Asymptotic-Preserving Neural Networks for Multiscale Time-Dependent Linear Transport Equations

Keke Wu

August 17, 2023

School of Mathematical Sciences, Shanghai Jiao Tong University

Joint work with **Shi Jin** and **Zheng Ma** (SJTU)

Outline

Introduction

Multiscale Time-Dependent Linear Transport Equation

$$\partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon}\mathcal{B}(f, f).$$

- f : distribution function of particles at time t, space position x and traveling in direction v
- $\mathcal{B}$: collision operator
- $\varepsilon > 0$: the mean free path (Knudsen number)

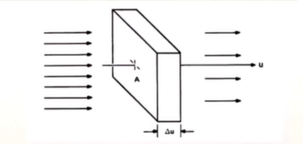


Figure 1: Particles travel through a medium with scattering and absorption.

Multiscale problem: the magnitude of ε from $\varepsilon = O(1)$ to $\varepsilon \ll 1$.

Nondimensionalization of the time-dependent linear transport equation

$$\varepsilon \partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon}\left(\frac{1}{2}\int_{-1}^{1} f \,\mathrm{d}v' - f\right).$$

Deep neural networks (DNNs)

An L-layer feed forward neural network with the set of parameters denoted by θ is defined recursively as,

$$
\begin{aligned}
f_\theta^{[0]}(x) &= x, \\
f_\theta^{[l]}(x) &= \sigma \circ (W^{[l-1]} f_\theta^{[l-1]}(x) + b^{[l-1]}),\ 1 \le l \le L-1, \\
f_\theta(x) = f_\theta^{[L]}(x) &= W^{[L-1]} f_\theta^{[L-1]}(x) + b^{[L-1]},
\end{aligned}
$$

where $W^{[l]} \in \mathbb{R}^{m_{l+1} \times m_l}$, $b^l \in \mathbb{R}^{m_{l+1}}$, m_0, m_L are the input and output dimension, σ is a scalar function and "$\circ$" means entry-wise operation.

Solving the linear transport equation by PINNs

$$\varepsilon \partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon}\left(\frac{1}{2}\int_{-1}^{1} f \,\mathrm{d}v' - f\right), \quad (t, x, v) \in \mathcal{T} \times \mathcal{D} \times \Omega.$$

Physics-informed neural networks(PINNs): approxmates the solution with deep neural networks

$$f_\theta^{\mathsf{NN}}(t, x, v) \approx f(t, x, v).$$

Procedure of solving PDEs by DNNs:

- Modeling: define the loss associated to a PDE;

- Architecture: build a deep neural network(function class) for the trail function;

- Optimization: minimize loss over the parameter space.

Solving the linear transport equation by PINNs

Take the least square of the residual equation as the loss function, together with boundary and initial conditions as penalty terms, that is,

$$
\begin{aligned}
\mathcal{R}^{\varepsilon}_{\mathsf{PINN}} = & \frac{1}{|\mathcal{T}\times\mathcal{D}\times\Omega|}\int_{\mathcal{T}}\int_{\mathcal{D}}\int_{\Omega}\left|\varepsilon^2\partial_t f_\theta^{\mathsf{NN}} + \varepsilon v\cdot\nabla_x f_\theta^{\mathsf{NN}} - \mathcal{L}f_\theta^{\mathsf{NN}} - \varepsilon^2 Q\right|^2 \mathrm{d}v\,\mathrm{d}x\,\mathrm{d}t \\
& + \frac{\lambda_1}{|\mathcal{T}\times\partial\mathcal{D}\times\Omega|}\int_{\mathcal{T}}\int_{\partial\mathcal{D}}\int_{\Omega}\left|\mathcal{B}f_\theta^{\mathsf{NN}} - F_{\mathsf{B}}\right|^2 \mathrm{d}v\,\mathrm{d}x\,\mathrm{d}t \\
& + \frac{\lambda_2}{|\mathcal{D}\times\Omega|}\int_{\mathcal{D}}\int_{\Omega}\left|\mathcal{I}f_\theta^{\mathsf{NN}} - f_0\right|^2 \mathrm{d}v\,\mathrm{d}x,
\end{aligned}
$$

where λ_1 and λ_2 are the penalty weights to be tuned.

