# Machine learning moment closure models for the radiative transfer equation I: directly learning a gradient based closure

Juntao Huang [1] Yingda Cheng [2] Andrew J. Christlieb [3] Luke F. Roberts [4]

**Abstract**

In this paper, we take a data-driven approach and apply machine learning to the moment closure problem for radiative transfer equation in slab geometry. Instead of learning the unclosed high order moment, we propose to directly learn the gradient of the high order moment using neural networks. This new approach is consistent with the exact closure we derive for the free streaming limit and also provides a natural output normalization. A variety of benchmark tests, including the variable scattering problem, the Gaussian source problem with both periodic and reflecting boundaries, and the two-material problem, show both good accuracy and generalizability of our machine learning closure model.

**Key Words: radiative transfer equation; moment closure; slab geometry; machine learning; neural network**

## 1 Introduction

The radiative transfer equation (RTE) describes particle propagation and interaction with a background medium. It has been widely applied in many fields of science and engineering including astrophysics [50], heat transfer [29], remote sensing [56], and medical imaging [28]. The RTE is a high-dimensional integro-differential kinetic equation. Common numerical methods for computing RTE can be classified into two categories: probabilistic methods such as the direct simulation Monte Carlo (DSMC) method [4], and deterministic schemes including the discrete ordinates method ($S_N$)

[1]Department of Mathematics, Michigan State University, East Lansing, MI 48824, USA. E-mail: huangj75@msu.edu

[2]Department of Mathematics, Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, MI 48824, USA. E-mail: ycheng@msu.edu. Research is supported by NSF grants DMS-2011838 and AST-2008004.

[3]Department of Computational Mathematics, Science and Engineering, Michigan State University, East Lansing, Michigan 48824, USA. E-mail: christli@msu.edu. Research is supported by: AFOSR grants FA9550-19-1-0281 and FA9550-17-1-0394; NSF grants DMS-1912183 and AST-2008004; and DoE grant DE-SC0017955.

[4]National Superconducting Cyclotron Laboratory and Department of Physics and Astronomy, Michigan State University, East Lansing, MI 48824, USA. E-mail: robertsl@nscl.msu.edu. Research is supported by: NSF grant AST-2008004; and DoE grant DE-SC0017955.

[34] and the moment method [10, 39] among others. In general, any mesh based numerical discretization faces formidable computational cost due to the curse of dimensionality.

While there have been advances in solving high dimensional models, such as the RTE, using various approaches in dimensional and model reduction [11, 12, 16, 17, 7, 57, 49], for a large class of problems, moment methods are the only tractable solution. Moment methods study the evolution of a finite number of moments of the specific intensity. Typically, in this scenario, the evolution of the $p^{th}$ moment depends on the $(p+1)^{th}$ moment, leading to what is known as the moment closure problem. Hence, one has to introduce suitable closure relations that relates the highest moment with the lower order moments in order to get a closed system of equations. A given closure relation makes assumptions about micro-physics, which may not be true in all settings. Therefore, the trade off in introducing a closure relation and solving a moment model instead of a kinetic equation, such as the RTE, is generic accuracy verses practical computability. Many moment closure strategies have been developed. Some of the best known methods include: the $P_N$ model [10]; the variable Eddington factor models [38, 47]; the entropy-based $M_N$ models [21, 2, 1]; the positive $P_N$ models [20]; the filtered $P_N$ models [45, 32, 51]; the $B_2$ models [3]; and the $MP_N$ model [13, 14, 41]. Newly developed theory around the generic closure problem suggests an approach to constructing analytical closure models based on gradients that lead to globally hyperbolic moment models [41]. This approach encompasses many of the well known closure models and offers insight into this challenging problem.

Recently, thanks to the rapid development of machine learning (ML) [35] and data-driven modeling [6, 52, 18, 60], a new approach to solving the moment closure problem has emerged based on ML. In [19], the authors introduced a framework to construct machine learning moment closure models for kinetic problems. They first learned a set of generalized moments using the auto-encoder to optimally represent the underlying velocity distribution, and then learned the moment closure model for the generalized moments with the aim of best capturing the associated dynamics of the kinetic equation. This framework was further applied to the Williams-Boltzmann equation for polydisperse evaporating sprays in [54]. In [25], based on the conservation-dissipation formalism [61] of irreversible thermodynamics, the authors proposed a stable closure model parametrized by multilayer perceptron (MLP) for the Boltzmann BGK equation. In [5], a nonlocal closure was proposed for the Vlasov-Poisson system using a convolutional neural network (CNN). In [43, 58], the authors applied MLP, CNN and a discrete Fourier transform (DFT) network to learn the well-known Hammett–Perkins Landau fluid closure. In [44], the capability of neural networks to reproduce some known magnetized plasma closures was further investigated. We also note that in addition to the closure problem, ML is being investigated as a method for directly solving high

dimensional kinetic equations. The physics informed neural networks (PINN) was applied to solve forward and inverse problems for kinetic equations including the Boltzmann BGK model [42], the phonon Boltzmann equation [40] and also the RTE [46]. In [59], the full Boltzmann collision operator was approximated by a neural network with the aim of reducing the computational cost.

In this work, we focus on using ML as a tool for model reduction to address the moment closure problem of the RTE. To close the moment model deduced from kinetic equations, the conventional approach is to provide an approximation to the unclosed high order moment. In optically thick regimes or intermediate regimes, it is easy to find an accurate closure relation. However, in optically thin regimes (or even the free streaming limit), the kinetic model does not possess intrinsic low dimensional structure, which makes any attempt at model reduction difficult [36, 37, 27, 53]. To address this problem, we start from investigating the RTE in the free streaming limit and derive the exact closure relations with isotropic initial conditions. Motivated by this closure relation, we propose to directly learn the gradient of the unclosed moment using neural networks for the RTE in slab geometry. The advantages of our approach are twofold. First, the functional form of the model is consistent with the exact closure for the free streaming limit. Thus, it is expected to gain better accuracy using this ansatz, especially in the optically thin regime. Second, the unclosed high order moments usually have a wide range of magnitudes and become very small in the optically thick regime. Such a target function makes the neural network difficult to learn, unless an appropriate output normalization is applied [5]. Our approach in learning gradient provides a natural output normalization since the magnitude of $\partial_x m_{N+1}$ is close to that of $\partial_x m_N$, see equation (2.31) in Section 2.3. In addition, we incorporate the scale invariance of the closure model into the neural networks by learning the normalized gradient. We demonstrate numerically that enforcing scale invariance in the closure model makes the model more generalizable, especially when applied to initial data whose dynamic range is outside the training set used to create our ML closure model.

In our numerical tests, the training data is generated by initial conditions consisting of a truncated Fourier series with random coefficients [43, 5] and constant scattering and absorption coefficients. The well-trained model is uniformly accurate in the optically thick regime, intermediate regime and the optically thin regime. Moreover, the accuracy of our model is much better than the approach based on creating a ML closure directly trained to match the moments, as well as the conventional $P_N$ closure and the filtered $P_N$ closure [51]. This is demonstrated on a wide range of 1D test problems, including the variable scattering problem, the Gaussian source problem, the two-material problem, and the reflective boundary conditions. An important observation is that our ML closure is able to nearly exactly reproduce the moments for the kinetic equation, even for the two-material problem, with a small number ($N = 5$) of moments. The motivation for this

choice is that, as shown below, at minimum four degrees of freedom are required to exactly close the moment equations in the free streaming limit with isotropic initial conditions. Therefore, it is natural to expect that more degrees of freedom are required in the closure for the variable scattering and absorption setting. In our numerical tests, we find that giving our ML closure the freedom to relate the gradient of the sixth-order moment to the gradient of the first six moments is enough to produce accurate results for a variety of different regimes.

Hyperbolicity is another important property in moment closure models, which is difficult to enforce for traditional closure models [9, 41] as well as ML models [25]. Our ML closure model is not able to preserve hyperbolicity. We numerically stabilize the model by adding more numerical diffusion with larger penalty constants in the Lax-Friedrichs numerical flux. How to incorporate hyperbolicity in the ML closure model is certainly an interesting topic. We also remark that we use ML to learn the gradient to close our system, while the contribution in [41] enforces the hyperbolicity with knowledge of the gradients, but does not involve ML. Incorporating the hyperbolicity in the ML closure model is an interesting topic, which is discussed in our subsequent works [24, 23].

The remainder of this paper is organized as follows. In Section 2, we introduce the moment closure problem for the RTE in slab geometry. We derive the exact closure for the free streaming limit with isotropic initial conditions and propose the approach to directly learn the gradient of the unclosed high order moment using ML. In Section 3, we present the details in data generation and the training of the neural networks. The effectiveness of our ML closure model is demonstrated through extensive numerical results in Section 4. Some concluding remarks are given in Section 5.

## 2 Moment closure for radiative transfer equation

In this section, we motivate a range of possible ML models. We start by introducing the moment method for the RTE in slab geometry. Next, we derive the exact closure for the free streaming limit with isotropic initial conditions. The ML closure formulations we produce are inspired by the exact free streaming closure. This includes the approach we propose based on directly learning the gradient of the unclosed high order moments using ML. At the end of this section, we go over the proposed functional forms for the various ML models.

### 2.1 Moment method

The time-dependent RTE for a gray medium in slab geometry has the form:

$$\partial_t f + v\partial_x f = \sigma_s \left( \frac{1}{2} \int_{-1}^{1} f dv - f \right) - \sigma_a f, \tag{2.1}$$

where $f = f(x, v, t)$ is the specific intensity of radiation. The variable $v \in [-1, 1]$ is the cosine of the angle between the photon velocity and the $x$-axis. $\sigma_s = \sigma_s(x) \geq 0$ and $\sigma_a = \sigma_a(x) \geq 0$ are the scattering and absorption coefficients.

It is common to take moments of the RTE against Legendre polynomials. We denote the $k$-th order Legendre polynomial by $P_k = P_k(x)$ for $k \geq 0$. Next, we define the $k$-th order moment of the gray model as

$$m_k(x, t) = \frac{1}{2} \int_{-1}^{1} f(x, v, t) P_k(v) dv, \quad k \geq 0. \tag{2.2}$$

Multiplying (2.1) by $P_k(v)$, then integrating over $v \in [-1, 1]$ and using Bonnet's recursion formula, we derive the moment equations up to moment $m_N$ as

$$\begin{aligned} \partial_t m_0 + \partial_x m_1 &= -\sigma_a m_0, \\ \partial_t m_1 + \frac{1}{3}\partial_x m_0 + \frac{2}{3}\partial_x m_2 &= -(\sigma_s + \sigma_a) m_1, \\ &\cdots \\ \partial_t m_N + \frac{N}{2N+1}\partial_x m_{N-1} + \frac{N+1}{2N+1}\partial_x m_{N+1} &= -(\sigma_s + \sigma_a) m_N. \end{aligned} \tag{2.3}$$

This truncated system is clearly not closed, since in the last equation the evolution of $m_N$ depends on $m_{N+1}$. There are various ways to close the system, including the classical $P_N$ model [10]; the entropy-based $M_N$ models [21, 2, 1]; the variable Eddington factor models [38, 47]; the positive $P_N$ model [20]; the filtered $P_N$ ($FP_N$) models [45, 51, 32]; the $B_2$ models [3]; and the $MP_N$ model [13, 14, 41].

In the numerical tests in Section 4, we will compare our ML closure model with the $P_N$ model [10] and the $FP_N$ model [51]. For the sake of completeness, we present the $P_N$ model and the $FP_N$ model here. The $P_N$ model assumes an ansatz of orthogonal polynomials in velocity space and the closure relation is $m_{N+1} = 0$. Therefore, the resulting $P_N$ model is

$$\partial_t \boldsymbol{m} + A \partial_x \boldsymbol{m} = S \boldsymbol{m}, \tag{2.4}$$

with $\boldsymbol{m} = (m_0, m_1, \cdots, m_N)^T$ and the coefficient matrix $A \in \mathbb{R}^{(N+1)\times(N+1)}$:

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} & 0 & \dots & 0 \\ 0 & \frac{2}{5} & 0 & \frac{3}{5} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \frac{N-1}{2N-1} & 0 & \frac{N}{2N-1} \\ 0 & 0 & \dots & 0 & \frac{N+1}{2N+1} & 0 \end{pmatrix} \tag{2.5}$$

and

$$S = \operatorname{diag}(-\sigma_a, -(\sigma_s + \sigma_a), \cdots, -(\sigma_s + \sigma_a)). \tag{2.6}$$

The $FP_N$ model proposed in [51] reads as

$$\partial_t \boldsymbol{m} + A\partial_x \boldsymbol{m} = S\boldsymbol{m} - \nu L\boldsymbol{m}, \tag{2.7}$$

with $\boldsymbol{m} = (m_0, m_1, \cdots, m_N)^T$. Here, the matrice $A$ and $S$ are the same with the $P_N$ closure. In the additional source term, $\nu > 0$ is a tunable parameter estimating the effective opacity of the filter and

$$L = \text{diag}\,(l_0, l_1, l_2, \cdots, l_N) \tag{2.8}$$

with

$$l_k = \frac{\log \rho(\frac{k}{N+1})}{\log \rho(\frac{N}{N+1})}, \quad k = 0, 1, \cdots, N \tag{2.9}$$

and $\rho$ is the filter function. In our numerical test, we follow [51] and take $\rho(\eta) = \frac{1}{1+\eta^4}$.

