JOURNAL OF INTEGRAL
EQUATIONS AND APPLICATIONS
Volume 36 (2024), No. 1, 23–55

DEEP NEURAL NETWORK SOLUTIONS FOR OSCILLATORY FREDHOLM INTEGRAL EQUATIONS

JIE JIANG AND YUESHENG XU

We studied the use of deep neural networks (DNNs) in the numerical solution of the oscillatory Fredholm integral equation of the second kind. It is known that the solution of the equation exhibits certain oscillatory behaviors due to the oscillation of the kernel. It was pointed out recently that standard DNNs favor low frequency functions, and as a result, they often produce poor approximation for functions containing high frequency components. We addressed this issue in this study. We first developed a numerical method for solving the equation with DNNs as an approximate solution by designing a numerical quadrature that tailors to computing oscillatory integrals involving DNNs. We proved that the error of the DNN approximate solution of the equation is bounded by the training loss and the quadrature error. We then proposed a multigrade deep learning (MGDL) model to overcome the spectral bias issue of neural networks. Numerical experiments demonstrate that the MGDL model is effective in extracting multiscale information of the oscillatory solution and overcoming the spectral bias issue from which a standard DNN model suffers.

1. Introduction

The goal of this paper is to develop an effective deep neural network (DNN) method for the numerical solution of the oscillatory Fredholm integral equation of the second kind. The integral equation has wide applications in physics and engineering, such as electromagnetic scattering [3; 6]. Even though the numerical solution for the equation is a classical research topic [1; 3; 5; 23; 24], it is desirable to investigate how DNNs can be effectively used to represent its numerical solution, given the recent success of the use of DNNs in the numerical solution of partial differential equations [20; 21] and variational problems [9].

The solution of the oscillatory Fredholm integral equation of the second kind exhibits certain oscillation due to the oscillation of its kernel, see [23]. In other words, the solution of the equation contains high frequency components as a result of its involvement of an oscillatory kernel. A similar circumstance also exists in the solution of the oscillatory Volterra integral equation [4]. This fact places a barrier for using DNNs as an approximate solution of the equation because of the spectral bias phenomenon of neural networks. The spectral bias phenomenon was discussed in [19], which showed that while neural networks can approximate arbitrary functions, they favor low frequency ones, and as a result, they exhibit a bias towards smooth functions. It is the aim of this paper to investigate how to overcome the spectral bias barrier.

Because of the nonlinearity of DNNs, solving the integral equation for a DNN approximate solution boils down to solve a non-convex optimization problem by an iterative scheme. Every step of the iteration requires to evaluate the objective function which involves oscillatory integrals. Hence, it is inevitable to

2020 AMS *Mathematics subject classification:* 65R20.
Keywords and phrases: oscillatory Fredholm integral equation, deep neural network, spectral bias.
Received by the editors on January 6, 2024, and in revised form on January 29, 2024.

DOI: 10.1216/jie.2024.36.23

compute oscillatory integrals at every step of the iteration. There is a large literature on the numerical quadrature of oscillatory integrals. Computing integrals of the product of a smooth non-oscillatory function and a structured oscillatory kernel function has a long history in numerical analysis. For example, computing integrals involving the Fourier kernel, the Bessel kernel and the Gauss kernel was studied in [12; 10], [7; 22] and [16], respectively. Existing quadrature methods for oscillatory integrals may be divided into four major categories: asymptotic methods [12], Filon-type methods [10; 15; 16], Levin-type methods [14; 17; 18] and numerical steepest descent methods [2]. All these methods require to know the oscillation level of the integrand in advance. The oscillatory integrals appearing in the context of using DNNs as approximate solutions are generated during the iteration and thus, their oscillatory levels cannot be determined exactly before hand. Therefore, aforementioned methods cannot be applied directly to the current situation. We will design a numerical quadrature that tailors to oscillatory integrals generated during the iteration process and estimate its error bound. Moreover, we will establish the error estimate that bounds the error of the DNN solution by the training loss and the quadrature error.

It is well understood that DNNs have the great expressiveness to present different scales of a function. However, the standard DNN model is not always competent to extract multiscale information of a function as discussed previously. The multigrade deep learning (MGDL) model tailors to the needs of the extraction. The effectiveness of the MGDL model was demonstrated in [25; 26] and [27] in the context of function approximation and numerical solutions of partial differential equations, respectively. We will develop a MGDL model for the DNN solution of the oscillatory Fredholm integral equation and demonstrate numerically that the proposed MGDL model can effectively extract multiscale components of its oscillatory solution, leading to a promising method, unlike the standard DNN model which suffers from the spectral bias.

We organize this paper in eight sections. In Section 2, we outline the oscillatory Fredholm integral equations under consideration and discuss the oscillatory property of its solution. In Section 3, we describe the DNN model for the numerical solution of the oscillatory integral equation and its associated optimization problem. Section 4 is devoted to the development of a numerical quadrature scheme for computing the oscillatory integrals involving DNNs generated during the iteration that solves the optimization problem, and its error analysis. In Section 5, we establish the error estimate for the DNN approximate solution bounded by the training loss and the quadrature error. We describe in Section 6 the MGDL model for the numerical solution of the integral equation. Numerical results are presented in Section 7. Finally, we make conclusive remarks in Section 8.

2. Oscillatory Fredholm integral equation of the second kind

In this section, we describe the Fredholm integral equation of the second kind with an oscillatory kernel to be considered in this paper.

Let $I := [-1, 1]$. We denote by $C(I)$ the space of continuous complex-valued functions defined on I and $C(I^2)$ the space of continuous complex-valued bivariate functions on I^2. Suppose that $K \in C(I^2)$ and $f \in C(I)$ are given. We consider the oscillatory integral equation

$$y(s) - \int_I K(s,t) e^{i\kappa|s-t|} y(t) \mathrm{d}t = f(s), \quad s \in I, \tag{1}$$

where $\kappa \geqslant 1$ is the wavenumber and $y \in C(I)$ is the solution to be solved. For simplicity, in this paper

we always assume the kernel function

$$K(s,t) := \lambda, \quad (s,t) \in I \times I$$

where $\lambda \in \mathbb{C}$, without loss of generality. By defining the integral operator $\mathscr{K}$ for $h \in C(I)$ by

$$(\mathscr{K}h)(s) := \int_I e^{i\kappa|s-t|} h(t)\mathrm{d}t, \quad s \in I,$$

integral equation (1) can be written in its operator form

$$(\mathscr{I} - \lambda\mathscr{K})y = f, \tag{2}$$

where $\mathscr{I}$ denotes the identity operator on $C(I)$.

It is known that the solution y of (2) in general exhibits certain oscillatory behavior. It is advantageous to classify the degree of oscillation for a given function. We recall a definition from [23].

Definition 2.1. Let n be a positive number and $(X, \|\cdot\|_X)$ be a normed space. A function u is called κ-oscillatory of order n in X if it satisfies

(1) u is κ-oscillatory in X, that is, u is κ-parametrized and $u \in X$ for any κ,

(2) there exist a positive constant c such that for all $\kappa > 1$,

$$\kappa^{-n}\|u\|_X \leqslant c.$$

When $n = 0$, we say that u is non-κ-oscillatory in X.

