. However, there has been a lack of analysis in how DLRA integrators behave when equilibria depend on the presence of external sources. We believe this analysis is fruitful since, even when applied to linear problems, the nonlinear DLRA model, in the continuum, is only well-posed up to finite time [4, 25]. This finite time condition coincides with the solution leaving the fixed-rank manifold – a property that is influenced by external sources. Error analysis has shown that DLRA integrators [10, 29] are robust in regimes where the DLRA solution might fail to exist in the continuum, but this analysis assumes a sufficiently small timestep in order to control the consistency error introduced by these integrators. This manuscript provides conditions to guarantee convergence of the SIBUG to an equilibrium solution from external sources that is valid for larger timesteps as is often taken with an implicit method. The analysis is technical and requires estimates on each substep of the SIBUG. Moreover, unlike standard DG methods where estimates for one timestep can be easily bootstrapped into multi-step estimates, the temporally varying reduced-order basis generated by DLRA requires conditions, given in this work, to extend one-step estimates over multiple timesteps.

The rest of the paper is organized as follows. In Section 2, we introduce the space homogeneous kinetic equation, the full-rank DG discretization, and summarize the properties of the full-rank DG solution. In Section 3, we formulate the matrix differential equation associated with the DG scheme and introduce its weighted DLRA. In Section 4, we introduce the SIBUG and the equivalent DLR-DG scheme. By analysis of the DLR-DG scheme, we prove the well-posedness of the SIBUG and the convergence of the DLR-DG solution to the equilibrium. Numerical examples illustrating the theoretical results are given in Section 5.

2. Background

2.1. **Model equations.** The space homogeneous kinetic equation modeling the emission and absorption of particles by a material background at rest can be written as (see, e.g., [31])

$$\begin{aligned} \partial_t f(x,\varepsilon,\vartheta,\varphi,t) =& \mathcal{C}(f)(x,\varepsilon,\vartheta,\varphi,t), \\ f(x,\varepsilon,\vartheta,\varphi,t=0) =& f_0(x,\varepsilon,\vartheta,\varphi), \end{aligned} \tag{2.1}$$

where $f \geq 0$ is the phase-space distribution function depending on position $x \in D_x \subset \mathbb{R}^3$, and spherical-polar momentum coordinates $(\varepsilon, \vartheta, \varphi)$, and time $t \geq 0$. Here, $\varepsilon \geq 0$ is the particle energy, $\vartheta \in [0, \pi]$ is the latitudinal angle, and $\varphi \in [0, 2\pi)$

the azimuthal angle. We also introduce the latitudinal angle cosine $\mu = \cos(\theta) \in [-1, 1]$.

Since we consider the space homogeneous case in this paper, we will suppress the explicit dependence on the position coordinate x from hereon. Furthermore, we impose axial symmetry in the azimuthal direction in momentum space (i.e., f is independent of φ). We then write (2.1) as

$$\partial_t f(\varepsilon, \mu, t) = \mathcal{C}(f)(\varepsilon, \mu, t), \tag{2.2a}$$

$$f(\mu, \varepsilon, t = 0) = f_0(\mu, \varepsilon), \tag{2.2b}$$

where the collision operator on the right-hand side is given by

$$\mathcal{C}(f)(\varepsilon, \mu, t) = \eta(\varepsilon) - \chi(\varepsilon) f(\mu, \varepsilon, t), \tag{2.3}$$

where $\eta > 0$ is the emissivity and $\chi > 0$ is the opacity. Both the emissivity and opacity are assumed to be independent of the momentum space angle cosine μ, as is often done when the particle-material coupling is modeled in the material rest frame [31]. The specific dependence of the opacity χ on the particle energy ε depends on the details of the particle-material interaction process. The problem (2.2) is well-posed if $\chi(\varepsilon) \in L^\infty(\mathbb{R})$ and $\int_{\mathbb{R}} \varepsilon^2 \eta^2 d\varepsilon < \infty$. Generally, we make Assumption 2.1 on χ.

Assumption 2.1. There exist constants $\chi_{\min}, \chi_{\max}$ such that

$$0 < \chi_{\min} \leq \chi \leq \chi_{\max}. \tag{2.4}$$

Under Assumption 2.1, it can be verified that as $t \to \infty$ the solution $f(\mu, \varepsilon, t)$ to (2.2) converges to the *isotropic* equilibrium solution $f^{\mathsf{Eq}}(\varepsilon) = \eta(\varepsilon)/\chi(\varepsilon)$, which is the solution to the steady equation

$$\mathcal{C}(f^{\mathsf{Eq}}) = 0. \tag{2.5}$$

Since f^{Eq} is low-dimensional (independent of μ), and the true solution is given by $f(\mu, \varepsilon, t) = f_0 e^{-\chi t} + (1 - e^{-\chi t}) f^{\mathsf{Eq}}$, it is expected that f will become independent of μ when $\chi t \gg 1$.

There is no coupling across energies in the collision term on the right-hand side of Eq. (2.2), and ε is simply a parameter of the model. However, we include the energy dimension in the DG discretization to develop a more general framework that can accommodate coupling in energy and angle — either through the inclusion of inelastic scattering or external fields. In addition, discretizing in both energy and angle allows us to capture momentum space structures.

2.2. **DG discretization and matrix equations.** Given $\varepsilon_{\max} > 0$, we denote the computational domain by $\Omega = \{(\mu, \varepsilon) : \mu \in [-1, 1], \varepsilon \in [0, \varepsilon_{\max}]\}$ with volume measure $\mathrm{d}\Omega = \varepsilon^2 \mathrm{d}\varepsilon \mathrm{d}\mu$.[1] Let $L^2(\Omega)$ be the Hilbert space of square integrable functions defined on Ω with respect to the measure $\mathrm{d}\Omega$, and inner product denoted by

$$(v, w; \varepsilon^2)_\Omega := \int_\Omega vw \, \mathrm{d}\Omega = \int_0^{\varepsilon_{\max}} \int_{-1}^{1} vw \, \varepsilon^2 d\mu d\varepsilon. \tag{2.6}$$

The associated norm on $L^2(\Omega)$ is given by $\|\varepsilon w\|^2_{L^2(\Omega)} := (w, w; \varepsilon^2)_\Omega$.

[1]The Lebesgue measure for axially symmetric functions defined on a ball centered at 0 in $\mathbb{R}^3$ is $2\pi\varepsilon^2 \mathrm{d}\varepsilon \mathrm{d}\mu$, but we drop the 2π as each integral will have it as a common factor.

We write $\Omega = \Omega_\mu \times \Omega_\varepsilon$, where $\Omega_\varepsilon = [0, \varepsilon_{\max}]$ and $\Omega_\mu = [-1, 1]$ with measures $\varepsilon^2 \mathrm{d}\varepsilon$ and $\mathrm{d}\mu$, respectively. Let $(\cdot,\cdot;\varepsilon^2)_{\Omega_\varepsilon}$ and $(\cdot,\cdot)_{\Omega_\mu}$ be the L^2 inner products induced from the given measures.

Given $N_\varepsilon \in \mathbb{N}$ and $N_\mu \in \mathbb{N}$, we partition Ω_ε and Ω_μ into N_ε and N_μ cells, respectively. Denote these partitions by

$$0 = \varepsilon_{1/2} < \varepsilon_{3/2} < \ldots < \varepsilon_{N_\varepsilon - 1/2} < \varepsilon_{N_\varepsilon + 1/2} = \varepsilon_{\max}, \tag{2.7}$$

$$-1 = \mu_{1/2} < \mu_{3/2} < \ldots < \mu_{N_\mu - 1/2} < \mu_{N_\mu + 1/2} = 1. \tag{2.8}$$

We partition the domain Ω into logical rectangles given by

$$K_{\mathfrak{ij}} = \{(\mu, \varepsilon) : \mu \in K^\mu_{\mathfrak{i}},\ \varepsilon \in K^\varepsilon_{\mathfrak{j}}\}, \tag{2.9}$$

where $K^\mu_{\mathfrak{i}} = [\mu_{\mathfrak{i}-1/2}, \mu_{\mathfrak{i}+1/2}]$ for $1 \le \mathfrak{i} \le N_\mu$, and $K^\varepsilon_{\mathfrak{j}} = [\varepsilon_{\mathfrak{j}-1/2}, \varepsilon_{\mathfrak{j}+1/2}]$ for $1 \le \mathfrak{j} \le N_\varepsilon$.

We now define the discontinuous Galerkin finite element space in each direction as

$$V_{z,h} \coloneqq \{\phi \in L^2(\Omega_v) : \phi|_{K^z_{\mathfrak{i}}} \in \mathcal{P}_k(K^z_{\mathfrak{i}}), 1 \le \mathfrak{i} \le N_z\}, \tag{2.10}$$

where $z = \mu, \varepsilon$, and $\mathcal{P}_k$ denotes polynomials of maximal degree k. The discontinuous Galerkin finite element space is defined as

$$V_h = V_{\mu,h} \otimes V_{\varepsilon,h} = \{v : v|_{K_{\mathfrak{ij}}} \in \mathcal{Q}_k(K_{\mathfrak{ij}}), 1 \le \mathfrak{i} \le N_\mu, 1 \le \mathfrak{j} \le N_\varepsilon\},$$

where $\mathcal{Q}_k$ denotes the space of tensor-product polynomials of degree at most k for each variable defined on $K_{\mathfrak{ij}}$.

Generally, for a scalar function v and vector valued functions $\mathcal{V} = [v_1, \ldots, v_m]^\top \in \mathbb{R}^m$ and $\mathcal{W} = [w_1, \ldots, w_n]^\top \in \mathbb{R}^n$, defined on $D \subseteq \Omega$, we define

$$\begin{aligned} (v, \mathcal{W}; \phi)_D &= (\mathcal{W}, v; \phi)_D = [(v, w_j; \phi)_D]_{n\times 1} \in \mathbb{R}^n, \\ (\mathcal{V}, \mathcal{W}^\top; \phi)_D &= [(v_i, w_j; \phi)_D]_{m\times n} \in \mathbb{R}^{m\times n}, \end{aligned} \tag{2.11}$$

where $\phi = \phi(\varepsilon) > 0$ is a specified weighting function.

2.2.1. *Semi-discrete full-rank DG scheme.* The standard semi-discrete DG scheme, which we call the semi-discrete full-rank DG scheme for (2.2), together with the initial data (2.2b), is to find $f_h(\mu, \varepsilon, t) \in V_h$ such that

$$(\partial_t f_h, w_h; \varepsilon^2)_\Omega = \mathcal{A}(f_h, w_h), \quad \forall w_h \in V_h, \tag{2.12a}$$

$$(f_h|_{t=0}, w_h; \varepsilon^2)_\Omega = (f_0, w_h; \varepsilon^2)_\Omega, \quad \forall w_h \in V_h, \tag{2.12b}$$

where $\mathcal{A} : L^2(\Omega) \times L^2(\Omega) \to \mathbb{R}$ is defined by

$$\mathcal{A}(f_h, w_h) = (\mathcal{C}(f_h), w_h; \varepsilon^2)_\Omega = \left(\eta, w_h; \varepsilon^2\right)_\Omega - \left(\chi f_h, w_h; \varepsilon^2\right)_\Omega. \tag{2.13}$$

Remark 2.2. We use the term *full-rank* throughout the paper to refer to a standard discontinuous Galerkin discretization with no low-rank techniques applied.

Definition 2.1. The discrete equilibrium $f_h^{\mathsf{Eq}} \in V_h$ is the solution to the variational problem

$$\mathcal{A}(f_h^{\mathsf{Eq}}, w_h) = 0 \quad \forall w_h \in V_h. \tag{2.14}$$

Moreover, the following statement holds for the discrete equilibrium.

Lemma 2.3. *Under Assumption* 2.1*, Eq.* (2.14) *admits a unique solution* f_h^{Eq}*, which is a quasi-optimal approximation to* f^{Eq} *in* $L^2(\Omega)$*, i.e.,*

$$\|(f^{\mathsf{Eq}} - f_h^{\mathsf{Eq}})\varepsilon\|_{L^2(\Omega)} \leq \frac{\chi_{\max}}{\chi_{\min}} \inf_{w_h \in V_h} \|(f^{\mathsf{Eq}} - w_h)\varepsilon\|_{L^2(\Omega)}. \tag{2.15}$$

Proof. From Assumption 2.1, the bilinear form $(\chi \cdot, \cdot)_\Omega$ is coercive on $V_h \times V_h$; therefore the well-posedness of (2.14) is immediate. Subtracting (2.14) from the variational formulation of (2.5) gives a Galerkin orthogonality condition

$$\left(\chi(f^{\mathsf{Eq}} - f_h^{\mathsf{Eq}}), w_h; \varepsilon^2\right)_\Omega = 0, \quad \forall w_h \in V_h. \tag{2.16}$$

Assumption 2.1 and (2.16) are then used in a standard finite element argument (see [7, (2.8.1)]) that yields the Céa-type estimate (2.15). □

2.2.2. *Fully-discrete full-rank DG scheme.* We wish to employ implicit time discretization methods because the short time scales induced by collision operators can render explicit methods inefficient. For $\mathfrak{n} \geq 0$, let $f_h^{\mathfrak{n}} = f_h(\mu, \varepsilon, t^{\mathfrak{n}}) \in V_h$ be an approximation of $f(\mu, \varepsilon, t^{\mathfrak{n}})$, where $t^{\mathfrak{n}} = \mathfrak{n}\Delta t$ and $\Delta t > 0$ is a specified time step. We apply a backward Euler time discretization to the semi-discrete full-rank DG scheme (2.12). For simplicity, we denote

$$D_t v^{\mathfrak{n}+1} = \frac{v^{\mathfrak{n}+1} - v^{\mathfrak{n}}}{\Delta t}, \tag{2.17}$$

where v can be any function (or matrix in the later sections). Then, the first-order fully-discrete full-rank DG scheme for (2.2) is to find $f_h^{\mathfrak{n}+1} \in V_h$ such that

$$\left(D_t f_h^{\mathfrak{n}+1}, w_h; \varepsilon^2\right)_\Omega = \mathcal{A}(f_h^{\mathfrak{n}+1}, w_h) \quad \forall w_h \in V_h. \tag{2.18}$$

We now give the following result detailing the well-posedness, and convergence to the discrete equilibrium of the fully-discrete full-rank DG scheme. For brevity, we omit the proof, since in Section 4, we prove a similar result in the low-rank setting.

Proposition 2.4. *For any* $\Delta t > 0$*, there exists a unique solution* $f_h^{\mathfrak{n}+1}$ *of the fully-discrete, full-rank DG scheme* (2.18) *such that*

(i) *The solution* $f_h^{\mathfrak{n}+1}$ *is* L^2 *stable in the following sense:*

$$\|\varepsilon f_h^{\mathfrak{n}+1}\|_{L^2(\Omega)} \leq c^{\mathfrak{n}+1}\|\varepsilon f_0\|_{L^2(\Omega)} + \frac{1}{\chi_{\min}}(1 - c^{\mathfrak{n}+1})\|\varepsilon\eta\|_{L^2(\Omega)}, \tag{2.19}$$

where the parameter c *is given by*

$$c = \frac{1}{1 + \Delta t \chi_{\min}}. \tag{2.20}$$

(ii) *The distance between* $f_h^{\mathfrak{n}+1}$ *and the discrete equilibrium* f_h^{Eq} *is geometrically decreasing:*

$$\|\varepsilon(f_h^{\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \leq c^{\mathfrak{n}+1}\|\varepsilon(f_h^0 - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}, \tag{2.21}$$

where f_h^{Eq} *satisfies* (2.14)*.*

Remark 2.5. For large Δt, (2.21) implies that $f_h^{\mathfrak{n}}$ converges to f_h^{Eq} at a rate $O(\Delta t^{-\mathfrak{n}})$ for any $\mathfrak{n} \geq 1$.

The main objective of this paper is to establish results analogous to Proposition 2.4 when the dynamical low-rank approximation is applied to the DG scheme. These are given in Section 4.

3. Dynamical low-rank formulation

In this section, we formulate low-rank approximations to (2.12).

3.1. **Formulation of the matrix differential equation.** In order to apply the dynamical low-rank approximation, we first convert (2.12) into an equivalent matrix differential equation via a basis expansion. Let $\{x_i(\mu)\}_{i=1}^m$ and $\{y_j(\varepsilon)\}_{j=1}^n$ be bases for the finite element spaces $V_{\mu,h}$ and $V_{\varepsilon,h}$, respectively. Here, $m = (k+1)N_\mu$ and $n = (k+1)N_\varepsilon$. We construct these bases using local Legendre polynomials on the local cells K_{i}^μ and $K_{\mathrm{j}}^\varepsilon$ that are orthonormal with respect to the local inner products $L^2(K_{\mathrm{i}}^\mu)$ and $L^2(K_{\mathrm{j}}^\varepsilon)$, respectively. With this choice $\{x_i(\mu)\}_{i=1}^m$ forms on an orthonormal basis for $V_{\mu,h}$. However, $\{y_j(\varepsilon)\}_{j=1}^n$ does not form on an orthonormal basis for $V_{\varepsilon,h}$ due to the weight ε^2 in the inner product (cf. (2.6)). This fact has technical consequences for the remainder of the paper.

Given a function $w_h \in V_h$, its basis expansion can be written as

$$w_h = \sum_{i=1}^m \sum_{j=1}^n W_{ij}(t) x_i(\mu) y_j(\varepsilon) = X^\top(\mu) W(t) Y(\varepsilon), \tag{3.1}$$

where $X : \Omega_\mu \to \mathbb{R}^m$ and $Y : \Omega_\varepsilon \to \mathbb{R}^n$ are defined by

$$X(\mu) = [x_1(\mu), \dots, x_m(\mu)]^\top \text{ and } Y(\varepsilon) = [y_1(\varepsilon), \dots, y_n(\varepsilon)]^\top. \tag{3.2}$$

We call $W = [W_{ij}] \in \mathbb{R}^{m\times n}$ the *coefficient matrix of* w_h (*with respect to the bases* $\{x_i(\mu)\}_{i=1}^m$ *and* $\{y_j(\varepsilon)\}_{j=1}^n$). For each fixed i, W satisfies

$$\sum_{j'=1}^n (y_j, y_{j'}; \varepsilon^2)_{\Omega_\varepsilon} W_{ij'} = (w_h, x_i y_j; \varepsilon^2)_\Omega, \quad j = 1, \dots, n. \tag{3.3}$$

Definition 3.1. Given matrices $A, B \in \mathbb{R}^{m\times n}$ with entries A_{ij} and B_{ij}, their Frobenius inner product is $(A, B)_{\mathrm{F}} = \operatorname{tr}(A^\top B) = \sum_{i=1}^m \sum_{j=1}^n A_{ij} B_{ij}$. The Frobenius norm of A is $\|A\|_{\mathrm{F}} = \sqrt{(A, A)_{\mathrm{F}}}$.

Lemma 3.1 relates weighted inner products of DG functions to weighted Frobenius inner products of the associated coefficient matrices. It follows from a direct calculation using (3.1).

Lemma 3.1. *Let $Z \in \mathbb{R}^{m\times n}$ and $W \in \mathbb{R}^{m\times n}$ be the coefficient matrices of $z_h \in V_h$ and $w_h \in V_h$, respectively, and let $\phi = \phi(\varepsilon)$ be a scalar function. Then*

$$(\phi(\varepsilon) z_h, w_h; \varepsilon^2)_\Omega = (I_m Z A_\phi, W)_{\mathrm{F}} = (Z A_\phi, W)_{\mathrm{F}}, \tag{3.4}$$

where I_m is the $m \times m$ identity matrix and the symmetric matrix

$$A_\phi = (\phi(\varepsilon) Y^\top(\varepsilon), Y(\varepsilon); \varepsilon^2)_{\Omega_\varepsilon} \in \mathbb{R}^{n\times n}, \tag{3.5}$$

is block diagonal due to the locality of the basis. If further $\phi(\varepsilon) > 0$, then A_ϕ is also positive-definite.