Finally, Adam optimizer is used to find the global minimum of loss.

Motivation

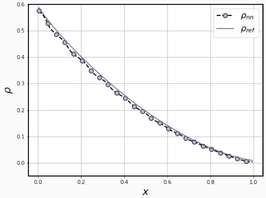
Consider the one-dimensional transport equation in slab geometry

$$\varepsilon \partial_t f + v \cdot \nabla_x f = \frac{1}{\varepsilon}\left(\rho(t,x) - f\right), \rho(t,x) = \frac{1}{2}\int_{-1}^{1} f \, \mathrm{d}v', \; x_L < x < x_R, \; -1 \le v \le 1,$$

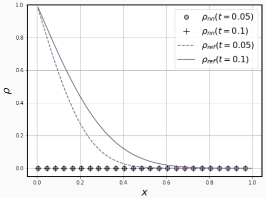
with initial condition $f_0(x,v) = 0$ and in-flow boundary conditions as,

$$\begin{cases} f(t, x_L, v) = 1 \text{ for } v > 0, \\ f(t, x_R, v) = 0 \text{ for } v < 0. \end{cases}$$

Case I: Kinetic regime ($\varepsilon = 1$)



Case II: Diffusion regime ($\varepsilon = 10^{-8}$)



PINN fails to obtain the ground truth !

The failure of PINN loss to resolve small scales

$$\mathcal{R}^{\varepsilon}_{\text{PINN}} = \frac{1}{|\mathcal{T}\times\mathcal{D}\times\Omega|}\int_{\mathcal{T}}\int_{\mathcal{D}}\int_{\Omega}\left|\varepsilon^2\partial_t f^{\text{NN}}_{\theta} + \varepsilon v\cdot\nabla_x f^{\text{NN}}_{\theta} - \left(\frac{1}{2}\int_{-1}^{1} f^{\text{NN}}_{\theta}\mathrm{d}v' - f^{\text{NN}}_{\theta}\right)\right|^2 \mathrm{d}v\mathrm{d}x\mathrm{d}t$$
$$+ \frac{\lambda_1}{|\mathcal{T}\times\partial\mathcal{D}\times\Omega|}\int_{\mathcal{T}}\int_{\partial\mathcal{D}}\int_{\Omega}|\mathcal{B}f^{\text{NN}}_{\theta} - F_{\text{B}}|^2\mathrm{d}v\mathrm{d}x\mathrm{d}t$$
$$+ \frac{\lambda_2}{|\mathcal{D}\times\Omega|}\int_{\mathcal{D}}\int_{\Omega}|\mathcal{I}f^{\text{NN}}_{\theta} - f_0|^2\mathrm{d}v\mathrm{d}x.$$

We only need to focus on the first term and taking $\varepsilon \to 0$, this will led to

$$\mathcal{R}^{0}_{\text{PINN}} = \frac{1}{|\mathcal{T}\times\mathcal{D}\times\Omega|}\int_{\mathcal{T}}\int_{\mathcal{D}}\int_{\Omega}\left|-\left(\frac{1}{2}\int_{-1}^{1} f^{\text{NN}}_{\theta}\mathrm{d}v' - f^{\text{NN}}_{\theta}\right)\right|^2 \mathrm{d}v\mathrm{d}x\mathrm{d}t,$$

which can be viewed as the PINN loss of the equilibrium equation

$$f = \frac{1}{2}\int_{-1}^{1} f\,\mathrm{d}v'.$$

Next, we show the limit equation of the linear transport equation is the *diffusion equation*.

