A DEEP LEARNING APPROXIMATION OF NON-STATIONARY SOLUTIONS TO WAVE KINETIC EQUATIONS*

Steven Walton[a,*], Minh-Binh Tran[b], Alain Bensoussan[c,d]

[a]*Southern Methodist University, Dallas, Texas, USA*
[b]*Texas A&M University, College Station, Texas, USA*
[c]*University of Texas Dallas, Dallas, Texas, USA*
[d]*City University of Hongkong, Hongkong*

Abstract

We present a deep learning approximation, stochastic optimization based, method for wave kinetic equations. To build confidence in our approach, we apply the method to a Smoluchowski coagulation equation with multiplicative kernel for which an analytic solution exists. Our deep learning approach is then used to approximate the non-stationary solution to a 3-wave kinetic equation corresponding to acoustic wave systems. To validate the neural network approximation, we compare the decay rate of the total energy with previously obtained theoretical results. A finite volume solution is presented and compared with the present method.

Keywords: Deep Learning, Function Approximation, Stochastic Optimization, Wave Turbulence, Partial Differential Equations, 3-Wave Equation

*S.W. was supported by U.S. National Science Foundation under the SMU Research Training Group grant DMS-1840260. M.-B. T. was funded in part by the NSF Grant DMS-1854453, Humboldt Fellowship, NSF CAREER DMS-2044626, NSF Grant DMS-2204795. A.B. was funded in part by the NSF Grant DMS-1905449, NSF Grant DMS-2204795 and a grant from the SAR Hong Kong RGC GRF 14301321.

*Corresponding author

Email addresses: stevenw@smu.edu (Steven Walton), minhbinh@tamu.edu (Minh-Binh Tran), alain.bensoussan@utdallas.edu, abensous@cityu.edu.hk (Alain Bensoussan)

1. Introduction

For more than 60 years, the theory of wave turbulence has been proved to have a wide range applications [1, 2, 3, 4, 5, 6, 7, 8]. Among the most important classes of wave kinetic equations (WKEs), 3-wave kinetic equations read

$$\begin{aligned} \partial_t f(t,k) &= Q_{3w}[f](t,k), \\ f(0,k) &= f_0(k), \end{aligned} \tag{1}$$

in which $f(t,k)$ is the nonnegative wave density at wavenumber $k \in \mathbb{R}^N$, $N \geq 2$; $f_0(k)$ is the initial condition. The quantity $Q[f]$ is of the form

$$Q_{3w}[f](k) = \iint_{\mathbb{R}^{2N}} \Big[\mathcal{R}_{k,k_1,k_2}[f] - \mathcal{R}_{k_1,k,k_2}[f] - \mathcal{R}_{k_2,k,k_1}[f]\Big] k^N k_1 d^N k_2, \tag{2}$$

with

$$\mathcal{R}_{k,k_1,k_2}[f] := |V_{k,k_1,k_2}|^2 \delta(k - k_1 - k_2)\delta(\omega - \omega_1 - \omega_2)(f_1 f_2 - f f_1 - f f_2),$$

with the short-hand notation $f = f(t,k)$, $\omega = \omega(p)$ and $f_j = f(t,k_j)$, $\omega_j = \omega(k_j)$, for wavenumbers k, k_j, $j \in \{1,2\}$. The function $\omega(k)$ is the dispersion relation of the wave system. This type of wave kinetic equations has several applications from ocean waves, acoustic waves, gravity capillary waves to Bose-Einstein condensates and many others (see [3, 4, 9, 10, 11, 12, 13, 14] and references therein). In the isotropic case, we identify $f(t,k)$ with $f(t,\omega)$ and the isotropic 3-wave kinetic equation now has the following form

$$\begin{aligned} &\partial_t f(t,\omega) = Q[f](t,\omega), \quad \omega \in \mathbb{R}_+, \\ &f(0,p) = f_0(p), \\ &Q[f](t,\omega) = \int_0^\infty \int_0^\infty [R(\omega,\omega_1,\omega_2) - R(\omega_1,\omega,\omega_2) - R(\omega_2,\omega_1,\omega)] d\omega_1 d\omega_2, \\ &R(\omega,\omega_1,\omega_2) := \delta(\omega - \omega_1 - \omega_2)\left[U(\omega_1,\omega_2) f_1 f_2 - U(\omega,\omega_1) f f_1 - U(\omega,\omega_2) f f_2\right], \end{aligned} \tag{3}$$

where U satisfies $|U(\omega_1,\omega_2)| = (\omega_1\omega_2)^{\gamma/2}$, in which γ is a non-negative constant which plays an important role in the sequel.

In [15], the authors show that if we define the energy of the solution (3) as $g(t,\omega) = \omega f(t,\omega)$, the energy is conserved. It has been proved in [15] that g can be decomposed into two parts

$$g(t,\omega) = \bar{g}(t,\omega) + \tilde{g}(t)\delta_{\{\omega=\infty\}}, \tag{4}$$

where $\bar{g}(t,\omega) \geq 0$ is a non-negative function, and $\tilde{g}(t)\delta_{\{\omega=\infty\}}$, is a measure. At time $t = 0$, we assume $\bar{g}(0,\omega) = g(0,\omega)$ and $\tilde{g}(0) = 0$. It has been proved in [15], that there exists *infinitely many blow-up times*

$$0 < t_1^* < t_2^* < \cdots < t_n^* < \cdots, \tag{5}$$

such that

$$\bar{g}(t_1^*,\omega) > \bar{g}(t_2^*,\omega) > \cdots > \bar{g}(t_n^*,\omega) > \cdots \to 0, \tag{6}$$

and

$$0 < \tilde{g}(t_1^*) < \tilde{g}(t_2^*) < \cdots < \tilde{g}(t_n^*) < \cdots \tag{7}$$

In the limit $t \to \infty$, all of the energy will be accumulated to the measure $\tilde{g}(t)\delta_{\omega=\infty}$, while the function will vanish $\tilde{g}(t)\delta_{\omega=\infty} \to 0$. We refer to time t_1^* as the first blow-up time. We define $\chi_{[0,R]}(\omega)$ be a cut-off function of ω on the finite domain $[0,R]$, the *multiple blow-up time* phenomenon (4)-(5)-(6)-(7), with the decay rate $O(\frac{1}{\sqrt{t}})$, can be observed as the decay of the total energy on any finite interval $[0,R]$

$$\int_0^R g(t,\omega)d\omega \;=\; \int_{\mathbb{R}_+} \chi_{[0,R]}(\omega)g(t,\omega)d\omega \leq\; O\Big(\frac{1}{\sqrt{t}}\Big) \text{ as } t \to \infty, \tag{8}$$

for all truncated parameter R. Inequality (8) simply means that the energy of the solution will move away from any truncated finite interval $[0,R]$ as $t \to \infty$ with the rate $O\Big(\frac{1}{\sqrt{t}}\Big)$. This theoretical finding has been numerically verified via a Finite Volume Scheme in [16].

In the important works [17, 18, 19], several numerical experiments were designed to investigate time dependent solutions of isotropic 3-wave equations. We refer to [16] for a detailed comparison between the different results of [17, 18, 19, 15, 16]. Below, we recall a brief comparison. The works [17, 18, 19] and [15, 16] complement each other as they consider very different scenarios of the solutions of (3). The works [15, 16] focuses on the finite capacity case ($\gamma > 1$), under the condition that the energy of the solution f of (3) is conserved in time

$$\int_0^\infty \omega f(t,\omega)d\omega = \text{ constant}. \tag{9}$$

and shows that there exist an infinite series of blow up times (4)-(5)-(6)- (7) (or inequality (8)). The works [17, 18, 19] focus on both cases - finite capacity ($\gamma > 1$) and infinite capacity ($0 \leq \gamma \leq 1$). In the finite capacity case ($\gamma > 1$), the solution

is studied before the first blow-up time $t < t_1^*$, rigorously proved later in [15] under the self-similar assumption

$$f(t,\omega) \approx s(t)^a F\left(\frac{\omega}{s(t)}\right). \tag{10}$$

and it is also assumed that the energy grows linearly in time

$$\int_0^\infty \omega f(t,\omega) d\omega = Jt. \tag{11}$$

Solving for the self-similar profiles before the first blow-up times t_1^* is an interesting direction of research. We mention, for instance, [20], where a self-similar profile of the solution for a different finite capacity system - the Alfven wave turbulence kinetic equation - is computed before the first blow-up time t_1^*. In addition, a recently published paper [21] presents a numerical method for for a collision integral 4-wave kinetic equation based on Chebyshev approximations.