Let m be a fixed positive integer. For Sobolev space $H^m(I) := \{u \in L^2(I) : u^{(n)} \in L^2(I), n \in \mathbb{Z}_{m+1}\}$ where $\mathbb{Z}_{m+1} := \{0, 1, 2, \ldots, m\}$ with the norm

$$\|u\|_{H^m} := \left(\sum_{j\in\mathbb{Z}_{m+1}} \left\|u^{(j)}\right\|_2^2\right)^{1/2},$$

define the space of the oscillatory functions $H^m_{\kappa,0}(I)$ as

$$H^m_{\kappa,0}(I) := \{u_1 + u_2 e^{i\kappa\cdot} + u_3 e^{-i\kappa\cdot} : u_j \text{ is non-}\kappa\text{-oscillatory in } H^m(I), \ j \in \mathbb{N}_3\},$$

where $\mathbb{N}_n := \{1, 2, \ldots, n\}$. Then Theorem 5.8 of [23] shows that the solution y exhibits certain oscillatory behavior. We state the result below for convenience of reference. To this end, for each $n \in \mathbb{N}$, we denote by $C^{[n]}(I^2)$ the space of functions whose derivatives of order up to n with respect to both the first and second variables are continuous.

Theorem 2.1. *Let y be the solution of (1). If $K \in C^{[m]}(I^2)$ is independent of κ and $f \in H^m_{\kappa,0}(I)$, then $y \in H^m_{\kappa,0}(I)$.*

This theorem shows the solution y is in the space of oscillatory function $H^m_{\kappa,0}(I)$, which is the basis for our further discussion and numerical experiments.

For numerical solutions of the integral equation (2), traditional methods typically focus on constructing an appropriate function space where a viable solution exists, and subsequently seek the solution within this space. Commonly utilized function spaces encompass polynomial spaces, trigonometric function spaces, piecewise polynomial spaces, and multi-scale spaces. The construction of a function space

typically necessitates a profound understanding of the solution's characteristics. In contrast, the DNN method directly seeks the solution through training data. Specifically, this methodology employs a DNN with unknown parameters to approximate the desired solution. The equation and constraints are then transformed into an optimization problem. Finally, the optimal parameters of the network are determined. In the subsequent section, we will introduce the DNN method to solve (2).

3. DNN learning model

In this section, we describe a DNN learning model for solving the oscillatory Fredholm integral equation.

We begin by describing DNNs to be used as approximate solutions of the integral equation. We adopt the notation from [29; 30]. A DNN is a function formed by compositions of vector-valued functions, each of which is defined by an activation function applied to an affine map. Given a uni-variate activation function $\sigma : \mathbb{R} \to \mathbb{R}$, the vector-valued activation function $\boldsymbol{\sigma} : \mathbb{R}^d \to \mathbb{R}^d$ is defined as

$$\boldsymbol{\sigma}(\boldsymbol{v}) := [\sigma(v_1), \sigma(v_2), \ldots, \sigma(v_d)]^T, \quad \text{for} \quad \boldsymbol{v} := [v_1, v_2, \ldots, v_d]^T \in \mathbb{R}^d.$$

For vector-valued functions f_j, $j \in \mathbb{N}_n$, satisfying the condition that the range of f_j is contained in the domain of f_{j+1}, their consecutive composition is denoted by

$$\bigodot_{j=1}^{n} f_j := f_n \circ f_{n-1} \circ \cdots \circ f_1.$$

Suppose that positive integers m_j, $j \in \mathbb{Z}_{n+1}$, with $m_0 := 1, m_n := 2$, are chosen. For weight matrices $\boldsymbol{W}_j \in \mathbb{R}^{m_j \times m_{j-1}}$ and the bias vectors $\boldsymbol{b}_j \in \mathbb{R}^{m_j}$, $j \in \mathbb{N}_n$, a DNN is a function defined by

$$\mathscr{N}_n\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; s\big) := \left(\boldsymbol{W}_n \bigodot_{j=1}^{n-1} \boldsymbol{\sigma}(\boldsymbol{W}_j \cdot + \boldsymbol{b}_j) + \boldsymbol{b}_n\right)(s), \quad s \in I. \tag{3}$$

The goal of this paper is to construct approximate solutions, of integral equation (1), which take the form of (3). Note that the solution of (1) is a complex-valued function. This requires us to define an operator $\mathscr{T}$ that transforms a real two dimensional vector-valued function to a complex-valued function. Specifically, for a two-dimensional vector-valued function $\boldsymbol{f}(s) := [f_1(s), f_2(s)]^T$, $s \in I$, with f_1, f_2 being real-valued functions defined on I, the operator $\mathscr{T}$ is defined as $(\mathscr{T}\boldsymbol{f})(s) := f_1(s) + if_2(s)$, $s \in I$. With this operator, we define the loss function as

$$e\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; s\big) := (\mathscr{I} - \lambda\mathscr{K})\mathscr{T}\mathscr{N}_n\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; \cdot\big)(s) - f(s), \quad s \in I. \tag{4}$$

We then find the optimal parameters $\{\boldsymbol{W}_j^*, \boldsymbol{b}_j^*\}_{j=1}^n$ by solving the optimization problem

$$\min_{\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n} \big\| e\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; \cdot\big) \big\|_2^2. \tag{5}$$

Once the optimal parameters $\{\boldsymbol{W}_j^*, \boldsymbol{b}_j^*\}_{j=1}^n$ is obtained, the function $y^* := \mathscr{T}\mathscr{N}_n\big(\{\boldsymbol{W}_j^*, \boldsymbol{b}_j^*\}_{j=1}^n; \cdot\big)$ provides a DNN solution of (1).

Numerical implementation for solving optimization problem (5) requires the availability of collocation data $\{(x_j, f(x_j)), j \in \mathbb{N}_N\}$ and discretization of the integral operator 2. The L_2 norm used in (5) may be

approximated by its discrete form

$$\|\phi\|_N := \sqrt{\frac{1}{N}\sum_{j\in\mathbb{N}_N} |\phi(x_j)|^2}, \quad \text{for } \phi \in C(I),$$

where $|\cdot|$ denotes the modulus of a complex number. It is clear that the functional $\|\cdot\|_N$ is a seminorm on $C(I)$, but it is not a norm since $\|\phi\|_N = 0$ does not imply $\phi = 0$. This seminorm $\|\cdot\|_N$ is associated with the l_2 norm in $\mathbb{C}^N$. For any $\phi \in C(I)$, defining $\boldsymbol{v}_\phi := [\phi(x_j),\, j \in \mathbb{N}_N]$, it follows from the definition of $\|\cdot\|_N$ that

$$\|\phi\|_N = \frac{\|\boldsymbol{v}_\phi\|_{\ell_2}}{\sqrt{N}}. \tag{6}$$

We then assume that there is a discrete oscillatory integral operator $\mathscr{K}_{p_\kappa}$ that approximates the operator $\mathscr{K}$ defined by 2, where p_κ is a positive integer dependent on κ. We postpone the construction of operator $\mathscr{K}_{p_\kappa}$ to the next section.

