Micro-Macro Neural galerkin method for kinetic equations

H. Ellabib

July 16, 2024

Abstract

This work is devoted to present, a Micro-Macro Neural Galerking method to solve some kinetic equations and more specifically the M1 model for radiative transfer. To do so, we first introduce the general method to reformulate the initial equation into a Micro-Macro model. We proceed by projecting the solution onto a vector space in a first hand, in another hand by applying a formulation in term of the moments of the solution. In the second section, we give some examples of equations that can be solved using the approach introduce in part 1. In the third part is the main one. Indeed, we present the M1 model in radiative transfer and we apply the method to this equation. The last part is dedicated to the presentation of the numerical scheme used to solve the M1 model.

Contents

1 Introduction

We want to solve the following type of equations :

$$\partial_t f(t,\boldsymbol{x},\boldsymbol{v}) + \boldsymbol{v}\cdot\nabla_x f(t,\boldsymbol{x},\boldsymbol{v}) + G(x)\cdot\nabla_v f(t,\boldsymbol{x},\boldsymbol{v}) = \sigma\left(M_{eq}(\boldsymbol{U}(t,\boldsymbol{x}),\boldsymbol{v}) - f(t,\boldsymbol{x},\boldsymbol{v})\right) \tag{1}$$

with $\boldsymbol{x}\in\mathbb{R}^d$, $\boldsymbol{v}\in\Omega\subset\mathbb{R}^d$. In this case we fix $\sigma\in\mathbb{R}$ and $M_{eq}(\boldsymbol{U}(t,\boldsymbol{x}),\boldsymbol{v})$ is the equilibrium distribution which is given. We aim to put (1) in the form of a moment model,

$$\partial_t \boldsymbol{U}(t,\boldsymbol{x}) + \nabla\cdot\boldsymbol{F}(\boldsymbol{U}) = \boldsymbol{S}(\boldsymbol{U}) \tag{2}$$

where,

$$\begin{aligned}
\boldsymbol{U}(t,x) &= \int_\Omega \boldsymbol{m}(\boldsymbol{v}) f(t,\boldsymbol{x},\boldsymbol{v}) d\boldsymbol{v}\\
\boldsymbol{F}(\boldsymbol{U}(t,x)) &= \int_\Omega \boldsymbol{m}(\boldsymbol{v})\boldsymbol{v} M(\boldsymbol{U}(t,\boldsymbol{x}),\boldsymbol{v}) d\boldsymbol{v}\\
\boldsymbol{S}(\boldsymbol{U}(t,x)) &= \int_\Omega \boldsymbol{m}(\boldsymbol{v})\nabla_v M(\boldsymbol{U}(t,\boldsymbol{x}),\boldsymbol{v}) d\boldsymbol{v}
\end{aligned}$$

We choose a specific distribution $M(\boldsymbol{U}(t,\boldsymbol{x}),\boldsymbol{v})$. The main difficulty is to choose in every equation that we want to solve a relevant form for M. We suppose that the solution can be written as :

$$f(t,\boldsymbol{x},\boldsymbol{v}) = M(\boldsymbol{U}(t,\boldsymbol{x}),\boldsymbol{v}) + g(t,\boldsymbol{x},\boldsymbol{v}) \tag{3}$$

with $\displaystyle\int_\Omega \boldsymbol{m}(\boldsymbol{v}) g(t,\boldsymbol{x},\boldsymbol{v}) d\boldsymbol{v} = 0$.

We introduce the space $L^2(M^{-1}d\boldsymbol{v})$ which is the space associated to the scalar product,

$$\langle\phi,\psi\rangle_{L^2(M^{-1})} = \langle\phi,\psi M^{-1}\rangle_{L^2(\Omega)}$$

.

We then project orthogonally the equation (1) onto $\mathcal{N} = Span(m_0(\boldsymbol{v})M,\dots,m_k(\boldsymbol{v})M)$. We denote by Π_M this projection. We inject (3) in (1) we get (we here forgot all the variables inside the functions).

$$\partial_t(M+g) + \boldsymbol{v}\cdot\nabla_x(M+g) + G(x)\cdot\nabla_v(M+g) = \sigma\left(M_{eq}(\boldsymbol{U},\boldsymbol{v}) - M - g\right) \tag{4}$$

Notice that we have some properties. $M\in\mathcal{N}$ implies that $\Pi_M(M) = M$. Then we get that, $(I_d - \Pi_M)M = 0$. Also because $M\in\mathcal{N}$ and $f = M+g$, we then get that $\Pi_M(g) = 0$. Hence $(I_d-\Pi_M)g = g$ and by continuity of the operator $(I_d-\Pi_M)$, we have that $(I_d-\Pi_M)\partial_t g = \partial_t(I_d-\Pi_M)g = \partial_t g$ and $(I_d-\Pi_M)\partial_t M = 0$ for the same reason.

Multiplying (4) by the vector $\boldsymbol{m}(\boldsymbol{v})$ and integrating the equation with respect to $\boldsymbol{v}$ leads to :

$$\int_\Omega \boldsymbol{m}(\boldsymbol{v})\partial_t(M+g)d\boldsymbol{v} + \int_\Omega \boldsymbol{m}(\boldsymbol{v})\boldsymbol{v}\cdot\nabla_x(M+g)d\boldsymbol{v} + \int_\Omega \boldsymbol{m}(\boldsymbol{v})G(x)\cdot\nabla_v(M+g)d\boldsymbol{v} = \int_\Omega \boldsymbol{m}(\boldsymbol{v})\sigma\left(M_{eq}(\boldsymbol{U},\boldsymbol{v}) - M - g\right)d\boldsymbol{v}$$

$$\partial_t\int_\Omega \boldsymbol{m}(\boldsymbol{v})(M+g)d\boldsymbol{v} + \nabla_x\cdot\int_\Omega \boldsymbol{m}(\boldsymbol{v})\boldsymbol{v}(M+g)d\boldsymbol{v} + G(x)\cdot\int_\Omega \boldsymbol{m}(\boldsymbol{v})\nabla_v(M+g)d\boldsymbol{v} = \sigma\int_\Omega \boldsymbol{m}(\boldsymbol{v})\left(M_{eq}(\boldsymbol{U},\boldsymbol{v}) - M - g\right)d\boldsymbol{v}$$

$$\partial_t\boldsymbol{U} + \nabla_x\cdot\boldsymbol{F}(\boldsymbol{U}) + \nabla_x\cdot\langle\boldsymbol{m}(\boldsymbol{v})\boldsymbol{v},g\rangle + G(x)\cdot\boldsymbol{S}(\boldsymbol{U}) + G(x)\cdot\langle\boldsymbol{m}(\boldsymbol{v}),\nabla_v g\rangle = \sigma(\langle M_{eq}(\boldsymbol{U},\boldsymbol{v}),\boldsymbol{m}(\boldsymbol{v})\rangle - \boldsymbol{U})$$

$$\partial_t\boldsymbol{U} + \nabla_x\cdot\boldsymbol{F}(\boldsymbol{U}) + \nabla_x\cdot\langle\boldsymbol{m}(\boldsymbol{v})\boldsymbol{v},g\rangle + G(x)\cdot\langle\boldsymbol{m}(\boldsymbol{v}),\nabla_v g\rangle = -G(x)\cdot\boldsymbol{S}(\boldsymbol{U}) + \sigma(\langle M_{eq}(\boldsymbol{U},\boldsymbol{v}),\boldsymbol{m}(\boldsymbol{v})\rangle - \boldsymbol{U})$$

We now apply the operator $I_d - \Pi_M$ to the equation (4), using the properties that we mentionned just before we obtain :

$$(I_d - \Pi_M)\partial_t(M+g) + (I_d - \Pi_M)(\boldsymbol{v}\cdot\nabla_x(M+g)) + (I_d - \Pi_M)(G(x)\cdot\nabla_v(M+g)) = (I_d - \Pi_M)(\sigma\left(M_{eq}(\boldsymbol{U},\boldsymbol{v}) - M - g\right))$$

$$\iff \partial_t g + (I_d - \Pi_M)(\boldsymbol{v}\cdot\nabla_x g + G(x)\cdot\nabla_v g) + (I_d - \Pi_M)(\boldsymbol{v}\cdot\nabla_x M + G(x)\cdot\nabla_v M) = \sigma(I_d - \Pi_M)\left(M_{eq}(\boldsymbol{U},\boldsymbol{v}) - M - g\right)$$

$$\iff \partial_t g + (I_d - \Pi_M)\mathcal{T} g = \sigma(I_d - \Pi_M)(M_{eq}(\boldsymbol{U},\boldsymbol{v}) - g) - (I_d - \Pi_M)\mathcal{T} M.$$

Where $\mathcal{T} = \boldsymbol{v}\cdot\nabla_x + G(x)\cdot\nabla_v$ is the transport operator.
We then have the following system to solve :

$$\begin{cases} \partial_t \boldsymbol{U} + \nabla_x\cdot\boldsymbol{F}(\boldsymbol{U}) + \nabla_x\cdot\langle\boldsymbol{m}(\boldsymbol{v})\boldsymbol{v}, g\rangle + G(x)\cdot\langle\boldsymbol{m}(\boldsymbol{v}), \nabla_v g\rangle = -G(x)\cdot\boldsymbol{S}(\boldsymbol{U}) + \sigma(\langle M_{eq}(\boldsymbol{U},\boldsymbol{v}), \boldsymbol{m}(\boldsymbol{v})\rangle - \boldsymbol{U}) \\ \partial_t g + (I_d - \Pi_M)\mathcal{T} g = \sigma(I_d - \Pi_M)(M_{eq}(\boldsymbol{U},\boldsymbol{v}) - g) - (I_d - \Pi_M)\mathcal{T} M \end{cases} \tag{5}$$

We now have a Micro-Macro model for the equation (1). In the system, the part with $\boldsymbol{U}$ is the Macro part that can be solve with classic method and the equation which relates to g is the Micro part that we will solve in the M1 problem using neural network.

