A MICRO-MACRO DECOMPOSED REDUCED BASIS METHOD FOR THE TIME-DEPENDENT RADIATIVE TRANSFER EQUATION

ZHICHAO PENG*, YANLAI CHEN†, YINGDA CHENG ‡, AND FENGYAN LI§

Abstract. Kinetic transport equations are notoriously difficult to simulate because of their complex multiscale behaviors and the need to numerically resolve a high dimensional probability density function. Past literature has focused on building reduced order models (ROM) by analytical methods. In recent years, there is a surge of interest in developing ROM using data-driven or computational tools that offer more applicability and flexibility. This paper is a work towards that direction.

Motivated by our previous work of designing ROM for the stationary radiative transfer equation in [34] by leveraging the low-rank structure of the solution manifold induced by the angular variable, we here further advance the methodology to the time-dependent model. Particularly, we take the celebrated reduced basis method (RBM) approach and propose a novel micro-macro decomposed reduced basis method (MMD-RBM). The MMD-RBM is constructed by exploiting, in a greedy fashion, the low-rank structures of both the micro- and macro-solution manifolds with respect to the angular and temporal variables. Our reduced order surrogate consists of: reduced bases for reduced order subspaces and a reduced quadrature rule in the angular space. The proposed MMD-RBM features several structure-preserving components: 1) an equilibrium-respecting strategy to construct reduced order subspaces which better utilize the structure of the decomposed system, and 2) a recipe for preserving positivity of the quadrature weights thus to maintain the stability of the underlying reduced solver. The resulting ROM can be used to achieve a fast online solve for the angular flux in angular directions outside the training set and for arbitrary order moment of the angular flux.

We perform benchmark test problems in 2D2V, and the numerical tests show that the MMD-RBM can capture the low rank structure effectively when it exists. A careful study in the computational cost shows that the offline stage of the MMD-RBM is more efficient than the proper orthogonal decomposition (POD) method, and in the low rank case, it even outperforms a standard full order solve. Therefore, the proposed MMD-RBM can be seen both as a surrogate builder and a low-rank solver at the same time. Furthermore, it can be readily incorporated into multi-query scenarios to accelerate problems arising from uncertainty quantification, control, inverse problems and optimization.

1. Introduction.

In this paper, we design a reduced order model (ROM) for a class of kinetic transport equation: the time-dependent radiative transfer equation (RTE), which provides prototype models for optical tomography [2], radiative transfer [40], remote sensing [43] and neutron transport [26] etc. The isotropic time-dependent RTE under the diffusive scaling is written as:

$$\varepsilon\partial_t f + \boldsymbol{v}\cdot\nabla_{\mathbf{x}} f = \frac{\sigma_s}{\varepsilon}(\langle f\rangle - f) - \varepsilon\sigma_a f + \varepsilon G. \tag{1.1}$$

It features three independent variables, $t\in\mathcal{R}^+, \mathbf{x}\in\Omega_{\mathbf{x}}, \boldsymbol{v}\in\Omega_v$, denoting the time, spatial location, and angular direction. For the full model considered in this paper, $\Omega_v = \mathbb{S}^2$ is the unit sphere. The equation models the transport and the interaction of the particles (e.g. photons) with the background media (e.g. through the scattering and absorption). The unknown $f(\mathbf{x},\boldsymbol{v},t)$ is the angular flux (also called the distribution of particles). $\mathcal{L}_{\text{collision}} f = \sigma_s(\langle f\rangle - f)$ is the scattering operator, where $\langle f\rangle = \frac{1}{|\Omega_v|}\int_{\Omega_v} f(\mathbf{x},\boldsymbol{v},t)d\boldsymbol{v}$ is the scalar flux (also the density) which is the average of f in the angular space. $G(\mathbf{x})$ is an isotropic source term. In (1.1), $\sigma_s(\mathbf{x})\geq 0$ and $\sigma_a(\mathbf{x})\geq 0$ are, respectively, the scattering and absorption cross sections. The Knudsen number ε is the non-dimensional mean free path of the particles. The main challenges for numerically solving this equation come from its high dimensional and multiscale nature. First, the angular flux f depends on the phase variable $(\mathbf{x},\boldsymbol{v})$ and the time. Therefore, any standard grid-based method will suffer from the curse of dimensionality. Second, the solution crosses different regimes thanks to its dependence on the non-dimensionalized mean free path ε. When ε is $O(1)$, the problem is transport dominant. When $\varepsilon\to 0$ and $\sigma_s > 0$, equation (1.1) converges to its diffusion limit:

$$\partial_t\rho - \nabla_{\mathbf{x}}\cdot(\sigma_s^{-1}D\nabla_{\mathbf{x}}\rho) = -\sigma_a\rho + G, \tag{1.2}$$

where $\rho(\mathbf{x},t) = \langle f\rangle$ and $D = \text{diag}(\langle v_x^2\rangle, \langle v_y^2\rangle, \langle v_z^2\rangle)$. This trans-regime behavior presents itself as both a challenge and an opportunity.

To leverage the opportunity presented by the inherent structure of the equation in the diffusive regime and address the challenge especially of high dimensionality, projection based ROMs and tensor decomposition based low rank algorithms have been designed for the stationary and time-dependent RTE. Projection based ROMs typically have two stages: an Offline stage, where a problem-specific low rank approximation is built, and an Online stage, where important physical quantities can be

*Department of Mathematics, Michigan State University, East Lansing, MI 48824 U.S.A. Email: pengzhic@msu.edu.

†Department of Mathematics, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747, USA. Email: yanlai.chen@umassd.edu. Research is partially supported by National Science Foundation grant DMS-2208277, and by the UMass Dartmouth Marine and UnderSea Technology (MUST) Research Program made possible via an Office of Naval Research grant N00014-20-1-2849.

‡Department of Mathematics, Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824 U.S.A. Email: ycheng@msu.edu. Research is supported by NSF grants DMS-2011838 and AST-2008004.

§Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, U.S.A. Email: lif@rpi.edu. Research is supported by NSF grant DMS-1913072.

predicted or reconstructed more efficiently. Tensor decomposition based low rank algorithms are offline-free but use less problem specific low rank approximations. The computational cost of the online stage of ROMs are usually determined by the low rank, while the computational cost of tensor decomposition based algorithms usually depend on the tensor format and the number of degrees of freedom in certain directions. One potential research direction lacking investigation is to combine the power of the two methods by utilizing tensor based low rank algorithms as the full order solver in the offline stage of ROMs. Along the line of low rank algorithms based on tensor decomposition, dynamical low rank algorithm (DLRA) [39, 14, 38] and the proper generalized decomposition (PGD) [1, 41, 13] have been designed. Projection based ROMs have also been actively developed in the recent few years, for example the proper orthogonal decomposition (POD) and its variations [5, 11, 12, 44, 3, 10, 20], the dynamical mode decomposition (DMD) [29, 30]. Among those work, the POD methods in [5, 45, 20] and our previous work in reduced basis method (RBM) for the steady state problem [34] make explicit use of the low rank structure of the solution manifold induced by the angular variable, namely, the ROM built is based on treating the angular variable as the "parameter" of the model. Once such ROM surrogate is constructed, it can be used to achieve a fast online calculation of the angular flux in an angular direction outside the training set. We will also show in this paper that a fast calculation of high order moments of the angular flux can be obtained by using the ROM surrogate. Moreover, the ROM can be further incorporated to multi-query scenarios to accelerate calculations in inverse problems and uncertainty quantification.

In this paper, we continue our effort in [34] and take the RBM approach [33, 42, 18], which is a projection-based model order reduction strategy for parametric problems and consists of Offline and Online stages. In the Offline stage, it constructs a low-dimensional reduced order subspace to approximate the underlying solution manifold of the parametric problem. In the Online stage, the reduced order solution for unseen parameter values is sought through a (Petrov-)Galerkin projection into the low-dimensional surrogate subspace constructed offline. RBM utilizes a greedy algorithm for constructing the surrogate subspace offline. It iteratively augments the reduced order subspace by greedily identifying the snapshot, via an error estimator or an error / importance indicator, corresponding to the most under-resolved parameter (were the current reduced space to be adopted) in the training set until the stopping criteria is satisfied.

While the angular variable is treated as the parameter of the model in our previous work in [34] for the stationary RTE, here for the time-dependent RTE, we regard both the angular $\boldsymbol{v}$ and temporal t variables as parameters and build a RBM by leveraging the low-rank structure of the $(\boldsymbol{v}, t)$-induced solution manifold. As observed in [34] for the stationary case, the solution of the time-dependent RTE corresponding to different angular directions $\boldsymbol{v}$ are not decoupled, due to the integral operator for the scattering. This makes our problem very different from the standard parametric problems the vanilla RBM is applied to. Compared to [34], the present work presents several significant algorithmic advances. Our full order and reduced order models are based on the micro-macro decomposition of the RTE [27] instead of the original form in (1.1) for directly solving f. To improve the performance in the diffusive and intermediate regime, we design an equilibrium-respecting strategy to construct reduced order subspaces which better utilize the structure of the decomposed system. We call the proposed method micro-macro decomposed reduced basis method (MMD-RBM). Furthermore, sampled angular variables are typically unstructured, and a direct robust and accurate quadrature rule to compute angular integrals is lacking. This is in particular crucial for time-dependent problems because it relates to the stability of the ROM. A recipe for constructing such quadrature rules preserving positivity of the weights is provided.

The rest of the paper is organized as follows. In Section 2, we present the micro-macro decomposition and the associated full order solver. In Section 3, we present Offline and Online stages of the MMD-RBM and estimate the computational cost. In Section 4, the performance of the proposed methods is demonstrated through a series of numerical examples. At last, we draw conclusions in Section 5.

2. Micro-macro decomposed RTE and its discretization.

The radiative transfer equation (RTE) in (1.1) is multiscale in nature. When $\varepsilon = O(1)$, it is transport dominant. On the other hand when $\varepsilon \to 0$, the model converges to its diffusion limit, and this can be illustrated through the micro-macro decomposition [27]. Define Π as the orthogonal projection onto the null space of the collision operator $\text{Null}(\mathcal{L}_{\text{collision}})$ in $L^2(\Omega_v)$. With the isotropic scattering being considered here, $\Pi f = \langle f \rangle$. We decompose f as $f = \Pi f + (I - \Pi)f = \rho(\mathbf{x}, t) + \varepsilon g(\mathbf{x}, \boldsymbol{v}, t)$, with $\rho(\mathbf{x}, t) = \langle f \rangle$ as the scalar flux (or called density). Equation (1.1) can then be rewritten as the micro-macro decomposed system:

$$\partial_t \rho + \nabla_{\mathbf{x}} \cdot \langle \boldsymbol{v} g \rangle = -\sigma_a \rho + G, \tag{2.1a}$$

$$\varepsilon^2 \partial_t g + \varepsilon (I - \Pi)(\boldsymbol{v} \cdot \nabla_{\mathbf{x}} g) + \boldsymbol{v} \cdot \nabla_{\mathbf{x}} \rho = -\sigma_s g - \varepsilon^2 \sigma_a g. \tag{2.1b}$$

As $\varepsilon \to 0$ and with $\sigma_s(\mathbf{x}) > 0$, (2.1b) becomes the local equilibrium

$$g = -\frac{1}{\sigma_s} \boldsymbol{v} \cdot \nabla_{\mathbf{x}} \rho. \tag{2.2}$$

Substitute (2.2) to (2.1a), we obtain the diffusion limit:

$$\partial_t \rho - \nabla_{\mathbf{x}} \cdot (\sigma_s^{-1} D \nabla_{\mathbf{x}} \rho) = -\sigma_a \rho + G,$$

where $D = \text{diag}(\langle v_x^2 \rangle, \langle v_y^2 \rangle, \langle v_z^2 \rangle)$.

2.1. Fully discretized micro-macro decomposed system.

When standard numerical methods are applied to solve (1.1), the computational cost can be prohibitive when $\varepsilon \ll 1$, as the mesh sizes smaller than ε are often needed for both accuracy and stability [6, 31]. A numerical method for (1.1) is said to be asymptotic preserving (AP) [22] if it preserves the asymptotic limit as $\varepsilon \to 0$ at the discrete level, namely, as $\varepsilon \to 0$ the method becomes a consistent and stable discretization for the limiting model. AP methods can work uniformly well for the model with a broad range of ε, particularly with $\varepsilon \ll 1$ on under-resolved meshes. This type of methods will be our choice as full order methods. In particular, in this work we adapt the

IMEX-DG-S method [37] to multiple dimensions. The method is AP, with desirable time step conditions for stability, specifically, it is unconditionally stable in the diffusive regime ($\varepsilon \ll 1$) and conditionally stable with a hyperbolic-type CFL condition in the transport regime ($\varepsilon = O(1)$). Alternatively, one can use other AP schemes based on the micro-macro decomposition as the full order model, such as [25, 21, 35], which can have different stability property in the diffusive regime ($\varepsilon \ll 1$).

In this work, we assume all unknowns are independent of the z variable, namely, $\partial_z \rho = \partial_z f = \partial_z g = 0$. With this, we consider $\Omega_{\mathbf{x}} = [x_L, x_R] \times [y_L, y_R]$ in two space dimensions (with $d = 2$) and $\Omega_v = \mathbb{S}^2$ as the angular space. The methodology developed here can be extended to $\Omega_{\mathbf{x}}$ in three dimensions straightforwardly. Next, we will present our full order method, starting from the time discretization.

Time discretization: To achieve unconditional stability in the diffusion dominant regime as well as the AP property, the time discretization is defined as follows. Given the solutions ρ^n and g^n at $t^n = n\Delta t$, we seek ρ^{n+1} and g^{n+1} such that

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} + \nabla_{\mathbf{x}} \cdot \langle \boldsymbol{v} g^{n+1} \rangle = -\sigma_a \rho^{n+1} + G^{n+1}, \tag{2.3a}$$

$$\varepsilon^2 \frac{g^{n+1} - g^n}{\Delta t} + \varepsilon (I - \Pi)(\boldsymbol{v} \cdot \nabla_{\mathbf{x}} g^n) + \boldsymbol{v} \cdot \nabla_{\mathbf{x}} \rho^{n+1} = -\sigma_s g^{n+1} - \varepsilon^2 \sigma_a g^{n+1}. \tag{2.3b}$$

As $\varepsilon \to 0$ and with $\sigma_s > 0$, (2.3b) becomes

$$g^{n+1} = -\frac{1}{\sigma_s} \boldsymbol{v} \cdot \nabla_{\mathbf{x}} \rho^{n+1}. \tag{2.4}$$

Substituting (2.4) into (2.3a), we obtain the limit of scheme (2.3) as $\varepsilon \to 0$,

$$\frac{\rho^{n+1} - \rho^n}{\Delta t} - \nabla_{\mathbf{x}} \cdot (\sigma_s^{-1} D \nabla_{\mathbf{x}} \rho^{n+1}) = -\sigma_a \rho^{n+1} + G^{n+1}.$$

This is nothing but the backward Euler method for the diffusion limit in (1.2). Hence, this time discretization is AP.

Angular discretization: In the angular space, we apply the discrete ordinates (S_N) method [40]. Let $\{\boldsymbol{v}_j\}_{j=1}^{N_v}$ be a set of quadrature points in Ω_v and $\{\omega_j\}_{j=1}^{N_v}$ be the corresponding quadrature weights, satisfying $\sum_{j=1}^{N_v} \omega_j = 1$. The semi-discrete system (2.3) is further discretized in the angular variable, following a collocation approach, by being evaluated at $\{\boldsymbol{v}_j\}_{j=1}^{N_v}$, with the integral operator $\langle \cdot \rangle$ approximated by its discrete analogue:

$$\langle f \rangle \approx \langle f \rangle_h = \sum_{j=1}^{N_v} \omega_j f(\cdot, \boldsymbol{v}_j, \cdot). \tag{2.5}$$

We require the quadrature rule to satisfy

$$\langle v_\xi v_\eta \rangle_h = \langle v_\xi v_\eta \rangle = \frac{1}{3} \delta_{\xi\eta},\ \xi, \eta\ \in \{x, y, z\},\ \delta_{\xi\eta} = \begin{cases} 1, & \xi = \eta \\ 0, & \xi \neq \eta \end{cases}, \tag{2.6}$$

so the coefficient matrix $D = \mathrm{diag}(\langle v_x^2 \rangle, \langle v_y^2 \rangle, \langle v_z^2 \rangle)$ will be exact, and the correct diffusion limit will be obtained for the full order model without cross-derivative terms (see Section 2.2). Particularly, with $\Omega_v = \mathbb{S}^2$, we use the Lebedev quadrature rule [24] in our fully-discrete method unless otherwise specified.

Spatial discretization: In the physical space, we apply a discontinuous Galerkin (DG) discretization. Letting

$$\mathcal{I}_h = \left\{ \mathcal{I}_{kl} = [x_{k-\frac{1}{2}}, x_{k+\frac{1}{2}}] \times [y_{l-\frac{1}{2}}, y_{l+\frac{1}{2}}], 1 \le k \le N_x, 1 \le l \le N_y \right\}$$

be a partition of the physical domain $\Omega_{\mathbf{x}}$, we define the discrete space as

$$U_h^K(\Omega_{\mathbf{x}}) := \{ u(\mathbf{x}) : u(\mathbf{x})|_{\mathcal{I}_{kl}} \in Q^K(\mathcal{I}_{kl}), 1 \le k \le N_x, 1 \le l \le N_y \},$$

where $Q^K(\mathcal{I}_{kl})$ is the bi-variate polynomial space with the degree in each direction at most K on the element $\mathcal{I}_{kl}$. We also write $\phi(x_0^{\pm}, y) = \lim_{x \to x_0^{\pm}} \phi(x, y)$ and $\phi(x, y_0^{\pm}) = \lim_{y \to y_0^{\pm}} \phi(x, y)$.