With the availability of the discrete form of the L_2-norm and the discrete oscillatory integral operator $\mathscr{K}_{p_\kappa}$, the discrete loss function is defined by

$$\tilde{e}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; s\big) := \big(f - (\mathscr{I} - \lambda\mathscr{K}_{p_\kappa})\mathscr{T}\mathscr{N}_n\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; \cdot\big)\big)(s), \quad s \in I, \tag{7}$$

which approximates the loss function defined by (4). We then find the optimal parameters $\{\tilde{\boldsymbol{W}}_j^*, \tilde{\boldsymbol{b}}_j^*\}_{j=1}^n$ by solving the discrete optimization problem

$$\arg\min_{\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n} \big\|\tilde{e}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; \cdot\big)\big\|_N^2, \tag{8}$$

where

$$\big\|\tilde{e}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; \cdot\big)\big\|_N^2 := \frac{1}{N}\sum_{l\in\mathbb{N}_N} |\tilde{e}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^n; x_l\big)|^2.$$

Upon finding the parameters $\{\tilde{\boldsymbol{W}}_j^*, \tilde{\boldsymbol{b}}_j^*\}_{j=1}^n$, we obtain a numerical solution of (2) given by

$$\tilde{y}^* := \mathscr{T}\mathscr{N}_n\big(\{\tilde{\boldsymbol{W}}_j^*, \tilde{\boldsymbol{b}}_j^*\}_{j=1}^n; \cdot\big) \tag{9}$$

with an error defined by

$$\tilde{e}^*(s) := \tilde{e}\big(\{\tilde{\boldsymbol{W}}_j^*, \tilde{\boldsymbol{b}}_j^*\}_{j=1}^n; s\big), \quad s \in I. \tag{10}$$

Optimization problem (8) is often solved by employing gradient-based algorithms. This motivates us to construct the discrete operator $\mathscr{K}_{p_\kappa}$ via a numerical integration method.

4. Numerical quadrature of oscillatory integrals

This section is devoted to the development of the discrete oscillatory integral operator $\mathscr{K}_{p_\kappa}$ and its error estimates.

The discrete oscillatory integral operator $\mathscr{K}_{p_\kappa}$ that approximates the oscillatory integral operator $\mathscr{K}$ is a key attribute of the DNN model. The discrete operator $\mathscr{K}_{p_\kappa}$ should effectively approximate the continuous oscillatory integral operator $\mathscr{K}$. Typically, in the optimization process, the optimization problem (8) is

solved by using gradient-based optimization algorithms. These algorithms iteratively update the model parameters $\theta_j := \{\boldsymbol{W}_{l,j}, \boldsymbol{b}_{l,j}\}_{l=1}^n$, where $j \in \mathbb{N} := \{1, 2, 3, \dots\}$. By defining

$$g_j(s) := \mathscr{T}\mathscr{N}_n(\theta_j; s) \in C(I), \quad j \in \mathbb{N}, \quad s \in I,$$

we will construct $\mathscr{K}_{p_\kappa}$ in a way that the errors $|(\mathscr{K} g_j - \mathscr{K}_{p_\kappa} g_j)(x_l)|, l \in \mathbb{N}_N, j \in \mathbb{N}$ are as small as possible. Noting that for any $l \in \mathbb{N}_N$,

$$|(\mathscr{K} g_j - \mathscr{K}_{p_\kappa} g_j)(x_l)| \leqslant \|\mathscr{K} g_j - \mathscr{K}_{p_\kappa} g_j\|_\infty, \quad j \in \mathbb{N},$$

our objective is to construct the operator $\mathscr{K}_{p_\kappa}$ to control $\|\mathscr{K}\chi - \mathscr{K}_{p_\kappa}\chi\|_\infty$ when wavenumber κ is large for χ in the class of functions $\{g_j\}_{j\in\mathbb{N}}$.

The behavior regarding the level of oscillation in the function sequence $\{g_j\}_{j\in\mathbb{N}}$ is a crucial consideration, especially given the use of the sin function as the activation function in the DNN, which directly influences the oscillatory behavior through the parameter θ_j, $j \in \mathbb{N}$. The initial guess θ_1 proposed in [11] leads to a non-oscillatory function g_1 independent of κ, setting the starting point for the iteration. If the iteration converges, the oscillation of the generated function sequence $\{g_j\}_{j\in\mathbb{N}}$ will not deviate significantly from the oscillation of the exact solution y. Below, we define a class of oscillatory functions with a relaxed oscillation level so that the class contains the exact solution y and approximate solutions g_j of the j-th iteration steps.

Motivated by Theorem 2.1, we assume that there exist functions u_j, $j \in \mathbb{N}_3$, satisfying the condition for some $\tau > 0$,

$$|u_j^{(l)}(s)| \leqslant \tau, \quad \text{for all } \ s \in I, \ \ l \in \mathbb{Z}_{m+1},$$

such that the solution y of integral equation (2) is in $H^m_{\kappa,0}(I)$. In other words, y can be expressed as

$$y(s) = u_1(s) + u_2(s)e^{i\kappa s} + u_3(s)e^{-i\kappa s}, \quad s \in I. \tag{11}$$

The functions g_j are not exactly in $H^m_{\kappa,0}(I)$ but in its perturbation. Below, we define a perturbation class of $H^m_{\kappa,0}(I)$ that contains the functions g_j. Let $\Gamma \geqslant 0$ and $r \in \mathbb{N}$. Suppose that $\alpha_j \in \mathbb{R}$, $j \in \mathbb{N}_r$, are unknown but satisfy $|\alpha_j| \leqslant 1 + \Gamma$, and functions w_j, $j \in \mathbb{N}_r$, satisfy

$$|w_j^{(l)}(s)| \leqslant \tau, \quad s \in I, \ \ l \in \mathbb{Z}_{m+1}. \tag{12}$$

The functions χ_κ in the perturbation class have the form

$$\chi_\kappa(s) := \sum_{j=1}^{r} w_j(s)e^{i\alpha_j\kappa s}, \quad s \in I. \tag{13}$$

We next introduce a sequence of discrete operators $\mathscr{K}_p$, for $p \in \mathbb{N}$, which approximate the integral operator $\mathscr{K}$ by using the compound trapezoidal quadrature formula. Specifically, for each $p \in \mathbb{N}$, let $h := 2/p$, $s_j := -1 + jh$ for $j \in \mathbb{Z}_{p+1}$, and for any $F \in C(I)$, we define

$$(\mathscr{K}_p F)(s) := \frac{h}{2}\left[F(s_0)e^{i\kappa|s-s_0|} + 2\sum_{j=1}^{p-1} F(s_j)e^{i\kappa|s-s_j|} + F(s_p)e^{i\kappa|s-s_p|}\right], \quad s \in I. \tag{14}$$

In the rest of this section, we choose the value of p according to the growth of the error $\|\mathscr{K}\chi_\kappa - \mathscr{K}_p\chi_\kappa\|_\infty$ with respect to the wavenumber κ. Error analysis of the compound trapezoidal quadrature formula for

smooth non-oscillatory integrands is well-understood (for example, see [8]). However, since the integrand that we consider here is oscillatory and non-differentiable, we will estimate the error by taking the special property of the integrand into account.

When analyzing the error $\|\mathscr{K}\chi_\kappa - \mathscr{K}_p\chi_\kappa\|_\infty$, due to (13), it suffices to consider

$$\tilde{\chi}_\kappa(t) := w(t)e^{i\alpha\kappa t}, \quad t \in I, \tag{15}$$

where $\alpha \in \mathbb{R}$ satisfies $|\alpha| \leqslant 1+\Gamma$, and function w is m times differentiable and satisfies (12). Our goal is to develop an effective quadrature scheme for numerical computation of the integral

$$(\mathscr{K}w(\cdot)e^{i\alpha\kappa\cdot})(s) := \int_I \left[w(t)e^{i\kappa\alpha t}\right] e^{i\kappa|s-t|}\mathrm{d}t, \quad s \in I,$$

for the purpose of constructing the discrete operator $\mathscr{K}_{p_\kappa}$.