Concerning the analysis of 3-wave kinetic equation, we mention the work [22], where a 3-wave kinetic equation, derived from the elastic beam wave equation on the lattice, has been studied. The global existence of 3-wave kinetic equations has been investigated in [23, 24] and the link to reaction networks has also been pointed out in [25, 26].

The goal of our current paper is to present an alternative numerical method, based on deep learning of the solution of the conservative form of equation (3) *introduced in [16].* It is difficult to exaggerate the impact and scope of machine learning within applied mathematics and scientific computing more broadly. A very active branch of scientific machine learning is the development of neural network approximations of the solutions to partial differential equations. As opposed to data-driven discovery of the dynamics of a physical process in which known governing equations may not be used, a physics-informed neural network (PINN) [27, 28] trains the neural network by minimizing the residual of the governing PDE along with minimization terms for the initial and boundary conditions. For completeness, we describe briefly the general idea.

Let $\mathcal{L}$ be a linear or non-linear (integro-)differential operator. Let $(t,x) \in [0,T] \times \Omega \subset \mathbb{R}^+ \times \mathbb{R}^d$ for $d \geq 1$ and $\Gamma = \partial\Omega$ smooth. Assume for simplicity that $u(t,x) \in L^2((0,T);\Omega)$ is the solution to the evolution equation

$$\begin{aligned} (\partial_t + \mathcal{L})u &= 0, \\ \mathcal{B}u &= u_\Gamma, \\ u(0,x) &= u_0(x), \end{aligned} \tag{12}$$

where $\mathcal{B}$ denotes a boundary operator and $u_0 \in L^2(\Omega)$.

If we wish to approximate the solution to (12) by a neural network $n(t, x; \theta)$ where $\theta \in \Theta$ the set of weights and biases of the neural network, then a common approach is to first define the residual operator, $\mathcal{R}$, such that

$$n(t, x; \theta) \in \mathcal{H}_{\mathcal{R}} = \{\varphi(t, x) \mid \varphi(t, x) \in L^2((0, T); \Omega),\ \mathcal{R}\varphi(t, x) \in L^2((0, T); \mathcal{R})\},$$

with

$$\|\varphi(t, x)\|_{\mathcal{H}_{\mathcal{R}}} = \|\varphi(t, x)\|_{L^2((0,T);\Omega)} + \|\mathcal{R}\varphi(t, x)\|_{L^2((0,T);\Omega)},$$

for

$$\mathcal{R} = (\partial_t + \mathcal{L}). \tag{13}$$

We then define the functional

$$J_\theta[n](t, x) = \|\mathcal{R}n(t, x; \theta)\|^2_{L^2((0,T);\Omega)} + \|\mathcal{B}n(t, x; \theta) - u_\Gamma\|^2_{L^2(\Gamma)} + \|n(0, x; \theta) - u_0\|^2_{L^2(\Omega)}. \tag{14}$$

To obtain the neural network approximation to $u(t, x)$, one then solves the following stochastic optimization problem

$$\theta^* = \arg\min_\theta J_\theta[n](t, x), \tag{15}$$

for the (probably local and non-unique) minimizer θ^* from which we obtain

$$n(t, x; \theta^*) \approx u(t, x).$$

In solving (15), one has many choices even when one excludes the usual hyper-parameter tuning that comes with selecting an architecture. In the absence of empirical measurements, a major choice to be made is how one selects inputs $(t, x) \in (0, T) \times \Omega$ for the residual and initial and boundary conditions. A common approach is to take samples from the uniform distribution and approximate the integrals in (14) by taking expectations ([27, 28] and many, many others). As in [29], one may use knowledge about the underlying distribution of the sample space to select a sampling distribution which more closely mimics empirical measurements and may be more relevant to the approximated solution.

In the present article, we choose neither of the previous approaches and opt for a quasi-Monte Carlo (qMC) approximation of the integrals in (14) selecting a low-discrepency sequence as our sample points, the Sobol sequence [30] (see figure 4).

We find the qMC approach to be appropriate for WKEs for two reasons. The first is that given the highly non-local nature of the collision term, the approximation of this operator is expensive and when coupled with the number of samples necessary to obtain an accurate approximation of the collision term via MC methods, we find a uniform sampling procedure to be impractical for our needs. The latter method introduces interesting challenges in that the distribution with which we would like to sample the wavenumber domain in order to take expectations is in fact the solution we want to find in the first place.

Thus a qMC approach which requires fewer sample points and being deterministic requires no knowledge of the underlying distribution is a simple and natural choice. For a more detailed analysis of qMC methods in neural network approximations to pdes with applications to fluid dynamics the reader is referred to the works [31, 32].

Outline

We now give a short overview of the remainder of the article. To build confidence in our approach, we present a neural network (NN) approximation to the solution of a Smoluchowski coagulation equation (SCE) in section 2. We choose this examples because, 1) as has been discussed elsewhere ([18, 15]) the SCE can be considered as a special case of a 3-WKE and 2) unlike for the 3-WKE the SCE has a known analytic solution which we may compare our approximation against. In section 3, we present results for a neural network representation of the non-stationary solution to a 3-WKE. The results are compared with previously derived theoretical results for the decay rate of the total energy in any finite interval of the wavenumber domain. As a means of validation for the neural network model, we give solutions for the same 3-WKE obtained via a finite volume scheme (FVS).

2. The Smoluchowski Coagulation Equation

Before giving results for wave kinetic equations, we provide a check of the method on a similar type of equation, the Smoluchowski coagulation equation, which has a known analytic solution in contrast to wave kinetic equations. Comparisons of the SCE and 3-WKE can be found in [15, 18, 16] and references therein.

The SCE may be written in the form [33]

$$\begin{aligned} v\partial_t f(t,v) &= -\partial_v Q_{Smol}[f](t,v) \quad (t,v) \in \mathbb{R}^+ \times \mathbb{R}^+, \\ f(0,v) &= f_0(v) \geq 0 \quad v \in \mathbb{R}^+ := (0,\infty), \end{aligned} \tag{16}$$

where the $f(t, v) \geq 0$ gives the density of particles at time t with volume v and

$$Q_{Smol}[f](t, v) = \int_0^v \int_{v-v_1}^{\infty} a(v_1, v_2) v_1 f(t, v_1) f(t, v_2) \mathrm{d}v_2 \mathrm{d}v_1, \tag{17}$$

where the kernel is given by $a(v_1, v_2) = v_1 v_2$ in what follows. Let us define

$$m(t, v) = v f(t, v),$$

then the total volume can be defined as

$$\mathcal{V}(t) = \int_0^{\infty} m(t, v) dv.$$

We will consider the case where

$$f_0(v) = \frac{e^{-v}}{v}, \tag{18}$$

i.e.,

$$m_0(v) = e^{-v}, \tag{19}$$

and the analytic solution corresponding to this initial condition is given by [33]

$$f(t, v) = e^{-Tv} \frac{I_1(2vt^{1/2})}{v^2 t^{1/2}}, \tag{20}$$

where

$$I_1(v) = \frac{1}{\pi} \int_0^{\pi} e^{v \cos\theta} \cos\theta \mathrm{d}\theta,$$

is the modified Bessel function of the first kind and

$$T = \begin{cases} 1 + t & t \leq T_{gel} \\ 2t^{1/2} & t > T_{gel} \end{cases},$$

in equation (20) above.

This SCE undergoes a gelation phenomenon at time $T_{gel} = 1$, that is,the particle number density is entirely concentrated at $v = \infty$ (again, see [33] and the references therein). This also means that $\mathcal{V}(t) = 1$ if $t \in [0, T_{gel})$ and $\mathcal{V}(t) = t^{-1/2}$ if $t \geq T_{gel}$.