2 Examples

2.1 Radiative transfer equation with diffusion reduced dynamic

We aim to solve the following equation,

$$\partial_t I(t,\boldsymbol{x},\boldsymbol{v}) + \boldsymbol{v}\cdot\nabla_x I(t,\boldsymbol{x},\boldsymbol{v}) = \sigma(E - I)$$

Here we suppose that, $U_0(t,\boldsymbol{x}) = E(t,\boldsymbol{x}) = \int_\Omega m_0(\boldsymbol{v}) I(t,\boldsymbol{x},\boldsymbol{v}) d\boldsymbol{v}$. $m_0(\boldsymbol{v}) = 1$, $\Omega = \mathbb{S}^{d-1}$. The equilibrium distribution is given by $M_{eq}(E(t,\boldsymbol{x}),\boldsymbol{v}) = E(t,\boldsymbol{x})$ and we choose to set $M(E(t,\boldsymbol{x}),\boldsymbol{v}) = E(t,x)$. We also have that $G(x) = 0$. In this case, $\mathcal{N} = Span(M)$. If we orthogonalize the basis of $\mathcal{N}$ we have that $\mathcal{N} = Span\left(\frac{1}{\langle M\rangle}M\right) = Span\left(\frac{1}{\sqrt{E\lambda\left(\mathbb{S}^{d-1}\right)}}M\right)$. We then express the projector in the orthogonal basis we get that,

$$\begin{aligned}\Pi_M(\phi) &= \left\langle \phi, \frac{1}{\sqrt{E\lambda\left(\mathbb{S}^{d-1}\right)}}M \right\rangle_{L^2(M^{-1}d\boldsymbol{v})} \frac{1}{\sqrt{E\lambda\left(\mathbb{S}^{d-1}\right)}}M \\ &= \frac{1}{\lambda\left(\mathbb{S}^{d-1}\right)}\langle\phi\rangle.\end{aligned}$$

We also have that,

$$\begin{aligned}F(U_0(t,\boldsymbol{x})) &= \int_\Omega \boldsymbol{v} M(E(t,\boldsymbol{x}),\boldsymbol{v}) d\boldsymbol{v} \\ &= E(t,\boldsymbol{x})\int_\Omega \boldsymbol{v} d\boldsymbol{v} = 0\end{aligned}$$

Hence, considering all the results in the first section and the fact that in this case, $\mathcal{T} M = \boldsymbol{v}\cdot\nabla_x E$ and $\Pi_M \mathcal{T} M = 0$, the model can be written as,

$$\begin{cases} \partial_t E + \nabla_x\cdot\langle\boldsymbol{v} g\rangle = 0 \\ \partial_t g + (I_d - \Pi_M)\mathcal{T} g = -\sigma g + \boldsymbol{v}\cdot\nabla_x E \end{cases} \tag{6}$$

2.2 Radiative transfert with P1 reduce dynamic

The equation that we want to solve is still the same as in the first example,

$$\partial_t I(t,\boldsymbol{x},\boldsymbol{v}) + \boldsymbol{v}\cdot\nabla_x I(t,\boldsymbol{x},\boldsymbol{v}) = \sigma(E-I)$$

In this case we still have, $\Omega = \mathbb{S}^{d-1}$ and :

$$U_0(t,\boldsymbol{x}) = E(t,x) = \int_\Omega m_0(\boldsymbol{v}) I(t,\boldsymbol{x},\boldsymbol{v}) d\boldsymbol{v},\ m_0(\boldsymbol{v}) = 1$$

$$U_1(t,\boldsymbol{x}) = \boldsymbol{F}(t,\boldsymbol{x}) = \int_\Omega m_1(\boldsymbol{v}) I(t,\boldsymbol{x},\boldsymbol{v}) d\boldsymbol{v},\ m_1(\boldsymbol{v}) = \boldsymbol{v}.$$

The equilibrium distribution is given by $M_{eq}(E(t,\boldsymbol{x}),\boldsymbol{v}) = E(t,\boldsymbol{x})$. We choose the distribution in the antzat,

$$M(E(t,\boldsymbol{x}),\boldsymbol{v}) = E(t,\boldsymbol{x}) + \boldsymbol{v}\cdot\boldsymbol{F}.$$

Where $\cdot$ denote the euclidean scalar product of $\mathbb{R}^d$ and $\|\cdot\|$ its associated norm. The space on which we want to project is, $\mathcal{N} = Span(1,\boldsymbol{v})$. By orthogonalizing the basis of $\mathcal{N}$ for the inner product of $L^2(\Omega)$ we get, $\mathcal{N} = Span\left(\frac{1}{\lambda\left(\mathbb{S}^{d-1}\right)}, \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{L^2}}\right)$. And, in the P1 method, we can write $\|\boldsymbol{v}\|$ in term of the presure tensor $\boldsymbol{P}$. In fact, this tensor is define as the second moment of the distribution M i.e,

$$\boldsymbol{P} = \int_\Omega \boldsymbol{v}\otimes\boldsymbol{v} M d\boldsymbol{v}.$$

In this case, there exist a simple expression for $\boldsymbol{P}$ which is the following :

$$\boldsymbol{P} = \frac{E}{3} I_d.$$

But, if we notice that $\|\boldsymbol{v}\|_{L^2}^2 = \langle\|\boldsymbol{v}\|^2\rangle = \langle \mathrm{Tr}(\boldsymbol{v}\otimes\boldsymbol{v})\rangle$ we have,

$$\begin{aligned}
\|\boldsymbol{v}\|_{L^2}^2 &= \langle \mathrm{Tr}(\boldsymbol{v}\otimes\boldsymbol{v})\rangle \\
&= \mathrm{Tr}(\langle \boldsymbol{v}\otimes\boldsymbol{v}\rangle) \\
&= \mathrm{Tr}\left(\int_\Omega \boldsymbol{v}\otimes\boldsymbol{v} d\boldsymbol{v}\right) \\
&= \mathrm{Tr}\left(\frac{1}{3} I_d\right) \\
&= \frac{d}{3}
\end{aligned}$$

We then get that $\|\boldsymbol{v}\|_{L^2} = \sqrt{\frac{d}{3}}$. Then in the orthonormal basis, the orthogonal projector is given by :

$$\begin{aligned}
\Pi_M(\phi) &= \left\langle \phi, \frac{1}{\lambda\left(\mathbb{S}^{d-1}\right)}\right\rangle_{L^2} \frac{1}{\lambda\left(\mathbb{S}^{d-1}\right)} + \left\langle \phi, \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{L^2}}\right\rangle_{L^2} \cdot \frac{\boldsymbol{v}}{\|\boldsymbol{v}\|_{L^2}} \\
&= \frac{1}{\lambda\left(\mathbb{S}^{d-1}\right)^2}\langle\phi\rangle + \frac{3}{d}\,\boldsymbol{v}\cdot\langle\phi\boldsymbol{v}\rangle.
\end{aligned}$$

We notice that we also have :

$$\begin{aligned}
\boldsymbol{F}(U_0(t,\boldsymbol{x})) &= \int_\Omega m_0(\boldsymbol{v})\boldsymbol{v} M(E(t,\boldsymbol{x}),\boldsymbol{v}) d\boldsymbol{v} \\
&= \int_\Omega \boldsymbol{v}(E(t,\boldsymbol{x}) + \boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}) d\boldsymbol{v} \\
&= \int_\Omega \boldsymbol{v} E(t,\boldsymbol{x}) d\boldsymbol{v} + \int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x})) d\boldsymbol{v}.
\end{aligned}$$

We can compute, then :

$$\int_\Omega \boldsymbol{v} E(t,\boldsymbol{x}) d\boldsymbol{v} = E(t,\boldsymbol{x}) \int_\Omega \boldsymbol{v} d\boldsymbol{v} = 0$$

and,

$$\int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}))d\boldsymbol{v} = \boldsymbol{F}(t,\boldsymbol{x})$$

Indeed, if we denote A the expression $A = \int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}))d\boldsymbol{v} - \boldsymbol{F}(t,\boldsymbol{x})$. We get that,

$$\begin{aligned}
A &= \int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}))d\boldsymbol{v} - \boldsymbol{F}(t,\boldsymbol{x}) \\
&= \int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}))d\boldsymbol{v} - \int_\Omega \boldsymbol{v} I(t,\boldsymbol{x},\boldsymbol{v})d\boldsymbol{v} \\
&= \int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}) - I(t,\boldsymbol{x},\boldsymbol{v}))d\boldsymbol{v} \\
&= \int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}) - M(E(t,\boldsymbol{x}),\boldsymbol{v}) - g(t,\boldsymbol{x},\boldsymbol{v}))d\boldsymbol{v} \\
&= \int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}) - M(E(t,\boldsymbol{x}),\boldsymbol{v}))d\boldsymbol{v}, \ \left(\text{because } \int_\Omega \boldsymbol{v} g(t,\boldsymbol{x},\boldsymbol{v})d\boldsymbol{v} = 0\right) \\
&= \int_\Omega \boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}) - E(t,\boldsymbol{x}) - \boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}))d\boldsymbol{v} \\
&= -\int_\Omega \boldsymbol{v} E(t,\boldsymbol{x})d\boldsymbol{v} = 0.
\end{aligned}$$

The next step is to make the same calculation for U_1. Thus, we have :

$$\begin{aligned}
\boldsymbol{F}(U_1(t,\boldsymbol{x})) &= \int_\Omega m_1(\boldsymbol{v})\boldsymbol{v} M(E(t,\boldsymbol{x}),\boldsymbol{v})d\boldsymbol{v} \\
&= \int_\Omega \boldsymbol{v}\otimes\boldsymbol{v}(E(t,\boldsymbol{x}) + \boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}))d\boldsymbol{v} \\
&= \int_\Omega \boldsymbol{v}\otimes\boldsymbol{v} E(t,\boldsymbol{x})d\boldsymbol{v} + \int_\Omega \boldsymbol{v}\otimes\boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}))d\boldsymbol{v}.
\end{aligned}$$