Next, we estimate $\|\mathscr{K}\tilde{\chi}_\kappa - \mathscr{K}_p\tilde{\chi}_\kappa\|_\infty$. To this end, for a fixed $s \in I$, we define

$$\phi(t) := w(t)e^{i\kappa(\alpha t+|s-t|)}, \quad t \in I. \tag{16}$$

Then, $(\mathscr{K}\tilde{\chi}_\kappa - \mathscr{K}_p\tilde{\chi}_\kappa)(s)$ can be expressed as

$$(\mathscr{K}\tilde{\chi}_\kappa - \mathscr{K}_p\tilde{\chi}_\kappa)(s) = \sum_{j=0}^{p-1}\left[\int_{s_j}^{s_{j+1}}\phi(t)\mathrm{d}t - \frac{h}{2}\left(\phi(s_j)+\phi(s_{j+1})\right)\right], \tag{17}$$

and for any $j \in \mathbb{Z}_p$, there holds that $s_{j+1} - s_j = h$.

Since the integrand ϕ defined by (16) is m times differentiable at $I \setminus \{s\}$ for any positive integer m, we partition interval I into three subintervals such that except for the interval containing the point s, the function ϕ remains m times differentiable over the other two intervals. Specifically, we choose

$$j^* := \begin{cases} \frac{s+1}{h} - 1, & \text{if } \frac{s+1}{h} \in \mathbb{N}, \\ \lfloor \frac{s+1}{h} \rfloor, & \text{otherwise}, \end{cases}$$

where for $x \in \mathbb{R}$, $\lfloor x \rfloor$ represents the largest integer not exceeding x. Thus, $j^* \in \mathbb{Z}_p$. Noticing

$$|s-t| = \begin{cases} s-t, & t \in [s_0, s_{j^*}], \\ t-s, & t \in [s_{j^*+1}, s_p], \end{cases} \tag{18}$$

together with the definition (16) of function ϕ, we conclude that ϕ is analytic in the intervals $[s_0, s_{j^*}]$ and $[s_{j^*+1}, s_p]$. If we define an m-times differentiable function

$$\psi(t) := \begin{cases} w(t)e^{i(\alpha-1)\kappa t}, & t \in [s_0, s_{j^*}], \\ w(t)e^{i(\alpha+1)\kappa t}, & t \in [s_{j^*+1}, s_p], \end{cases} \tag{19}$$

together with the relation (18), we obtain

$$\phi(t) = \begin{cases} e^{i\kappa s}\psi(t), & t \in [s_0, s_{j^*}], \\ e^{-i\kappa s}\psi(t), & t \in [s_{j^*+1}, s_p]. \end{cases}$$

Upon defining

$$T_1 := e^{i\kappa s} \sum_{j=0}^{j^*-1} \left[\int_{s_j}^{s_{j+1}} \psi(t)\,\mathrm{d}t - \frac{h}{2}\left(\psi(s_j) + \psi(s_{j+1})\right)\right], \tag{20}$$

$$T_2 := \int_{s_{j^*}}^{s_{j^*+1}} \phi(t)\mathrm{d}t - \frac{h}{2}\left(\phi(s_{j^*}) + \phi(s_{j^*+1})\right), \tag{21}$$

$$T_3 := e^{-i\kappa s} \sum_{j=j^*+1}^{p-1} \left[\int_{s_j}^{s_{j+1}} \psi(t)\,\mathrm{d}t - \frac{h}{2}\left(\psi(s_j) + \psi(s_{j+1})\right)\right], \tag{22}$$

equation (17) can be rewritten as

$$(\mathscr{K}\tilde{\chi}_\kappa - \mathscr{K}_p\tilde{\chi}_\kappa)(s) = T_1 + T_2 + T_3.$$

Using the triangle inequality in the above equation, it yields that

$$|(\mathscr{K}\tilde{\chi}_\kappa - \mathscr{K}_p\tilde{\chi}_\kappa)(s)| \leqslant |T_1| + |T_2| + |T_3|. \tag{23}$$

We will estimate T_j, $j = 1, 2, 3$, separately.

We first estimate $|T_2|$ in the following lemma.

Lemma 4.1. *If T_2 is defined in (21), then*

$$|T_2| \leqslant \frac{4\tau}{p}.$$

Proof. Applying the triangle inequality to the definition of T_2, it yields that

$$|T_2| \leqslant \int_{s_{j^*}}^{s_{j^*+1}} |\phi(t)|\ \mathrm{d}t + h\|\phi\|_\infty \leqslant (s_{j^*+1} - s_{j^*} + h)\|\phi\|_\infty.$$

This together with the relation $s_{j^*+1} - s_{j^*} = h$ leads to the bound

$$|T_2| \leqslant 2h\|\phi\|_\infty. \tag{24}$$

Using the inequality (12) with $l := 0$, we know that $\|w\|_\infty \leqslant \tau$. Combining this with the definition of ϕ, we observe that $\|\phi\|_\infty = \|w\|_\infty \leqslant \tau$. Thus, the inequality (24) reduces to $|T_2| \leqslant 2\tau h$, which together with the fact $h = 2/p$ yields the desired result. □

We next estimate the terms $|T_1|$ and $|T_3|$. In the following consideration, we unify the cases $|T_1|$ and $|T_3|$ into one. For $\mu, \nu \in \mathbb{Z}_p$ satisfying $0 \leqslant \mu \leqslant \nu \leqslant p-1$ with $j^* \notin [\mu, \nu]$, we consider

$$T := \sum_{j=\mu}^{\nu} \left[\int_{s_j}^{s_{j+1}} \psi(t)\,\mathrm{d}t - \frac{h}{2}\left(\psi(s_j) + \psi(s_{j+1})\right)\right]. \tag{25}$$

Clearly, if we choose $\mu := 0$, $\nu := j^* - 1$, then $|T| = |T_1|$ by (20), and if we choose $\mu := j^* + 1$, $\nu := p - 1$, then $|T| = |T_3|$ by (22).

To estimate $|T|$, we will make use of the Taylor expansion of the function ψ at the midpoint of each subinterval. Specifically, for each $j \in \mathbb{Z}_{[\mu,\nu]} := \{\mu, \mu+1, \ldots, \nu\}$, we let $s_{j+1/2} := \frac{1}{2}(s_j + s_{j+1})$, and

write $\psi^{(l)}_{j+1/2} := \psi^{(l)}(s_{j+1/2})$ for $l \in \mathbb{Z}_{m+1}$. By the Taylor theorem, for each $t \in [s_j, s_{j+1}]$, there exists a ξ_t between t and $s_{j+1/2}$ such that

$$\psi(t) = \psi^{(0)}_{j+1/2} + \sum_{l=1}^{m-1} \frac{(t - s_{j+1/2})^l}{l!} \psi^{(l)}_{j+1/2} + \frac{(t - s_{j+1/2})^m}{m!} \psi^{(m)}(\xi_t). \tag{26}$$

For each $j \in \mathbb{Z}_{[\mu,\nu]}$, we let

$$D_{j,1} := \frac{(-h)^m}{2^m m!} \psi^{(m)}(\xi_{s_j}) + \frac{h^m}{2^m m!} \psi^{(m)}(\xi_{s_{j+1}}). \tag{27}$$