Let the numerical solution at t^n be $\rho_h^n(\cdot) \approx \rho(\cdot, t^n)$ and $g_{h,j}^n(\cdot) \approx g(\cdot, \boldsymbol{v}_j, t^n), \forall j = 1, \ldots, N_v$. With a DG discretization applied in space, we reach our fully-discrete scheme: given $\rho_h^n \in U_h^K, \{g_{h,j}^n\}_{j=1}^{N_v} \subset U_h^K$, we seek $\rho_h^{n+1} \in U_h^K, \{g_{h,j}^{n+1}\}_{j=1}^{N_v} \subset U_h^K$, satisfying

the following equations $\forall k = 1, \dots, N_x, l = 1, \dots, N_y$,

$$\int_{\mathcal{I}_{kl}} \frac{\rho_h^{n+1} - \rho_h^n}{\Delta t} \phi_h d\mathbf{x} + \sum_{\gamma=1}^{N_v} \omega_\gamma \int_{\mathcal{I}_{kl}} \left(\mathcal{D}_x^g(v_{\gamma,x} g_{h,\gamma}^{n+1}; \rho_h^{n+1}) + \mathcal{D}_y^g(v_{\gamma,y} g_{h,\gamma}^{n+1}; \rho_h^{n+1}) \right) \phi_h d\mathbf{x}$$

$$= \int_{\mathcal{I}_{kl}} (-\sigma_a \rho_h^{n+1} + G^{n+1}) \phi_h d\mathbf{x}, \quad \forall \phi_h \in U_h^K, \tag{2.7a}$$

$$\varepsilon^2 \int_{\mathcal{I}_{kl}} \frac{g_{h,j}^{n+1} - g_{h,j}^n}{\Delta t} \psi_h d\mathbf{x} + \int_{\mathcal{I}_{kl}} \left(v_{j,x} \mathcal{D}_x^- + v_{j,y} \mathcal{D}_y^- \right) \rho_h^{n+1} \, \psi_h d\mathbf{x}$$

$$+ \varepsilon \sum_{\gamma=1}^{N_v} (\delta_{j\gamma} - \omega_\gamma) \int_{\mathcal{I}_{kl}} \left(\mathcal{D}_x^{\mathrm{up}}(v_{\gamma,x}, g_{h,\gamma}^n) + \mathcal{D}_y^{\mathrm{up}}(v_{\gamma,y}, g_{h,\gamma}^n) \right) \psi_h d\mathbf{x}$$

$$= - \int_{\mathcal{I}_{kl}} (\sigma_s + \varepsilon^2 \sigma_a) g_{h,j}^{n+1} \psi_h d\mathbf{x}, \qquad \forall \psi_h \in U_h^K, \forall j = 1, \dots N_v. \tag{2.7b}$$

Here $\delta_{j\gamma}$ is the Kronecker delta, $\mathcal{D}_x^-(\cdot), \mathcal{D}_y^-(\cdot), \mathcal{D}_x^g(\cdot;\cdot), \mathcal{D}_y^g(\cdot;\cdot), \mathcal{D}_x^{\mathrm{up}}(\cdot,\cdot), \mathcal{D}_y^{\mathrm{up}}(\cdot,\cdot) \in U_h^K$ are all discrete (partial) derivatives, and they can be expressed in terms of $\mathcal{D}_x^\pm(\cdot), \mathcal{D}_y^\pm(\cdot) \in U_h^K$ that are defined as follows

$$\int_{\mathcal{I}_{kl}} \mathcal{D}_x^\pm \phi_h \psi_h d\mathbf{x} = - \int_{\mathcal{I}_{kl}} \phi_h \partial_x \psi_h d\mathbf{x} + \int_{y_{l-\frac{1}{2}}}^{y_{l+\frac{1}{2}}} \phi_h(x^\pm_{k+\frac{1}{2}}, y) \psi_h(x^-_{k+\frac{1}{2}}, y) dy$$

$$- \int_{y_{l-\frac{1}{2}}}^{y_{l+\frac{1}{2}}} \phi_h(x^\pm_{k-\frac{1}{2}}, y) \psi_h(x^+_{k-\frac{1}{2}}, y) dy, \qquad \forall \psi_h \in U_h^K, \tag{2.8a}$$

$$\int_{\mathcal{I}_{kl}} \mathcal{D}_y^\pm \phi_h \psi_h d\mathbf{x} = - \int_{\mathcal{I}_{kl}} \phi_h \partial_y \psi_h d\mathbf{x} + \int_{x_{k-\frac{1}{2}}}^{x_{k+\frac{1}{2}}} \phi_h(x, y^\pm_{l+\frac{1}{2}}) \psi_h(x, y^-_{l+\frac{1}{2}}) dx$$

$$- \int_{x_{k-\frac{1}{2}}}^{x_{k+\frac{1}{2}}} \phi_h(x, y^\pm_{l-\frac{1}{2}}) \psi_h(x, y^+_{l-\frac{1}{2}}) dx, \qquad \forall \psi_h \in U_h^K. \tag{2.8b}$$

With $\boldsymbol{v} \cdot \nabla_{\mathbf{x}} g^n$ in (2.3b) discretized following an upwind mechanism, we set

$$\mathcal{D}_x^{\mathrm{up}}(v_x, \phi_h) = v_x \mathcal{D}_x^\star(\phi_h), \qquad \text{with } \star = \begin{cases} -, & v_x \geq 0, \\ +, & v_x < 0, \end{cases}$$

$$\mathcal{D}_y^{\mathrm{up}}(v_y, \phi_h) = v_y \mathcal{D}_y^\star(\phi_h), \qquad \text{with } \star = \begin{cases} -, & v_y \geq 0, \\ +, & v_y < 0. \end{cases}$$

Moreover, we take

$$\mathcal{D}_\xi^g(v_{\gamma,\xi} g_{h,\gamma}; \rho_h) = v_{\gamma,\xi} \mathcal{D}_\xi^+ g_{h,\gamma} + \alpha_\xi \mathcal{D}_\xi^{\mathrm{jump}} \rho_h, \quad \text{with } \xi = x, y. \tag{2.9}$$

Here, $\mathcal{D}_x^{\mathrm{jump}}(\cdot) \in U_h^K$, given locally on the element $\mathcal{I}_{kl}$ by

$$\int_{\mathcal{I}_{kl}} \mathcal{D}_x^{\mathrm{jump}}(\rho_h) \psi_h d\mathbf{x} = \int_{y_{l-\frac{1}{2}}}^{y_{l+\frac{1}{2}}} \left(\rho_h(x^-_{k+\frac{1}{2}}, y) - \rho_h(x^+_{k+\frac{1}{2}}, y) \right) \psi_h(x^-_{k+\frac{1}{2}}, y) dy$$

$$- \int_{y_{l-\frac{1}{2}}}^{y_{l+\frac{1}{2}}} \left(\rho_h(x^-_{k-\frac{1}{2}}, y) - \rho_h(x^+_{k-\frac{1}{2}}, y) \right) \psi_h(x^+_{k-\frac{1}{2}}, y) dy, \quad \forall \psi_h \in U_h^K,$$

and equivalently,

$$\mathcal{D}_x^{\mathrm{jump}}(\rho_h) = \mathcal{D}_x^-(\rho_h) - \mathcal{D}_x^+(\rho_h).$$

Similarly

$$\mathcal{D}_y^{\mathrm{jump}}(\rho_h) = \mathcal{D}_y^-(\rho_h) - \mathcal{D}_y^+(\rho_h). \tag{2.10}$$

The jump operators are added in (2.9) to maintain accuracy in the case of the Dirichlet boundary conditions [7]. As shown in [7], the constants α_x, α_y in (2.9) need to be $O(1)$ and positive. In this paper, we consider the vacuum boundary condition. In all the discrete derivatives, when the data from the outside of the domain is needed for the solution, we directly set it as 0.

From here on, we refer to the fully-discrete method (2.7) along with (2.8)-(2.10) as the full order model denoted as FOM. Given that our plan is to treat the angular variable $\boldsymbol{v}$ as a parameter to formulate reduced order models, when we want to emphasize the set of the angular values $\mathcal{V}$ (and its "associated" quadrature weights in (2.5)) used to define (2.7), we also write it as FOM($\mathcal{V}$). As an example, we have $\mathcal{V} = \{\boldsymbol{v}_j\}_{j=1}^{N_v}$ for (2.7).

2.2. Matrix-vector form and Schur complement. Though $\langle vg\rangle$ is treated implicitly in (2.7a), we only need to invert a discrete heat operator for ρ with the help of the Schur complement, and this will be demonstrated next via the matrix-vector form of the scheme. Let $\{e_l(\mathbf{x})\}_{l=1}^{N_\mathbf{x}}$ be a basis of the DG space U_h^K, then ρ_h^n and $g_{h,j}^n$ can be expanded as $\rho_h^n(\mathbf{x}) = \sum_{l=1}^{N_\mathbf{x}} \rho_l^n e_l(\mathbf{x})$ and $g_{h,j}^n(\mathbf{x}) = \sum_{l=1}^{N_\mathbf{x}} g_{l,j}^n e_l(\mathbf{x})$. Defining $\boldsymbol{\rho}^n = (\rho_1^n, \dots, \rho_{N_\mathbf{x}}^n)^T$ and $\boldsymbol{g}_j^n = (g_{1,j}^n, \dots, g_{N_\mathbf{x},j}^n)^T$, we are ready to rewrite (2.7) into its matrix-vector formulation:

$$\mathcal{A}\left(\boldsymbol{\rho}^{n+1}, \boldsymbol{g}_1^{n+1}, \boldsymbol{g}_2^{n+1}, \dots \boldsymbol{g}_{N_v}^{n+1}\right)^T = \left(\mathbf{b}_\rho^n, \mathbf{b}_{g1}^n, \mathbf{b}_{g2}^n, \dots, \mathbf{b}_{g_{N_v}}^n\right)^T, \tag{2.11a}$$

$$\mathcal{A} = \begin{pmatrix} M + \Delta t\Sigma_a + \Delta t D^{\text{jump}} & \Delta t\omega_1(v_{1,x}D_x^+ + v_{1,y}D_y^+) & \dots & \Delta t\omega_{N_v}(v_{N_v,x}D_x^+ + v_{N_y,y}D_y^+) \\ \Delta t(v_{1,x}D_x^- + v_{1,y}D_y^-) & \Theta & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \Delta t(v_{N_v,x}D_x^- + v_{N_v,y}D_y^-) & 0 & \dots & \Theta \end{pmatrix}, \tag{2.11b}$$

$$\mathbf{b}_\rho^n = M\boldsymbol{\rho}^n + \Delta t\boldsymbol{G}^{n+1}, \tag{2.11c}$$

$$\mathbf{b}_{g_j}^n = \varepsilon^2 M\boldsymbol{g}_j^n - \varepsilon\Delta t\sum_{\gamma=1}^{N_v}(\delta_{j\gamma} - \omega_\gamma)(D_{x,v_{\gamma,x}}^{\text{up}} + D_{y,v_{\gamma,y}}^{\text{up}})\boldsymbol{g}_\gamma^n, \quad j = 1, \dots, N_v \tag{2.11d}$$

Here M is the mass matrix, Σ_s (resp. Σ_a) is the scattering (resp. absorption) matrix, D^{jump} is the jump matrix, $D_\xi^\pm$, $D_{\xi,v_{\gamma,\xi}}^{\text{up}}$ ($\xi = x, y, \gamma = 1, \dots, N_v$) are discrete derivatives matrices, all being of the size $N_\mathbf{x} \times N_\mathbf{x}$ ($N_\mathbf{x}$ is the number of degrees of freedom resulting from the spatial discretization), with their (kl)-th entry given as:

$$M_{kl} = \int_{\Omega_x} e_l e_k d\mathbf{x}, \quad (\Sigma_s)_{kl} = \int_{\Omega_x} \sigma_s e_l e_k d\mathbf{x}, \quad (\Sigma_a)_{kl} = \int_{\Omega_x} \sigma_a e_l e_k d\mathbf{x},$$

$$(D_\xi^\pm)_{kl} = \int_{\Omega_x} \mathcal{D}_\xi^\pm e_l e_k d\mathbf{x}, \quad (D_{\xi,v_{\gamma,\xi}}^{\text{up}})_{kl} = \int_{\Omega_x} \mathcal{D}_\xi^{\text{up}}(v_{\gamma,\xi}, e_l) e_k d\mathbf{x}, \quad (\text{with } \xi = x, y),$$

$$D^{\text{jump}} = \alpha_x(D_x^- - D_x^+) + \alpha_y(D_y^- - D_y^+).$$

In addition, $\boldsymbol{G}^{n+1}$ is the source vector, with its k-th entry $\int_{\Omega_x} G^{n+1} e_k d\mathbf{x}$, and $\Theta = \varepsilon^2(M + \Delta t\Sigma_a) + \Delta t\Sigma_s$. Using the standard choices of the basis of U_h^K (e.g. with the support of each basis function being one mesh element), the matrices M, Σ_s, Σ_a and Θ are block-diagonal. When the boundary conditions are periodic or vacuum in space, one can easily show $D_\xi^+ = -(D_\xi^-)^T$ with $\xi = x, y$ (see [37] for details).

To avoid inverting the big matrix $\mathcal{A}$ directly, we apply the Schur complement. Noticing that

$$\boldsymbol{g}_j^{n+1} = \Theta^{-1}\left(\mathbf{b}_{g_j}^n - \Delta t(v_{j,x}D_x^- + v_{j,y}D_y^-)\boldsymbol{\rho}^{n+1}\right), \quad \forall j = 1, \dots, N_v, \tag{2.12}$$

we eliminate $\boldsymbol{g}_j^{n+1}$ terms in the equation determined by the first line of $\mathcal{A}$ and obtain

$$\mathcal{H}\boldsymbol{\rho}^{n+1} = \widetilde{\mathbf{b}}_\rho^n, \tag{2.13}$$

where

$$\mathcal{H} = M + \Delta t\Sigma_a + \Delta t D^{\text{jump}} - \Delta t^2\sum_j \omega_j(v_{j,x}D_x^+ + v_{j,y}D_y^+)\Theta^{-1}(v_{j,x}D_x^- + v_{j,y}D_y^-)$$

$$= M + \Delta t\Sigma_a + \Delta t D^{\text{jump}} - \Delta t^2(\langle v_x^2\rangle_h D_x^+\Theta^{-1}D_x^- + \langle v_y^2\rangle_h D_y^+\Theta^{-1}D_y^-).$$

The second line above is a direct result of $\langle v_x v_y\rangle_h = \langle v_x v_y\rangle = 0$ in (2.6). With (2.13), we only need to invert a linear system (2.13) of a much smaller size for ρ. Moreover, $\mathcal{H}$ is a discrete heat operator, and it is symmetric positive definite due to $D_\xi^+ = -(D_\xi^-)^T$ with $\xi = x, y$, and hence can be efficiently inverted, e.g. by the conjugate gradient (CG) method with algebraic multigrid (AMG) preconditioners. Once $\boldsymbol{\rho}^{n+1}$ is available, $\boldsymbol{g}_j^{n+1}$ can be obtained from (2.12), and this can be carried out in a parallel fashion, given that Θ is block-diagonal and the equations (2.12) in j are decoupled.

2.3. Stability. When U_h^K with $K = 0$ is used (as numerically tested in Section 4), our FOM method is first order accurate, and its stability can be established by following similar techniques in [37], and this result will play an important role in the design of the ROM. The key to prove the stability in [37] is to introduce the following discrete energy:

$$E_h^n = ||\rho_h^n||^2 + \varepsilon^2\sum_{j=1}^{N_v}\omega_j||g_{h,j}^n||^2 + \Delta t\sum_{j=1}^{N_v}\omega_j\int_{\Omega_x}\sigma_s(g_{h,j}^n)^2 d\mathbf{x}, \tag{2.14}$$

where $||\cdot||$ is the standard L^2 norm in $L^2(\Omega_x)$. With $\sigma_s \geq 0$, the term E_h^n is non-negative and gives a well-defined energy. Using similar techniques in [36, 37], we can extend the Theorem 5.4 in [37] from 1D to 2D. We next state this result, presented in the context of the current work.

Theorem 2.1. ***(Stability condition)***[1] *Suppose* $\omega_j \geq 0, \forall 1 \leq j \leq N_v$, *and* $\sigma_s \geq \sigma_m > 0$. *Let* $h = \min(\min_{1\leq i\leq N_x}(x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}), \min_{1\leq i\leq N_y}(y_{i+\frac{1}{2}} - y_{i-\frac{1}{2}}))$, *we have that*

(1) when $\frac{\varepsilon}{\sigma_m h} \leq \frac{1}{4\max_{1\leq j\leq N_v}|\boldsymbol{v}_j|_\infty}$, $E_h^{n+1} \leq E_h^n \ \ \forall \Delta t > 0$;

(2) when $\frac{\varepsilon}{\sigma_m h} > \frac{1}{4\max_{1\leq j< N_v}|\boldsymbol{v}_j|_\infty}$, $E_h^{n+1} \leq E_h^n$ *under the time step condition*

$$\Delta t \leq \frac{\varepsilon h}{4\max_{1\leq j\leq N_v}|\boldsymbol{v}_j|_\infty - \sigma_m h/\varepsilon}.$$

The theorem implies that the scheme is unconditionally stable in the diffusive regime (i.e. when $\varepsilon/(\sigma_m h)$ is small enough), and the stability condition in the transport regime (i.e. $\varepsilon = O(1)$) is on the same level as the standard CFL condition $\Delta t = O(\varepsilon h)$.