All of these methods relate the $(p+1)^{th}$ moment to the $p^{th}$ moment as a way of closing the system. Each of them have a set of pros and cons, but none of them do very well in the optically thin limit. Our proposed ML closures have a functional form that is motivated by a limiting case, which we introduce in the next section.

### 2.2 Exact closure for the free streaming limit

For all moment closure models, it is very challenging to accurately compute the free streaming limit, i.e. when $\sigma_s$ is close to zero in (2.1). In such a transport dominated regime, the model does not possess an intrinsic low dimensional structure, which makes any attempt at model reduction difficult. In this part, we aim to find an exact closure in such a challenging case under some suitable assumptions. This will motivate the functional form of our ML model.

We focus on the free-streaming limit in the simplified case of 1D:

$$\partial_t f + v\partial_x f = 0, \tag{2.10}$$

with an isotropic initial condition

$$f(x, v, 0) = f_0(x), \tag{2.11}$$

which serves as a baseline for our ML closure models. Here, we do not pay attention to boundary conditions, and the solution is considered to be either periodic or compactly supported.

The exact solution to (2.10)-(2.11) is

$$f(x, v, t) = f_0(x - vt). \tag{2.12}$$

Here, instead of using $P_k$, we define the $k$-th order moment by the projection with respect to a monomial basis:

$$n_k(x, t) = \int_{-1}^{1} f(x, v, t) v^k dv, \quad k \geq 0. \tag{2.13}$$

We note that the monomial basis is equivalent to those defined by Legendre polynomials in (2.2). Here, we use the monomial basis moments since it is easier to derive the exact closure relations. In the free-streaming limit, these moments satisfy the equations:

$$\partial_t n_k + \partial_x n_{k+1} = 0, \quad k \geq 0. \tag{2.14}$$

Plugging (2.12) into (2.13), we have

$$\begin{aligned} n_k(x,t) &= \int_{-1}^{1} f_0(x-vt)v^k dv \\ &= \int_{x+t}^{x-t} f_0(w)\left(\frac{x-w}{t}\right)^k \left(-\frac{1}{t}\right) dw \\ &= t^{-(k+1)} \int_{x-t}^{x+t} f_0(w)(x-w)^k dw, \end{aligned} \tag{2.15}$$

which implies

$$t^{k+1} n_k(x,t) = \int_{x-t}^{x+t} f_0(w)(x-w)^k dw. \tag{2.16}$$

Taking a time derivative to the above equation yields

$$t^{k+1}\partial_t n_k(x,t) + (k+1)t^k n_k(x,t) = f_0(x+t)(-t)^k + f_0(x-t)t^k. \tag{2.17}$$

We immediately obtain

$$t\partial_t n_k(x,t) + (k+1)n_k(x,t) = f_0(x+t)(-1)^k + f_0(x-t). \tag{2.18}$$

We notice that the right-hand-side of the above equation only depends on $f_0(x \pm t)$ and the fact that $k$ is even or odd. Thus, we have that: for even $k$,

$$t\partial_t n_k + (k+1)n_k = t\partial_t n_0 + n_0, \tag{2.19}$$

and for odd $k$

$$t\partial_t n_k + (k+1)n_k = t\partial_t n_1 + 2n_1, \tag{2.20}$$

i.e. for any $k \geq 0$,

$$t\partial_t n_k + (k+1)n_k = \frac{1+(-1)^k}{2}(t\partial_t n_0 + n_0) + \frac{1-(-1)^k}{2}(t\partial_t n_1 + 2n_1). \tag{2.21}$$

Next, in the above equation, we replace the time derivatives by the spatial derivatives using (2.14):

$$-t\partial_x n_{k+1} + (k+1)n_k = \frac{1+(-1)^k}{2}(-t\partial_x n_1 + n_0) + \frac{1-(-1)^k}{2}(-t\partial_x n_2 + 2n_1), \tag{2.22}$$

and then derive

$$\begin{aligned}\partial_x n_{k+1} = &\frac{1+(-1)^k}{2}\partial_x n_1 + \frac{1-(-1)^k}{2}\partial_x n_2 \\ &+ \frac{1}{t}\left((k+1)n_k - \frac{1+(-1)^k}{2}n_0 - (1-(-1)^k)n_1\right).\end{aligned} \tag{2.23}$$

Notice that the above relation provides an exact closure for $\partial_x n_{k+1}$. However, the closure has dependence on $t$. Next, we remove the dependence on $t$. Taking $k = 2$ in (2.23) yields

$$\partial_x n_3 = \partial_x n_1 + \frac{1}{t}(3n_2 - n_0), \tag{2.24}$$

then eliminating $t$ in (2.23)

$$\begin{aligned}\partial_x n_{k+1} = &\frac{1+(-1)^k}{2}\partial_x n_1 + \frac{1-(-1)^k}{2}\partial_x n_2 \\ &+ \frac{(k+1)n_k - \frac{1+(-1)^k}{2}n_0 - (1-(-1)^k)n_1}{3n_2 - n_0}\left(\partial_x n_3 - \partial_x n_1\right)\end{aligned} \tag{2.25}$$

This is an exact closure for $\partial_x n_{k+1}$. We remark that this closure holds for any $k \geq 0$ but it reduces to the trivial case for $k = 0, 1, 2$ and only make senses for $k \geq 3$. Since the two sets of moments defined in (2.2) and (2.13) are equivalent, it is easy to derive similar closure relations for $\partial_x m_{k+1}$.

We also point out that the exact closure (2.25) cannot be written into a conservative form:

**Theorem 2.1.** *The exact closure* (2.25) *for* $k \geq 3$ *cannot be written into a conservative form. To be more precise, there exists no smooth function* $F$ *such that* (2.25) *can be written as*

$$n_{k+1} = F(n_0, n_1, \cdots, n_k). \tag{2.26}$$

*Proof.* We prove by contradiction. Assume that there exist a smooth function $F$ such that (2.26) holds true. Using the equality of mixed partial derivatives, we have

$$\frac{\partial}{\partial n_3}\left(\frac{1-(-1)^k}{2}\right) = \frac{\partial}{\partial n_2}\left(\frac{(k+1)n_k - \frac{1+(-1)^k}{2}n_0 - (1-(-1)^k)n_1}{3n_2 - n_0}\right) \tag{2.27}$$

This is equivalent to

$$(k+1)n_k - \frac{1+(-1)^k}{2}n_0 - (1-(-1)^k)n_1 = 0 \tag{2.28}$$

Plugging (2.12)-(2.13) into the above equality, we have

$$\begin{aligned}0 &= (k+1)\int_{-1}^{1} f_0(x-vt)v^k dv - \frac{1+(-1)^k}{2}\int_{-1}^{1} f_0(x-vt)dv - (1-(-1)^k)\int_{-1}^{1} f_0(x-vt)vdv \\ &= \int_{-1}^{1} f_0(x-vt)\left((k+1)v^k - \frac{1+(-1)^k}{2} - (1-(-1)^k)v\right)dv.\end{aligned}$$

For any fixed $k \geq 3$, we can always find $f_0$ such that the above relation does not hold true. □

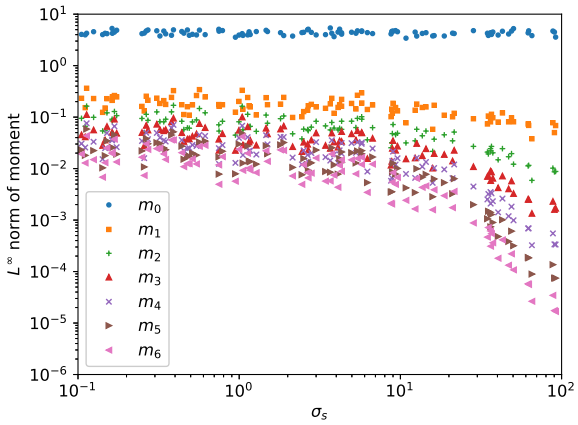



Figure 2.1: Magnitudes of moments $m_k$ for $k = 0, \cdots, 6$ with different scattering coefficients in the training dataset.

## 2.3 Machine learning closure model

In this section, armed with what we just learned about the functional form of moment closures in the free-streaming limit in 1D, we introduce several models we will explore as possible closure models. As part of this work on closing moment based models of the RTE, we explore the effectiveness of each of these models.

A standard approach for moment closures is to find a relation between $m_{N+1}$ and the lower order moments:

$$m_{N+1} = \mathcal{N}(m_0, m_1, \cdots, m_N). \tag{2.29}$$

The ML based approach will find $\mathcal{N} : \mathbb{R}^{N+1} \to \mathbb{R}$ that is represented by a neural network and trained from data. This is the regression in supervised learning and also a part of the end-to-end learning procedure in [19]. We call this approach Learning the Moment (LM). However, in our numerical implementation, we find that the training process usually gets stuck in local minimum when using this approach. Hence, the model has difficulty fitting the data well, see the detailed discussions of Figure 3.3 in Section 3.

We also experiment with another approach, which we refer to as the weighted moment model. The weighted moment model recognizes the fact that equation (2.25) could loosely be viewed as relating many weighted moments as a closure. The Learning Weighted Moment (LWM) model takes the following form:

$$m_{N+1} = \sum_{k=0}^{N} \mathcal{N}_k(m_0, m_1, \ldots, m_N) m_k. \tag{2.30}$$

Here $\mathcal{N} = (\mathcal{N}_0, \ldots, \mathcal{N}_N) : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ will be represented by a neural network. This formulation

can deal with normalization issues. However, as shown in Figure 3.3 in Section 3, its behaviour is similar to that of the simpler Learn Moment model, in (2.29).

Motivated by the exact closure (2.25) for the free streaming limit, we propose to directly learn the gradient of the unclosed moment. Specifically, we assume that $\partial_x m_{N+1}$ depends linearly on the gradients of the lower order moments with the coefficients being functions of the lower order moments:

$$\partial_x m_{N+1} = \sum_{k=0}^{N} \mathcal{N}_k(m_0, m_1, \dots, m_N)\partial_x m_k. \tag{2.31}$$

Here $\mathcal{N} = (\mathcal{N}_0, \dots, \mathcal{N}_N) : \mathbb{R}^{N+1} \to \mathbb{R}^{N+1}$ will be represented by a neural network and trained from data. We call this approach learning the gradient (LG).

The advantages of directly learning the gradient in (2.31) are twofold. First, this ansatz (2.31) is consistent with the exact closure for the free streaming limit (2.25). Therefore, it is expected to have better accuracy, especially in the optically thin regime. Second, the unclosed high order moments usually have a wide range of magnitudes and become very small in the optically thick regime. In Figure 2.1, we show the $L^\infty$-norm (in space and time) of moments $m_k$ for $k = 0, \cdots, 6$ with different scattering coefficients. It is observed that the magnitude of $m_6$ ranges from $10^{-5}$ to $10^{-2}$. Such a target function makes the neural network difficult to learn, unless appropriate output normalization is applied. Similar problems were also noticed in closing the Vlasov-Poisson equation in [5], where the output normalization technique was applied to the heat flux with its estimation given by the Navier-Stokes approximation. Our antasz (2.31) provides a natural output normalization since the magnitude of $\partial_x m_{N+1}$ is similar to that of $\partial_x m_N$.

We further incorporate the scale invariance of the closure model into the neural networks and learn the gradient with normalized moments (LGNM):

$$\partial_x m_{N+1} = \sum_{k=0}^{N} \mathcal{N}_k\left(\frac{m_1}{m_0}, \frac{m_2}{m_0}, \dots, \frac{m_N}{m_0}\right)\partial_x m_k. \tag{2.32}$$

Here $\mathcal{N} = (\mathcal{N}_0, \dots, \mathcal{N}_N) : \mathbb{R}^{N} \to \mathbb{R}^{N+1}$ will be replaced by a neural network and trained from data. Given the linearity of the RTE in (2.1), LGNM in (2.32) is expected to have better performance than LG in (2.31): (1) If $f$ is scaled by a constant in (2.1), the antasz (2.32) can provide exactly the same prediction; (2) LGNM in (2.32) removes the linear redundancy of the training data and should be more data efficient; (3) LGNM in (2.32) does not sacrifice any expressive ability. This results in a ML model with better generalization performance, especially for testing data with a totally different magnitude. This will be further investigated numerically, see Figure 4.15 in Example 4.3 in Section 4.