By the Taylor expansion (26), we obtain that

$$\frac{h}{2}[\psi(s_j) + \psi(s_{j+1})] = h\psi^{(0)}_{j+1/2} + \sum_{l=1}^{\tilde{m}} \frac{2}{(2l)!} \left(\frac{h}{2}\right)^{2l+1} \psi^{(2l)}_{j+1/2} + \frac{hD_{j,1}}{2}, \tag{28}$$

where $\tilde{m} := \lfloor (m-1)/2 \rfloor$. Meanwhile, by integrating both sides of (26) from s_j to s_{j+1}, and letting

$$D_{j,2} := \int_{s_j}^{s_{j+1}} \frac{(t - s_{j+1/2})^m}{m!} \psi^{(m)}(\xi_t) \mathrm{d}t, \tag{29}$$

we have

$$\int_{s_j}^{s_{j+1}} \psi(t)\, \mathrm{d}t = h\psi^{(0)}_{j+1/2} + \sum_{l=1}^{\tilde{m}} \frac{2}{(2l+1)!} \left(\frac{h}{2}\right)^{2l+1} \psi^{(2l)}_{j+1/2} + D_{j,2}. \tag{30}$$

Substituting (28) and (30) into the right-hand side of (25) yields

$$T = \sum_{j=\mu}^{\nu} \left(- \sum_{l=1}^{\tilde{m}} \frac{2l h^{2l+1}}{4^l (2l+1)!} \psi^{(2l)}_{j+1/2} + D_{j,2} - \frac{hD_{j,1}}{2} \right). \tag{31}$$

Upon defining

$$U_l := h \sum_{j=\mu}^{\nu} \psi^{(2l)}_{j+1/2} \quad \text{and} \quad \eta_l := \frac{2l}{4^l (2l+1)!}, \quad l \in \mathbb{N}_{\tilde{m}}, \tag{32}$$

we rewrite (31) as

$$T = - \sum_{l=1}^{\tilde{m}} \eta_l h^{2l} U_l + \sum_{j=\mu}^{\nu} \left(D_{j,2} - \frac{hD_{j,1}}{2} \right). \tag{33}$$

We begin estimating $|T|$ by establishing the inequality

$$|T| \leqslant \left| T + \sum_{l=1}^{\tilde{m}} \eta_l h^{2l} U_l \right| + \left| \sum_{l=1}^{\tilde{m}} \eta_l h^{2l} U_l \right|. \tag{34}$$

The next lemma estimates the first term of the right-hand side of (34).

Lemma 4.2. *If T is defined as in (25), then*

$$\left| T + \sum_{l=1}^{\tilde{m}} \eta_l h^{2l} U_l \right| \leqslant 3(\nu - \mu + 1) \|\psi^{(m)}\|_{L_\infty([s_\mu, s_{\nu+1}])} \left(\frac{h}{2}\right)^{m+1}. \tag{35}$$

Proof. By (33) and the triangle inequality, we have

$$\left|T+\sum_{l=1}^{\tilde{m}}\eta_l h^{2l}U_l\right| \leqslant \sum_{j=\mu}^{\nu}\left(|D_{j,2}|+\frac{h|D_{j,1}|}{2}\right). \tag{36}$$

It suffices to estimate the remainders $D_{j,1}$, $D_{j,2}$, $j\in\mathbb{Z}_{[\mu,\nu]}$, defined respectively in (27) and (29), of the Taylor expansion. Clearly, for each $j\in\mathbb{Z}_{[\mu,\nu]}$, we have the estimates

$$|D_{j,1}| \leqslant \frac{h^m\|\psi^{(m)}\|_\infty}{2^{m-1}m!},$$

where $\|\psi^{(m)}\|_\infty := \|\psi^{(m)}\|_{L_\infty([s_\mu,s_{\nu+1}])}$ and

$$|D_{j,2}| \leqslant \frac{\|\psi^{(m)}\|_\infty}{m!}\int_{s_j}^{s_{j+1}}|t-s_{j+1/2}|^m\mathrm{d}t = \frac{h^{m+1}\|\psi^{(m)}\|_\infty}{2^m(m+1)!}.$$

Substituting the above two estimates into the right-hand side of inequality (36) with noticing $\frac{m+2}{(m+1)!}\leqslant\frac{3}{2}$ yields the desired result. □

We now consider the second term of the right-hand side of (34), which involves U_l defined by (32). Note that when $h=\frac{2}{p}$ and $\nu-\mu\leqslant p-1$, there holds

$$|U_l| = \left|h\sum_{j=\mu}^{\nu}\psi_{j+1/2}^{(2l)}\right| \leqslant \frac{2(\nu-\mu+1)\|\psi^{(2l)}\|_\infty}{p} \leqslant 2\|\psi^{(2l)}\|_\infty, \quad l\in\mathbb{N}_{\tilde{m}}.$$

We will see later that $\|\psi^{(2l)}\|_\infty$ is in the order of κ^{2l}, which will lead to the conclusion that the second term of the right-hand side of (34) has a leading term in the order of κ^2. We would like to obtain a better estimate so that the leading term is in the order of κ, not κ^2. To this end, we need the Taylor expansion of the function $\psi^{(2l)}$, $l\in\mathbb{N}_{\tilde{m}}$, at the midpoint of the every subinterval. For each $j\in\mathbb{Z}_{[\mu,\nu]}$, and $l\in\mathbb{N}_{\tilde{m}}$, the Taylor theorem ensures that for each $t\in[s_j,s_{j+1}]$, there exists a $\xi_{l,t}$ between t and $s_{j+1/2}$ such that

$$\psi^{(2l)}(t) = \psi_{j+1/2}^{(2l)} + \sum_{b=1}^{m-2l-1}\frac{(t-s_{j+1/2})^{(b)}}{b!}\psi_{j+1/2}^{(2l+b)} + \frac{(t-s_{j+1/2})^{m-2l}}{(m-2l)!}\psi^{(m)}(\xi_{l,t}). \tag{37}$$

Integrating both sides of (37) from s_j to s_{j+1} and defining

$$R_{j,l} := \int_{s_j}^{s_{j+1}}\frac{(t-s_{j+1/2})^{m-2l}}{(m-2l)!}\psi^{(m)}(\xi_{l,t})\mathrm{d}t,$$

we obtain

$$\psi^{(2l-1)}(s_{j+1})-\psi^{(2l-1)}(s_j) = h\psi_{j+1/2}^{(2l)} + \sum_{b=1}^{\tilde{m}-l}\frac{2}{(2b+1)!}\left(\frac{h}{2}\right)^{2b+1}\psi_{j+1/2}^{(2l+2b)} + R_{j,l}, \quad j\in\mathbb{Z}_{[\mu,\nu]}. \tag{38}$$

Summing up (38) for $j\in\mathbb{Z}_{[\mu,\nu]}$, together with the definition (32) of η_l, we obtain that

$$\psi^{(2l-1)}(s_{\nu+1})-\psi^{(2l-1)}(s_\mu) = h\sum_{j=\mu}^{\nu}\psi_{j+1/2}^{(2l)} + \sum_{j=\mu}^{\nu}\sum_{b=1}^{\tilde{m}-l}\frac{\eta_b}{2b}h^{2b+1}\psi_{j+1/2}^{(2l+2b)} + \sum_{j=\mu}^{\nu}R_{j,l}, \quad l\in\mathbb{N}_{\tilde{m}}.$$

With the definition of $U_l, l \in \mathbb{N}_{\tilde{m}}$ in (32), the above equation can be rewritten as

$$U_l + \sum_{b=1}^{\tilde{m}-l} \frac{\eta_b}{2b} h^{2b} U_{b+l} = \psi^{(2l-1)}(s_{\nu+1}) - \psi^{(2l-1)}(s_\mu) - \sum_{j=\mu}^{\nu} R_{j,l}, \quad l \in \mathbb{N}_{\tilde{m}}. \tag{39}$$

We now return to investigating the second term of the right-hand side of (34).