Corollary 3.2. *Let $F \in \mathbb{R}^{m\times n}$ be the coefficient matrix of the DG solution $f_h \in V_h$ in* (2.12)*, and $W \in \mathbb{R}^{m\times n}$ be the coefficient matrix of any function $w_h \in V_h$. Then the semi-discrete DG scheme* (2.12) *is equivalent to the following problem: Find $F(t) \in \mathbb{R}^{m\times n}$ such that*

$$(\partial_t F(t) A_{\mathbf{1}}, W)_{\mathrm{F}} = (G(F), W)_{\mathrm{F}}, \quad \forall W \in \mathbb{R}^{m\times n}, \tag{3.6a}$$

$$F(0) = F_0, \tag{3.6b}$$

where $F_0 \in \mathbb{R}^{m\times n}$ *is the coefficient matrix of* $f_h(\mu,\varepsilon,0)$ *obtained by solving* (2.12b). *Here* $A_{\mathbf{1}}$ *is the symmetric positive-definite, block-diagonal matrix defined by* (3.5) *with* $\phi = 1$, *and* G *is the affine function defined by*

$$G(F) = L_0 L_\eta^\top - F A_\chi, \tag{3.7}$$

where

$$L_0 = (1, X)_{\Omega_\mu} \in \mathbb{R}^{m\times 1}, \quad L_\eta = (\eta, Y; \varepsilon^2)_{\Omega_\varepsilon} \in \mathbb{R}^{n\times 1}, \tag{3.8}$$

and the symmetric positive-definite, block-diagonal matrix A_χ *is defined by* (3.5) *with* $\phi = \chi$.

The variational problem (3.6) immediately yields the following matrix-valued ODE:

$$\partial_t F = G(F) A_{\mathbf{1}}^{-1}. \tag{3.9}$$

3.2. **Weighted dynamical low-rank approximation.** Let $\mathcal{M}_r \subset \mathbb{R}^{m\times n}$ be the manifold of rank-r matrices ($r \le \min\{m,n\}$). The Dynamical Low-Rank Approximation (DLRA) is traditionally formulated by evolving the matrix-valued ODE (3.9) on $\mathcal{M}_r$ by a Galerkin projection of $\partial_t F$ onto the tangent space of $\mathcal{M}_r$ centered at F (see e.g., [25]). This projection is on the space of $m \times n$ matrices and is traditionally orthogonal with respect to the standard Frobenius inner product in Definition 3.1. However, such a formulation will not preserve the natural equivalence between the Galerkin equation of the DRLA and the matrix variational problem in (3.6). In order to maintain this equivalence in the DLRA framework, we propose a modification to the standard DLRA approach that uses the weight $A_{\mathbf{1}}$ to characterize the tangent space.

Definition 3.2. For any $Z, W \in \mathbb{R}^{m_1\times n}$, $1 \le m_1 \le m$, and any symmetric positive definite matrix $M \in \mathbb{R}^{n\times n}$ with Cholesky factorization $M = C^\top C$, the M-weighted Frobenius inner product and its induced norm on $\mathbb{R}^{m_1\times n}$ are given by

$$(Z, W)_M := (ZM, W)_{\mathrm{F}} = (ZC^\top, WC^\top)_{\mathrm{F}} \quad \text{and} \quad \|W\|_M^2 := (W, W)_M. \tag{3.10}$$

Remark 3.3. The weighted Frobenius norm serves two purposes. The first is to introduce the matrix weight induced by the ε^2 integration weight in the definition of $\mathcal{A}$; see (2.13). The second is to introduce linear operations on the energy basis that, due to the transpose that appears in the rank-based representation of a matrix (e.g., the matrix $E^\top$ in (3.12)), are often represented by left matrix multiplication. Thus for consistency, we reserve the usual vector norm $\|\cdot\|$ on $\mathbb{R}^n$ for column vectors $x \in \mathbb{R}^{n\times 1}$ and use the Frobenius norm for row vectors $x^\top \in \mathbb{R}^{1\times n}$, $\|x^\top\|_{\mathrm{F}}^2 = \mathrm{tr}(xx^\top) = \|x\|^2$.

Definition 3.3. Let $\hat{F}_0 \in \mathcal{M}_r$ be given. The (weighted) dynamical low-rank approximation to (3.9) is given by the solution $\hat{F} \in \mathcal{M}_r$ (where $\hat{F}$ approximates F) of the differential equation

$$\partial_t \hat{F} = \underset{\delta\hat{F}\in\mathcal{T}_{\hat{F}}\mathcal{M}_r}{\arg\min}\ J(\delta\hat{F}), \quad \text{where} \quad J(\delta\hat{F}) = \|\delta\hat{F} - G(\hat{F})A_{\mathbf{1}}^{-1}\|_{A_{\mathbf{1}}}, \tag{3.11}$$

with initial condition $\hat{F}(0) = \hat{F}_0$. Here, $\mathcal{T}_{\hat{F}}\mathcal{M}_r$ is the tangent space of $\mathcal{M}_r$ at $\hat{F}$.

Remark 3.4. The initial condition $\hat{F}_0$ should be a rank-r approximation to $F(0)$. We delay the choice of $\hat{F}_0$ until the end of this section.

Like the usual DLRA [25], (3.11) can be rewritten into an equivalent system that updates the components of the low-rank decomposition of $\hat{F}$ in time; this equivalent system is often called the *equations of motion*. Let $\hat{F}$ have the rank-r decomposition

$$\hat{F} = USE^\top, \text{ where } U^\top U = E^\top A_{\mathbf{1}} E = I_r, \tag{3.12}$$

with $U \in \mathbb{R}^{m\times r}$, $S \in \mathbb{R}^{r\times r}$, and $E \in \mathbb{R}^{n\times r}$ all full-rank matrices.[2] In terms of U, S, and E, the tangent space of $\mathcal{M}_r$ at $\hat{F}$ is (see e.g., [25]):

$$\mathcal{T}_{\hat{F}}\mathcal{M}_r = \{\delta USE^\top + U\delta SE^\top + US\delta E^\top : U^\top \delta U = 0,\ E^\top A_{\mathbf{1}} \delta E = 0\}, \tag{3.13}$$

where $\delta U \in \mathbb{R}^{m\times r}$, $\delta S \in \mathbb{R}^{r\times r}$, and $\delta E \in \mathbb{R}^{n\times r}$. Due to the gauge conditions $U^\top \delta U = E^\top A_{\mathbf{1}} \delta E = 0$ in (3.13), any matrix $\delta\hat{F} \in \mathcal{T}_{\hat{F}}\mathcal{M}_r$ has the unique decomposition

$$\delta\hat{F} = \delta USE^\top + U\delta SE^\top + US\delta E^\top = P_U^\perp \delta\hat{F} A_{\mathbf{1}} P_E + P_U \delta\hat{F} A_{\mathbf{1}} P_E + P_U \delta\hat{F}^\top A_{\mathbf{1}} P_E^\perp, \tag{3.14}$$

where

$$\delta U = P_U^\perp \delta\hat{F} A_{\mathbf{1}} E S^{-1},\ \delta S = U^\top \delta\hat{F} A_{\mathbf{1}} E, \text{ and } \delta E = P_E^\perp A_{\mathbf{1}} \delta\hat{F}^\top U S^{-T} \tag{3.15}$$

with symmetric matrices

$$P_U = UU^\top \text{ and } P_U^\perp = I_m - P_U, \tag{3.16}$$

and

$$P_E = EE^\top \text{ and } P_E^\perp = A_{\mathbf{1}}^{-1} - P_E. \tag{3.17}$$

The matrix P_U is the orthogonal projection onto the column space of U with respect to the standard inner product on $\mathbb{R}^m$ and $P_U^\perp$ is its orthogonal complement. The matrix $P_E A_{\mathbf{1}}$ is the orthogonal projection onto the column space of E with respect to the inner product on $\mathbb{R}^n$ with weight $A_{\mathbf{1}}$. Moreover, for any $Z, W \in \mathbb{R}^{\ell\times n}$, $1 \le \ell \le m$,

$$(ZA_{\mathbf{1}}P_E, W)_{A_{\mathbf{1}}} = (Z, WA_{\mathbf{1}}P_E)_{A_{\mathbf{1}}}, \tag{3.18}$$

where $(\cdot,\cdot)_{A_{\mathbf{1}}}$ is the Frobenius inner product defined in Definition 3.2.

We now give several equivalent formulations of the weighted DLRA solution $\hat{F}$ in Definition 3.3.

Proposition 3.5. *The solution $\hat{F} = USE^\top \in \mathcal{M}_r$ of Definition* 3.3*, with initial data $\hat{F}(0) = U^0 S^0 (E^0)^\top \in \mathcal{M}_r$ where $(U^0)^\top U^0 = (E^0)^\top A_{\mathbf{1}} E^0 = I_r$, satisfies the equivalent problems* [25]

(i) *$\partial_t \hat{F} \in \mathcal{T}_{\hat{F}}\mathcal{M}_r$ is the solution of the Galerkin condition*

$$\left(\partial_t\hat{F} - G(\hat{F})A_{\mathbf{1}}^{-1}, \delta\hat{F}\right)_{A_{\mathbf{1}}} = 0, \quad \forall \delta\hat{F} \in \mathcal{T}_{\hat{F}}\mathcal{M}_r, \tag{3.19a}$$

$$\hat{F}(0) = U^0 S^0 (E^0)^\top. \tag{3.19b}$$

(ii) *The factors of $\hat{F}$ satisfy the equations of motion given by*

$$\dot{U} = P_U^\perp G(\hat{F}) E S^{-1}, \quad \dot{S} = U^\top G(\hat{F}) E, \quad \dot{E} = P_E^\perp G(\hat{F})^\top U S^{-T}, \tag{3.20a}$$

$$U(0) = U^0, \quad S(0) = S^0, \quad E(0) = E^0, \tag{3.20b}$$

[2]Unless otherwise stated, any matrices denoted with U and E satisfy $U^\top U = I_r$ and $E^\top A_{\mathbf{1}} E = I_r$, respectively.

where $P_U^\perp$ and $P_E^\perp$ are defined in (3.16) *and* (3.17)*, respectively.*

(iii) *The matrices $\mathbf{K} = US \in \mathbb{R}^{m\times r}$, $\mathbf{L} = ES^\top \in \mathbb{R}^{n\times r}$, and S satisfy the coupled ODE system*[3]

$$\dot{\mathbf{K}} = G(\mathbf{K}E^\top)E, \quad \dot{\mathbf{L}} = A_{\mathbf{1}}^{-1}G(U\mathbf{L}^\top)^\top U, \quad \dot{S} = U^\top G(USE^\top)E, \tag{3.21a}$$

$$\mathbf{K}(0) = U^0S^0, \quad \mathbf{L}(0) = E^0(S^0)^\top, \quad S(0) = S^0. \tag{3.21b}$$

Proof. We give a short sketch.

- Definition 3.3 $\Leftrightarrow$ (i). The minimization problem (3.11) is unchanged if J is replaced by $\frac{1}{2}J^2$. The minimization of this strongly convex quadratic functional over the linear subspace $\mathcal{T}_{\hat{F}}\mathcal{M}_r$ is equivalent to the Galerkin condition (i).
- (i) $\Rightarrow$ (ii). Since $\partial_t\hat{F} = \dot{U}SE^\top + U\dot{S}E^\top + US\dot{E}^\top$, the equations for $\dot{U}$, $\dot{S}$, and $\dot{E}$ in (3.20) can be found from (3.19a) by testing against

$$\delta USE^\top = P_U^\perp U_W S^{-\top}E^\top, \quad U\delta SE^\top = US_W E^\top, \quad US(\delta E)^\top = US^{-\top}E_W^\top A_{\mathbf{1}}P_E^\perp, \tag{3.22}$$

  respectively, where $U_W \in \mathbb{R}^{m\times r}$, $S_W \in \mathbb{R}^{r\times r}$, and $E_W \in \mathbb{R}^{n\times r}$ are arbitrary. By the arbitrariness of U_W, S_W, E_W, and the gauge condition $U^\top\dot{U} = E^\top A_{\mathbf{1}}\dot{E} = 0$, (3.19a) reduces (3.20a).
- (ii) $\Leftrightarrow$ (iii). Direct calculation: Take the derivative of $\mathbf{K}$ and $\mathbf{L}$ and use the product rule, (3.16), and (3.17).
- (ii) $\Rightarrow$ (i). From the equations of motion (3.20),

$$\partial_t\hat{F} = \dot{U}SE^\top + U\dot{S}E^\top + US\dot{E}^\top = P_U^\perp GP_E + P_UGP_E + P_UGP_E^\perp. \tag{3.23}$$

  Plugging (3.23) and (3.14) into (3.19a), using (3.16) and (3.17), verifies the result.

□

Remark 3.6. With the DLRA defined in Definition 3.3, the semi-discrete DG scheme in matrix formulation (3.6a) is identical to the Galerkin equation of the DRLA (3.19a) when the coefficient matrix of the DG solution possesses a rank-r decomposition and evolves tangentially to $\mathcal{M}_r$.

4. Fully discrete dynamical low-rank DG schemes

In this section, we propose a fully discrete dynamical low-rank DG (DLR-DG) method. Similar to Proposition 2.4 for the full-rank scheme, we investigate the well-posedness of the DLR-DG method and show the convergence of its solution to the equilibrium for a sufficiently large time step.

4.1. The fully discrete DLR-DG schemes.

Applying a numerical integrator to the equations of motion in the form of Eq. (3.20) will produce an unstable method unless Δt is of the same order as the smallest singular value of S [29]. Several DLRA temporal integrators have been developed with timestep restrictions that are much more reasonable [10, 24, 29]. Here we choose the basis-update & Galerkin (BUG) integrator [10], which is easily combined with the backward Euler method.

[3]We use bold notation to represent key matrices formed by the product of matrices.

4.1.1. *A semi-implicit BUG integrator.* The BUG integrator of [10] can be viewed as a splitting method applied to the KLS system in Eq. (3.21), where the K and L equations are decoupled and updated independently, followed by an update using the S equation. We use backward (implicit) Euler for the underlying numerical integrator for all equations as collision operators generally induce timescales that cannot be efficiently advanced with an explicit method. Given $\Delta t > 0$ and the factored rank-r matrix $\hat{F}^{\mathtt{n}} = U^{\mathtt{n}} S^{\mathtt{n}} (E^{\mathtt{n}})^\top$ with factors satisfying

$$(U^{\mathtt{n}})^\top U^{\mathtt{n}} = I_r, \quad (E^{\mathtt{n}})^\top A_{\mathbf{1}} E^{\mathtt{n}} = I_r, \tag{4.1}$$

one step of the method generates a new rank-r matrix factorization

$$\hat{F}^{\mathtt{n}+1} = U^{\mathtt{n}+1} S^{\mathtt{n}+1} (E^{\mathtt{n}+1})^\top \tag{4.2}$$

with factors satisfying

$$(U^{\mathtt{n}+1})^\top U^{\mathtt{n}+1} = I_r, \quad (E^{\mathtt{n}+1})^\top A_{\mathbf{1}} E^{\mathtt{n}+1} = I_r. \tag{4.3}$$

Algorithm 4.1 precisely defines the semi-implicit basis-update & Galerkin integrator.

Algorithm 4.1. *A semi-implicit basis-update & Galerkin (SIBUG) integrator.*

- *Input:* $U^{\mathtt{n}}, S^{\mathtt{n}}, E^{\mathtt{n}}$, Δt; *output:* $U^{\mathtt{n}+1}, S^{\mathtt{n}+1}, E^{\mathtt{n}+1}$.
- ***Step 1:*** *Update* $U^{\mathtt{n}} \to U^{\mathtt{n}+1}$ *and* $E^{\mathtt{n}} \to E^{\mathtt{n}+1}$ *in parallel:*
  - K***-step:***
    - *Solve for* $\mathbf{K}^{\mathtt{n}+1}$ *from the* $m \times r$ *matrix equation*

$$D_t \mathbf{K}^{\mathtt{n}+1} = G(\mathbf{K}^{\mathtt{n}+1}(E^{\mathtt{n}})^\top) E^{\mathtt{n}}, \quad \mathbf{K}^{\mathtt{n}} = U^{\mathtt{n}} S^{\mathtt{n}}. \tag{4.4}$$

    - *Perform a QR factorization* $\mathbf{K}^{\mathtt{n}+1} = U^{\mathtt{n}+1} R_{\mathbf{K}}$.
    - *Compute the* $r \times r$ *matrix* $M^{\mathtt{n}+1} = (U^{\mathtt{n}+1})^\top U^{\mathtt{n}}$.
  - L***-step:***
    - *Solve for* $\mathbf{L}^{\mathtt{n}+1}$ *from the* $n \times r$ *matrix equation*

$$D_t \mathbf{L}^{\mathtt{n}+1} = A_{\mathbf{1}}^{-1} G(U^{\mathtt{n}}(\mathbf{L}^{\mathtt{n}+1})^\top)^\top U^{\mathtt{n}}, \quad \mathbf{L}^{\mathtt{n}} = E^{\mathtt{n}}(S^{\mathtt{n}})^\top. \tag{4.5}$$

    - *Perform a generalized QR factorization (Algorithm* B.3*)* $\mathbf{L}^{\mathtt{n}+1} = E^{\mathtt{n}+1} R_{\mathbf{L}}$.
    - *Compute the* $r \times r$ *matrix* $N^{\mathtt{n}+1} = (E^{\mathtt{n}+1})^\top A_{\mathbf{1}} E^{\mathtt{n}}$.
- ***Step 2:*** *Update* $S^{\mathtt{n}} \to S^{\mathtt{n}+1}$*:*
  - S***-step:***
    - *Project* $S^{\mathtt{n}}$ *to the new bases*

$$S^{\mathtt{n},*} = M^{\mathtt{n}+1} S^{\mathtt{n}} (N^{\mathtt{n}+1})^\top. \tag{4.6}$$

    - *Solve for* $S^{\mathtt{n}+1}$ *from the* $r \times r$ *matrix equation*

$$\frac{S^{\mathtt{n}+1} - S^{\mathtt{n},*}}{\Delta t} = (U^{\mathtt{n}+1})^\top G(U^{\mathtt{n}+1} S^{\mathtt{n}+1} (E^{\mathtt{n}+1})^\top) E^{\mathtt{n}+1}. \tag{4.7}$$

Remark 4.1. The following remarks apply to Algorithm 4.1.

(a) The choice of bases $U^{\mathtt{n}+1}$ and $E^{\mathtt{n}+1}$ used in the S-step is not unique. For any unitary matrices $V_U, V_E \in \mathbb{R}^{r \times r}$, the matrices $U^{\mathtt{n}+1} V_U$ and $E^{\mathtt{n}+1} V_E$ could replace $U^{\mathtt{n}+1}$ and $E^{\mathtt{n}+1}$, respectively, without changing $\hat{F}^{\mathtt{n}+1}$.

(b) The algorithm is semi-implicit since it uses explicit evaluation of the bases $U^{\mathtt{n}}$ and $E^{\mathtt{n}}$ in Eqs. (4.4) and (4.5), respectively, but makes implicit updates for $\mathbf{K}^{\mathtt{n}+1}$, $\mathbf{L}^{\mathtt{n}+1}$, and $S^{\mathtt{n}+1}$.

(c) $S^{\mathfrak{n},*}$ in (4.6) is the projection of $S^{\mathfrak{n}}$ under the new bases $U^{\mathfrak{n}+1}$ and $E^{\mathfrak{n}+1}$. Thus, $\|S^{\mathfrak{n},*}\|_{\mathrm{F}} \leq \|S^{\mathfrak{n}}\|_{\mathrm{F}}$. For sufficiently large Δt, the projection error does not affect the SIBUG solution's convergence to an equilibrium.