But, we know that the function, $\varphi(\boldsymbol{v}) = \boldsymbol{v}\otimes\boldsymbol{v}(\boldsymbol{v}\cdot\boldsymbol{F}(t,\boldsymbol{x}))$ is odd (using the properties of the tensor and the dot product). Then we get that :

$$\int_\Omega \boldsymbol{v}\otimes\boldsymbol{v}(\boldsymbol{v},\boldsymbol{F}(t,\boldsymbol{x}))d\boldsymbol{v} = 0.$$

Which leads to,

$$\begin{aligned}
\boldsymbol{F}(U_1(t,\boldsymbol{x})) &= E(t,\boldsymbol{x})\int_\Omega \boldsymbol{v}\otimes\boldsymbol{v}d\boldsymbol{v} \\
&= \frac{E(t,\boldsymbol{x})}{3} I_d
\end{aligned}$$

Just before we write the final micro-macro model, notice that we can make some simplication. Indeed, in this case we have $\langle \boldsymbol{v} g\rangle = 0$ and $(I_d - \Pi_M)E = 0$. But can also compute the operator $(I_d - \Pi_M)\mathcal{T}M$. Here we have $\mathcal{T}M = \boldsymbol{v}\cdot\nabla_x M = \boldsymbol{v}\cdot\nabla_x E$. It leads to :

$$
\begin{aligned}
\Pi_M \mathcal{T} M &= \frac{1}{\lambda\left(\mathbb{S}^{d-1}\right)^2}\langle \boldsymbol{v}\cdot\nabla_x M\rangle + \frac{3}{d}\,\boldsymbol{v}\cdot\langle(\boldsymbol{v}\cdot\nabla_x M)\boldsymbol{v}\rangle \\
&= \frac{1}{\lambda\left(\mathbb{S}^{d-1}\right)^2}\langle \boldsymbol{v}\rangle\cdot\nabla_x E + \frac{3}{d}\,\boldsymbol{v}\cdot\langle \boldsymbol{v}(\nabla_x E\cdot\boldsymbol{v})\rangle \\
&= \frac{3}{d}\,\boldsymbol{v}\cdot\langle \boldsymbol{v}\boldsymbol{v}^T\nabla_x E\rangle \\
&= \frac{3}{d}\,\boldsymbol{v}\cdot\langle \boldsymbol{v}\otimes\boldsymbol{v}\rangle\nabla_x E \\
&= \frac{3}{d}\,\boldsymbol{v}\cdot\frac{1}{3}I_d\nabla_x E \\
&= \frac{1}{d}\,\boldsymbol{v}\cdot\nabla_x E.
\end{aligned}
$$

Hence,

$$
(I_d - \Pi_M)\mathcal{T} M = \frac{d-1}{d}\,\boldsymbol{v}\cdot\nabla_x E \tag{7}
$$

Finally, we obtain the following model

$$
\begin{cases}
\partial_t E + \nabla_x\cdot\boldsymbol{F} = 0 \\
\partial_t \boldsymbol{F} + \nabla_x\cdot\frac{E}{3}I_d + \nabla_x\cdot\langle \boldsymbol{v}\otimes\boldsymbol{v}g\rangle = -\sigma\boldsymbol{F} \\
\partial_t g + (I_d - \Pi_M)\,\mathcal{T} g = -\sigma g + \dfrac{d-1}{d}\,\boldsymbol{v}\cdot\nabla_x E
\end{cases} \tag{8}
$$

2.3 Vlassov-Poisson equation

We consider the following equation :

$$
\partial_t f(t,\boldsymbol{x},\boldsymbol{v}) + \boldsymbol{v}\cdot\nabla_x f(t,\boldsymbol{x},\boldsymbol{v}) = \frac{1}{\varepsilon}Q(f,f) \tag{9}
$$

We now consider $\Omega = \mathbb{R}^d$, the vector $\boldsymbol{m} = (1,\boldsymbol{v})^T$ the function $\rho(t,\boldsymbol{x})$ which is the mean of f over the domain Ω, $\rho(t,\boldsymbol{x}) = \displaystyle\int_\Omega f(t,\boldsymbol{x},\boldsymbol{v})d\boldsymbol{v}$. $\boldsymbol{U}$ is define as the vector of the first moments of f,

$$
\boldsymbol{U} = \int_\Omega \begin{pmatrix}1\\ \boldsymbol{v}\end{pmatrix} f(t,\boldsymbol{x},\boldsymbol{v})d\boldsymbol{v} = \begin{pmatrix}\rho\\ \rho u\end{pmatrix}. \tag{10}
$$

Here, $\mathcal{N} = Span(M, \boldsymbol{v}M)$. In this method, we obtain M as the minimizer of the system entropy such that M as the same moments of the solution f. Then, M has now the following form :

$$
M(\boldsymbol{U})(\boldsymbol{v}) = \frac{\rho}{(2\pi T)^{\frac{d}{2}}}\exp\left(-\frac{|\boldsymbol{v}-u|^2}{2T}\right) \tag{11}
$$

Where ρ, u and T are density, mean velocity and temperature associated to $\boldsymbol{U}$ by the relation between moments :

$$
\langle mM(\boldsymbol{U})\rangle = \boldsymbol{U}. \tag{12}
$$

In fact, for the first coefficient we have :

$$
\int_\Omega M(\boldsymbol{U})(\boldsymbol{v})d\boldsymbol{v} = \frac{\rho}{(2\pi T)^{\frac{d}{2}}}\int_\Omega \exp\left(-\frac{|\boldsymbol{v}-u|^2}{2T}\right)d\boldsymbol{v} = \rho
$$

For the second one we get :

$$\begin{aligned}\int_\Omega \boldsymbol{v}M(\boldsymbol{U})(\boldsymbol{v})d\boldsymbol{v} &= \frac{\rho}{(2\pi T)^{\frac{d}{2}}}\int_\Omega \boldsymbol{v}\exp\left(-\frac{|\boldsymbol{v}-u|^2}{2T}\right)d\boldsymbol{v}\\ &= \frac{\rho}{(2\pi T)^{\frac{d}{2}}}\left(\int_\Omega(\boldsymbol{v}-u)\exp\left(-\frac{|\boldsymbol{v}-u|^2}{2T}\right)d\boldsymbol{v}+\int_\Omega u\exp\left(-\frac{|\boldsymbol{v}-u|^2}{2T}\right)d\boldsymbol{v}\right)\\ &= \frac{\rho u}{(2\pi T)^{\frac{d}{2}}}\int_\Omega\exp\left(-\frac{|\boldsymbol{v}-u|^2}{2T}\right)d\boldsymbol{v}\\ &= \rho u\end{aligned}$$

which is the wanted result. Now, we have to find the orthogonal projector onto $\mathcal{N}$. To do so, we orthonormalize the basis (M, vM) using Gram-Schimdt algorithm. First we set,

$$e_1 = \frac{M}{\|M\|_{L^2(M^{-1})}}$$

Then we compute,

$$\|M\|^2_{L^2(M^{-1})} = \langle M, M\rangle_{L^2(M^{-1})} = \langle M\rangle = \rho.$$

It leads to, $e_1 = \dfrac{M}{\sqrt{\rho}}$. The second vector is then define by $e_2 = \dfrac{b}{\|b\|_{L^2(M^{-1})}}$, where b is define by :

$$\begin{aligned}b &= \boldsymbol{v}M - \langle \boldsymbol{v}M, e_1\rangle_{L^2(M^{-1})}e_1\\ &= \boldsymbol{v}M - \frac{1}{\rho}\langle \boldsymbol{v}M, M\rangle_{L^2(M^{-1})}M\\ &= \boldsymbol{v}M - \frac{1}{\rho}\langle \boldsymbol{v}M\rangle M\\ &= \boldsymbol{v}M - \frac{1}{\rho}\rho u M\\ &= (\boldsymbol{v}-u)M\end{aligned}$$

We now need to compute the norm of b,

$$\begin{aligned}\|b\|^2_{L^2(M^{-1})} &= \langle(\boldsymbol{v}-u)M, (\boldsymbol{v}-u)M\rangle_{L^2(M^{-1})}\\ &= \langle(\boldsymbol{v}-u)M, (\boldsymbol{v}-u)\rangle_{L^2}\\ &= \langle(\boldsymbol{v}-u)\cdot(\boldsymbol{v}-u)M\rangle\end{aligned}$$

But the pressure tensor is define, as $\boldsymbol{P} = \langle(\boldsymbol{v}-u)\otimes(\boldsymbol{v}-u)M\rangle$. In our case we have $\boldsymbol{P} = \rho T$. Then we just define the vector $e_2 = \dfrac{b}{\sqrt{\rho T}}$ We finally have the expression of the projection Π_M for every φ :

$$\Pi_M(\varphi) = \langle\varphi, e_1\rangle_{L^2(M^{-1})}e_1 + \langle\varphi, e_2\rangle_{L^2(M^{-1})}e_2.$$

Hence, $\mathcal{N} = Span(e_1, e_2)$ and for every φ we have an analytical formula for the orthogonal projection onto $\mathcal{N}$:

$$\Pi_M(\varphi) = \frac{1}{\rho}\left(\langle\varphi\rangle + \frac{(\boldsymbol{v}-u)\cdot\langle(\boldsymbol{v}-u)\varphi\rangle}{T}\right)M. \tag{13}$$

3 Radiative transfer equation with M1 model

We aim to solve :

$$\partial_t I(t,\boldsymbol{x},\boldsymbol{v}) + \boldsymbol{v}\cdot\nabla_x I(t,\boldsymbol{x},\boldsymbol{v}) = \sigma(E(t,\boldsymbol{x}) - I(t,\boldsymbol{x},\boldsymbol{v}))$$