3. The micro-macro decomposed reduced basis method.

Our proposed MMD-RBM algorithm consists of an Offline stage, which constructs the low dimensional subspaces and a reduced quadrature rule, and an Online stage which features a surrogate solver capable of efficiently computing moments of f and predicting the angular flux f corresponding to angular directions unseen during the Offline stage. In this section, we outline the entire algorithm in Section 3.1. In particular, we provide a high-level sketch in Figure 1 to assist reading. We then discuss each step of the Online and Offline stages in Sections 3.2 and 3.3, respectively. A computational complexity analysis is provided in Section 3.4 relating the cost of MMD-RBM with those of vanilla POD and brute force FOM.

3.1. Outline of the MMD-RBM algorithm.

The flowchart of the entire algorithm is summarized in Figure 1. Other than the clear distinction of Offline and Online stages, another feature of this algorithm is that

$$\mathrm{ROM}(\cdot;\cdot,\cdot),$$

representing our reduced order (thus online) solver, appears offline too, albeit with a pair of dynamically expanding surrogate spaces as the second and third input. Being a critical step in the greedy algorithm, this solver helps to recursively build the reduced parameter sets and augment the surrogate spaces in a greedy fashion. For this reason, before we dive into the detailed description of the Offline stage in Section 3.3, we first introduce in Section 3.2 this reduced formulation which corresponds to the full-order scheme (2.7).

Specifically, in Section 3.2, we introduce our projection-based reduced formulation $\mathrm{ROM}(\mathcal{V}; U^\rho_{h,r}, U^g_{h,r})$. Here $U^\rho_{h,r}$ is the reduced order space for ρ, $U^g_{h,r}$ is the reduced order space for g, and $\mathcal{V}$ is the angular set used in the angular discretization. We assume that there are quadrature weights $\{\omega_{\boldsymbol{v}}\}_{\boldsymbol{v}\in\mathcal{V}}$ associated with $\mathcal{V}$, and the discrete analogue $\langle\cdot\rangle_{h,\mathcal{V}}$ for the integral operator $\langle\cdot\rangle$. In the online surrogate solver, we solve $\mathrm{ROM}(\mathcal{V}_{\mathrm{rq}}; U^\rho_{h,r}, U^g_{h,r})$ with the terminal $U^\rho_{h,r}$ and $U^g_{h,r}$; and in the greedy sampling offline, we solve $\mathrm{ROM}(\mathcal{V}_{\mathrm{train}}; U^\rho_{h,r}, U^g_{h,r})$ with the current (and to-be-updated) $U^\rho_{h,r}$ and $U^g_{h,r}$. Here, $\mathcal{V}_{\mathrm{rq}}$ is the (usually unstructured) set of angular values identified by the Offline algorithm while $\mathcal{V}_{\mathrm{train}}$ denotes the (usually structured) training set of the angular directions specified at the beginning of the Offline algorithm.

In the Online stage (the pink block of the flowchart, to be described in Section 3.2), our ROM can be utilized to predict f at angular directions outside the training set as well as some moments of f with significantly fewer degrees of freedom. In the Offline stage (the blue block of the flowchart, to be described in Section 3.3), after initializing the quadrature nodes of the reduced quadrature rule $\mathcal{V}_{\mathrm{rq}}$ and the set of sampled parameters $\mathcal{T}^\rho_{\mathrm{rb}}$ and $\mathcal{TV}^g_{\mathrm{rb}}$, we use a greedy algorithm to iteratively construct the subspace $U^\rho_{h,r}$ and $U^g_{h,r}$. The main steps are

- described in Section 3.3.1, solving $\mathrm{ROM}(\mathcal{V}_{\mathrm{train}}; U^\rho_{h,r}, U^g_{h,r})$ to identify the most under-resolved angular and temporal samples, t^{new}_ρ for ρ and $(t^{\mathrm{new}}_g, \boldsymbol{v}^{\mathrm{new}}_g)$ pair for g, based on an importance indicator. Updating the set of sampled parameters $\mathcal{T}^\rho_{\mathrm{rb}}$ with t^{new}_ρ and $\mathcal{TV}^g_{\mathrm{rb}}$, in a symmetry-enhancing fashion, with $(t^{\mathrm{new}}_g, \pm\boldsymbol{v}^{\mathrm{new}}_g)$.
- described in Section 3.3.2, updating the corresponding reduced quadrature rule $\langle\cdot\rangle_{h,\mathcal{V}_{\mathrm{rq}}}$ preserving weight positivity via a novel least squares strategy.
- described in Section 3.3.3, updating the RB spaces $(U^\rho_{h,r}, U^g_{h,r})$.

3.2. Reduced MMD formulation and online functionalities.

Reduced MMD formulation $\mathrm{ROM}(\mathcal{V}; U^\rho_{h,r}, U^g_{h,r})$. We present the reduced MMD formulation in its matrix-vector form. Toward this end, we assume that $B_\rho \in \mathbb{R}^{N_{\boldsymbol{x}}\times r_\rho}$ and $B_g \in \mathbb{R}^{N_{\boldsymbol{x}}\times r_g}$ contain the orthonormal basis of $U^\rho_{h,r}$ and $U^g_{h,r}$, respectively, as their columns, and look for the reduced solution $\boldsymbol{\rho}_r = B_\rho \boldsymbol{c}_\rho$ for ρ, and $\boldsymbol{g}_{\boldsymbol{v},r} = B_g \boldsymbol{c}_{g_{\boldsymbol{v}}}$ for g at $\boldsymbol{v} \in \mathcal{V}$. More specifically: given $\boldsymbol{c}^n_\rho \in \mathbb{R}^{r_\rho}$ and $\boldsymbol{c}^n_{g_{\boldsymbol{v}}} \in \mathbb{R}^{r_g}\ \forall \boldsymbol{v} \in \mathcal{V}$, we seek $\boldsymbol{c}^{n+1}_\rho \in \mathbb{R}^{r_\rho}$

[1] This theorem can be established by following the proofs of Theorem 5.3 and Theorem 5.4 in [37] for the one spatial dimension case. The only difference is that, due to the extra dimension in space, there will be two extra terms similar to equations (5.7) and (5.8) of [37] in an equality similar to equation (5.5) of [36].

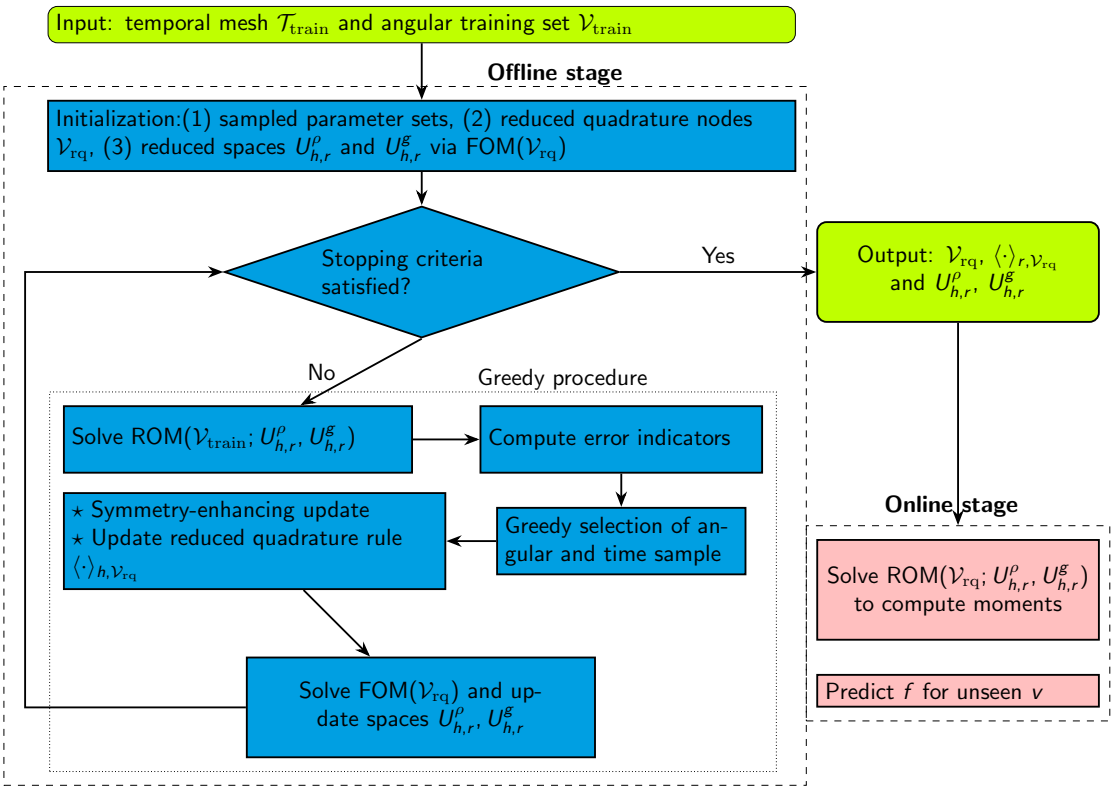

Fig. 1: The flowchart of the proposed MMD-RBM algorithm.

and $\boldsymbol{c}^{n+1}_{g_{\boldsymbol{v}}} \in \mathbb{R}^{r_g}$ $\forall \boldsymbol{v} \in \mathcal{V}$, satisfying

$$B_\rho^T M B_\rho \frac{\boldsymbol{c}^{n+1}_\rho - \boldsymbol{c}^n_\rho}{\Delta t} + \sum_{\boldsymbol{v}=(v_x,v_y)\in\mathcal{V}} \omega_{\boldsymbol{v}} B_\rho^T (v_x D_x^+ + v_y D_y^+) B_g \boldsymbol{c}^{n+1}_{g_\gamma}$$

$$+ B_\rho^T D^{\text{jump}} B_\rho \boldsymbol{c}^{n+1}_\rho = -B_\rho^T \Sigma_a B_\rho \boldsymbol{c}^{n+1}_\rho + B_\rho^T \boldsymbol{G}^{n+1}, \tag{3.1a}$$

$$\varepsilon^2 B_g^T M B_g \frac{\boldsymbol{c}^{n+1}_{g_{\boldsymbol{v}}} - \boldsymbol{c}^n_{g_{\boldsymbol{v}}}}{\Delta t} + \varepsilon \sum_{\boldsymbol{\mu}=(\boldsymbol{\mu}_x,\boldsymbol{\mu}_y)\in\mathcal{V}} (\delta_{\boldsymbol{v}\boldsymbol{\mu}} - \omega_{\boldsymbol{\mu}}) B_g^T (D^{\text{up}}_{x,\boldsymbol{\mu}_x} + D^{\text{up}}_{y,\boldsymbol{\mu}_y}) B_g \boldsymbol{c}^n_{g_\gamma}$$

$$+ B_g^T (v_x D_x^- + v_y D_y^-) B_\rho \boldsymbol{c}^{n+1}_\rho = -B_g^T (\Sigma_s + \varepsilon^2 \Sigma_a) B_g \boldsymbol{c}^{n+1}_{g_{\boldsymbol{v}}}. \tag{3.1b}$$

Similar to the FOM, the Schur complement can again be applied when solving the linear system (3.1), and the resulting $r_\rho \times r_\rho$ problem is in the form: $\mathcal{H}^\rho_r \boldsymbol{c}^{n+1}_\rho = \text{RHS}^n_{r,\rho}$. Here

$$\begin{aligned}\mathcal{H}^\rho_r =& B_\rho^T (M + \Delta t \Sigma_a + \Delta t D^{\text{jump}}) B_\rho \\ & - \Delta t^2 (\langle v_x^2 \rangle_{h,\mathcal{V}} D^+_{r,\rho g,x} (\Theta_{r,g})^{-1} D^-_{r,\rho g,x} + \langle v_y^2 \rangle_{h,\mathcal{V}} D^+_{r,\rho g,y} (\Theta_{r,g})^{-1} D^-_{r,\rho g,y}),\end{aligned} \tag{3.2}$$

where $D^+_{r,\rho g,\xi} = B_\rho^T D^+_\xi B_g$ and $D^-_{r,\rho g,\xi} = B_g^T D^-_\xi B_\rho$ with $\xi = x, y$ and $\Theta_{r,g} = B_g^T (\varepsilon^2 M + \Delta t \Sigma_s + \varepsilon^2 \Delta t \Sigma_a) B_g$, therefore $\mathcal{H}^\rho_r$ is symmetric positive definite, just like its FOM counterpart.

Online functionalities. This reduced MMD formulation is iteratively called in the Offline training stage, as to be seen in Section 3.3. At each iteration, the spaces $U^\rho_{h,r}$ and $U^g_{h,r}$ are augmented and the reduced quadrature rule $\langle\cdot\rangle_{h,\mathcal{V}_{\text{rq}}}$ is updated. At the end of this process with the terminal surrogate spaces $U^\rho_{h,r}$ and $U^g_{h,r}$, ROM($\mathcal{V}_{\text{rq}}$; $U^\rho_{h,r}$, $U^g_{h,r}$) can be utilized as a surrogate solver for two purposes. First, we can reconstruct the scalar flux ρ and high order moments of f; and second, we can predict solutions f for $\boldsymbol{v}$ unseen in the offline process. We next detail these two functionalities.

To reconstruct ρ and compute the high order moments, we solve ROM($\mathcal{V}_{\text{rq}}$; $U^\rho_{h,r}$, $U^g_{h,r}$) to compute $\boldsymbol{c}^n_\rho$ and $\boldsymbol{c}^n_g$. The scalar

flux, the first and the second order moments are approximated as:

$$\rho^n \approx B_\rho \boldsymbol{c}_\rho^n, \tag{3.3a}$$

$$\langle f^n v_\xi \rangle = \langle (\rho^n + \varepsilon g^n) v_\xi \rangle = \varepsilon \langle g^n v_\xi \rangle \approx \varepsilon B_g \langle v_\xi \boldsymbol{c}_{g_{\boldsymbol{v}}}^n \rangle_{h,\mathcal{V}_{\rm rq}}, \quad \xi = x, y, z, \tag{3.3b}$$

$$\begin{aligned} \langle f^n v_\xi v_\eta \rangle &= \langle (\rho^n + \varepsilon g^n) v_\xi v_\eta \rangle = \langle v_\xi v_\eta \rangle \rho^n + \varepsilon \langle g^n v_\xi v_\eta \rangle \\ &\approx \langle v_\xi v_\eta \rangle B_\rho \boldsymbol{c}_\rho^n + \varepsilon B_g \langle v_\xi v_\eta \boldsymbol{c}_{g_{\boldsymbol{v}}}^n \rangle_{h,\mathcal{V}_{\rm rq}}, \quad \xi, \eta = x, y, z. \end{aligned} \tag{3.3c}$$

Moreover, higher order moments can be computed similarly by integrating, *using the reduced quadrature rule* $\langle \cdot \rangle_{h,\mathcal{V}_{rq}}$, the corresponding quantities involving the reduced order solutions. We note that the advantages to reconstruct ρ and high order moments with $\mathrm{ROM}(\mathcal{V}_{\rm rq}; U_{h,r}^\rho, U_{h,r}^g)$ include computation efficiency, resulting from the adoption of the reduced quadrature rule, and memory saving[2].

When predicting f for an unseen angular direction $\boldsymbol{v}^{\rm un}$, we solve

$$\begin{aligned} &\varepsilon^2 B_g^T M B_g \frac{\boldsymbol{c}_{g_{\boldsymbol{v}^{\rm un}}}^{n+1} - \boldsymbol{c}_{g_{\boldsymbol{v}^{\rm un}}}^n}{\Delta t} + \varepsilon \left(B_g^T (D_{x,v_x^{\rm un}}^{\rm up} + D_{y,v_y^{\rm un}}^{\rm up}) B_g \boldsymbol{c}_{g_{\boldsymbol{v}}^{\rm un}}^n - \boldsymbol{c}_{\langle \boldsymbol{v} \cdot \nabla_x g \rangle}^{n,\rm upwind} \right) \\ &\quad + B_g^T (v_x^{\rm un} D_x^- + v_y^{\rm un} D_y^-) B_\rho \boldsymbol{c}_\rho^{n+1} = -B_g^T (\Sigma_s + \varepsilon^2 \Sigma_a) B_g \boldsymbol{c}_{g_{\boldsymbol{v}^{\rm un}}}^{n+1}, \\ &\text{with } \boldsymbol{c}_{\langle \boldsymbol{v} \cdot \nabla_x g \rangle}^{n,\rm upwind} = \langle B_g^T (D_{x,\boldsymbol{\mu}_x}^{\rm up} + D_{y,\boldsymbol{\mu}_y}^{\rm up}) B_g \boldsymbol{c}_{g_{\boldsymbol{v}}}^n \rangle_{h,\mathcal{V}_{\rm rq}}. \end{aligned} \tag{3.4}$$

In equation (3.4), c_ρ^{n+1} and $\boldsymbol{c}_{\langle v \cdot \nabla_x g \rangle}^{n,\rm upwind}$ can be obtained through pre- or on-the-fly computations by solving $\mathrm{ROM}(\mathcal{V}_{\rm rq}; U_{h,r}^\rho, U_{h,r}^g)$. The angular flux f for $\boldsymbol{v}^{\rm un}$ is approximated by $\boldsymbol{f}_{\boldsymbol{v}^{\rm un}}^n \approx B_\rho \boldsymbol{c}_\rho^n + \varepsilon B_g \boldsymbol{c}_{g_{\boldsymbol{v}^{\rm un}}}^n$.