2.1. NN Representation of the SCE

To begin, instead of (16), we solve the evolution equation for $m(t, v)$. Then,

$$\begin{aligned} &\partial_t m(t, v) = -\partial_v Q_{Smol}[m](t, v) \quad (t, v) \in \mathbb{R}^+ \times \mathbb{R}^+, \\ &m(0, v) = m_0(v) = v f_0(v) \geq 0 \qquad\qquad\qquad v \in \mathbb{R}^+, \end{aligned} \tag{21}$$

and the Smoluchowski collision operator is then written as

$$Q_{Smol}[m](t, v) = \int_0^v \int_{v-v_1}^{\infty} a(v_1, v_2) m(t, v_1) \frac{m(t, v_2)}{v_2} \mathrm{d}v_2 \mathrm{d}v_1. \tag{22}$$

Let $m(t, v; \theta)$ be the neural network approximation to the solution of (21) with θ denoting the weights and biases of the neural network. The NN approximation is computed for $(t, v) \in [0, T] \times [0, R]$ for some time $T > 0$ and truncation value $R > 0$.

To compute the NN approximation $m(t, v; \theta)$, the problem (21) is recast as the following functional minimization problem [34, 35]

$$m(t, v; \theta^*) = \min_{\theta \in \Theta} J_{Smol}[m](t, v; \theta), \tag{23}$$

with $\theta^* \in \Theta$ a minimizing set of parameters and $J_{Smol}[m](t, v; \theta)$ is

$$J_{Smol}[m](t, v; \theta) = \|\mathcal{R}m(t, v; \theta)\|^2_{L^2((0,T]\times[0,R])} + \|m(0, v; \theta) - m_0(v)\|^2_{L^2([0,R])}, \tag{24}$$

where $\mathcal{R}$ is the residual operator defined by

$$\mathcal{R}m(t, v; \theta) = \partial_t m(t, v; \theta) + \partial_v Q[m](t, v; \theta). \tag{25}$$

The functional $J_{Smol}[m]$ and collision term $Q_{Smol}[m]$ are discretized via a Quasi-Monte Carlo method ([31], [32])). Sample points for the functional are drawn from the Sobol sequence ([36]) in the unit square which are then mapped to the discrete set $S \subset (0, T] \times [0, R]$. For the collision term, for each v, we draw samples from the Sobol sequence in the unit square as before, but map the variables (v_1, v_2) to the discrete intervals $V_1 \subset [0, v]$ and, for each v_1, $V_2 \subset [v - v_1, R]$, respectively. The discretized collision operator is then given by

$$\hat{Q}_{Smol}[m](t, v; \theta) = \frac{v}{|V_1|} \sum_{v_1 \in V_1} v_1 m(t, v_1; \theta) \left(\frac{|R + v_1 - v|}{|V_2|} \sum_{v_2 \in V_2} m(t, v_2; \theta) \right), \tag{26}$$

which leads to the semi-discretized residual

$$\hat{r}(t, v; \theta) = \partial_t m(t, v; \theta) + \partial_v \hat{Q}_{Smol}[m](t, v; \theta), \tag{27}$$

and so the semi-discrete functional $\hat{J}_{Smol}[m](t, v; \theta)$ to be minimized is

$$\hat{J}_{Smol}[m](t, v; \theta) = \frac{1}{|S|} \sum_{(t,v)\in S} \hat{r}(t, v; \theta)^2 + \frac{1}{|S_0|} \sum_{v\in S_0} (m(0, v;\ \theta) - m_0(v))^2, \tag{28}$$

where the set $S_0 \subset S$, $|S_0| \leq |S|$ denotes a subset of the volume samples in S.

3. A 3-Wave Kinetic Equation

In [16], a new identity for the energy of the solutions to 3-WKEs [15] is presented which takes the form of a conservation law [33]. This is the equation we shall study in the present paper due to the simplicity provided by this form of the equation. Specifically, there is no need to compute the resonant manifolds of the system, though we emphasize that the method does not depend on this simplification.

Let us write the equation to be solved. The equation is equivalent to (3) and is explained in [15, 16]

$$\begin{aligned} \partial_t g(t, p) = p\partial_p Q[g](t, p) \quad (t, p) \in \mathbb{R}^+ \times \mathbb{R}^+, \\ g(0, p) = g_0(p) \geq 0 \quad (t, p) \in \{0\} \times \mathbb{R}^+, \end{aligned} \tag{29}$$

where the collision term is given by

$$Q[g](t, p) = -2 \int_0^p \int_0^p (p_1, p_2)^{\frac{\gamma}{2}-1} g_1 g_2 \chi_{p,1,2} \mathrm{d}p_{2,1} + \int_0^\infty \int_0^\infty (p_1, p_2)^{\frac{\gamma}{2}-1} g_1 g_2 \chi_{p,1,2} \mathrm{d}p_{2,1}, \tag{30}$$

where we have used the notation $g_i = g(t, p_i)$ for $i = 1, 2$, $\mathrm{d}p_{2,1} = \mathrm{d}p_2\mathrm{d}p_1$ and $\chi_{p,1,2} = \chi\{p < p_1 + p_2\}$ where $\chi\{A\}$ is the set characteristic function of some set A. The parameter γ is the degree of the kernel as discussed in the introduction and $\gamma = 2$ in the present work which corresponds to acoustic wave systems.

As in the previous section, we define a neural network approximation, $g(t, p; \theta)$ to be a solution to the optimization problem [37]

$$g(t, p; \theta^*) = \min_{\theta\in\Theta} J[g](t, p; \theta), \tag{31}$$

with $\theta^* \in \Theta$ a minimizing set of parameters of the functional

$$J[g](t, p; \theta) = \|r(t, p; \theta)\|^2_{L^2(\mathbb{R}^+\times\mathbb{R}^+)} + \|g(0, p; \theta) - g_0(p)\|^2_{L^2(\mathbb{R}^+)}, \tag{32}$$

where, again, $\mathcal{R}$ denotes the residual operator of the evolution equation defined by

$$\mathcal{R}g(t, p; \theta) = \partial_t g(t, p; \theta) - p\partial_p Q[g](t, p; \theta). \tag{33}$$

The functional (32) is approximated via a Quasi-Monte Carlo method with sample points drawn from the Sobol sequence in the unit square and then transformed to some truncated rectangle of the time, wavenumber domain, i.e. we generate the set $W \sim Sobol([0,T]\times[0,R])$, for $T, R > 0$ truncation parameters of the time and wavenumber domain, respectively. The residual (33) is approximated similarly where the set of sample points for (p_1, p_2) are given by $P_1 \sim Sobol(0, p)$ for each p, and $P_2 \sim Sobol(p - p_1, p)$ for each p, p_1 in the first term of (30) and $\hat{P}_2 \sim Sobol(p - p_1, R)$ for each (p, p_1).

To see how these sample sets are defined, note that the collision term can be rewritten as

$$Q[g](t,p) = -2\int_0^p \int_{p-p_1}^p (p_1 p_2)^{\frac{\gamma}{2}-1} g_1 g_2 \mathrm{d}p_{2,1} + \int_0^p \int_{p-p_1}^R (p_1 p_2)^{\frac{\gamma}{2}-1} g_1 g_2 \mathrm{d}p_{2,1}, \tag{34}$$

where we have applied the truncation parameter of the wavenumber domain to the second expression and enforced the restrictions $p - p_1 < p_2$ in both terms to satisfy the characteristic set function and $p_1 < p$ to guarantee positivity of the variable p_2 in the second term. Thus, using the above collision operator and sample sets defined in the previous paragraph, we can define the discrete collision operator to be

$$\hat{Q}[g](t,p) = \frac{-2p}{|P_1||P_2|}\sum_{p_1\in P_1} p_1^{\frac{\gamma}{2}} g_1 \Bigg(\sum_{p_2\in P_2} p_2^{\frac{\gamma}{2}-1} g_2\Bigg) + \frac{p}{|P_1||\hat{P}_2|}\sum_{p_1\in P_1} p_1^{\frac{\gamma}{2}-1} g_1 \Bigg[(R-p+p_1)\Big(\sum_{p_2\in \hat{P}_2} p_2^{\frac{\gamma}{2}-1} g_2\Big)\Bigg]. \tag{35}$$

Using the discrete collision operator, we can define the semi-discrete residual by

$$\hat{\mathcal{R}} g(t,p;\theta) = \partial_t g(t,p;\theta) - p\partial_p \hat{Q}[g](t,p;\theta), \tag{36}$$

and therefore we can minimize the semi-discrete functional

$$\hat{J}[g](t,p);\theta) = \frac{1}{|W|}\sum_{(t,p)\in W} \hat{\mathcal{R}} g(t,p;\theta)^2 + \frac{1}{|W_0|}\sum_{p\in W_0} (g(0,p;\theta) - g_0(p))^2, \tag{37}$$

where $W_0 \subset W$, $|W_0| \leq |W|$ contains only sample points for the wavenumber.