Lemma 4.3. *If*

$$\delta_l := \eta_l - \sum_{b=1}^{l-1} \frac{\eta_{l-b}\delta_b}{2l-2b}, \quad l \in \mathbb{N}, \tag{40}$$

then

$$\sum_{l=1}^{\tilde{m}} \eta_l h^{2l} U_l = \sum_{l=1}^{\tilde{m}} \delta_l h^{2l} \left(\psi^{(2l-1)}(s_{\nu+1}) - \psi^{(2l-1)}(s_\mu) - \sum_{j=\mu}^{\nu} R_{j,l} \right). \tag{41}$$

Proof. According to (39), it suffices to prove that

$$\sum_{l=1}^{\tilde{m}} \eta_l h^{2l} U_l = \sum_{l=1}^{\tilde{m}} \delta_l h^{2l} \left(U_l + \sum_{b=1}^{\tilde{m}-l} \frac{\eta_b}{2b} h^{2b} U_{b+l} \right). \tag{42}$$

By direct computation, we observe that

$$\sum_{l=1}^{\tilde{m}} \delta_l h^{2l} \left(U_l + \sum_{b=1}^{\tilde{m}-l} \frac{\eta_b}{2b} h^{2b} U_{b+l} \right) = \sum_{l=1}^{\tilde{m}} \left(\delta_l + \sum_{b=1}^{l-1} \frac{\eta_{l-b}\delta_b}{2l-2b} \right) h^{2l} U_l. \tag{43}$$

By the definition (40) of δ_l, $l \in \mathbb{N}$, we have

$$\eta_l = \delta_l + \sum_{b=1}^{l-1} \frac{\eta_{l-b}\delta_b}{2l-2b}, \quad l \in \mathbb{N}.$$

Substituting this into the right-hand side of (43) yields the desired (42). □

Substituting (41) into the second term in the right-hand side of inequality (34) yields

$$|T| \leqslant \left| T + \sum_{l=1}^{\tilde{m}} \eta_l h^{2l} U_l \right| + \sum_{l=1}^{\tilde{m}} |\delta_l| h^{2l} \left(|\psi^{(2l-1)}(s_{\nu+1}) - \psi^{(2l-1)}(s_\mu)| + \sum_{j=\mu}^{\nu} |R_{j,l}| \right). \tag{44}$$

Substituting

$$|\psi^{(2l-1)}(s_{\nu+1}) - \psi^{(2l-1)}(s_\mu)| \leqslant 2\|\psi^{(2l-1)}\|_\infty,$$

$$|R_{j,l}| \leqslant \frac{\|\psi^{(m)}\|_\infty}{(m-2l)!} \int_{s_j}^{s_{j+1}} |t - s_{j+1/2}|^{m-2l} \mathrm{d}t = \frac{h^{m-2l+1}\|\psi^{(m)}\|_\infty}{2^{m-2l}(m-2l+1)!} \leqslant 2\|\psi^{(m)}\|_\infty \left(\frac{h}{2}\right)^{m-2l+1},$$

and estimate (35) into the right-hand side of inequality (44), we obtain

$$|T| \leqslant 2 \sum_{l=1}^{\tilde{m}} h^{2l} |\delta_l| \|\psi^{(2l-1)}\|_\infty + (\nu - \mu + 1) \left(3 + 2 \sum_{l=1}^{\tilde{m}} 4^l |\delta_l| \right) \|\psi^{(m)}\|_\infty \left(\frac{h}{2}\right)^{m+1}. \tag{45}$$

By the above inequality, we estimate the growth of derivatives of function ψ in the following lemma.

Lemma 4.4. *If ψ is defined in (19), then*

$$\|\psi^{(l)}\|_\infty \leqslant \tau[(\Gamma+2)\kappa+1]^l, \quad l \in \mathbb{Z}_{m+1}, \tag{46}$$

where $\|\psi^{(m)}\|_\infty := \|\psi^{(m)}\|_{L_\infty([s_\mu, s_{\nu+1}])}$.

Proof. Without loss of generality, we will assume that $0 \leqslant \mu \leqslant \nu \leqslant j^* - 1$. In this case, the definition (19) of ψ yields that $\psi(t) = w(t)e^{i(\alpha-1)\kappa t}$, $t \in [s_\mu, s_{\nu+1}]$. Repeatedly differentiating ψ yields

$$\psi^{(l)}(s) = e^{i(\alpha-1)\kappa s} \sum_{d=0}^{l} C_{l,d} w^{(d)}(s), \quad s \in [s_\mu, s_\nu], \quad l \in \mathbb{Z}_{m+1}, \tag{47}$$

where $C_{0,0} := 1$ and for $l \in \mathbb{N}_m$,

$$C_{l,d} := \begin{cases} i(\alpha-1)\kappa C_{l-1,d}, & d = 0, \\ i(\alpha-1)\kappa C_{l-1,d} + C_{l-1,d-1}, & d = 1, 2, \ldots, l-1, \\ C_{l-1,d-1}, & d = l. \end{cases}$$

It can be shown by induction on l that

$$\sum_{d=0}^{l} |C_{l,d}| \leqslant (1 + |\alpha - 1|\kappa)^l, \quad l \in \mathbb{Z}_{m+1}. \tag{48}$$

Now, by applying the L_∞-norm of $\psi^{(l)}$ to (47), we obtain that

$$\|\psi^{(l)}\|_\infty \leqslant \sum_{d=0}^{l} |C_{l,d}| \|w^{(d)}\|_\infty, \quad l \in \mathbb{Z}_{m+1}. \tag{49}$$

Substituting (12) and (48) into the right-hand side of (49) yields

$$\|\psi^{(l)}\|_\infty \leqslant \tau(1 + |\alpha - 1|\kappa)^l, \quad l \in \mathbb{Z}_{m+1}.$$

The above inequality with $|\alpha - 1| \leqslant |\alpha| + 1 \leqslant \Gamma + 2$ leads to the desired estimate. □

We next estimate $|T|$. To this end, we define $\Delta(\tilde{m}) := \sum_{l=1}^{\tilde{m}} 4^l |\delta_l|$ and $G_{p,\kappa} := \dfrac{(\Gamma+2)\kappa+1}{p}$.