We also point out a disadvantage of the ML closure model with LG in (2.31) or LGNM in (2.32). This breaks the conservation property of the last equation in (2.3) and generates a partially-conservative system. This might lead to problems in numerical implementations, especially when shocks exists [31, 30].

# 3 Training of the neural network

In this section, we present the details of generating the training data and the training of the neural network. These details are provided for reproduce-ability of the results.

## 3.1 Data preparation

In this section, we outline how we generate the data we use for training our ML model. The data comes from simulating a simple 1D slab RTE problem over a range of initial conditions and scattering and absorption coefficients. The data is curated and used to create $M$ time snapshots of the $N$ moments, $m_{N+1}^{true}$, of the kinetic solution over a fixed time window. This data is used in training our ML model (see section 3.2). We now go over the details for the creation of this data.

We consider the unit interval $[0, 1]$ in the physical domain with periodic boundary conditions. Following [43, 5], we take the initial conditions to be an isotropic distribution in the form of a truncated Fourier series:

$$f_0(x, v) = a_0 + \sum_{k=1}^{k_{\max}} a_k \sin(2k\pi x + \phi_k). \tag{3.1}$$

Here, we take $k_{\max} = 10$ in our dataset. For $k \geq 1$, $a_k$ and $\phi_k$ are random variables sampled from the uniform distributions on $[-\frac{1}{k}, \frac{1}{k}]$ and $[0, 2\pi]$, respectively. We take $a_0 = c + \sum_{k=1}^{k_{\max}} \frac{1}{k}$ with $c$ a random variable sampled from the uniform distributions on $[0, 1]$. This guarantees the positivity of the distribution function. Both $\sigma_s$ and $\sigma_a$ are constants over the domain for each run. The scattering coefficient $\sigma_s$ is sampled from a log-uniform distribution on $[0.1, 100]$. The absorption coefficient $\sigma_a$ are randomly sampled from $[0, 10]$. A possible function generated with (3.1) are shown in Figure 3.2. In the current work, we train with 100 different initial data.

The space time discontinuous Galerkin (DG) method [11, 12] is applied to solve the RTE in slab geometry (2.1). Piecewise polynomials of degree 2 in space and degree 1 in time are applied. We take 64 Gauss-Legendre quadrature points to discretize the velocity space. We take the number of grid points in space to be $N_x = 512$. The CFL condition is taken to be $\Delta t = 8\Delta x$ and the final time is $t = 1$. In this work, we use the $N$ moments of every time step taken for each of the 100 initial conditions to form our ground truth data set.

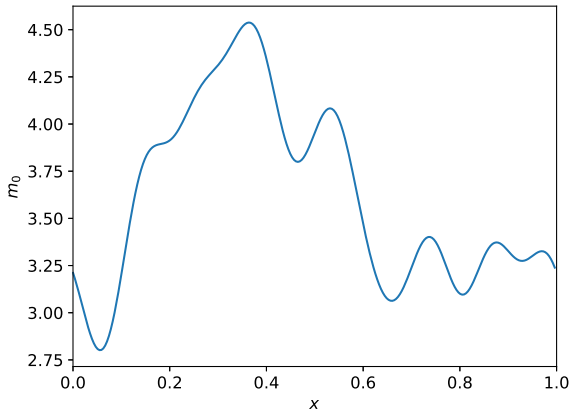



Figure 3.2: A possible initial condition generated from the truncated Fourier series in the training dataset.

### 3.2 Training

For our neural network model, we use a fully-connected neural network and choose to use the hyperbolic tangent for our activation function. The number of layers is taken to be 6 and the number of nodes in each hidden layer is taken to be 256, unless otherwise stated. The input normalization is applied: each component of the input is linearly scaled to have zero mean value and unit variance. For the training of the neural network, we take 1000 total epochs (the number of iterations in the optimization process). The learning rate is set to be $10^{-3}$ in the initial epoch and decays by 0.35 every 100 epochs. $L^2$ regularization is applied with weight $10^{-7}$. The batch size is taken to be 1024. The training is implemented within the PyTorch framework [48]. We refer readers to [22] for the details on the basic concepts of ML.

For the moment model and weighted moment model, (2.29) and (2.30), the loss function is taken to be the mean squared error (MSE):

$$\mathcal{L} = \frac{1}{N_{\text{data}}} \sum_{j,n} \left| m_{N+1}^{\text{true}}(x_j, t_n) - m_{N+1}^{\text{appx}}(x_j, t_n) \right|^2 . \tag{3.2}$$

Here, $m_{N+1}^{\text{true}}(x_j, t_n)$ denote the $(N+1)$-th order moment at $x = x_j$ and $t = t_n$ computed from the kinetic solver and $m_{N+1}^{\text{appx}}(x_j, t_n)$ is the neural network given in equations (2.29) or (2.30), the moment model and weighted moment model respectively. Here, $N_{\text{data}}$ denotes the total count of the data used in training the neural network. For the gradient model and normalized gradient model, (2.31) and (2.32), the loss function is taken to be:

$$\mathcal{L} = \frac{1}{N_{\text{data}}} \sum_{j,n} \left| \partial_x m_{N+1}^{\text{true}}(x_j, t_n) - \partial_x m_{N+1}^{\text{appx}}(x_j, t_n) \right|^2 . \tag{3.3}$$

Here, $\partial_x m_{N+1}^{\text{true}}(x_j, t_n)$ denotes the spatial derivative of $(N+1)$-th order moment at $x = x_j$ and $t = t_n$ computed from the kinetic solver and $\partial_x m_{N+1}^{\text{appx}}(x_j, t_n)$ comes from the evaluation of the neural network using (2.31) or (2.32).

In the training process, we find that the approach to learn the moment in (2.29) and learn the weighted moment in (2.30) usually get stuck in a local minimum and does not fit the data well, no matter how we tune the hyperparameters. As a comparison, the approach to learn the gradient in (2.31) and (2.32) has much smaller relative errors, see Figure 3.3. Moreover, we observe that increasing the number of moments will result in the smaller training error in Figure 3.3. The relative $L^2$ error in Figure 3.3 is defined to be the relative error between the target function and the approximated function. Specifically, for the moment model, (2.29) and (2.30), the relative $L^2$ error for the moment models is defined to be

$$E_2 = \sqrt{\frac{\sum_{j,n}(m_{N+1}^{\text{true}}(x_j, t_n) - m_{N+1}^{\text{appx}}(x_j, t_n))^2}{\sum_{j,n}(m_{N+1}^{\text{true}}(x_j, t_n))^2}}. \tag{3.4}$$

Likewise, for the gradient model, (2.31) and (2.32), the relative $L^2$ error for the gradient model is defined to be

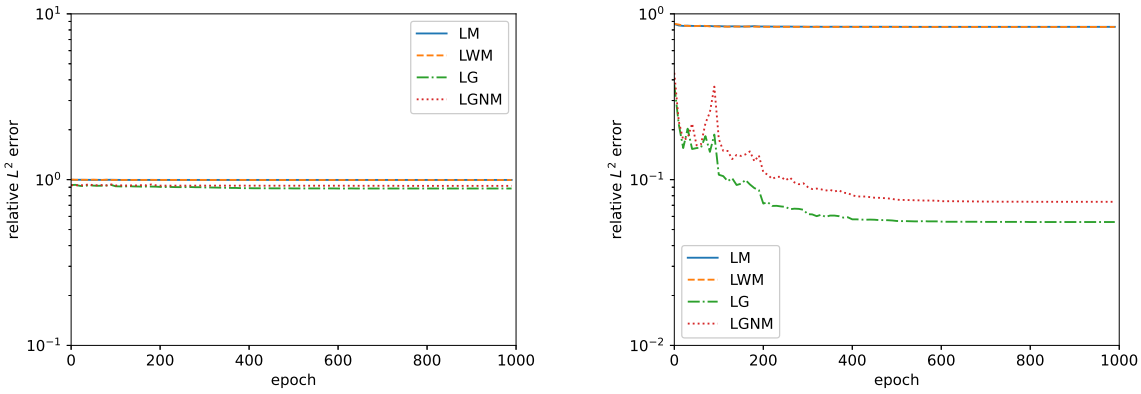
$$E_2 = \sqrt{\frac{\sum_{j,n}(\partial_x m_{N+1}^{\text{true}}(x_j, t_n) - \partial_x m_{N+1}^{\text{appx}}(x_j, t_n))^2}{\sum_{j,n}(\partial_x m_{N+1}^{\text{true}}(x_j, t_n))^2}}. \tag{3.5}$$

The depth and width of neural networks (i.e., the number of hidden layers and the number of nodes in the hidden layers) are also crucial hyperparameters in a neural network. Here, we test the number of layers to be $2, 3, \cdots, 7$ and number of nodes in hidden layers to be $8, 16, \cdots, 1024$. The results are shown in Figure 3.4. The error decreases when we increase the number of layers and nodes in hidden layers and saturate when they reach a certain level. These tests indicate that taking number of layers to be 6 and number of nodes to be 256 are good hyperparameters for our neural network. As such these are the values used in this work unless otherwise stated.

# 4 Numerical tests

In this section, we show the performance of our ML closure model on a variety of benchmark tests, including problems with constant scattering and absorption coefficients, variable scattering problems, Gaussian source problems and two-material problems. In all the numerical examples, we take the physical domain to be the unit interval $[0, 1]$. We consider periodic boundary conditions and reflective boundary conditions.

To numerically solve the ML moment closure system, we apply the fifth-order finite difference WENO scheme [26] with a Lax–Friedrichs flux splitting for the spatial discretization. For the time discretization, we employ the third order strong-stability-preserving Runge-Kutta (RK) scheme




(a) $N = 1$ (b) $N = 3$

(c) $N = 5$ (d) $N = 7$

Figure 3.3: Relative $L^2$ error of the target function during the training process. Four approaches are compared here: learning the moment (LM) in (2.29); learning the weighted moment (LWM) in (2.30); learning the gradient (LG) in (2.31); learning the gradient with normalized moments (LGNM) in (2.32).

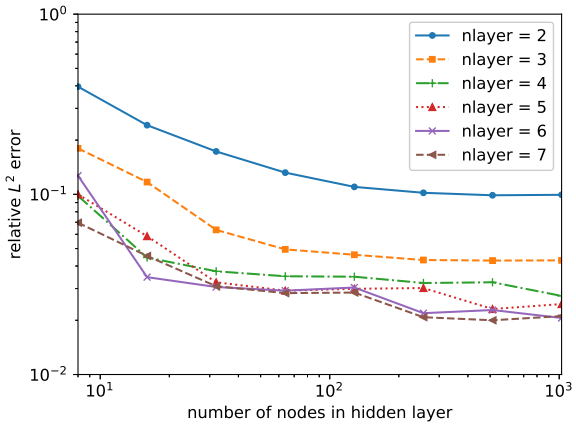



Figure 3.4: Relative $L^2$ error in the training data with different depths and widths of the neural networks. The number of layers: $2, 3, \ldots, 7$; the number of nodes in the hidden layers: $8, 16, \ldots, 1024$. The number of moments in the system is $N = 5$.

[55]. We take the grid number in space to be $N_x = 256$. The CFL condition is taken to be $\Delta t = 0.1\Delta x$. The penalty constant in the Lax-Friedrichs numerical flux is taken to be $\alpha_{\text{LF}} = 5$, unless otherwise stated.

We mainly focus on the comparison of the ML closure based on Learning the Gradient with Normalized Moments (LGNM) (2.32), ML closure based on Learning the Moments (LM) (2.29), and the $P_N$ closure [10]. In addition, we consider the $FP_N$ closure [51] and make comparisons with the other three models in the context of modeling transport on bounded domains with reflecting walls in Example 4.3 and in the study of the two-material problem in Example 4.5. We note that, in most cases, the ML model based on LGNM (2.32) has better accuracy than the Learn Gradient (LG) model (2.31). In particular, the benefit of using (2.32) will be illustrated using the Gaussian source problem in Example 4.3.