4. Numerical Results

In this section, we provide numerical results for the SCE with initial condition (18) and compare with the anaylitc solution (20). A 3-WKE with initial condition

(38) is then solved and the decay rate of the total energy is compared with the results obtained in [15]. The same initial condition (38) is then used to obtain a solution with the finite volume scheme developed in [16]. The results of the two methods are discussed.

All deep learning tests were implemented in TensorFlow [38] with Keras [39]. Training was accelerated using multiple GPUs utilizing Horovod [40]. All computations were performed on SMU's computing cluster MII. The finite volume computations were performed in Matlab.

4.1. Test 1

A simple, yet effective, architecture is chosen. Namely, it is enough to employ a feedforward network with only two hidden layers with 128 hidden units in each layer utilizing sigmoidal activation functions [41]. Given its simplicity, we can write the neural network (NN) out explicitly as

$$m(t, v; \theta) = \mathbf{W}\sigma(\mathbf{W}_2\sigma(\mathbf{W}_1\vec{x} + \mathbf{b}_1) + \mathbf{b}_2) + b,$$

with the entries of $\mathbf{W}_i, \mathbf{b}_i \sim \mathcal{N}(0, 1)$ for $i = 1, 2$, $\mathbf{W}, b \sim \mathcal{N}(0, 1)$ and $\sigma(\cdot)$ denoting the sigmoid function. To fully discretize the functional (28), we employ TensorFlow's built-in automatic differentiation method.

The training samples were drawn from the rectangle $(0, T] \times [0, R]$ using the strategy described above for $T = 0.8$ and $R = 8$. To achieve the results shown in Figure 1, only the first 32 Sobol points were used to minimize the residual term (16 time samples and 16 volume samples) and only the first 16 Sobol points were needed to train on the initial data.

Figure 1 shows the neural network predictions on unseen data within the training interval. The inputs to the neural network to produce the plot were the first 128 Sobol points transformed to the volume training interval and $t = 0.0, 0.2, 0.4$ and 0.62. We see the NN solution gives a good approximation to the analytic solution.

In Figure 2, we quantify the match of the NN approximation to the analytic solution by giving the sup norm of the error at each snapshot. Here, we supply the first 2^{10} Sobol points within the volume training interval and use the same points in time $t = 0.0, 0.2, 0.4, 0.62$. We note here that while the NN performs well on unseen points within the volume interval $v \in [0, R]$, for $v > R$ we generalization error is larger. This is to be expected given the small training sample size used (16 volume points). We show the absolute error in Figure 3 for each snapshot in time over the interval $v \in [0, 1000]$ with 2^{13} Sobol points in this volume interval. This

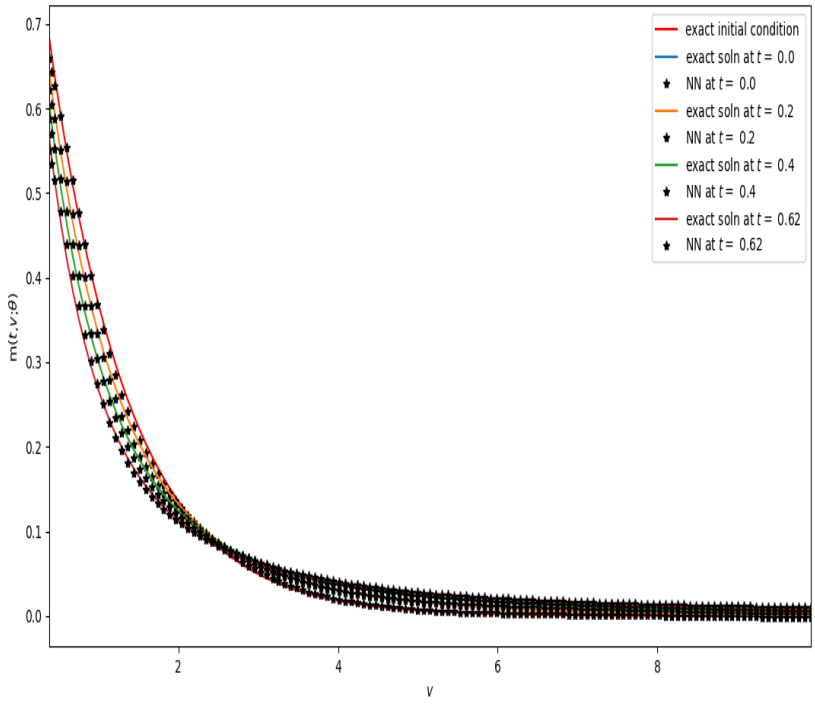


Figure 1: Neural Network (NN) approximation to the Smoluchoswki equation (21). The solid lines denote the analytic solution (20) and $*$ markers denote the NN approximation at equivalent snapshots in time.

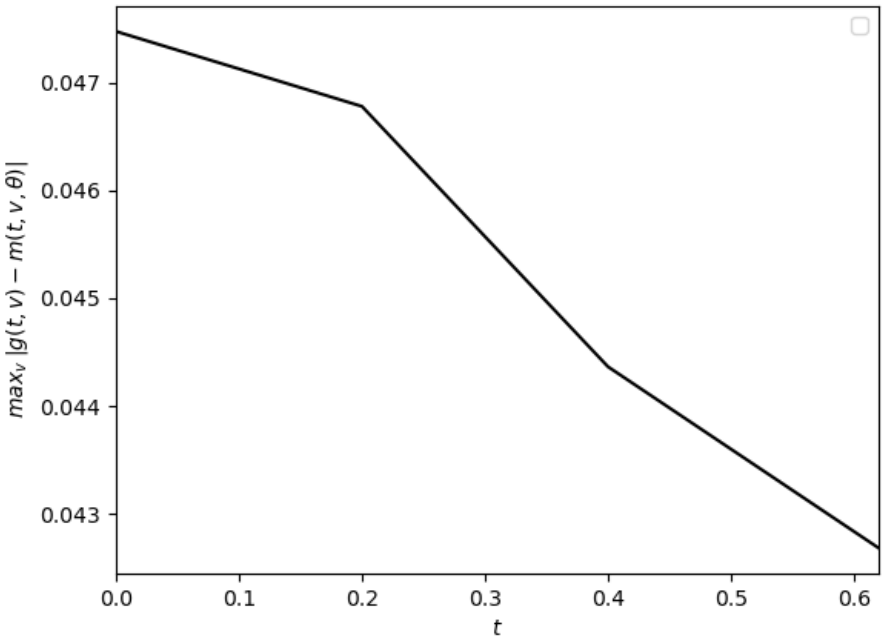


Figure 2: Sup norm of the error at $t = 0.0, 0.2, 0.4$ and 0.62 where the maximum is taken over the volume domain $v \in [0, 10]$. This shows the accuracy of the method on unseen data points within the training interval.