With, $\Omega = \mathbb{S}^{d-1}$ and,

$$U_0(t,x) = E(t,x) = \int_\Omega m_0(\boldsymbol{v}) I(t,\boldsymbol{x},\boldsymbol{v}) d\boldsymbol{v},\ m_0(\boldsymbol{v}) = 1$$
$$U_1(t,x) = \boldsymbol{F}(t,x) = \int_\Omega m_1(\boldsymbol{v}) I(t,\boldsymbol{x},\boldsymbol{v}) d\boldsymbol{v},\ m_1(\boldsymbol{v}) = \boldsymbol{v}.$$

The equilibrium distribution is given by : $M_{eq}(E(t,\boldsymbol{x}),\boldsymbol{v}) = E(t,\boldsymbol{x})$. We choose the distribution M in the antzat,

$$M(E(t,\boldsymbol{x}),\boldsymbol{F}(t,\boldsymbol{x}),\boldsymbol{v}) = \frac{2h\nu^3}{c^2}\left(\exp\left(\frac{h\nu}{kT_r^*}\left(1 - \frac{2-\sqrt{4-3\|\boldsymbol{f}\|^2}}{\|\boldsymbol{f}\|^2}(\boldsymbol{f},\boldsymbol{v})\right)\right) - 1\right)^{-1} \tag{14}$$

Where we use the notations $\boldsymbol{f} = \dfrac{\boldsymbol{F}}{cE}$,

$$T_r^* = \frac{2}{\|\boldsymbol{f}\|}\left(\sqrt{4-3\|\boldsymbol{f}\|^2} - 1\right)^{\frac{1}{4}}\sqrt{\|\boldsymbol{f}\|^2 - 2 + \sqrt{4-3\|\boldsymbol{f}\|^2}}T_r$$
$$\text{and } T_r = \left(\frac{E}{a}\right)^{\frac{1}{4}}$$

Notice that here, $\|\cdot\|$ refer to the ℓ^2 norm. This distribution is choosen by minimizing the entropy of the system under the constraint that M as the same moments as I, i.e :

$$M = \arg\min_{\varphi}\left\{\int_\Omega \left(\varphi\log\varphi - (\varphi+1)\log(\varphi+1)\right)d\boldsymbol{v}\,;\ \langle \boldsymbol{m}\varphi\rangle = \langle \boldsymbol{m}I\rangle\right\}.$$

Thus, by definition, we can use that $\langle M\rangle = E(t,\boldsymbol{x})$ and, $\langle \boldsymbol{v}M\rangle = \boldsymbol{F}(t,\boldsymbol{x})$. We aslo want to orthogonaly project onto the space $\mathcal{N} = Span(M, \boldsymbol{v}M)$. As usual, we othonormalize using Gram-Schmidt. $e_1 = \dfrac{M}{\|M\|_{L^2(M^{-1})}}$ and,

$$\begin{aligned}\|M\|^2_{L^2(M^{-1})} &= \langle M, M\rangle_{L^2(M^{-1})}\\ &= \langle M\rangle\\ &= E\end{aligned}$$

One, we get : $e_1 = \dfrac{M}{\sqrt{E}}$. Then, introducing the vector $b = \boldsymbol{v}M - \langle \boldsymbol{v}M, e_1\rangle_{L^2(M^{-1})}e_1$ which is orthognal to e_1. We obtain :

$$\begin{aligned}b &= \boldsymbol{v}M - \frac{1}{E}\langle \boldsymbol{v}, M\rangle_{L^2}M\\ &= \boldsymbol{v}M - \frac{1}{E}\langle \boldsymbol{v}M\rangle M\\ &= \boldsymbol{v}M - \frac{1}{E}\boldsymbol{F}M\\ &= (\boldsymbol{v} - c\boldsymbol{f})M.\end{aligned}$$

By construction, b is orthogonal with respect to e_1. We can compute to check :

$$\begin{aligned}
\langle e_1, b\rangle_{L^2(M^{-1})} &= \frac{1}{\sqrt{E}}\langle M, (\boldsymbol{v}-c\boldsymbol{f})M\rangle_{L^2(M^{-1})} \\
&= \frac{1}{\sqrt{E}}\langle M, \boldsymbol{v}-c\boldsymbol{f}\rangle_{L^2} \\
&= \frac{1}{\sqrt{E}}\left(\langle M, \boldsymbol{v}\rangle_{L^2} - c\langle M, \boldsymbol{f}\rangle_{L^2}\right) \\
&= \frac{1}{\sqrt{E}}\left(\boldsymbol{F} - c\boldsymbol{f}\langle M\rangle\right) \\
&= \frac{1}{\sqrt{E}}\left(\boldsymbol{F} - c\boldsymbol{f}E\right) \\
&= 0.
\end{aligned}$$

We now compute the norm of b :

$$\begin{aligned}
\|b\|^2_{L^2(M^{-1})} &= \langle b, b\rangle_{L^2(M^{-1})} \\
&= \langle (\boldsymbol{v}-c\boldsymbol{f})M, (\boldsymbol{v}-c\boldsymbol{f})M\rangle_{L^2(M^{-1})} \\
&= \langle (\boldsymbol{v}-c\boldsymbol{f})M, \boldsymbol{v}-c\boldsymbol{f}\rangle_{L^2} \\
&= \langle (\boldsymbol{v}-c\boldsymbol{f})\cdot(\boldsymbol{v}-c\boldsymbol{f})M\rangle \\
&= \left\langle \|\boldsymbol{v}-c\boldsymbol{f}\|^2 M\right\rangle.
\end{aligned}$$

But, we can rewrite the euclidean norm $\|\boldsymbol{v}-c\boldsymbol{f}\|^2$ in term of the trace of the Eddington tensor which is define as :

$$\boldsymbol{P} = \int_\Omega \boldsymbol{v}\otimes\boldsymbol{v}M(E(t,\boldsymbol{x}),\boldsymbol{F}(t,\boldsymbol{x}),\boldsymbol{v})d\boldsymbol{v}.$$

In the case of the M1 model, we have the following formula for $\boldsymbol{P}$,

$$\boldsymbol{P} = \frac{1}{2}\left((1-\chi(\boldsymbol{f}))I_d + (3\chi(\boldsymbol{f})-1)\frac{\boldsymbol{F}\otimes\boldsymbol{F}}{\|\boldsymbol{F}\|^2}\right)E.$$

Where $\chi(\boldsymbol{f})$ is the eddington factor, define by :

$$\chi(\boldsymbol{f}) = \frac{3+4\|\boldsymbol{f}\|^2}{5+2\sqrt{4-3\|\boldsymbol{f}\|^2}}.$$

We will use the following relation :

$$\|\boldsymbol{v}-c\boldsymbol{f}\|^2 = \mathrm{Tr}\left((\boldsymbol{v}-c\boldsymbol{f})\otimes(\boldsymbol{v}-c\boldsymbol{f})\right).$$

We now can compute,

$$(\boldsymbol{v}-c\boldsymbol{f})\otimes(\boldsymbol{v}-c\boldsymbol{f}) = \boldsymbol{v}\otimes\boldsymbol{v} - c(\boldsymbol{v}\otimes\boldsymbol{f}+\boldsymbol{f}\otimes\boldsymbol{v}) + c^2\boldsymbol{f}\otimes\boldsymbol{f}.$$

It leads to,

$$\begin{aligned}
\|\boldsymbol{v}-c\boldsymbol{f}\|^2 &= \mathrm{Tr}\left(\boldsymbol{v}\otimes\boldsymbol{v} - c(\boldsymbol{v}\otimes\boldsymbol{f}+\boldsymbol{f}\otimes\boldsymbol{v}) + c^2\boldsymbol{f}\otimes\boldsymbol{f}\right) \\
&= \mathrm{Tr}(\boldsymbol{v}\otimes\boldsymbol{v}) - c\left(\mathrm{Tr}(\boldsymbol{v}\otimes\boldsymbol{f}) + \mathrm{Tr}(\boldsymbol{f}\otimes\boldsymbol{v})\right) + c^2\mathrm{Tr}(\boldsymbol{f}\otimes\boldsymbol{f}) \\
&= \mathrm{Tr}(\boldsymbol{v}\otimes\boldsymbol{v}) - 2c\,\mathrm{Tr}(\boldsymbol{f}\otimes\boldsymbol{v}) + c^2\,\mathrm{Tr}(\boldsymbol{f}\otimes\boldsymbol{f})
\end{aligned}$$

the last line comming from the fact that $\boldsymbol{f}\otimes\boldsymbol{v}=\boldsymbol{f}\boldsymbol{v}^T=\left(\boldsymbol{v}\boldsymbol{f}^T\right)^T=(\boldsymbol{v}\otimes\boldsymbol{f})^T$. Hence $\mathrm{Tr}(\boldsymbol{v}\otimes\boldsymbol{f})=\mathrm{Tr}(\boldsymbol{f}\otimes\boldsymbol{v})$. Using that we can invert the integral and the Tr application we get :

$$\begin{aligned}
\|b\|^2_{L^2(M^{-1})} &= \langle \mathrm{Tr}((\boldsymbol{v}-c\boldsymbol{f})\otimes(\boldsymbol{v}-c\boldsymbol{f}))M\rangle \\
&= \mathrm{Tr}(\langle(\boldsymbol{v}-c\boldsymbol{f})\otimes(\boldsymbol{v}-c\boldsymbol{f})M\rangle) \\
&= \mathrm{Tr}(\langle(\boldsymbol{v}\otimes\boldsymbol{v}-c(\boldsymbol{v}\otimes\boldsymbol{f}+\boldsymbol{f}\otimes\boldsymbol{v})+c^2\boldsymbol{f}\otimes\boldsymbol{f})M\rangle) \\
&= \mathrm{Tr}(\langle\boldsymbol{v}\otimes\boldsymbol{v}M\rangle)-2c\,\mathrm{Tr}(\langle\boldsymbol{f}\otimes\boldsymbol{v}M\rangle)+c^2\,\mathrm{Tr}(\langle\boldsymbol{f}\otimes\boldsymbol{f}M\rangle) \\
&= \mathrm{Tr}(\boldsymbol{P})-2c\,\mathrm{Tr}(\boldsymbol{f}\otimes\langle\boldsymbol{v}M\rangle)+c^2\,\mathrm{Tr}(\boldsymbol{f}\otimes\boldsymbol{f}\langle M\rangle) \ \text{ (Because } \boldsymbol{f} \text{ only depends of } \boldsymbol{x} \text{ and } t) \\
&= \mathrm{Tr}(\boldsymbol{P})-2c\,\mathrm{Tr}(\boldsymbol{f}\otimes\boldsymbol{F})+Ec^2\,\mathrm{Tr}(\boldsymbol{f}\otimes\boldsymbol{f}) \\
&= \mathrm{Tr}(\boldsymbol{P})-\frac{2}{E}\,\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F})+\frac{1}{E}\,\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F}) \\
&= \mathrm{Tr}(\boldsymbol{P})-\frac{1}{E}\,\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F})
\end{aligned}$$