An important observation is that our ML closure is able to nearly exactly reproduce the moments for the kinetic equation, even for the two-material problem, with a small number ($N = 5$) of moments. In our numerical tests, we find that giving our ML closure the freedom to relate the gradient of the sixth-order moment to the gradient of the $0^{th}$ through $5^{th}$ moments is enough to produce accurate results for different regimes. Moreover, we numerically observe that if we use fewer than $N = 5$ moments in our model, where the ML closure would relate the gradient of the highest moment to the gradient of the lower moments, the method can still produce accurate results in the intermediate regime, but struggles to describe the solution near the optically thin regime.

**Example 4.1** (constant scattering and absorption coefficients)**.** The setup of this example is the same as the data preparation. The scattering and absorption coefficients are taken to be constants

over the domain.

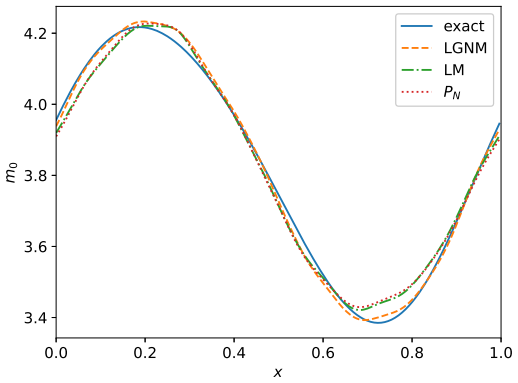
We test three different regimes: the optically thick regime ($\sigma_s = \sigma_t = 100$); the intermediate regime ($\sigma_s = \sigma_t = 10$); and the optically thin regime ($\sigma_s = \sigma_t = 1$). All closures work well in the optically thick regime and thus we omit the results and only focus on the intermediate and optically thin regimes.

In Figure 4.5, we show the numerical solutions of $m_0$ and $m_1$ at $t = 0.5$ with two moments ($N = 1$) in the system and in the closure. We observe that, in the intermediate regime ($\sigma_s = \sigma_t = 10$), the solution generated by the closure model with LGNM agrees well with the solution to the kinetic model. As a comparison, there exist some deviations for the other two closure models, the LM model and the $P_N$ model. However, in the optically thin regime ($\sigma_s = \sigma_t = 1$), all closures fail to capture the correct physical phenomenon for the kinetic model. This indicates that taking only two moments is not enough to close the kinetic equation in this regime. This is also consistent with what we discovered in Section 2.2, that is, as we move to the free streaming limit, the closure is related to the gradients of many moments.

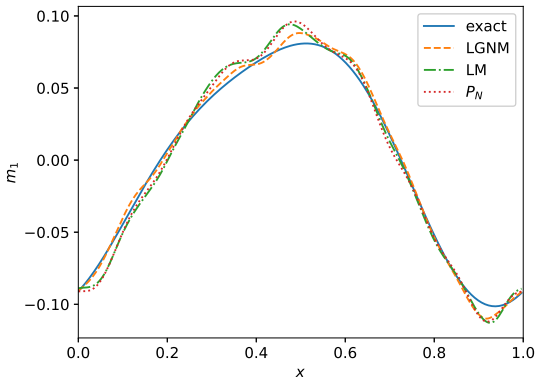
In Figure 4.6, we show the numerical solutions of $m_0$ and $m_1$ at $t = 0.5$ with $N = 5$. In the intermediate regime, all the closures predict the solution quite well. Moreover, it is observed that the ML closure model based on the LGNM formulation has the smallest error (see the zoomed-in figure). In the optically thin regime, only the closure model based on the LGNM formulation agrees well with the kinetic equation, while the other two closures have large deviations in the moments.

In Figure 4.7 and Figure 4.8, we display the log–log scatter plots of the relative $L^2$ error versus the scattering coefficient for $N = 1$ and $N = 5$, respectively. In the case of $N = 1$, we observe that building a closure model based on the LG and LGNM formulations does not always result in smaller errors than the $P_N$ closure, see Figure 4.7. This indicates that it seems impossible to find a model with only two moments that will be able to approximate the moments accurately in the optically thin case. In the case of $N = 5$, the LG and LGNM models have a much smaller error than the LM model and the $P_N$ model, especially for smaller scattering coefficients, see Figure 4.8. Moreover, we notice that, in most cases, the model based on LGNM has better accuracy than the LG model.

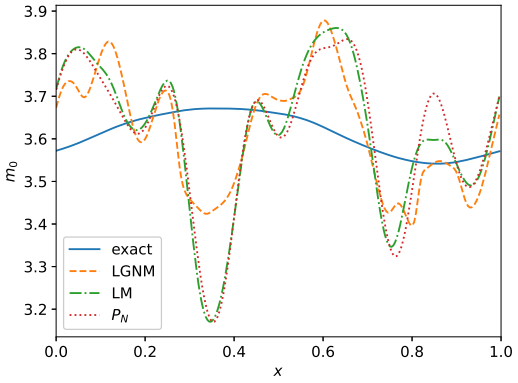
Similar to many analytical closure models, our ML closure models do not guarantee hyperbolicity. We now use this example to investigate the numerical stability of our closure model. We take the penalty constant in the Lax-Friedrichs flux to be $\alpha_{\text{LF}} = 2$. In Figure 4.9, we output the number of grid points with imaginary eigenvalues and $L^\infty$-norm of the numerical solution during the time evolution. We find that this model does not preserve the hyperbolicity property and the numerical solution starts to blow up at around $t = 0.27$. We then increase the penalty constant to $\alpha_{\text{LF}} = 5$.



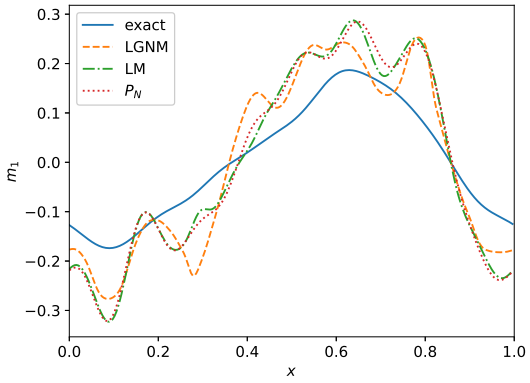

(a) $m_0$, $\sigma_s = \sigma_t = 10$




(b) $m_1$, $\sigma_s = \sigma_t = 10$

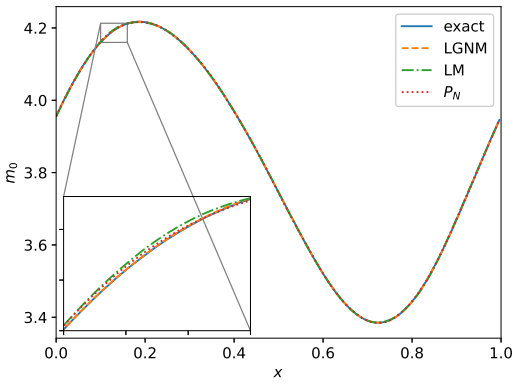



(c) $m_0$, $\sigma_s = \sigma_t = 1$

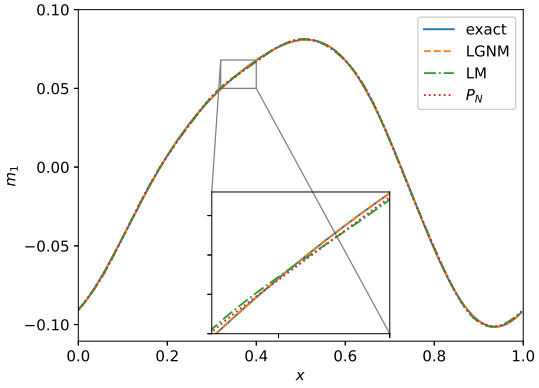



(d) $m_1$, $\sigma_s = \sigma_t = 1$

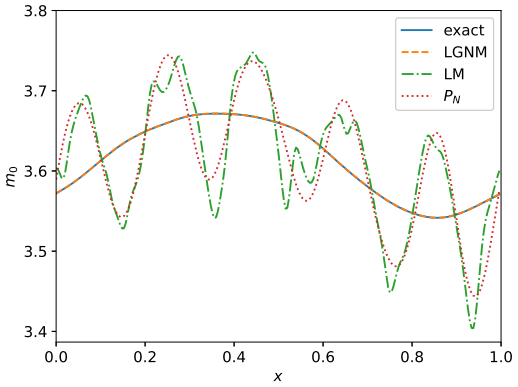
Figure 4.5: Example 4.1: constant scattering and absorption coefficients. Here we are plotting the numerical solutions of $m_0$ and $m_1$ for three moment closures, including $P_N$, Learning Moment (LM) and Learning Gradient with Normalized Moments (LGNM), at $t = 0.5$ with $N = 1$ in the intermediate regime ($\sigma_s = \sigma_t = 10$) and the optically thin regime ($\sigma_s = \sigma_t = 1$). We note that the closure based on the LGNM performs slightly better than the other two methods for the $N = 1$ case.



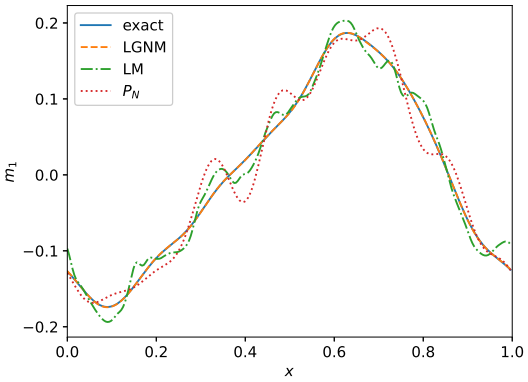

(a) $m_0$, $\sigma_s = \sigma_t = 10$




(b) $m_1$, $\sigma_s = \sigma_t = 10$

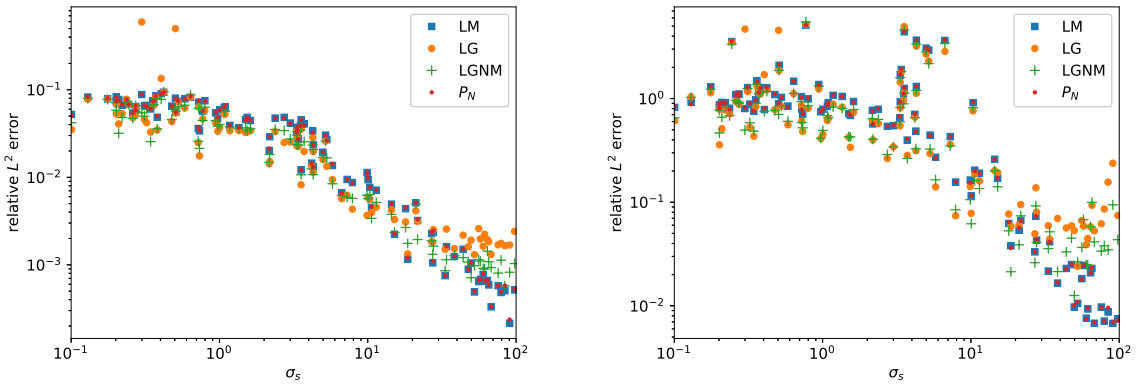



(c) $m_0$, $\sigma_s = \sigma_t = 1$



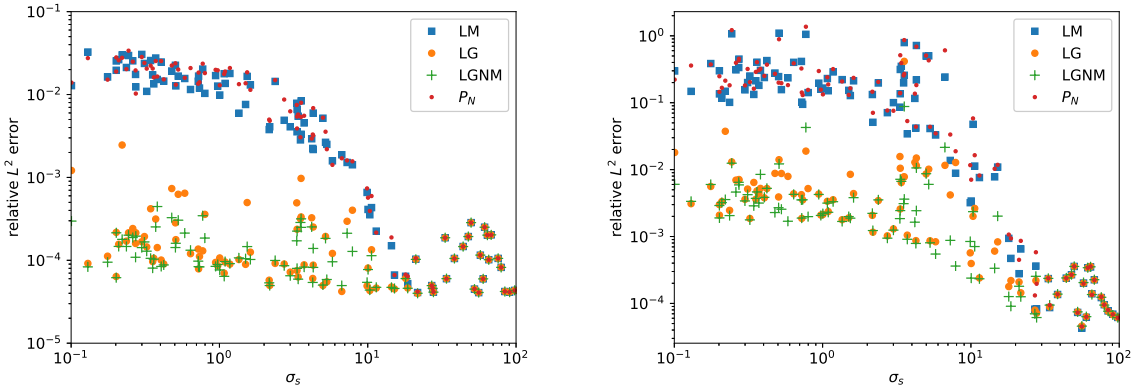

(d) $m_1$, $\sigma_s = \sigma_t = 1$

Figure 4.6: Example 4.1: constant scattering and absorption coefficients. Here we are plotting the numerical solutions of $m_0$ and $m_1$ for three moment closures, including $P_N$, Learning Moment (LM) and Learning Gradient with Normalized Moments (LGNM), at $t = 0.5$ with $N = 5$ in the intermediate regime ($\sigma_s = \sigma_t = 10$) and the optically thin regime ($\sigma_s = \sigma_t = 1$). We note that the closure based on the LGNM performs better than the other two methods for the $N = 5$ case.



