A Reduced Basis Method for Radiative Transfer Equation

Zhichao Peng[1] · Yanlai Chen[2] · Yingda Cheng[3] · Fengyan Li[4]

Abstract

Linear kinetic transport equations play a critical role in optical tomography, radiative transfer and neutron transport. The fundamental difficulty hampering their efficient and accurate numerical resolution lies in the high dimensionality of the physical and velocity/angular variables and the fact that the problem is multiscale in nature. Leveraging the existence of a hidden low-rank structure hinted by the diffusive limit, in this work, we design and test the angular-space reduced order model for the linear radiative transfer equation, the first such effort based on the celebrated reduced basis method (RBM). Our method is built upon a high-fidelity solver employing the discrete ordinates method in the angular space, an asymptotic preserving upwind discontinuous Galerkin method for the physical space, and an efficient synthetic accelerated source iteration for the resulting linear system. Addressing the challenge of the parameter values (or angular directions) being coupled through an integration operator, the first novel ingredient of our method is an iterative procedure where the macroscopic density is constructed from the RBM snapshots, treated explicitly and allowing a transport sweep, and then updated afterwards. A greedy algorithm can then proceed to adaptively select the representative samples in the angular space and form a surrogate solution space. The second novelty is a least squares density reconstruction strategy, at each of the relevant physical locations, enabling the robust and accurate integration over an arbitrarily unstructured set of angular samples toward the macroscopic density. Numerical experiments indicate that our method is effective for computational cost reduction in a variety of regimes.

Keywords kinetic · reduced order model · reduced basis · multiscale · greedy algorithm · least squares · asymptotic preserving

✉ Zhichao Peng
　 pengzhic@msu.edu

Extended author information available on the last page of the article

　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　　🍦 Springer

1 Introduction

Linear kinetic transport equations model particles propagating through, and interacting with, background media. They provide prototype models for optical tomography [5], radiative transfer [35, 41] and neutron transport [31]. In this work, we consider the following steady-state linear radiative transfer equation

$$\mathbf{\Omega} \cdot \nabla f = \sigma_s \langle f \rangle - \sigma_t f + G, \quad \forall \mathbf{x} \in \mathbf{X}, \ \mathbf{\Omega} \in \mathbb{S}^{d-1}, \tag{1.1a}$$

whose solution $f = f(\mathbf{x}, \mathbf{\Omega})$ delineates particle distribution at location $\mathbf{x}$ in a physical domain $\mathbf{X} \subset \mathbb{R}^d$, and $\mathbf{\Omega} \in \mathbb{S}^{d-1}$ denotes the angular variable. We enforce a Dirichlet boundary condition on the inflow boundary

$$f(\mathbf{x}, \mathbf{\Omega}) = f_{\text{inflow}}(\mathbf{x}, \mathbf{\Omega}), \ \mathbf{x} \in \partial\mathbf{X} \ \text{and} \ \mathbf{\Omega} \cdot \mathbf{n}(\mathbf{x}) < 0. \tag{1.1b}$$

Here $\mathbf{n}(\mathbf{x})$ stands for the unit outward normal on $\partial\mathbf{X}$. In addition, $\sigma_s(\mathbf{x}) \geq 0$ is the scattering cross section, $\sigma_t(\mathbf{x}) = \sigma_s(\mathbf{x}) + \sigma_a(\mathbf{x})$ is the total cross section, with $\sigma_a(\mathbf{x}) \geq 0$ being the absorption cross section, and $G(\mathbf{x})$ is the source. $\langle \cdot \rangle$ encodes a normalized integration in the angular space, namely,

$$\langle f \rangle = \frac{1}{|\mathbb{S}^{d-1}|} \int_{\mathbb{S}^{d-1}} f \, d\mathbf{\Omega}. \tag{1.2}$$

This gives the macroscopic density, $\rho(\mathbf{x}) = \langle f \rangle$, defined on $\mathbf{X}$. A fundamental difficulty in numerically resolving (1.1) originates from the high dimension $(2d - 1)$ of $(\mathbf{x}, \mathbf{\Omega})$ space.

Given the non-dimensional mean free path length $\varepsilon > 0$, one can further define the rescaled cross sections as $\sigma_s = \frac{\widehat{\sigma}_s}{\varepsilon}$, $\sigma_a = \varepsilon\widehat{\sigma}_a$, and $G = \varepsilon\widehat{G}$. As proved in [6], when $\varepsilon \to 0$, $f(\mathbf{x}, \mathbf{\Omega}) \to \rho(\mathbf{x})$, and $\rho(\mathbf{x})$ solves the diffusion equation

$$\nabla \cdot \left(\widehat{\sigma}_s^{-1} D \nabla \rho \right) = \widehat{\sigma}_a \rho + \widehat{G}. \tag{1.3}$$

Here $D = \text{diag}\left(\langle \Omega_1^2 \rangle, \ldots, \langle \Omega_d^2 \rangle \right)$, with Ω_i being the i-th component of $\mathbf{\Omega} \in \mathbb{S}^{d-1}$. As a result, when the problem is in its diffusive regime, the solution is close to a rank-1 manifold in the angular space (i.e. the diffusion limit). Hence, in or near such regime, it is possible to design a reduced order model (ROM) to capture the low-rank structure of the solution, and this provides an opportunity to mitigate the curse of dimensionality. On the other hand, when ε varies and possibly spans through several magnitude in the physical domain, the problem is multiscale in its nature. This poses challenges not only for traditional numerical schemes but also for any attempt to design an effective ROM. It is well known that a standard numerical method may fail to capture the diffusion limit and lead to unphysical solutions when spatial meshes are under-resolved with respect to ε, i.e. when the mesh size h satisfies $h \gg \varepsilon$ [29, 36]. This issue should be addressed before designing any ROM. One solution is to apply the asymptotic preserving (AP) schemes [27], which preserve the asymptotic limit on the discrete level and work well for both the kinetic regime and the diffusive regime. In other words, AP methods are able to capture the diffusion limit even on under-resolved meshes. In this paper, we apply the discrete ordinates (S_N) method [40] in the angular space and the upwind discontinuous Galerkin (DG) method in physical space. Upwind DG method was investigated for (1.1) in [1, 28, 29], and it is proved to be AP [23] when the approximation space contains continuous functions that are at least linear on each mesh element. To solve the resulting linear system, we apply iterative solvers based on source iteration [31]. Standard source iteration converges slowly when the problem is scattering dominant [2], we here apply the synthetic accelerated source iteration (SASI) [2] to achieve good efficiency for various regimes.

As discussed above, the existence of a low-rank structure of the solution manifold in the diffusive regime gives hope to the success of ROM techniques for kinetic transport equations (1.1), the focus of our paper. This is by no means the first such attempt. In fact, ROM has become an increasingly popular technique in kinetic simulations in the last few years. The proper orthogonal decomposition (POD) type techniques were applied to (1.1) and its time transient counterpart in [7, 9, 16]. Other related works include space-time POD [15], AP random singular value decomposition (RSVD) [10], dynamic mode decomposition (DMD) [34], proper generalized decomposition (PGD) [4, 18, 19, 42], and dynamical low rank approximations (DLRA) [17, 20, 38]. These existing approaches either lack the hallmark greedy algorithm or is not a method of snapshots. A POD method with the greedy algorithm to adaptively select angular samples is proposed in [46], but [46] only considers problems without the scattering effect. In this work, taking the scattering effect into account, we design and test a model reduction technique for the linear radiative transfer equation (1.1) that features both ingredients, namely a celebrated greedy algorithm adaptively selecting the representative samples in the angular space and a resulting surrogate solution space spanned by the corresponding snapshots. Indeed, it is under the framework of the reduced basis (RB) method [24, 26, 43, 44] and to the best of our knowledge, the very first such attempt, which takes the scattering effect into account.

RB method (RBM) has become the go-to option for efficiently simulating parametric partial differential equation (PDE). Its hallmark feature is a greedy algorithm embedded in an offline-online decomposition procedure. The offline (i.e. training) stage is devoted to a judicious exploration of the parameter-induced solution manifold. It adaptively selects a number of representative parameter values via a mathematically rigorous greedy algorithm [8]. Solution snapshots for these parameter values are then obtained through a user-specified (potentially expensive) full order (i.e. accurate) solver. The buildup of the surrogate solution space spanned by these snapshots is done step-by-step. Each iteration of the greedy algorithm adds the parameter value which is the maximizer of a mathematically rigorous *a posteriori* error estimator or an effective error indicator were the current surrogate space used as a reduced solver space via e.g. a Galerkin or Petrov-Galerkin projection. For parametric systems bearing a small Kolmogorov N-width [39], the dimension of the surrogate space is orders of magnitude smaller than the total degrees of freedom for the full model in order to reach a high degree of accuracy. This difference in size leads to a dramatic decrease in computation time for the online simulations when a *reduced* solution is sought in the terminal surrogate space for each parameter value. Moreover, unlike other ROM techniques (e.g. POD-based approaches), the number of full order inquiries RBM takes offline is minimum i.e. equal to the dimension of the surrogate space.

Leveraging the low-rank structure induced by the angular space, it is natural to treat the angular variable $\boldsymbol{\Omega}$ as our parameter. What prevents a direct application of RBM is the integral operator (1.2) which poses a two-fold extra challenge. First, unlike the standard setting when RBM applies, the solutions for different parameter values $\{f(\cdot, \boldsymbol{\Omega}) : \boldsymbol{\Omega} \in \mathbb{S}^{d-1}\}$ are all coupled through (1.2). A milder version of this type of coupling is addressed in the stochastic PDE setting [32]. Second, from a practical viewpoint, the robustness and the efficiency of the SASI iterative linear solver is highly sensitive to the quality of the density approximation $\rho = \langle f \rangle$. In particular, the unstructured nature of the selected RB parameter samples in the angular variable prevents a robust and accurate numerical integration which usually requires a structured set of quadrature points.

To address these challenges, we effectively decouple the solutions for different angular samples by designing a greedy iterative procedure where an approximation of the macro-scopic density ρ is constructed from the RB snapshots at the beginning of each iteration. It is

frozen during the iteration and reconstructed before the next. To resolve the lack of structure for a robust numerical integration, we develop a least squares density reconstruction strategy capable of integrating over an arbitrary set of selected angular samples robustly. Our numerical experiments show that the proposed method is effective for the radiative transfer problems in the scattering dominant and intermediate regimes, as well as for the multiscale problems with large scattering dominant subregions. The proposed RB method (including both the offline and online stages) can be seen as a surrogate model for the full order upwind DG solver. The online stage of the proposed method can be further applied to predict solutions at "unseen" angular samples. The proposed method can also be utilized as a building block to construct ROMs for problems with essential physical parameters such as the magnitude of the scattering cross section whose "multi-query" nature will lead to more pronounced saving for our approach.

The rest of the paper is organized as follows. We introduce the full order numerical scheme and the SASI iterative solver in Sect. 2. Section 3 is devoted to the least squares density reconstruction and the RB algorithm. We demonstrate the performance of the proposed method through a series of numerical experiments in Sect. 4. Finally, conclusions are made in Sect. 5.

2 Full Order Numerical Method

In this section, we describe our full order numerical scheme for solving (1.1) focusing on the 1D slab geometry and the 2D case with $\mathbf{\Omega} \in \mathbb{S}^1$. The equation (1.1) on the 1D slab geometry is given as

$$v\partial_x f = \frac{\sigma_s}{2} \int_{-1}^{1} f\,dv - \sigma_t f + G, \quad v \in [-1, 1], \quad x \in \mathbf{X}, \tag{2.1}$$

where we write Ω as v here following convention. The 2D equation with $\mathbf{\Omega} \in \mathbb{S}^1$ can be written as

$$\cos(\theta)\partial_x f + \sin(\theta)\partial_y f = \sigma_s \langle f \rangle - \sigma_t f + G, \quad \theta \in [0, 2\pi], \quad \mathbf{x} \in \mathbf{X}. \tag{2.2}$$

In the next three subsections, we detail the discrete ordinates method [40] for the angular discretization, the upwind DG method for the spatial discretization, and finally the SASI iterative solver.

2.1 Angular Discretization

In the angular space, we apply the discrete ordinates method [40] by sampling f at quadrature points $\{\mathbf{\Omega}_j\}_{j=1}^{N_{\mathbf{\Omega}}}$. Letting $\{\omega_j\}_{j=1}^{N_{\mathbf{\Omega}}}$ be the corresponding normalized quadrature weights, we can discretize the integral operator as

$$\langle f \rangle \approx \langle f \rangle_h = \sum_{j=1}^{N_{\mathbf{\Omega}}} \omega_j f(\cdot, \mathbf{\Omega}_j), \quad \sum_{j=1}^{N_{\mathbf{\Omega}}} \omega_j = 1,$$

and, as a consequence, Eq. (1.1) can be discretized as

$$\mathbf{\Omega}_j \cdot \nabla f_j = \sigma_s \sum_{i=1}^{N_{\mathbf{\Omega}}} \omega_i f_i - \sigma_t f_j + G, \quad j = 1, \ldots, N_{\mathbf{\Omega}}, \tag{2.3}$$

where $f_j(\mathbf{x}) \approx f(\mathbf{x}, \mathbf{\Omega}_j)$. This method is also referred to as the S_N method, if one use $N_\Omega = N$ quadrature points for the 1D slab geometry or $N_\mathbf{\Omega} = 2N$ quadrature points for the 2D case. Particularly, we use the following quadrature rules.