We can now calculate the trace of $\boldsymbol{P}$ in term of the Eddington factor and the first moment of I.

$$\begin{aligned}
\mathrm{Tr}(\boldsymbol{P}) &= \frac{E}{2}\mathrm{Tr}\left((1-\chi(\boldsymbol{f}))I_d+(3\chi(\boldsymbol{f})-1)\frac{\boldsymbol{F}\otimes\boldsymbol{F}}{\|\boldsymbol{F}\|^2}\right) \\
&= \frac{E}{2}d(1-\chi(\boldsymbol{f}))+\frac{E}{2\|\boldsymbol{F}\|^2}(3\chi(\boldsymbol{f})-1)\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F})
\end{aligned}$$

Putting together with the previous result, we have :

$$\begin{aligned}
\|b\|^2_{L^2(M^{-1})} &= \frac{E}{2}d(1-\chi(\boldsymbol{f}))+\frac{E}{2\|\boldsymbol{F}\|^2}(3\chi(\boldsymbol{f})-1)\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F})-\frac{1}{E}\,\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F}) \\
&= \frac{E}{2}d(1-\chi(\boldsymbol{f}))+\left(\frac{E}{2\|\boldsymbol{F}\|^2}(3\chi(\boldsymbol{f})-1)-\frac{1}{E}\right)\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F})
\end{aligned}$$

We then define e_2, the second vector of the orthonormal basis.

$$e_2=\frac{\boldsymbol{v}-c\boldsymbol{f}}{\sqrt{\frac{E}{2}d(1-\chi(\boldsymbol{f}))+\left(\frac{E}{2\|\boldsymbol{F}\|^2}(3\chi(\boldsymbol{f})-1)-\frac{1}{E}\right)\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F})}}M.$$

We can now compute the orthogonal projection onto $\mathcal{N}=Span(e_1,e_2)$, we obtain that for every φ :

$$\begin{aligned}
\Pi_M(\varphi) &= \langle\varphi,e_1\rangle_{L^2(M^{-1})}e_1+\langle\varphi,e_2\rangle_{L^2(M^{-1})}e_2 \\
&= \frac{\langle\varphi\rangle}{E}M+\frac{\langle\varphi,\boldsymbol{v}-c\boldsymbol{f}\rangle_{L^2}}{\frac{E}{2}d(1-\chi(\boldsymbol{f}))+\left(\frac{E}{2\|\boldsymbol{F}\|^2}(3\chi(\boldsymbol{f})-1)-\frac{1}{E}\right)\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F})}\cdot(\boldsymbol{v}-c\boldsymbol{f})M.
\end{aligned}$$

We can put this expression in the form :

$$\Pi_M(\varphi)=\left(\frac{\langle\varphi\rangle}{E}+(\boldsymbol{v}-c\boldsymbol{f})\cdot\frac{\langle\boldsymbol{v}\varphi\rangle-c\langle\varphi\rangle\boldsymbol{f}}{\frac{E}{2}d(1-\chi(\boldsymbol{f}))+\left(\frac{E}{2\|\boldsymbol{F}\|^2}(3\chi(\boldsymbol{f})-1)-\frac{1}{E}\right)\mathrm{Tr}(\boldsymbol{F}\otimes\boldsymbol{F})}\right)M. \tag{15}$$

We also have :

$$\boldsymbol{F}(U_0(t,\boldsymbol{x}))=\int_\Omega \boldsymbol{v}M(E(t,\boldsymbol{x}),\boldsymbol{F}(t,\boldsymbol{x}),\boldsymbol{v})d\boldsymbol{v}=\langle\boldsymbol{v}M\rangle.$$

But, by construction, M as the same moment of I. Hence, we have, $\langle\boldsymbol{v}M\rangle=\boldsymbol{F}$. Then, $\boldsymbol{F}(U_0(t,\boldsymbol{x}))=\boldsymbol{F}(t,\boldsymbol{x})$. And :

$$\boldsymbol{F}(U_1(t,\boldsymbol{x}))=\int_\Omega \boldsymbol{v}\otimes\boldsymbol{v}M(E(t,\boldsymbol{x}),\boldsymbol{F}(t,\boldsymbol{x}),\boldsymbol{v})d\boldsymbol{v}=\boldsymbol{P} \tag{16}$$

Finally, we can write the model :

$$\begin{cases} \partial_t E + \nabla_x \cdot \boldsymbol{F} = 0 \\ \\ \partial_t \boldsymbol{F} + \nabla_x \cdot \boldsymbol{P} + \nabla_x \cdot \langle \boldsymbol{v} \otimes \boldsymbol{v} g \rangle = -\sigma \boldsymbol{F} \\ \\ \partial_t g + (I_d - \Pi_M)\mathcal{T} g = \sigma((I_d - \Pi_M)E - g) - (I_d - \Pi_M)\mathcal{T} M \end{cases} \tag{17}$$

Now that we wrote a Micro-Macro model for the M1 problem, we want to solve it numerically, this is the purpose of the last section.

4 Numerical study of the problem

4.1 Scheme

Now that we presented the general theory and that we have the system of PDEs that we want to solve, we can discretize the system to obtain an approximation of the solution. We write on the general form of the system for the M1 problem :

$$\begin{cases} \partial_t \boldsymbol{U} + \nabla_x \cdot \boldsymbol{F}(\boldsymbol{U}) + \nabla_x \cdot \langle \boldsymbol{m}(\boldsymbol{v}) g \rangle = 0 \\ \\ \partial_t g + (I_d - \Pi_M)\mathcal{T} g = \sigma((I_d - \Pi_M)E - g) - (I_d - \Pi_M)\mathcal{T} M \end{cases} \tag{18}$$

Here, $\boldsymbol{U}$ is the macro part and g the micro one. The principle is to solve the macro part using finite difference method and to use a Neural Galerkin method for the kinetic part. We first write the discretize system that approximate the solution of 18,

$$\begin{cases} \boldsymbol{U}^{n+1} = \boldsymbol{U}^n - \Delta t \nabla \cdot \boldsymbol{F}(\boldsymbol{U}^n) - \Delta t \nabla_x \cdot \langle \boldsymbol{m}(\boldsymbol{v}) g^{n+1} \rangle = 0 \\ \\ g^{n+1} = g^n - \Delta t (I_d - \Pi_M)(\mathcal{T} g^n - \Delta t \sigma g^n + \Delta t (I_d - \Pi_M)\mathcal{T} M^n \end{cases} \tag{19}$$

Where $\boldsymbol{U}^{n+1}$ (resp. g^{n+1}) is the approximation of $\boldsymbol{U}(t_n, \boldsymbol{x}, \boldsymbol{v})$ (resp. $g(t_n, \boldsymbol{x}, \boldsymbol{v})$). We then approximate g^n by a neural network g_{θ_n}. So that we have for every integer n, $g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) \approx g^n(\boldsymbol{x}, \boldsymbol{v})$ where all the time t_n are choosed by discretizing the time domain. We are using Neural Galerkin method here, it means that to compute the new time we will find the solve the minimisation problem :

$$\min_{\theta_{n+1}} \int_{\mathbb{R}^d} \int_{\Omega} \left\| g_{\theta_{n+1}}(\boldsymbol{x}, \boldsymbol{v}) - g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) - \Delta t (I_d - \Pi_M)(\mathcal{T} g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) - \Delta t \sigma g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) + \Delta t (I_d - \Pi_M)\mathcal{T} M^n(\boldsymbol{x}, \boldsymbol{v})) \right\|^2 d\boldsymbol{x} d\boldsymbol{v}. \tag{20}$$

After this, we can linearize to obtain the approximation,

$$g_{\theta_{n+1}}(\boldsymbol{x}, \boldsymbol{v}) \approx g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) + \nabla_{\theta_n} g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) \cdot (\theta_{n+1} - \theta_n).$$

So we rewrite the minimasation problem as :

$$\min_{\theta_{n+1}} \int_{\mathbb{R}^d} \int_{\Omega} \left\| \nabla_{\theta_n} g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) \cdot (\theta_{n+1} - \theta_n) - \Delta t (I_d - \Pi_M)(\mathcal{T} g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) - \Delta t \sigma g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) + \Delta t (I_d - \Pi_M)\mathcal{T} M^n(\boldsymbol{x}, \boldsymbol{v})) \right\|^2 d\boldsymbol{x} d\boldsymbol{v}. \tag{21}$$

After this, we can calculate the gradient of the fonctionnal that we want to minimize with respect to θ_n and see in which θ_n it is vanishing. We obtain the following system of ODEs :

$$\mathcal{J}(\theta(t)) \frac{d\theta(t)}{dt} = f(\theta(t)). \tag{22}$$

Where $\theta(t)$ is the vector of all the parameters of the network at the time step t. Using explicit euler scheme leads to the linear system :

$$\mathcal{J}(\theta_n)\theta_{n+1} = \mathcal{J}(\theta_n)\theta_n + \Delta t f(\theta_n) \tag{23}$$

With,

$$\mathcal{J}(\theta_n) = \int_{\mathbb{R}^d} \int_{\Omega} \nabla_{\theta_n} g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) \otimes \nabla_{\theta_n} g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) d\boldsymbol{x} d\boldsymbol{v} \tag{24}$$

and,

$$f(\theta_n) = \int_{\mathbb{R}^d} \int_{\Omega} \nabla_{\theta_n} g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v})^T - \Delta t (I_d - \Pi_M)(\mathcal{T} g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) - \Delta t \sigma g_{\theta_n}(\boldsymbol{x}, \boldsymbol{v}) + \Delta t (I_d - \Pi_M) \mathcal{T} M^n(\boldsymbol{x}, \boldsymbol{v})) d\boldsymbol{x} d\boldsymbol{v}. \tag{25}$$

All the integral are approximate by Monte-Carlo method.