Algorithm 3.1 Offline algorithm

1: **Input:** the training parameter sets $\mathcal{T}_{\rm train}$ and $\mathcal{V}_{\rm train}$
2: **Step 1 (initialization):** Initialize sampled parameter sets $\mathcal{T}_{\rm rb}^\rho = \emptyset$ and $\mathcal{TV}_{\rm rb}^g = \emptyset$, the reduced quadrature nodes set $\mathcal{V}_{\rm rq}$, and the reduced spaces $U_{h,r}^\rho$ and $U_{h,r}^g$.
3: **Step 2 (greedy iteration):**
4: **for** $i = 1$: max number of iterations **do**
5: **if** the stopping criteria are satisfied **then**
6: Stop.
7: **else**
8: (i) solve the reduced order problem $\mathrm{ROM}(\mathcal{V}_{\rm train}; U_{h,r}^\rho, U_{h,r}^g)$;
 (ii) compute the values of the L^1 importance indicators for ρ and g, and greedily pick the most under-resolved time $t_\rho^{\rm new}$ for ρ and the most under-resolved $(t_g^{\rm new}, \boldsymbol{v}_g^{\rm new})$ pair for g;
 (iii) update the parameter sets $\mathcal{T}_{\rm rb}^\rho$ and $\mathcal{TV}_{\rm rb}^g$ with symmetry-enhancing strategy;
 (iv) update the reduced quadrature set $\mathcal{V}_{\rm rq}$ and the corresponding quadrature rule $\langle \cdot \rangle_{h,\mathcal{V}_{\rm rq}}$;
 (v) perform the full order solve with the reduced quadrature rule $\mathrm{FOM}(\mathcal{V}_{\rm rq})$ and update the reduced spaces $U_{h,r}^\rho$ and $U_{h,r}^g$, and the corresponding basis matrices.
9: **end if**
10: **end for**
11: **Output:** a reduced order solver, determined by $\mathcal{V}_{\rm rq}$, $\langle \cdot \rangle_{h,\mathcal{V}_{\rm rq}}$, and $U_{h,r}^\rho$, $U_{h,r}^g$.

3.3. Offline algorithm. Summarized in Algorithm 3.1, the Offline algorithm starts with the training sets for t and $\boldsymbol{v}$, given as

$$\mathcal{T}_{\rm train} = \{t^n, 0 \le n \le N_t\}, \quad \mathcal{V}_{\rm train} = \{\boldsymbol{v}_j : 1 \le j \le N_v\},$$

with some prescribed cardinalities N_t and N_v. In preparing for the greedy iteration, we initialize the sampled parameter sets, $\mathcal{T}_{\rm rb}^\rho \subset \mathcal{T}_{\rm train}$ and $\mathcal{TV}_{\rm rb}^g \subset \mathcal{T}_{\rm train} \otimes \mathcal{V}_{\rm train}$, as empty. We use a low order Lebdev quadrature rule (i.e. nodes and weights) to

[2]For the FOM, the memory to save the time history of ρ and the high order moments is of $O(N_{\mathbf{x}} N_t)$. In the reduced order reconstruction, $O(N_{\mathbf{x}} r_\rho)$ and $O(N_{\mathbf{x}} r_g)$ are needed to save B_ρ and B_g, while $O(N_t r_\rho)$ and $O(N_t r_g)$ are assigned for the time history of $\boldsymbol{c}_\rho^n$ and moments of $\boldsymbol{c}_{g_{\boldsymbol{v}}}^n$ (e.g. $\langle v_x \boldsymbol{c}_{g_{\boldsymbol{v}}}^n \rangle_{h,\mathcal{V}_{\rm rq}}$). The total memory needed by the reduced order model to reconstruct the time history of ρ is of $O(r_\rho(N_{\mathbf{x}} + N_t))$, and that for the $k^{\rm th}$ order moments following (3.3) is of $O(r_g(N_{\mathbf{x}} + N_t))$ (k odd) and $O((r_\rho + r_g)(N_{\mathbf{x}} + N_t))$ (k even) respectively. These are all significantly smaller than their FOM counterparts assuming $r_\rho, r_g \ll N_{\mathbf{x}}$ or N_t.

initialize the set of reduced quadrature nodes $\mathcal{V}_{\rm rq}$ and the associated quadrature rule $\langle\cdot\rangle_{h,\mathcal{V}_{\rm rq}}$. Given $\mathcal{V}_{\rm rq}$, we call the full order solver FOM($\mathcal{V}_{\rm rq}$) with the integral replaced by $\langle\cdot\rangle_{h,\mathcal{V}_{\rm rq}}$, and obtain the numerical solution $\{\boldsymbol{\rho}^n, \boldsymbol{g}_{\boldsymbol{v}}^n : 1 \le n \le N_t, \forall \boldsymbol{v} \in \mathcal{V}_{\rm rq}\}$ which allows us to initiate the reduced spaces and the corresponding snapshot matrices

$$U_{h,r}^{\rho} = \text{span}\{\boldsymbol{\rho}^{N_t}\}, \quad U_{h,r}^{g} = \text{span}\{\boldsymbol{g}_{\boldsymbol{v}}^{N_t},\ \boldsymbol{v} \in \mathcal{V}_{\rm rq}\},$$

$$S_\rho = [\boldsymbol{\rho}^{N_t}] \in \mathbb{R}^{N_{\mathbf{x}} \times 1}, \quad S_g = [\boldsymbol{g}_{\boldsymbol{v}}^{N_t}]_{\boldsymbol{v} \in \mathcal{V}_{\rm rq}} \in \mathbb{R}^{N_{\mathbf{x}} \times |\mathcal{V}_{\rm rq}|}.$$

The initial basis matrix B_η is obtained by orthonormalizing the columns of S_η for $\eta = \rho, g$. We are now ready for details of the greedy iteration, with its main components presented below according to the order summarized at the end of Section 3.1.

3.3.1. L1 importance indicator and symmetry-enhancing parameter selection. At every greedy step, the most under-resolved parameter values for ρ and g (were the current reduced spaces to be adopted) will be determined by the L^1 importance indicator [9, 8]. Indeed, given the reduced order space $U_{h,r}^{\eta}$ ($\eta = \rho, g$), its snapshot and orthonormal matrices S_η and B_η, together with the sampled parameter set $\mathcal{T}_{\rm rb}^{\rho} \subset \mathcal{T}_{\rm train}$ and $\mathcal{TV}_{\rm rb}^{g} \subset \mathcal{T}_{\rm train} \otimes \mathcal{V}_{\rm train}$, we invoke ROM($\mathcal{V}_{\rm train}; U_{h,r}^{\rho}, U_{h,r}^{g}$) to obtain the reduced order solution $\{(\boldsymbol{\rho}_r^n, \boldsymbol{g}_{\boldsymbol{v},r}^n) :\ \forall n = 1, \dots N_t, \forall \boldsymbol{v} \in \mathcal{V}_{\rm train}\}$. They are expanded under the two basis systems as

$$\left\{\left(\boldsymbol{\rho}_r^n = B_\rho \boldsymbol{c}_\rho^n = S_\rho \tilde{\boldsymbol{c}}_\rho^n,\ \boldsymbol{g}_{\boldsymbol{v},r}^n = B_g \boldsymbol{c}_{g_{\boldsymbol{v}}}^n = S_g \tilde{\boldsymbol{c}}_{g_{\boldsymbol{v}}}^n\right) : \forall n = 1, \dots N_t, \forall \boldsymbol{v} \in \mathcal{V}_{\rm train}\right\}. \tag{3.5}$$

The L^1 importance indicator is defined as:

$$\Delta_\rho^n = ||\widetilde{\boldsymbol{c}}_\rho^n||_1, \quad \Delta_{g_{\boldsymbol{v}}}^n = ||\widetilde{\boldsymbol{c}}_{g_{\boldsymbol{v}}}^n||_1.$$

Here $||\cdot||_1$ represents the ℓ^1-norm. As shown in [8], $\widetilde{\boldsymbol{c}}_\rho^n$ (resp. $\widetilde{\boldsymbol{c}}_{g_{\boldsymbol{v}}}^n$) represents a Lagrange interpolation basis in the parameter induced solution space $\{\boldsymbol{\rho}_r^n : 1 \le n \le N_t\}$ (resp. $\{\boldsymbol{g}_{\boldsymbol{v},r}^n : 1 \le n \le N_t, \boldsymbol{v} \in \mathcal{V}_{\rm train}\}$), implying that the indicator Δ_ρ^n (resp. $\Delta_{g_{\boldsymbol{v}}}^n$) represents the corresponding Lebesgue constant. The following strategy to select the parameter sample then amounts to controlling the growth of the Lebesgue constants and hence is key toward accurate interpolation.

$$t_\rho^{\rm new} = \text{argmax}_{t^n \in \mathcal{T}_{\rm train} \setminus \mathcal{T}_{\rm rb}^{\rho}} \Delta_\rho^n,$$
$$(t_g^{\rm new}, \boldsymbol{v}_g^{\rm new}) = \text{argmax}_{(t^n, \boldsymbol{v}) \in \mathcal{T}_{\rm train} \otimes \mathcal{V}_{\rm train} \setminus \mathcal{TV}_{\rm rb}^{g}} \Delta_{g_{\boldsymbol{v}}}^n.$$

Once these greedy picks are determined, the parameter sample sets will be updated

$$\mathcal{T}_{\rm rb}^{\rho} \leftarrow \{t_\rho^{\rm new}\} \bigcup \mathcal{T}_{\rm rb}^{\rho}, \quad \mathcal{TV}_{\rm rb}^{g} \leftarrow \left\{(t_g^{\rm new}, \boldsymbol{v}_g^{\rm new}), (t_g^{\rm new}, -\boldsymbol{v}_g^{\rm new})\right\} \bigcup \mathcal{TV}_{\rm rb}^{g}.$$

Similar to the steady state problem [34], a symmetry enhancing strategy is applied when updating $\mathcal{TV}_{\rm rb}^{g}$ by adding both $\boldsymbol{v}_g^{\rm new}$ and its opposite angular direction $-\boldsymbol{v}_g^{\rm new}$. This strategy improves the robustness and accuracy of the reduced quadrature rule, especially in the early stage of the greedy algorithm.

Remark 3.1. *The main advantage of the L^1 importance indicator is that it is residual free and can be computed fast (also see (3.11)). One can alternatively use the residual as an error estimator. However, the RTE is a multiscale transport system and the residual of its numerical method is not a sharp error estimator. Sharper error estimators can be constructed for transport problems by solving the adjoint problems [19], and this requires extra cost and will not be pursued in this paper.*

3.3.2. Reduced quadrature rule construction. When $\boldsymbol{v}_g^{\rm new} \notin \mathcal{V}_{\rm rq}$, we update the set of the reduced quadrature nodes as

$$\mathcal{V}_{\rm rq} \leftarrow \{\boldsymbol{v}_g^{\rm new}, -\boldsymbol{v}_g^{\rm new}\} \cup \mathcal{V}_{\rm rq}.$$

Though with some symmetry built-in at each step, the angular samples in $\mathcal{V}_{\rm rq}$ that are greedily picked offline are in general unstructured. A stable and accurate numerical quadrature rule associated with these samples, although important to the robustness and accuracy of the proposed reduced order solver, may not naturally exist. To fill this void, we design a least squares strategy to construct a reduced quadrature rule, similar to that for mesh-free numerical methods [16] and further propose an algorithm capable of preserving weight positivity.

Theorem 3.2. *Given an integrable function $f(\boldsymbol{v}) : \mathbb{S}^2 \to \mathbb{R}$ and a positive integer M, let $Y_{m,l}$ be the real-valued spherical harmonic function of degree m and order l with $0 \le m \le M$ and $-m \le l \le m$. On a (possibly unstructured) grid $\mathcal{V}_{rq}$ with cardinality N_v^{rq} and nodes having spherical coordinates $\{(\theta_k, \phi_k)\}_{k=1}^{N_v^{rq}}$, the following reduced quadrature rule*

$$\langle f\rangle_{h,\mathcal{V}_{rq}} = \sum_{k=1}^{N_v^{rq}} \omega_k f(\boldsymbol{v}(\theta_k, \phi_k)),\ \textit{with}\ \omega_k = \frac{1}{\sqrt{4\pi}} \mathbb{I}_{1,k}^{\dagger} \tag{3.7}$$

has a degree of exactness M. Here $\mathbb{I}$ is a matrix of size $N_v^{rq} \times (M+1)^2$ with $\mathbb{I}_{ij} = Y_{ml}(\theta_i, \phi_i)$ and $j = m^2 + l + m + 1$. It is assumed $(M+1)^2 \le N_v^{rq}$.

Proof: We note that $\mathbb{S}^2 = \{\boldsymbol{v} = \boldsymbol{v}(\theta, \phi) = (\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta)), \theta \in [0, \pi], \phi \in [0, 2\pi]\}$ and the real-valued spherical harmonics form an orthogonal basis of $L^2(\mathbb{S}^2)$. We define the following ansatz of order M,

$$f_{\boldsymbol{\beta}}(\boldsymbol{v}(\theta, \phi)) = \sum_{m=0}^{M} \sum_{l=-m}^{m} \beta_{m,l} Y_{m,l}(\theta, \phi), \tag{3.8}$$

and seek a particular such function with coefficient being the solution to the least squares problem:

$$\boldsymbol{\beta}_{\mathrm{LS}} = \arg\min_{\boldsymbol{\beta}} \sum_{i=1}^{N_v^{\mathrm{rq}}} \left| f_{\boldsymbol{\beta}}(\boldsymbol{v}(\theta_i,\phi_i)) - f(\boldsymbol{v}(\theta_i,\phi_i)) \right|^2 = \arg\min_{\boldsymbol{\beta}} ||\mathbb{I}\boldsymbol{\beta} - \boldsymbol{f}||,$$

where $\mathbb{I} \in \mathbb{R}^{N_v^{\mathrm{rq}} \times (M+1)^2}$ and $\boldsymbol{f} \in \mathbb{R}^{N_v^{\mathrm{rq}}}$ satisfy $\mathbb{I}_{ij} = Y_{ml}(\theta_i,\phi_i)$, with $j = m^2 + l + m + 1$, and $\boldsymbol{f}_i = f(\boldsymbol{v}(\theta_i,\phi_i))$. One can easily see that $\boldsymbol{\beta}_{\mathrm{LS}} = \mathbb{I}^\dagger \boldsymbol{f}$, where $\mathbb{I}^\dagger$ is the pseudo inverse of $\mathbb{I}$. The integral $\langle f \rangle$ is now approximated by the reduced quadrature rule $\langle f \rangle_{h,\mathcal{V}_{\mathrm{rq}}}$ which is nothing but the exact integration of the least squares approximation

$$\begin{aligned}\langle f \rangle_{h,\mathcal{V}_{\mathrm{rq}}} =& \frac{1}{4\pi}\int_0^\pi \int_0^{2\pi} f_{\boldsymbol{\beta}_{\mathrm{LS}}}(\boldsymbol{v}(\theta,\phi))\sin(\theta) d\theta d\phi = \frac{1}{4\pi}\Big(\int_0^\pi \int_0^{2\pi} \beta_{\mathrm{LS},00} Y_{0,0}(\theta,\phi)\sin(\theta) d\theta d\phi \\ &+ \sum_{m=1}^{M}\sum_{l=-m}^{m} \int_0^\pi \int_0^{2\pi} \beta_{\mathrm{LS},ml} Y_{m,l}(\theta,\phi)\sin(\theta) d\theta d\phi \Big) \\ =& \frac{1}{\sqrt{4\pi}} \beta_{\mathrm{LS},00} = \sum_{k=1}^{N_v^{\mathrm{rq}}} \frac{1}{\sqrt{4\pi}} \mathbb{I}^\dagger_{1k} f(\boldsymbol{v}(\theta_k,\phi_k)).\end{aligned}$$

From the construction above, one can see that the reduced quadrature rule is exact for polynomials (in $\boldsymbol{v}$) up to degree M, hence of accuracy order M. □

We emphasize that, just like any numerical integration of interpolatory type, the weights are independent of the integrand f. In this work, we always assume $M \geq 3$. As a result, $\langle v_\xi^2 \rangle$ with $\xi = x, y, z$ are computed exactly and they will appear in the diffusion limit. Additionally $\langle v_x v_y \rangle = \langle v_x v_z \rangle = \langle v_y v_z \rangle = 0$ is also exactly computed, and this will ensure the absence of the cross-derivatives of second order in the reduced order problems (3.2), as illustrated in (2.13). We also note that the proposed algorithm can be easily generalized to the 1D slab geometry $\Omega_v = [-1, 1]$ and the unit circle $\Omega_v = \mathbb{S}^1$ by replacing the spherical harmonic expansion in (3.8) with expansions of Legendre polynomials and trigonometric functions, respectively.

While the construction of the reduced quadrature has spectral accuracy, it does not guarantee the associated quadrature weights to be non-negative. It is observed numerically that the reduced and full order solvers could blow up when some of quadrature weights are negative. The root of this instability is that the discrete energy $\mathcal{E}_h^n$ defined in (2.14) can be negative in the presence of negative quadrature weights. To preserve stability, we propose a strategy, described in Algorithm 3.2, to generate the reduced quadrature rule with non-negative weights. The basic idea is to decrease the order M, when negative weights are present, until either all the weights are non-negative for the first time or M reaches a prescribed minimal value $M_{\min} \geq 3$. If taking $M = M_{\min}$ still results in negative weights, we simply use the same quadrature rule as the previous greedy iteration and set the weights associated with the newly added angular samples to be 0. Recall that the initial quadrature rule is chosen as a low order Lebedev quadrature rule with positive quadrature weights. Therefore, the proposed strategy always results in non-negative reduced quadrature weights during the greedy iterations.

Algorithm 3.2 Iterative procedure to construct reduced quadrature rule with non-negative weights.