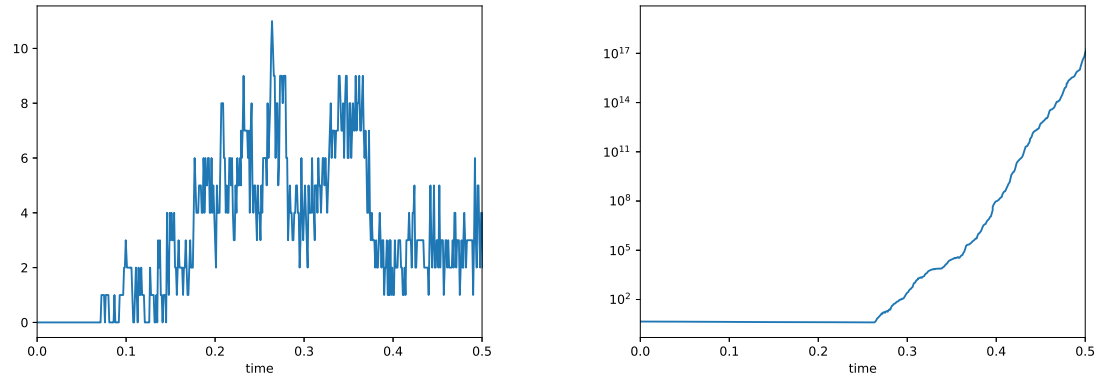

(a) $m_0$ (b) $m_1$

Figure 4.7: Example 4.1: constant scattering and absorption coefficients. In this figure we are plotting the relative $L^2$ error of $m_0$ and $m_1$ with different scattering coefficient at $t = 0.5$ with $N = 1$. All methods perform about the same for $N = 1$.




(a) $m_0$ (b) $m_1$

Figure 4.8: Example 4.1: constant scattering and absorption coefficients. The figure plots the relative $L^2$ error of $m_0$ and $m_1$ with different scattering coefficient at $t = 0.5$ with $N = 5$. Here Learning Gradient (LG) and Learn Gradient with Normalized Moments (LGNM) methods perform better than Learning Moment (LM) and the $P_N$ closure.




(a) number of grid points with imaginary eigenvalues

(b) $L^\infty$ norm of numerical solutions

Figure 4.9: Example 4.1: constant scattering and absorption coefficients. The number of grid points with imaginary eigenvalues and $L^\infty$ norm of numerical solutions during the time evolution in the optically thin regime ($\sigma_s = \sigma_t = 1$) with $N = 5$ and $\alpha_{\text{LF}} = 2$.

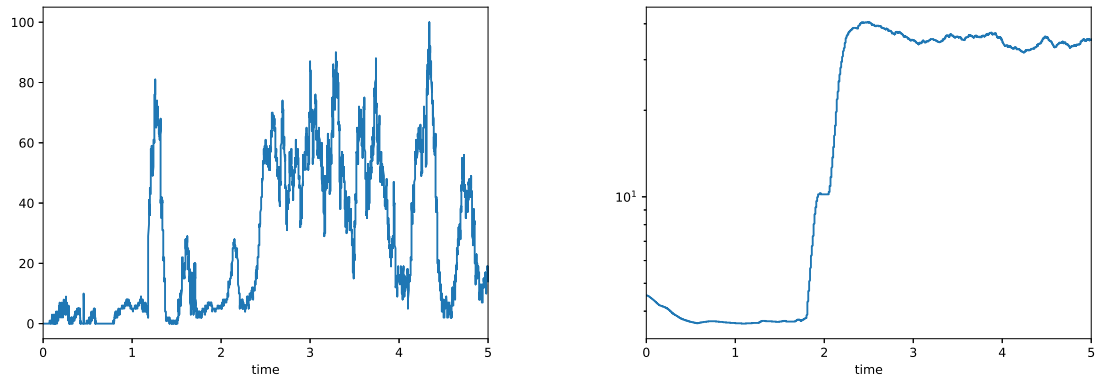
As shown in Figure 4.10, this helps stabilize the model numerically. Even though there are still a large number of grid points with imaginary eigenvalues during the time evolution, the numerical solution does not blow up. We conclude that incorporating hyperbolicity in the ML closure model is an important topic, which is discussed in a subsequent work [24, 23].

**Example 4.2** (variable scattering problem)**.** In this example, we investigate the performance and generalizability of our closure models by testing them on problems that have spatially varying scattering coefficients. In these tests, the scattering coefficient is taken to have the following form

$$\sigma_s(x) = c_1(\tanh(1 + c_2(x - x_0)) + \tanh(1 - c_2(x - x_0))) + \sigma_{s,\text{base}}, \tag{4.1}$$

with $c_1$, $c_2$, $\sigma_{s,\text{base}}$ and $x_0$ being constants. Here, the parameters are taken to be $c_1 = c_2 = 15$, $x_0 = 0.5$ and $\sigma_a = 1$. We test two cases with $\sigma_{s,\text{base}} = 1$ and $\sigma_{s,\text{base}} = 10$. The profiles of scattering coefficients with $\sigma_{s,\text{base}} = 1$ and $\sigma_{s,\text{base}} = 10$ are shown in Figure 4.11. It is observed that, in the case of $\sigma_{s,\text{base}} = 1$, the middle part of the domain is in the intermediate regime and the domain near the boundary is in the optically thin regime. On the other hand, the whole domain is in the intermediate regime when $\sigma_{s,\text{base}} = 10$.

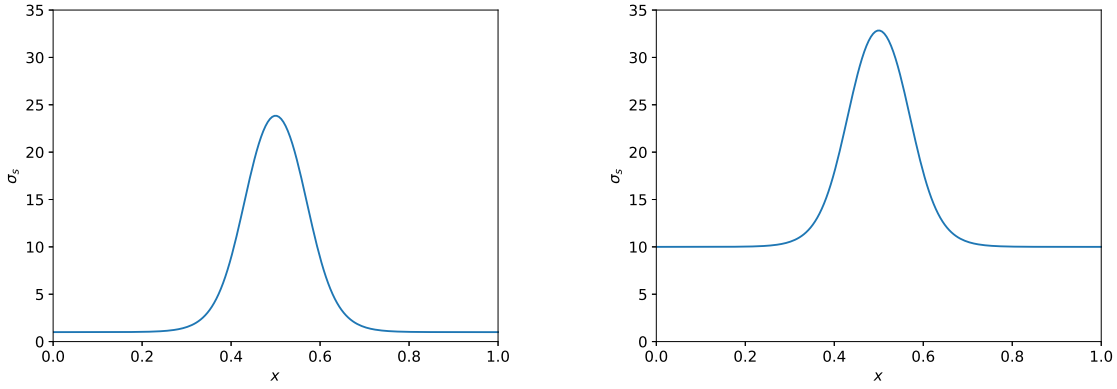
The profiles of numerical solutions with $\sigma_{s,\text{base}} = 1$ and $\sigma_{s,\text{base}} = 10$ are presented in Figure 4.12 and Figure 4.13, respectively. In the test case that spans the optically thin to intermediate regime, we observe good agreement between our LGNM closure model and the kinetic model, see Figure 4.12, while the other two closure models do not obtain satisfactory solution profiles. In Figure 4.13, we plot the results when $\sigma_{s,\text{base}} = 10$. In this case, all closure models work well. We note that our LGNM model still has the smallest error.




(a) number of grid points with imaginary eigenvalues

(b) $L^\infty$ norm of numerical solutions

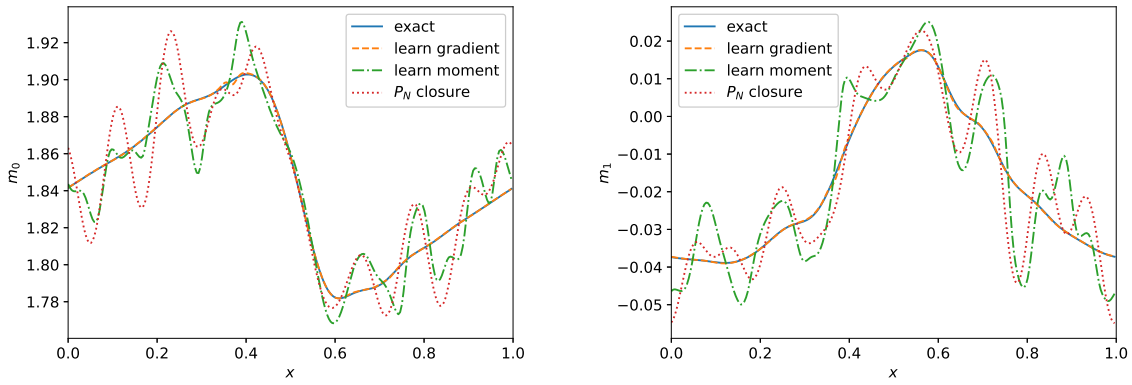
Figure 4.10: Example 4.1: constant scattering and absorption coefficients. The number of grid points with imaginary eigenvalues and $L^\infty$ norm of numerical solutions during the time evolution in the optically thin regime ($\sigma_s = \sigma_t = 1$) with $N = 5$ and $\alpha_{\text{LF}} = 5$.




(a) $\sigma_{s,\text{base}} = 1$

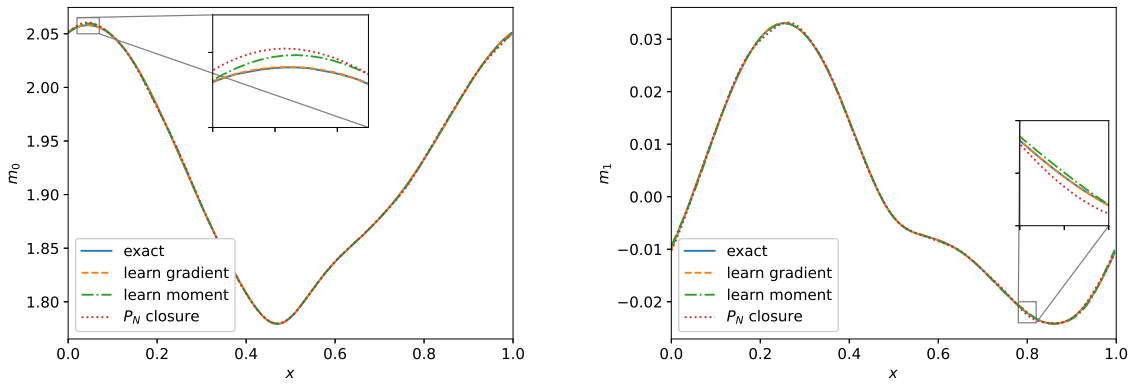
(b) $\sigma_{s,\text{base}} = 10$

Figure 4.11: Example 4.2: variable scattering problem. Plots of the profiles of scattering coefficient functions given in equation (4.1), with $\sigma_{s,\text{base}} = 1$ and $\sigma_{s,\text{base}} = 10$ respectively.



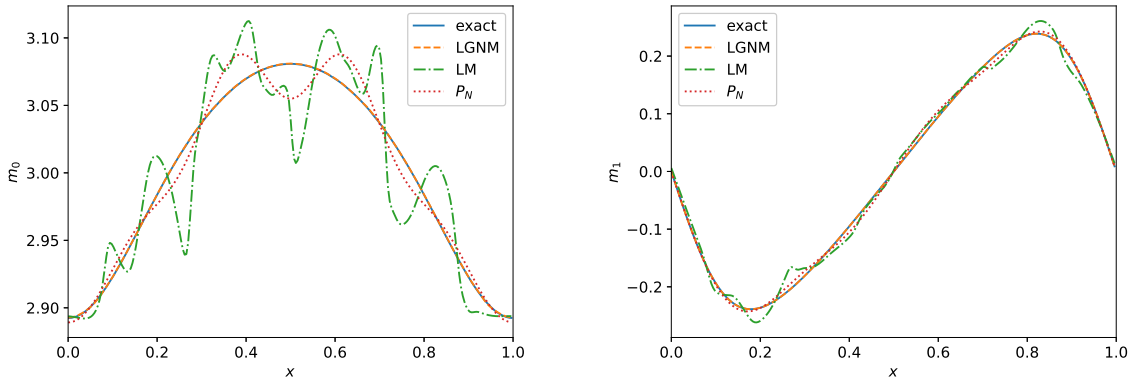

(a) $m_0$ (b) $m_1$

Figure 4.12: Example 4.2: variable scattering problem. Numerical solutions of $m_0$ and $m_1$ at $t = 0.5$ with $N = 5$ and $\sigma_{s,\text{base}} = 1$.