(d) Other than Algorithm B.3, the factorization $\mathbf{L}^{\mathfrak{n}+1} = E^{\mathfrak{n}+1}R_{\mathbf{L}}$ in the L**-step** can alternatively be computed by a regular QR factorization $\mathbf{L}^{\mathfrak{n}+1} = \tilde{E}^{\mathfrak{n}+1}\tilde{R}_{\mathbf{L}}$ with $(\tilde{E}^{\mathfrak{n}+1})^{\top}\tilde{E}^{\mathfrak{n}+1} = I_r$, followed by the weighted Gram–Schmidt decomposition $\tilde{E}^{\mathfrak{n}+1} = E^{\mathfrak{n}+1}\bar{R}_{\mathbf{L}}$ with $E^{\mathfrak{n}+1}$ satisfying (4.3), and then setting $R_{\mathbf{L}} = \bar{R}_{\mathbf{L}}\tilde{R}_{\mathbf{L}}$. The stability of this alternative factorization has been numerically verified, and we recover the same results as when using Algorithm B.3.

(e) If L_0, defined in (3.8), is in the span of the columns of $U^{\mathfrak{n}}$, then $U^{\mathsf{Eq}} = L_0/\|L_0\| = U^{\mathfrak{n}}z$ for some vector $z \in \mathbb{R}^{r\times 1}$. In this case, (4.4) reduces to

(4.8)
$$\mathbf{K}^{\mathfrak{n}+1} = U^{\mathfrak{n}}\bar{R}, \quad \text{where } \bar{R} = (S^{\mathfrak{n}} + \Delta t\|L_0\| z L_\eta^{\top}E^{\mathfrak{n}})\left(I_r + \Delta t(E^{\mathfrak{n}})^{\top}A_\chi E^{\mathfrak{n}}\right)^{-1} \in \mathbb{R}^{r\times r}.$$

Thus, the K**-step** can be omitted, and we can set $U^{\mathfrak{n}+1} = U^{\mathfrak{n}}$. (See also Remark 4.9, following Lemma 4.8.)

(f) The matrix $A_{\mathbf{1}}$, whose inverse is needed in the L**-step** in Algorithm 4.1, is positive definite and block diagonal. For a given mesh, its smallest eigenvalue is bounded away from zero. The inverse, $A_{\mathbf{1}}^{-1}$, can be computed (once at program startup) by inverting each $(k+1)\times(k+1)$ block independently.

4.1.2. *DG formulation of the SIBUG.* Given a low-rank approximation $\hat{f}_h^{\mathfrak{n}}$ with coefficient matrix $\hat{F}^{\mathfrak{n}} = U^{\mathfrak{n}}S^{\mathfrak{n}}E^{\mathfrak{n}}$, define the following subspaces of V_h (which depend on $\hat{f}_h^{\mathfrak{n}}$):

$$V_0^{\mathfrak{n}} = \left\{v \mid v(\mu,\varepsilon) = X^{\top}(\mu)U^{\mathfrak{n}}S(E^{\mathfrak{n}})^{\top}Y(\varepsilon), \quad \forall S \in \mathbb{R}^{r\times r}\right\}, \tag{4.9a}$$

$$V_1^{\mathfrak{n}} = \left\{v \mid v(\mu,\varepsilon) = X^{\top}(\mu)\mathbf{K}(E^{\mathfrak{n}})^{\top}Y(\varepsilon), \quad \forall \mathbf{K} \in \mathbb{R}^{m\times r}\right\}, \tag{4.9b}$$

$$V_2^{\mathfrak{n}} = \left\{v \mid v(\mu,\varepsilon) = X^{\top}(\mu)U^{\mathfrak{n}}\mathbf{L}^{\top}Y(\varepsilon), \quad \forall \mathbf{L} \in \mathbb{R}^{n\times r}\right\}. \tag{4.9c}$$

It is easy to check that $\hat{f}_h^{\mathfrak{n}} = X^{\top}U^{\mathfrak{n}}S^{\mathfrak{n}}(E^{\mathfrak{n}})^{\top}Y \in V_0^{\mathfrak{n}} \cap V_1^{\mathfrak{n}} \cap V_2^{\mathfrak{n}}$, but $\hat{f}_h^{\mathfrak{n}} \notin V_0^{\mathfrak{n}+1}$. However,

$$f_S^{\mathfrak{n},*} := X^{\top}U^{\mathfrak{n}+1}S^{\mathfrak{n},*}(E^{\mathfrak{n}+1})^{\top}Y \in V_0^{\mathfrak{n}+1}, \tag{4.10}$$

where $S^{\mathfrak{n},*}$ is given in (4.7). Moreover, $f_S^{\mathfrak{n},*}$ is the L^2 projection of $\hat{f}_h^{\mathfrak{n}}$ onto $V_0^{\mathfrak{n}+1}$:

$$(f_S^{\mathfrak{n},*}, w_h; \varepsilon^2)_\Omega = (\hat{f}_h^{\mathfrak{n}}, w_h; \varepsilon^2)_\Omega, \quad \forall w_h \in V_0^{\mathfrak{n}+1}. \tag{4.11}$$

Lemma 4.2 establishes an equivalent DG formulation for (4.4)–(4.7).

Lemma 4.2. *The matrices* $\mathbf{K}^{\mathfrak{n}+1}, \mathbf{L}^{\mathfrak{n}+1}, S^{\mathfrak{n}+1}$ *are solutions to* (4.4), (4.5), (4.7), *respectively, iff* $f_{\mathbf{K}}^{\mathfrak{n}+1} := X^{\top}(\mu)\mathbf{K}^{\mathfrak{n}+1}(E^{\mathfrak{n}})^{\top}Y(\varepsilon) \in V_1^{\mathfrak{n}}$, $f_{\mathbf{L}}^{\mathfrak{n}+1} := X^{\top}(\mu)U^{\mathfrak{n}}(\mathbf{L}^{\mathfrak{n}+1})^{\top}Y(\varepsilon) \in V_2^{\mathfrak{n}}$, *and* $f_S^{\mathfrak{n}+1} := X^{\top}(\mu)U^{\mathfrak{n}+1}S^{\mathfrak{n}+1}(E^{\mathfrak{n}+1})^{\top}Y(\varepsilon) \in V_0^{\mathfrak{n}+1}$ *solve the following DLR-DG scheme*

$$\left(D_t f_{\mathbf{K}}^{\mathfrak{n}+1}, w_1; \varepsilon^2\right)_\Omega = \mathcal{A}(f_{\mathbf{K}}^{\mathfrak{n}+1}, w_1), \quad \forall w_1 \in V_1^{\mathfrak{n}}, \tag{4.12a}$$

$$\left(D_t f_{\mathbf{L}}^{\mathfrak{n}+1}, w_2; \varepsilon^2\right)_\Omega = \mathcal{A}(f_{\mathbf{L}}^{\mathfrak{n}+1}, w_2), \quad \forall w_2 \in V_2^{\mathfrak{n}}, \tag{4.12b}$$

$$\left(D_t f_S^{\mathfrak{n}+1}, w_0; \varepsilon^2\right)_\Omega = \mathcal{A}(f_S^{\mathfrak{n}+1}, w_0), \quad \forall w_0 \in V_0^{\mathfrak{n}+1}, \tag{4.12c}$$

where $f_{\mathbf{K}}^{\mathfrak{n}} = f_{\mathbf{L}}^{\mathfrak{n}} = f_S^{\mathfrak{n}} = \hat{f}_h^{\mathfrak{n}}$.

Proof. We only prove the equivalence between (4.7) and (4.12c); the others can be proved similarly. Suppose $f_S^{\mathfrak{n}+1}$ solves (4.12c). Then, by Lemma 3.1, $S^{\mathfrak{n}+1}$ solves

$$\begin{aligned}&\left(U^{\mathfrak{n}+1}D_tS^{n+1}(E^{\mathfrak{n}+1})^\top A_{\mathbf{1}}, U^{\mathfrak{n}+1}W_0(E^{\mathfrak{n}+1})^\top\right)_{\mathrm{F}}\\ &\quad=\left(G(U^{\mathfrak{n}+1}S^{\mathfrak{n}+1}(E^{\mathfrak{n}+1})^\top), U^{\mathfrak{n}+1}W_0(E^{\mathfrak{n}+1})^\top\right)_{\mathrm{F}},\end{aligned}\tag{4.13}$$

for all $W_0 \in \mathbb{R}^{r\times r}$. The matrix form of (4.11):

$$\left(U^{\mathfrak{n}}S^{\mathfrak{n}}(E^{\mathfrak{n}})^\top A_{\mathbf{1}}, U^{\mathfrak{n}+1}W_0(E^{\mathfrak{n}+1})^\top\right)_{\mathrm{F}} = \left(U^{\mathfrak{n}+1}S^{\mathfrak{n},*}(E^{\mathfrak{n}+1})^\top A_{\mathbf{1}}, U^{\mathfrak{n}+1}W_0(E^{\mathfrak{n}+1})^\top\right)_{\mathrm{F}},\tag{4.14}$$

can be used to replace $S^{\mathfrak{n}}$ by $S^{\mathfrak{n},*}$ in (4.13). Then applying Lemma A.1 and (4.3) gives

$$\left(\frac{S^{\mathfrak{n}+1}-S^{\mathfrak{n},*}}{\Delta t}, W_0\right)_{\mathrm{F}} = \left((U^{\mathfrak{n}+1})^\top G(U^{\mathfrak{n}+1}S^{\mathfrak{n}+1}(E^{\mathfrak{n}+1})^\top)E^{\mathfrak{n}+1}, W_0\right)_{\mathrm{F}}.\tag{4.15}$$

Since W_0 is arbitrary, (4.15) is equivalent to (4.7). □

4.2. **Well-posedness.** We now obtain an analog of Proposition 2.4(i) for the SIBUG listed in Algorithm 4.1 – namely that the DLR-DG scheme is uniquely solvable and uniformly stable.

Lemma 4.3. *Given the low-rank representation $\hat{f}_h^{\mathfrak{n}}$, from which $f_{\mathbf{K}}^{\mathfrak{n}}$, $f_{\mathbf{L}}^{\mathfrak{n}}$, and $f_S^{\mathfrak{n}}$ can be computed, the first order fully discrete DG scheme* (4.12) *admits a unique solution $(f_{\mathbf{K}}^{\mathfrak{n}+1}, f_{\mathbf{L}}^{\mathfrak{n}+1}, f_S^{\mathfrak{n}+1}) \in V_1^{\mathfrak{n}} \times V_2^{\mathfrak{n}} \times V_0^{\mathfrak{n}+1}$ for any $\Delta t > 0$. Equivalently, Algorithm* 4.1 *admits a unique matrix solution $(\mathbf{K}^{\mathfrak{n}+1}, \mathbf{L}^{\mathfrak{n}+1}, S^{\mathfrak{n}+1})$.*

Proof. We only prove the existence and uniqueness for $f_S^{\mathfrak{n}+1}$; the corresponding results for $f_{\mathbf{K}}^{\mathfrak{n}+1}$ and $f_{\mathbf{L}}^{\mathfrak{n}+1}$ can be proved in a similar way. Since (4.12c) is a linear system in a finite dimensional space where the domain and codomain have the same dimension, existence is equivalent to uniqueness. Let $\delta f_S^{\mathfrak{n}+1} \in V_0^{n+1}$ be the difference between two possible solutions to (4.12c). Then

$$\left(\delta f_S^{\mathfrak{n}+1}, w_0; \varepsilon^2\right)_\Omega = -\Delta t(\chi(\varepsilon)\delta f_S^{\mathfrak{n}+1}, w_0; \varepsilon^2)_\Omega \quad \forall w_0 \in V_0^{\mathfrak{n}+1}.\tag{4.16}$$

If $w_0 = \delta f_S^{\mathfrak{n}+1}$, then $\|\varepsilon\delta f_S^{\mathfrak{n}+1}\|_{L^2(\Omega)}^2 + \Delta t\|\varepsilon\sqrt{\chi(\varepsilon)}\delta f_S^{\mathfrak{n}+1}\|_{L^2(\Omega)}^2 = 0$, which implies $\delta f_S^{\mathfrak{n}+1} = 0$. Therefore, the DG scheme (4.12c) admits a unique solution. The uniqueness of $(f_{\mathbf{K}}^{\mathfrak{n}+1}, f_{\mathbf{L}}^{\mathfrak{n}+1}, f_S^{\mathfrak{n}+1})$ and the equivalence established by Lemma 4.2 imply that Algorithm 4.1 admits the unique matrix solution $(\mathbf{K}^{\mathfrak{n}+1}, \mathbf{L}^{\mathfrak{n}+1}, S^{\mathfrak{n}+1})$. □

Definition 4.1. We define the DG approximation $\hat{f}_h^{\mathfrak{n}+1} = f_S^{\mathfrak{n}+1}$ as the DLR-DG solution, and the subspace $V_0^{\mathfrak{n}+1}$ as the DLR-DG space.

The L^2 stability of the DLR-DG solution $\hat{f}_h^{\mathfrak{n}+1}$ is established by Lemma 4.4.

Lemma 4.4. *Suppose that $\|\varepsilon\hat{f}_h^0\|_{L^2(\Omega)} \le \|\varepsilon f_h^0\|_{L^2(\Omega)}$. Then the solution of the DG scheme* (4.12) *is stable in the following sense*

$$\|\varepsilon\hat{f}_h^{\mathfrak{n}+1}\|_{L^2(\Omega)} \le c^{\mathfrak{n}+1}\|\varepsilon f_0\|_{L^2(\Omega)} + \frac{1}{\chi_{\min}}(1-c^{\mathfrak{n}+1})\|\varepsilon\eta\|_{L^2(\Omega)},\tag{4.17}$$

where c is given in (2.20).

Proof. Setting $w_0 = f_S^{\mathfrak{n}+1} \equiv \hat{f}_h^{\mathfrak{n}+1}$ (see Definition 4.1) in (4.12c) gives

$$\left((1+\Delta t\chi)\hat{f}_h^{\mathfrak{n}+1}, \hat{f}_h^{\mathfrak{n}+1}; \varepsilon^2\right)_\Omega = (\hat{f}_h^{\mathfrak{n}}, \hat{f}_h^{\mathfrak{n}+1}; \varepsilon^2)_\Omega + \Delta t(\eta, \hat{f}_h^{\mathfrak{n}+1}; \varepsilon^2)_\Omega, \tag{4.18}$$

which, with Cauchy-Schwartz, leads to

$$\|\varepsilon \hat{f}_h^{\mathfrak{n}+1}\|_{L^2(\Omega)} \le c\|\varepsilon \hat{f}_h^{\mathfrak{n}}\|_{L^2(\Omega)} + c\Delta t\|\varepsilon\eta\|_{L^2(\Omega)}, \tag{4.19}$$

where c is given by (2.20). Applying (4.19) recursively gives

$$\begin{aligned}\|\varepsilon \hat{f}_h^{\mathfrak{n}+1}\|_{L^2(\Omega)} \le& c^{\mathfrak{n}+1}\|\varepsilon \hat{f}_h^{0}\|_{L^2(\Omega)} + \Delta t\|\varepsilon\eta\|_{L^2(\Omega)}\sum_{i=1}^{\mathfrak{n}+1} c^i \\ \le& c^{\mathfrak{n}+1}\|\varepsilon \hat{f}_h^{0}\|_{L^2(\Omega)} + \frac{1}{\chi_{\min}}(1-c^{\mathfrak{n}+1})\|\varepsilon\eta\|_{L^2(\Omega)}.\end{aligned} \tag{4.20}$$

Thus the estimate (4.17) follows from (4.20), the assumption on the initial data, and the fact that $\|\varepsilon f_h^0\|_{L^2(\Omega)} \le \|\varepsilon f_0\|_{L^2(\Omega)}$. □

4.3. **Convergence to the equilibrium distribution.** The convergence result in Proposition 2.4(ii) follows from the fact that the discrete equilibrium f_h^{Eq} is in the trial space of the fully discrete full-rank DG scheme (2.18). However, for the DLR-DG scheme, the space of trial functions may not contain the discrete equilibrium. In this subsection, we provide additional conditions to ensure convergence of the DLR-DG solution $\hat{f}_h^{\mathfrak{n}}$ to f_h^{Eq}. We first evaluate the error between the equilibrium solution f_h^{Eq} and its projection in the DLR-DG space and then investigate the convergence of $\hat{f}_h^{\mathfrak{n}}$ to this projection.

The equilibrium solution of the steady state equation (2.14) has the form $f_h^{\mathsf{Eq}} = X^\top(\mu)F^{\mathsf{Eq}}Y(\varepsilon)$, where $G(F^{\mathsf{Eq}}) = 0$ and G is given in (3.7). Equivalently,

$$F^{\mathsf{Eq}} = L_0(L_\eta)^\top(A_\chi)^{-1} \quad \text{and} \quad G = (F^{\mathsf{Eq}} - F)(A_\chi)^{-1}, \tag{4.21}$$

where the vectors L_0, L_η, and the matrix A_χ are given in Corollary 3.2. The matrix F^{Eq} in (4.21) is a rank-1 matrix that can be decomposed as

$$F^{\mathsf{Eq}} = U^{\mathsf{Eq}}S^{\mathsf{Eq}}(E^{\mathsf{Eq}})^\top, \tag{4.22}$$

where $U^{\mathsf{Eq}} \in \mathbb{R}^{m\times 1}$, $E^{\mathsf{Eq}} \in \mathbb{R}^{n\times 1}$, and $S^{\mathsf{Eq}} \in \mathbb{R}^{1\times 1}$ are given by

$$U^{\mathsf{Eq}} = \frac{L_0}{\|L_0\|}, \quad E^{\mathsf{Eq}} = \frac{(A_\chi)^{-1}L_\eta}{\|(L_\eta)^\top(A_\chi)^{-1}\|_{A_{\mathbf{1}}}}, \quad S^{\mathsf{Eq}} = \|L_0\|\|(L_\eta)^\top(A_\chi)^{-1}\|_{A_{\mathbf{1}}}. \tag{4.23}$$

The vectors U^{Eq} and E^{Eq} satisfy the orthogonality conditions $(U^{\mathsf{Eq}})^\top U^{\mathsf{Eq}} = 1$ and $(E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E^{\mathsf{Eq}} = 1$, and the scalar S^{Eq} satisfies the following estimate.