Lemma 4.5. *If $p \in \mathbb{N}$ is chosen to satisfy $p \geqslant (\Gamma+2)\kappa + 1$, then*

$$|T| \leqslant \frac{\tau}{p}\big(2\Delta(\tilde{m})G_{p,\kappa} + (3 + 2\Delta(\tilde{m}))\, G_{p,\kappa}^m\big). \tag{50}$$

Proof. We prove this lemma by employing inequality (45). As for the first term in the right-hand side of inequality (45), by inequality (46) in Lemma 4.4 and the fact $h = 2/p$, with noting $G_{p,\kappa} \leqslant 1$ by the choice of p, we observe that

$$\begin{aligned} 2\sum_{l=1}^{\tilde{m}} h^{2l} |\delta_l| \|\psi^{(2l-1)}\|_\infty &\leqslant \frac{2\tau((\Gamma+2)\kappa+1)}{p^2} \sum_{l=1}^{\tilde{m}} 4^l |\delta_l| \left(\frac{(\Gamma+2)\kappa+1}{p}\right)^{2l-2} \\ &= \frac{2\tau G_{p,\kappa}}{p} \sum_{l=1}^{\tilde{m}} 4^l |\delta_l| G_{p,\kappa}^{2l-2} \leqslant \frac{2\tau G_{p,\kappa}\Delta(\tilde{m})}{p}. \end{aligned}$$

As for the second term in the right-hand side of inequality (45), for the same reason, we obtain

$$(3+2\Delta(\tilde{m}))\,\|\psi^{(m)}\|_\infty \left(\frac{h}{2}\right)^{m+1} \leqslant \frac{\tau\,(3+2\Delta(\tilde{m}))\,G^m_{p,\kappa}}{p}.$$

Substituting the above two inequalities into the right-hand side of inequality (45) yields the desired estimate (50). □

Note that $\Delta(\tilde{m}) \leqslant \sum_{l=1}^{\infty} 4^l|\delta_l|$, which is estimated in the next lemma.

Lemma 4.6. *If δ_l, $l \in \mathbb{N}$, is the sequence defined in (40), then*

$$\sum_{l=1}^{\infty} 4^l|\delta_l| \leqslant \frac{6}{5}.$$

Proof. Defining

$$\tilde{\delta}_l := \frac{1}{(2l)!} + \sum_{b=1}^{l-1} \frac{\tilde{\delta}_{l-b}}{(2b)!}, \quad l \in \mathbb{N}, \tag{51}$$

this lemma may be proved by establishing the two estimates

$$\sum_{l=1}^{\infty} \tilde{\delta}_l \leqslant \frac{6}{5} \qquad \text{and} \qquad 4^l|\delta_l| \leqslant \tilde{\delta}_l, \quad l \in \mathbb{N}. \tag{52}$$

First, we establish the latter inequality via induction on $l \in \mathbb{N}$. By definition (40), the sequence δ_l, $l \in \mathbb{N}$, can be written as

$$\delta_l = \eta_l - \sum_{b=1}^{l-1} \frac{\delta_{l-b}\eta_b}{2b}, \quad l \in \mathbb{N}. \tag{53}$$

For $l = 1$, it is clear that $4|\delta_1| = \frac{1}{3} \leqslant \tilde{\delta}_1 = \frac{1}{2}$. Now, we assume for some $l^* \in \mathbb{N}$ that

$$4^b|\delta_b| \leqslant \tilde{\delta}_b, \quad b \in \mathbb{N}_{l^*} \tag{54}$$

and consider the case $l = l^* + 1$. Using (53) with $l := l^* + 1$ and the fact $4^b\eta_b = \dfrac{2b}{(2b+1)!} \leqslant \dfrac{1}{(2b)!}$, $b \in \mathbb{N}$, we derive that

$$4^{l^*+1}|\delta_{l^*+1}| \leqslant \frac{1}{(2l^*+2)!} + \sum_{b=1}^{l^*} \frac{4^{l^*+1-b}|\delta_{l^*+1-b}|}{(2b)!}.$$

Using (54), this becomes

$$|4^{l^*+1}\delta_{l^*+1}| \leqslant \frac{1}{(2l^*+2)!} + \sum_{b=1}^{l^*} \frac{\tilde{\delta}_{l^*+1-b}}{(2b)!},$$

whose right-hand side is exactly equal to $\tilde{\delta}_{l^*+1}$ by definition (51). Thus, we have proved the desired inequality for the case $l := l^* + 1$. By the induction principle, the inequality holds for all $l \in \mathbb{N}$.

We now prove the first inequality in (52). By defining a positive monotonically increasing sequence

$$\Delta_l := \sum_{b=1}^{l} \tilde{\delta}_b, \quad l \in \mathbb{N}, \tag{55}$$

it suffices to show

$$\lim_{l\to\infty} \Delta_l \leqslant \frac{6}{5}. \tag{56}$$

We first derive a recursive formula for the sequence Δ_l, $l \in \mathbb{N}$. To this end, we substitute definition (51) of $\tilde{\delta}_l$ into definition (55) of Δ_l to obtain that

$$\Delta_l = \sum_{b=1}^{l} \tilde{\delta}_b = \sum_{b=1}^{l} \frac{1}{(2b)!} + \sum_{d=1}^{l} \sum_{b=1}^{d-1} \frac{\tilde{\delta}_{d-b}}{(2b)!}, \quad l \in \mathbb{N}. \tag{57}$$

Since for all $l \in \mathbb{N}$,

$$\sum_{d=1}^{l} \sum_{b=1}^{d-1} \frac{\tilde{\delta}_{d-b}}{(2b)!} = \sum_{b=1}^{l-1} \sum_{d=b+1}^{l} \frac{\tilde{\delta}_{d-b}}{(2b)!} = \sum_{b=1}^{l-1} \frac{\Delta_{l-b}}{(2b)!},$$

where the last equality holds due to definition (55) of Δ_l, equation (57) becomes a recursion of Δ_l:

$$\Delta_l = \sum_{b=1}^{l} \frac{1}{(2b)!} + \sum_{b=1}^{l-1} \frac{\Delta_{l-b}}{(2b)!}, \quad l \in \mathbb{N}. \tag{58}$$

We observe the sequence Δ_l, $l \in \mathbb{N}$ is monotonically increasing. In particular, for $l \geqslant 2$, we have $\Delta_b \leqslant \Delta_l$ for all $b \in \mathbb{N}_{l-1}$. Consequently, (58) implies that

$$\Delta_l \leqslant \sum_{b=1}^{l} \frac{1}{(2b)!} + \Delta_l \sum_{b=1}^{l-1} \frac{1}{(2b)!}, \quad l \geqslant 2.$$

Since $\Delta_l > 0$ for $l \in \mathbb{N}$ and $\sum_{b=1}^{\infty} \frac{1}{(2b)!} = \cosh(1) - 1$, we deduce

$$\Delta_l \leqslant (\cosh(1) - 1)(\Delta_l + 1), \quad l \geqslant 2,$$

namely,

$$\Delta_l \leqslant \frac{\cosh(1) - 1}{2 - \cosh(1)} \leqslant \frac{6}{5}, \quad l \geqslant 2.$$

Thus, by the monotone convergence theorem, we conclude the existence of the limit for sequence Δ_l, $l \in \mathbb{N}$, and confirm inequality (56). □

Combining Lemmas 4.5 and 4.6 yields the following estimate of $|T|$.

Lemma 4.7. *If $p \in \mathbb{N}$ is chosen to satisfy $p \geqslant (\Gamma + 2)\kappa + 1$, then*

$$|T| \leqslant \frac{\tau}{5p}\big(12G_{p,\kappa} + 27(\nu - \mu + 1)G^m_{p,\kappa}\big). \tag{59}$$

Finally, we have the next proposition for an estimate of $\|\mathscr{K}\chi_\kappa - \mathscr{K}_p\chi_\kappa\|_\infty$.