(a) $m_0$ (b) $m_1$

Figure 4.13: Example 4.2: variable scattering problem. Numerical solutions of $m_0$ and $m_1$ at $t = 0.5$ with $N = 5$ and $\sigma_{s,\text{base}} = 10$.




(a) $m_0$ (b) $m_1$

Figure 4.14: Example 4.3: Gaussian source problem. Numerical solutions of $m_0$ and $m_1$ at $t = 0.5$ with $N = 5$ in the optically thin regime ($\sigma_s = 1$).

**Example 4.3** (Gaussian source problem)**.** In this example, we investigate the Gaussian source problem, which simulates particles with an initial intensity that is a Gaussian distribution in space [15, 13]

$$f_0(x, v) = \frac{c_1}{(2\pi\theta)^{1/2}} \exp\left(-\frac{(x - x_0)^2}{2\theta}\right) + c_2. \tag{4.2}$$

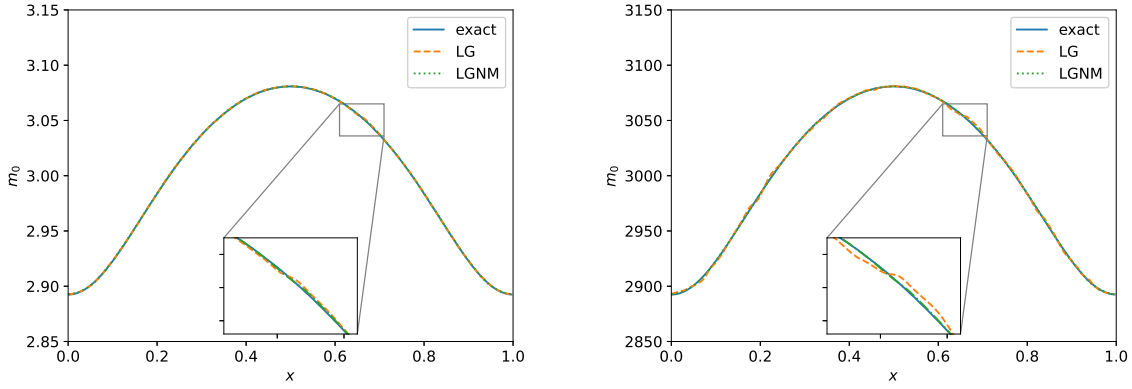
In this test, we take $c_1 = 0.5$, $c_2 = 2.5$, $x_0 = 0.5$ and $\theta = 0.01$.

In Figure 4.14, we present the results obtained using various closure models. We observe good agreement between the LGNM closure model and the kinetic model, while the other two closure models have large deviations from the kinetic model. This illustrates that the ML closure model exhibits good generalization to other types of initial conditions beyond the training data.

From (3.1), it is easy to see that the upper bound of the initial conditions in our training dataset is $1 + 2\sum_{k=1}^{k_{\max}} \frac{1}{k} \approx 6.86$. To investigate the generalization of our model to problems that have a magnitude that is outside of the range of training data, we amplify the initial condition (4.2) by 1000 times. That is, in this test we take the initial condition to be

$$f_0(x, v) = 1000 \left( \frac{c_1}{(2\pi\theta)^{1/2}} \exp\left(-\frac{(x - x_0)^2}{2\theta}\right) + c_2 \right) \tag{4.3}$$

with the same parameters $c_1$, $c_2$, $x_0$ and $\theta$ as in (4.2). For the original initial condition (4.2), the ML closure model based on the LG and LGNM both predict the solution well, see Figure 4.15 (a). The relative $L^2$ errors are $1.05 \times 10^{-4}$ and $6.42 \times 10^{-5}$ for LG and LGNM approaches respectively. When we amplify the initial condition by 1000 times, the two models still generate good predictions, see Figure 4.15 (b). The relative $L^2$ errors are $3.89 \times 10^{-4}$ and $6.38 \times 10^{-5}$ for LG and LGNM approaches respectively. This implies that the model without the scale invariance




(a) original initial condition

(b) original initial condition × 1000

Figure 4.15: Example 4.3: Gaussian source problem. Numerical solutions of $m_0$ and $m_1$ at $t = 0.5$ with $N = 5$ in the optically thin regime ($\sigma_s = 1$). Left: results from the original initial condition, equation (4.2); right: results from the original initial condition × 1000, equation (4.3).

constraint automatically learns this invariance property from data. Nevertheless, enforcing the scale invariance property will lead to better results when the magnitude of the problem initial conditions is very different from that used in the training dataset.

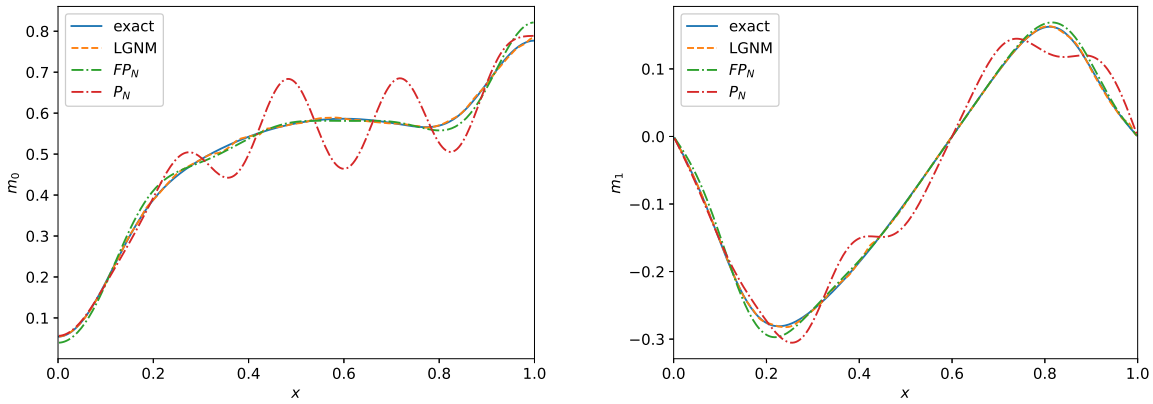
We also test the performance of the ML closure model with non-periodic boundary conditions. In the initial Gaussian distribution (4.2), we take $c_1 = 0.5$, $c_2 = 10^{-6}$, $x_0 = 0.6$ and $\theta = 0.005$. The reflective boundary conditions are imposed on both the left and right boundaries. Assume that the grid points in the interior domain are $x_j = (j + \frac{1}{2})\Delta x$ with $j = 0, \cdots, N_x - 1$. To treat the reflective boundary conditions for the moment closure system, we set several ghost points outside of the domain: $x_j = (j + \frac{1}{2})\Delta x$ with $j = -3, -2, -1$ on the left and $j = N_x, N_x + 1, N_x + 2$ on the right. In each time step, the moments at the ghost points are updated by

$$m_k(x_j, t_n) = (-1)^k m_k(x_{-j-1}, t_n), \quad j = -3, -2, -1 \tag{4.4}$$

and

$$m_k(x_j, t_n) = (-1)^k m_k(x_{2N_x-j-1}, t_n), \quad j = N_x, N_x + 1, N_x + 2. \tag{4.5}$$

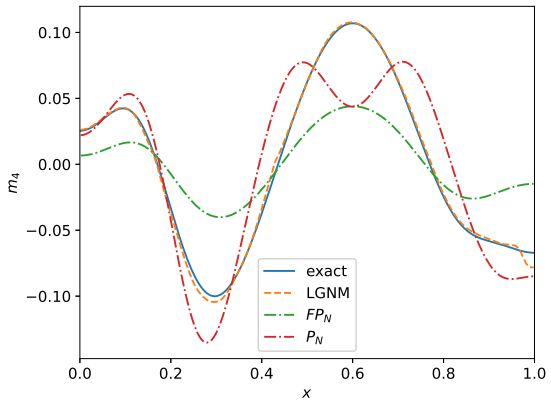
The numerical results with $N = 5$ in the optically thin regime ($\sigma_s = \sigma_t = 1$) are shown in Figure 4.16. We observe that the $P_N$ closure has large deviations from the exact solution to the kinetic model. The $FP_N$ closure is more accurate than $P_N$ in $m_0$ and $m_1$, but has relatively larger error compared to that of the LGNM model, especially near the boundaries. Moreover, the LGNM model predicts the higher order moments $m_4$ and $m_5$ much more accurately than both the $P_N$ and $FP_N$ models.



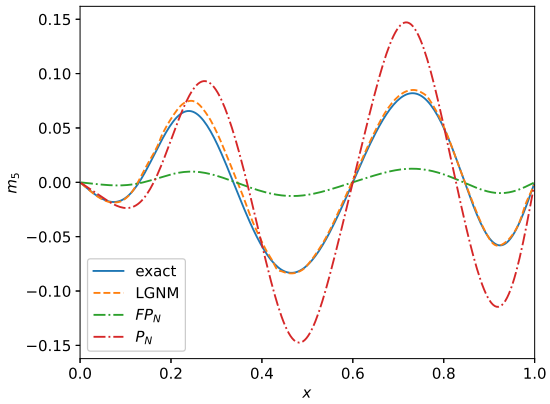

(a) $m_0$




(b) $m_1$




(c) $m_4$




(d) $m_5$

Figure 4.16: Example 4.3: Gaussian source problem with reflective boundary conditions. Numerical solutions of $m_0$, $m_1$, $m_4$ and $m_5$ at $t = 0.5$ with $N = 5$ in the optically thin regime ($\sigma_s = \sigma_t = 1$).

We also remark that there exist a lot of work on inflow boundary conditions (including vacuum boundary condition as a special case), see e.g. [21, 8, 14]. This remains an open problem, even for analytical moment models. This is not the focus of the current paper and we leave it to our future work.

**Example 4.4** (higher wave number test)**.** To further test the generalizability of our ML closure model, we consider initial conditions across a range of wave numbers that extends outside of the wave numbers present in the training data:

$$f_0(x, v) = 2 + \sin(2\pi k x + \phi). \tag{4.6}$$

Here $k$ is the wave number of the initial data and $\phi$ is a random number from $[0, 2\pi)$. In the test, we take $k = 1, 2, \cdots, 25$. This is a challenging test, since the magnitudes in the training data in (3.1) decay with the wave number, so it is naturally difficult for the well-trained model to capture the correct behavior for the initial condition in (4.6) with non-decaying magnitudes. We test the optically thin regime, i.e., $\sigma_s = \sigma_t = 1$, since the intermediate regime and the optically thick regime are relatively easy to capture for the moment closure model. We run the simulations to time $t = 0.4$.

Figure 4.17 shows the relative $L^2$ error of $m_0$ when using our LGNM closure model with $N = 5$. Overall, the method performs well for these waves that are outside of the training data. For low wave numbers ($1 \le k \le 10$), the error stays in the magnitude of $10^{-4}$ to $10^{-3}$. For high wave numbers ($11 \le k \le 25$), the error slightly increases and saturates in the magnitude of $10^{-3}$ to $10^{-2}$. This shows that our ML closure model generalizes well when the initial conditions are outside of the training data.

**Example 4.5** (two-material problem)**.** The two-material problem models a domain with a discontinuous material cross sections [33]. In our problem setup, there exist two discontinuities $0 < x_1 < x_2 < 1$ in the domain, and $\sigma_s$ and $\sigma_a$ are piecewise constant functions:

$$\sigma_s(x) = \begin{cases} \sigma_{s1}, & x_1 < x < x_2, \\ \sigma_{s2}, & 0 \le x < x_1 \text{ or } x_2 \le x < 1. \end{cases}$$

and

$$\sigma_a(x) = \begin{cases} \sigma_{a1}, & x_1 < x < x_2, \\ \sigma_{a2}, & 0 \le x < x_1 \text{ or } x_2 \le x < 1. \end{cases}$$

Specifically, we take $x_1 = 0.3$, $x_2 = 0.7$, $\sigma_{s1} = 1$, $\sigma_{s2} = 10$ and $\sigma_{a1} = \sigma_{a2} = 0$.