1: **Input**: Current sampled angular points $\mathcal{V}_{\mathrm{rq}} = \{\boldsymbol{v}_{k_j}\}_{j=1}^{N_v^{\mathrm{rq}}}$ and the sampled angular points for the previous iteration $\mathcal{V}_{\mathrm{rq}}^{\mathrm{old}}$. Let the reduced quadrature rule for $\mathcal{V}_{\mathrm{rq}}^{\mathrm{old}}$ be $\{\boldsymbol{v}_{k_j}^{\mathrm{old}}, \omega_j^{\mathrm{old}}\}_{j=1}^{N_v^{\mathrm{rq,old}}}$ with $\omega_j^{\mathrm{old}} \geq 0$, $\forall j$, the order $M_{\min}$ and $M_{\max}$.
2: Initialize the bool variable Failure = *true*.
3: **for** $M = M_{\max} : -1 : M_{\min}$ **do**
4: Use equation (3.7) to construct an order M reduced quadrature rule $\langle \cdot \rangle_{h,\mathcal{V}_{\mathrm{rq}}}$.
5: **if** All the quadrature weights are non-negative, **then**
6: set Failure = false, and break.
7: **end if**
8: **end for**
9: **if** Failure **then**
10: set the quadrature weight ω_j^{new} for $\boldsymbol{v}_{k_j} \in \mathcal{V}_{\mathrm{rq}}$ as

$$\omega_j^{\mathrm{new}} = \begin{cases} 0, & \text{if} \quad \boldsymbol{v}_{k_j} \notin \mathcal{V}_{\mathrm{rq}}^{\mathrm{old}}, \\ \omega_j^{\mathrm{old}}, & \text{if} \quad \boldsymbol{v}_{k_j} \in \mathcal{V}_{\mathrm{rq}}^{\mathrm{old}}. \end{cases}$$

11: **end if**
12: **Output**: the quadrature rule $\{\boldsymbol{v}_{k_j}, \omega_j^{\mathrm{new}}\}_{j=1}^{N_v^{\mathrm{rq}}}$ for $\mathcal{V}_{\mathrm{rq}}$ with non-negative weights.

3.3.3. Update of the reduced order spaces. Given the sampled parameter set $\{\mathcal{T}^\rho_{\text{rb}}, \mathcal{TV}^g_{\text{rb}}\}$, reduced quadrature nodes $\mathcal{V}_{\text{rq}}$ containing the $\boldsymbol{v}-$components of $\mathcal{TV}^g_{\text{rb}}$, and the associated quadrature rule $\langle\cdot\rangle_{h,\mathcal{V}_{\text{rq}}}$, we augment the reduced order space $U^\eta_{h,r}$ $(\eta = \rho, g)$ and its corresponding matrices S_η and B_η. Indeed, we perform FOM($\mathcal{V}_{\text{rq}}$) which is affordable thanks to the small size of $\mathcal{V}_{\text{rq}}$ to obtain the solution snapshots $\boldsymbol{\rho}^n, \boldsymbol{g}^n_{\boldsymbol{v}}, \forall n = 1, \ldots, N_t, \forall \boldsymbol{v} \in \mathcal{V}_{\text{rq}}$. We are then ready for the updates.

Update $U^\rho_{h,r}$ and B_ρ. This will be done in a straightforward manner, namely $U^\rho_{h,r} = \text{span}\{\boldsymbol{\rho}^m : t^m \in \mathcal{T}^\rho_{\text{rb}}\}$. Correspondingly, the snapshot matrix S_ρ is assembled. We then orthonormalize S_ρ through the (reduced) singular value decomposition (SVD):

$$S_\rho = B_\rho \Lambda_\rho V_\rho^T \in \mathbb{R}^{N_{\mathbf{x}} \times r_\rho}, \tag{3.9}$$

where $B_\rho \in \mathbb{R}^{N_{\mathbf{x}} \times r_\rho}$, $V_\rho \in \mathbb{R}^{r_\rho \times r_\rho}$, satisfying $B_\rho^T B_\rho = V_\rho^T V_\rho = I_{r_\rho}$, and $\Lambda_\rho \in \mathbb{R}^{r_\rho \times r_\rho}$ is a diagonal matrix. The columns of B_ρ form an orthonormal basis of $U^\rho_{h,r}$. As one will see, the singular values in Λ_ρ can be further utilized in the stopping criteria.

Update $U^g_{h,r}$ and B_g via an equilibrium respecting strategy. The update of the reduced order space for g is more subtle. Particularly, we set

$$U^g_{h,r} = \text{span}\left\{\{\Delta t \Theta^{-1} D_x^- \boldsymbol{\rho}^m, \Delta t \Theta^{-1} D_y^- \boldsymbol{\rho}^m : t^m \in \mathcal{T}^\rho_{\text{rb}}\} \cup \{\boldsymbol{g}^m_{\boldsymbol{v}} : (t^m, \boldsymbol{v}) \in \mathcal{TV}^g_{\text{rb}}\}\right\}.$$

That is, the reduced order space for g includes not only the sampled g-snapshots but also the scaled discrete derivatives of the sampled ρ-snapshots. Correspondingly, the snapshot matrix S_ρ is assembled which is further orthonormalized through its own SVD

$$S_g = B_g \Lambda_g V_g^T \in \mathbb{R}^{N_{\mathbf{x}} \times r_g}, \tag{3.10}$$

where $B_g \in \mathbb{R}^{N_{\mathbf{x}} \times r_g}$, $V_g \in \mathbb{R}^{r_g \times r_g}$, satisfying $B_g^T B_g = V_g^T V_g = I_{r_g}$. The columns of B_g form an orthogonal basis of $U^g_{h,r}$.

Fast computation of L^1 error indicator. Using the SVD in (3.9) and (3.10), one can show that $\widetilde{\boldsymbol{c}}^n_\rho$ and $\widetilde{\boldsymbol{c}}^n_{g_{\boldsymbol{v}}}$ in (3.5) satisfy

$$\widetilde{\boldsymbol{c}}^n_\rho = V_\rho \Lambda_\rho^{-1} c^n_\rho, \quad \widetilde{\boldsymbol{c}}^n_{g_{\boldsymbol{v}}} = V_g \Lambda_g^{-1} c^n_{g_{\boldsymbol{v}}},$$

and as a result Δ^n_ρ and $\Delta^n_{g_{\boldsymbol{v}}}$ can be computed efficiently as

$$\Delta^n_\rho = ||V_\rho \Lambda_\rho^{-1} \boldsymbol{c}^n_\rho||_1 \quad \text{and} \quad \Delta^n_{g_{\boldsymbol{v}}} = ||V_g \Lambda_g^{-1} \boldsymbol{c}^n_{g_{\boldsymbol{v}}}||_1. \tag{3.11}$$

Remark 3.3. *The equilibrium respecting strategy is designed to improve the performance of our method especially in the diffusive regime. To see the motivation, note that as $\varepsilon \to 0$ and with $\sigma_s > 0$, we have*

$$\boldsymbol{g}^m_{\boldsymbol{v}} \to -\Sigma_s^{-1}(v_x D_x^- + v_y D_y^-)\boldsymbol{\rho}^m.$$

That is, in the diffusion limit, $\boldsymbol{g}^m$ is a linear combination of the scaled derivatives of $\boldsymbol{\rho}^n$. In general, ε is small in the diffusive regime yet nonzero, and one would want to consider the relation in (2.12) instead. Hence $\Delta t \Theta^{-1} D_x^- \boldsymbol{\rho}^m$ and $\Delta t \Theta^{-1} D_y^- \boldsymbol{\rho}^m$ are included to enrich the reduced order space for g. Another benefit of such enrichment over including $\Sigma_s^{-1} D_x^- \boldsymbol{\rho}^m$ and $\Sigma_s^{-1} D_y^- \boldsymbol{\rho}^m$ is to be able to handle the case when σ_s is zero in some subregion(s) and the associated Σ_s is singular. It is easy to see that $\lim_{\varepsilon \to 0} \Theta = \Delta t \Sigma_s$.

Remark 3.4. *In this paper we always add scaled ρ derivative terms to the subspace for g. To improve the efficiency one can add scaled derivative terms only at (x, t) where the solution is close to the local equilibrium. One potential strategy is to utilize indicators measuring the distance between the solution and the local equilibrium. Such indicators have been designed in hybrid solvers adapting between kinetic and fluid/diffusion solvers [15, 23]. We first identify (x, t) corresponding to solutions which are close enough to the local equilibrium and only add scaled derivative terms at those points. Another approach is to apply the reduced basis element method [28] which divides the computational domain to small subdomains and build basis functions for each subdomain. We only add derivative terms for subdomains where scattering effect is strong. These approaches to adaptively adding ρ derivatives are left for future investigation.*

Remark 3.5. *We orthornormalize S_ρ and S_g with SVD, and one can alternatively orthornormalize them with the QR decomposition. The SVD decomposition provides singular values which can be utilized in the stopping criteria and furnishes a mechanism for efficiently computing the error indicators.*

Remark 3.6. *We note that the dimension of $U^g_{h,r}$ resulting from the first greedy iteration will be smaller than its initial dimension. After the first greedy iteration, $U^g_{h,r}$ is determined by the sampled parameter set $\mathcal{T}^\rho_{rb}$ and $\mathcal{TV}^g_{rb}$, while the initial $U^g_{h,r}$ is not and its initial dimension is $|\mathcal{V}_{rq}|$. In the first greedy iteration, $\max\{|\mathcal{T}^\rho_{rb}|, |\mathcal{TV}^g_{rb}|\} < |\mathcal{V}_{rq}|$, and this leads to the reduction of dimension of $U^g_{h,r}$ compared with its initialization.*

3.3.4. Stopping criteria. The L^1 importance indicator identifies the most under-resolved parameter sample(s), but it does not inform us the magnitude of the error. To effectively stop the Offline greedy algorithm, we design the following two-fold stopping criteria. The first criterion, based on the spectral ratio, measures how much new information is added in each greedy iteration. The second criterion, an approximate relative error at the final time, can be computed efficiently. The Offline greedy algorithm stops when both criteria are satisfied.

1. **Spectral ratio stopping criterion:** Similar to [34], we use the spectral ratio as one stopping criterion measuring how much new information is gained by expanding the reduced subspaces. Suppose we are in the m-th greedy iteration, with all notation now having a superscript m. Let Λ_ρ^m and Λ_g^m be the diagonal matrix from the SVD in (3.9) and (3.10), with the last diagonal entry as $\sigma_{r_\rho^m}^{\rho,m}$ and $\sigma_{r_g^m}^{g,m}$, respectively. We define two spectral ratios:

$$\text{ratio}_\rho^m = \frac{\sigma_{r_\rho^m}^{\rho,m}}{Tr(\Lambda_\rho^m)}, \qquad \text{ratio}_g^m = \frac{\sigma_{r_g^m}^{g,m}}{Tr(\Lambda_g^m)},$$

and check whether $\max\{\text{ratio}_\rho^m, \text{ratio}_g^m\} < \text{tol}_{\text{ratio}}$ is satisfied.
The spectral ratio criterion itself does not directly estimate the error in the reduced order approximations. For that, we propose the second criterion.

2. **Approximate relative error at the final time with a coarse mesh in Ω_v:** Recall that in each greedy iteration, we have two sets of approximations for ρ and $g(\cdot,\boldsymbol{v},\cdot)$ $\forall \boldsymbol{v} \in \mathcal{V}_{\text{rq}}$. One set, denoted as $\rho_{h,r}^n, g_{h,\boldsymbol{v},r}^n$ $\forall \boldsymbol{v} \in \mathcal{V}_{\text{train}}$, is obtained by calling the reduced order solve $\text{ROM}(\mathcal{V}_{\text{train}}; U_{h,r}^\rho, U_{h,r}^g)$ in the greedy sampling. The other set, denoted as $\rho_{h,\mathcal{V}_{\text{rq}}}^{n,\text{FOM}}, g_{h,\boldsymbol{v},\mathcal{V}_{\text{rq}}}^{n,\text{FOM}}$ $\forall \boldsymbol{v} \in \mathcal{V}_{\text{rq}}$, is obtained when updating the reduced order spaces by calling the full order solve $\text{FOM}(\mathcal{V}_{\text{rq}})$, with a reduced quadrature rule associated with $\mathcal{V}_{\text{rq}}$. Based on these approximations, we define the following to measure the relative errors at the final time t^{N_t}:

$$\text{Estimator}_\rho = \frac{||\rho_{h,r}^{N_t} - \rho_{h,\mathcal{V}_{\text{rq}}}^{N_t,\text{FOM}}||}{||\rho_{h,\mathcal{V}_{\text{rq}}}^{N_t,\text{FOM}}||}, \tag{3.12a}$$

$$\text{Estimator}_f = \max_{\boldsymbol{v}\in\mathcal{V}_{\text{rq}}\cap\mathcal{V}_{\text{train}}} \frac{||\rho_{h,r}^{N_t} + \varepsilon g_{h,\boldsymbol{v},r}^{N_t} - \rho_{h,\mathcal{V}_{\text{rq}}}^{N_t,\text{FOM}} - \varepsilon g_{h,\boldsymbol{v},\mathcal{V}_{\text{rq}}}^{N_t,\text{FOM}}||}{||\rho_{h,\mathcal{V}_{\text{rq}}}^{N_t,\text{FOM}} + \varepsilon g_{h,\boldsymbol{v},\mathcal{V}_{\text{rq}}}^{N_t,\text{FOM}}||}, \tag{3.12b}$$

and check whether $\text{Estimator}_\rho < \text{tol}_{\text{error},\rho}$ and $\text{Estimator}_f < \text{tol}_{\text{error},f}$ are satisfied.

The reason why we still need the spectral ratio criterion is that $\mathcal{V}_{\text{rq}}$ is a coarse mesh in Ω_v, and in the early stage of the greedy algorithm, the full order solution associated with this mesh may not be accurate enough to approximate the full order solution corresponding to the training set which has high resolution in Ω_v. We also want to point out that this error approximation strategy can not be used in the greedy sampling step, as we need an error indicator for all the $v \in \mathcal{V}_{\text{train}}$ while the full order solution is only available for $\boldsymbol{v} \in \mathcal{V}_{\text{rq}}$ which have already been sampled.

3.4. Computational cost.

Now, we summarize the computational cost of the Online and Offline stages. We will start with the computational cost of the reduced order problem $\text{ROM}(\mathcal{V}; U_{h,r}^\rho, U_{h,r}^g)$, which will be used both online and offline. This cost consists of two parts. Firstly, before time marching begins, one needs to assemble the reduced order discrete operators such as $B_\rho^T M B_\rho$, $D_{r,\rho g,x}^{\pm}$ etc, and the leading order of the cost is $O(\max\{r_\rho, r_g\}^2 N_{\mathbf{x}})$. Additionally, one needs to invert $\Theta_{r,g}$ and $\mathcal{H}_r^\rho$. With Cholesky factorization, the associated cost will be $O(r_g^3)$ and $O(r_\rho^3)$, respectively. Secondly, in each time step, with the precomputed Cholesky factor, the cost to solve (3.1) for $\boldsymbol{c}_\rho^{n+1}$ is $O(r_\rho^2)$, and the cost to update $\boldsymbol{c}_{g_{\boldsymbol{v}}}^{n+1}$ for all $\boldsymbol{v} \in \mathcal{V}$ based on the known $\boldsymbol{c}_\rho^{n+1}$ is $O(\max(r_\rho, r_g) r_g |\mathcal{V}|)$. Hence the total cost over N_t time steps is $O\big((r_\rho^2 + \max(r_\rho, r_g) r_g |\mathcal{V}|) N_t\big)$ once the reduced order operators are computed prior to the time marching.

Online Cost. The computational cost of the Online stage comes from solving the $\text{ROM}(\mathcal{V}_{\text{rq}}; U_{h,r}^\rho, U_{h,r}^g)$ in (3.1) from $t = 0$ to $N_t \Delta t$, and it is $O((r_\rho^2 + \max(r_\rho, r_g) r_g N_v^{\text{rq}}) N_t)$ with $N_v^{\text{rq}} = |\mathcal{V}_{\text{rq}}|$. The computational cost to predict f for an unseen angular direction from $n = 0$ to N_t by solving (3.4) is $O(\max(r_\rho, r_g) r_g N_t)$. Here we assume that the reduced order operators are available.

Offline Cost. We denote the reduced orders for ρ and g in the m-th greedy iteration as r_ρ^m and r_g^m, and the number of reduced quadrature nodes by $N_{v,m}^{\text{rq}}$. We let $r_m = \max(r_\rho^m, r_g^m)$ and $N_v^{\text{train}} = |\mathcal{V}_{\text{train}}|$. The cost of the m-th iteration of the offline greedy procedure in Algorithm 3.1 is summarized in Table 1, in particular the total computational cost of the Offline stage of the m-th iteration is

$$\sum_{m=1}^{N_{\text{iter}}} \left(O(r_m^2 (N_{\mathbf{x}} + N_v^{\text{train}} N_t)) + O(N_{v,m}^{\text{rq}} N_{\mathbf{x}} N_t) \right).$$

To estimate the overall offline cost, we assume that the final reduced orders are r_ρ and r_g, and let $r = \max(r_\rho, r_g)$. Given that the total number of greedy iterations N_{iter} scales linearly with r, that r_m scales linearly with m, and that in the worst scenario $N_{v,m}^{\text{rq}} (\leq N_v^{\text{train}})$ scales linearly with m, we conclude that

$$\text{Offline time of MMD-RBM} = O(r^3 (N_{\mathbf{x}} + N_v^{\text{train}} N_t)) + O(r^2 N_{\mathbf{x}} N_t)). \tag{3.13}$$

To put this estimate into context, we compare it with the costs of the POD and the full order model. The offline cost of the vanilla POD is dominated by computing the SVD of the snapshot matrix which is of size $N_{\mathbf{x}} \times (N_t N_v^{\text{train}})$. That cost (of obtaining U and Σ in $U\Sigma V^T$) is $O(\max(N_{\mathbf{x}}, N_v^{\text{train}} N_t) \times (\min(N_{\mathbf{x}}, N_v^{\text{train}} N_t))^2)$ [17]. Therefore, the relative offline computational time of the MMD-RBM and the vanilla POD is

$$\frac{\text{Offline time of MMD-RBM}}{\text{Offline time of vanilla POD}} = O\left(\frac{r^2}{N_v^{\text{train}} N_\star} + \frac{r^3}{N_\star^2} \right),$$

	Leading order of the cost
Greedy sampling:	
Assemble reduced order operators	$O(r_m^2 N_{\mathbf{x}})$
Compute Cholesky factorization of $\mathcal{H}_r^\rho$ and $\Theta_{r,g}$	$O(r_m^3)$
Compute ROM$(\mathcal{V}_{\text{train}}; U_{h,r}^\rho, U_{h,r}^g)$ and error indicators	$O(r_m^2 N_v^{\text{train}} N_t)$
Update $\mathcal{V}_{\mathbf{rq}}$ and $\langle\cdot\rangle_{h,\mathcal{V}_{\mathbf{rq}}}$ if necessary	$O(N_{v,m}^{\text{rq}})$
Update reduced order spaces and basis:	
Solve FOM$(\mathcal{V}_{\text{rq}})$ with AMG- preconditioned CG	$O(N_{v,m}^{\text{rq}} N_{\mathbf{x}} N_t)$
Update basis with SVD	$O(r_m^2 N_{\mathbf{x}})$
Check stopping criteria	$O(r_m + N_{\mathbf{x}})$
Total cost for the m-th iteration	$O(r_m^2(N_{\mathbf{x}} + N_v^{\text{train}} N_t)) + O(N_{v,m}^{\text{rq}} N_{\mathbf{x}} N_t)$

Table 1: The computational cost of the m-th greedy iteration of the Offline algorithm.