Micro-macro decomposition

$$\varepsilon\partial_t f + v\cdot\nabla_x f = \frac{1}{\varepsilon}\left(\rho(t,x) - f\right),\ \rho(t,x) = \langle f\rangle := \frac{1}{2}\int_{-1}^{1} f\,\mathrm{d}v'$$

Decompose f into the equilibrium $\rho(t,x)$ and the non-equilibrium part $g(t,x,v)$:

$$f(t,x,v) = \rho(t,x) + \varepsilon g(t,x,v).$$

The non-equilibrium part g clearly satisfies $\langle g\rangle = 0$. Subsititing $f = \rho + \varepsilon g$ into the linear transport equation yields

$$\varepsilon\partial_t\rho + \varepsilon^2\partial_t g + v\cdot\nabla_x\rho + \varepsilon v\cdot\nabla_x g = -g. \tag{1}$$

Integrating this equation with respect to v :

$$\left\langle \varepsilon\partial_t\rho + \varepsilon^2\partial_t g + v\cdot\nabla_x\rho + \varepsilon v\cdot\nabla_x g\right\rangle = 0,$$

i.e.,

$$\partial_t\rho + \langle v\cdot\nabla_x g\rangle = 0.$$

Define operator $\Pi(\cdot)(v) : \langle\cdot\rangle$ and I the identity operator, then apply the projection operator $I - \Pi$ to equation (1):

$$(I - \Pi)\left(\varepsilon\partial_t\rho + \varepsilon^2\partial_t g + v\cdot\nabla_x\rho + \varepsilon v\cdot\nabla_x g\right) = -\left(I - \Pi\right)(g),$$

i.e.,

$$\varepsilon^2\partial_t g + \varepsilon\left(I - \Pi\right)\left(v\cdot\nabla_x g\right) + v\cdot\nabla_x\rho = -g. \tag{2}$$

Thus, one can get the micro-macro system for the linear trasport equation

$$\begin{cases} \partial_t\rho + \langle v\cdot\nabla_x g\rangle = 0, \\ \\ \varepsilon^2\partial_t g + \varepsilon\left(I - \Pi\right)\left(v\cdot\nabla_x g\right) + v\cdot\nabla_x\rho = -g. \end{cases}$$

The micro-macro system for the linear trasport equation

$$\begin{cases} \partial_t \rho + \langle v \cdot \nabla_x g \rangle = 0, \\ \\ \varepsilon^2 \partial_t g + \varepsilon (I - \Pi)(v \cdot \nabla_x g) + v \cdot \nabla_x \rho = -g. \end{cases}$$

Sending $\varepsilon \to 0$, the above system formally approaches

$$\begin{cases} \partial_t \rho + \langle v \cdot \nabla_x g \rangle = 0, \\ -v \cdot \nabla_x \rho = g. \end{cases}$$

Plugging the second equation into the first equation gives the diffusion equation

$$\rho_t - \frac{1}{3} \rho_{xx} = 0.$$

What kind of loss is “good” ?

Asymptotic-Preserving neural networks

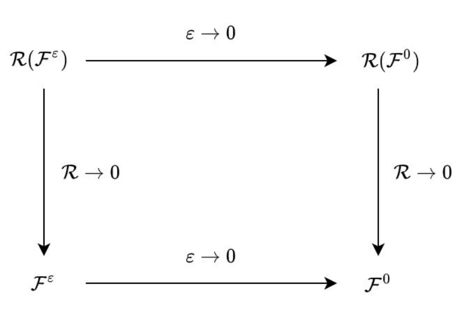


Figure 2: Illustration of APNNs. $\mathcal{F}^\varepsilon$ is the microscopic equation that depends on the small scale parameter ε and $\mathcal{F}^0$ is its macroscopic limit as $\varepsilon \to 0$, which is independent of ε. The latent solution of $\mathcal{F}^\varepsilon$ is approximated by neural networks with its measure(risk) denoted by $\mathcal{R}(\mathcal{F}^\varepsilon)$. The asymptotic limit of $\mathcal{R}(\mathcal{F}^\varepsilon)$ as $\varepsilon \to 0$, if exists, is denoted by $\mathcal{R}(\mathcal{F}^0)$. If $\mathcal{R}(\mathcal{F}^0)$ is a good measure(risk) of $\mathcal{F}^0$, then it is called AP.

APNNs for linear transport equation

$$\begin{cases} \partial_t \rho + \langle v \cdot \nabla_x g \rangle = 0, \\ \varepsilon^2 \partial_t g + \varepsilon (I - \Pi)(v \cdot \nabla_x g) + v \cdot \nabla_x \rho = -g. \end{cases} \xrightarrow{\varepsilon \to 0} \rho_t - \frac{1}{3}\rho_{xx} = 0.$$

Our APNN method rewrites the original PDE into a system in AP form and takes their mean-square residual error as the loss.