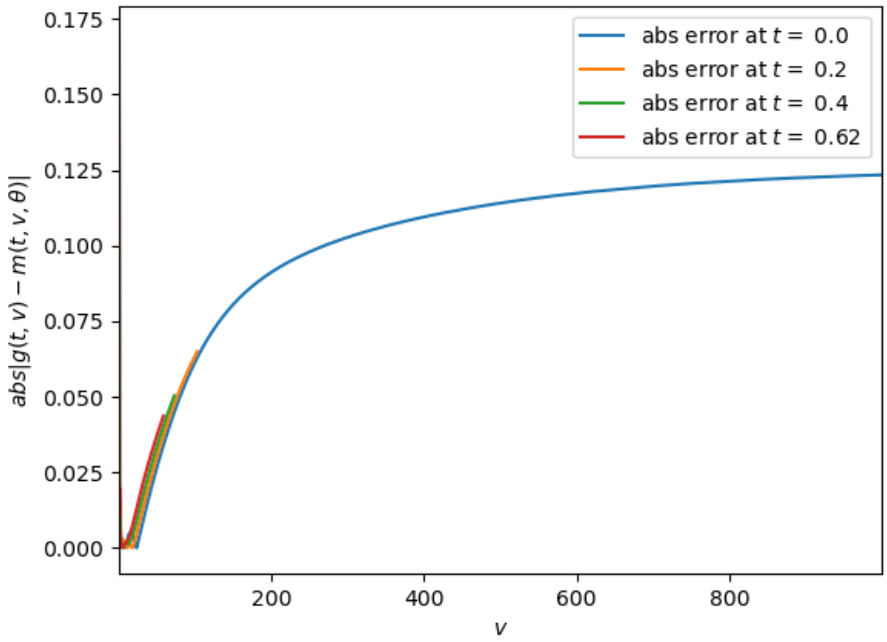


Figure 3: Absolute error of the NN approximation versus the analytic solution over the volume interval $v \in [0, 1000]$ with 2^{13} volume samples given as input to the NN. We see a slow growth in the error for large volume numbers.

result informed our decision to apply a batch sampling procedure for the wave kinetic equation, the details of which are discussed below.

4.2. Test 2

We choose the initial condition

$$g_0(p) = \sqrt{\frac{7}{2\pi}} e^{-\frac{7(p-2)^2}{2}}. \tag{38}$$

The neural network was trained using samples from the Sobol sequence in the rectangle $W \sim Sobol((0, T] \times [0, R])$ for $T = 10$ and $R = 10$. As illustrated in figure 4, the size of the sample set $|W|$ was 2^{15}, which was broken up into smaller training batches.

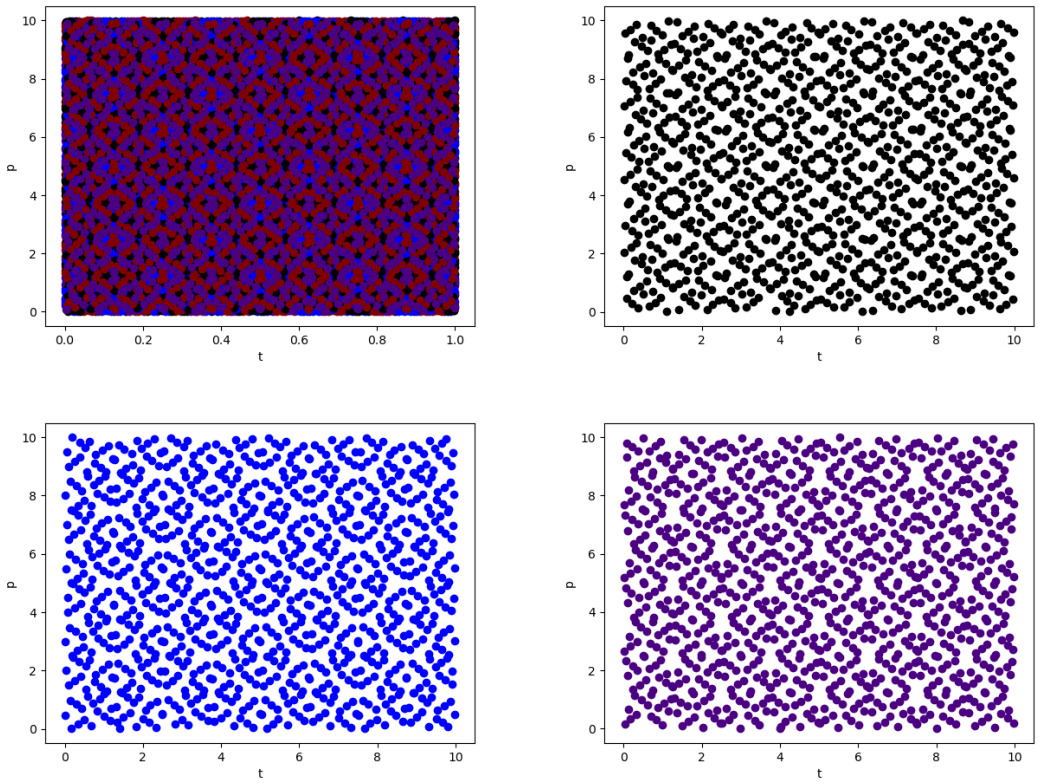


Figure 4: Batched samples of Sobol points. (Top Left) All sample points. (Top Right and Bottom Row) Example batches of sampled points.

The architecture was again chosen to have 2 hidden layers, each with 128 units and sigmoidal activation functions. The loss was again minimized using tensorflows implementation of ADAM. The collision terms were approximated with the sampling procedure described above with each sample comprised of 32 Sobol points. We consider only the case $\gamma = 2$ here.

We show a few early snapshots of the approximated solution in Figure 5. The behaviour seems consistent with predictions in that the L^∞ norm decays as time increases. This is more clearly evidenced by the total energy discussed in the following paragraph. In Figure 5, we also present a comparison with the solution computed the Finite Volume Scheme of [16]. This comparison will be described later in this section, around equation (39)-(40).

The total energy of the solutions was computed and the log of the total energy is plotted against the log of t. The comparison between the decay of the numerical solution is made with the theoretical decay rate of $t^{-1/2}$, obtained in [15] (see (8)), in Figure 6. The total energy was predicted up to $t = 148$. From Figure 6, the energy decays and the decay is in good agreement with the theoretical rate obtained in [15].

The neural network is able to make reasonable predictions for very large, unseen in training, samples in the wave number domain. In Figure 7, we see the predicted solution for wavenumber values up to 1e+06. In contrast, the most we could push for reasonable predicitons in time was $t = 148$ as previously reported in the results shown in figure 6. For better predictions in time, we needed to enforce more dense sampling in the time domain. Thus, the neural network was not only trained on the set W for T= 10, but also for $T = 5$ and $T = 2$ with the same number of sample points, 2^{15}.

Comparison with a Finite Volume Scheme

As mentioned earlier, to validate the results, we provide a comparison with another model. Here we use the finite volume scheme presented in [16], which we briefly outline below.

The discretization of the wavenumber domain $p \in [0, R]$, is given as follows. Let $i \in \{0, 1, 2, \ldots, M\} = I_h^M$, with $h \in (0, 1)$ fixed and $M = M(h)$. Define the set $S_d = \{0, \ldots, R\}$ to be the discretization of the interval $[0, R]$. Let

$$S_d = \{p_{i+1/2}\}_{i \in I_h^M},\ \{p_i\}_{i \in I_h^M \setminus \{0\}} = \frac{p_{i+1/2} + p_{i-1/2}}{2},\ \{\Delta p_i\}_{i \in I_h^M \setminus \{0\}} = p_{i+1/2} - p_{i-1/2} \leq h, \tag{39}$$