	Leading order of the cost
Solving ROM$(\mathcal{V}_{\text{rq}}; U_{h,r}^\rho, U_{h,r}^g)$	$O(r^2 N_v^{\text{rq}} N_t)$

Table 2: The computational cost of the Online algorithm.

where $N_\star = \min(N_{\mathbf{x}}, N_v^{\text{train}} N_t)$. Moreover, we have

$$\frac{\text{Offline time of MMD-RBM}}{\text{Time of solving FOM}(\mathcal{V}_{\text{train}})} = O\left(\frac{r^2}{N_v^{\text{train}}} + \frac{r^3}{N_{\mathbf{x}}} + \frac{r^3}{N_v^{\text{train}} N_t}\right).$$

Remark 3.7. *SVD can be computed incrementally [4], and hence the POD can be more efficient. If the low rank of the snapshot matrix, which is determined by the tolerance in the incremental SVD, is r, the associated cost will be $O(N_{\mathbf{x}} N_v^{\text{train}} N_t r)$. With the same r, the relative offline computation between our method and the POD with the incremental SVD is*

$$\frac{\text{Offline time of MMD-RBM}}{\text{Offline time of POD with incremental SVD}} = O\left(\frac{r}{N_v^{\text{train}}} + \frac{r^2}{N_v^{\text{train}} N_t} + \frac{r^2}{N_{\mathbf{x}}}\right).$$

One can see that as long as $r \ll \min(\sqrt{N_{\mathbf{x}}}, N_v^{\text{train}}, \sqrt{N_v^{\text{train}} N_t})$, the Offline stage of our method is faster than the POD method with the incremental SVD.

Remark 3.8. *In this work, we build the reduced basis spaces for both ρ and g. Given the computational cost of the full order solve for ρ is much smaller than that for the microscopic part, in practice one can choose to build the reduced basis space only for g, especially when the cost of the full order solve for ρ is deemed affordable.*

4. Numerical examples.

We demonstrate the performance of the proposed MMD-RBM through a series of numerical examples. Throughout this section, the angular training set $\mathcal{V}_{\text{train}}$ is the set of $N_v = 590$ Lebedev quadrature points. We use piece-wise constant polynomials, i.e. $K = 0$ in space. When σ_s is constant, we use the following time step to guarantee stability,

$$\Delta t = \begin{cases} h, & \text{if} \quad \varepsilon < 0.25\sigma_s h \\ 0.25 \min(\frac{h}{\sqrt{2}}, \frac{\varepsilon h}{\sqrt{2}\sigma_s}), & \text{otherwise,} \end{cases}$$

where $h = \min(\min_{1\le i\le N_x}(x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}), \min_{1\le i\le N_y}(y_{i+\frac{1}{2}} - y_{i-\frac{1}{2}}))$. When σ_s is spatially dependent, we use the smallest time step size allowed by all σ_s values. Throughout this section, vacuum boundary conditions are considered. The constants in the numerical flux (2.9) are taken to be $\alpha_x = 1/\langle v_x^2\rangle_h$ and $\alpha_y = 1/\langle v_y^2\rangle_h$. We measure the absolute errors and the relative errors of the scalar flux ρ and first order moment $\langle \boldsymbol{v} f\rangle$ as follows, by evaluating the difference between the reduced order solution and a reference solution which is computed by the full order solver with $N_v^{\text{test}} = 2072$ Lebedev points denoted collectively as $\mathcal{V}_{\text{test}}$,

$$\text{(4.1a)} \qquad \mathcal{E}_\rho = \sqrt{\Delta t \sum_{n=1}^{N_t} ||\rho_{h,\text{ROM}}^n - \rho_{h,\text{FOM}}^n||^2}, \qquad \mathcal{R}_\rho = \frac{\mathcal{E}_\rho}{\sqrt{\Delta t \sum_{n=1}^{N_t} ||\rho_{h,\text{FOM}}^n||^2}},$$

$$\text{(4.1b)} \qquad \mathcal{E}_{\langle \boldsymbol{v} f\rangle} = \sqrt{\Delta t \sum_{n=1}^{N_t} ||\langle \boldsymbol{v} f\rangle_{h,\text{ROM}}^n - \langle \boldsymbol{v} f\rangle_{h,\text{FOM}}^n||^2}, \qquad \mathcal{R}_{\langle \boldsymbol{v} f\rangle} = \frac{\mathcal{E}_{\langle \boldsymbol{v} f\rangle}}{\sqrt{\Delta t \sum_{n=1}^{N_t} ||\langle \boldsymbol{v} f\rangle_{h,\text{FOM}}^n||^2}}.$$

Here $\|\cdot\|$ denotes the L^2 norm which is computed as $\|\rho\| = \sqrt{\int_{\Omega_x} \rho^2 d\mathbf{x}}$ for the scalar function ρ and $\|\langle \boldsymbol{v} f\rangle\| = \sqrt{\int_{\Omega_x} \langle \boldsymbol{v}_x f\rangle^2 + \langle \boldsymbol{v}_y f\rangle^2 d\mathbf{x}}$ for the vector function $\langle \boldsymbol{v} f\rangle = (\langle \boldsymbol{v}_x f\rangle, \langle \boldsymbol{v}_x f\rangle)^T$. Moreover, we have $|\mathcal{V}_{\text{test}} \backslash \mathcal{V}_{\text{train}}| = 2058$. To demonstrate the ability of our method to predict the angular fluxes at angular directions outside the training set, we solve for $\{f(\boldsymbol{v}) : \boldsymbol{v} \in \mathcal{V}_{\text{test}}\}$ with our ROM and evaluate the worst case absolute and relative errors,

$$\mathcal{E}_f = \max_{\boldsymbol{v}} \sqrt{\Delta t \sum_{n=1}^{N_t} ||f^n_{h,\boldsymbol{v},\text{ROM}} - f^n_{h,\boldsymbol{v},\text{FOM}}||^2}, \ \mathcal{R}_f = \frac{\mathcal{E}_f}{\max_{\boldsymbol{v}} \sqrt{\Delta t \sum_{n=1}^{N_t} ||f^n_{h,\boldsymbol{v},\text{FOM}}||^2}}.$$

We recall that r_ρ and r_g are the dimensions of the reduced order subspace for ρ and g. N_v^{rq} is the number of nodes in the reduced quadrature rule. Finally, we keep track of the data compression efficiency of our ROM via recording the compression ratio (C-R)

$$\text{C-R} = \frac{\text{DOFs of ROM}(\mathcal{V}_{\text{rq}}; U^\rho_{h,r}, U^g_{h,r})}{\text{DOFs of FOM}(\mathcal{V}_{\text{train}})} = \frac{r_\rho + N_v^{\text{rq}} r_g}{(N_v^{\text{train}} + 1) N_{\mathbf{x}}}.$$

All these quantities will appear in the tables of this section documenting the performance of the proposed MMD-RBM on various examples. We implement our solvers in the `Julia` programming language. When comparing offline computational cost with the vanilla POD in Section 4.1, the code was run on Michigan State University's HPCC cluster. All the other tests were performed on a Macbook Air laptop with a M1 chip.

4.1. Homogeneous media. In the first example, we consider a homogeneous media with $\sigma_s = 1$ and $\sigma_a = 0$ on the computational domain $[0, 2]^2$, uniformly partitioned into 80×80 rectangular elements. We adopt an initial condition $f(\mathbf{x}, \boldsymbol{v}, 0) = 0$ and a Gaussian source $G(x) = \exp\left(-100((x-1)^2 + (y-1)^2)\right)$. Different values of the Knudsen number $\varepsilon = 1.0$ (transport regime), $\varepsilon = 0.1$ (intermediate regime) and $\varepsilon = 0.005$ (diffusive regime) are considered to benchmark the performance of the proposed algorithm. The final time is $T = 0.25$ for $\varepsilon = 1.0$ and 0.1, and it is $T = 1.5$ for $\varepsilon = 0.005$. The reduced quadrature rule and reduced spaces are initialized with 26 Lebedev points. For the stopping criteria, we set $\text{tol}_{\text{ratio}}$ as $1\text{e}-4$, $\text{tol}_{\text{error},\rho} = 1.0\%$, and $\text{tol}_{\text{error},f} = 2.0\%$.

Performance of the MMD-RBM: The results of the MMD-RBM are presented in Table 3 and Figure 2. In the top row of Figure 2, we observe that the reduced order solutions match the full order solutions well. As shown in Table 3, the MMD-RBM achieves small relative errors in the scalar flux, the first order moment, and f (w.r.t $\boldsymbol{v} \in \mathcal{V}_{\text{test}}$). The C-R in the ROM is consistently below 0.08%. The reduced dimensions r_ρ and r_g decrease as ε decreases showcasing our method's capability of numerically capturing the fact that the problem approaches its diffusive limit.

	r_ρ	r_g	N_v^{rq}	C-R	$\mathcal{E}_\rho$	$\mathcal{R}_\rho$	$\mathcal{E}_{\langle \boldsymbol{v} f\rangle}$	$\mathcal{R}_{\langle \boldsymbol{v} f\rangle}$	$\mathcal{E}_f$	$\mathcal{R}_f$
$\varepsilon = 1$	13	52	48	0.07%	1.29e-5	0.22%	1.99e-5	1.29%	1.21e-4	1.74%
$\varepsilon = 0.1$	8	32	40	0.03%	1.44e-5	0.48%	6.48e-6	1.34%	1.05e-4	3.16%
$\varepsilon = 0.005$	3	12	32	0.01%	7.86e-5	0.48%	1.29e-6	1.43%	7.90e-5	0.48%

Table 3: Dimensions of the reduced order subspaces, r_ρ, r_g, the number of reduced quadrature nodes N_v^{rq}, the testing error and the compression ratio for the homogeneous media example with the MMD-RBM.

In the middle row of Figure 2, we present the training history of convergence. The relative training errors at the final time are defined as

$$\mathcal{R}^{N_t}_\rho = ||\rho^{N_t}_{h,\text{ROM}} - \rho^{N_t}_{h,\text{FOM}}||/||\rho^{N_t}_{h,\text{FOM}}||, \quad \mathcal{E}^{N_t}_f = \max_{v \in \mathcal{V}_{\text{train}}} ||f^{N_t}_{h,\boldsymbol{v},\text{ROM}} - f^{N_t}_{h,\boldsymbol{v},\text{FOM}}||/||f^{N_t}_{h,\boldsymbol{v},\text{FOM}}||. \tag{4.2}$$

The training errors at the final time and the error estimators in (3.12) are plotted with respect to the number of greedy iterations. We can see that as the number of greedy iterations grows, our estimators approximate the relative training errors at the final time well. Overall, the relative training errors for ρ and f decrease. In the bottom row of Figure 2, we plot the error history, as time evolves, of ρ, $\langle \boldsymbol{v} f\rangle$ and f (w.r.t $\boldsymbol{v} \in \mathcal{V}_{\text{test}}$). It is clear that, across different regimes, the errors either grow and then plateaus at the level of the prescribed error threshold, or decrease from that level.

In Figure 3, we present the sampled angular points when the stopping criteria are satisfied. The number of quadrature points in the reduced quadrature rule generated by MMD-RBM are 48 for $\varepsilon = 1$, 40 for $\varepsilon = 0.1$ and 32 for $\varepsilon = 0.005$. We can see that the sample points are fairly uniform on the sphere for this homogeneous case.

Benefit of the equilibrium-respecting strategy: We demonstrate the benefit of the equilibrium respecting strategy, that is the inclusion of $\{\Delta t \Theta^{-1} D_x^- \boldsymbol{\rho}^m, \Delta t \Theta^{-1} D_y^- \boldsymbol{\rho}^m, t^m \in \mathcal{T}^\rho_{\text{rb}}\}$ when updating the reduced order space $U^g_{h,r}$. Without these extra functions, we report in Table 4 the dimensions of the reduced order subspaces and the errors when the stopping criteria are the same. Comparing with Table 3, we see that when $\varepsilon = 0.1$ and $\varepsilon = 0.005$ including derivatives of ρ in $U^g_{h,r}$ leads to smaller values of r_ρ, N_v^{rq} and comparable errors. Having smaller r_ρ values is particularly beneficial since the cost of solving the reduced order problem for one time step scales roughly as $O(r^3 N_v^{\text{rq}})$ and the size of the reduced order operator in (3.2) is $r_\rho \times r_\rho$. This advantage is particularly pronounced in the more diffusive regime with $\varepsilon = 0.005$.

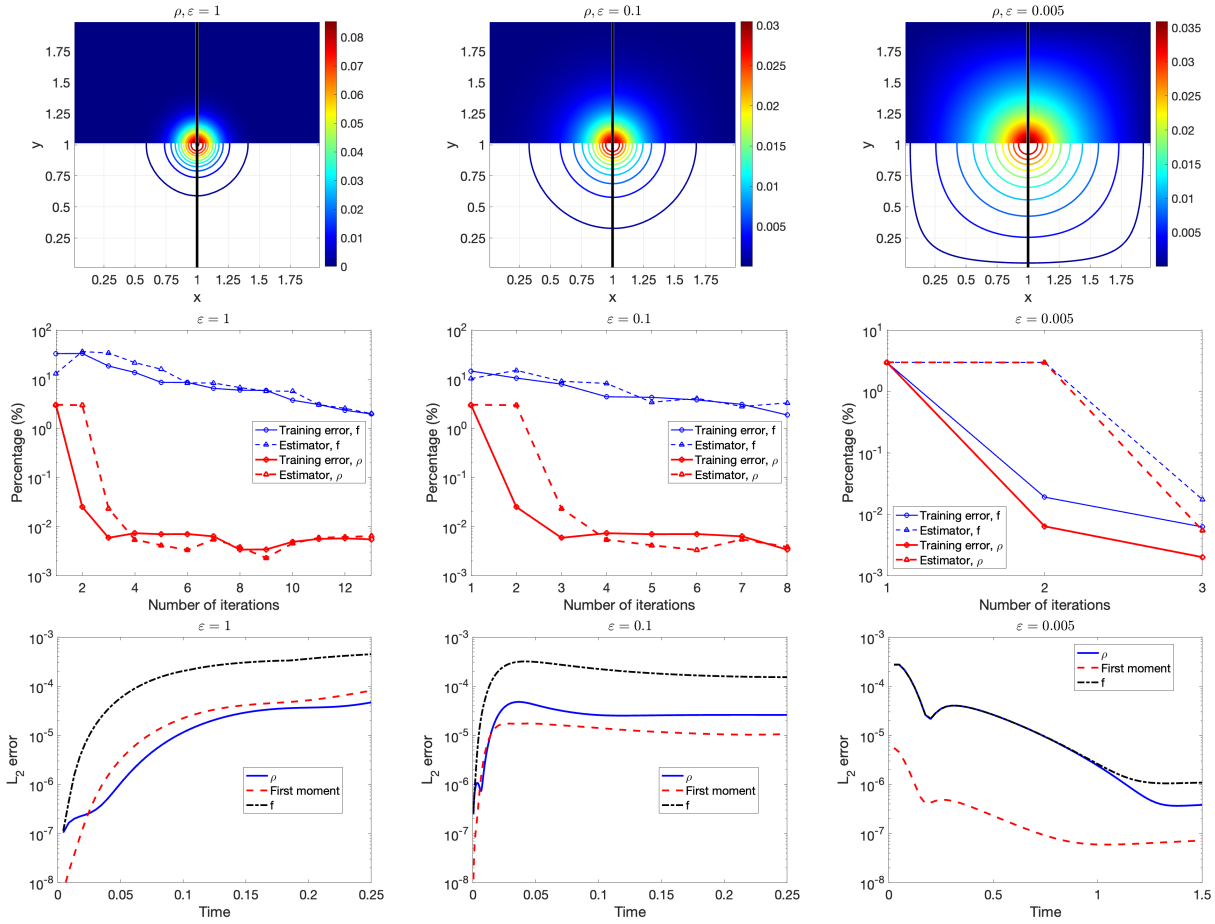

Fig. 2: Results for the homogeneous media example. Shown on the top are the reduced order solutions (left) and the full order solutions (right). In the middle row are the relative training errors of ρ and f at the final time and values of our error estimators. Shown on the bottom are the error histories with respect to time, when we compute the scalar flux ρ, first order moment $\langle \boldsymbol{v} f \rangle$ and predict f at unseen angular directions $\boldsymbol{v} \in \mathcal{V}_{\text{test}}$.