Two networks are used to parametrize two functions ρ and g :

$$\rho_\theta^{\mathrm{NN}}(t,x) := \exp\left(-\tilde{\rho}_\theta^{\mathrm{NN}}(t,x)\right) \approx \rho(t,x),$$

$$g_\theta^{\mathrm{NN}}(t,x,v) := \tilde{g}_\theta^{\mathrm{NN}}(t,x,v) - \langle \tilde{g}_\theta^{\mathrm{NN}} \rangle(t,x) \approx g(t,x,v).$$

Here $\tilde{\rho}$ and $\tilde{g}$ are both fully-connected neural netwroks.

Notice that $g_\theta^{\mathrm{NN}}(t,x,v)$ automatically satisfy the constraint :

$$\langle g_\theta^{\mathrm{NN}} \rangle = \langle \tilde{g}_\theta^{\mathrm{NN}} \rangle - \langle \tilde{g}_\theta^{\mathrm{NN}} \rangle = 0, \quad \forall\ t, x.$$

Loss of APNNs

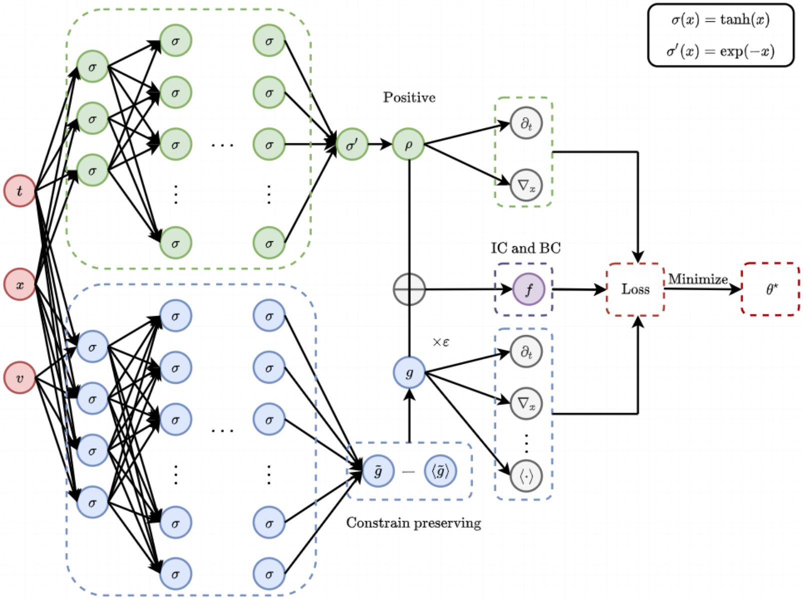
Then we propose the least square of the residual of the micro-macro system as the APNN loss,

$$\begin{aligned}\mathcal{R}^{\varepsilon}_{\mathsf{APNN}} = &\frac{1}{|\mathcal{T}\times\mathcal{D}|}\int_{\mathcal{T}}\int_{\mathcal{D}}|\partial_t\rho_\theta^{\mathsf{NN}} + \nabla_x\cdot\left\langle v g_\theta^{\mathsf{NN}}\right\rangle|^2\,\mathrm{d}x\,\mathrm{d}t\\ &+\frac{1}{|\mathcal{T}\times\mathcal{D}\times\Omega|}\int_{\mathcal{T}}\int_{\mathcal{D}}\int_{\Omega}|\varepsilon^2\partial_t g_\theta^{\mathsf{NN}} + \varepsilon(I-\Pi)(v\cdot\nabla_x g_\theta^{\mathsf{NN}})\\ &\quad + v\cdot\nabla_x\rho_\theta^{\mathsf{NN}} - \mathcal{L}g_\theta^{\mathsf{NN}}|^2\,\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}t\\ &+\frac{\lambda_1}{\mathcal{T}\times\partial\mathcal{D}\times\Omega|}\int_{\mathcal{T}}\int_{\partial\mathcal{D}}\int_{\Omega}|\mathcal{B}(\rho_\theta^{\mathsf{NN}}+\varepsilon g_\theta^{\mathsf{NN}}) - F_{\mathsf{B}}|^2\,\mathrm{d}v\,\mathrm{d}x\,\mathrm{d}t\\ &+\frac{\lambda_2}{|\mathcal{D}\times\Omega|}\int_{\mathcal{D}}\int_{\Omega}|\mathcal{I}(\rho_\theta^{\mathsf{NN}}+\varepsilon g_\theta^{\mathsf{NN}}) - f_0|^2\,\mathrm{d}v\,\mathrm{d}x.\end{aligned}$$