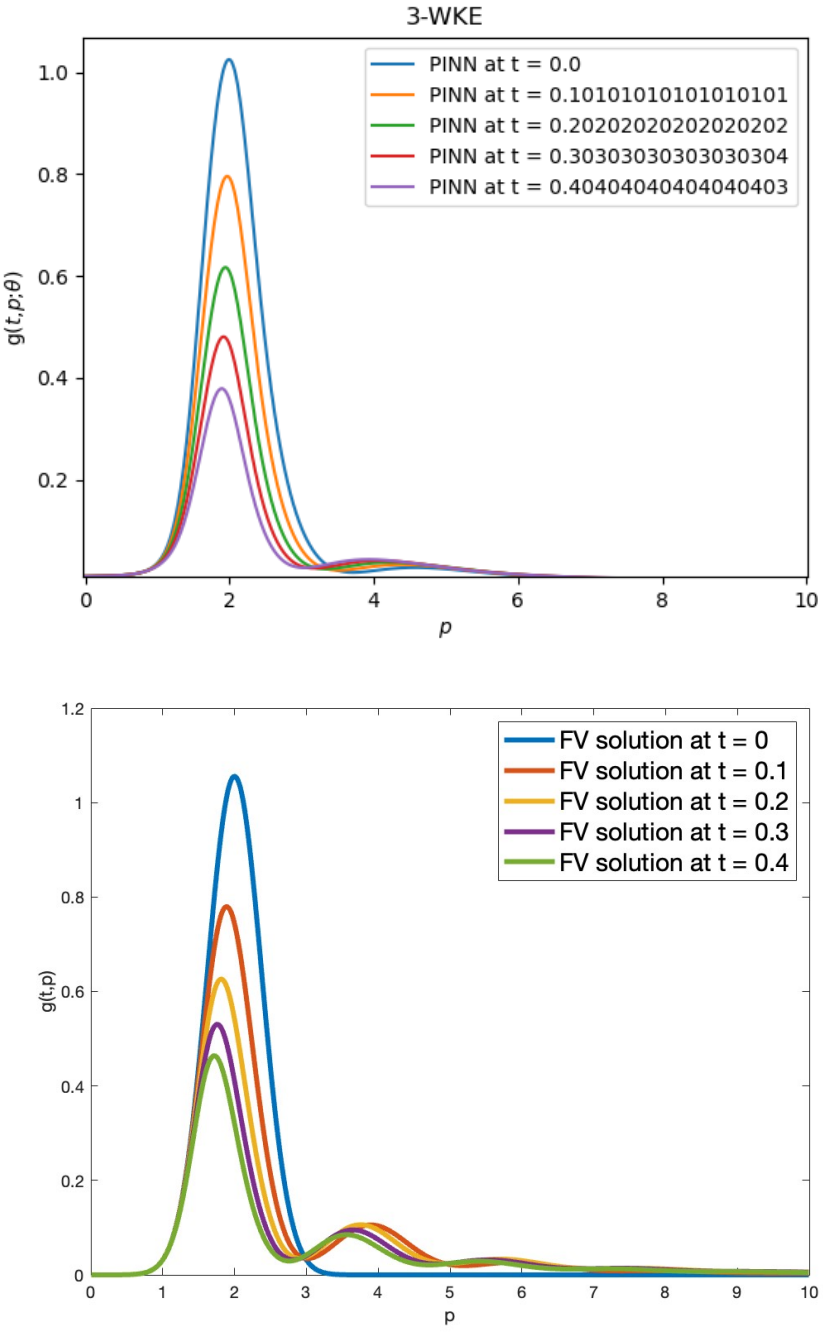
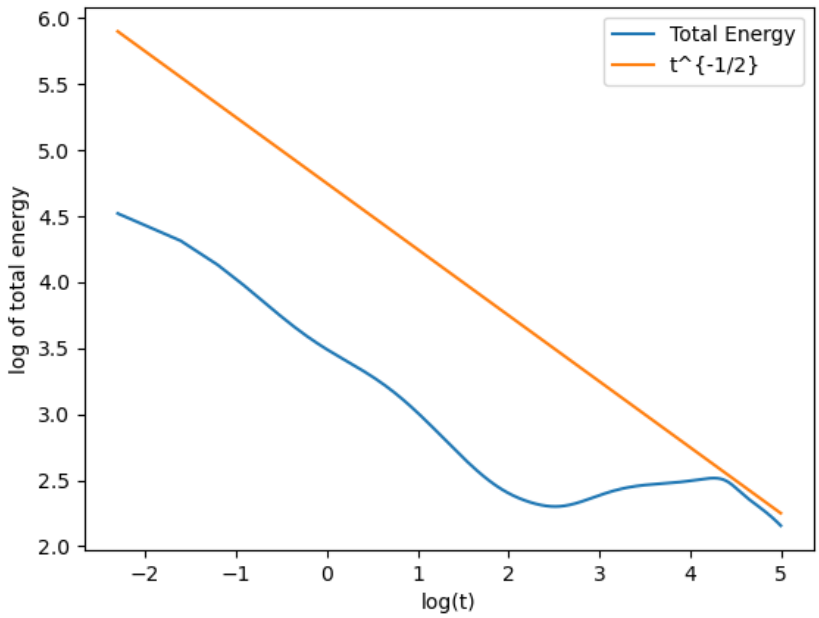


Figure 5: Top Picture: A few snapshots of the NN approximation corresponding to initial condition (38). Bottom Picture: Comparative snapshots of the FVS solution for the same initial condition (38).



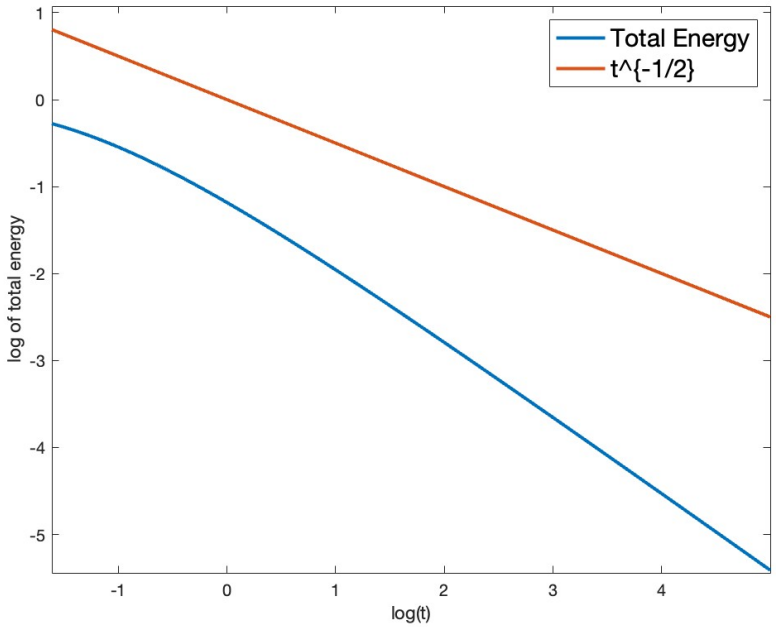


Figure 6: Top Picture: Log-Log plot of the total energy corresponding to initial condition (38) as predicted up to $t = 148$ the NN. Bottom Picture: The total energy computed for $t \in [0, 148]$ with the FVS.