4.2 Tests and results

The purpose of this section is to show some numerical results that we obtained for solving the transfer radiative equation. Notice that for the rest of the document, we did not use the micro-macro method that we developped in the last section. Here, we approximate the whole solution of (18) by a network. We are only using Neural-Galerkin method. Here are the results. We are here interested in the following problem :

$$\begin{cases} \dfrac{\partial u}{\partial t}(t, \boldsymbol{x}, \boldsymbol{v}) + \boldsymbol{v} \cdot \nabla_x u(t, \boldsymbol{x}, \boldsymbol{v}) = 0 \\ u(0, \boldsymbol{x}, \boldsymbol{v}) = u_0(\boldsymbol{x}, \boldsymbol{v}) \end{cases} \tag{26}$$

Where $t \in [0, T[$, $\boldsymbol{x} \in \mathbb{R}^2$ and $\boldsymbol{v} \in \mathbb{S}^1 \subset \mathbb{R}^2$.
Using the method of characteristics, we know that the soluton of this equation can be written for every $t, \boldsymbol{x}$ and $\boldsymbol{v}$ as :

$$u(t, \boldsymbol{x}, \boldsymbol{v}) = u_0(\boldsymbol{x} - t\boldsymbol{v}, \boldsymbol{v}). \tag{27}$$

We did implement an approximation of the analytic solution using neural galerkin schemes. Here are the results for several networks. In all our tests, we choose to take u_0 as a product of two 2-dimensional normal distribution. Thus $u_0(x, v) = f_1(x) \cdot f_2(v)$, where f_1 (resp f_2) is a 2-dimensionnal normal distribution with mean μ_1 (resp μ_2) and covariance matrix Σ_1 (resp Σ_2). Notice that for every test, the code is running for 4000 epochs and we take 15000 points at each epochs. Also, except the size of the neural network and the learning rate, we only vary the covariance matrix of the second gaussian density f_2 that we aim to make tend towards a dirac mass so that the velocity vector is concentrated in one direction. We choose to take a sample of 3 covariance matrix define by :

$$\Sigma_v^{(1)} = 2\pi \frac{5}{10} I_2,\ \Sigma_v^{(2)} = 2\pi \frac{8}{10^2} I_2,\ \Sigma_v^{(3)} = 2\pi \frac{3}{10^2} I_2.$$

Where I_2 is the 2 dimensional identity matrix. The 2π factor is here to scale the variance of the velocity Gaussian which is define on the sphere $\mathbb{S}^1$ so that our network approximate well the solution. The neural Galerkin scheme consist in solving a system of ODEs for network parameters. We solve this system by using the forward Eueler scheme which is an explicit scheme. We did use a time step $\Delta t = 0,005$ and a final time $T = 0,2$ in all our test.
In every test we set μ_2, the mean of the normal distribution f_2 equal to $\mu_2 = (0, 1)$ so that we project the velocity in the direction μ_2. The first step will be to see what is happening when we look in the direction μ_2. If the network approximate well the analytic solution or not. The second step will be to see what is happening in a neighborhood of μ_2. Indeed, in theory the far we are getting of μ_2, the less we are getting mass transport. This is what we expect from the network.

First Network : We first choose to consider a neural-network with 4 hidden layers which are composed of 40 neurons each. We fix the learning rate of the initial condition to $3 \cdot 10^{-2}$. By training the initial condition for 4000 epochs with a sample of 15000 colocation points at each epoch, we obtain the following results :

Covariance matrix of the velocity	Best loss after 4000 epochs (15000 points per epoch)	Global error $\|u_\theta - u\|^2_{\ell^2}$
$\Sigma_v^{(1)}$	$1,11 \cdot 10^{-6}$	$1,79 \cdot 10^{-6}$
$\Sigma_v^{(2)}$	$4,80 \cdot 10^{-6}$	$1,85 \cdot 10^{-6}$
$\Sigma_v^{(3)}$	$7,68 \cdot 10^{-6}$	$2,21 \cdot 10^{-6}$

We can see that the first network approximate quite well the analytic solution with low global error. As the covariance of the velocity is getting lower, the network as some difficulties to approach the analytic solution

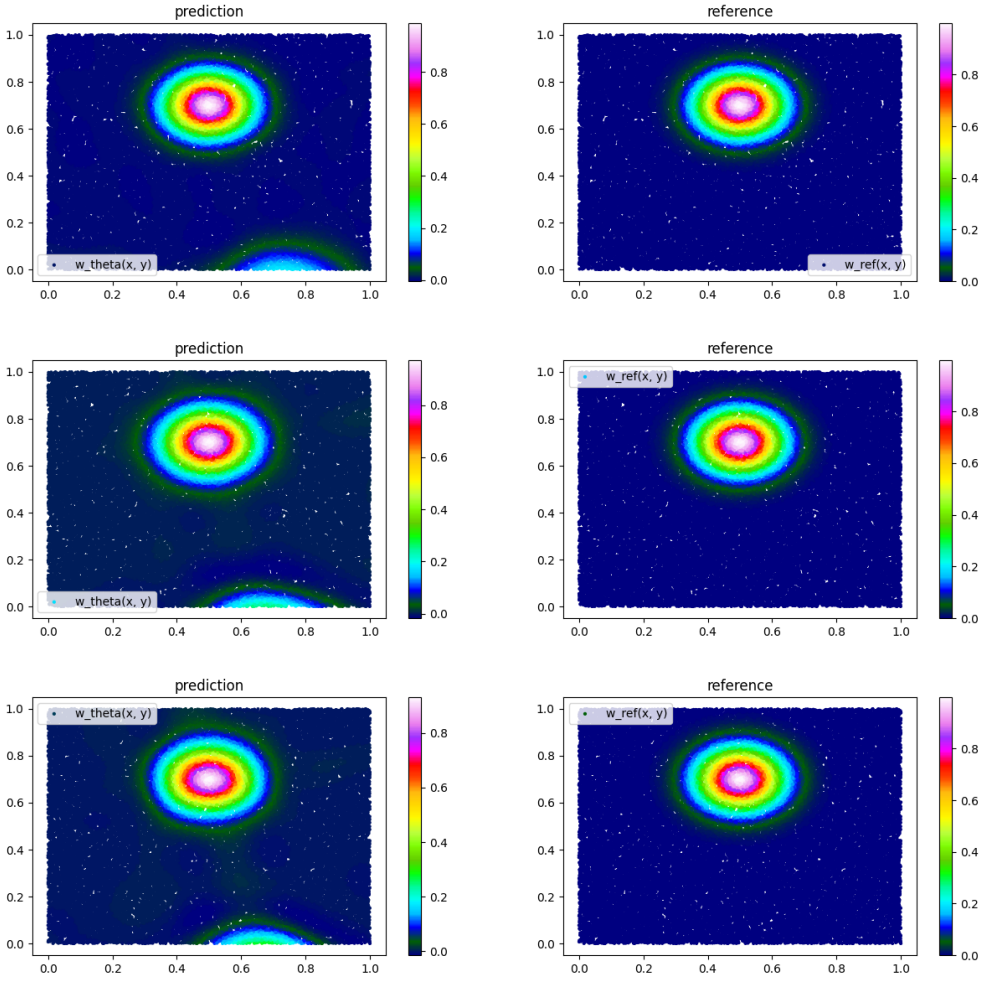

Figure 1: Results of the approximation of the transport equation with a neural network composed of 4 layers and 40 neurons in each layer when the covariance matrix of the velocity is $\Sigma_v^{(1)}$, $\Sigma_v^{(2)}$ and $\Sigma_v^{(3)}$ respectively.

which is normal but here there are no big diference between case and error are quite low.