Sending $\varepsilon\to 0$, one have

$$\begin{aligned}\mathcal{R}^{0}_{\mathsf{APNN}} = &\frac{1}{|\mathcal{T}\times\mathcal{D}|}\int_{\mathcal{T}}\int_{\mathcal{D}}|\partial_t\rho_\theta^{\mathsf{NN}} + \nabla_x\cdot\left\langle v g_\theta^{\mathsf{NN}}\right\rangle|^2\mathrm{d}x\mathrm{d}t\\ &+\frac{1}{|\mathcal{T}\times\mathcal{D}\times\Omega|}\int_{\mathcal{T}}\int_{\mathcal{D}}\int_{\Omega}|v\cdot\nabla_x\rho_\theta^{\mathsf{NN}} + g_\theta^{\mathsf{NN}}|^2\mathrm{d}v\mathrm{d}x\mathrm{d}t.\end{aligned}$$

This loss is exactly what we want.


$\sigma(x) = \tanh(x)$
$\sigma'(x) = \exp(-x)$
t
x
v
σ
σ'
ρ
Positive
∂_t
∇_x
IC and BC
f
Loss
Minimize
$\theta^\star$
$\times\varepsilon$
g
$\langle\cdot\rangle$
$\tilde{g}$ — $\langle\tilde{g}\rangle$
Constrain preserving

Numerical results

Problem I. Smooth initial data with periodic BC

Consider periodic boundary condition with a smooth initial data as follows

$$f_0(x,v) = \frac{\rho(x)}{\sqrt{2\pi}} e^{-\frac{v^2}{2}}, \ \rho(x) = 1 + \cos(4\pi x).$$

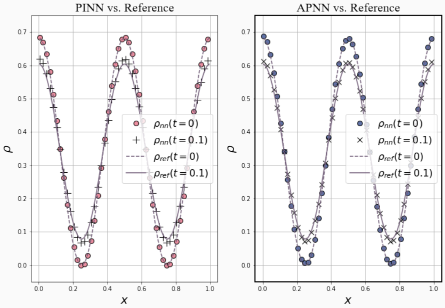
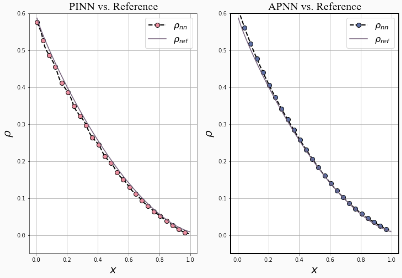


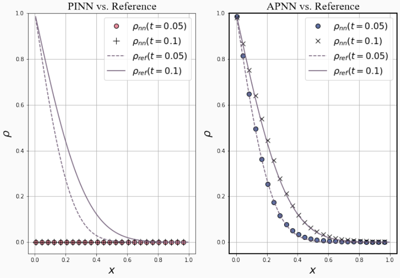
Figure 4: Plots of approximate density $\rho(t,x)$ by PINN, APNN vs. Reference.

Problem II. In-flow BC: from kinetic regime to diffusion regime

$$\begin{cases} f(t, x_L, v) = 1 \text{ for } v > 0, \\ f(t, x_R, v) = 0 \text{ for } v < 0. \end{cases} \quad \text{and} \quad f_0(x, v) = 0.$$



(a) Kinetic regime with $\varepsilon = 1$



(b) Diffusion regime with $\varepsilon = 10^{-8}$

Figure 5: Plots of approximate density $\rho(t, x)$ by PINN, APNN vs. Reference.