The cost of the Offline stage: In Figure 4, the offline computational time of our MMD-RBM is reported along with the computational time of FOM($\mathcal{V}_{\text{train}}$) and a vanilla POD strategy that computes the SVD of all the snapshots from FOM($\mathcal{V}_{\text{train}}$). All reported times are normalized by that of the full order solve in each case. Here, for comparison purpose, we implement the offline algorithm with 50 or 100 greedy iterations even though the stopping criteria are satisfied much sooner. For the first 50 iterations, we see that the offline computational time of the MMD-RBM scales roughly as r^2 (with $r = r_\rho + r_g$) which is faster than the $O(r^3)$ cost suggested by (3.13). As shown in the bottom right picture of Figure 4, the offline cost transitions from $O(r^2)$ to $O(r^3)$ as greedy procedure continues to 100 iterations, and it eventually scales slightly close to $O(r^3)$. We also label the location, via a vertical line, when the stopping criteria are satisfied. For all ε's, the offline cost of our method is smaller than the cost of vanilla POD. Moreover, for $\varepsilon = 0.005$, it is even smaller than the time of FOM($\mathcal{V}_{\text{train}}$). This shows the effectiveness of the greedy RB procedure in producing a low rank numerical solver.

4.2. Anisotropic initial condition. To demonstrate the ability of our method in adaptively sampling physically important angular directions, we consider the initial condition with anisotropy in the angular variable for g, namely, $g(x, y, \boldsymbol{v}, 0) = u(\boldsymbol{v}(\theta, \phi))\rho(x, y, 0)$ with

$$\rho(x, y, 0) = \begin{cases} \exp(-1.0/(0.5 - x^2 - y^2)), & \text{if } x^2 + y^2 < 0.5, \\ 0.0, & \text{else} \end{cases} \qquad \text{and}$$

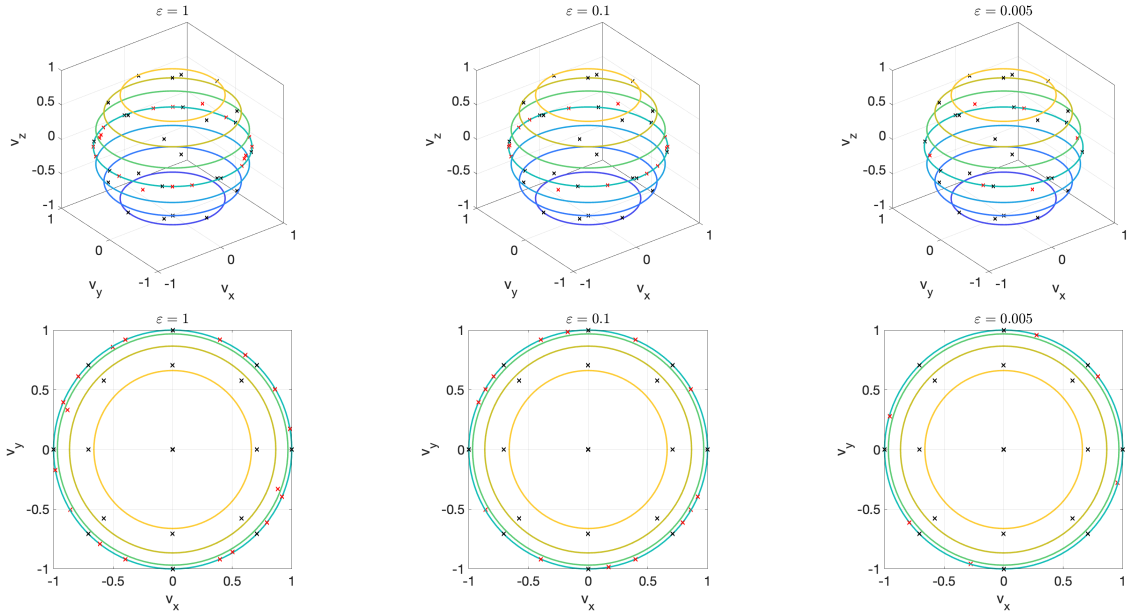

Fig. 3: The reduced quadrature nodes on the unit sphere (Black for points in the initial reduced quadrature nodes, and Red for those sampled by the greedy algorithm) and these nodes with a view from the north pole. $\varepsilon = 1.0, 0.1, 0.005$ from left to right.

	r_ρ	r_g	N_v^{rq}	C-R	$\mathcal{E}_\rho$	$\mathcal{R}_\rho$	$\mathcal{E}_{\langle \boldsymbol{v} f \rangle}$	$\mathcal{R}_{\langle \boldsymbol{v} f \rangle}$	$\mathcal{E}_f$	$\mathcal{R}_f$
$\varepsilon = 1$	14	28	52	0.04%	1.01e-5	0.18%	2.05e-5	1.33%	1.40e-4	2.01%
$\varepsilon = 0.1$	16	32	50	0.04%	2.15e-5	0.72%	8.19e-6	1.70%	3.96e-5	1.20%
$\varepsilon = 0.005$	9	18	38	0.02%	2.18e-5	0.13%	4.77e-7	0.53%	2.18e-5	0.13%

Table 4: Dimensions of the reduced order subspaces, r_ρ, r_g, the number of reduced quadrature nodes N_v^{rq}, the testing error and the compression ratio for the homogeneous media example with the ROM constructing the reduced space for g only with snapshots of g.

$$u(\boldsymbol{v}(\theta,\phi)) = \begin{cases} \exp\left(\frac{-1}{\frac{\pi^2}{16} - (\phi - \frac{\pi}{4})^2}\right), & \text{if } v_x > 0, v_y > 0, \\ -\exp\left(\frac{-1}{\frac{9\pi^2}{16} - (\phi + \frac{3\pi}{4})^2}\right), & \text{if } v_x < 0, v_y < 0, \\ 0.0, & \text{else.} \end{cases}$$

The computational domain is $[-1,1]^2$. The Knudsen number is $\varepsilon = 1.0$ and the final time is $T = 0.5$. As shown in the top left picture of Figure 5, $u(\boldsymbol{v})$ in the initial condition $g(\mathbf{x}, \boldsymbol{v}, 0)$ has more features when v_x and v_y are both positive or negative. We set $\text{tol}_{\text{ratio}} = 1\text{e}-4$, $\text{tol}_{\text{error},\rho} = 1.25\%$ and $\text{tol}_{\text{error},f} = 1.25\%$. The initial reduced quadrature rule is a Lebedev quadrature with 26 points. We consider different scattering cross sections $\sigma_s = 5,\ 1,\ 0.01$ with zero absorption $\sigma_a = 0$. Our MMD-RBM produces less than 1.44% relative error when reconstructing ρ online and less than 2.27% relative error when predicting f for unseen angular directions. In Figure 5, we also present $\int_{\Omega_x} g(\mathbf{x}, \boldsymbol{v}, 0.5) d\mathbf{x}$ and the sampled angular directions. When $\sigma_s = 5$, $\int_{\Omega_x} g(\mathbf{x}, \boldsymbol{v}, 0.5) d\mathbf{x}$ is almost isotropic w.r.t $\boldsymbol{v}$ due to the strong scattering. Indeed, the sampled angular directions are more uniformly distributed. As σ_s becomes smaller, the problem becomes more transport dominant and we observe that more angular directions are sampled in the first and third quadrants, where g has more features.

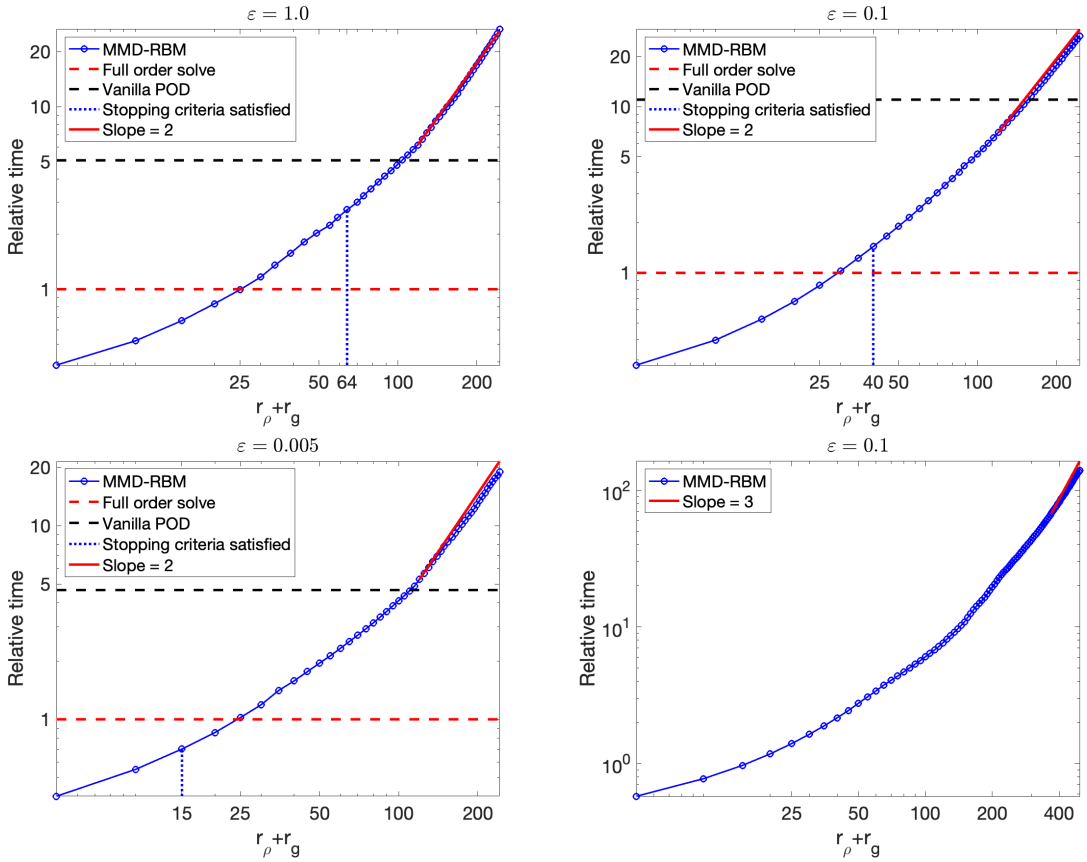

Fig. 4: Relative offline computational time with respect to the reduced order $r_\rho + r_g$ for the homogeneous media example. Note the computational time is normalized by the full order solve in each case. Bottom right: 100 greedy iterations; Others: 50 greedy iterations.

4.3. A multiscale problem with a spatially dependent scattering. Now, we consider a spatially-dependent scattering cross section [14]

$$\sigma_s(x,y) = \begin{cases} 0.999r^4(r+\sqrt{2})^2(r-\sqrt{2})^2 + 0.001, \text{ with } r = \sqrt{x^2+y^2} < 1, \\ 1, \text{ otherwise,} \end{cases}$$

on the computational domain $[-1,1]^2$ with $\varepsilon = 0.01$. The effective Knudsen number for this problem ε/σ_s smoothly varies from 10 to 0.01 indicating a smooth transition from a transport dominant region in the center to a scattering dominant region in the outer part of the computational domain. The initial value for this problem is $f(\mathbf{x},\boldsymbol{v},0) = \frac{5}{\pi}\exp(-25(x^2+y^2))$. We use a uniform mesh of 80×80 uniform rectangular elements to partition the computational domain. The final time is $T = 0.05$. The parameters in the stopping criteria are $\text{tol}_{\text{ratio}} = 1\text{e}-4$, $\text{tol}_{\text{error},\rho} = 1.5\%$ and $\text{tol}_{\text{error},f} = 2.5\%$. The greedy iteration is initialized with the 11-th order 50 points Lebedev quadrature rule. The configuration of $\sigma_s(x,y)$, the FOM and the ROM solutions are presented on the top row of Figure 6. ROM solution matches the FOM solution well. In the bottom left of Figure 6, the 94 sampled angular points are presented. In the bottom right, we present the relative training error at the final time and the values of error estimators as a function of the number of greedy iterations. Overall, the error estimator provides a reasonable approximation to the relative training error at the final time. The errors are shown in Table 5. It is clear that this example requires a higher rank representation for the reduced solution than the previous examples due to the large effective Knudsen number in the center region. The MMD-RBM produces numerical solutions with relative error below 0.8% for the scalar flux with only 0.27% degrees of freedom in comparison to the full model.

4.4. A lattice problem. The last example is a two-material lattice problem with $\varepsilon = 1$. The geometry set-up is shown in the middle of the top row of Figure 7. The black region is pure absorption with $\sigma_s = 0$ and $\sigma_a = 100$, while the rest is pure

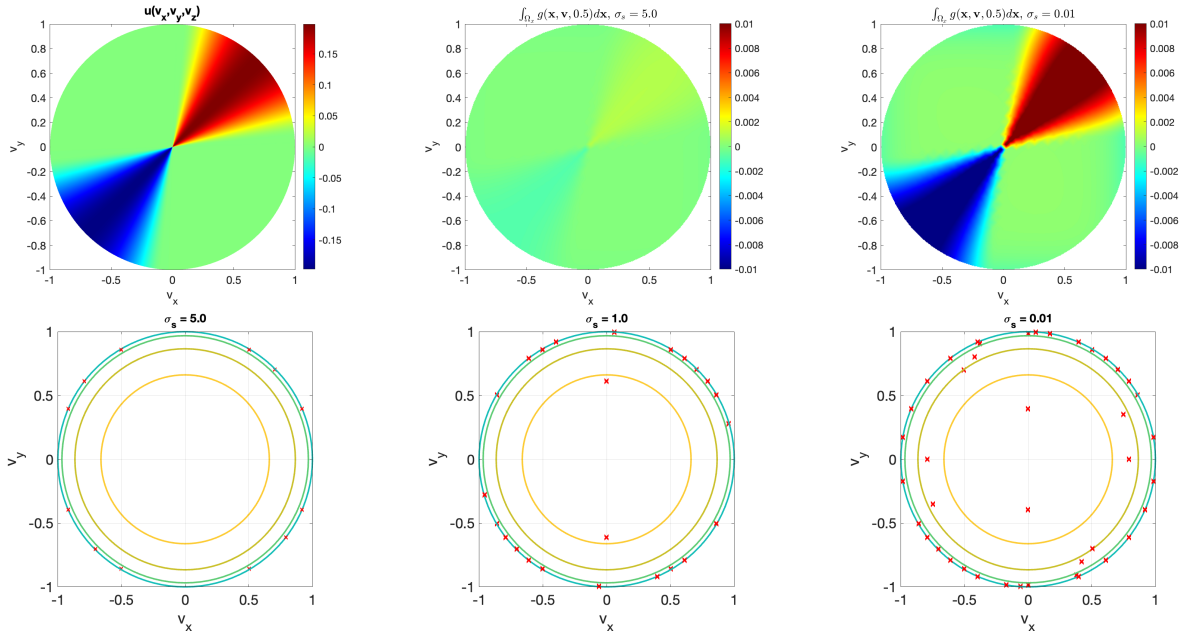

Fig. 5: Shown on top are the configuration $u(\boldsymbol{v})$ in the initial condition of $g(\mathbf{x},\boldsymbol{v},0) = u(\boldsymbol{v})\rho(\mathbf{x},0)$ and $\int_{\Omega_x} g(\mathbf{x},\boldsymbol{v},0.5)d\mathbf{x}$ for $\sigma_s = 5.0,\ 0.01$ (view from the north pole). On the bottom are the sampled angular directions (view from the north pole) for the example in Section 4.2 with various σ_s values.

r_ρ	r_g	N_v^{rq}	C-R	$\mathcal{E}_\rho$	$\mathcal{R}_\rho$	$\mathcal{E}_{\langle \boldsymbol{v} f\rangle}$	$\mathcal{R}_{\langle \boldsymbol{v} f\rangle}$	$\mathcal{E}_f$	$\mathcal{R}_f$
27	108	94	0.27%	3.00e-4	0.75%	8.32e-5	1.33%	1.18e-3	1.69%

Table 5: Dimensions of the reduced order subspaces, r_ρ, r_g, the number of reduced quadrature nodes N_v^{rq}, the testing error and the compression ratio for the multiscale example with the MMD-RBM.

scattering with $\sigma_s = 1$ and $\sigma_a = 0$. In the orange region, a constant source is imposed:

$$G(x,y) = \begin{cases} 1.0, & \text{if} \quad |x-2.5| < 0.5 \text{ and } \quad |y-2.5| < 0.5, \\ 0, & \text{otherwise.} \end{cases}$$

A uniform mesh of 100×100 rectangular elements is used to partition the computational domain. The final time is $T = 1.7$. The tolerances in the stopping criteria are $\text{tol}_{\text{ratio}}$ =1e-3, $\text{tol}_{\text{error},\rho} = 1.5\%$ and $\text{tol}_{\text{error},f} = 3.0\%$. When initializing the RBM offline, we use the 11-th order 50 point Lebedev quadrature rule.

r_ρ	r_g	N_v^{rq}	C-R	$\mathcal{E}_\rho$	$\mathcal{R}_\rho$	$\mathcal{E}_{\langle \boldsymbol{v} f\rangle}$	$\mathcal{R}_{\langle \boldsymbol{v} f\rangle}$	$\mathcal{E}_f$	$\mathcal{R}_f$
31	124	102	0.21%	1.85e-3	0.27%	4.45e-3	2.41%	2.38e-2	2.71%

Table 6: Dimensions of the reduced order subspaces, r_ρ, r_g, the number of reduced quadrature nodes N_v^{rq}, the testing error and the compression ratio for the lattice example with the MMD-RBM.