1D: $\{\mathbf{\Omega}_j\}_{j=1}^N$ is the collection of N-point Gauss-Legendre quadrature points on $[-1, 1]$.

2D: $\mathbf{\Omega}_j = \left(\cos(\theta_j), \sin(\theta_j)\right)^T$, with $\theta_j = \frac{(2j-1)\pi}{2N}$, $\omega_j = \frac{1}{2N}$, $j = 1, \ldots, N_\mathbf{\Omega}$.

2.2 Spatial Discretization

It is well known that the asymptotic preserving (AP) schemes [27, 36] can capture the correct diffusion limit without a highly refined mesh resolving the small ε-scale. We adopt the upwind DG method [23, 28, 29], which has been proven to be AP if the approximation space contains continuous functions that are at least linear on each mesh element [1, 23]. Without loss of generality, we assume $\mathbf{X} = [x_L, x_R]$ in 1D and $\mathbf{X} = [x_L, x_R] \times [y_L, y_R]$ in 2D. Let $\mathbf{X}_h = \{T_i, i = 1, \ldots, N_x\}$ be a partition of $\mathbf{X}$, with each T_i being an interval in 1D or a rectangle in 2D. We introduce a discrete space

$$U_h^K = \{u(\mathbf{x}) : u(\mathbf{x})|_{T_i} \in Q^K(T_i), i = 1, \ldots, N_x\}, \quad K \geq 1, \tag{2.4}$$

where $Q^K(T_i)$ denotes the d-variate polynomial on T_i with degree up to K for each variable. We are now ready to state the upwind DG spatial discretization: we seek $\{f_h(\mathbf{x}, \mathbf{\Omega}_j) \in U_h^K : j = 1, \ldots, N_\mathbf{\Omega}\}$, such that

$$-\int_{T_i} \left(\mathbf{\Omega}_j \cdot \nabla\phi_h(\mathbf{x})\right) f_h(\mathbf{x}, \mathbf{\Omega}_j) d\mathbf{x} + \int_{\partial T_i} \widehat{\mathbf{H}}(\mathbf{\Omega}_j, f_h, \mathbf{n}_i)\phi_h(\mathbf{x})ds + \int_{T_i} \sigma_t f_h(\mathbf{x}, \mathbf{\Omega}_j)\phi_h(\mathbf{x})d\mathbf{x}$$

$$= \sum_{k=1}^{N_\mathbf{\Omega}} \omega_k \int_{T_i} \sigma_s f_h(\mathbf{x}, \mathbf{\Omega}_k)\phi_h(\mathbf{x})d\mathbf{x} + \int_{T_i} G(\mathbf{x})\phi_h(\mathbf{x})d\mathbf{x}, \quad \forall \phi_h \in U_h^K. \tag{2.5}$$

Here $\widehat{\mathbf{H}}(\mathbf{\Omega}_j, f_h, \mathbf{n}_i)$ is the upwind numerical flux along ∂T_i, that is defined, for an element $T_i = T^-$ with the neighboring element T^+, as

$$\widehat{\mathbf{H}}(\mathbf{\Omega}_j, f_h, \mathbf{n}_i) = \frac{\mathbf{\Omega}_j \cdot \mathbf{n}_i}{2}\left(f_h^+(\mathbf{x}, \mathbf{\Omega}_j) + f_h^-(\mathbf{x}, \mathbf{\Omega}_j)\right)$$

$$+ \frac{|\mathbf{\Omega}_j \cdot \mathbf{n}_i|}{2}\left(f_h^-(\mathbf{x}, \mathbf{\Omega}_j) - f_h^+(\mathbf{x}, \mathbf{\Omega}_j)\right). \tag{2.6}$$

We use $f_h^\pm$ to denote the restriction of f_h to $T^\pm$, while $\mathbf{n}_i$ is the unit outward normal on ∂T_i. Based on f_h, the density ρ is further approximated by

$$\rho_h(\mathbf{x}) = \langle f_h \rangle_h = \sum_{j=1}^{N_\mathbf{\Omega}} \omega_j f_h(\mathbf{x}, \mathbf{\Omega}_j). \tag{2.7}$$

Next we will rewrite the DG scheme (2.5) into its matrix-vector form. To this end, we assume that $\{\phi_k\}_{k=1}^{N_{\text{dof}}}$ is a basis for U_h^K, and $f_h(\mathbf{x}, \mathbf{\Omega}_j)$, $\rho_h(\mathbf{x})$ are then expanded as

$$f_h(\cdot, \mathbf{\Omega}_j) = \sum_{k=1}^{N_{\text{dof}}} \alpha_k^f(\mathbf{\Omega}_j)\phi_k, \quad \rho_h = \sum_{k=1}^{N_{\text{dof}}} \alpha_k^\rho \phi_k, \text{ with } \alpha_k^\rho = \sum_{j=1}^{N_\mathbf{\Omega}} \omega_j \alpha_k^f(\mathbf{\Omega}_j).$$

The last equality is due to the numerical quadrature (2.7). We further define

$$\mathbf{f}_{\mathbf{\Omega}_j} = \left(\alpha_1^f(\mathbf{\Omega}_j), \ldots, \alpha_{N_{\mathrm{dof}}}^f(\mathbf{\Omega}_j) \right)^T \in \mathbb{R}^{N_{\mathrm{dof}}}, \quad \mathbf{F} = \left(\mathbf{f}_{\mathbf{\Omega}_1}, \ldots, \mathbf{f}_{\mathbf{\Omega}_{N_\Omega}} \right) \in \mathbb{R}^{N_{\mathrm{dof}} \times N_\Omega},$$

$$\boldsymbol{\rho} = \left(\alpha_1^\rho, \ldots, \alpha_{N_{\mathrm{dof}}}^\rho \right)^T \in \mathbb{R}^{N_{\mathrm{dof}}},$$

and $\mathbf{U}_{\mathbf{\Omega}_j}, j = 1, \ldots, N_\Omega, \boldsymbol{\Sigma}_t, \boldsymbol{\Sigma}_s \in \mathbb{R}^{N_{\mathrm{dof}} \times N_{\mathrm{dof}}}$ and $\mathbf{g} \in \mathbb{R}^{N_{\mathrm{dof}}}$ as

$$(\mathbf{U}_{\mathbf{\Omega}_j})_{kl} = \sum_{i=1}^{N_x} \left(-\int_{T_i} (\mathbf{\Omega}_j \cdot \nabla \phi_k(\mathbf{x})) \phi_l(\mathbf{x}) d\mathbf{x} + \int_{\partial T_i} \widehat{\mathbf{H}} \left(\mathbf{\Omega}_j, \phi_l, \mathbf{n}_i \right) \phi_k(\mathbf{x}) ds \right),$$

$$(\boldsymbol{\Sigma}_t)_{kl} = \sum_{i=1}^{N_x} \left(\int_{T_i} \sigma_t \phi_k(\mathbf{x}) \phi_l(\mathbf{x}) d\mathbf{x} \right), \quad (\boldsymbol{\Sigma}_s)_{kl} = \sum_{i=1}^{N_x} \left(\int_{T_i} \sigma_s \phi_k(\mathbf{x}) \phi_l(\mathbf{x}) d\mathbf{x} \right),$$

$$(\mathbf{g})_k = \sum_{i=1}^{N_x} \left(\int_{T_i} G(\mathbf{x}) \phi_k(\mathbf{x}) d\mathbf{x} \right).$$

We here adopt the commonly used basis functions $\{\phi_k\}_{k=1}^{N_{\mathrm{dof}}}$, with each being nonzero only on one mesh element as a scaled Legendre polynomial or its tensor version. With such a choice, $\boldsymbol{\Sigma}_t$ and $\boldsymbol{\Sigma}_s$ are block-diagonal, symmetric and semi-positive definite. If mesh elements in space are suitably reordered, each $\mathbf{U}_{\mathbf{\Omega}_j}$ can be block lower triangular. With the notation above, the DG scheme (2.5) can be rewritten into its matrix-vector form:

$$\mathbf{U}_{\mathbf{\Omega}_j} \mathbf{f}_{\mathbf{\Omega}_j} + \boldsymbol{\Sigma}_t \mathbf{f}_{\mathbf{\Omega}_j} - \boldsymbol{\Sigma}_s \boldsymbol{\rho} = \mathbf{g}, \quad \forall j = 1, \ldots, N_\Omega. \tag{2.8}$$

We end this subsection by noting that the AP upwind DG scheme (2.5) also exists for unstructured meshes and general geometries.

2.3 Synthetic Accelerated Source Iteration

Due to the high (i.e. $2d - 1$) dimensional nature of the problem, iterative methods must be adopted when solving (2.8). However, when the problem is scattering dominant, iterative solvers such as the standard source iterations may converge slowly [2]. To efficiently solve (2.8), we apply the synthetic accelerated source iteration (SASI) scheme [2]. Each iteration of a typical SASI scheme consists of two main steps. The first step is a transport sweep based on the known $\boldsymbol{\rho}^k$ from the previous iteration. More specifically, using the given data $\boldsymbol{\rho}^k$ and $\mathbf{g}$, we invert $\mathbf{U}_{\mathbf{\Omega}_j} + \boldsymbol{\Sigma}_t$ in (2.8) and obtain $\mathbf{f}_{\mathbf{\Omega}_j}^{k+1}$ for each $\mathbf{\Omega}_j$. After that, we numerically integrate $f_{\mathbf{\Omega}_j}^{k+1}$ in the angular space to obtain an initial update of the density $\boldsymbol{\rho}^{k,*}$. The second step is to compute a correction, $\boldsymbol{\rho}^{k,c}$, for the density by a computationally less expensive procedure. One can then update $\boldsymbol{\rho}^{k+1} = \boldsymbol{\rho}^{k,*} + \boldsymbol{\rho}^{k,c}$ and proceed to the next iteration.

In this work, we mainly focus on the S_2 synthetic acceleration (S2SA) method following [2, 33]. To elaborate the detail of the second step, we assume f^{k+1} is the solution to

$$\mathbf{\Omega} \cdot \nabla f^{k+1} + \sigma_t f^{k+1} = \sigma_s \rho^k + G, \tag{2.9}$$

and define $\rho^{k,*} = \langle f^{k+1} \rangle$. Let $\delta f^k = f - f^k$. By subtracting (2.9) from (1.1), we obtain the equation for the correction δf^{k+1}, namely

$$\mathbf{\Omega} \cdot \nabla(\delta f^{k+1}) + \sigma_t \delta f^{k+1} - \sigma_s \langle \delta f^{k+1} \rangle = \sigma_s(\rho^{k,*} - \rho^k). \tag{2.10}$$

The correction of the density can then be calculated as $\rho^{k,c} = \langle \delta f^{k+1} \rangle$. The main idea of the S2SA is to apply the S_2 approximation in the angular space when solving the correction equation (2.10). This means that we only work with N_d quadrature points in the angular space, with $N_d = 2$ in 1D and $N_d = 4$ in 2D. As a result, the direct solver in the correction step can be implemented very efficiently. With the DG spatial discretization, the discretized linear system for the correction step is

$$\mathbf{U}^d_{\mathbf{\Omega}_j}(\delta \mathbf{f}_{\mathbf{\Omega}_j}) + \mathbf{\Sigma}_t(\delta \mathbf{f}_{\mathbf{\Omega}_j}) - \mathbf{\Sigma}_s \langle \delta \rho \rangle_h = \mathbf{\Sigma}_s(\rho^{k,*} - \rho^k), \ \forall j = 1, \ldots, N_d, \qquad (2.11)$$

where $\mathbf{U}^d_{\mathbf{\Omega}_j}$ is the "upwind" matrix corresponding to the angular samples of the S_2 approximation. Details of the algorithm are presented in Algorithm 1.

Remark 2.1 An alternative synthetic acceleration strategy is the diffusion synthetic acceleration (DSA) [2, 3, 47], which approximates the correction equation (2.10) through a diffusion model. It is well known that a so-called "consistent" discretization must be applied to the diffusion model, otherwise the source iteration with the DSA may converge slowly or even diverge [2, 3, 47]. With the S2SA, one can reuse the kinetic solver with fewer angular samples, but it does have more degrees of freedom compared with the DSA. In our reduced order algorithm, we observe that the S2SA is more robust for different regimes and slightly more accurate than the DSA, though the full order solvers with the DSA and S2SA are comparable with respect to the robustness and accuracy. More details of the DSA and the comparison between the RB method with both acceleration strategies can be found in Appendix A.