Problem II. In-flow BC: from kinetic regime to diffusion regime

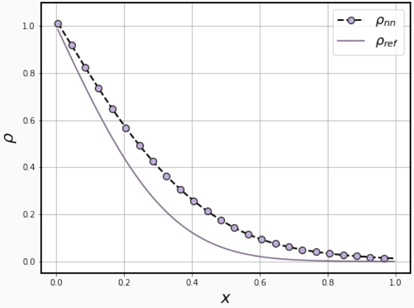
Table 1: Comparison of PINN and APNN from kinetic regime to diffusion regime.

QoIs \ ε	1		10^{-1}		10^{-3}		10^{-8}	
	PINN	APNN	PINN	APNN	PINN	APNN	PINN	APNN
Relative ℓ^2 error	4.01 e-2	1.36 e-2	1.17 e-1	3.30 e-2	2.17 e-1	1.98 e-2	9.40 e-1	2.76 e-2

For the constraint $\langle g \rangle = 0$, one way is to construct a novel neural network for g such that it exactly satisfies $\langle g \rangle = 0$ (hard constraint). The other way is to treat it as a soft constraint with parameter λ_3, we use $\tilde{g}_\theta^{\text{NN}}$ and modifies the loss as

$$\mathcal{R}_{\text{APNN}} + \frac{\lambda_3}{|\mathcal{T} \times \mathcal{D}|} \int_{\mathcal{T}} \int_{\mathcal{D}} \left| \left\langle \tilde{g}_\theta^{\text{NN}} \right\rangle - 0 \right|^2 \mathrm{d}x \mathrm{d}t \quad \text{(soft constraint)}.$$

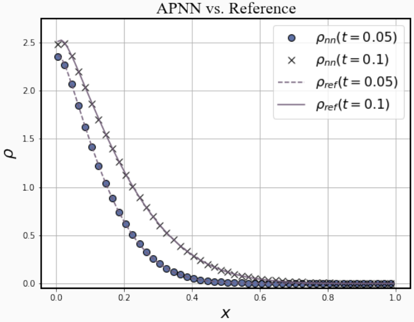
The following figure shows the approximate density at $t = 0.1$ by APNN with soft constraint for $\varepsilon = 10^{-8}$.



Problem III. A problem with boundary layer

$$\begin{cases} f(t, x_L, v) = 5\sin(v) \text{ for } v > 0, \\ f(t, x_R, v) = 0 \text{ for } v < 0. \end{cases} \quad \text{and} \quad f_0(x, v) = 0.$$

The following figure shows the approximate density at $t = 0.05, 0.1$ by APNN for $\varepsilon = 5 \times 10^{-2}$.



The two-dimensional case is similar except the velocity/angular variables are constrained in the unit circle

$$\varepsilon\partial_t f + v\cdot\partial_x f = \frac{1}{\varepsilon}\left(\frac{1}{2\pi}\int_{|v|=1} f\,\mathrm{d}v' - f\right) + \varepsilon Q, \quad x\in\Gamma\subset\mathbb{R}^2, \quad |v|=1,$$

with the in-flow boundary condition

$$f(t,x,v)=0 \quad \text{for} \quad n\cdot v<0, \quad x\in\partial\Gamma,$$

where n is the outer normal of the boundary. The initial condition is $f_0(x,v)=0$

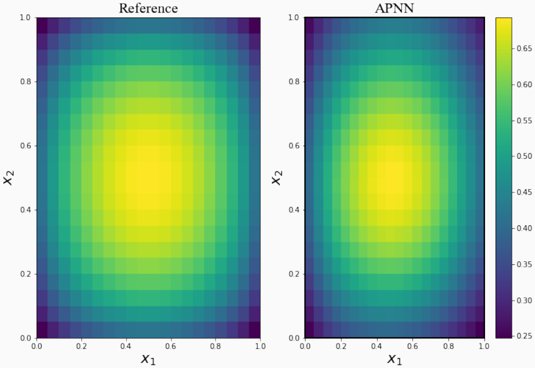
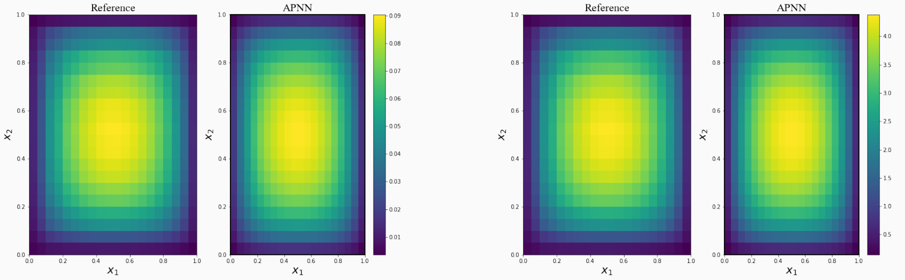