Lemma 4.5. *The scalar S^{Eq} is uniformly bounded in the following sense:*

$$|S^{\mathsf{Eq}}| = \|\varepsilon f_h^{\mathsf{Eq}}\|_{L^2(\Omega)} \le \chi_{\min}^{-1/2}\|\varepsilon\eta\|_{L^2(\Omega)}. \tag{4.24}$$

Proof. Setting $w_h = f_h^{\mathsf{Eq}}$ in (2.14), it is easy to show that

$$\chi_{\min}^{1/2}\|\varepsilon f_h^{\mathsf{Eq}}\|_{L^2(\Omega)} \le \|\varepsilon\chi^{1/2} f_h^{\mathsf{Eq}}\|_{L^2(\Omega)} \le \|\varepsilon\eta\|_{L^2(\Omega)}. \tag{4.25}$$

A direct calculation using Lemma A.1 gives $\|\varepsilon f_h^{\mathsf{Eq}}\|^2_{L^2(\Omega)} = \left(F^{\mathsf{Eq}}, F^{\mathsf{Eq}}\right)_{A_{\mathbf{1}}} = |S^{\mathsf{Eq}}|^2$ which, when substituted into (4.25), recovers the estimate (4.24). □

Let $f_S^{\mathsf{Eq},\mathfrak{n}}$ be the projection of f_h^{Eq} onto the DLR-DG space $V_0^{\mathfrak{n}}$ that is orthogonal with respect to the inner product $(\cdot,\cdot;\varepsilon^2)_\Omega$, defined in (2.6), that is

$$(f_S^{\mathsf{Eq},\mathfrak{n}}, w_h; \varepsilon^2)_\Omega = (f_h^{\mathsf{Eq}}, w_h; \varepsilon^2)_\Omega \quad \forall w_h \in V_0^{\mathfrak{n}}.$$

Taking the test functions $w_h = X^\top(\mu)U^{\mathfrak{n}}S(E^{\mathfrak{n}})^\top Y(\varepsilon),\ \forall S \in \mathbb{R}^{r\times r}$ implies that the projection $f_S^{\mathsf{Eq},\mathfrak{n}}$ has an expansion of the form

$$f_S^{\mathsf{Eq},\mathfrak{n}} = X^\top(\mu)P_{U^{\mathfrak{n}}}F^{\mathsf{Eq}}A_{\mathbf{1}}P_{E^{\mathfrak{n}}}Y(\varepsilon) \in V_0^{\mathfrak{n}}. \tag{4.26}$$

4.3.1. *Projection error of the equilibrium in the DLR-DG space.* The projection of U^{Eq} onto the columns of $U^{\mathfrak{n}}$ is $P_{U^{\mathfrak{n}}}U^{\mathsf{Eq}} := U^{\mathfrak{n}}(U^{\mathfrak{n}})^\top U^{\mathsf{Eq}}$, and the projection error is

$$\|U^{\mathsf{Eq}} - P_{U^{\mathfrak{n}}}U^{\mathsf{Eq}}\|^2 = 1 - \|P_{U^{\mathfrak{n}}}U^{\mathsf{Eq}}\|^2 = 1 - \|(U^{\mathfrak{n}})^\top U^{\mathsf{Eq}}\|^2 \in [0,1]. \tag{4.27}$$

Similarly, the (weighted) projection of E^{Eq} onto the space spanned by the columns of $E^{\mathfrak{n}}$ is $P_{E^{\mathfrak{n}}}A_{\mathbf{1}}E^{\mathsf{Eq}} := E^{\mathfrak{n}}(E^{\mathfrak{n}})^\top A_{\mathbf{1}}E^{\mathsf{Eq}}$, and the (weighted) projection error is

$$\|(E^{\mathsf{Eq}})^\top - (E^{\mathsf{Eq}})^\top A_{\mathbf{1}}P_{E^{\mathfrak{n}}}\|_{A_{\mathbf{1}}}^2 = 1 - \|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}}P_{E^{\mathfrak{n}}}\|_{A_{\mathbf{1}}}^2 = 1 - \|(E^{\mathfrak{n}})^\top A_{\mathbf{1}}E^{\mathsf{Eq}}\|^2 \in [0,1]. \tag{4.28}$$

Lemma 4.6 and Lemma 4.7 provide upper bounds for the projection errors in (4.28) and (4.27), respectively. Their proofs can be found in Appendix C.1 and Appendix C.2, respectively.

Lemma 4.6. *Assume that for some constant $\beta \in (0,1]$,*

$$\|P_{U^{\mathfrak{n}}}U^{\mathsf{Eq}}\| \geq \beta. \tag{4.29}$$

Then, for any $\delta > 0$ and any $\Delta t \geq \Delta t_1 = \frac{\sqrt{r}}{\beta\delta\chi_{\min}}$,

$$1 - \|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}}P_{E^{\mathfrak{n}+1}}\|_{A_{\mathbf{1}}}^2 \leq \frac{\delta^2}{|S^{\mathsf{Eq}}|^2}\|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}^2. \tag{4.30}$$

Moreover, if $\hat{f}_h^{\mathfrak{n}} = f_h^{\mathsf{Eq}}$, then for any $\Delta t > 0$,

$$\|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}}P_{E^{\mathfrak{n}+1}}\|_{A_{\mathbf{1}}} = 1. \tag{4.31}$$

Define the symmetric matrix

$$P_E^\chi = E(E^TA_\chi E)^{-1}E^T. \tag{4.32}$$

Then $P_E^\chi A_\chi$ is the orthogonal projection onto the column space of E with respect to the inner product on $\mathbb{R}^n$ with weight A_χ.

Lemma 4.7. *Assume there exists a constant $\alpha > 0$ such that*

$$\|(E^{\mathsf{Eq}})^\top A_\chi P_{E^{\mathfrak{n}}}^\chi\|_{A_{\mathbf{1}}} \geq \alpha. \tag{4.33}$$

Then for any $\delta > 0$ and any $\Delta t \geq \Delta t_2 = \frac{r^{1/2}\chi_{\max}^{1/2}}{\alpha\delta\chi_{\min}^{3/2}}$,

$$1 - \|P_{U^{\mathfrak{n}+1}}U^{\mathsf{Eq}}\| \leq \frac{\delta^2}{|S^{\mathsf{Eq}}|^2}\|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}^2. \tag{4.34}$$

Moreover, if $\hat{f}_h^{\mathfrak{n}} = f_h^{\mathsf{Eq}}$, then for any $\Delta t > 0$,

$$\|P_{U^{\mathfrak{n}+1}}U^{\mathsf{Eq}}\| = 1. \tag{4.35}$$

Lemma 4.6 and Lemma 4.7 can be used to bound the projection error of the equilibrium with respect to $U^{\mathfrak{n}+1}$ and $E^{\mathfrak{n}+1}$.

Lemma 4.8. *Assume there exist constants $\beta \in (0,1]$ and $\alpha > 0$ such that $\|P_{U^{\mathfrak{n}}} U^{\mathsf{Eq}}\| \geq \beta$ and $\|(E^{\mathsf{Eq}})^\top A_\chi P_{E^{\mathfrak{n}}}^\chi\|_{A_{\mathbf{1}}} \geq \alpha$. Then for any $\delta > 0$, there exists*

$$\Delta t_0 = \frac{\sqrt{2}}{\delta} \max\left\{ \frac{r^{1/2}}{\beta \chi_{\min}}, \frac{r^{1/2} \chi_{\max}^{1/2}}{\alpha \chi_{\min}^{3/2}} \right\} \tag{4.36}$$

such that when $\Delta t \geq \Delta t_0$,

$$\|\varepsilon(f_S^{\mathsf{Eq},\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \leq \delta \|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}. \tag{4.37}$$

Moreover, if $\hat{f}_h^{\mathfrak{n}} = f_h^{\mathsf{Eq}}$, it follows that for any $\Delta t > 0$,

$$f_S^{\mathsf{Eq},\mathfrak{n}+1} = f_h^{\mathsf{Eq}}. \tag{4.38}$$

Proof. By (4.22), Lemma 3.1, and Lemma A.1,

(4.39)

$$\begin{aligned}
&\|\varepsilon(f_h^{\mathsf{Eq}} - f_S^{\mathsf{Eq},\mathfrak{n}+1})\|_{L^2(\Omega)}^2 = \|F^{\mathsf{Eq}} - P_{U^{\mathfrak{n}+1}} F^{\mathsf{Eq}} A_{\mathbf{1}} P_{E^{\mathfrak{n}+1}}\|_{A_{\mathbf{1}}}^2 \\
&= |S^{\mathsf{Eq}}|^2 \|(I - P_{U^{\mathfrak{n}+1}}) U^{\mathsf{Eq}} (E^{\mathsf{Eq}})^\top + P_{U^{\mathfrak{n}+1}} U^{\mathsf{Eq}} (E^{\mathsf{Eq}})^\top (I - A_{\mathbf{1}} P_{E^{\mathfrak{n}+1}})\|_{A_{\mathbf{1}}}^2 \\
&= |S^{\mathsf{Eq}}|^2 \left[\|(I - P_{U^{\mathfrak{n}+1}}) U^{\mathsf{Eq}}\|^2 + \|P_{U^{\mathfrak{n}+1}} U^{\mathsf{Eq}} (E^{\mathsf{Eq}})^\top (I - A_{\mathbf{1}} P_{E^{\mathfrak{n}+1}})\|_{A_{\mathbf{1}}}^2 \right] \\
&= |S^{\mathsf{Eq}}|^2 \left[\left(1 - \|P_{U^{\mathfrak{n}+1}} U^{\mathsf{Eq}}\|^2\right) + \|P_{U^{\mathfrak{n}+1}} U^{\mathsf{Eq}}\|^2 \left(1 - \|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}} P_{E^{\mathfrak{n}+1}}\|_{A_{\mathbf{1}}}^2\right) \right],
\end{aligned}$$

where orthogonality is used in obtaining the third equality.

By Lemma 4.6 and Lemma 4.7 (with δ being replaced by $\delta/\sqrt{2}$), it follows that when $\Delta t \geq \Delta t_0$, where Δt_0 is given by Eq. (4.36), the following estimates hold

$$1 - \|P_{U^{\mathfrak{n}+1}} U^{\mathsf{Eq}}\|^2 \leq \frac{\delta^2}{2|S^{\mathsf{Eq}}|^2} \|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}^2, \tag{4.40a}$$

$$1 - \|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}} P_{E^{\mathfrak{n}+1}}\|_{A_{\mathbf{1}}}^2 \leq \frac{\delta^2}{2|S^{\mathsf{Eq}}|^2} \|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}^2, \tag{4.40b}$$

which, when substituted into (4.39), yields (4.37). Then (4.38) follows from (4.39), using (4.31) and (4.35). □

Remark 4.9. If $\|P_{U^{\mathfrak{n}}} U^{\mathsf{Eq}}\| = \|(U^{\mathfrak{n}})^\top U^{\mathsf{Eq}}\| = 1$, then $U^{\mathsf{Eq}} = U^{\mathfrak{n}} z$ for some vector $z \in \mathbb{R}^{r \times 1}$, and (4.8) implies that the ***K*-step** can be omitted for Algorithm 4.1 by simply taking $U^{\mathfrak{n}+1} = U^{\mathfrak{n}}$. Then for any $\delta > 0$, there exists $\Delta t_0 = \frac{r^{1/2}}{\delta \chi_{\min}}$, such that when $\Delta t \geq \Delta t_0$, (4.37) holds.

4.3.2. *Convergence of the DLR-DG solution to the equilibrium.* We estimate the convergence of the DLR-DG solution $\hat{f}_h^{\mathfrak{n}+1}$ to the equilibrium f_h^{Eq}. We first provide a one-step estimate.

Theorem 4.10. *Suppose the assumptions in Lemma* 4.8 *hold. For any $\delta > 0$, let Δt_0 be given in* (4.36)*. Then for any $\Delta t \geq \Delta t_0$,*

$$\|\varepsilon(\hat{f}_h^{\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \leq (c + \delta_\chi) \|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}, \tag{4.41}$$

where c is given by (2.20) *and $\delta_\chi = \left(1 + \frac{\chi_{\max}}{\chi_{\min}}\right) \delta$. Moreover, if $\hat{f}_h^{\mathfrak{n}} = f_h^{\mathsf{Eq}}$, then for any $\Delta t > 0$,*

$$\hat{f}_h^{\mathfrak{n}+1} = f_h^{\mathsf{Eq}}. \tag{4.42}$$

Proof. Since $\eta(\varepsilon) = \chi(\varepsilon) f^{\mathsf{Eq}}(\varepsilon)$, the DG scheme (4.12c) can be written using ((4.11) and (2.14) as

$$\left((1+\Delta t\chi) f_S^{\mathfrak{n}+1}, w_h; \varepsilon^2\right)_\Omega = \left(\Delta t \chi f_h^{\mathsf{Eq}} + f_S^{\mathfrak{n},*}, w_h; \varepsilon^2\right)_\Omega \quad \forall w_h \in V_0^{\mathfrak{n}+1}. \tag{4.43}$$

Subtracting $((1+\Delta t\chi) f_S^{\mathsf{Eq},\mathfrak{n}+1}, w_h; \varepsilon^2)_\Omega$ from (4.43) yields

$$\begin{aligned} &\left((1+\Delta t\chi)(f_S^{\mathfrak{n}+1} - f_S^{\mathsf{Eq},\mathfrak{n}+1}), w_h; \varepsilon^2\right)_\Omega \\ &\qquad = \left(\Delta t\chi(f_h^{\mathsf{Eq}} - f_S^{\mathsf{Eq},\mathfrak{n}+1}) + (f_S^{\mathfrak{n},*} - f_S^{\mathsf{Eq},\mathfrak{n}+1}), w_h; \varepsilon^2\right)_\Omega. \end{aligned} \tag{4.44}$$

Setting $w_h = f_S^{\mathfrak{n}+1} - f_S^{\mathsf{Eq},\mathfrak{n}+1} \in V_0^{\mathfrak{n}+1}$ in (4.44) and applying the Cauchy–Schwarz inequality gives

$$\|\varepsilon(f_S^{\mathfrak{n}+1} - f_S^{\mathsf{Eq},\mathfrak{n}+1})\|_{L^2(\Omega)} \le c\|\varepsilon(f_S^{\mathfrak{n},*} - f_S^{\mathsf{Eq},\mathfrak{n}+1})\|_{L^2(\Omega)} + c\Delta t\|\varepsilon\chi(f_S^{\mathsf{Eq},\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}, \tag{4.45}$$

where c is given by (2.20). By the triangle inequality and (4.45),

$$\begin{aligned} &\|\varepsilon(f_S^{\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \\ &\quad \le \|\varepsilon(f_S^{\mathfrak{n}+1} - f_S^{\mathsf{Eq},\mathfrak{n}+1})\|_{L^2(\Omega)} + \|\varepsilon(f_S^{\mathsf{Eq},\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \\ &\quad \le c\|\varepsilon(f_S^{\mathfrak{n},*} - f_S^{\mathsf{Eq},\mathfrak{n}+1})\|_{L^2(\Omega)} + (1 + c\Delta t\chi_{\max})\|\varepsilon(f_S^{\mathsf{Eq},\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}. \end{aligned} \tag{4.46}$$

Additionally, $f_S^{\mathfrak{n},*}$ and $f_S^{\mathsf{Eq},\mathfrak{n}+1}$ are both L^2 projections of $\hat{f}_h^{\mathfrak{n}}$ and f_h^{Eq} onto $V_0^{\mathfrak{n}+1}$; therefore

$$\|\varepsilon(f_S^{\mathfrak{n},*} - f_S^{\mathsf{Eq},\mathfrak{n}+1})\|_{L^2(\Omega)} \le \|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}. \tag{4.47}$$

By (4.46), (4.47), and Definition 4.1, the stability estimate follows:

$$\begin{aligned} &\|\varepsilon(\hat{f}_h^{\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \\ &\qquad \le c\|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} + \left(1 + \frac{\chi_{\max}}{\chi_{\min}}\right)\|\varepsilon(f_S^{\mathsf{Eq},\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}. \end{aligned} \tag{4.48}$$

If $\Delta t \ge \Delta t_0$, then (4.37) holds and, when substituted into (4.48), gives (4.41). If $\hat{f}_h^{\mathfrak{n}} = f_h^{\mathsf{Eq}}$, then (4.38) and (4.48) imply (4.42). □

Remark 4.11. To obtain the one step projection error estimate (4.41) in Theorem 4.10, we can set $\beta = \|P_{U^{\mathfrak{n}}} U^{\mathsf{Eq}}\|$ in Lemma 4.6 and $\alpha = \|(E^{\mathsf{Eq}})^\top A_\chi P_{E^{\mathfrak{n}}}^\chi\|_{A_{\mathbf{1}}}$ in Lemma 4.7, and these values can be calculated from η and χ. For the multi-step error estimate, α and β are determined by the initial bases U^0 and E^0, as described in Theorem 4.13. These conditions are numerically computed in Example 5.2.

Unlike the full-rank case, the one-step estimate in Theorem 4.10 cannot be trivially extended to a multi-step estimate. This is because of the disconnect between the conclusion of Lemma 4.6 and the hypothesis of Lemma 4.7, which bound the projection with respect to the $A_{\mathbf{1}}$- and A_χ-inner products respectively. In order to bootstrap the one-step estimate further, we require Lemma 4.12 which controls $\|(E^{\mathsf{Eq}})^\top A_\chi P_{E^{\mathfrak{n}}}^\chi\|_{A_{\mathbf{1}}}$ by $\|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}} P_{E^{\mathfrak{n}}}\|_{A_{\mathbf{1}}}$, where $P_{E^{\mathfrak{n}}}^\chi$ is defined in (4.32). This estimate depends on $\frac{\chi_{\max}}{\chi_{\min}}$, the weighted condition number of A_χ.

Lemma 4.12. *For any $\alpha \in (0,1)$, there exists $\gamma^* \in (\alpha, 1)$, dependent only on $\frac{\chi_{\max}}{\chi_{\min}}$ and α, such that if $E \in \mathbb{R}^{n\times r}$ with $E^T A_{\mathbf{1}} E = I_r$ and $\|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}} P_E\|_{A_{\mathbf{1}}} \ge \gamma^*$, then $\|(E^{\mathsf{Eq}})^\top A_\chi P_E^\chi\|_{A_{\mathbf{1}}} \ge \alpha$.*

Proof. Decompose E^{Eq} as

$$E^{\mathsf{Eq}} = E_1^{\mathsf{Eq}} + E_2^{\mathsf{Eq}}, \tag{4.49}$$

where $E_1^{\mathsf{Eq}} = P_E A_{\mathbf{1}} E^{\mathsf{Eq}}$ is the orthogonal projection of E^{Eq} onto the column space of E and $E_2^{\mathsf{Eq}} = P_E^{\perp} A_{\mathbf{1}} E^{\mathsf{Eq}}$ is the orthogonal complement satisfying $\|(E_1^{\mathsf{Eq}})^\top\|_{A_{\mathbf{1}}}^2 + \|(E_2^{\mathsf{Eq}})^\top\|_{A_{\mathbf{1}}}^2 = 1$. Since $P_E^\chi A_\chi$ is also a projection onto the column space of E,

$$P_E^\chi A_\chi E_1^{\mathsf{Eq}} = E_1^{\mathsf{Eq}}. \tag{4.50}$$

Suppose $\|(E_1^{\mathsf{Eq}})^\top\|_{A_{\mathbf{1}}} =: \gamma \in (\alpha, 1]$. By Lemma C.5,

$$\|(E_2^{\mathsf{Eq}})^\top A_\chi P_E^\chi\|_{A_{\mathbf{1}}}^2 \le \frac{\chi_{\max}}{\chi_{\min}} \|(E_2^{\mathsf{Eq}})^\top\|_{A_{\mathbf{1}}}^2 = (1-\gamma^2)\frac{\chi_{\max}}{\chi_{\min}}. \tag{4.51}$$

By (4.49), (4.50), (4.51), Hölder's and Young's inequalities, for any $\tau(\gamma) \in (0,1)$,

$$\begin{aligned}
&\|(E^{\mathsf{Eq}})^\top A_\chi P_E^\chi\|_{A_{\mathbf{1}}}^2 \\
&\quad = \|(E_1^{\mathsf{Eq}})^\top A_\chi P_E^\chi + (E_2^{\mathsf{Eq}})^\top A_\chi P_E^\chi\|_{A_{\mathbf{1}}}^2 = \|(E_1^{\mathsf{Eq}})^\top + (E_2^{\mathsf{Eq}})^\top A_\chi P_E^\chi\|_{A_{\mathbf{1}}}^2 \\
&\quad = \|(E_1^{\mathsf{Eq}})^\top\|_{A_{\mathbf{1}}}^2 + \|(E_2^{\mathsf{Eq}})^\top A_\chi P_E^\chi\|_{A_{\mathbf{1}}}^2 + 2\left((E_1^{\mathsf{Eq}})^\top, (E_2^{\mathsf{Eq}})^\top A_\chi P_E^\chi\right)_{A_{\mathbf{1}}} \\
&\quad \ge \|(E_1^{\mathsf{Eq}})^\top\|_{A_{\mathbf{1}}}^2 (1-\tau(\gamma)) + \|(E_2^{\mathsf{Eq}})^\top A_\chi P_E^\chi\|_{A_{\mathbf{1}}}^2 \left(1 - \frac{1}{\tau(\gamma)}\right) \\
&\quad \ge \gamma^2(1-\tau(\gamma)) + (1-\gamma^2)\frac{\chi_{\max}}{\chi_{\min}}\left(1 - \frac{1}{\tau(\gamma)}\right) =: g(\gamma).
\end{aligned} \tag{4.52}$$

Let $\tau(\gamma) = \frac{1}{2}(1 - \frac{\alpha^2}{\gamma^2})$. Then for every $\gamma \in (\alpha, 1]$, τ satisfies $0 < \tau(\gamma) < 1$ and $1 - \tau(\gamma) = \frac{1}{2} + \frac{\alpha^2}{\gamma^2} > \frac{\alpha^2}{\gamma^2}$. Since g is continuous at $\gamma = 1$ and $g(1) = 1 - \tau(1) > \alpha^2$, there exists $\gamma^* \in (\alpha, 1)$, dependent on α and $\frac{\chi_{\max}}{\chi_{\min}}$, such that for any $\gamma^* \le \gamma \le 1$, $g(\gamma) \ge \alpha^2$. Therefore by (4.52) the result follows. □

We now have the following multi-step estimate.