Algorithm 1: Synthetic accelerated source iteration to solve (2.8)

> Given initial guess: $\boldsymbol{\rho}^0$
> **Transport sweep**: solve $(\mathbf{U}_{\mathbf{\Omega}_j} + \mathbf{\Sigma}_t)\mathbf{f}^1_{\mathbf{\Omega}_j} = \mathbf{g} + \mathbf{\Sigma}_s \boldsymbol{\rho}^0$, $j = 1, \ldots, N_{\mathbf{\Omega}}$.
> Reconstruct $\boldsymbol{\rho}^{0,*}$ through numerical integration based on $\mathbf{f}^1_{\mathbf{\Omega}_j}$, $j = 1, \ldots, N_{\mathbf{\Omega}}$.
> **Correction**: obtain the correction $\boldsymbol{\rho}^{0,c}$. Let $\boldsymbol{\rho}^1 = \boldsymbol{\rho}^{0,*} + \boldsymbol{\rho}^{0,c}$. Set $k = 1$.
> **while** $||\rho^k - \rho^{k-1}||_\infty > \text{error}_{\text{tol}}$ and $k \leq \text{iter}_{\text{tol}}$ **do**
> > **Transport sweep**: solve $(\mathbf{U}_{\mathbf{\Omega}_j} + \mathbf{\Sigma}_t)\mathbf{f}^{k+1}_{\mathbf{\Omega}_j} = \mathbf{g} + \mathbf{\Sigma}_s \boldsymbol{\rho}^k$, $j = 1, \ldots, N_{\mathbf{\Omega}}$
> > Reconstruct $\boldsymbol{\rho}^{k,*}$ through numerical integration based on $\mathbf{f}^{k+1}_{\mathbf{\Omega}_j}$, $j = 1, \ldots, N_{\mathbf{\Omega}}$.
> > **Correction**: obtain the correction $\boldsymbol{\rho}^{k,c}$ by solving (2.11). Let $\boldsymbol{\rho}^{k+1} = \boldsymbol{\rho}^{k,*} + \boldsymbol{\rho}^{k,c}$.
> > Set $k := k + 1$.
> **end while**

---

3 Reduced Basis Method in the Angular Space

In this section, we briefly review the basics of RBM and describe the main challenges for designing a RBM for (1.1) in Sect. 3.1. We then present all elements of our algorithm in Sect. 3.2.

3.1 Background

RBM is a popular approach for obtaining reduced order models for a parametric differential equation

$$r(\eta; \mu) = \mathcal{L}(\eta; \mu) - g_\mu = 0, \ \mu \in \mathcal{X}_{\mathcal{P}}, \ \eta(\cdot; \mu) \in \Gamma_\mu, \qquad (3.1)$$

where $\mu \in \mathcal{X}_\mathcal{P}$ stands for a parameter, $\mathcal{L}(\cdot ; \mu)$ encodes a steady-state or time-dependent parametric differential operator. $\eta(\cdot ; \mu) \in \Gamma_\mu$ denotes the solution corresponding to μ and is often referred to as a snapshot. The method assumes a full order (potentially expensive) solver of high accuracy for (3.1) which, for simplicity, we write in a strong form

$$r_h(\eta_h ; \mu) = 0. \tag{3.2}$$

Its hallmark feature is a greedy algorithm embedded in an offline-online decomposition procedure. Given a (sufficiently fine) training set $\mathcal{T} \subset \mathcal{X}_\mathcal{P}$. The judicious exploration offline aims to build a low-dimensional surrogate for the $\mathcal{T}$-induced solution manifold $\{\eta_h(\mu) : \mu \in \mathcal{T}\}$. This surrogate space is iteratively constructed via a hierarchical series of reduced basis $\{\eta_j\}_{j=1}^r$ by a greedy algorithm. At each iteration the snapshot corresponding to the most under-represented parameter value (were the current reduced space to be adopted), as identified by an error indicator or *a posteriori* error estimator, is added to the current set of bases. These snapshots are obtained through the full order solver (3.2). A defining feature of RBM is that the number of full model solves is minimum, i.e. the same as the surrogate space dimension. For parametric systems bearing a small Kolmogorov N-width [39], the dimension of the surrogate space is orders of magnitude smaller than the total degrees of freedom for the full model in order to reach a high degree of accuracy. This difference in size leads to a dramatic decrease in computation time for the online simulations when a *reduced* solution is sought in the terminal surrogate space for each parameter value as the Galerkin or Petrov-Galerkin projection into the reduced space $\Gamma^r = \mathrm{span}\{\eta_j, j = 1, \dots, r\}$ constructed offline.

Leveraging the low-rank structure induced by the angular space for our problem, it is natural to treat the angular variable $\boldsymbol{\Omega}$ as our parameter. What prevents a direct application of RBM is the integral operator (1.2) which poses a two-fold extra challenge. First, unlike the standard setting when RBM applies, the solutions for different parameter values $\{f(\cdot, \boldsymbol{\Omega}) : \boldsymbol{\Omega} \in \mathbb{S}^{d-1}\}$ are all coupled through (1.2) or its discrete counterpart $\langle f \rangle \approx \sum_{j=1}^{N_{\boldsymbol{\Omega}}} \omega_j f(\cdot, \boldsymbol{\Omega}_j)$. Second, the robustness and the efficiency of the SASI iterative solver relies on the high quality of the density approximation $\rho = \langle f \rangle$. In particular, the unstructured nature of the selected RB parameter samples in the angular variable prevents a robust and accurate numerical integration which usually requires a structured set of quadrature points.

3.2 The Proposed Algorithm

We propose to effectively decouple the solutions for different angular samples by designing an iterative procedure where an approximation of the macroscopic density ρ is constructed from the RB snapshots and gradually refined as the RB space is built and RB solutions get more accurate. This iterative procedure manifests the first novel ingredient of our method in that a quantity that is indirectly dependent on the parameter (i.e. the macroscopic density defined as an integral over the parameter domain) is fixed during one greedy iteration and updated only at the end of such iteration. It allows the greedy algorithm to proceed efficiently which in turn enriches the surrogate solution space rendering the RBM solutions and the dependent macroscopic quantity more accurate. We emphasize that once the greedy algorithm converges, we consider the macroscopic density well-resolved and will adopt its terminal value online for any new parameter value. This is reasonable since the macroscopic density is not directly dependent on a particular parameter value. To resolve the lack of structure for a robust numerical integration, we develop a least squares density reconstruction strategy capable of integrating over an arbitrary set of selected angular samples. In the remaining

part of this subsection, we first describe our online solver with any given RB space which is repeatedly called offline to construct the terminal RB space from scratch. Next, the least squares reconstruction algorithm is presented and an L^1-based residual-free error indicator is reviewed. Finally, we finish by detailing our algorithm and making a few relevant remarks.

Online Stage

Assuming that the (discrete) reduced basis space corresponds to the column space of $\mathcal{U}_{RB}$ and the currently reconstructed density is $\boldsymbol{\rho}_{RB}$, the online solver amounts to seeking a Galerkin projection of $\mathbf{f}_{\boldsymbol{\Omega}}$ into $\mathcal{U}_{RB}$ that satisfies the weak formulation for any given $\boldsymbol{\Omega}$. That is, we assume

$$\mathbf{f}_{\boldsymbol{\Omega}} \approx \mathcal{U}_{RB}\mathbf{c}_{RB}(\boldsymbol{\Omega}), \tag{3.3}$$

and compute $\mathbf{c}_{RB}(\boldsymbol{\Omega})$ by solving the reduced formulation of (2.8), namely

$$(\mathcal{U}_{RB})^T \mathbf{U}_{\boldsymbol{\Omega}_i} \mathcal{U}_{RB}\mathbf{c}_{RB}(\boldsymbol{\Omega}) + (\mathcal{U}_{RB})^T \boldsymbol{\Sigma}_t \mathcal{U}_{RB}\mathbf{c}_{RB}(\boldsymbol{\Omega}) = (\mathcal{U}_{RB})^T \boldsymbol{\Sigma}_s \boldsymbol{\rho}_{RB} + (\mathcal{U}_{RB})^T \mathbf{g}. \tag{3.4}$$

Remark 3.1 Note that the online solver (3.4) is repeatedly called during an iterative procedure offline to build up $\mathcal{U}_{RB}$. By assuming that the approximation for the macroscopic density $\boldsymbol{\rho}_{RB}$ is given and performing a reduced transport sweep (i.e. obtaining $\mathbf{c}_{RB}(\boldsymbol{\Omega})$ for all $\boldsymbol{\Omega}$) with a fixed known density, we effectively decouple the angular dependence of the system. Without this technique, all $f_{\boldsymbol{\Omega}}$ are coupled and this leads to a larger system being inverted. The resulting increase in online (and thus offline) time is not amenable. An implication of this strategy is that we would need an *initial guess* for $\boldsymbol{\rho}_{RB}$. We noticed that a poor initial guess may lead to inaccurate reduced approximations. In this paper, we sample at a group of fixed but small number of quadrature points $\mathcal{P}_0 = \{\widehat{\boldsymbol{\Omega}}_k\}_{k=1}^{N_0}$ to perform a fully coupled full order solve to obtain this initial guess.

Least Squares Density Reconstruction

The existence of a low rank structure for $\mathbf{f}_{\boldsymbol{\Omega}}$ in its angular dependence, at least in the diffusive regime, indicates that the density can be accurately captured by the reduced space constructed offline. However, one main challenge in the offline stage is that the selected angular samples may not automatically give us a robust numerical quadrature formula in the angular space. Nevertheless, careful design of a density reconstruction algorithm approximating $\langle \cdot \rangle$ is essential for the robustness and efficiency of the SASI solver. In this paper, we propose a reconstruction algorithm based on a least squares procedure. Indeed, we first fix a group of robust high order quadrature points $\{(\bar{\boldsymbol{\Omega}}_j, \bar{\omega}_j)\}_{j=1}^{\bar{N}_{\boldsymbol{\Omega}}}$. Given an arbitrary group of angular samples $\mathcal{P}_{RB}$ and corresponding distribution functions $\{f_{\boldsymbol{\Omega}}\}$, we construct a least squares approximation of f, denoted by f_{ls} in the angular space. We then calculate the macroscopic density ρ based on f_{ls}.

$$\rho = \langle f \rangle \approx \sum_{j=1}^{\bar{N}_{\boldsymbol{\Omega}}} \bar{\omega}_j f_{\mathrm{ls}}(\bar{\boldsymbol{\Omega}}_j). \tag{3.5}$$

A natural set of space to consider in 1D are polynomial function space. In 2D, to suit the periodic structure, we use trigonometric function space. The details of the f_{ls} reconstruction are as follows. Here we denote the vector of degrees of freedom in space for $\mathbf{f}_{\boldsymbol{\Omega}}$ as $\left(\mathbf{f}_{\boldsymbol{\Omega}}(1), \ldots, \mathbf{f}_{\boldsymbol{\Omega}}(N_{dof})\right)^T$.

1D: With $\mathbf{\Omega} = v$, we find $\mathbf{f}_{ls}(v) = (p_1(v), \ldots, p_{N_{dof}}(v))^T$ such that

$$p_i(v) = \arg \min_{p \in P^s([-1,1])} \sum_{v \in \mathcal{P}_{RB}} (p(v) - \mathbf{f}_v(i))^2, \; i = 1, \ldots, N_{dof}, \tag{3.6}$$

where $P^s([-1,1])$ denotes the set of polynomials on $[-1, 1]$ of degree at most s.

2D: With $\mathbf{\Omega} = (\cos(\theta), \sin(\theta))$, we find $\mathbf{f}_{ls}(\mathbf{\Omega}) = \mathbf{f}_{ls}(\theta) = (t_1(\theta), \ldots, t_{N_{dof}}(\theta))^T$ such that,

$$t_i(\theta) = a_0 + \sum_{k=1}^{s} a_k \cos(k\theta) + \sum_{k=1}^{s-1} b_k \sin(k\theta), \; \text{and} \; (a_0, \ldots, a_s, b_1, \ldots, b_{s-1}) \; \text{solves}$$

$$\arg\min_{\widetilde{a}_0, \ldots, \widetilde{a}_s, \widetilde{b}_1, \ldots, \widetilde{b}_{s-1}} \sum_{\theta \in \mathcal{P}_{RB}} \left(\widetilde{a}_0 + \sum_{k=1}^{s} \widetilde{a}_k \cos(k\theta) + \sum_{k=1}^{s-1} \widetilde{b}_k \sin(k\theta) - \mathbf{f}_\theta(i) \right)^2. \tag{3.7}$$

We use the trigonometric least squares approximation (3.7), as $\theta = 0$ and $\theta = 2\pi$ represent the same point on the unit circle. We note that the resulting $\mathbf{f}_{ls}$ is a 2π-periodic function with respect to θ, preserving the property of the original distribution function.

The choice of s will be specified for our numerical experiments and discussed in Sect. 4.

L^1 Residual-Free Error Indicator

A critical piece for the RB greedy algorithm is an *a posteriori* error estimator which guides the surrogate space construction and certifies the accuracy of the RB solution. It is often residual-based and can be derived by mimicking the *a posteriori* error analysis of the underlying full order scheme [26, 43], with the RB solution taking the place of finite element solution which plays the role of the exact solution. *A posteriori* error analysis of the streamline upwind finite element method for the kinetic Eq. (1.1) is considered in [21]. With the need of solving its dual problem [25], its extension to the RB setting is not computationally appealing. As a result, we turn to the highly efficient and provably reliable L^1 residual-free error indicator proposed in [14].