Figure 6: Plot of density $\rho(x_1, x_2)$ with $\varepsilon = 1$, $Q = 1$ at $t = 1.0$ by APNN vs. Reference. The relative ℓ^2 error is 5.78×10^{-2}.

Problem IV. Diffusion regime



(a) $Q = 1$ (b) $Q = 50$

Figure 7: Plot of density $\rho(x_1, x_2)$ with $\varepsilon = 10^{-8}$ at $t = 0.1$ by APNN vs. Reference. The relative ℓ^2 error are 1.29×10^{-2} and 4.35×10^{-2} for source $Q = 1, 50$, respectively.

Problem V. Uncertainty quantification (UQ) problem

For the UQ problem we consider the linear transport equation with a Gaussian scattering function ($\varepsilon = 10^{-8}$):

$$\varepsilon \partial_t f + v \partial_x f = \frac{\sigma_S(\boldsymbol{z})}{\varepsilon} \left(\frac{1}{2} \int_{-1}^{1} f \, dv' - f \right), \quad x_L < x < x_R, \quad -1 \leq v \leq 1,$$

with scattering coefficients

$$\sigma_S(\boldsymbol{z}) = 1 + 0.3 \exp\left(-\frac{|\boldsymbol{z}|^2}{2} \right), \ \boldsymbol{z} = (z_1, z_2, \cdots, z_{20}) \sim \mathcal{U}([-3, 3]^{20}).$$

The boundary and initial condition are

$$\begin{cases} f(t, x_L, v) = 1 \text{ for } v > 0, \\ f(t, x_R, v) = 0 \text{ for } v < 0, \end{cases} \quad \text{and} \quad f_0(x, v) = 0.$$

Problem V. Diffusion regime

The following figure shows the approximate density by taking expectation for z at $t = 0.05, 0.1$ by APNN for $\varepsilon = 10^{-8}$.

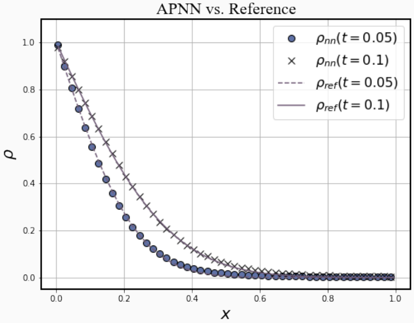


Figure 8: Plot of density ρ by taking expectation for z at $t = 0.05, 0.1$ by APNN vs. Reference.The relative ℓ^2 error is 1.51×10^{-2}.

Conclusions

Conclusions

- PINNs can only capture the leading order/single scale depending on how to design the loss behavior.
- Our work defines the loss based on an AP strategy, which captures the correct asymptotic behavior when the physical scaling parameters become small.
- Beyong the AP property, we propose a conservation mechanism in our loss to satisfy physical constraints.

References

- Maziar Raissi, Paris Perdikaris, and George Em 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, 201
- Shi Jin. Asymptotic preserving (ap) schemes for multiscale kinetic and hyperbolic equations: a review. Lecture notes for summer school on methods and models of kinetic theory (M&MKT), Porto Ercole (Grosseto, Italy), pages 177–216, 2
- Mohammed Lemou and Luc Mieussens. A new asymptotic preserving scheme based on micro-macro formulation for linear kinetic equations in the diffusion limit. SIAM Journal on Scientific Computing, 31(1):334–368, 20
- Shi Jin, Zheng Ma, and Keke Wu. Asymptotic-preserving neural networks for multiscale time-dependent linear transport equations. Journal of Scientific Computing, 94(3):57, 20