Theorem 4.13. *Assume there exist constants $\beta \in (0,1)$ and $\alpha \in (0,1)$ such that $\|P_{U^0} U^{\mathsf{Eq}}\| \ge \beta$ and $\|(E^{\mathsf{Eq}})^\top A_\chi P_{E^0}^\chi\|_{A_{\mathbf{1}}} \ge \alpha$. Let $\gamma^* \in (\alpha, 1)$ be given in Lemma* 4.12. *Then for any*

$$0 < \delta < \min\left\{(1-c)\left(1 + \frac{\chi_{\max}}{\chi_{\min}}\right)^{-1}, \frac{\sqrt{2(1-\max\{\gamma^*, \beta\}^2)}\|\varepsilon f_h^{\mathsf{Eq}}\|_{L^2(\Omega)}}{\|\varepsilon(\hat{f}_h^0 - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}}\right\}, \tag{4.53}$$

and Δt_0 given in Theorem 4.10, *when $\Delta t \ge \Delta t_0$,*

$$\|\varepsilon(\hat{f}_h^{\mathfrak{n}+1} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \le (c + \delta_\chi)^{\mathfrak{n}+1} \|\varepsilon(\hat{f}_h^0 - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \quad \forall \mathfrak{n} \ge 1, \tag{4.54}$$

where c is given by (2.20) *and $\delta_\chi = \left(1 + \frac{\chi_{\max}}{\chi_{\min}}\right)\delta$. Moreover, if $\hat{f}_h^0 = f_h^{\mathsf{Eq}}$, then for any $\Delta t > 0$, $\hat{f}_h^{\mathfrak{n}+1} = f_h^{\mathsf{Eq}}$.*

Proof. We prove the result by the method of induction. For $\mathfrak{n} = 0$, (4.54) follows from the one-step result in Theorem 4.10; see (4.41). We assume that (4.54) holds for some $\mathfrak{n} \ge 1$, that is

$$\|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)} \le (c + \delta_\chi)^{\mathfrak{n}} \|\varepsilon(\hat{f}_h^0 - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}. \tag{4.55}$$

By (4.53), $(c+\delta_\chi)<1$; thus $\|\varepsilon(\hat{f}_h^{\mathfrak{n}}-f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}\le\|\varepsilon(\hat{f}_h^{0}-f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}$. Then for $\mathfrak{n}+1$, the bounds in (4.40), the fact that $|S^{\mathsf{Eq}}|=\|\varepsilon f_h^{\mathsf{Eq}}\|_{L^2(\Omega)}$ (see Lemma 4.5), and the definition of δ in (4.53) imply that

$$\|P_{U^{\mathfrak{n}+1}}U^{\mathsf{Eq}}\|^2\ge 1-\frac{\delta^2}{2\|\varepsilon f_h^{\mathsf{Eq}}\|^2_{L^2(\Omega)}}\|\varepsilon(\hat{f}_h^{\mathfrak{n}}-f_h^{\mathsf{Eq}})\|^2_{L^2(\Omega)}\ge\beta^2, \tag{4.56a}$$

$$\|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}}P_{E^{\mathfrak{n}+1}}\|^2_{A_{\mathbf{1}}}\ge 1-\frac{\delta^2}{2\|\varepsilon f_h^{\mathsf{Eq}}\|^2_{L^2(\Omega)}}\|\varepsilon(\hat{f}_h^{\mathfrak{n}}-f_h^{\mathsf{Eq}})\|^2_{L^2(\Omega)}\ge(\gamma^*)^2. \tag{4.56b}$$

By Lemma 4.12, (4.56b) implies $\|(E^{\mathsf{Eq}})^\top A_\chi P^\chi_{E^{\mathfrak{n}+1}}\|_{A_{\mathbf{1}}}\ge\alpha$. Therefore, the one-step estimate (4.41) holds. The estimate (4.54) then follows from (4.41) and (4.55). Finally, if $\hat{f}_h^0=f_h^{\mathsf{Eq}}$, by (4.42), $\hat{f}_h^{\mathfrak{n}+1}=\hat{f}_h^{\mathfrak{n}}=\ldots=\hat{f}_h^0=f_h^{\mathsf{Eq}}$. □

5. Numerical results

In this section, we present numerical examples to validate our theoretical findings. For all the numerical tests in this section, we construct initial data $\hat{F}(0)=U^0S^0(E^0)^\top\in\mathcal{M}_r$ for Algorithm 4.1 by applying the generalized singular value decomposition (GSVD) [1] (Algorithm B.2) to $F(0)$, followed by truncation.

Example 5.1. In this example, we test the performance of the dynamical low-rank DG scheme in (4.12), or Algorithm 4.1, by comparing with the full-rank DG scheme in (2.18). We let $\varepsilon_{\max}=1$, and set the opacity $\chi(\varepsilon)=4+\frac{\varepsilon^2}{2}$ and the emissivity $\eta(\varepsilon)=f^{\mathsf{Eq}}(\varepsilon)\chi(\varepsilon)$, where

$$f^{\mathsf{Eq}}(\varepsilon)=\frac{1}{\varepsilon^2+1} \tag{5.1}$$

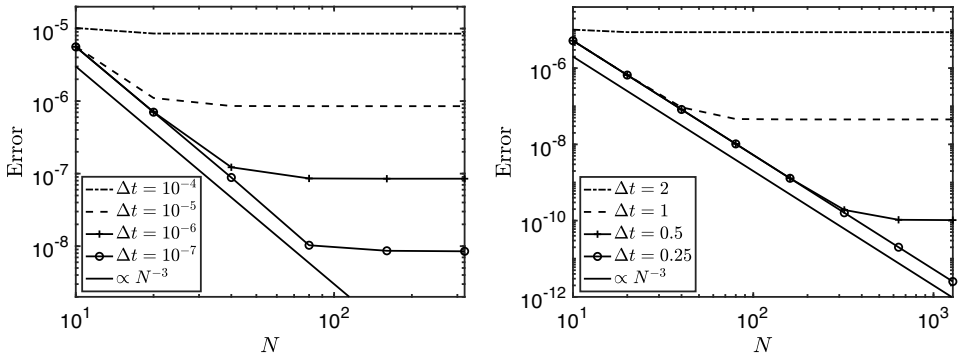
is the rank-1 equilibrium distribution. With initial data $f(\mu,\varepsilon,0)=\frac{1}{\varepsilon^2+1}+\frac{1}{\mu^2+\varepsilon^2+1/2}$, the exact solution to (2.2) is

$$f(\mu,\varepsilon,t)=\frac{1}{\varepsilon^2+1}+\frac{1}{\mu^2+\varepsilon^2+1/2}e^{-\chi(\varepsilon)t}. \tag{5.2}$$

We use $\mathcal{Q}_2$ polynomials for all the tests in this example.

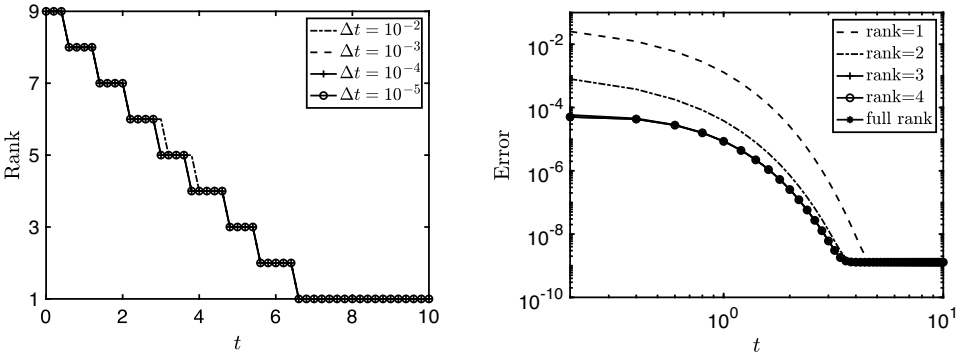
To establish a baseline, we first test the spatial and temporal accuracy of the full-rank DG scheme in (2.18) with $N=N_\mu=N_\varepsilon$ cells in each direction. Errors at $t=1$ are shown in Figure 1(a). The convergence rate of the full rank DG scheme (2.18) is first-order in time (as expected with backward Euler time stepping) and third-order in phase-space (as expected with $\mathcal{Q}_2$ polynomials) until saturation due to the temporal error. Errors at $t=10$ are shown in Figure 1(b). In this case, the phase-space convergence rate is still third-order for sufficiently small Δt, but the temporal accuracy is super linear due to the fact that the solution is very near the time-independent equilibrium distribution. Thus the error follows the bound in (2.21), which decreases geometrically.

Second, we show the evolution of the rank of the coefficient matrix F^n for the full-rank DG scheme (2.18), using a mesh with $N_\mu=N_\varepsilon=160$. The numerical rank is calculated with the Matlab function `rank`$(F^n,\ 10^{-12})$, which returns the total number of singular values of F^n that are larger than 10^{-12}. The results with different time steps are plotted in Figure 2(a). We observe that the numerical rank of the coefficient matrix decreases from $r=9$ at the initial condition to $r=1$ as the solution approaches equilibrium.



(a) Errors at $t = 1$ (b) Errors at $t = 10$

FIGURE 1. Error, $\|\varepsilon(f - f_h^{\mathrm{n}})\|_{L^2(\Omega)}$, for the full-rank DG scheme in (2.18) versus number of elements, $N = N_\mu = N_\varepsilon$, for two different time step sizes. The scheme uses $\mathcal{Q}_2$ polynomials in phase-space and backward Euler time stepping. In each panel, the solid lines without symbols are reference lines proportional to N^{-3}.



(a) (b)

FIGURE 2. (a) Evolution of the numerical rank for the coefficient matrix F^n of the full-rank DG scheme, plotted vs. time using various time step sizes. (b) Weighted L^2 errors of the DLR-DG method (using $r = 1$, 2, 3, and 4) and the full rank DG scheme (2.18) relative to the exact solution versus time, using $\Delta t = 10^{-4}$ and $N_\mu \times N_\varepsilon = 160 \times 160$.

Third, we solve (2.2) using both the DLR-DG scheme in Algorithm 4.1 and the full-rank DG scheme (2.18). The purpose of this test is to compare the DLR-DG solution with the solution of the full-rank DG scheme (2.18) as the rank r in Algorithm 4.1 increases. The L^2 errors of the numerical solutions are plotted in Figure 2(b). These errors decrease as the rank r increases. In particular, Algorithm 4.1 with $r = 3$ and time step $\Delta t = 10^{-4}$ produces numerical solutions that are practically identical to that of the full-rank scheme (2.18). All low-rank solutions eventually give accurate equilibrium approximations.

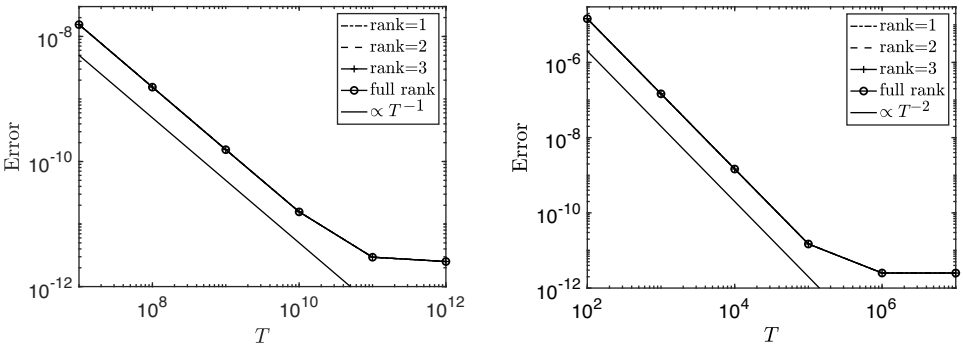


FIGURE 3. Weighted L^2 errors of the dynamical low-rank DG method (with $r = 1$, 2, and 3) and the full-rank DG scheme relative to the exact solution versus final time T, computed with one time step (with $\Delta t = T$; left panel) and two time steps (with $\Delta t = T/2$; right panel)

Fourth, we test the convergence of the dynamical low-rank DG solution and the full-rank DG solution to the equilibrium with one time step $\Delta t = T$, and two time steps $\Delta t = T/2$ for some final time T. The L^2 error of the numerical solution as a function of T is plotted in Figure 3. The results show that both algorithms converge up to discretization error, and the convergence rates of both algorithms to the equilibrium are equal to the total number of the time steps (i.e., $\propto T^{-1}$ for one step and $\propto T^{-2}$ for two steps), which is consistent with the theoretical results of Theorem 4.13, regarding the low-rank scheme, and Proposition 2.4(ii), regarding the full-rank scheme. The L^2 error saturates for large T, when it becomes dominated by the projection error of the equilibrium (around 10^{-12}).

Finally, we test the convergence of the dynamical low-rank DG solution and the full-rank DG solution to the equilibrium after $\mathfrak{n}$ steps, using two different time step sizes: $\Delta t = 2$ and $\Delta t = 10$. We show the L^2 error between the numerical solution and the discrete equilibrium f_h^{Eq} versus $\mathfrak{n}$ in Figure 4. The results show that both algorithms converge with convergence rates equal to the decay rate $c = \frac{1}{1+\Delta t \chi_{\min}} = \frac{1}{1+4\Delta t}$ (i.e., $\propto 9^{-1}$ for $\Delta t = 2$ and $\propto 41^{-1}$ for $\Delta t = 10$), which is consistent with the theoretical results of Theorem 4.13, regarding the low-rank scheme, and Proposition 2.4(ii), regarding full-rank scheme.

Example 5.2. The purpose of this example is to demonstrate how the condition given in Theorem 4.10 affects the convergence of the DLR-DG solution to the equilibrium. We solve (2.2) with the same parameters as in Example 5.1, but with $\mathcal{Q}_1$ polynomials and different initial conditions. The equilibrium is given in (5.1) and is independent of the initial data.

To construct different initial conditions, we first prepare some basis functions.

(i) Let $\mathbf{K}^0 = [U^{\mathsf{Eq}}, \widetilde{U}^0]$, and $\mathbf{L}^0 = [E^{\mathsf{Eq}}, \widetilde{E}^0]$, where U^{Eq}, E^{Eq} are given in (4.23), and $\widetilde{U}^0$ and $\widetilde{E}^0$ are rank-2 matrices, computed from Algorithm 4.1 using the initial data from Example 5.1.

(ii) Perform a QR factorization to obtain $\mathbf{K}^0 = \hat{U}^0 R_U^0$, where $\hat{U}^0 = [U^{\mathsf{Eq}}, \hat{U}_2^0, \hat{U}_3^0]$.

(iii) Perform an $A_{\mathbf{1}}$-weighted Gram–Schmidt decomposition to obtain $\mathbf{L}^0 = \hat{E}^0 R_E^0$, where $\hat{E}^0 = [E^{\mathsf{Eq}}, \hat{E}_2^0, \hat{E}_3^0]$.

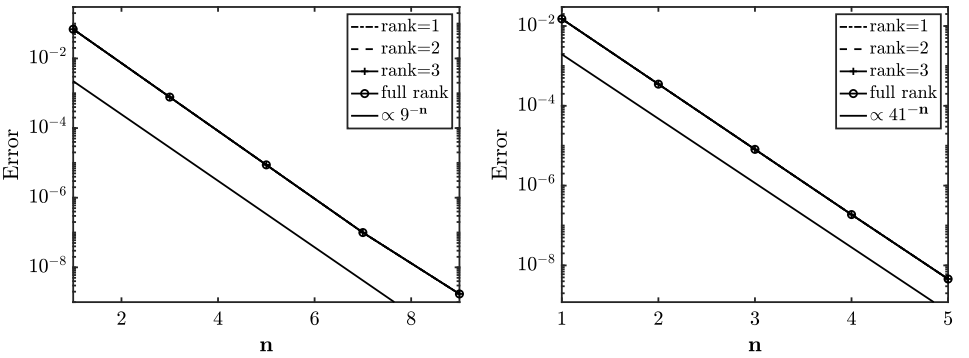


FIGURE 4. Weighted L^2 errors of the dynamical low-rank DG method (with $r = 1$, 2, and 3) and the full-rank DG scheme relative to the exact solution versus the total number of steps $\mathfrak{n}$, computed with ($\Delta t = 2$; left panel) and ($\Delta t = 10$; right panel). In both panels, we compare the numerical results with the predicted decay rate $c = \frac{1}{1+\Delta t \chi_{\min}} = \frac{1}{1+4\Delta t}$.

(iv) Generate $\check{U} = \frac{U^{\mathsf{Eq}} + \hat{U}_2^0}{\|U^{\mathsf{Eq}} + \hat{U}_2^0\|}$ and $\check{E} = \frac{E^{\mathsf{Eq}} + \hat{E}_2^0}{\|E^{\mathsf{Eq}} + \hat{E}_2^0\|_{A_{\mathbf{1}}}}$.

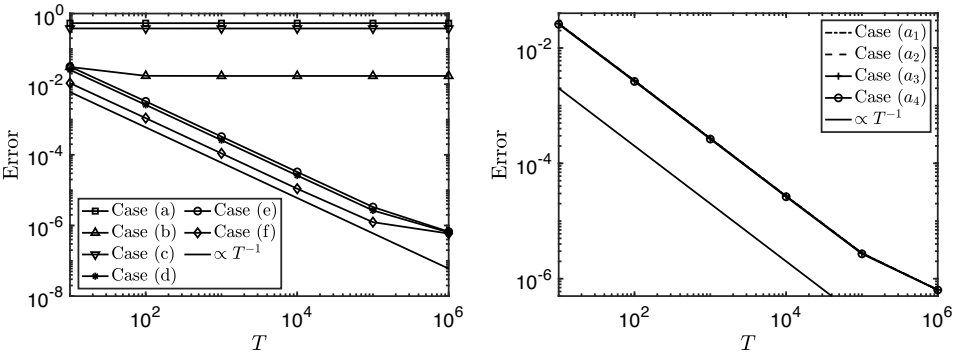
(v) Perform an A_χ-weighted Gram–Schmidt decomposition to obtain $\mathbf{L}^0 = \tilde{E}^0 \tilde{R}_E^0$, where $\tilde{E}^0 = [\check{E}^{\mathsf{Eq}}, \tilde{E}_2^0, \tilde{E}_3^0]$. Then perform an $A_{\mathbf{1}}$-weighted Gram–Schmidt decomposition to obtain $[\tilde{E}_2^0, \tilde{E}_3^0] = [\bar{E}_2^0, \bar{E}_3^0]\bar{R}_E^0$. Here, we expect $\|(\tilde{E}_j^0)^\top A_\chi E^{\mathsf{Eq}}\|$ to be close to zero for $j = 2, 3$.