Indeed, using the notation from Sect. 3.1, we assume the r-dimensional RB space is given by $\Gamma^r = \text{span}\{\eta_j, j = 1, \ldots, r\}$ where η_j is the solution to (3.1) when the parameter takes value μ^j. For a new parameter value μ whose corresponding RB solution is identified as

$$\eta(\mu) = \sum_{j=1}^{r} c_j(\mu)\eta_j.$$

The greedy choice informed by the L^1 error indicator then proceeds as follows[1]

$$\mu^{r+1} := \arg\max_\mu \Delta_r(\mu), \; \text{where} \; \Delta_r(\mu) = \sum_{j=1}^{r} |c_j(\mu)|. \tag{3.8}$$

As shown in [14], the $\{c_j(\mu)\}_{j=1}^r$ is the Lagrange interpolation basis in the parameter space. Taking the maximizer of $\Delta_r(\mu)$ then amounts to controlling the growth of the Lebesgue constant. This lead to its effectiveness for selecting the RB snapshots [14] including for nonlinear steady-state or time-dependent problems [11, 13]. Finally, it is imperative to note that this indicator is straightforward to implement and essentially free to compute.

---

[1] In actual computation, the basis $\{\eta_j\}$ will be processed e.g. subject to a Gram-Schmidt orthonormalization for numerical stability. The subsequent L^1 error indicator will have to incorporate the transformation matrix.

Greedy Algorithm

We are now ready to describe the greedy algorithm for iteratively constructing the reduced basis in the offline stage. It starts with a small set of quadrature points $\mathcal{P}_0 = \{\widehat{\boldsymbol{\Omega}}_k\}_{k=1}^{N_0}$ with the initial reduced space constructed as the span of the resulting snapshots and the initial density $\boldsymbol{\rho}^0$ computed accordingly. In each iteration, we calculate the projection of solution for an unselected parameter into the current reduced space and greedily expand the reduced space according to (3.8). With matrix-vector formulation, details of this algorithm is presented in Algorithm 2.

Algorithm 2: RBM greedy algorithm for radiative transfer equation.

1: **Input:** iteration number tolerance M_{tol}, spectral ratio tolerance r_{tol}, training set $\mathcal{P} = \{\boldsymbol{\Omega}_1, \ldots, \boldsymbol{\Omega}_{N_{\boldsymbol{\Omega}}}\}$, and an initial quadrature rule $\mathcal{P}_0 = \{\widehat{\boldsymbol{\Omega}}_k\}_{k=1}^{N_0}$.

2: Solve (globally coupled) (2.8) with a direct solver for $\boldsymbol{\Omega} \in \mathcal{P}_0$ to obtain $\mathbf{f}_{\widehat{\boldsymbol{\Omega}}_k}^0$ and $\boldsymbol{\rho}^0$.

3: **Iteration:**

4: Set $m = 0$, $r^m = 2r_{\text{tol}}$, $\mathbf{F}_{RB}^0 = (\mathbf{f}_{\widehat{\boldsymbol{\Omega}}_1}^0, \ldots, \mathbf{f}_{\widehat{\boldsymbol{\Omega}}_{N_0}}^0)$ and $\mathcal{P}_{RB} = \{\widehat{\boldsymbol{\Omega}}_k, k = 1, \ldots, N_0\}$.

5: **while** $m \leq M_{\text{tol}}$ and $r^m > r_{\text{tol}}$ **do**

6:    Perform an SVD $\mathbf{F}_{RB}^m = \mathcal{U}_{RB}^m \boldsymbol{\Lambda}_{RB}^m (\mathcal{V}_{RB}^m)^T$.

7:    Calculate the spectral ratio $r^m = \frac{\lambda_{\min}^m}{tr(\boldsymbol{\Lambda}_{RB}^m)}$ where $\lambda_{\min}^m$ is the minimal singular value.

8:    **for** $i = 1 : N_{\boldsymbol{\Omega}}$ **do**

9:      **if** $\boldsymbol{\Omega}_i \notin \mathcal{P}_{RB}$, **then**

10:        **Compute the RBM solution for** $\boldsymbol{\Omega}_i$, $\mathbf{f}_{\boldsymbol{\Omega}_i}^m = \mathcal{U}_{RB}^m \mathbf{c}_{RB}^m(\boldsymbol{\Omega}_i)$ with $\mathbf{c}_{RB}^m(\boldsymbol{\Omega}_i)$ solving

$$(\mathcal{U}_{RB}^m)^T \mathbf{U}_{\boldsymbol{\Omega}_i} \mathcal{U}_{RB}^m \mathbf{c}_{RB}^m(\boldsymbol{\Omega}_i) + (\mathcal{U}_{RB}^m)^T \boldsymbol{\Sigma}_t \mathcal{U}_{RB}^m \mathbf{c}_{RB}^m(\boldsymbol{\Omega}_i) = (\mathcal{U}_{RB}^m)^T \boldsymbol{\Sigma}_s \boldsymbol{\rho}^m + (\mathcal{U}_{RB}^m)^T \mathbf{g}.$$

11:        **Calculate the L^1 error indicator** $\mathcal{E}_{\boldsymbol{\Omega}_i}^m = \|\mathcal{V}_{RB}^n (\Lambda_{RB}^n)^{-1} \mathbf{c}_{RB}^m(\boldsymbol{\Omega}_i)\|_{\ell^1}$.

12:      **end if**

13:    **end for**

14:    Set $i_{\text{new}} = \arg\max_i\{\mathcal{E}_{\boldsymbol{\Omega}_i}^m\}$, $m := m + 1$, and $\mathcal{P}_{RB} = \mathcal{P}_{RB} \bigcup \{\boldsymbol{\Omega}_{i_{\text{new}}}, \widetilde{\boldsymbol{\Omega}}_{i_{\text{new}}}\}$.

15:    Execute the SASI method to solve (2.8) and update $\mathbf{f}_{\boldsymbol{\Omega}_{j_k}}^m$ for $\boldsymbol{\Omega}_{j_k} \in \mathcal{P}_{RB}$ to assemble

$$\mathbf{F}_{RB}^m = \left(\mathbf{f}_{\boldsymbol{\Omega}_{j_1}}^m, \ldots, \mathbf{f}_{\boldsymbol{\Omega}_{j_{N_m}}}^m\right), \quad N_m = 2m + N_0 \quad \text{and} \quad \mathcal{P}_{RB} = \{\boldsymbol{\Omega}_{j_k}\}_{k=1}^{N_m}.$$

16:    Reconstruct $\boldsymbol{\rho}^m$ based on $\mathbf{f}_{\boldsymbol{\Omega}}^m$ for $\boldsymbol{\Omega} \in \mathcal{P}_{RB}$ via (3.5) and (3.6) (or (3.7)).

17: **end while**

18: **Output:** the reduced basis $\mathcal{U}_{RB} = \mathcal{U}_{RB}^m$ and $\boldsymbol{\rho}_{RB} = \boldsymbol{\rho}^m$.

Remark 3.2 We emphasize two features of the greedy algorithm. The first feature is the *symmetry-enhancing greedy addition*. When we augment the RB space, in addition to the maximizer of the error indicator $\boldsymbol{\Omega}_{i_{\text{new}}}$ as determined by (3.8), we include its symmetric counterpart $\widetilde{\boldsymbol{\Omega}}_{i_{\text{new}}}$. Assuming that the training set is $\mathcal{P} = \{\boldsymbol{\Omega}_j\}_{j=1}^{N_{\boldsymbol{\Omega}}}$. For each $\boldsymbol{\Omega}$, in 1D, we define $\widetilde{\Omega} = -\Omega$. In 2D, given $\boldsymbol{\Omega} = (\cos\theta, \sin\theta)$ with $\theta \in (0, 2\pi)$, we define $\widetilde{\boldsymbol{\Omega}} = (\cos(\widetilde{\theta}), \sin(\widetilde{\theta}))$ with $\widetilde{\theta} = \mod(\theta + \pi, 2\pi)$. By construction, if $\boldsymbol{\Omega} \in \mathcal{P}$ then $\widetilde{\boldsymbol{\Omega}} \in \mathcal{P}$. In our numerical simulations, we observe that this method is more robust than adding $\boldsymbol{\Omega}_{i_{\text{new}}}$ into $\mathcal{P}_{RB}$ alone. Indeed, for the 2D examples in Sect. 4.2, if one angular sample is added per iteration, the SASI iterative solver may fail to converge.

The second feature is a *spectral ratio stopping criteria*. The purpose is to mitigate the fact that our L^1 residual-free error indicator, albeit highly effective in identifying the next representative parameter value, is not an error estimator. Inspired by the POD method, we monitor a spectral ratio as an additional stopping criteria. We define the spectral ratio for the m-th iteration r^m as

$$r^m = \frac{\lambda_{\min}^m}{Tr(\Lambda_{RB}^m)}, \tag{3.9}$$

where $\lambda_{\min}^m$ is the smallest diagonal element of Λ_{RB}^m, and $Tr(\cdot)$ is the trace operator.

4 Numerical Results

In this section, we present a series of one- and two-dimensional numerical examples to showcase the performance of the proposed RB method. For the underlying DG spatial discretization, the discrete space in (2.4) with $K = 1$ is used. With the consideration for the efficiency and robustness of the SASI method, we set the degree parameter s in the density reconstruction as $s = m + 1$ for 1D slab geometry and $s = \min(5, m + 1)$ in 2D during the m-th iteration of the greedy algorithm. Throughout the experiments, we measure the following absolute and relative L^2 errors.

$$\mathcal{E}_f = \max_j \| f_{\mathrm{F}}(\mathbf{\Omega}_j) - f_{\mathrm{R}}(\mathbf{\Omega}_j)\|, \qquad R_f = \max_j \frac{\| f_{\mathrm{F}}(\mathbf{\Omega}_j) - f_{\mathrm{R}}(\mathbf{\Omega}_j)\|}{\| f_{\mathrm{F}}(\mathbf{\Omega}_j)\|},$$

$$\mathcal{E}_\rho = \|\rho_{\mathrm{F}} - \rho_{\mathrm{R}}\|, \qquad R_\rho = \frac{\|\rho_{\mathrm{F}} - \rho_{\mathrm{R}}\|}{\|\rho_{\mathrm{F}}\|}.$$

Here f_{F}, ρ_{F} denote the full order numerical solutions. The RB solutions are f_{R}, ρ_{R} and $\|\cdot\|$ is the standard L^2 norm of $L^2(\mathbf{X})$.

In the current setting, there are no essential physical parameters (e.g. scattering cross section, boundary conditions) whose "multi-query" nature will make the (one-time) offline investment more worthwhile. However, we still compare the proposed RB method (including both the offline and online stages) against (a single query of) the full order DG solver using the training set for angular space discretization. We note that the online stage can be utilized to predict solutions at angular samples outside of the training set, a feat out of reach by the full order DG scheme. When the proposed method is utilized as a surrogate for the full order solve with respect to the training set, we concern both the offline and online efficiency. When its online stage is applied to predict solutions at "unseen" angular samples, we only take the online efficiency into account. We call the error associated with the training set "training error" and the error associated with the test set "testing error". Here and below, the training set refers to the set of parameter values of the angular variable used during the offline stage to build the surrogate space, while the test set refers to that used during the online stage to test the performance of the RB method.

4.1 One-Dimensional Examples

We perform the one-dimensional experiments on a slab geometry domain $\mathbf{X} = [x_L, x_R]$, which is discretized by a uniform mesh with $\Delta x = 0.125$. The initial guess for our algorithm is obtained by 2 Gauss-Legendre points, i.e. $N_0 = 2$. We consider 5 examples and conduct two tests for each of them. The training set consists of 24 Gauss-Legendre points. We consider two

different r_{tol} as 10^{-4} and 10^{-6}. The algorithm stops once r_{tol} is reached leading to different RB dimensions for different examples. The testing errors associated with a test set will be reported. For the second test, we ask the algorithm to generate reduced bases with $M_{\text{tol}} = 12$ for all problems and record the training error between the RB solutions and the full order solutions. In the end, we demonstrate the robustness of the algorithm by varying the strength of the scattering cross section σ_s. These tests aim at showing the capability of our method to predict solutions at angular samples outside of the training set and the effectiveness of the L^1 residual-free error indicator.

Example 1 (Scattering dominant)

$$\mathbf{X} = [0, 10], \ G = 0.01, \ \sigma_t = 100, \ \sigma_s = 100, \ f(0, v) = 0$$
with $v > 0$, $f(10, v) = 0$ with $v \leq 0$.