Second network : We then consider another network. Here we propose a little bit more complex neural network with 4 hidden layers and each layer have 60 neurons. The learning rate is fixed to $2 \cdot 10^{-2}$. We then have the following results :

Covariance matrix of the velocity	best loss after 4000 epochs (15000 points per epoch)	Global error $\|u_\theta - u\|_{\ell^2}^2$
$\Sigma_v^{(1)}$	$1,91 \cdot 10^{-6}$	$5,35 \cdot 10^{-7}$
$\Sigma_v^{(2)}$	$3,43 \cdot 10^{-6}$	$8,68 \cdot 10^{-7}$
$\Sigma_v^{(3)}$	$7,79 \cdot 10^{-6}$	$9,01 \cdot 10^{-7}$

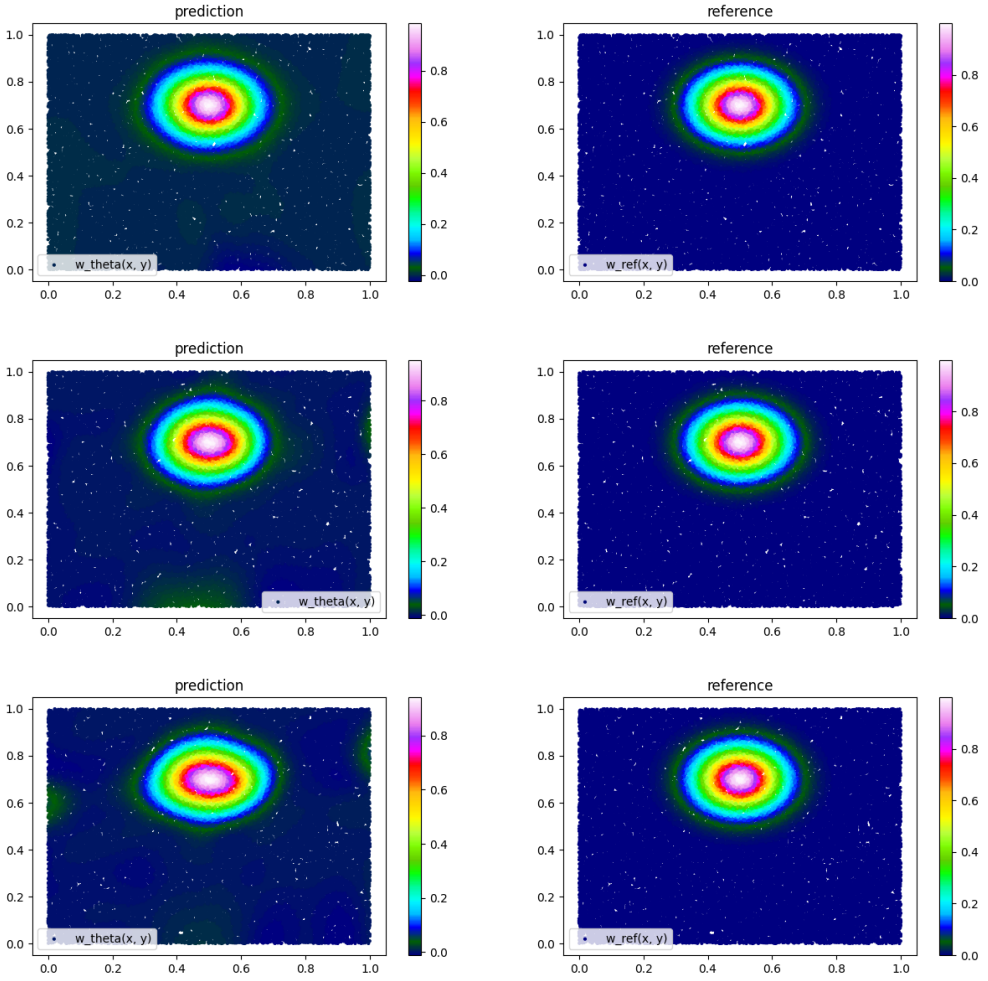

Figure 2: Results of the approximation of the transport equation with a neural network composed of 4 layers and 60 neurons in each layer when the covariance matrix of the velocity is $\Sigma_v^{(1)}$, $\Sigma_v^{(2)}$ and $\Sigma_v^{(3)}$ respectively.

As expected, because we put more neurons on each layers, we have much more parameters to approach so the approximation is much better with at least one order of magnitude lower for the global error. Compare to the first network, here we can notice that there is a difference between cases. In fact, the network has more difficulties to approximate the solution has the distribution of the velocity approaches a Dirac mass. However, the results are still satisfactory.

Third network : The last example that we consider will be a network with 6 hidden layers always with 26 neurons each. We decrease the learning rate again for this example, we set it at $1 \cdot 10^{-2}$. The results are also good as we can see :

Covariance matrix of the velocity	best loss after 4000 epochs (15000 points per epoch)	Global error $\|u_\theta - u\|_{\ell^2}^2$
$\Sigma_v^{(1)}$	$8,36 \cdot 10^{-6}$	$3,18 \cdot 10^{-5}$
$\Sigma_v^{(2)}$	$5,83 \cdot 10^{-6}$	$2,22 \cdot 10^{-7}$
$\Sigma_v^{(3)}$	$6,44 \cdot 10^{-6}$	$3,66 \cdot 10^{-7}$

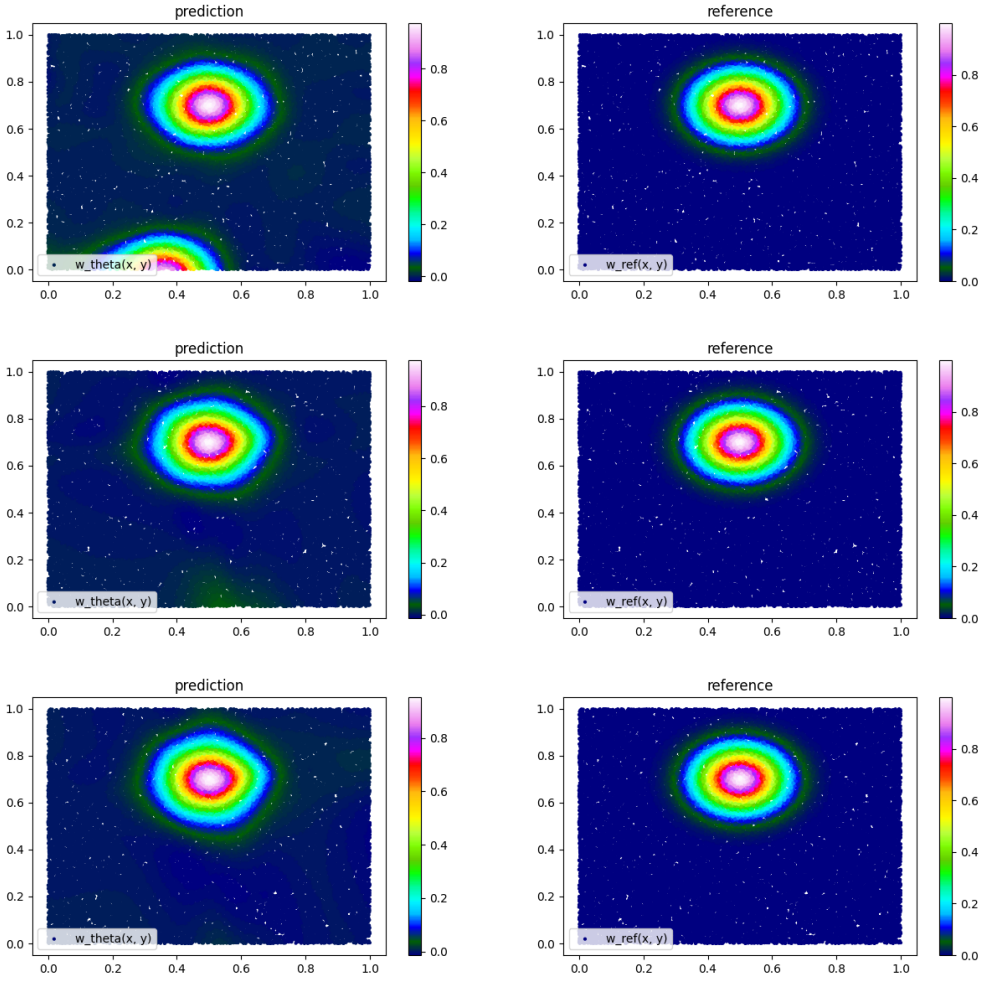

Figure 3: Results of the approximation of the transport equation with a neural network composed of 4 layers and 80 neurons in each layer when the covariance matrix of the velocity is $\Sigma_v^{(1)}$, $\Sigma_v^{(2)}$ and $\Sigma_v^{(3)}$ respectively.

Here we can make the same conclusion as the last test, we did increase the number of neurons in each layer so it implies to highly increase the number of parameters in the network then we get a better precision with this network. Even with a very low covariance $\Sigma_v^{(3)}$, the result is very good. It shows that neural network can approximate well the normal distribution when the velocity is very concentrated which is exactly what we want. We can notice big differences between approximation by the first and the third network, which is normal because the number of parameters in the third one is much higher.

Now that we have good results on our tests, want to see if the method keeps the intuition that when you propagate mostly in one direction there are less mass that sent on neighboring direction. Here, as in the first part of the test, we propagate the velocity in the direction $(0, 1)$. And we are watching to what is happens in other direction as we are moving away from $(0, 1)$. In all the following test, we are using the second network, ie a network with 4 layers and 60 neurons per layer. And we are using the covariance matrix $\Sigma_v^{(2)} = 2\pi\frac{8}{10^2}I_2$ so that the velocity is still concentrated in one point. Here are the results for several direction.

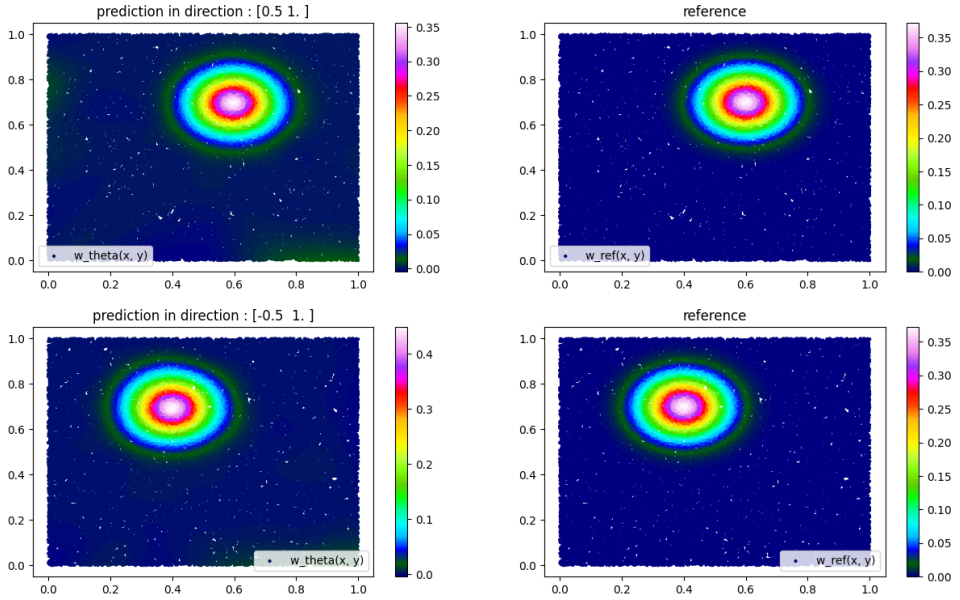

Figure 4: Approximation of the solution in the direction $(0.5, 1)$ and $(-0.5, 1)$ for the velocity.