Test Case 5.2-1. We use these different matrices to construct the various initial conditions given in the second and third rows of Table 1, with $S^0 = 1$. We solve (2.2) with rank-1 initial conditions given in Table 1, using Algorithm 4.1 with $r = 1$, $\mathcal{Q}_1$ polynomials, and a mesh size of $N_\mu = N_\varepsilon = 160$. We show the one time step ($\Delta t = T$) convergence of the dynamical low-rank DG solution to the equilibrium in Figure 5(a). In Table 1, we show the initial basis U^0 and E^0, the values in (4.29) and (4.33), whether the assumptions of Theorem 4.10 are satisfied (✓) or not (✗), and whether the scheme converges to the equilibrium (C) or not (NC). For Cases (a)-(c) in Table 1, the conditions for convergence in Theorem 4.10 are not satisfied, and the corresponding solution in Figure 5(a) does not converge to the equilibrium. Case (d) is a special case that is addressed in Remark 4.9. Specifically, $\|(E^{\mathsf{Eq}})^\top A_\chi P_{E^0}^\chi\|_{A_{\mathbf{1}}}$ is zero (to algorithmic precision) and hence does not satisfy the associated condition in Lemma 4.8. However, because $\left\|P_{U^0} U^{\mathsf{Eq}}\right\| = 1$, the corresponding numerical solution in Figure 5(a) still converges to the equilibrium with a first-order convergence rate. Cases (e)-(f) satisfy the conditions of Theorem 4.10, and, as expected, the corresponding numerical solutions in Figure 5(a) converge to the equilibrium solution with a first-order convergence rate. All these results indicate that the conditions given in Theorem 4.10, or Remark 4.9, are sufficient to determine the convergence of the one-step DLR-DG solution to the equilibrium.

Test Case 5.2-2. Though the conditions for convergence in Theorem 4.10 are not satisfied for Cases (a)-(c), the initial bases can be manually adjusted to yield a convergent algorithm. In the following, we take Case (a) in Table 1 as an example and show how to modify Algorithm 4.1 such that the solution converges to the

TABLE 1. Initial bases for Algorithm 4.1 used in Test Case 5.2-1, the values for the conditions in (4.29) and (4.33), whether the conditions of Theorem 4.10 are satisfied (✓) or not (✗), and the observed numerical behavior: Convergence (C) or no convergence (NC)

	Case (a)	Case (b)	Case (c)	Case (d)	Case (e)	Case (f)
U^0	$\hat{U}_2^0$	$\hat{U}_2^0$	$\check{U}$	U^{Eq}	$\check{U}$	U^{Eq}
E^0	$\bar{E}_2^0$	E^{Eq}	$\bar{E}_2^0$	$\bar{E}_2^0$	$\check{E}$	E^{Eq}
$\left\|P_{U^0}U^{\mathsf{Eq}}\right\|$	0	0	$\sqrt{2}/2$	1	$\sqrt{2}/2$	1
$\|(E^{\mathsf{Eq}})^{\dagger}A_\chi P_{E^0}^{\chi}\|_{A_\mathbf{1}}$	1.7699e-15	1	1.7699e-15	1.7699e-15	0.7119	1
Theorem 4.13	✗	✗	✗	✓(Remark 4.9)	✓	✓
Figure 5(a)	NC	NC	NC	C	C	C



(a) Example 5.2 Test Case 5.2-1 (b) Example 5.2 Test Case 5.2-2

FIGURE 5. The weighted L^2 errors between the dynamical low-rank DG solution and the equilibrium solution based on $\mathcal{Q}_1$ polynomials and $N_\mu = N_\varepsilon = 160$

equilibrium. (Similar modifications can be applied to Case (b) and Case (c).) To achieve convergence, we increase the rank to $r = 2$ and append the basis such that the conditions in Theorem 4.10 or Remark 4.9 are satisfied. Let x and y be scalar parameters (not both zero), and define the functions

$$U_\perp(x,y) := \frac{xU^{\mathsf{Eq}} + y\hat{U}_3^0}{\|xU^{\mathsf{Eq}} + y\hat{U}_3^0\|} \quad \text{and} \quad E_\perp(x,y) := \frac{xE^{\mathsf{Eq}} + y\bar{E}_3^0}{\|xE^{\mathsf{Eq}} + y\bar{E}_3^0\|_{A_\mathbf{1}}}. \tag{5.3}$$

Then $\{\hat{U}_2^0, U_\perp(x,y)\}$ and $\{\bar{E}_2^0, E_\perp(x,y)\}$ are orthonormal and $A_\mathbf{1}$-orthonormal bases, respectively.

We use $U_\perp$ and $E_\perp$ to generate different initial bases for Algorithm 4.1; these are listed as Cases (a_1)-(a_3) in Table 2. Case (a_4) is different: we randomly generate the basis functions by calling `randn`$((k+1)N,1)$ in Matlab and apply the QR decomposition followed by an $A_\mathbf{1}$-weighted Gram–Schmidt decomposition to obtain the random basis functions U_{rand} and E_{rand}, respectively. We set $S^0 = \text{diag}(1,0)$, so that the initial matrix F^0 is unchanged after the basis enrichment.

The results are shown in Figure 5(b), from which we see that after adding an additional component to the original bases, all of the initial conditions satisfy the

TABLE 2. Modified bases and the corresponding values for the condition in (4.29) and (4.33)

	Case (a_1)	Case (a_2)	Case (a_3)	Case (a_4)
U^0	$[\hat{U}_2^0,\ U_\perp(1,1)]$	$[\hat{U}_2^0,\ U_\perp(1,0)]$	$[\hat{U}_2^0,\ U_\perp(0.1,10)]$	$[\hat{U}_2^0,\ U_{\text{rand}}]$
E^0	$[\hat{E}_2^0,\ E_\perp(1,1)]$	$[\hat{E}_2^0,\ E_\perp(0,1)]$	$[\hat{E}_2^0,\ U_\perp(0.1,10)]$	$[\hat{E}_2^0,\ E_{\text{rand}}]$
$\left\|\left\|P_{U^0}U^{\mathsf{Eq}}\right\|\right\|$	0.7071	1	0.01	0.1601
$\|(E^{\mathsf{Eq}})^\top A_\chi P_{E^0}^\chi\|_{A_1}$	0.7046	3.0119e-13	0.01	0.0869

conditions in Theorem 4.10, and consequently converge to the equilibrium. In addition, we also repeated Case (a_4) for more than 1000 times with different random basis functions, and observed that all converged to the equilibrium. This is not surprising as the probability of drawing a random vector that is orthogonal to the equilibrium is very small.

6. Conclusions

In this paper, we have proposed a semi-implicit dynamical low-rank, discontinuous Galerkin (DLR-DG) method for a space homogeneous kinetic equation with a relaxation operator that models the emission and absorption of particles by a background medium. We have derived a weighted dynamical low-rank approximation (DLRA) that is consistent with the matrix differential equation of the DG scheme. A semi-implicit unconventional integrator (SIUI) is used to integrate the DLRA, and we show that the solution is identical to the solution of a DLR-DG scheme in a DLR-DG space. We have shown the well-posedness of the fully discrete DLR-DG scheme and identified a sufficient condition on the time step size, together with conditions on the DLR-DG basis, such that the distance between the DLR-DG solution and the equilibrium solution decays geometrically with the number of time steps. Numerical results show that the DLR-DG solution is comparable to the full-rank DG solution and converges to the equilibrium solution when the bases satisfy the conditions of the theory.

In future work, it would be interesting to apply the proposed DLR-DG method to more general kinetic equations, e.g., that model scattering with a background. Then, in addition to the properties stated in Proposition 2.4 for the kinetic equation modeling emission and absorption, the conservation of particles in the scattering process should be captured. It may be challenging for the proposed DLR-DG scheme to conserve particles, but extensions inspired by ideas proposed in [5, 18] may be fruitful. We will investigate this in future works.

Appendix A. Some useful matrix results

From Lemma A.1 to Lemma A.3, we assume that m, n, and r are some positive integers satisfying $r \le \min\{m, n\}$. If $A \in \mathbb{R}^{n\times n}$ is a symmetric and positive definite matrix, then Cholesky decomposition implies that there exists a nonsingular matrix $C \in \mathbb{R}^{n\times n}$ such that

$$A = C^\top C. \tag{A.1}$$

Lemma A.1. *For any matrices $A \in \mathbb{R}^{m\times r}$, $B \in \mathbb{R}^{n\times r}$, and $D \in \mathbb{R}^{m\times n}$,*

$$\left(AB^\top, D\right)_{\mathrm{F}} = \left(B^\top, A^\top D\right)_{\mathrm{F}} = (A, DB)_{\mathrm{F}}\,. \tag{A.2}$$

Lemma A.2. *Let $a \in \mathbb{R}$ and $b \in \mathbb{R}$ be constants satisfying $0 \le a \le b$. Suppose $D \in \mathbb{R}^{n\times n}$ is a symmetric positive semi-definite matrix with eigenvalues $\{\lambda_i\}_{i=1}^n$ satisfying $a \le \lambda_1 \le \cdots \le \lambda_n \le b$. Then for any nonzero $Z \in \mathbb{R}^{m\times n}$,*

$$a \le \frac{(ZD, Z)_{\mathrm{F}}}{(Z, Z)_{\mathrm{F}}} \le b. \tag{A.3}$$

Proof. For any nonzero $z \in \mathbb{R}^{n\times 1}$, the Rayleigh quotient satisfies

$$a \le \frac{(Dz, z)}{(z, z)} = \frac{(z^\top D^\top, z^\top)}{(z^\top, z^\top)} \le b. \tag{A.4}$$

Set $Z^\top = [z_1, \ldots, z_m]$ where each $z_j \in \mathbb{R}^{n\times 1}$. Then

$$\frac{(ZD, Z)_{\mathrm{F}}}{(Z, Z)_{\mathrm{F}}} = \frac{\sum_{j=1}^m (Dz_j, z_j)}{\sum_{j=1}^m (z_j, z_j)}, \tag{A.5}$$

which gives (A.3) by applying (A.4) to each term in the sum of the numerator. □

Lemma A.3. *Let $A \in \mathbb{R}^{n\times n}$ be a symmetric positive definite matrix. Suppose $D \in \mathbb{R}^{n\times n}$ with eigenvalues $\{\lambda_i\}_{i=1}^n$ satisfying $a \le \lambda_1 \le \cdots \le \lambda_n \le b$. Then for any $Z \in \mathbb{R}^{m\times n}$,*

$$0 \le (ZD^\top A, ZD^\top)_{\mathrm{F}} \le b^2 (ZA, Z)_{\mathrm{F}}. \tag{A.6}$$

Proof. Let (λ, q) be an eigenpair of the matrix D so that $Dq = \lambda q$. If $q' = Cq$ for C given in (A.1), then $CDC^{-1}q' = \lambda q'$, which implies that λ is also the eigenvalue of the matrix CDC^{-1} and that the symmetric positive-definite matrix $(CDC^{-1})^T(CDC^{-1})$ has an eigenvalue $\lambda^2 \in [0, b^2]$. Let $Z' \in \mathbb{R}^{m\times n}$ be any matrix. By Lemma A.2, we have

$$0 \le (Z'(CDC^{-1})^T(CDC^{-1}), Z')_{\mathrm{F}} \le b^2 (Z', Z')_{\mathrm{F}}, \tag{A.7}$$

which can be reformulated as (A.6) by taking $Z' = ZC^T$. □

Appendix B. Some useful algorithms

Motivated by [1], we introduce the generalized singular value decomposition (GSVD) and the generalized QR factorization (GQR). Let $\mathbb{S}_{++}^n$ be the set of $n \times n$ symmetric positive definite matrices.

Algorithm B.1 (Matrix square root)**.** *Input: $A_{\mathbf{1}} \in \mathbb{S}_{++}^n$. Output: $A_{\mathbf{1}}^{\pm\frac{1}{2}} \in \mathbb{S}_{++}^n$.*

- *Apply the eigen-decomposition (`svd` in MATLAB) to $A_{\mathbf{1}}$ and obtain*

$$A_{\mathbf{1}} = \Phi\Lambda\Phi^\top, \tag{B.1}$$

  where Φ satisfies $\Phi^\top\Phi = I_n$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with $\lambda_i > 0$.
- *Compute $\Lambda^{\pm\frac{1}{2}} = \mathrm{diag}(\lambda_1^{\pm\frac{1}{2}}, \ldots, \lambda_n^{\pm\frac{1}{2}})$.*
- *Compute the symmetric matrix $A_{\mathbf{1}}^{\pm\frac{1}{2}} = \Phi\Lambda^{\pm\frac{1}{2}}\Phi^\top$.*

Algorithm B.1 gives

$$A_{\mathbf{1}} = A_{\mathbf{1}}^{\frac{1}{2}}(A_{\mathbf{1}}^{\frac{1}{2}})^\top = A_{\mathbf{1}}^{\frac{1}{2}}A_{\mathbf{1}}^{\frac{1}{2}}, \quad A_{\mathbf{1}}^{\frac{1}{2}}A_{\mathbf{1}}^{-\frac{1}{2}} = I_n. \tag{B.2}$$

Algorithm B.2 (GSVD)**.** *Input: $F \in \mathbb{R}^{m\times n}$, $A_{\mathbf{1}}^{\pm\frac{1}{2}} \in \mathbb{S}_{++}^n$, $r \le \min\{m, n\}$. Output: $U \in \mathbb{R}^{m\times r}$, $S \in \mathbb{R}^{r\times r}$ and $E \in \mathbb{R}^{n\times r}$.*

- *Apply the SVD decomposition to* $FA_{\mathbf{1}}^{\frac{1}{2}}$ *and obtain*

$$FA_{\mathbf{1}}^{\frac{1}{2}} = US\hat{E}^\top, \tag{B.3}$$

where U *satisfies*

$$U^\top U = I_r, \tag{B.4}$$

and $\hat{E}$ *satisfies* $\hat{E}^\top \hat{E} = I_r$.
- *Compute* $E = A_{\mathbf{1}}^{-\frac{1}{2}} \hat{E}$.

Algorithm B.2 gives the GSVD

$$F = USE^T, \tag{B.5}$$

where U satisfies (B.4) and E satisfies

$$E^\top A_{\mathbf{1}} E = \hat{E}^\top A_{\mathbf{1}}^{-\frac{1}{2}} A_{\mathbf{1}} A_{\mathbf{1}}^{-\frac{1}{2}} \hat{E} = I_r. \tag{B.6}$$

Algorithm B.3 (GQR). *Input:* $\mathbf{L} \in \mathbb{R}^{n\times r}$, $A_{\mathbf{1}}^{\pm\frac{1}{2}} \in \mathbb{S}_{++}^n$. *Output:* $E \in \mathbb{R}^{n\times r}$.

- *Apply the QR decomposition to* $A_{\mathbf{1}}^{\frac{1}{2}}\mathbf{L}$ *and obtain*

$$A_{\mathbf{1}}^{\frac{1}{2}}\mathbf{L} = \hat{E}R, \tag{B.7}$$

where $\hat{E}$ *satisfies* $\hat{E}^\top \hat{E} = I_r$.
- *Compute* $E = A_{\mathbf{1}}^{-\frac{1}{2}} \hat{E}$.

Algorithm B.3 gives the generalized QR factorization

$$\mathbf{L} = ER, \tag{B.8}$$

where E also satisfies (B.6).

Appendix C. Technical proofs

In this section, we present the proofs to some lemmas. For any nonzero function $w_h = X^\top(\mu)WY(\varepsilon) \in V_h$ for some nonzero $W \in \mathbb{R}^{m\times n}$, let

$$R_\chi(w_h) = \frac{(\chi w_h, w_h; \varepsilon^2)_\Omega}{(w_h, w_h; \varepsilon^2)_\Omega} = \frac{(WA_\chi, W)_{\mathrm{F}}}{(WA_{\mathbf{1}}, W)_{\mathrm{F}}} = \frac{\|W\|_{A_\chi}^2}{\|W\|_{A_{\mathbf{1}}}^2}. \tag{C.1}$$

Lemma C.1. *Let* m_1 *be an integer satisfying* $1 \le m_1 \le m$. *Then for any nonzero matrix* $Z \in \mathbb{R}^{m_1\times n}$,

$$\chi_{\min} \le \frac{\|Z\|_{A_\chi}^2}{\|Z\|_{A_{\mathbf{1}}}^2} \le \chi_{\max}. \tag{C.2}$$

Therefore if λ *is an eigenvalue of the matrix* $(A_\chi)^{-1}A_{\mathbf{1}}$ *or* $A_{\mathbf{1}}(A_\chi)^{-1}$, *then*

$$\chi_{\max}^{-1} \le \lambda \le \chi_{\min}^{-1}. \tag{C.3}$$

Proof. A consequence of Assumption 2.1 is that

$$\chi_{\min} \le R_\chi(w_h) \le \chi_{\max}. \tag{C.4}$$

The inequality in (C.2) follows from setting $w_h = X^\top(\mu)WY(\varepsilon)$ in (C.4), where $W^\top = [Z^\top, Z_1^\top]$ and $Z_1 = 0 \in \mathbb{R}^{(m-m_1)\times n}$. Inverting (C.2) gives

$$\chi_{\max}^{-1} \le \frac{\|Z\|_{A_{\mathbf{1}}}^2}{\|Z\|_{A_\chi}^2} \le \chi_{\min}^{-1} \quad \text{for all nonzero } Z \in \mathbb{R}^{m_1\times n}. \tag{C.5}$$

The inequalities in (C.3) follow immediately by setting $Z^\top$ in (C.5) to be an eigenvector of $(A_\chi)^{-1}A_{\mathbf{1}}$. □

C.1. **Proof of Lemma 4.6.** We will first need a rather technical lemma.