Example 2 (Spatially varying scattering coefficient)

$$\mathbf{X} = [0, 10], \ G = 0.01, \ \sigma_t = 100(1 + x), \ \sigma_s = 100(1 + x),$$
$$f(0, v) = 0 \text{ with } v > 0, \ f(10, v) = 0 \text{ with } v \leq 0.$$

Example 3 (Two-material problem 1)

$$\mathbf{X} = [0, 20], \ G = \begin{cases} 5, \ 0 < x < 10, \\ 0, \ 10 < x < 20, \end{cases} \quad \sigma_t = 100, \ \sigma_s = \begin{cases} 90, \ 0 < x < 10, \\ 100, \ 10 < x < 20, \end{cases}$$
$$f(0, v) = 0 \text{ with } v > 0, \ f(20, v) = 0 \text{ with } v \leq 0.$$

Example 4 (Two-material problem 2)

$$\mathbf{X} = [0, 11], \ G = 0, \ \sigma_t = \begin{cases} 100, \ 1 < x < 11, \\ 2, \ 0 < x < 1, \end{cases} \quad \sigma_s = \begin{cases} 100, \ 1 < x < 11, \\ 0, \ 0 < x < 1, \end{cases}$$
$$f(0, v) = 5 \text{ with } v > 0, \ f(11, v) = 0 \text{ with } v \leq 0.$$

Example 5 (Transport dominant)

$$\mathbf{X} = [0, 10], \ G = 0.01, \ \sigma_t = 1.2, \ \sigma_s = 1, \ f(0, v) = 0 \text{ with } v > 0, \ f(10, v) = 0$$
with $v \leq 0$.

Online prediction accuracy We first present the prediction accuracy for a test set with 32 quadrature points in the angular variable, that are different from the angular samples in the training set. With $r_{\text{tol}} = 10^{-4}$, the comparison between the density obtained by the RB solutions and that by the full order solutions is presented in Fig. 1. For all five examples, the RB solutions match the full order solutions well. We present the testing errors in f and training errors in ρ for these examples in Tables 1 and 2. One can see that the RB method is most effective for scattering dominant problems (i.e. Examples 1-3). It achieves very good degree of accuracy (3 to 4 digits) for both f and ρ with just 4 to 6 reduced basis functions. The solution of the two-material problem (Example 4), that involves a purely absorbing subregion without scattering, has more complicated structure. Consequently, more basis functions are needed, and the accuracy of f is worse than the other examples. Example 5 is transport dominant leading to distribution functions being far from the macroscopic density. As a result, it demands the most number of RB functions. It is worth noting that, for all five examples, the method commits less than 1% relative L^2 error for ρ. To reach this accuracy,

(a) Example 1, 4 reduced basis   **(b)** Example 2, 4 reduced basis   **(c)** Example 3, 4 reduced basis

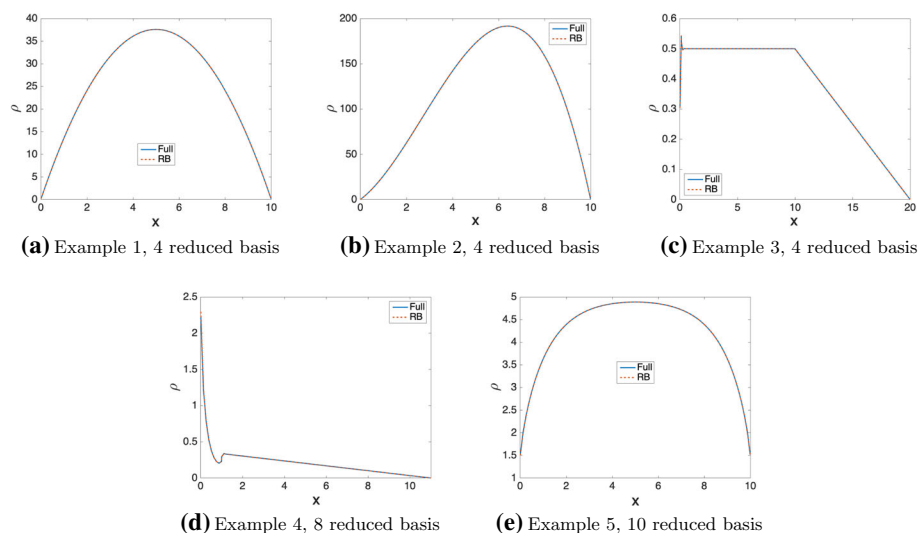**(d)** Example 4, 8 reduced basis   **(e)** Example 5, 10 reduced basis

Fig. 1 The true (i.e. full order) and RB densities for the one-dimensional test examples 1-5 of 1D slab geometry. The number of reduced basis is determined by $r_{\text{tol}} = 10^{-4}$

Table 1 Testing error for f and training error for ρ with $r_{\text{tol}} = 10^{-4}$, 1D slab geometry

	RB dimension	$\mathcal{E}_f$	R_f	$\mathcal{E}_\rho$	$R_\rho(\%)$
Example 1	4	8.04e-3	9.26e-3%	8.00e-3	9.21e-3
Example 2	4	8.51e-3	2.05e-3%	8.49e-3	2.04e-3
Example 3	4	9.35e-4	5.13e-2%	8.25e-4	4.53e-2
Example 4	8	9.99e-2	1.28e+1%	2.59e-3	2.54e-1
Example 5	10	7.50e-2	5.39e-1%	4.37e-3	3.21e-2

Table 2 Testing error for f and training error for ρ with $r_{\text{tol}} = 10^{-6}$, 1D slab geometry

	RB dimension	$\mathcal{E}_f$	R_f	$\mathcal{E}_\rho$	$R_\rho(\%)$
Example 1	6	3.34e-3	3.84e-3%	3.31e-3	3.80e-3
Example 2	4	8.51e-3	2.05e-3%	8.49e-3	2.05e-3
Example 3	6	4.31e-4	4.83e-2%	3.77e-4	2.07e-2
Example 4	10	4.73e-2	6.09e-0%	1.91e-3	1.87e-1
Example 5	14	2.78e-2	2.00e-1%	1.93e-3	1.42e-2

the full model needs 160 degrees of freedom for each angular sample, while our RBM uses at most 14 global reduced basis functions.

Effectiveness of the error indicator In Fig. 2, we present the histories of the convergence of the L^2 error of f and the spectral ratio r^m as the number of iteration grows. We observe that the error decays as the dimension of the reduced space increases with a terminal error smaller than 10^{-6} for the worst case. We also note that the spectral ratio r^m decays exponentially, albeit with different speed for different examples. The pattern, across different problems, of
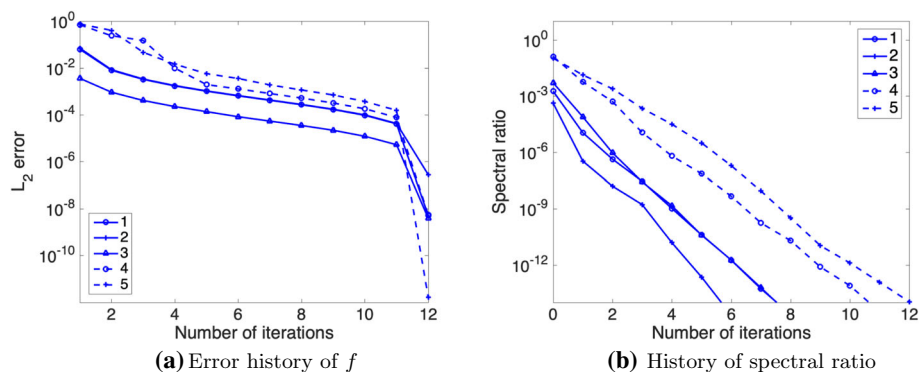
(a) Error history of f

(b) History of spectral ratio

Fig. 2 Histories of convergence of the L^2 error of f and the monitored spectral ratio during training for 1D Examples 1–5

(a) Dimension of reduced space for $r_{\text{tol}} = 10^{-8}$
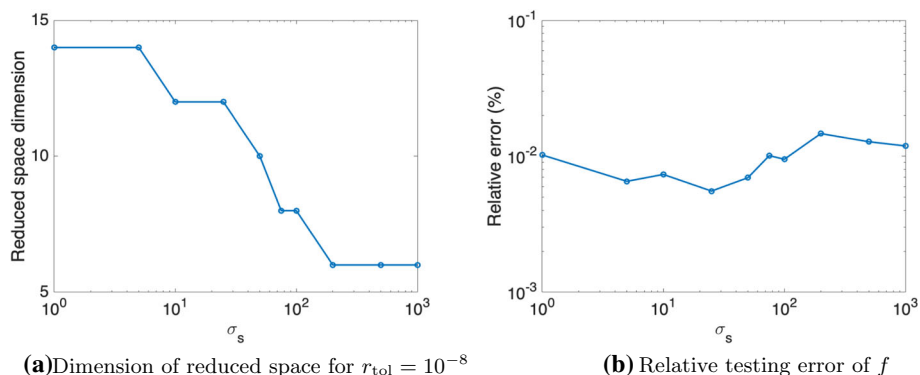
(b) Relative testing error of f

Fig. 3 RBM robustness test: Dimension of the reduced space and the relative testing error versus different scattering cross section

the spectral ratio decay leads to the different terminal RB dimensions for different problems as shown in Tables 1 and 2.

Robustness test with respect to σ_s: To showcase the robustness of our method, we consider the following example by varying strength of the scattering cross section σ_s,

$$\mathbf{X} = [0, 10], \ G = 0.01, \ \sigma_t \equiv C + 0.5, \ \sigma_s \equiv C, \ f(0, v) = 0 \text{ with } v > 0, \ f(10, v) = 0$$
$$\text{with } v \leq 0,$$

where $C = 1, 5, 10, 25, 50, 75, 100, 200, 500, 1000$. We use 40 Gauss-Legendre points as the training set, 32 Gauss-Legendre points as the test set, and take $r_{\text{tol}} = 10^{-8}$. The dimensions of the resulting RB spaces and the corresponding relative testing errors are reported in Fig. 3. We clearly observe that the more scattering dominant the problem is, the fewer reduced basis functions are needed. The fact that the relative testing error is on the same level for different σ_s, even though we only monitor the spectral ratio, attests to the robustness of our method and the reliability of our error indicators.

Fig. 4 Geometry of the checkerboard (2D Example 1). Black region: scattering dominant; white region: transport dominant
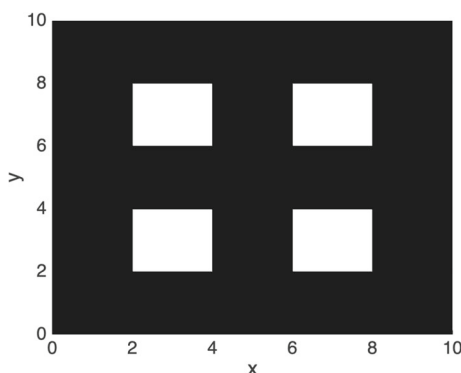
4.2 Two-Dimensional Examples

In this section, we consider the following four examples, all with the source term $G(x, y) = \exp\left(-100((x - 5)^2 + (y - 5)^2)\right)$ and the zero inflow boundary condition on the computational domain $\mathbf{X} = [0, 10]^2$.

Example 1 (Checkerboard) This is a multiscale problem, with part of the domain being scattering dominant and the rest being transport dominant. The checkerboard geometry is shown in Fig. 4. The white region is defined as $\bigcup_{i,j=1}^2 \{\max(|x - x_i|, |y - y_j|) < 1\}$,   with $x_1 = y_1 = 3$, $x_2 = y_2 = 7$, where we set $\sigma_s(x, y) = 1$ and $\sigma_t(x, y) = \sigma_s(x, y) + 1$. In the black region, we have $\sigma_s(x, y) = \sigma_t(x, y) = 100$.

Example 2 (Scattering dominant) $\sigma_s(x, y) = \sigma_t(x, y) = 100$.

Example 3 (Intermediate regime) $\sigma_s(x, y) = \sigma_t(x, y) = 10$.

Example 4 (Transport dominant) $\sigma_s(x, y) = \sigma_t(x, y) = 1$.

In our experiments, a uniform rectangular mesh of 40×40 is used for the DG scheme. To generate the initial guess, we start from $\{\theta_j\}_{j=1}^{N_0}$, with $\theta_j = \frac{2(j-1)\pi}{N_0}$, and use $N_0 = 4$ for Examples 1-3 and $N_0 = 8$ for Example 4.

We start with applying the proposed RB algorithm with a training set of $N_{\boldsymbol{\Omega}} = 32$ quadrature points in the angular variable, and obtain a surrogate model for the full order upwind DG solver. To show the efficiency and accuracy of this surrogate solver, we first present the relative computational time and the accuracy with respect to a full order solve based on the same $N_{\boldsymbol{\Omega}} = 32$ quadrature points, and then use the RB method to predict f at angular samples that are not included in the training set.