In Figure 4.18 and Figure 4.19, we compare our LGNM moment closure with the $P_N$ closure and the $FP_N$ closure for $N = 5$ and $N = 9$, respectively. The gray background region is in the

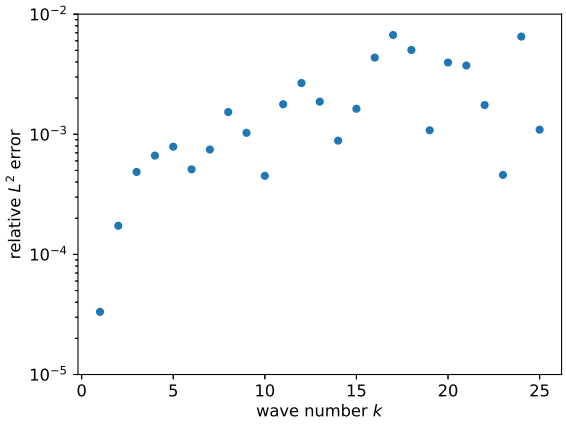
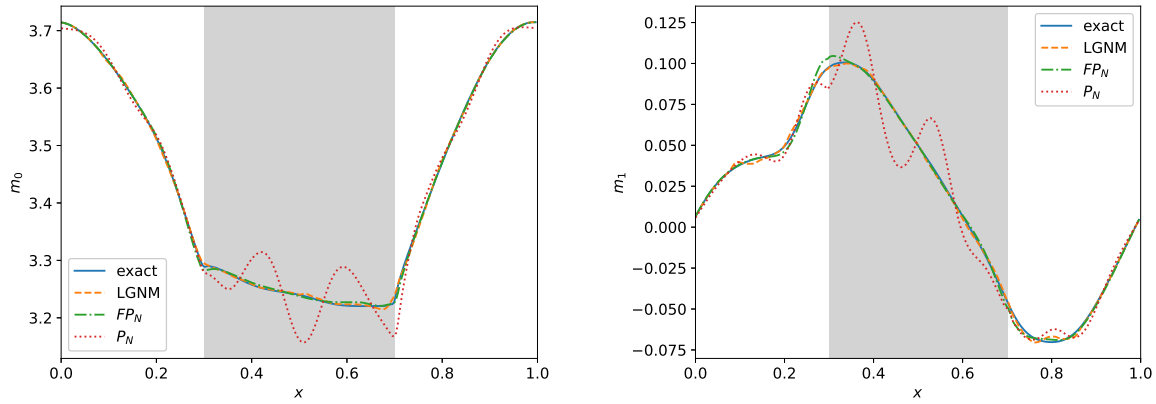


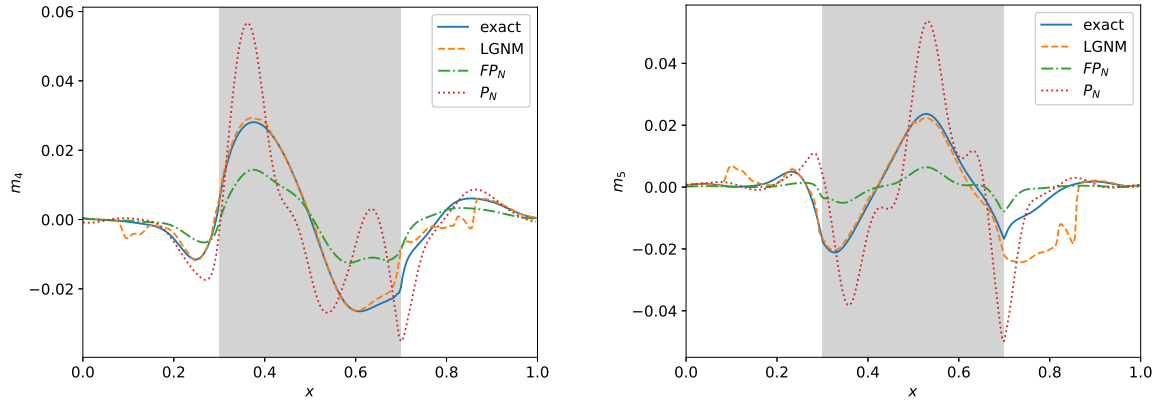

Figure 4.17: Example 4.4: higher wave number test. relative $L^2$ error of $m_0$ at $t = 0.4$ and optically thin regime $\sigma_s = \sigma_t = 1$. The number of moments in the system is $N = 5$.

optically thin regime and the white background region is in the intermediate regime. Here, we tune the parameter $\nu$ in the $FP_N$ closure (2.7) in the range of $[1, 100]$ and find that $\nu = 20$ is the optimal parameter for the best performance of $m_0$ in $FP_N$ closure. For both $N = 5$ and $N = 9$, we observe that the $P_N$ closure has large deviations from the kinetic solution in the optically thin portion of the domain, see the gray part in Figure 4.18 and Figure 4.19. As a comparison, in the ML closure and the $FP_N$ closure, the low order moments $m_0$ and $m_1$ agree well with the kinetic model, see Figure 4.18 (a) and Figure 4.18 (b) for $N = 5$ and Figure 4.19 (a) and Figure 4.19 (b) for $N = 9$. For the high order moments, the ML closure behaves slightly better than the $FP_N$ closure in $m_4$ and $m_5$ for $N = 5$, see Figure 4.18 (c) and Figure 4.18 (d). Moreover, for $N = 9$, the ML closure behaves much better than the $FP_N$ closure in the high order moments $m_4$, $m_5$, $m_8$ and $m_9$. This illustrates the potential for better performance of our ML closure over the traditional approaches including the $P_N$ and $FP_N$ closures, especially for the case of capturing the higher order moments.

We further investigate the performance of the ML closure quantitatively. In Table 4.5, we show the relative $L^2$ errors of $m_0$ and $m_N$ moments at $t = 0.4$ with $N = 5, 6, 7, 8, 9$. We first note that all three closure models show convergence with an increasing number of moments. Second, we note that for the zeroth order moment $m_0$, the $FP_N$ closure and the ML closure have similar performance, and both are more accurate than $P_N$ closure. Third, we note that for high order moments, $m_N$, the ML closure is much more accurate than both the $P_N$ and $FP_N$ closures.




(a) $m_0$ (b) $m_1$

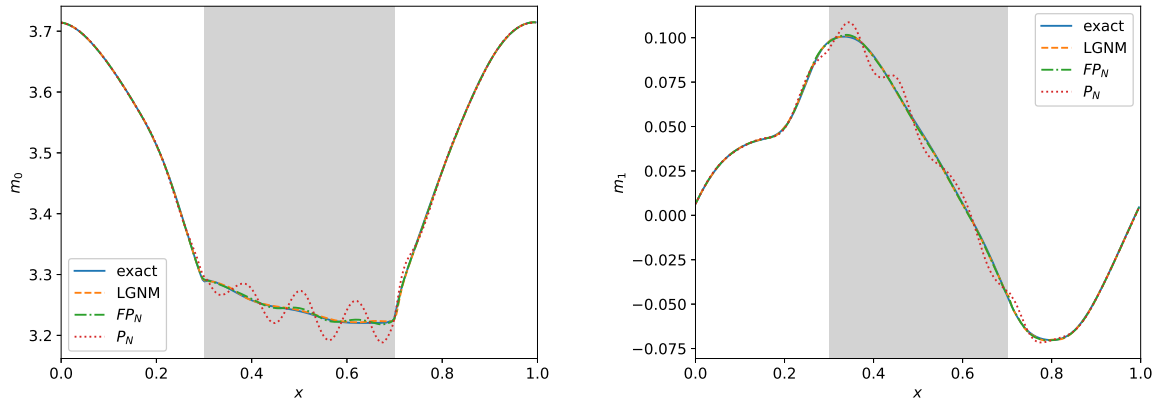



(c) $m_4$ (d) $m_5$

Figure 4.18: Example 4.5: two-material problem. Numerical solutions of $m_0$, $m_1$, $m_4$ and $m_5$ at $t = 0.4$ with $N = 5$. We take $\nu = 20$ in the $FP_N$ closure. The gray part in the middle is in the optically thin regime and the other part is in the intermediate regime.

| $N$ | relative $L^2$ error of $m_0$ | | | relative $L^2$ error of $m_N$ | | |
|---|---|---|---|---|---|---|
| | $P_N$ | $FP_N$ | LGNM | $P_N$ | $FP_N$ | LGNM |
| 5 | 8.61e-03 | 9.55e-04 | 7.28e-04 | 1.14e+00 | 7.72e-01 | 6.06e-01 |
| 6 | 4.93e-03 | 6.40e-04 | 6.74e-04 | 1.01e+00 | 7.79e-01 | 7.81e-01 |
| 7 | 5.84e-03 | 6.83e-04 | 5.71e-04 | 9.36e-01 | 7.66e-01 | 1.13e-01 |
| 8 | 5.56e-03 | 6.78e-04 | 4.52e-04 | 9.03e-01 | 7.68e-01 | 8.92e-02 |
| 9 | 4.00e-03 | 5.25e-04 | 4.21e-04 | 8.24e-01 | 7.71e-01 | 5.06e-02 |

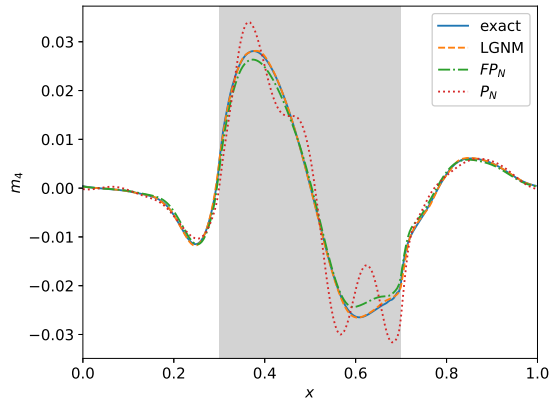
Table 4.1: Example 4.5: two-material problem. The relative $L^2$ errors of $m_0$ and $m_N$ at $t = 0.4$ with $N = 5, 6, 7, 8, 9$. We take $\nu = 20$ in the $FP_N$ closure.



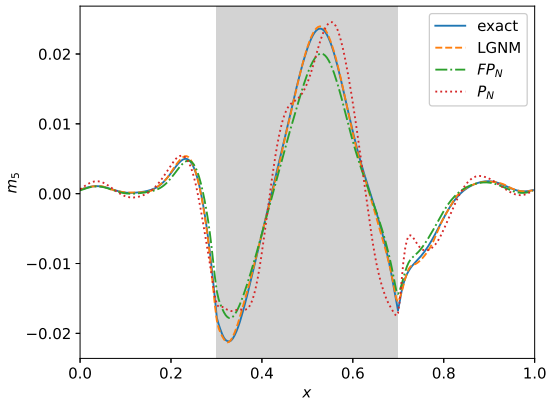

(a) $m_0$



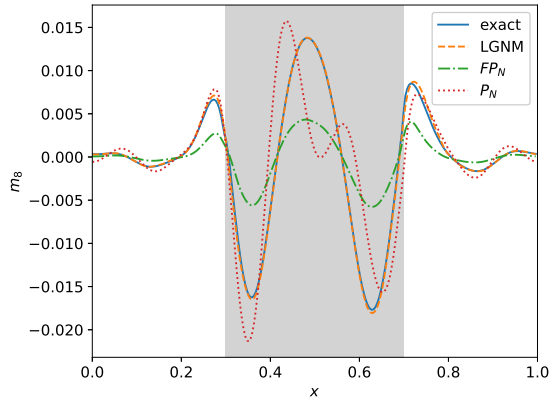

(b) $m_1$



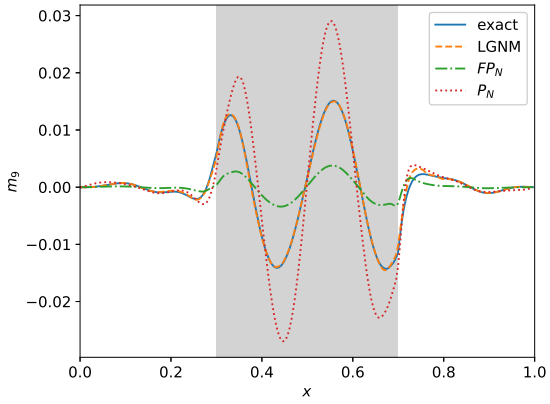

(c) $m_4$




(d) $m_5$




(e) $m_8$




(f) $m_9$

Figure 4.19: Example 4.5: two-material problem. Numerical solutions of $m_0$, $m_1$, $m_4$, $m_5$, $m_8$ and $m_9$ at $t = 0.4$ with $N = 9$. We take $\nu = 20$ in the $FP_N$ closure. The gray part in the middle is in the optically thin regime and the other part is in the intermediate regime.