Lemma C.2. *Let $E^{\mathsf{Eq}} \in \mathbb{R}^{n\times 1}$, $B_{\mathbf{L}} = [b_1 \dots, b_r] \in \mathbb{R}^{n\times r}$, and $l = [l_1, \dots, l_r]$ be a nonzero vector, where $b_i \in \mathbb{R}^{n\times 1}$ and $l_i \in \mathbb{R}$ for $i = 1, \dots, r$. Assume that the matrix*

$$\mathbf{L}^{\mathfrak{n}+1} = \left[l_1 E^{\mathsf{Eq}} + \frac{1}{\Delta t} b_1, \cdots, l_r E^{\mathsf{Eq}} + \frac{1}{\Delta t} b_r \right] \in \mathbb{R}^{n\times r} \tag{C.6}$$

has a decomposition $\mathbf{L}^{\mathfrak{n}+1} = E^{\mathfrak{n}+1} S_{\mathbf{L}}^{\mathfrak{n}+1}$ with $E^{\mathfrak{n}+1} = \left[E_1^{\mathfrak{n}+1}, \cdots, E_r^{\mathfrak{n}+1}\right]$ satisfying (4.3). Then

$$1 - \|(E^{\mathsf{Eq}})^\top A_{\mathbf{1}} P_{E^{\mathfrak{n}+1}}\|_{A_{\mathbf{1}}}^2 = 1 - \|(E^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E^{\mathsf{Eq}}\|^2 \le \frac{\|B_{\mathbf{L}}^\top\|_{A_{\mathbf{1}}}^2}{\Delta t^2 \|l\|_\infty^2}. \tag{C.7}$$

Proof. As long as (4.3) holds, (C.7) is independent of the choice of basis for the span of $\mathbf{L}^{\mathfrak{n}+1}$. Hence without loss of generality, we assume a weighted Gram-Schmidt decomposition:

$$E_i^{\mathfrak{n}+1} = \frac{\mathbf{L}_i^{\mathfrak{n}+1} - \sum_{j=1}^{i-1} \left((\mathbf{L}_i^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1}\right) E_j^{\mathfrak{n}+1}}{\sqrt{(\mathbf{L}_i^{\mathfrak{n}+1})^\top A_{\mathbf{1}} \mathbf{L}_i^{\mathfrak{n}+1} - \sum_{j=1}^{i-1} \left((\mathbf{L}_i^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1}\right)^2}}, \tag{C.8}$$

where $\mathbf{L}_i^{\mathfrak{n}+1} = l_i E^{\mathsf{Eq}} + \frac{1}{\Delta t} b_i$. Then

$$\begin{aligned} &((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_i^{\mathfrak{n}+1})^2 \\ &= \frac{\left((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} \mathbf{L}_i^{\mathfrak{n}+1} - \sum_{j=1}^{i-1} \left((\mathbf{L}_i^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1}\right) (E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1}\right)^2}{(\mathbf{L}_i^{\mathfrak{n}+1})^\top A_{\mathbf{1}} \mathbf{L}_i^{\mathfrak{n}+1} - \sum_{j=1}^{i-1} \left((\mathbf{L}_i^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1}\right)^2} \\ &= \frac{\left(l_i \xi_i^2 + \frac{1}{\Delta t}\alpha_i\right)^2}{l_i^2 \xi_i^2 + \frac{1}{\Delta t}(2 l_i \alpha_i + \frac{1}{\Delta t}\gamma_i^2)}, \end{aligned} \tag{C.9}$$

where

$$\begin{aligned} \alpha_i &= (E^{\mathsf{Eq}})^\top A_{\mathbf{1}} b_i - \sum_{j=1}^{i-1} ((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})((E_j^{\mathfrak{n}+1})^\top A_{\mathbf{1}} b_i), \\ \gamma_i &= \left(b_i^\top A_{\mathbf{1}} b_i - \sum_{j=1}^{i-1} (b_i^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2 \right)^{\frac{1}{2}}, \\ \xi_i &= \left((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E^{\mathsf{Eq}} - \sum_{j=1}^{i-1} ((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2 \right)^{\frac{1}{2}} \\ &= \left(1 - \sum_{j=1}^{i-1} ((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2 \right)^{\frac{1}{2}} \end{aligned} \tag{C.10}$$

are all nonnegative. We extend the orthonormal basis $E_j^{\mathfrak{n}+1}$ from $1 \le j \le r$ to $1 \le j \le n$. Then E^{Eq} and b_i in (C.6) can be expressed in terms of the basis

functions $\{E_j^{\mathfrak{n}+1}\}_{j=1}^n$ as

$$E^{\mathsf{Eq}} = \sum_{j=1}^{n}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})E_j^{\mathfrak{n}+1}, \quad b_i = \sum_{j=1}^{n}(b_i^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})E_j^{\mathfrak{n}+1}, \tag{C.11}$$

which implies

$$\gamma_i^2 = \sum_{j=i}^{n}(b_i^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2, \qquad \xi_i^2 = \sum_{j=i}^{n}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2, \tag{C.12}$$

and

$$\alpha_i = \sum_{j=i}^{n}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})((E_j^{\mathfrak{n}+1})^\top A_{\mathbf{1}} b_i). \tag{C.13}$$

Therefore, we have

$$|\alpha_i| \le \left(\sum_{j=i}^{n}\left((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1}\right)^2\right)^{\frac{1}{2}}\left(\sum_{j=i}^{n}\left((E_j^{\mathfrak{n}+1})^\top A_{\mathbf{1}} b_i\right)^2\right)^{\frac{1}{2}} = \xi_i\gamma_i. \tag{C.14}$$

By (C.11) and (C.12), it follows that

$$\gamma_i^2 \le b_i^\top A_{\mathbf{1}} b_i \le (B_{\mathbf{L}}^\top A_{\mathbf{1}}, B_{\mathbf{L}}^\top)_{\mathrm{F}} = \|B_{\mathbf{L}}^\top\|_{A_{\mathbf{1}}}^2. \tag{C.15}$$

Meanwhile, the direct calculation gives

$$\|(E^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E^{\mathsf{Eq}}\|^2 = \sum_{j=1}^{r}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2. \tag{C.16}$$

Choose i such that $1 \le i \le r$ and $|l_i| = \|l\|_\infty := \max_{1\le j\le r}|l_j|$. We consider the following cases.

Case 1. If $l_i\xi_i^2 + \frac{1}{\Delta t}\alpha_i = 0$, that is $\xi_i^2 = -\frac{\alpha_i}{\Delta t l_i} = \frac{|\alpha_i|}{\Delta t|l_i|}$, then by (C.14),

$$0 \le \xi_i \le \frac{\gamma_i}{\Delta t|l_i|}, \tag{C.17}$$

which together with (C.15) and (C.16) implies that

$$1 - \|(E^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E^{\mathsf{Eq}}\|^2 \le 1 - \sum_{j=1}^{i-1}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2 = \xi_i^2 \le \frac{\gamma_i^2}{\Delta t^2 l_i^2} \le \frac{\|B_{\mathbf{L}}^\top\|_{A_{\mathbf{1}}}^2}{\Delta t^2\|l\|_\infty^2}. \tag{C.18}$$

Case 2. Now we consider $l_i\xi_i^2 + \frac{1}{\Delta t}\alpha_i \ne 0$.

Case 2.a. If $\xi_i = 0$, then

$$1 - \|(E^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E^{\mathsf{Eq}}\|^2 \le 1 - \sum_{j=1}^{i-1}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2 = \xi_i^2 = 0. \tag{C.19}$$

Therefore, the inequality (C.7) holds.

Case 2.b. If $\xi_i \ne 0$, we consider two cases:

Case 2.b.i. If $\gamma_i = 0$, then by (C.14), $\alpha_i = 0$. By (C.9), $((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_i^{\mathfrak{n}+1})^2 = \xi_i^2$, which implies

(C.20)
$$1 - \|(E^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E^{\mathsf{Eq}}\|^2 \le 1 - \sum_{j=1}^{i}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2 = \xi_i^2 - ((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_i^{\mathfrak{n}+1})^2 = 0.$$

Therefore, the inequality (C.7) still holds.

Case 2.b.ii. If $\gamma_i \neq 0$, by (C.14) there exists a parameter $\tau \in [-1, 1]$ such that

$$\alpha_i = \tau\gamma_i\xi_i. \tag{C.21}$$

Substituting (C.21) into (C.9) and rewriting yield

$$\begin{aligned}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_i^{\mathfrak{n}+1})^2 =& \xi_i^2 \frac{\left(l_i\xi_i + \frac{1}{\Delta t}\tau\gamma_i\right)^2}{\left(l_i\xi_i + \frac{1}{\Delta t}\tau\gamma_i\right)^2 + \frac{1}{(\Delta t)^2}(1-\tau^2)\gamma_i^2} \\ =& \xi_i^2 - g(\tau) = 1 - \sum_{j=1}^{i-1}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2 - g(\tau),\end{aligned} \tag{C.22}$$

where we have used (C.10) for the third equality and $g : [-1, 1] \to \mathbb{R}$ is a non-negative and differentiable function given by

$$g(\tau) = \frac{\xi_i^2 \frac{1}{(\Delta t)^2}(1-\tau^2)\gamma_i^2}{\left(l_i\xi_i + \frac{1}{\Delta t}\tau\gamma_i\right)^2 + \frac{1}{(\Delta t)^2}(1-\tau^2)\gamma_i^2}. \tag{C.23}$$

We wish to maximize g on $[-1, 1]$. Since $g(-1) = g(1) = 0$, we solve for the critical points τ^* satisfying

$$g'(\tau) = \frac{-\frac{2}{(\Delta t)^2}\xi_i^2\gamma_i^2\left(l_i\xi_i + \frac{1}{\Delta t}\tau\gamma_i\right)\left(\tau l_i\xi_i + \frac{1}{\Delta t}\gamma_i\right)}{\left(\left(l_i\xi_i + \frac{1}{\Delta t}\tau\gamma_i\right)^2 + \frac{1}{(\Delta t)^2}(1-\tau^2)\gamma_i^2\right)^2} = 0. \tag{C.24}$$

Since $\xi_i \neq 0$ and $\left(l_i\xi_i^2 + \frac{1}{\Delta t}\alpha_i\right)^2 = \xi_i^2\left(l_i\xi_i + \frac{1}{\Delta t}\tau\gamma_i\right)^2 \neq 0$, it follows that $l_i\xi_i + \frac{1}{\Delta t}\tau\gamma_i \neq 0$. Therefore, the only critical point for (C.24) is $\tau^* = -\frac{\gamma_i}{\Delta t l_i\xi_i} \in (-1, 1)$. Plugging in τ^* into (C.23) yields

$$g(\tau) \le g\left(-\frac{\gamma_i}{\Delta t l_i\xi_i}\right) = \frac{\gamma_i^2}{\Delta t^2 l_i^2}. \tag{C.25}$$

Therefore (C.22), (C.25), and (C.15) imply

(C.26)
$$1 - \|(E^{\mathfrak{n}+1})^\top A_{\mathbf{1}} E^{\mathsf{Eq}}\|^2 \le 1 - \sum_{j=1}^{i}((E^{\mathsf{Eq}})^\top A_{\mathbf{1}} E_j^{\mathfrak{n}+1})^2 = g(\tau) \le \frac{\gamma_i^2}{\Delta t^2 l_i^2} \le \frac{\|B_{\mathbf{L}}^\top\|_{A_{\mathbf{1}}}^2}{\Delta t^2 \|l\|_\infty^2}.$$

□

Next, we present the proof of Lemma 4.6.

Proof of Lemma 4.6. Start with (4.12b), which is equivalent to finding $\mathbf{L}^{\mathfrak{n}+1} \in \mathbb{R}^{n\times r}$ such that for any $\mathbf{L}_W \in \mathbb{R}^{n\times r}$,

$$\left(U^{\mathfrak{n}}\left(D_t\mathbf{L}^{\mathfrak{n}+1}\right)^\top A_{\mathbf{1}}, U^{\mathfrak{n}}\mathbf{L}_W^\top\right)_{\mathrm{F}} = \left(G(U^{\mathfrak{n}}(\mathbf{L}^{\mathfrak{n}+1})^\top), U^{\mathfrak{n}}\mathbf{L}_W^\top\right)_{\mathrm{F}}. \tag{C.27}$$

Set $\mathbf{L}_W = (A_\chi)^{-1}\mathbf{L}'_W$, where $\mathbf{L}'_W$ is arbitrary, into (C.27). Then use (4.21) for G, and apply Lemma A.1:

$$\left(\left(D_t\mathbf{L}^{\mathtt{n}+1}\right)^\top A_{\mathbf{1}}(A_\chi)^{-1}, (\mathbf{L}'_W)^\top\right)_{\mathrm{F}} = \left((U^{\mathtt{n}})^\top F^{\mathsf{Eq}} - (\mathbf{L}^{\mathtt{n}+1})^\top, (\mathbf{L}'_W)^\top\right)_{\mathrm{F}}. \tag{C.28}$$

Since $\mathbf{L}'_W$ is arbitrary, it follows that

$$\begin{aligned}&\left(I_n + \frac{1}{\Delta t}(A_\chi)^{-1}A_{\mathbf{1}}\right)\mathbf{L}^{\mathtt{n}+1} = (F^{\mathsf{Eq}})^\top U^{\mathtt{n}} + \frac{1}{\Delta t}(A_\chi)^{-1}A_{\mathbf{1}}\mathbf{L}^{\mathtt{n}}\\ &= \left(I_n + \frac{1}{\Delta t}(A_\chi)^{-1}A_{\mathbf{1}}\right)(F^{\mathsf{Eq}})^\top U^{\mathtt{n}} + \frac{1}{\Delta t}(A_\chi)^{-1}A_{\mathbf{1}}\left(\mathbf{L}^{\mathtt{n}} - (F^{\mathsf{Eq}})^\top U^{\mathtt{n}}\right),\end{aligned}$$

which gives

$$\mathbf{L}^{\mathtt{n}+1} = (F^{\mathsf{Eq}})^\top U^{\mathtt{n}} + \frac{1}{\Delta t}B_{\mathbf{L}}, \tag{C.29}$$

where

$$\begin{aligned}B_{\mathbf{L}} &:= D_{\mathbf{L}}\left(\mathbf{L}^{\mathtt{n}} - (F^{\mathsf{Eq}})^\top U^{\mathtt{n}}\right) \in \mathbb{R}^{n\times r},\\ D_{\mathbf{L}} &:= \left(I + \frac{1}{\Delta t}(A_\chi)^{-1}A_{\mathbf{1}}\right)^{-1}(A_\chi)^{-1}A_{\mathbf{1}} \in \mathbb{R}^{n\times n}.\end{aligned} \tag{C.30}$$

Because $U^{\mathtt{n}}$ is orthogonal and $U^{\mathtt{n}}(\mathbf{L}^{\mathtt{n}})^\top = U^{\mathtt{n}}S^{\mathtt{n}}(E^{\mathtt{n}})^\top = \hat{F}^{\mathtt{n}}$, it follows that

$$\begin{aligned}\|(\mathbf{L}^{\mathtt{n}})^\top - (U^{\mathtt{n}})^\top F^{\mathsf{Eq}}\|_{A_{\mathbf{1}}} &= \|\hat{F}^{\mathtt{n}} - U^{\mathtt{n}}(U^{\mathtt{n}})^\top F^{\mathsf{Eq}}\|_{A_{\mathbf{1}}} \le \|\hat{F}^{\mathtt{n}} - F^{\mathsf{Eq}}\|_{A_{\mathbf{1}}}\\ &= \|\varepsilon(\hat{f}_h^{\mathtt{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}.\end{aligned} \tag{C.31}$$

Any eigenvalue $\lambda_{D_{\mathbf{L}}}$ of the matrix $D_{\mathbf{L}}$ can be expressed in terms of the corresponding eigenvalue λ of $(A_\chi)^{-1}A_{\mathbf{1}}$ as follows

$$\lambda_{D_{\mathbf{L}}} = \frac{\lambda}{1 + \frac{1}{\Delta t}\lambda} = \frac{1}{\frac{1}{\lambda} + \frac{1}{\Delta t}}. \tag{C.32}$$

Therefore, according to (C.3), $\lambda_{D_{\mathbf{L}}}$ satisfies

$$0 < \frac{1}{\chi_{\max} + \frac{1}{\Delta t}} \le \lambda_{D_{\mathbf{L}}} \le \frac{1}{\chi_{\min} + \frac{1}{\Delta t}} < \frac{1}{\chi_{\min}}. \tag{C.33}$$

Together, (C.31), (C.33), and Lemma A.3 imply that

$$\|B_{\mathbf{L}}^\top\|_{A_{\mathbf{1}}}^2 \le \frac{1}{\chi_{\min}^2}\|\varepsilon(\hat{f}_h^{\mathtt{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}^2. \tag{C.34}$$

Let $S^{\mathsf{Eq}}(U^{\mathsf{Eq}})^\top U^{\mathtt{n}} = [l_1, \dots, l_r] = l \in \mathbb{R}^{1\times r}$ for scalars l_i $(i = 1, \dots, r)$. Using (4.22), (C.29) becomes

$$\mathbf{L}^{\mathtt{n}+1} = E^{\mathsf{Eq}}S^{\mathsf{Eq}}(U^{\mathsf{Eq}})^\top U^{\mathtt{n}} + \frac{1}{\Delta t}B_{\mathbf{L}} = \left[l_1E^{\mathsf{Eq}} + \frac{1}{\Delta t}b_1, \dots, l_rE^{\mathsf{Eq}} + \frac{1}{\Delta t}b_r\right], \tag{C.35}$$

where $b_i \in \mathbb{R}^n$ $(i = 1, \dots, r)$ are the column vectors of $B_{\mathbf{L}}$. Combining Lemma C.2 and the bound in (C.34) gives

$$1 - \|(E^{\mathtt{n}+1})^\top A_{\mathbf{1}}E^{\mathsf{Eq}}\|^2 \le \frac{\|B_{\mathbf{L}}^\top\|_{A_{\mathbf{1}}}^2}{\Delta t^2\|l\|_\infty^2} \le \frac{\|\varepsilon(\hat{f}_h^{\mathtt{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}^2}{\Delta t^2\|l\|_\infty^2\chi_{\min}^2}. \tag{C.36}$$

For l, it holds

(C.37)
$$\|l\|_\infty = |S^{\mathsf{Eq}}|\,\|(U^{\mathfrak{n}})^\top U^{\mathsf{Eq}}\|_\infty \geq \frac{|S^{\mathsf{Eq}}|\,\|(U^{\mathfrak{n}})^\top U^{\mathsf{Eq}}\|}{\sqrt{r}} = \frac{|S^{\mathsf{Eq}}|\|P_{U^{\mathfrak{n}}}U^{\mathsf{Eq}}\|}{\sqrt{r}} \geq \frac{\beta|S^{\mathsf{Eq}}|}{\sqrt{r}},$$

where the first inequality follows from the norm equivalence, and the last inequality follows from the assumption in (4.29). Thus, if $\Delta t \geq \frac{\sqrt{r}}{\beta\delta\chi_{\min}}$, the estimate (4.30) holds.

The equality (4.31) follows from (C.36) when $\hat{f}_h^{\mathfrak{n}} = f_h^{\mathsf{Eq}}$. □

C.2. **Proof of Lemma 4.7.** Similar to Lemma C.2, we prepare the following result.

Lemma C.3. *Let $U^{\mathsf{Eq}} \in \mathbb{R}^{n\times 1}$, $B_{\mathbf{K}} = [b_1 \dots, b_r] \in \mathbb{R}^{m\times r}$, and $l = [l_1, \dots, l_r]$ be a nonzero vector, where $b_i \in \mathbb{R}^{m\times 1}$ and $l_i \in \mathbb{R}$ for $i = 1, \dots, r$. Assume that the matrix*

(C.38)
$$\mathbf{K}^{\mathfrak{n}+1} = \left[l_1 U^{\mathsf{Eq}} + \frac{1}{\Delta t}b_1, \cdots, l_r U^{\mathsf{Eq}} + \frac{1}{\Delta t}b_r\right] \in \mathbb{R}^{m\times r}$$

has a decomposition $\mathbf{K}^{\mathfrak{n}+1} = U^{\mathfrak{n}+1}S_{\mathbf{K}}^{\mathfrak{n}+1}$ with $U^{\mathfrak{n}+1} = \left[U_1^{\mathfrak{n}+1}, \cdots, U_r^{\mathfrak{n}+1}\right]$ satisfying (4.3). *Then*

(C.39)
$$1 - \|P_{U^{\mathfrak{n}+1}}U^{\mathsf{Eq}}\|^2 = 1 - \|(U^{\mathfrak{n}+1})^\top U^{\mathsf{Eq}}\|^2 \leq \frac{\|B_{\mathbf{K}}\|_{\mathrm{F}}^2}{\Delta t^2\|l\|_\infty^2}.$$

For any $E \in \mathbb{R}^{n\times r}$ satisfying $E^\top A_{\mathbf{1}} E = I_r$, because the term $E^\top A_\chi E$ will appear frequently, we introduce the symmetric matrix

(C.40)
$$B = E^\top A_\chi E \in \mathbb{R}^{r\times r},$$

for which we have the following results.