Efficiency and the training accuracy as a surrogate solver In Fig. 5, the relative total computational time (including both the online and offline time) with respect to the full order solver based on $N_{\boldsymbol{\Omega}} = 32$ quadrature points as well as the relative L^2 error are presented as a function of r_{tol}. As expected, the error decays as r_{tol} becomes smaller. Examples 1-3 are in the diffusive regime, intermediate regime and as a multiscale problem with large scattering dominant subregions, respectively. For these problems, with a suitable r_{tol}, one can achieve less than 1% error with less than 18% relative computational time. Example 4 is transport dominant. To achieve less than 2% error for this example, we need to use $r_{\text{tol}} = 10^{-2}$ and the offline computational time is longer than that of a full order solve. To understand the poor
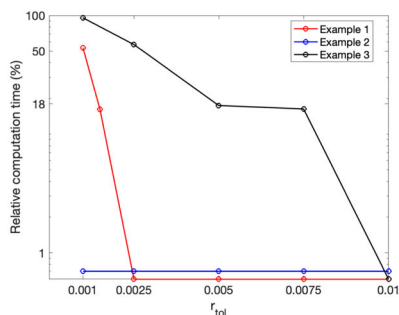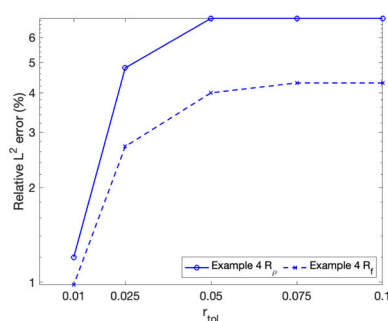
(a) Relative total computational time for Examples 1-3 **(b)** Relative total computational time for Example 4
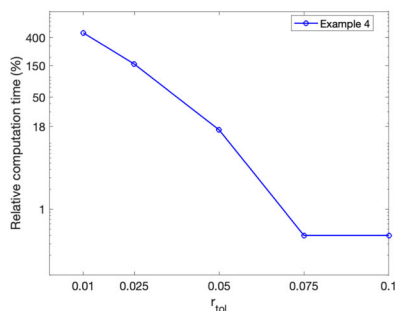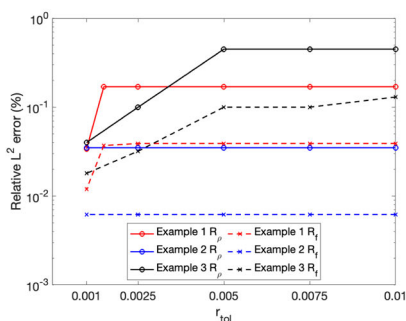
(c) Relative errors for Examples 1-3   **(d)** Relative errors for Example 4

Fig. 5 Relative total computational time and relative L^2 training errors of 2D examples with $N_{\mathbf{\Omega}} = 32$ and different r_{tol}

efficiency observed for Example 4, one can recall that at the end of each greedy iteration in the offline stage, a full order method will be applied based on the angular sample set $\mathcal{P}_{RB}$ selected so far. Though each problem is of small size, the transport-dominant nature of the model can require relatively more greedy iterations hence more solves of such small problems, due to the known slow decay of the Kolmogorov N-width for transport dominant problems [22, 37]. To improve the offline efficiency for such problems, nonlinear reduced order models may be needed.

In Table 3, the training errors and the dimensions of the reduced order model with $N_{\mathbf{\Omega}} = 32$ are presented. For Examples 1-3 with $r_{\text{tol}} = 10^{-3}$, we achieve less than 0.1% relative L^2 errors, and for Example 4 with $r_{\text{tol}} = 10^{-2}$, we achieve less than 2% relative L^2 errors. The scattering dominant problem, the multiscale checkerboard problem, and the intermediate regime problem all need relatively smaller RB spaces. The transport dominant example requires the most reduced basis functions, and has relatively lower accuracy. Nevertheless, the full order solve takes 6400 degrees of freedom for each sample of the angular variable. In comparison, our RB algorithm only needs fewer than 0.5% degrees of freedom online.

Online efficiency and the prediction accuracy With the reduced basis functions obtained from a training set of $N_{\mathbf{\Omega}} = 32$ quadrature points, we predict the solution f at other angular samples. In Fig. 6a, we present the relative testing errors for a group of test sets. As the size N_{test} of test sets varies, the relative errors stay at almost the same level as that for the original

Table 3 training errors for 2D Examples 1-3 with $r_{\text{tol}} = 10^{-3}$ and 2D Example 4 with $r_{\text{tol}} = 10^{-2}$ and $N_{\mathbf{\Omega}} = 32$

	RB dimension	$\mathcal{E}_f$	R_f	$\mathcal{E}_\rho$	$R_\rho(\%)$
Example 1	8	1.50e-3	1.21e-2%	7.69e-4	3.44e-2
Example 2	4	2.40e-3	6.23e-3%	2.40e-3	3.49e-2
Example 3	10	7.10e-4	1.83e-2%	2.77e-4	4.03e-2
Example 4	26	2.95e-4	9.97e-1%	9.72e-4	1.15e-0
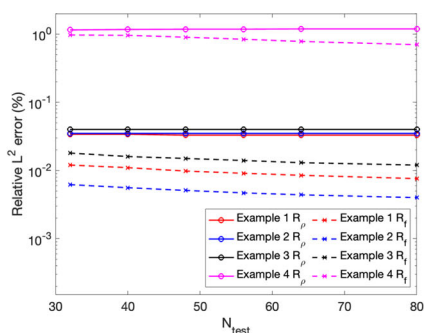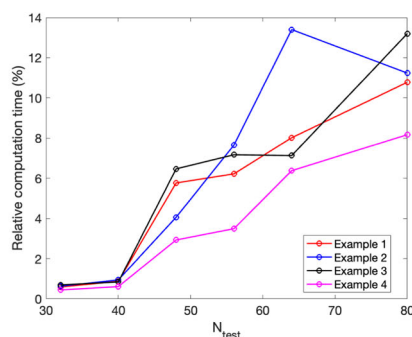
(a) Relative testing errors   **(b)** Relative online computational time

Fig. 6 Relative testing errors and relative online computational time for different test sets of size N_{test}, with $r_{\text{tol}} = 10^{-3}$ for 2D Examples 1–3 and $r_{\text{tol}} = 10^{-2}$ for 2D Example 4

training set. To illustrate the computational efficiency of online prediction, we further report in Fig. 6b the relative online computational time with respect to that of the full order solve (with the same angular quadrature points as the test set), and they are always below 14% as the size of test sizes varies. This implies that the proposed RB algorithm can be used as a building block to construct ROMs when the model has essential parameters (e.g. scattering or absorption cross sections, boundary data etc), and substantial saving can be expected for online computation.

The comparison between the densities obtained by the RB method (trained with $N_{\mathbf{\Omega}} = 32$) and the full order method (with 40 quadrature points) are presented in Fig. 7, with the former computed as the terminal density when $r_{\text{tol}} = 10^{-3}$ is reached for Examples 1-3 and $r_{\text{tol}} = 10^{-2}$ is reached for Example 4. We see that the RB solution and the full order solution match each other well. Moreover, the second row of Fig. 7 demonstrates the effectiveness of our method in mitigating the ray effect [30], which refers to the phenomenon that the particles mainly propagate along the directions of sampled angular directions and the numerical solution has noticeable unphysical oscillations. As shown in Fig. 7(d), the initial guess for the problem in the transport regime suffers severely from the ray effect. Even with an initial guess of poor quality, the ray effect in the reduced order solution is less pronounced as the dimension of the reduced space grows (see Figss 7e and Fig. 7f).

Effectiveness of the error indicator In the last test, we ask the algorithm to generate $M_{\text{tol}} = 16$ reduced bases for all problems and monitor the training errors. The histories of convergence for the L^2 errors of f and the spectral ratio r^m are presented in Fig. 8. Similar to the one-dimensional cases, we see that, as the dimension of the reduced space increases, the overall trend of the error and spectral ratio is decreasing for all examples, albeit slower than the
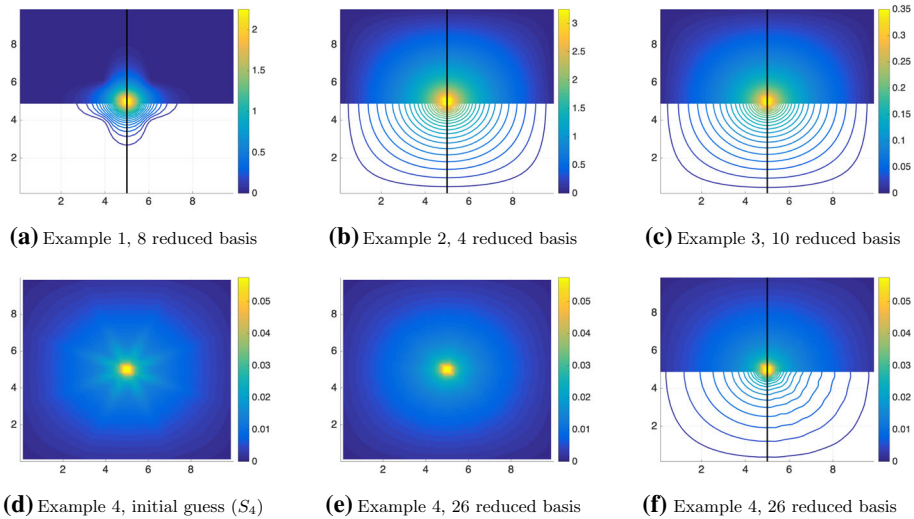
(a) Example 1, 8 reduced basis

(b) Example 2, 4 reduced basis

(c) Example 3, 10 reduced basis

(d) Example 4, initial guess (S_4)

(e) Example 4, 26 reduced basis

(f) Example 4, 26 reduced basis

Fig. 7 Density for 2D examples. Top row: Examples 1-3, $r_{tol} = 10^{-3}$; Bottom row: Example 4, $r_{tol} = 10^{-2}$. In each plot - left half: full order solution; right half: RB solution; Top half: full/RB solution; bottom half: contour of the full/RB solution

(a) Error history of f

(b) History of spectral ratio
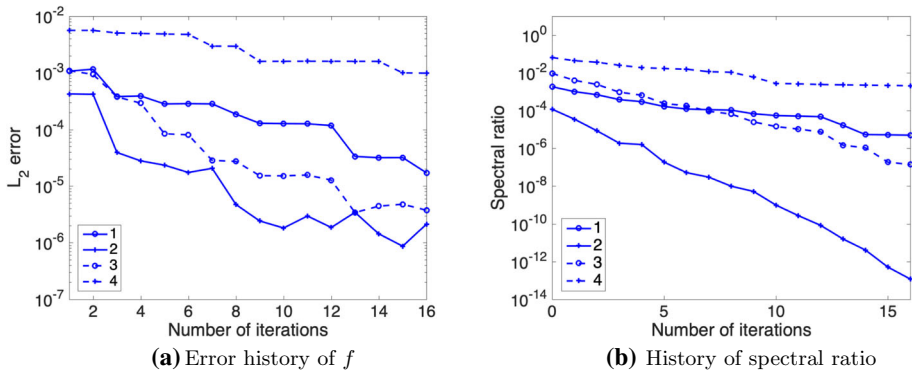
Fig. 8 Histories of convergence of the L^2 error of f and the monitored spectral ratio during training for 2D Examples 1–4

one-dimensional cases. Moreover, the different speed of decay of the spectral ratio r^m for different problems leads to the different terminal RB dimensions as shown in Tab. 3. The fact that these quantities decay slower for transport dominant problems is consistent with the slow decay of the Kolmogorov N-width for transport problems, see e.g. [22, 37]. As a result, the performance of our method and many other linear ROMs suffer.

Finally we want to mention that in Appendix B, we plot and comment about the leading reduced basis functions generated by the proposed algorithm for selected 1D and 2D examples.

5 Conclusion

In this paper, we design a RB method to construct an angular-space reduced order model for the linear radiative transfer equation. Unlike the standard setting where RBM applies, the solutions for different parameter values are coupled through an integration operator. This coupling makes impossible the direct inquiry of the snapshot for any particular parameter value. An additional challenge is that parameter ensemble identified by traditional RBM is usually unstructured, and thus may not form a set of quadrature points for a robust and accurate integration toward the macroscopic density which is however crucial for the efficiency and robustness of the full order SASI solver.

Via a careful iterative procedure where the macroscopic density is treated explicitly allowing a transport sweep and then updated afterwards, a least squares density reconstruction at each of the relevant physical locations, a L^1-based residual-free error indicator, and a symmetry enhancing greedy addition, we successfully designed the first RBM for the kinetic transport equation. Our numerical experiments indicate that the new method is highly effective for the scattering dominant problems, the intermediate regime problems, and the multiscale problems with scattering dominant subregions. Moreover, as the reduced space grows, the ray effect can be mitigated. While less efficient for transport dominant problems, it is capable of decreasing the problem size by more than one order of magnitude and achieving one digit of accuracy. Designing a RBM that works equally well for transport dominant problems, likely by constructing nonlinear reduced manifolds, constitutes our future work. Other future work include the application to the full 5D model and the extension to time dependent problems and other kinetic models.

Data Availability Enquiries about data availability should be directed to the authors.

Declarations

Conflict of interest The authors have not disclosed any competing interests.

Appendix

Appendix A RB Method with the Diffusion Synthetic Acceleration

In this work, we apply the S2SA to speed up the convergence of the source iteration when algebraically solving the upwind DG discretization. One main advantage of the S2SA is that the same kind of kinetic solver as the full order one is applied. Another type widely used synthetic acceleration is the diffusion synthetic acceleration (DSA) [2, 3, 47]. Instead of using a low order S_N model for a kinetic problem to approximate the correction Eq. (2.10) for δf^{k+1} first and then to compute $\rho^{k,c} = \langle \delta f^{k+1} \rangle$, the DSA method works with a discrete diffusion approximation, that is "consistent" (see [2] for the definition of the consistency), to approximate $\rho^{k,c}$ directly. It is known that the source iteration with "inconsistent" DSA may converge slowly or even diverge in some regimes [2]. Compared with a straightforward

S2SA method, the DSA will involve fewer degrees of freedom. Next we will use the 1D model on the slab geometry as an example to present a DSA method that is consistent to the upwind DG discretization, and then demonstrate and compare the performance of the RB method with both synthetic acceleration strategies.