## 5 Conclusion

In this work, we investigate the moment closure problem for the RTE in slab geometry and learn a closure relation from data. Instead of learning the moment itself, we use neural networks to directly learn its gradient. This new approach is consistent with the exact closure we derive for the free streaming limit and also provides a natural output normalization. Moreover, we incorporate the scale invariance of the closure model into the neural networks, which brings better generalization and performance, especially when applied to initial conditions that have their magnitude outside of the training data. A variety of benchmark tests, including the variable scattering problem, the Gaussian source problem and the two-material problem, were investigated. All tests show that the Learn Gradient ML closure model has both good accuracy as well as that the ML closure model using a simple training procedure has a strong generalization property, i.e., we did not need additional training to maintain accuracy even when applied to problems with discontinuities in the scattering cross section.

We also remark that the methodology of learning the gradients can be generalized to the multidimensional case. In this case, we would relate the spatial derivatives in each direction of the higher order moments to the derivatives of the lower order moments in all the spatial directions. We are currently working on this topic and hope to report the progress in the near future.

Finally, we point out that hyperbolicity is an important property in moment closure models, which is difficult to enforce for traditional closure models [9, 41] as well as ML models [25]. Our current ML closure model is not able to preserve hyperbolicity and thus have some numerical instabilities. How to incorporate hyperbolicity in the ML closure model is certainly an interesting topic and constitutes our ongoing work [24, 23].

## Acknowledgment

We thank Michael M. Crockatt in Sandia National Laboratories for providing numerical solver for the radiative transfer equation. We acknowledge the High Performance Computing Center (HPCC) at Michigan State University for providing computational resources that have contributed to the research results reported within this paper. We thank two anonymous reviewers for providing helpful comments on earlier draft of the manuscript.

## References

[1] G. W. Alldredge, C. D. Hauck, D. P. OLeary, and A. L. Tits. Adaptive change of basis in entropy-based moment closures for linear kinetic equations. *Journal of Computational Physics*,

258:489–508, 2014.

[2] G. W. Alldredge, C. D. Hauck, and A. L. Tits. High-order entropy-based closures for linear transport in slab geometry ii: A computational study of the optimization problem. *SIAM Journal on Scientific Computing*, 34(4):B361–B391, 2012.

[3] G. W. Alldredge, R. Li, and W. Li. Approximating the $M_2$ method by the extended quadrature method of moments for radiative transfer in slab geometry. *Kinetic & Related Models*, 9(2):237, 2016.

[4] G. Bird. *Molecular Gas Dynamics and the Direct Simulation of Gas Flows*. Clarendon Press, Oxford, United Kingdom, 1994.

[5] L. Bois, E. Franck, L. Navoret, and V. Vigon. A neural network closure for the Euler-Poisson system based on kinetic simulations. *arXiv preprint arXiv:2011.06242*, 2020.

[6] S. L. Brunton, J. L. Proctor, and J. N. Kutz. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the national academy of sciences*, 113(15):3932–3937, 2016.

[7] A. G. Buchan, A. Calloo, M. G. Goffin, S. Dargaville, F. Fang, C. C. Pain, and I. M. Navon. A POD reduced order model for resolving angular direction in neutron/photon transport problems. *Journal of Computational Physics*, 296:138–157, 2015.

[8] J. Bünger, N. Sarna, and M. Torrilhon. Stable boundary conditions and discretization for pn equations. *arXiv preprint arXiv:2004.02497*, 2020.

[9] Z. Cai, Y. Fan, and R. Li. Globally hyperbolic regularization of Grad's moment system. *Communications on pure and applied mathematics*, 67(3):464–518, 2014.

[10] S. Chandrasekhar. On the radiative equilibrium of a stellar atmosphere. *The Astrophysical Journal*, 99:180, 1944.

[11] M. M. Crockatt, A. J. Christlieb, C. K. Garrett, and C. D. Hauck. An arbitrary-order, fully implicit, hybrid kinetic solver for linear radiative transport using integral deferred correction. *Journal of Computational Physics*, 346:212–241, 2017.

[12] M. M. Crockatt, A. J. Christlieb, C. K. Garrett, and C. D. Hauck. Hybrid methods for radiation transport using diagonally implicit runge–kutta and space–time discontinuous galerkin time integration. *Journal of Computational Physics*, 376:455–477, 2019.

[13] Y. Fan, R. Li, and L. Zheng. A nonlinear hyperbolic model for radiative transfer equation in slab geometry. *SIAM Journal on Applied Mathematics*, 80(6):2388–2419, 2020.

[14] Y. Fan, R. Li, and L. Zheng. A nonlinear moment model for radiative transfer equation in slab geometry. *Journal of Computational Physics*, 404:109128, 2020.

[15] M. Frank, C. D. Hauck, and E. Olbrant. Perturbed, entropy-based closure for radiative transfer. *arXiv preprint arXiv:1208.0772*, 2012.

[16] W. Guo and Y. Cheng. A sparse grid discontinuous galerkin method for high-dimensional transport equations and its application to kinetic simulations. *SIAM Journal on Scientific Computing*, 38(6):A3381–A3409, 2016.

[17] W. Guo and Y. Cheng. An adaptive multiresolution discontinuous galerkin method for time-dependent transport equations in multidimensions. *SIAM Journal on Scientific Computing*, 39(6):A2962–A2992, 2017.

[18] J. Han, A. Jentzen, and W. E. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences*, 115(34):8505–8510, 2018.

[19] J. Han, C. Ma, Z. Ma, and W. E. Uniformly accurate machine learning-based hydrodynamic models for kinetic equations. *Proceedings of the National Academy of Sciences*, 116(44):21983–21991, 2019.

[20] C. Hauck and R. McClarren. Positive $P_N$ closures. *SIAM Journal on Scientific Computing*, 32(5):2603–2626, 2010.

[21] C. D. Hauck. High-order entropy-based closures for linear transport in slab geometry. *Communications in Mathematical Sciences*, 9(1):187–205, 2011.

[22] C. F. Higham and D. J. Higham. Deep learning: An introduction for applied mathematicians. *SIAM Review*, 61(4):860–891, 2019.

[23] J. Huang, Y. Cheng, A. J. Christlieb, and L. F. Roberts. Machine learning moment closure models for the radiative transfer equation III: enforcing hyperbolicity and physical characteristic speeds. *arXiv preprint arXiv:2109.00700*, 2021.

[24] J. Huang, Y. Cheng, A. J. Christlieb, L. F. Roberts, and W.-A. Yong. Machine learning moment closure models for the radiative transfer equation II: enforcing global hyperbolicity in gradient based closures. *arXiv preprint arXiv:2105.14410*, 2021.

[25] J. Huang, Z. Ma, Y. Zhou, and W.-A. Yong. Learning thermodynamically stable and Galilean invariant partial differential equations for non-equilibrium flows. *Journal of Non-Equilibrium Thermodynamics*, 2021.

[26] G.-S. Jiang and C.-W. Shu. Efficient implementation of weighted ENO schemes. *Journal of Computational Physics*, 126(1):202–228, 1996.

[27] Y. Kim, Y. Choi, D. Widemann, and T. Zohdi. A fast and accurate physics-informed neural network reduced order model with shallow masked autoencoder. *arXiv preprint arXiv:2009.11990*, 2020.

[28] A. D. Klose, U. Netz, J. Beuthan, and A. H. Hielscher. Optical tomography using the time-independent equation of radiative transfer—part 1: forward model. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 72(5):691–713, 2002.

[29] R. Koch and R. Becker. Evaluation of quadrature schemes for the discrete ordinates method. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 84(4):423–435, 2004.

[30] J. Koellermeier and M. Castro. High-order non-conservative simulation of hyperbolic moment models in partially-conservative form. *East Asian Journal on Applied Mathematics*, 2021.

[31] J. Koellermeier and M. Torrilhon. Numerical study of partially conservative moment equations in kinetic theory. *Communications in Computational Physics*, 21(4):981–1011, 2017.

[32] V. M. Laboure, R. G. McClarren, and C. D. Hauck. Implicit filtered $P_N$ for high-energy density thermal radiation transport using discontinuous galerkin finite elements. *Journal of Computational Physics*, 321:624–643, 2016.

[33] E. Larsen and J. Morel. Asymptotic solutions of numerical transport problems in optically thick, diffusive regimes ii. *Journal of Computational Physics*, 83(1), 1989.

[34] E. W. Larsen and J. E. Morel. Advances in discrete-ordinates methodology. *Nuclear Computational Science*, pages 1–84, 2010.

[35] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[36] K. Lee and K. Carlberg. Deep conservation: A latent-dynamics model for exact satisfaction of physical conservation laws. *arXiv preprint arXiv:1909.09754*, 2019.

[37] K. Lee and K. T. Carlberg. Model reduction of dynamical systems on nonlinear manifolds using deep convolutional autoencoders. *Journal of Computational Physics*, 404:108973, 2020.

[38] C. Levermore. Relating eddington factors to flux limiters. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 31(2):149–160, 1984.

[39] C. D. Levermore. Moment closure hierarchies for kinetic theories. *Journal of statistical Physics*, 83(5):1021–1065, 1996.

[40] R. Li, E. Lee, and T. Luo. Physics-informed neural networks for solving multiscale mode-resolved phonon Boltzmann transport equation. *arXiv preprint arXiv:2103.07983*, 2021.

[41] R. Li, W. Li, and L. Zheng. Direct flux gradient approximation to close moment model for kinetic equations. *arXiv preprint arXiv:2102.07641*, 2021.

[42] Q. Lou, X. Meng, and G. E. Karniadakis. Physics-informed neural networks for solving forward and inverse flow problems via the Boltzmann-BGK formulation. *arXiv preprint arXiv:2010.09147*, 2020.

[43] C. Ma, B. Zhu, X.-Q. Xu, and W. Wang. Machine learning surrogate models for Landau fluid closure. *Physics of Plasmas*, 27(4):042502, 2020.

[44] R. Maulik, N. A. Garland, J. W. Burby, X.-Z. Tang, and P. Balaprakash. Neural network representability of fully ionized plasma fluid model closures. *Physics of Plasmas*, 27(7):072106, 2020.

[45] R. G. McClarren and C. D. Hauck. Robust and accurate filtered spherical harmonics expansions for radiative transfer. *Journal of Computational Physics*, 229(16):5597–5614, 2010.

[46] S. Mishra and R. Molinaro. Physics informed neural networks for simulating radiative transfer. *arXiv preprint arXiv:2009.13291*, 2020.

[47] E. Murchikova, E. Abdikamalov, and T. Urbatsch. Analytic closures for M1 neutrino transport. *Monthly Notices of the Royal Astronomical Society*, 469(2):1725–1737, 2017.

[48] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala. Pytorch: An imperative style, high-performance deep learning library. *arXiv preprint arXiv:1912.01703*, 2019.

[49] Z. Peng, Y. Chen, Y. Cheng, and F. Li. A reduced basis method for radiative transfer equation. *arXiv preprint arXiv:2103.07574*, 2021.

[50] G. C. Pomraning. *The Equations of Radiation Hydrodynamics*. Pergamon Press, Oxford, UK, 1973.

[51] D. Radice, E. Abdikamalov, L. Rezzolla, and C. D. Ott. A new spherical harmonics scheme for multi-dimensional radiation transport I. static matter configurations. *Journal of Computational Physics*, 242:648–669, 2013.

[52] M. Raissi, P. Perdikaris, and G. E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019.

[53] D. Rim, L. Venturi, J. Bruna, and B. Peherstorfer. Depth separation for reduced deep networks in nonlinear model reduction: Distilling shock waves in nonlinear hyperbolic problems. *arXiv preprint arXiv:2007.13977*, 2020.

[54] J. B. Scoggins, J. Han, and M. Massot. Machine learning moment closures for accurate and efficient simulation of polydisperse evaporating sprays. In *AIAA Scitech 2021 Forum*, page 1786, 2021.

[55] C.-W. Shu and S. Osher. Efficient implementation of essentially non-oscillatory shock-capturing schemes. *Journal of computational physics*, 77(2):439–471, 1988.

[56] R. Spurr, T. Kurosu, and K. 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.

[57] J. Tencer, K. Carlberg, R. Hogan, and M. Larsen. 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.

[58] L. Wang, X. Xu, B. Zhu, C. Ma, and Y.-A. Lei. Deep learning surrogate model for kinetic Landau-fluid closure with collision. *AIP Advances*, 10(7):075108, 2020.

[59] T. Xiao and M. Frank. Using neural networks to accelerate the solution of the Boltzmann equation. *arXiv preprint arXiv:2010.13649*, 2020.

[60] L. Zhang, J. Han, H. Wang, R. Car, and W. E. Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics. *Physical Review Letters*, 120(14):143001, 2018.

[61] Y. Zhu, L. Hong, Z. Yang, and W.-A. Yong. Conservation-dissipation formalism of irreversible thermodynamics. *Journal of Non-Equilibrium Thermodynamics*, 40(2):67–74, 2015.