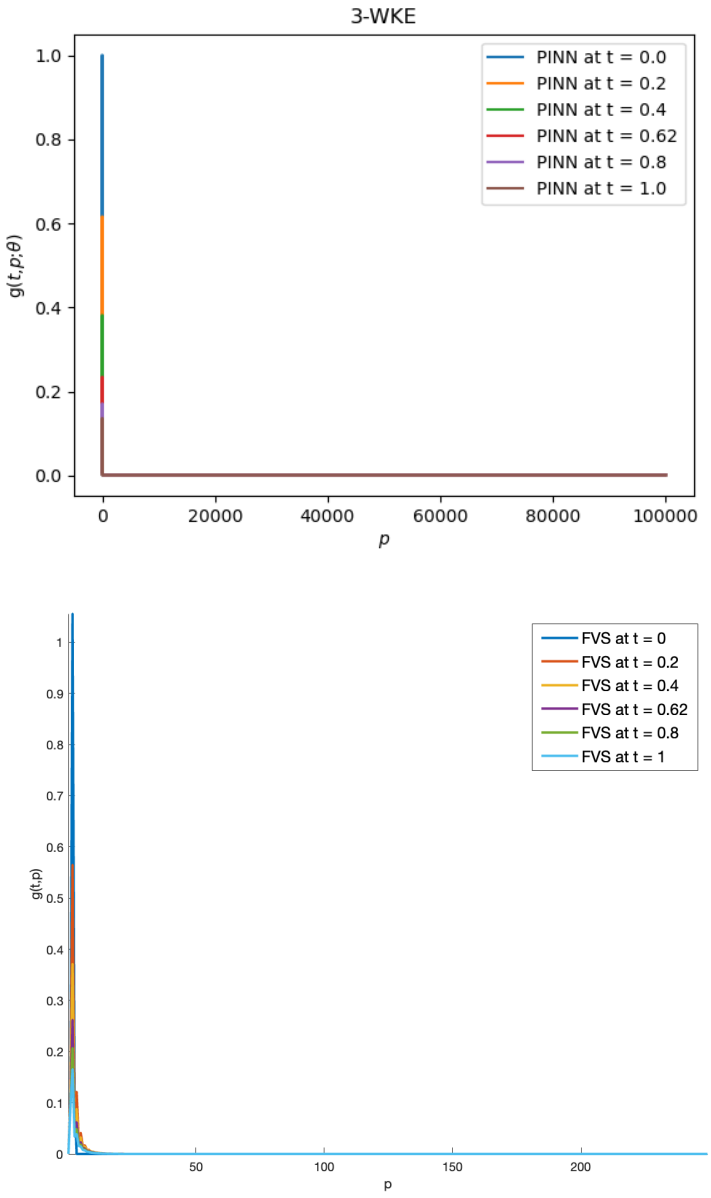


Figure 7: Top picture: Neural network prediction corresponding to initial condition (38) for wavenumbers up to 1e+06. Bottom picture: The finite volume solution for wavenumbers up to $p = 250$, the largest value for which stability is maintained without further decreasing Δt. The two figures highlight a key strength of the presented method, in that the neural network approximation is consistent with more traditional solvers but is able to produce results for computationally prohibitive values of the wavenumber for traditional methods while maintaining positivity and stability.

define the faces, pivots and step-size respectively, with $p_{1/2} = 0$ and $p_{M+1/2} = R$. For simplicity, we use a uniform grid and set $h = 0.01$ which leads to

$$S_d = \{ih\}_{i\in I_h^M}, \quad \{p_i\}_{i\in I_h^M\setminus\{0\}} = \frac{h}{2}(2i-1), \quad \{\Delta p_i\}_{i\in I_h^M\setminus\{0\}} = h \in (0,1).$$

The set $T_N = \{0,\dots,T\}$ with $N+1$ nodes and $T = 148$ is the maximum time. We fix the time step to be $\Delta t = \frac{T}{N} = 0.005$, and denote by $t_n = \Delta t \cdot n$ for $n \in \{0,\dots,N\}$. We approximate equation (29) by

$$g^{n+1}(p_i) = g^n(p_i) + \lambda_i\Big(Q^n_{i+1/2}\Big[\frac{g}{p}\Big] - Q^n_{i-1/2}\Big[\frac{g}{p}\Big]\Big), \tag{40}$$

where $\lambda_i = \frac{p_i\Delta t}{\Delta p_i}$, and

$$Q^n_{i+1/2}\Big[\frac{g}{p}\Big] - Q^n_{i-1/2}\Big[\frac{g}{p}\Big] = -2\Big(Q^n_{1,i+1/2}\Big[\frac{g}{p}\Big] - Q^n_{1,i-1/2}\Big[\frac{g}{p}\Big]\Big) + \Big(Q^n_{2,i+1/2}\Big[\frac{g}{p}\Big] - Q^n_{2,i-1/2}\Big[\frac{g}{p}\Big]\Big),$$

with

$$Q^n_{1,i+1/2}\Big[\frac{g}{p}\Big] = \sum_{m=1}^{i}\Delta p_m\frac{g^n(p_m)}{p_m}\Bigg(\sum_{j=1}^{i}\Delta p_j\frac{g^n(p_j)}{p_j}a(p_m,p_j)\chi\Big\{p_{i+1/2} < p_m + p_j\Big\}\Bigg), \tag{41}$$

$$Q^n_{2,i+1/2}\Big[\frac{g}{p}\Big] = \sum_{m=1}^{M}\Delta p_m\frac{g^n(p_m)}{p_m}\Bigg(\sum_{j=1}^{M}\Delta p_j\frac{g^n(p_j)}{p_j}a(p_m,p_j)\chi\Big\{p_{i+1/2} < p_m + p_j\Big\}\Bigg), \tag{42}$$

where we have used the midpoint rule to approximate the integrals in equation (29) and we choose an explicit time stepping method.

The initial condition (38) is approximated by

$$g^0(p_i) = \frac{1}{\Delta p_i}\int_{p_{i-1/2}}^{p_{i+1/2}} g_0(p)\mathrm{d}p \approx g_0(p_i),$$

by again employing the midpoint rule.

We draw the reader's attention again to Figure 5 where a few comparative snapshots of the solution are provided. Qualitatively, the two models seem to agree and capture the main features of the theorized behavior of solutions. Namely, an evacuation of the energy within any finite interval. Further, as is typical of simple feedforward architectures, the finer oscillations seen in the FVS solutions appear to be averaged out in the NN solution as expected. For our purposes, the

dynamics captured by the simple feed-forward architecture we have employed here are enough to confirm the theory presented in [15]. We leave it to a future work to investigate more complicated architectures. For example, the architecture described in [42] is able to capture highly oscillatory solutions for the stationary Navier-Stokes equations.

In Figure 6 (bottom) we see the decay rate of the total energy as provided by the computed solution of the FVS, which is compared with the *predicted* values by the NN (top). Both models are in good agreement with the theorized bound on the rate of decay [16, 15], though the FVS appears to capture a slightly faster rate of decay.

Looking back to Figure 7, we make a comparison with the computed solution of the finite volume method for large wavenumber values with predictions provided by the neural network. Here, we have increased h to 0.8 and $R = 250$ while keeping the timestep fixed at $\Delta t = 0.005$ as in the previous figures. The two figures highlight a key strength of the presented deep learning method, in that the neural network approximation is consistent with more traditional solvers but is able to produce results for computationally prohibitive values of the wavenumber for traditional methods, while maintaining positivity and stability. Indeed, constructing positivity preserving schemes for PDEs is a very important direction of research [43, 44, 45]. For the FVS (40), the positivity and thus stability is lost for wavenumber values larger than 250 with the timestep set at $\Delta t = 0.005$. In order to preserve the positivity of the solutions produced by the FVS (40), the CFL condition is restrictive and the time step Δt needs to be chosen sufficiently small as shown in Proposition 3.1 of [16]. As thus, being able to preserve the positivity of the solutions is indeed a very important feature of the presented deep learning approximation.

5. Conclusions

We present a deep learning approximation, stochastic optimization based, method for the 3-wave kinetic equation, studied theoretically in [15] and numerically in [16]. We first apply the method to a Smoluchowski coagulation equation with multiplicative kernel for which an analytic solution exists. The deep learning method is proved to give a good approximation of the analytic solution. Next, the learning approach is then used to approximate the solution of the 3-wave kinetic equation. The deep learning approximation is tested and proved to be as good as the Finite Volume approximation introduced in [16] for the same equation and both approximations are in good agreement with the theoretical results of [15].

Acknowledgments

Computational resources for this research were provided by SMU's Center for Research Computing. S.W. would like to thank Thom Hagstrom for many useful discussions.

References

[1] R. Peierls, Zur kinetischen theorie der warmeleitung in kristallen, Annalen der Physik 395 (8) (1929) 1055–1101.

[2] R. E. Peierls, Quantum theory of solids, in: Theoretical physics in the twentieth century (Pauli memorial volume), Interscience, New York, 1960, pp. 140–160.

[3] K. Hasselmann, On the non-linear energy transfer in a gravity-wave spectrum part 1. general theory, Journal of Fluid Mechanics 12 (04) (1962) 481–500.

[4] K. Hasselmann, On the spectral dissipation of ocean waves due to white capping, Boundary-Layer Meteorology 6 (1-2) (1974) 107–127.

[5] D. J. Benney, P. G. Saffman, Nonlinear interactions of random waves in a dispersive medium, Proc. R. Soc. Lond. A 289 (1418) (1966) 301–320.