A.1 A Consistent DSA Method

For the 1D slab geometry with $X = [0, 1]$, by using an ansatz $\delta f(x, v) = \rho^{k,c}(x) + 3vg(x)$ and taking the zeroth, first moments of the correction equation (2.10) in the angular variable, we obtain an approximated diffusion model

$$\underbrace{\langle v^2 \rangle}_{=\frac{1}{3}} \partial_x \ (\sigma_t^{-1} \partial_x \rho^{k,c}) + \sigma_a \rho^{k,c} = \sigma_s(\rho^{k,*} - \rho^k), \quad \text{on } X \tag{A.1a}$$

or, equivalently, in its first order form,

$$\partial_x g + \sigma_a \rho^{k,c} = \sigma_s(\rho^{k,*} - \rho^k), \quad \underbrace{\langle v^2 \rangle}_{=\frac{1}{3}} \partial_x \rho^{k,c} + \sigma_t g = 0. \tag{A.2a}$$

They will be complemented by the boundary conditions,

$$\langle v(\rho^{k,c} + 3vg) \rangle^+ = \frac{1}{4}(\rho^{k,c} + 2g) = \frac{1}{4}(\rho^{k,c} - \frac{2}{3\sigma_t}\partial_x \rho^{k,c}) = 0, \quad \text{at } x = 0, \tag{A.3a}$$

$$\langle v(\rho^{k,c} + 3vg) \rangle^- = -\frac{1}{4}(\rho^{k,c} - 2g) = -\frac{1}{4}(\rho^{k,c} + \frac{2}{3\sigma_t}\partial_x \rho^{k,c}) = 0, \quad \text{at } x = 1. \tag{A.3b}$$

Here $\langle \eta \rangle^+ = \frac{1}{2} \int_{v>0} \eta(v) dv$, $\langle \eta \rangle^- = \frac{1}{2} \int_{v<0} \eta(v) dv$.

Let $\{T_i = [x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}}], \ i = 1 \ldots, N_x\}$ be a partition of the domain $X = [0, 1]$, we then discretize (A.2) with a DG method: we seek $\rho_h^{k,c}, g_h \in U_h^K$, such that $\forall \phi_h, \psi_h \in U_h^K$, $\forall i = 1, \ldots, N_x$,

$$-\int_{I_i} g_h \partial_x \phi_h dx + (\widehat{g_h}\phi_h^-)_{i+\frac{1}{2}} - (\widehat{g_h}\phi_h^+)_{i-\frac{1}{2}} + \int_{I_i} \sigma_a \rho_h^{k,c} \phi_h dx = \int_{I_i} \sigma_s(\rho^{k,*} - \rho^k)\phi_h dx, \tag{A.4a}$$

$$\frac{1}{3}\left(-\int_{I_i} \rho_h^{k,c} \partial_x \psi_h dx + (\widehat{\rho_h^{k,c}}\psi_h^-)_{i+\frac{1}{2}} - (\widehat{\rho_h^{k,c}}\psi_h^+)_{i-\frac{1}{2}}\right) + \int_{I_i} \sigma_t g_h \psi_h dx = 0. \tag{A.4b}$$

The key to make the method (A.4) consistent to the upwind DG method for the transport sweep step lies in the numerical fluxes $(\widehat{g_h})_{i+\frac{1}{2}}$ and $(\widehat{\rho_h^{k,c}})_{i+\frac{1}{2}}$, that shall be based on the upwind flux for δf.[2] With this in mind, we take

$$(\widehat{g_h})_{i+\frac{1}{2}} = \langle v\widehat{\delta f}^{\text{upwind}} \rangle_{i+\frac{1}{2}} = \langle v(\rho_h^{k,c} + 3vg_h) \rangle^+(x_{i+\frac{1}{2}}^-) + \langle v(\rho_h^{k,c} + 3vg_h) \rangle^-(x_{i+\frac{1}{2}}^+)$$

$$= \{g_h\}_{i+\frac{1}{2}} - \frac{1}{4}[\rho_h^{k,c}]_{i+\frac{1}{2}}, \tag{A.5a}$$

$$\frac{1}{3}(\widehat{\rho_h^{k,c}})_{i+\frac{1}{2}} = \langle v^2\widehat{\delta f}^{\text{upwind}} \rangle_{i+\frac{1}{2}} = \langle v^2(\rho_h^{k,c} + 3vg_h) \rangle^+(x_{i+\frac{1}{2}}^-) + \langle v^2(\rho_h^{k,c} + 3vg_h) \rangle^-(x_{i+\frac{1}{2}}^+)$$

---

[2] In our RB method, we use the exact values of $\langle v^2 \rangle$ in (A.4) and $\langle v^k \rangle^\pm$, $k = 1, 2, 3$ in (A.5). In literature, numerical integrations with certain property are often used.

$$= \frac{1}{3}\left(\{\rho_h^{k,c}\}_{i+\frac{1}{2}} - \frac{9}{8}[g_h]_{i+\frac{1}{2}} \right). \tag{A.5b}$$

Here $u(x_{i+\frac{1}{2}}^-) = u_{i+\frac{1}{2}}^-$ (resp. $u(x_{i+\frac{1}{2}}^+) = u_{i+\frac{1}{2}}^+$) stands for the left (resp. right) limit of $u(x)$ at the cell interface $x_{i+\frac{1}{2}}$. And $\{u\}_{i+\frac{1}{2}} = \frac{1}{2}(u_{i+\frac{1}{2}}^+ + u_{i+\frac{1}{2}}^-)$, $[u]_{i+\frac{1}{2}} = u_{i+\frac{1}{2}}^+ - u_{i+\frac{1}{2}}^-$. At boundaries, the numerical fluxes are set as (A.5) with $\langle v^k(\rho_h^{k,c} + 3vg_h)\rangle^+(x_{\frac{1}{2}}^-) = 0$, $\langle v^k(\rho_h^{k,c} + 3vg_h)\rangle^-(x_{N_x+\frac{1}{2}}^+) = 0$, $k = 1, 2$. In actual simulation, we eliminate g_h in (A.4) at the algebraic level, leading to a smaller linear system for $\rho_h^{k,c}$ only, that is given in its matrix-vector form as follows,

$$\left(\Sigma_a - \frac{1}{4}\mathbf{D}_{\text{jump}} - \frac{1}{3}\mathbf{D}_c(\Sigma_t - \frac{3}{8}\mathbf{D}_{\text{jump}})^{-1}\mathbf{D}_c\rho^{k,c} \right)\rho^{k,c} = \Sigma_S(\rho^{k,*} - \rho^k), \tag{A.6}$$

with $\mathbf{D}_c = \frac{1}{2}\left(\mathbf{D}^+ + \mathbf{D}^- \right)$, $\mathbf{D}_{\text{jump}} = \mathbf{D}^+ - \mathbf{D}^-$, where

$$(\mathbf{D}^+)_{kl} = -\sum_{i=1}^{N_x}\int_{T_i}\partial_x\phi_k(x)\phi_l(x)dx + \sum_{i=1}^{N_x-1}\phi_l(x_{i+\frac{1}{2}}^+)\phi_k(x_{i+\frac{1}{2}}^-) - \sum_{i=1}^{N_x}\phi_l(x_{i-\frac{1}{2}}^+)\phi_k(x_{i-\frac{1}{2}}^+), \tag{A.7a}$$

$$(\mathbf{D}^-)_{kl} = -\sum_{i=1}^{N_x}\int_{T_i}\partial_x\phi_k(x)\phi_l(x)dx + \sum_{j=1}^{N_x}\phi_l(x_{i+\frac{1}{2}}^-)\phi_k(x_{i+\frac{1}{2}}^-) - \sum_{i=2}^{N_x}\phi_l(x_{i-\frac{1}{2}}^-)\phi_k(x_{i-\frac{1}{2}}^+). \tag{A.7b}$$

One example of "partially consistent" DSA methods is to use central fluxes as in [3]. Partially consistent DSA methods may result in slower convergence.

A.2 Comparison of the RB Method with the S2SA and the DSA

For the 1D examples, we replace the S2SA with the DSA in the source iteration, and compare the performance of the overall RB algorithm with the two different acceleration strategies.

Table 4 Errors of the RB Method with the DSA and S2SA for 1D examples

	RB dimension	$\mathcal{E}_f$-DSA	$\mathcal{E}_f$-S2SA	$\mathcal{E}_\rho$-DSA	$\mathcal{E}_\rho$-S2SA
Example 1	4	6.26e-2	8.04e-3	6.25e-2	8.00e-3
Example 2	4	6.99e-2	8.51e-3	6.98e-2	8.49e-3
Example 3	4	3.28e-3	9.35e-4	2.88e-3	8.25e-4
Example 4	RB-DSA not convergent				
Example 5	RB-DSA not convergent				

Table 5 Relative computational time for 1D examples: (time with S2SA)/(time with DSA)

	Example 1	Example 2	Example 3	Example 4	Example 5
RB: offline	4.68	4.67	3.37	RB-DSA not convergent	
Full	1.57	1.05	1.34	1.12	0.96

The stopping criteria is $r_{tol} \leq 10^{-4}$. For all examples, the full order solvers with the S2SA and the DSA in the source iteration always converge to the same result. The errors for the RB method with the S2SA and the DSA are reported in Table 4, and the relative computational time, defined as (time with S2SA)/(time with DSA), is summarized in Table 5. The full order model with the DSA is slightly more efficient in the diffusive regime, and it is comparable with the S2SA method in other regimes. For the diffusive and intermediate regimes (Examples 1-2), the RB method with the DSA is more efficient in both offline, but the errors are relatively larger. The angular samples picked by the greedy algorithm in the RB method are the same for both the DSA and S2SA, and hence the online computational costs of the RB method with the DSA and the S2SA are close to each other. For Example 4 (two-material problem) and Example 5 (transport regime), the RB method with the DSA fails to converge to the correct solution. It is known that reduced order methods can be more sensitive to the choice of preconditioners compared with full order solvers [12, 45, 48]. In summary, when combined with our RB method, the S2SA is more robust with various regimes and slightly more accurate. The RB method with the DSA is slightly more efficient if it converges, but it may fail to converge for problems with transport-dominant (sub)regions.

Appendix B Plots of Selected RB Functions

We here present some selected RB functions generated by the proposed algorithm. In Fig. 9, we plot the first four RB functions (after the SVD orthogonalization step) for 1D Examples 1, 4, 5 from Sect . 4.1. The most interesting example is the two-material problem in Example 4. Particularly, the presence of a material interface is captured by most RB functions. Moreover, all four RB functions behave fairly differently in the left subregion where the problem is transport dominant.

In Figs. 10, 11, we present the first four reduced basis functions of 2D Examples 1 and 2 from Sect. 4.2, again after the SVD orthogonalization step. The leading RB function captures the overall configuration of the density as in Fig. 7, while the remaining RB functions encode various multipole structures.
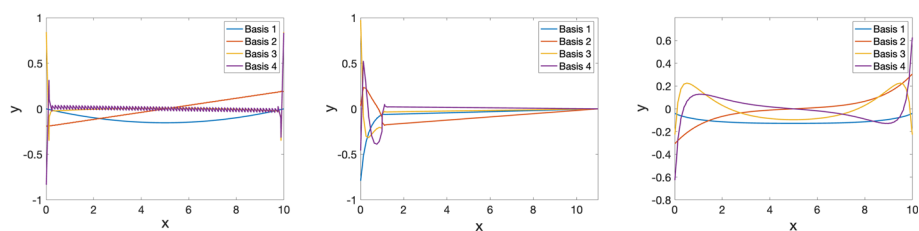
Fig. 9 First four reduced basis functions for 1D Examples 1 (left), 4 (middle), 5 (right)
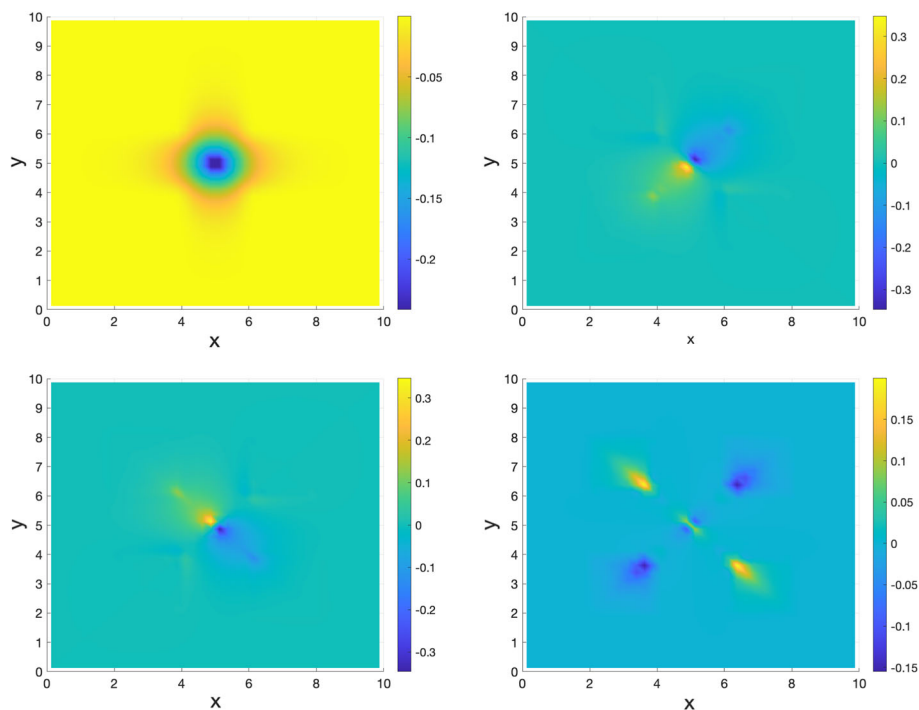