Lemma C.4. *Let (λ_B, q_B) be an eigenpair of B in* (C.40). *Then*

(C.41)
$$\chi_{\min} \leq \lambda_B = \frac{(Eq_B)^\top A_\chi E q_B}{(Eq_B)^\top A_{\mathbf{1}} E q_B} = \frac{\|(Eq_B)^\top\|_{A_\chi}^2}{\|(Eq_B)^\top\|_{A_{\mathbf{1}}}^2} \leq \chi_{\max}.$$

Proof. If (λ_B, q_B) is an eigenpair of B, then

(C.42)
$$E^\top A_\chi E q_B = B q_B = \lambda_B q_B = \lambda_B E^\top A_{\mathbf{1}} E q_B.$$

Left-multiplying (C.42) by $q_B^\top$ and applying (C.2) with $Z = (Eq_B)^\top$ gives (C.41). □

Lemma C.5. *Let $E \in \mathbb{R}^{n\times r}$ satisfy $E^\top A_{\mathbf{1}} E = I_r$, and recall the definition of P_E^χ from* (4.32). *Then for any $Z \in \mathbb{R}^{\ell\times n}$, $1 \leq \ell \leq m$,*

(C.43)
$$\|ZA_\chi P_E^\chi\|_{A_{\mathbf{1}}} \leq \sqrt{\frac{\chi_{\max}}{\chi_{\min}}}\|Z\|_{A_{\mathbf{1}}}.$$

Proof. Recall that A_χ is symmetric and positive definite, and thus can be decomposed as $A_\chi = C_\chi^\top C_\chi$ where C_χ is nonsingular. Let $D_\chi = C_\chi E B^{-1}$, where B is given in (C.40), and compute

(C.44)
$$\|ZA_\chi P_E^\chi\|_{A_{\mathbf{1}}} = \|ZA_\chi E B^{-1}\|_{\mathrm{F}} \leq \|ZC_\chi^\top\|_{\mathrm{F}}\|C_\chi E B^{-1}\| = \|Z\|_{A_\chi}\|D_\chi\|.$$

Since $\|D_\chi\|^2$ is the largest eigenvalue of $D_\chi^\top D_\chi$ and

(C.45)
$$D_\chi^\top D_\chi = (B^{-1})^\top E^\top C_\chi^\top C_\chi E B^{-1} = B^{-1}BB^{-1} = B^{-1},$$

then (C.41) implies $\|D_\chi\| \le \chi_{\min}^{-1/2}$, which along with (C.44) and (C.2) yields (C.43). □

Next, we present the proof of Lemma 4.7.

Proof of Lemma 4.7. The proof follows along the same lines as the proof of Lemma 4.6. (4.12a) is equivalent to finding $\mathbf{K}^{\mathfrak{n}+1} \in \mathbb{R}^{m\times r}$ such that for any $\mathbf{K}_W \in \mathbb{R}^{m\times r}$,

$$\left(D_t\mathbf{K}^{\mathfrak{n}+1}(E^{\mathfrak{n}})^\top A_{\mathbf{1}}, \mathbf{K}_W(E^{\mathfrak{n}})^\top\right)_{\mathrm{F}} = \left(G(\mathbf{K}^{\mathfrak{n}+1}(E^{\mathfrak{n}})^\top), \mathbf{K}_W(E^{\mathfrak{n}})^\top\right)_{\mathrm{F}}. \tag{C.46}$$

Applying (4.21) and Lemma A.1 to (C.46) gives

$$\left(D_t\mathbf{K}^{\mathfrak{n}+1}, \mathbf{K}_W\right)_{\mathrm{F}} = \left(F^{\mathsf{Eq}}A_\chi E^{\mathfrak{n}} - \mathbf{K}^{\mathfrak{n}+1}(E^{\mathfrak{n}})^\top A_\chi E^{\mathfrak{n}}, \mathbf{K}_W\right)_{\mathrm{F}}. \tag{C.47}$$

Let $\mathbf{K}_W = \mathbf{K}'_W(B^{\mathfrak{n}})^{-1}$ for any $\mathbf{K}'_W \in \mathbf{R}^{m\times r}$, where $B^{\mathfrak{n}} = (E^{\mathfrak{n}})^\top A_\chi E^{\mathfrak{n}}$. Then

$$\left((D_t\mathbf{K}^{\mathfrak{n}+1}(B^{\mathfrak{n}})^{-1}, \mathbf{K}'_W\right)_{\mathrm{F}} = \left(F^{\mathsf{Eq}}A_\chi E^{\mathfrak{n}}(B^{\mathfrak{n}})^{-1} - \mathbf{K}^{\mathfrak{n}+1}, \mathbf{K}'_W\right)_{\mathrm{F}}. \tag{C.48}$$

Since $\mathbf{K}'_W$ in arbitrary, it follows that

$$\mathbf{K}^{\mathfrak{n}+1} = F^{\mathsf{Eq}}A_\chi E^{\mathfrak{n}}(B^{\mathfrak{n}})^{-1} + \frac{1}{\Delta t}B_{\mathbf{K}}, \tag{C.49}$$

where

$$\begin{aligned} B_{\mathbf{K}} &:= [b_1, \dots, b_r] = \left(\mathbf{K}^{\mathfrak{n}} - F^{\mathsf{Eq}}A_\chi E^{\mathfrak{n}}(B^{\mathfrak{n}})^{-1}\right) D_{\mathbf{K}} \in \mathbb{R}^{m\times r}, \\ D_{\mathbf{K}} &:= (B^{\mathfrak{n}})^{-1}\left(I_r + \frac{(B^{\mathfrak{n}})^{-1}}{\Delta t}\right)^{-1} \in \mathbb{R}^{r\times r}. \end{aligned} \tag{C.50}$$

Since $\mathbf{K}^{\mathfrak{n}} = U^{\mathfrak{n}}S^{\mathfrak{n}}$, we can write

$$\mathbf{K}^{\mathfrak{n}} - F^{\mathsf{Eq}}A_\chi E^{\mathfrak{n}}(B^{\mathfrak{n}})^{-1} = \left(F^{\mathfrak{n}} - F^{\mathsf{Eq}}\right) A_\chi E^{\mathfrak{n}}(B^{\mathfrak{n}})^{-1}. \tag{C.51}$$

By (C.51), Lemma A.2, Lemma C.5, and Lemma 3.1,

$$\begin{aligned} \|\mathbf{K}^{\mathfrak{n}} - F^{\mathsf{Eq}}A_\chi E^{\mathfrak{n}}(B^{\mathfrak{n}})^{-1}\|_{\mathrm{F}} &= \|\left(F^{\mathfrak{n}} - F^{\mathsf{Eq}}\right) A_\chi P^{\chi}_{E^{\mathfrak{n}}}\|_{A_{\mathbf{1}}} \\ &\le \sqrt{\frac{\chi_{\max}}{\chi_{\min}}}\left\|F^{\mathfrak{n}} - F^{\mathsf{Eq}}\right\|_{A_{\mathbf{1}}} = \sqrt{\frac{\chi_{\max}}{\chi_{\min}}}\|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}. \end{aligned} \tag{C.52}$$

Any eigenvalue $\lambda_{D_{\mathbf{K}}}$ of $D_{\mathbf{K}}$ satisfies

$$0 < \frac{1}{\chi_{\max} + \frac{1}{\Delta t}} \le \lambda_{D_{\mathbf{K}}} \le \frac{1}{\chi_{\min} + \frac{1}{\Delta t}} < \frac{1}{\chi_{\min}}. \tag{C.53}$$

Then, by (C.52), (C.53), and Lemma A.3,

$$\|B_{\mathbf{K}}\|_{\mathrm{F}} \le \frac{1}{\chi_{\min}}\sqrt{\frac{\chi_{\max}}{\chi_{\min}}}\|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}. \tag{C.54}$$

Let $S^{\mathsf{Eq}}(E^{\mathsf{Eq}})^\top A_\chi E^{\mathfrak{n}}(B^{\mathfrak{n}})^{-1} = [l_1, \dots, l_r] \in \mathbb{R}^{1\times r}$ for scalars l_i $(i = 1, \dots, r)$. By (4.22) and (C.49),

$$\begin{aligned} \mathbf{K}^{\mathfrak{n}+1} &= U^{\mathsf{Eq}}S^{\mathsf{Eq}}(E^{\mathsf{Eq}})^\top A_\chi E^{\mathfrak{n}}(B^{\mathfrak{n}})^{-1} + \frac{1}{\Delta t}B_{\mathbf{K}} \\ &= \left[l_1U^{\mathsf{Eq}} + \frac{1}{\Delta t}b_1, \dots, l_rU^{\mathsf{Eq}} + \frac{1}{\Delta t}b_r\right], \end{aligned} \tag{C.55}$$

where $b_i \in \mathbb{R}^m$ $(i = 1, \dots, r)$ are the columns of $B_{\mathbf{K}}$. By Lemma C.3 and (C.54),

$$1 - \|(U^{\mathfrak{n}+1})^\top U^{\mathsf{Eq}}\|^2 \le \frac{\|B_{\mathbf{K}}\|_{\mathrm{F}}^2}{\Delta t^2 \|l\|_\infty^2} \le \frac{\chi_{\max}}{\Delta t^2 \|l\|_\infty^2 \chi_{\min}^3} \|\varepsilon(\hat{f}_h^{\mathfrak{n}} - f_h^{\mathsf{Eq}})\|_{L^2(\Omega)}^2. \tag{C.56}$$

By the assumption (4.33) and the fact that

$$\|(B^{\mathfrak{n}})^{-1}(E^{\mathfrak{n}})^\top A_\chi E^{\mathsf{Eq}}\| = \|(E^{\mathsf{Eq}})^\top A_\chi P_{E^{\mathfrak{n}}}^\chi\|_{A_{\mathbf{1}}},$$

(C.57)

$$\|l\|_\infty = |S^{\mathsf{Eq}}|\,\|(B^{\mathfrak{n}})^{-1}(E^{\mathfrak{n}})^\top A_\chi E^{\mathsf{Eq}}\|_\infty \ge \frac{|S^{\mathsf{Eq}}|\,\|(E^{\mathsf{Eq}})^\top A_\chi P_{E^{\mathfrak{n}}}^\chi\|_{A_{\mathbf{1}}}}{\sqrt{r}} \ge \frac{\alpha |S^{\mathsf{Eq}}|}{\sqrt{r}}.$$

Thus, if $\Delta t \ge \frac{\sqrt{r}\chi_{\max}^{1/2}}{\alpha\delta\chi_{\min}^{3/2}}$, estimate (4.34) holds.

The equality (4.35) follows from (C.56) when $\hat{f}_h^{\mathfrak{n}} = f_h^{\mathsf{Eq}}$. □

References

[1] H. Abdi, *Singular value decomposition (SVD) and generalized singular value decomposition (GSVD)*, Encyclopedia of Measurement and Statistics, vol. 907, 2007, p. 912.

[2] M. L. Adams, *Discontinuous finite element transport solutions in thick diffusive problems*, Nuclear Sci. Eng., **137** (2001), no. 3, 298–333.

[3] B. Ayuso, J. A. Carrillo, and C.-W. Shu, *Discontinuous Galerkin methods for the one-dimensional Vlasov-Poisson system*, Kinet. Relat. Models **4** (2011), no. 4, 955–989, DOI 10.3934/krm.2011.4.955. MR2861582

[4] M. Bachmayr, H. Eisenmann, E. Kieri, and A. Uschmajew, *Existence of dynamical low-rank approximations to parabolic problems*, Math. Comp. **90** (2021), no. 330, 1799–1830, DOI 10.1090/mcom/3626. MR4273116

[5] L. Baumann, L. Einkemmer, C. Klingenberg, and J. Kusch, *Energy stable and conservative dynamical low-rank approximation for the Su-Olson problem*, SIAM J. Sci. Comput. **46** (2024), no. 2, B137–B158, DOI 10.1137/23M1586215. MR4733957

[6] M. H. Beck, A. Jäckle, G. A. Worth, and H.-D. Meyer, *The multiconfiguration time-dependent Hartree (MCTDH) method: a highly efficient algorithm for propagating wavepackets*, Phys. Rep. **324** (2000), no. 1, 1–105.

[7] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Texts in Applied Mathematics, vol. 15, Springer, New York, 2008, DOI 10.1007/978-0-387-75934-0. MR2373954

[8] S. W. Bruenn, *Stellar core collapse - numerical model and infall epoch*, Astrophys. J. Suppl. Ser., **58** (1985), 771–841.

[9] A. Burrows, S. Reddy, and T. A. Thompson, *Neutrino opacities in nuclear matter*, Nuclear Phys. A, **777** (2006), 356–394.

[10] G. Ceruti and C. Lubich, *An unconventional robust integrator for dynamical low-rank approximation*, BIT **62** (2022), no. 1, 23–44, DOI 10.1007/s10543-021-00873-0. MR4375023

[11] Y. Cheng, I. M. Gamba, and P. J. Morrison, *Study of conservation and recurrence of Runge-Kutta discontinuous Galerkin schemes for Vlasov-Poisson systems*, J. Sci. Comput. **56** (2013), no. 2, 319–349, DOI 10.1007/s10915-012-9680-x. MR3071178

[12] Z. Ding, L. Einkemmer, and Q. Li, *Dynamical low-rank integrator for the linear Boltzmann equation: error analysis in the diffusion limit*, SIAM J. Numer. Anal. **59** (2021), no. 4, 2254–2285, DOI 10.1137/20M1380788. MR4301408

[13] P. A. M. Dirac, *Note on exchange phenomena in the Thomas atom*, Mathematical Proceedings of the Cambridge Philosophical Society, vol. 26, Cambridge University Press, 1930, pp. 376–385.

[14] L. Einkemmer, J. Hu, and J. Kusch, *Asymptotic-preserving and energy stable dynamical low-rank approximation*, SIAM J. Numer. Anal. **62** (2024), no. 1, 73–92, DOI 10.1137/23M1547603. MR4686848

[15] L. Einkemmer, J. Hu, and Y. Wang, *An asymptotic-preserving dynamical low-rank method for the multi-scale multi-dimensional linear transport equation*, J. Comput. Phys. **439** (2021), Paper No. 110353, 21, DOI 10.1016/j.jcp.2021.110353. MR4253315

[16] 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 J. Sci. Comput. **43** (2021), no. 5, B1057–B1080, DOI 10.1137/21M1392772. MR4315485

[17] L. Einkemmer and C. Lubich, *A low-rank projector-splitting integrator for the Vlasov-Poisson equation*, SIAM J. Sci. Comput. **40** (2018), no. 5, B1330–B1360, DOI 10.1137/18M116383X. MR3863075

[18] L. Einkemmer, A. Ostermann, and C. Scalone, *A robust and conservative dynamical low-rank algorithm*, J. Comput. Phys. **484** (2023), Paper No. 112060, 20, DOI 10.1016/j.jcp.2023.112060. MR4569231

[19] J. Frenkel, *Wave Mechanics*, Clarendon, 1934.

[20] L. Grasedyck, D. Kressner, and C. Tobler, *A literature survey of low-rank tensor approximation techniques*, GAMM-Mitt. **36** (2013), no. 1, 53–78, DOI 10.1002/gamm.201310004. MR3095914

[21] J.-L. Guermond and G. Kanschat, *Asymptotic analysis of upwind discontinuous Galerkin approximation of the radiative transport equation in the diffusive limit*, SIAM J. Numer. Anal. **48** (2010), no. 1, 53–78, DOI 10.1137/090746938. MR2608358

[22] J. S. Hesthaven and T. Warburton, *Nodal Discontinuous Galerkin Methods*, Texts in Applied Mathematics, vol. 54, Springer, New York, 2008. Algorithms, analysis, and applications, DOI 10.1007/978-0-387-72067-8. MR2372235

[23] T. Jahnke and W. Huisinga, *A dynamical low-rank approach to the chemical master equation*, Bull. Math. Biol. **70** (2008), no. 8, 2283–2302, DOI 10.1007/s11538-008-9346-x. MR2448010

[24] E. Kieri and B. Vandereycken, *Projection methods for dynamical low-rank approximation of high-dimensional problems*, Comput. Methods Appl. Math. **19** (2019), no. 1, 73–92, DOI 10.1515/cmam-2018-0029. MR3898206

[25] O. Koch and C. Lubich, *Dynamical low-rank approximation*, SIAM J. Matrix Anal. Appl. **29** (2007), no. 2, 434–454, DOI 10.1137/050639703. MR2318357

[26] J. Kusch, G. Ceruti, L. Einkemmer, and M. Frank, *Dynamical low-rank approximation for Burgers' equation with uncertainty*, Int. J. Uncertain. Quantif. **12** (2022), no. 5, 1–21, DOI 10.1615/int.j.uncertaintyquantification.2022039345. MR4466549

[27] E. W. Larsen and J. E. Morel, *Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes. II*, J. Comput. Phys. **83** (1989), no. 1, 212–236, DOI 10.1016/0021-9991(89)90229-5. MR1010164

[28] C. Lubich, *On variational approximations in quantum molecular dynamics*, Math. Comp. **74** (2005), no. 250, 765–779, DOI 10.1090/S0025-5718-04-01685-0. MR2114647

[29] C. Lubich and I. V. Oseledets, *A projector-splitting integrator for dynamical low-rank approximation*, BIT **54** (2014), no. 1, 171–188, DOI 10.1007/s10543-013-0454-0. MR3177960

[30] A. Mezzacappa, E. Endeve, O. E. Bronson Messer, and S. W. Bruenn, *Physical, numerical, and computational challenges of modeling neutrino transport in core-collapse supernovae*, Living Rev. Comput. Astrophys., **6** 2020, no. 1, 4.

[31] D. Mihalas and B. W. Mihalas, *Foundations of Radiation Hydrodynamics*, Oxford University Press, New York, 1984. MR781346

[32] Z. Peng and R. G. McClarren, *A high-order/low-order (HOLO) algorithm for preserving conservation in time-dependent low-rank transport calculations*, J. Comput. Phys. **447** (2021), Paper No. 110672, 22, DOI 10.1016/j.jcp.2021.110672. MR4316005

[33] Z. Peng and R. G. McClarren, *A sweep-based low-rank method for the discrete ordinate transport equation*, J. Comput. Phys. **473** (2023), Paper No. 111748, 18, DOI 10.1016/j.jcp.2022.111748. MR4509450

[34] Z. Peng, R. G. McClarren, and M. Frank, *A low-rank method for two-dimensional time-dependent radiation transport calculations*, J. Comput. Phys. **421** (2020), 109735, 18, DOI 10.1016/j.jcp.2020.109735. MR4132848

[35] B. Rivière, *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations: Theory and Implementation*, SIAM, 2008.

[36] S. Schotthöfer, E. Zangrando, J. Kusch, G. Ceruti, and F. Tudisco, *Low-rank lottery tickets: finding efficient low-rank neural networks via matrix differential equations*, Adv. Neural Inf. Process. Syst., **35** (2022), 20051–20063.

[37] C.-W. Shu, *Discontinuous Galerkin methods: general approach and stability*, Numerical Solutions of Partial Differential Equations, Adv. Courses Math. CRM Barcelona, Birkhäuser, Basel, 2009, pp. 149–201. MR2531713

Department of Mathematical Sciences, The University of Texas at El Paso, El Paso, Texas 79968
Email address: pyin@utep.edu

Mathematics in Computation Section, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831; and Department of Physics and Astronomy, University of Tennessee, Knoxville, 1408 Circle Drive, Knoxville, Tennessee 37996
Email address: endevee@ornl.gov

Mathematics in Computation Section, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831; and Department of Mathematics, University of Tennessee, Knoxville, 1403 Circle Drive, Knoxville, Tennessee 37996
Email address: hauckc@ornl.gov

Mathematics in Computation Section, Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831
Email address: schnakesr@ornl.gov