[6] B. B. Kadomtsev, Plasma turbulence, New York: Academic Press, 1965 (1965).

[7] V. E. Zakharov, V. S. L'vov, G. Falkovich, Kolmogorov spectra of turbulence I: Wave turbulence, Springer Science & Business Media, 2012.

[8] D. J. Benney, A. C. Newell, Random wave closures, Studies in Applied Mathematics 48 (1) (1969) 29–53.

[9] S. Nazarenko, A. Soffer, M.-B. Tran, On the wave turbulence theory for the nonlinear schrödinger equation with random potentials, Entropy 21 (9) (2019) 823.

[10] V. E. Zakharov, Stability of periodic waves of finite amplitude on the surface of a deep fluid, Journal of Applied Mechanics and Technical Physics 9 (2) (1968) 190–194.

[11] V. E. Zakharov, N. N. Filonenko, Weak turbulence of capillary waves, Journal of applied mechanics and technical physics 8 (5) (1967) 37–40.

[12] A. Soffer, M.-B. Tran, On the dynamics of finite temperature trapped bose gases, Advances in Mathematics 325 (2018) 533–607.

[13] Y. Pomeau, M.-B. Tran, Statistical physics of non equilibrium quantum phenomena, Lecture Notes in Physics, Springer (2019).

[14] V. E. Zakharov, Weak turbulence in media with a decay spectrum, Journal of Applied Mechanics and Technical Physics 6 (4) (1965) 22–24.

[15] A. Soffer, M.-B. Tran, On the energy cascade of 3-wave kinetic equations: beyond kolmogorov–zakharov solutions, Communications in Mathematical Physics (2019) 1–48.

[16] S. Walton, M.-B. Tran, A numerical scheme for wave turbulence: 3-wave kinetic equations, arXiv preprint arXiv:2110.10124 (2021).

[17] C. Connaughton, Numerical solutions of the isotropic 3-wave kinetic equation, Physica D: Nonlinear Phenomena 238 (23) (2009) 2282–2297.

[18] C. Connaughton, P. Krapivsky, Aggregation–fragmentation processes and decaying three-wave turbulence, Physical Review E 81 (3) (2010) 035303.

[19] C. Connaughton, A. C. Newell, Dynamical scaling and the finite-capacity anomaly in three-wave turbulence, Physical Review E 81 (3) (2010) 036303.

[20] N. K. Bell, V. N. Grebenev, S. B. Medvedev, S. V. Nazarenko, Self-similar evolution of alfven wave turbulence, Journal of Physics A: Mathematical and Theoretical 50 (43) (2017) 435501.

[21] B. V. Semisalov, V. N. Grebenev, S. B. Medvedev, S. V. Nazarenko, Numerical analysis of a self-similar turbulent flow in bose–einstein condensates, Communications in Nonlinear Science and Numerical Simulation 102 (2021) 105903.

[22] B. Rumpf, A. Soffer, M.-B. Tran, On the wave turbulence theory: ergodicity for the elastic beam wave equation, Submitted (2021).

[23] I. M. Gamba, L. M. Smith, M.-B. Tran, On the wave turbulence theory for stratified flows in the ocean, M3AS: Mathematical Models and Methods in Applied Sciences. Vol. 30, No. 1 105-137 (2020).

[24] T. T. Nguyen, M.-B. Tran, On the Kinetic Equation in Zakharov's Wave Turbulence Theory for Capillary Waves, SIAM J. Math. Anal. 50 (2) (2018) 2020–2047. doi:10.1137/17M1125042.
URL https://doi.org/10.1137/17M1125042

[25] G. Craciun, M.-B. Tran, A reaction network approach to the convergence to equilibrium of quantum boltzmann equations for bose gases, ESAIM: Control, Optimisation and Calculus of Variations (2021).

[26] M.-B. Tran, G. Craciun, L. M. Smith, S. Boldyrev, A reaction network approach to the theory of acoustic wave turbulence, Journal of Differential Equations 269 (5) (2020) 4332–4352.

[27] M. Raissi, P. Perdikaris, G. 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 (2019) 686–707. doi:https://doi.org/10.1016/j.jcp.2018.10.045.
URL https://www.sciencedirect.com/science/article/pii/S0021999118307125

[28] L. Lu, X. Meng, Z. Mao, G. E. Karniadakis, Deepxde: A deep learning library for solving differential equations, SIAM Review 63 (1) (2021) 208–228. arXiv:https://doi.org/10.1137/19M1274067, doi:10.1137/19M1274067.
URL https://doi.org/10.1137/19M1274067

[29] J. Sirignano, K. Spiliopoulos, Dgm: A deep learning algorithm for solving partial differential equations, Journal of Computational Physics 375 (2018) 1339–1364. doi:https://doi.org/10.1016/j.jcp.2018.08.029.
URL https://www.sciencedirect.com/science/article/pii/S0021999118305527

[30] I. Sobol', On the distribution of points in a cube and the approximate evaluation of integrals, USSR Computational Mathematics and Mathematical Physics 7 (4) (1967) 86–112. doi:https://doi.org/10.1016/0041-5553(67)90144-9.
URL https://www.sciencedirect.com/science/article/pii/0041555367901449

[31] S. Mishra, T. K. Rusch, Enhancing accuracy of deep learning algorithms by training with low-discrepancy sequences, SIAM Journal on Numerical Analysis 59 (3) (2021) 1811–1834. arXiv:https://doi.org/10.1137/20M1344883, doi:10.1137/20M1344883.
URL https://doi.org/10.1137/20M1344883

[32] K. O. Lye, S. Mishra, D. Ray, Deep learning observables in computational fluid dynamics, Journal of Computational Physics 410 (2020) 109339. doi:https://doi.org/10.1016/j.jcp.2020.109339.
URL https://www.sciencedirect.com/science/article/pii/S0021999120301133

[33] F. Filbet, P. Laurençot, Numerical simulation of the smoluchowski coagulation equation, SIAM Journal on Scientific Computing 25 (6) (2004) 2004–2028.

[34] A. Bensoussan, Lectures on stochastic control, in: Nonlinear filtering and stochastic control, Springer, 1982, pp. 1–62.

[35] A. Bensoussan, Y. Li, D. P. C. Nguyen, M.-B. Tran, S. C. P. Yam, X. Zhou, Machine learning and control theory, in: Handbook of Numerical Analysis, Elsevier BV, 2022, pp. 531–558.

[36] R. E. Cafliscsh, Monte carlo and quasi-monte carlo methods, Acta Numerica vol. 7, pp. 1-49. (1998).

[37] A. Bensoussan, Stochastic control of partially observable systems, Cambridge University Press, 1992.

[38] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, TensorFlow: Large-scale machine learning on heterogeneous systems, software available from tensorflow.org (2015).
URL https://www.tensorflow.org/

[39] F. Chollet, et al., Keras, https://keras.io (2015).

[40] A. Sergeev, M. D. Balso, Horovod: fast and easy distributed deep learning in TensorFlow, arXiv preprint arXiv:1802.05799 (2018).

[41] M. Hornik, K. Stinchcombe, H. White, Multilayer feedforward networks are universal approximators, Neural Networks, 2(5), 359–366. (1989).

[42] L. Liu, B. Wang, W. Cai, Linearized learning methods with multiscale deep neural networks for stationary navier-stokes equations with oscillatory solutions (2021). doi:10.48550/ARXIV.2102.03293.
URL https://arxiv.org/abs/2102.03293

[43] X. Y. Hu, N. A. Adams, C.-W. Shu, Positivity-preserving method for high-order conservative schemes solving compressible euler equations, Journal of Computational Physics 242 (2013) 169–180.

[44] J. Huang, C.-W. Shu, Positivity-preserving time discretizations for production–destruction equations with applications to non-equilibrium flows, Journal of Scientific Computing 78 (3) (2019) 1811–1839.

[45] J. Huang, W. Zhao, C.-W. Shu, A third-order unconditionally positivity-preserving scheme for production–destruction equations with applications to non-equilibrium flows, Journal of Scientific Computing 79 (2) (2019) 1015–1056.