We present the ROM and FOM solutions on the top row of Figure 7. Shown on the bottom are the 102 nodes of the reduced quadrature rule and the history of the relative training error at the final time and the values of error estimators. Our error estimators approximate the relative errors at the final time well and the MMD-RBM solution matches the FOM well. The errors are displayed in Table 6. We see that the ROM achieves 0.27% relative error for ρ with 0.21% DOFs w.r.t FOM($\mathcal{V}_{\text{train}}$), while the relative errors $\langle \boldsymbol{v} f\rangle$ and f on the test set stayabout 2% to 3%.

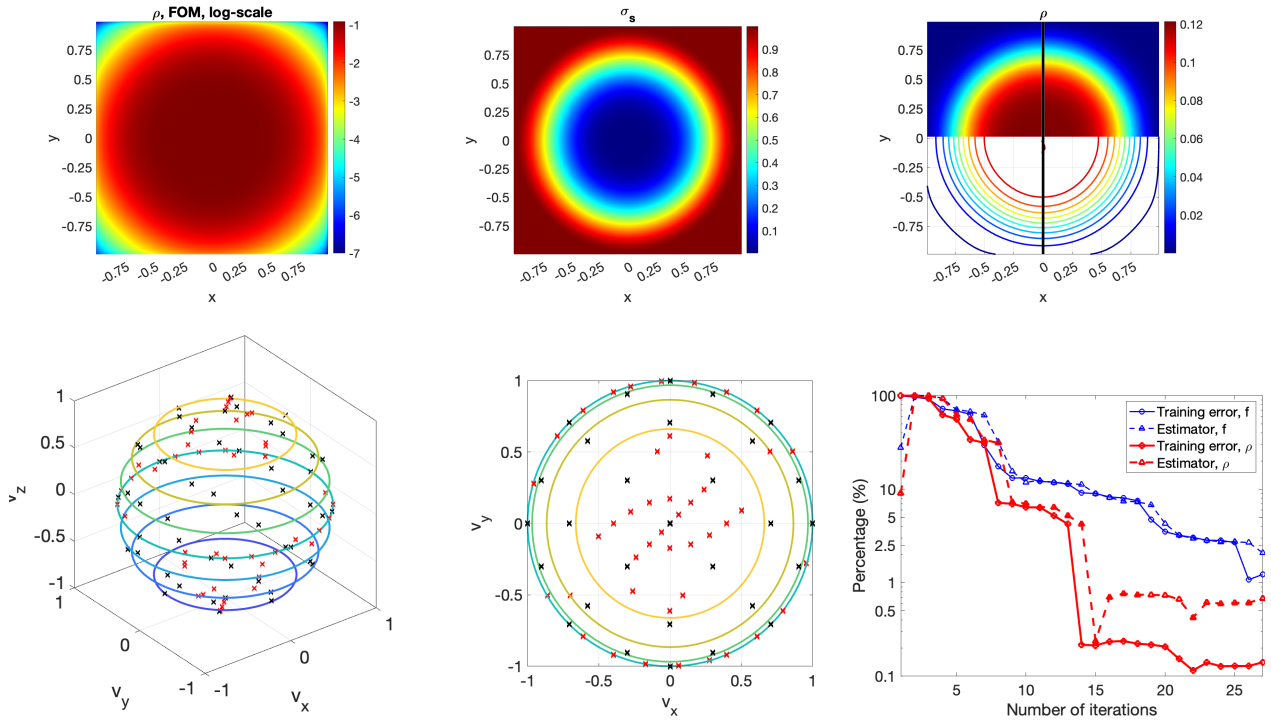

Fig. 6: Results for the multiscale example. Shown on top from left to right are the FOM solution in log scale, the function σ_s, and comparison between the ROM (right) and FOM (left) solutions. Shown on the bottom are reduced quadrature nodes on the unit sphere (Black for points in the initial reduced quadrature nodes, and Red for those sampled by the greedy algorithm), these nodes with a view from the north pole, and the history of the relative training error at the final time and the values of error estimators as a function of number of iterations.

5. Conclusion. In this paper, utilizing low rank structures with respect to the angular direction $\boldsymbol{v}$ and the temporal variable t, we developed a novel RBM to construct ROM for the time-dependent RTE based on the micro-macro decomposition. The proposed MMD-RBM is featured by an equilibrium-respecting strategy to construct reduced order subspaces and a reduced quadrature rule with non-negative weights preserving the stability of the underlying numerical solver. As demonstrated by our numerical tests, the Offline stage of the proposed method is more efficient than the vanilla POD method and sometimes even the standard full order solve, and the Online stage is able to efficiently predict angular fluxes for unseen angular directions and reconstruct the moments of the angular flux. The natural next step along this work is to use the proposed method as a building block to design ROMs for multi-query scenarios (e.g. inverse problems and uncertainty quantification) with essential physical parameters.

Our current ROM is a linear ROM utilizing the local equilibrium structure in the angular space. However when the scattering effect is not strong enough the system may be transport dominant and far from the local equilibrium. For a pure transport problem our method may need more careful choice of initial basis, initial quadrature rule and error indicator/estimator. Moreover it is also well known that linear ROM such as our method may be inefficient for transport problems [32]. One may design ROMs based on Lagrangian frameworks to conquer this issue and we leave it for the future investigation.

REFERENCES

[1] Anthony L Alberti and Todd S Palmer. Reduced-order modeling of nuclear reactor kinetics using proper generalized decomposition. Nuclear Science and Engineering, 194(10):837–858, 2020.

[2] Simon R Arridge and John C Schotland. Optical tomography: forward and inverse problems. Inverse problems, 25(12):123010, 2009.

[3] Patrick A Behne, Jean C Ragusa, and Jim E Morel. Model order reduction for Sn radiation transport. Nuclear Science and Engineering, 2021.

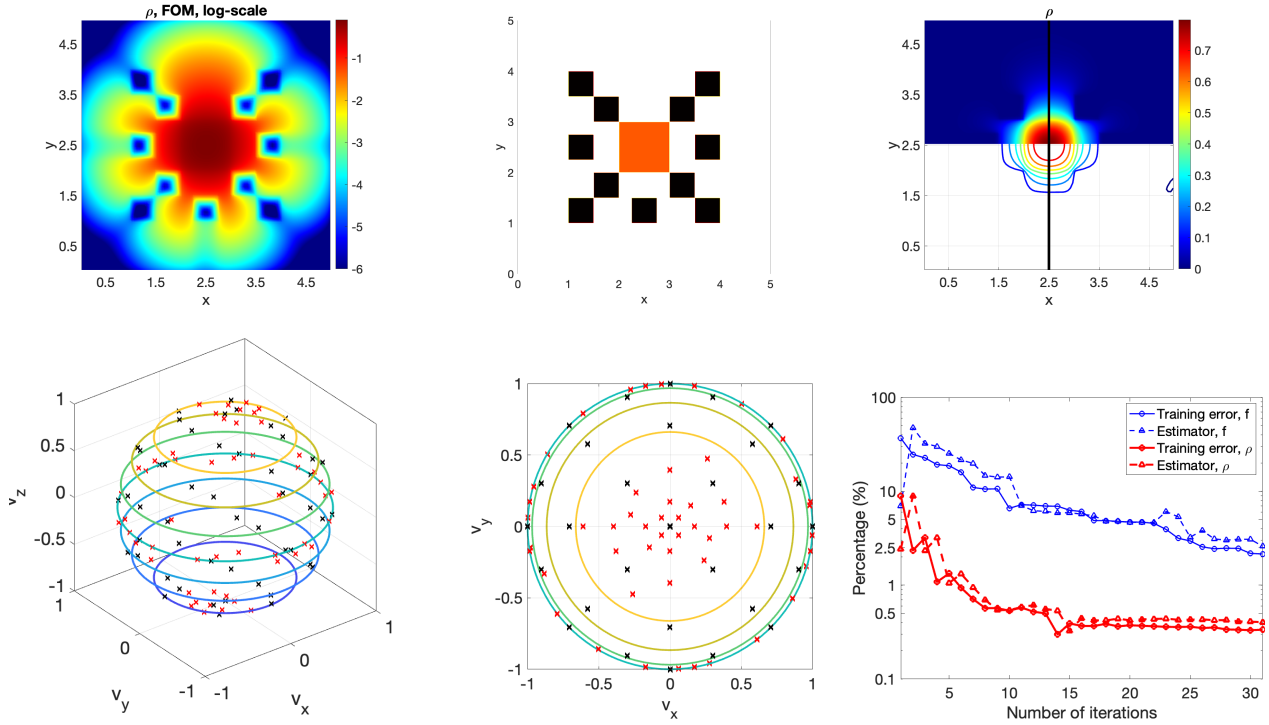

Fig. 7: Results for the lattice problem. Shown on top from left to right are FOM solution in the log scale, the domain setup (Black for pure absorption, White for pure scattering, Orange for a constant source and $\sigma_s = 1$, $\sigma_a = 0$), and comparison between the FOM (left) and ROM (right) solution. Shown on the bottom are reduced quadrature nodes on the unit sphere (Black for points in the initial reduced quadrature nodes, and Red for those sampled by the greedy algorithm), these nodes with a view from the north pole, and the history of the relative training error at the final time and the values of error estimators as a function of number of iterations.

[4] Matthew Brand. Incremental singular value decomposition of uncertain data with missing values. In European Conference on Computer Vision, pages 707–720. Springer, 2002.

[5] Andrew G Buchan, AA Calloo, Mark G Goffin, Steven Dargaville, Fangxin Fang, Christopher C Pain, and Ionel Michael Navon. A POD reduced order model for resolving angular direction in neutron/photon transport problems. Journal of Computational Physics, 296:138–157, 2015.

[6] Russel E Caflisch, Shi Jin, and Giovanni Russo. Uniformly accurate schemes for hyperbolic systems with relaxation. SIAM Journal on Numerical Analysis, 34(1):246–281, 1997.

[7] Paul Castillo, Bernardo Cockburn, Ilaria Perugia, and Dominik Schötzau. An a priori error analysis of the local discontinuous Galerkin method for elliptic problems. SIAM Journal on Numerical Analysis, 38(5):1676–1706, 2000.

[8] Yanlai Chen, Sigal Gottlieb, Lijie Ji, and Yvon Maday. An eim-degradation free reduced basis method via over collocation and residual hyper reduction-based error estimation. Journal of Computational Physics, 444:110545, 2021.

[9] Yanlai Chen, Lijie Ji, Akil Narayan, and Zhenli Xu. L1-based reduced over collocation and hyper reduction for steady state and time-dependent nonlinear equations. Journal of Scientific Computing, 87(1):1–21, 2021.

[10] Youngsoo Choi, Peter Brown, William Arrighi, Robert Anderson, and Kevin Huynh. Space–time reduced order model for large-scale linear dynamical systems with application to boltzmann transport problems. Journal of Computational Physics, 424:109845, 2021.

[11] Joseph Coale and Dmitriy Y Anistratov. A reduced-order model for thermal radiative transfer problems based on multilevel quasidiffusion method. In International Conference on Mathematics and Computational Methods Applied to Nuclear Science and Engineering, M and C, volume 2019, pages 278–287, 2019.

[12] Joseph Michael Coale. Reduced Order Models for Thermal Radiative Transfer Problems Based on Low-Order Transport Equations and the Proper Orthogonal Decomposition. 2019.

[13] Kurt A Dominesey and Wei Ji. Reduced-order modeling of neutron transport separated in space and angle via proper generalized decomposition. Nuclear Science and Engineering, 2022.

[14] Lukas Einkemmer, Jingwei Hu, and Yubo Wang. An asymptotic-preserving dynamical low-rank method for the multi-scale multi-dimensional linear transport equation. Journal of Computational Physics, 439:110353, 2021.
[15] Francis Filbet and Thomas Rey. A hierarchy of hybrid numerical methods for multiscale kinetic equations. SIAM Journal on Scientific Computing, 37(3):A1218–A1247, 2015.
[16] Bengt Fornberg and Jordan M Martel. On spherical harmonics based numerical quadrature over the surface of a sphere. Advances in Computational Mathematics, 40(5):1169–1184, 2014.
[17] Gene H Golub and Charles F Van Loan. Matrix computations. JHU press, 2013.
[18] Bernard Haasdonk. Reduced basis methods for parametrized PDEs–a tutorial introduction for stationary and instationary problems. Model reduction and approximation: theory and algorithms, 15:65, 2017.
[19] Ralf Hartmann and Paul Houston. Adaptive discontinuous Galerkin finite element methods for nonlinear hyperbolic conservation laws. SIAM Journal on Scientific Computing, 24(3):979–1004, 2003.
[20] Alexander C Hughes and Andrew G Buchan. An adaptive reduced order model for the angular discretization of the Boltzmann transport equation using independent basis sets over a partitioning of the space-angle domain. International Journal for Numerical Methods in Engineering, 2022.
[21] Juhi Jang, Fengyan Li, Jing-Mei Qiu, and Tao Xiong. High order asymptotic preserving DG-IMEX schemes for discrete-velocity kinetic equations in a diffusive scaling. Journal of Computational Physics, 281:199–224, 2015.
[22] Shi Jin. Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. Lecture notes for summer school on methods and models of kinetic theory (M&MKT), Porto Ercole (Grosseto, Italy), pages 177–216, 2010.
[23] Tino Laidin and Thomas Rey. Hybrid kinetic/fluid numerical method for the vlasov-poisson-bgk equation in the diffusive scaling. arXiv e-prints, pages arXiv–2303, 2023.
[24] Vyacheslav Ivanovich Lebedev. Quadratures on a sphere. USSR Computational Mathematics and Mathematical Physics, 16(2):10–24, 1976.
[25] Mohammed Lemou and Luc Mieussens. A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit. SIAM Journal on Scientific Computing, 31(1):334–368, 2008.
[26] Elmer Eugene Lewis and Warren F Miller. Computational methods of neutron transport. 1984.
[27] Tai-Ping Liu and Shih-Hsien Yu. Boltzmann equation: micro-macro decompositions and positivity of shock profiles. Communications in Mathematical Physics, 246(1):133–179, 2004.
[28] Yvon Maday and Einar M Rønquist. A reduced-basis element method. Journal of scientific computing, 17:447–459, 2002.
[29] Ryan G McClarren and Terry S Haut. Acceleration of source iteration using the dynamic mode decomposition. arXiv preprint arXiv:1812.05241, 2018.
[30] Ryan G McClarren and Terry S Haut. Data-driven acceleration of thermal radiation transfer calculations with the dynamic mode decomposition and a sequential singular value decomposition. Journal of Computational Physics, 448:110756, 2022.
[31] Giovanni Naldi and Lorenzo Pareschi. Numerical schemes for kinetic equations in diffusive regimes. Applied mathematics letters, 11(2):29–35, 1998.
[32] Mario Ohlberger and Stephan Rave. Reduced basis methods: Success, limitations and future challenges. arXiv preprint arXiv:1511.02021, 2015.
[33] Anthony T Patera, Gianluigi Rozza, et al. Reduced basis approximation and a posteriori error estimation for parametrized partial differential equations, 2007.
[34] Zhichao Peng, Yanlai Chen, Yingda Cheng, and Fengyan Li. A reduced basis method for radiative transfer equation. Journal of Scientific Computing, 91(1):1–27, 2022.
[35] Zhichao Peng, Yingda Cheng, Jing-Mei Qiu, and Fengyan Li. Stability-enhanced AP IMEX-LDG schemes for linear kinetic transport equations under a diffusive scaling. Journal of Computational Physics, 415:109485, 2020.
[36] Zhichao Peng, Yingda Cheng, Jing-Mei Qiu, and Fengyan Li. Stability-enhanced AP IMEX1-LDG method: energy-based stability and rigorous AP property. SIAM Journal on Numerical Analysis, 59(2):925–954, 2021.
[37] Zhichao Peng and Fengyan Li. Asymptotic preserving IMEX-DG-S schemes for linear kinetic transport equations based on Schur complement. SIAM Journal on Scientific Computing, 43(2):A1194–A1220, 2021.
[38] Zhuogang Peng and Ryan G McClarren. A high-order/low-order (HOLO) algorithm for preserving conservation in time-dependent low-rank transport calculations. Journal of Computational Physics, 447:110672, 2021.
[39] Zhuogang Peng, Ryan G McClarren, and Martin Frank. A low-rank method for two-dimensional time-dependent radiation transport calculations. Journal of Computational Physics, 421:109735, 2020.
[40] Gerald C. Pomraning. The equations of radiation hydrodynamics. International Series of Monographs in Natural Philosophy, Oxford: Pergamon Press, 1973.
[41] Zachary M Prince and Jean C Ragusa. Space-energy separated representations for multigroup neutron diffusion using proper generalized decompositions. Annals of Nuclear Energy, 142:107360, 2020.
[42] Gianluigi Rozza, Dinh Bao Phuong Huynh, and Anthony T Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. Archives of Computational Methods in Engineering, 15(3):229–275, 2008.
[43] RJD Spurr, TP Kurosu, and KV Chance. A linearized discrete ordinate radiative transfer model for atmospheric remote-sensing retrieval. Journal of Quantitative Spectroscopy and Radiative Transfer, 68(6):689–735, 2001.
[44] Mauricio Tano, Jean Ragusa, Dominic Caron, and Patrick Behne. Affine reduced-order model for radiation transport problems in cylindrical coordinates. Annals of Nuclear Energy, 158:108214, 2021.
[45] John Tencer, Kevin Carlberg, Roy Hogan, and Marvin Larsen. Reduced order modeling applied to the discrete ordinates

method for radiation heat transfer in participating media. In Heat Transfer Summer Conference, volume 50336, page V002T15A011. American Society of Mechanical Engineers, 2016.