Here are the plots for the case where we are looking for behave of the solution when the velocity is going to the direction $(0.5, 1)$ (the first picture) and $(-0.5, 1)$ (the second one). Basically what we are doing here is to watch a cone which is kind of projected onto the sphere and were look to what happen as we get to the extremal points of the cone. The result is that as we are getting far from the point where the veloicity is centered, the solution decrease this is exactly what is expected because the velocity follow a normal distribution. Also, the approximation has the same order of magnitued than the exact solution so the method is consistent. We can now look at more extremal directions.

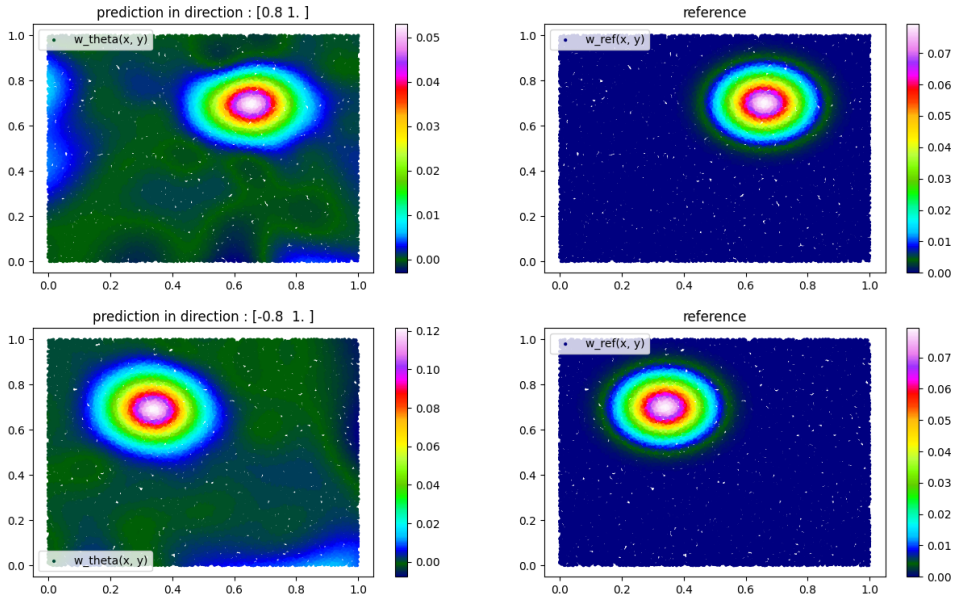

Figure 5: Approximation of the solution in the direction $(0.8, 1)$ and $(-0.8, 1)$ for the velocity.

As we can see, the approximation has the same order of magnitude than the exact solution, this is what is telling us that the approximation by the network behaves well. Indeed, at the peak of the solution, the function is at $0,07$ which is very low compare to the unit that appear in the solution when we are looking in the direction $(0, 1)$ for the velocity.

4.3 Comparison with classic method

We did implement a more classical method to show that Neural Galerkin method is much accurate and that there is a real interest in his use. The model that we consider here is a moments model define by,

$$u_\theta(t, \boldsymbol{x}, \boldsymbol{v}) = \alpha_\theta(t, \boldsymbol{x}) + \langle \beta_\theta(t, \boldsymbol{x}), \boldsymbol{v} \rangle \tag{28}$$

Where θ is the vector of all the parameters that we want to optimize using a neural network which is composed of 4 layers, with 60 neurons in each layer. For all those tests, we use the same parameter as we used in the with the second neural network in the second section. We show some result depending of the initial condition. The first case that we consider is the one where the initial condition is a gaussian in space and does not depend on v. It is the case where the variance of the velocity gaussian is infinite. Here are the results.

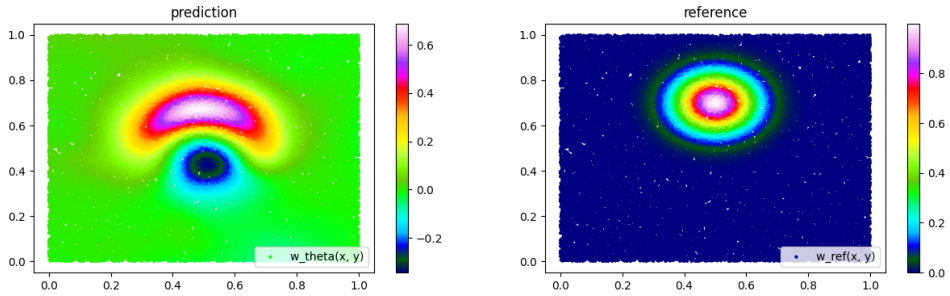

Figure 6: Approximation of the solution of (26) when the initial condition is a normal distribution that only depends on the space variable.

So here as we can see, the classic model does not approximate well the solution compare to the Nueral Galerkin one. However, the order of magnitude between the reference and the prediction is the same which is not bad. This was predictable because in this model, we are basically approximating by using a two order Taylor development. As the initial condition is written as $u_0(\boldsymbol{x}, \boldsymbol{v}) = f_1(\boldsymbol{x}) \cdot 1$, where f_1 is a normal distribution, it is easy to approximate the velocity part of u_0 using a two order Taylor developpement. It is not the case if we define u_0 as the product of two normal distribution as it is the case in the first part. Now, for u_0, we are taking the same form as previously, i.e $u_0(\boldsymbol{x}, \boldsymbol{v}) = f_1(\boldsymbol{x}) \cdot f_2(\boldsymbol{v})$ where f_1 and f_2 are two normal distribution and we vary the covariance matrix of f_2 taking $\Sigma_v^{(1)}$, $\Sigma_v^{(2)}$ and $\Sigma_v^{(3)}$. Here are the results of the approximation by the moments model.

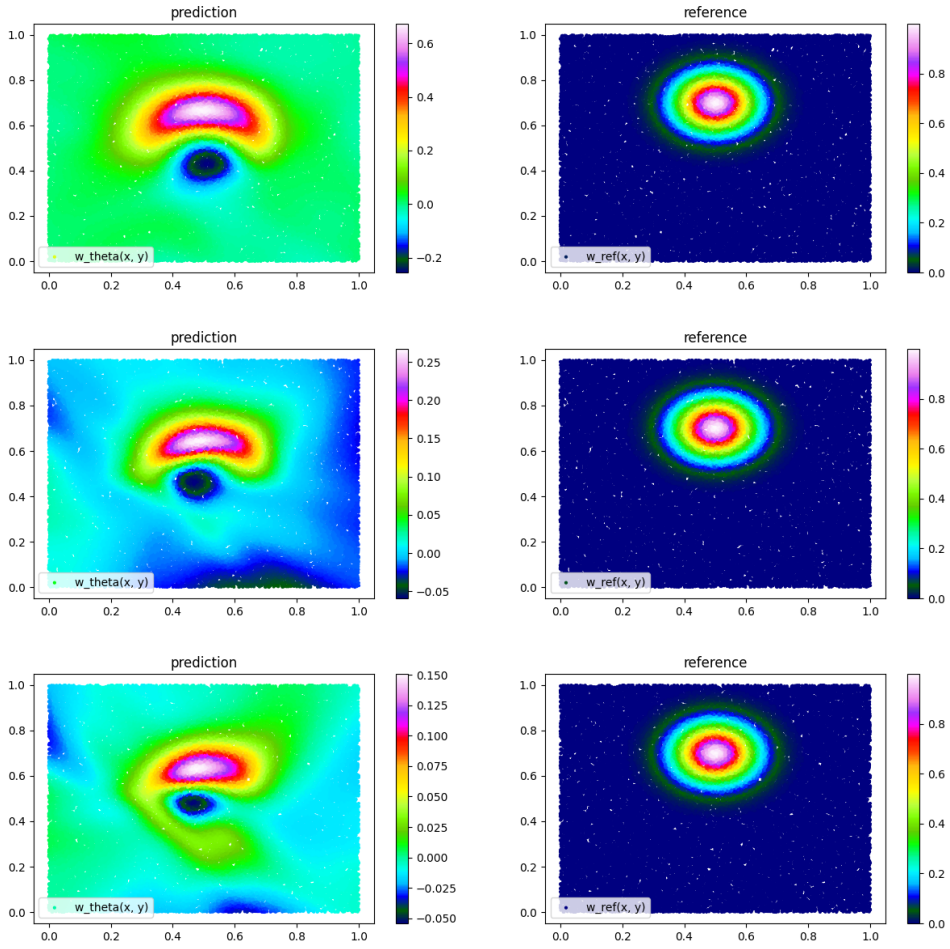

Figure 7: Approximation of the equation (26) using a moments model for the initial condition $u_0 = f_1 \cdot f_2$ when f_2 as for covariance matrix respectively $\Sigma_v^{(1)}$, $\Sigma_v^{(2)}$ and $\Sigma_v^{(3)}$.

As we can see all the approximation with moments model are not good compare to the one using Neural Galerkin model. The observation that we can make is that as the covariance of the velocity of the initial condition decrease, the approximation gets even worse. In fact, we can see that when the covariance matrix decrease, the order of magnitude of the approximation gets lower as the one of the reference do not. It is linked to the fact that as the covariance matrix decrease, f_2 tends to a dirac mass so it's much harder for a moments model because it is of the form of a Taylor development of order 1.