Fig. 10 First four reduced basis functions, ordered from left to right and from top to bottom, for 2D Example 1 (checkerboard problem)
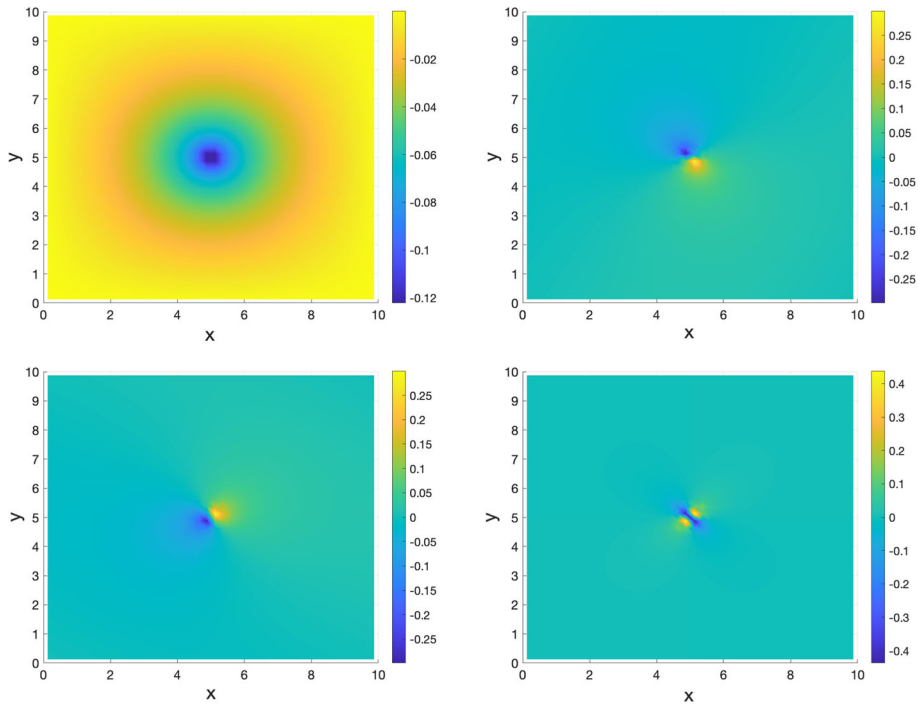
Fig. 11 First four reduced basis, ordered from left to right and from top to bottom, for 2D Example 2 (scattering dominant)

References

1. Adams, M.L.: Discontinuous finite element transport solutions in thick diffusive problems. Nucl. Sci. Eng. **137**, 298–333 (2001)
2. Adams, M.L, Larsen, E.W.: Fast iterative methods for discrete-ordinates particle transport calculations. Prog. Nucl. Energy **40**, 3–159 (2002)
3. Adams, M.L., Martin, W.R.: Diffusion synthetic acceleration of discontinuous finite element transport iterations. Nucl. Sci. Eng. **111**, 145–167 (1992)
4. Alberti, A.L., Palmer, T.S.: Reduced-order modeling of nuclear reactor kinetics using proper generalized decomposition. Nucl. Sci. Eng. **194**, 837–858 (2020)
5. Arridge, S.R., Schotland, J.C.: Optical tomography: forward and inverse problems. Inverse Probl. **25**, 123010 (2009)
6. Bardos, C., Santos, R., Sentis, R.: Diffusion approximation and computation of the critical size. Trans. Am. Math. Soc. **284**, 617–649 (1984)
7. Behne, P., Ragusa, J., Morel, J.: *Model-order reduction for sn radiation transport*. In : ANS International Conference on Mathematics and Computation (M&C). Portland, OR, USA (2019)
8. Binev, P., Cohen, A., Dahmen, W., Devore, R., Petrova, G., Wojtaszczyk, P.: Convergence rates for greedy algorithms in reduced basis methods. SIAM J. Math. Anal, **43**(3), 1457-1472
9. Buchan, A.G., Calloo, A., Goffin, M.G., Dargaville, S., Fang, F., Pain, C.C., Navon, I.M.: A POD reduced order model for resolving angular direction in neutron/photon transport problems. J. Comput. Phys. **296**, 138–157 (2015)
10. Chen, K., Li, Q., Lu, J., Wright, S.J.: Random sampling and efficient algorithms for multiscale pdes. SIAM J. Sci. Comput. **42**, A2974–A3005 (2020)
11. Chen, Y., Gottlieb, S., Ji, L., Maday, Y.: *An EIM-degradation free reduced basis method via over collocation and residual hyper reduction-based error estimation*, arXiv preprint arXiv:2101.05902
12. Chen, Y., Gottlieb, S., Maday, Y.: Parametric analytical preconditioning and its applications to the reduced collocation methods. C. R. Acad. Sci. Paris Ser. I **352**, 661–666 (2014)

13. Chen, Y., Ji, L., Narayan, A., Xu, Z.: L1-based reduced over collocation and hyper reduction for steady state and time-dependent nonlinear equations. J. Sci. Comput. **87**(1), 1–21

14. Chen, Y., Jiang, J., Narayan, A.: A robust error estimator and a residual-free error indicator for reduced basis methods. Comput. & Math. Appl. **77**, 1963–1979 (2019)

15. Choi, Y., Brown, P., Arrighi, W., Anderson, R., Huynh, K.: Space-time reduced order model for large-scale linear dynamical systems with application to boltzmann transport problems. J. Comput. Phys. **424**, 109845 (2020)

16. Coale, J., Anistratov, D.Y.: 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 2019, pp. 278–287 (2019)

17. Ding, Z., Einkemmer, L., Li, Q.: *Error analysis of an asymptotic preserving dynamical low-rank integrator for the multi-scale radiative transfer equation*, arXiv preprint arXiv:1907.04247 (2019)

18. Dominesey, K.A., Ji, W.: *Reduced-order modeling of neutron transport separated in space and angle via proper generalized decomposition*. In: ANS International Conference on Mathematics and Computation (M&C). Portland, OR, USA (2019)

19. Dominesey, K.A., Senecal, J.P.: A reduced-order neutron transport model separated in space and angle. Transactions **119**, 687–690 (2018)

20. Einkemmer, L., Hu, J., Wang, Y.: *An asymptotic-preserving dynamical low-rank method for the multi-scale multi-dimensional linear transport equation*, arXiv preprint arXiv:2005.06571 (2020)

21. Führer, C., Kanschat, G.: A posteriori error control in radiative transfer. Computing **58**, 317–334 (1997)

22. Greif, C., Urban, K.: Decay of the Kolmogorov N-width for wave problems. Appl. Math. Lett. **96**, 216–222 (2019)

23. Guermond, J.-L., Kanschat, G.: Asymptotic analysis of upwind discontinuous Galerkin approximation of the radiative transport equation in the diffusive limit. SIAM J. Numer. Anal. **48**, 53–78 (2010)

24. Haasdonk, B.: Reduced basis methods for parametrized PDEs-a tutorial introduction for stationary and instationary problems. Model Reduct Approx Theory Algorithms **15**, 65 (2017)

25. Hartmann, R., Houston, P.: Adaptive discontinuous galerkin finite element methods for nonlinear hyperbolic conservation laws. SIAM J. Sci. Comput. **24**, 979–1004 (2003)

26. Hesthaven, J.S., Rozza, G., Stamm, B., et al.: Certified Reduced Basis Methods for Parametrized Partial Differential Equations, vol. 590. Springer, Berlin (2016)

27. Jin, S.: Asymptotic preserving (AP) schemes for multiscale kinetic and hyperbolic equations: a review. In: Lecture Notes for Summer School on Methods and Models of Kinetic Theory (M&MKT), Porto Ercole (Grosseto, Italy), pp. 177–216 (2010)

28. Larsen, E.W., Morel, J.E.: Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes ii. J. Comput. Phys. **83**(1), (1989)

29. Larsen, E.W., Morel, J.E., Miller, W.F., Jr.: Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes. J. Comput. Phys. **69**, 283–324 (1987)

30. Lathrop, K.D.: Ray effects in discrete ordinates equations. Nucl. Sci. Eng. **32**, 357–369 (1968)

31. Lewis, E.E., Miller, W.F.: Computational methods of neutron transport. John Wiley & Sons, New York (1984)

32. Liu, Y., Chen, T., Chen, Y., Shu, C.-W.: Certified offline-free reduced basis (cofrb) methods for stochastic differential equations driven by arbitrary types of noise. J. Sci. Comput. **81**(3), 1210–1239 (2019)

33. Lorence, L.J., Jr., Morel, J., Larsen, E.W.: An S_2 synthetic acceleration scheme for the one-dimensional S_n equations with linear discontinuous spatial differencing. Nucl. Sci. Eng. **101**, 341–351 (1989)

34. McClarren, R.G.: Calculating time eigenvalues of the neutron transport equation with dynamic mode decomposition. Nucl. Sci. Eng. **193**, 854–867 (2019)

35. Mihalas, D., Mihalas, B.W.: Foundations of Radiation Hydrodynamics. Courier Corporation, Chelmsford (2013)

36. Naldi, G., Pareschi, L.: Numerical schemes for kinetic equations in diffusive regimes. Appl. Math. Lett. **11**, 29–35 (1998)

37. Ohlberger, M., Rave, S.: *Reduced basis methods: success, limitations and future challenges*, arXiv:1511.02021 (2015)

38. Peng, Z., McClarren, R.G., Frank, M.: A low-rank method for two-dimensional time-dependent radiation transport calculations. J. Comput. Phys. **421**, 109735 (2020)

39. Pinkus, A.: N-widths Approx. Theory. Springer, Berlin (1985)

40. Pomraning, G.C.: The equations of radiation hydrodynamics. In: International Series of Monographs in Natural Philosophy. Pergamon Press, Oxford (1973)

41. Pomraning, G.C.: Equ. Radiat. Hydrodyn. Courier Corporation, Chelmsford (2005)

42. Pringce, Z. Ragusa, J.: *Separated representation of spatial dimensions in sn neutron transport using the proper generalized decomposition*. In: ANS International Conference on Mathematics and Computation (M&C). Portland, OR, USA (2019)
43. Quarteroni, A., Manzoni, A., Negri, F.: Reduced Basis Methods for Partial Differential Equations: an Introduction, vol. 92. Springer, Berlin (2015)
44. Quarteroni, A., Rozza, G., Manzoni, A.: Certified reduced basis approximation for parametrized partial differential equations and applications. J. Math. Ind. **1**, 3 (2011)
45. Santo, N.D., Deparis, S., Manzoni, A., Quarteroni, A.: Multi space reduced basis preconditioners for large-scale parametrized pdes. SIAM J. Sci. Comput. **40**, A954–A983 (2018)
46. Tencer, J., Carlberg, K., Hogan, R., Larsen, M.: Reduced order modeling applied to the discrete ordinates method for radiation heat transfer in participating media. In: ASME 2016 Heat Transfer Summer Conference collocated with the ASME 2016 Fluids Engineering Division Summer Meeting and the ASME 2016 14th International Conference on Nanochannels, Microchannels, and Minichannels, American Society of Mechanical Engineers Digital Collection (2016)
47. Wareing, T.A.: *New Diffusion-Synthetic Accelerations Methods for the SN Equations with Corner Balance Spatial Differencing*, Joint Int. Conf. Mathematical Methods and Supercomputing in Nuclear Applications, (Karlsruhe, Germany), April 19–23, 1993, Vol. 2, p. 500 (1993)
48. Zahm, O., Nouy, A.: Interpolation of inverse operators for preconditioning parameter-dependent equations. SIAM J. Sci. Comput. **38**, A1044–A1074 (2016)

Authors and Affiliations

Zhichao Peng[1] · Yanlai Chen[2] · Yingda Cheng[3] · Fengyan Li[4]

✉ Zhichao Peng
  pengzhic@msu.edu

  Yanlai Chen
  yanlai.chen@umassd.edu

  Yingda Cheng
  ycheng@msu.edu

  Fengyan Li
  lif@rpi.edu

[1] Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA

[2] Department of Mathematics, University of Massachusetts Dartmouth, 285 Old Westport Road, North Dartmouth, MA 02747, USA

[3] Department of Mathematics, Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA

[4] Department of Mathematical Sciences, Rensselaer Polytechnic Institute, Troy, NY 12180, USA