Proposition 4.1. *For a complex-valued analytic χ_κ as defined in (13), for any $\kappa \geqslant 1$, if parameter $p \in \mathbb{N}$ is selected to satisfy $p \geqslant (\Gamma + 2)\kappa + 1$, then*

$$\|\mathscr{K}\chi_\kappa - \mathscr{K}_p\chi_\kappa\|_\infty \leqslant \frac{44r\tau}{5p} + \frac{27r\tau}{5}G^m_{p,\kappa}.$$

Proof. It suffices to prove, for $\tilde{\chi}_\kappa$ as defined in (15), that

$$\|\mathscr{K}\tilde{\chi}_\kappa - \mathscr{K}_p\tilde{\chi}_\kappa\|_\infty \leqslant \frac{44\tau}{5p} + \frac{27\tau}{5}G_{p,\kappa}^m. \tag{60}$$

By letting $\mu := 0$, $\nu := j^* - 1$ and $\mu := j^* + 1$, $\nu := p - 1$ in the definition of T in (25), the inequality (59) in Lemma 4.7 holds for $|T| = |T_1|$ and $|T| = |T_3|$ respectively, namely

$$|T_1| \leqslant \frac{\tau}{5p}\big(12G_{p,\kappa} + 27j^*G_{p,\kappa}^m\big) \quad \text{and} \quad |T_3| \leqslant \frac{\tau}{5p}\big(12G_{p,\kappa} + 27(p - j^* - 1)G_{p,\kappa}^m\big).$$

Together with the estimate of $|T_2|$ in Lemma 4.1 and the fact $j^* + (p - j^* - 1) \leqslant p$, inequality (23) implies that

$$|(\mathscr{K}\tilde{\chi}_\kappa - \mathscr{K}_p\tilde{\chi}_\kappa)(s)| \leqslant \frac{4\tau}{p}\left(1 + \frac{6}{5}G_{p,\kappa}\right) + \frac{27\tau}{5}G_{p,\kappa}^m.$$

The above equality holds for any fixed $s \in I$, with noting $G_{p,\kappa} \leqslant 1$ by the choice of p, yielding the estimate (60). □

Proposition 4.1 elucidates the connection among the error, the sample size p and the wavenumber κ. Next, we aim at ascertaining the value of p as a function of the wavenumber κ such that the error tends to zero as κ approaches infinity. This will be expounded upon in the following proposition. Before this, for any $x \in \mathbb{R}$, $\lceil x \rceil$ denotes the smallest integer greater than or equal to x.

Proposition 4.2. *For the function χ_κ defined in (13), suppose the parameters β and γ are chosen to satisfy $\beta \geqslant 1$ and $\gamma \geqslant \Gamma + 3$. If $p_\kappa := \lceil \gamma\kappa^\beta \rceil$, then for any $\kappa \geqslant 1$,*

$$\|\mathscr{K}\chi_\kappa - \mathscr{K}_{p_\kappa}\chi_\kappa\|_\infty \leqslant \frac{44r\tau}{5\gamma\kappa^\beta} + \frac{27r\tau(\Gamma + 3)^m}{5\gamma^m\kappa^{m(\beta-1)}}.$$

Proof. First, we need to verify $p_\kappa \geqslant (\Gamma + 2)\kappa + 1$ which is the condition of Proposition 4.1. For $\beta \geqslant 1$, we consider $\kappa^{\beta-1}G_{p_\kappa,\kappa}$, $\kappa \geqslant 1$. By the definition of p_κ, it holds that

$$\kappa^{\beta-1}G_{p_\kappa,\kappa} \leqslant \frac{(\Gamma + 2)\kappa + 1}{\gamma\kappa}, \quad \kappa \geqslant 1.$$

Noting that the right hand side of above inequality equal to $\frac{\Gamma+2}{\gamma} + \frac{1}{\gamma\kappa}$ is a decreasing function for $\kappa \geqslant 1$, the above inequality implies

$$\kappa^{\beta-1}G_{p_\kappa,\kappa} \leqslant \frac{\Gamma + 3}{\gamma}, \quad \kappa \geqslant 1.$$

This is,

$$G_{p_\kappa,\kappa} \leqslant \frac{\Gamma + 3}{\gamma\kappa^{\beta-1}}, \quad \kappa \geqslant 1. \tag{61}$$

Since $\gamma \geqslant \Gamma + 3$, the above equation implies that $p_\kappa \geqslant (\Gamma + 2)\kappa + 1$. Hence, by Proposition 4.1 with $p := p_\kappa$, we obtain that

$$\|\mathscr{K}\chi_\kappa - \mathscr{K}_{p_\kappa}\chi_\kappa\|_\infty \leqslant \frac{44r\tau}{5p_\kappa} + \frac{27r\tau}{5}G_{p_\kappa,\kappa}^m.$$

Substituting $p_\kappa \geqslant \gamma\kappa^\beta$ and inequality (61) into the two terms in the right hand side of above inequality respectively yields the desired result directly. □

As for $y \in H^m_{\kappa,0}(I)$ as shown in (11), we have the next proposition.

Proposition 4.3. *Suppose that the function y has the form (11) and $\Gamma \geqslant 0$ is a given relaxation factor. If the parameters β and γ are chosen to satisfy $\beta \geqslant 1$, $\gamma \geqslant \Gamma + 3$ and $p_\kappa := \lceil \gamma\kappa^\beta \rceil$, then for any $\kappa \geqslant 1$,*

$$\|\mathscr{K} y - \mathscr{K}_{p_\kappa} y\|_\infty \leqslant \frac{132\tau}{5\gamma\kappa^\beta} + \frac{81\tau(\Gamma+3)^m}{5\gamma^m \kappa^{m(\beta-1)}}. \tag{62}$$

Proof. The function y having the form (11) can be expressed in the form χ_κ by letting $r := 3$, $\alpha_1 := 0$, $\alpha_2 := -1$, $\alpha_3 := 1$, and $w_j := u_j$ for $j \in \mathbb{N}_3$. Then, (62) can be obtained by setting $\chi_\kappa = y$ in Proposition 4.2. □

For the purpose of approximating the operator $\mathscr{K}$, by Proposition 4.3, we propose the discrete oscillatory integral operator $\mathscr{K}_{p_\kappa}$ by taking

$$p_\kappa := \lceil \gamma\kappa^\beta \rceil, \tag{63}$$

where the parameters β and γ are chosen to satisfy

$$\beta \geqslant 1, \quad \gamma \geqslant \Gamma + 3. \tag{64}$$

Up to now, with the discrete oscillatory integral operator $\mathscr{K}_{p_\kappa}$, the DNN learning model is well described.

In the next section, we will estimate the error of the solution obtained from the DNN model (8). To distinguish it from a MGDL model to be described in section 6, we will refer to the standard DNN model (8) as the single-grade learning model.

5. Analysis of single-grade learning model

The purpose of this section is to bound the error of the DNN approximate solution by the training loss and the quadrature error.

Throughout this section, we assume that the solution y has the form (11), the parameters β and γ are chosen according to rule (64), and p_κ is defined as in (63). The operator $\mathscr{K}_{p_\kappa}$ is defined by setting $p := p_\kappa$ as described in (14).

We first derive an equivalent form of the fully discrete minimization problem (8). To this end, for each $\kappa \geqslant 1$ and a fixed $q \in \mathbb{N}$, we let $N_\kappa := qp_\kappa + 1$ and

$$\omega_\kappa := e^{\frac{2\kappa}{qp_\kappa} i}. \tag{65}$$

We then define the matrix

$$\boldsymbol{B}_\kappa := (b_{j,l}(\kappa))_{j,l\in\mathbb{N}_{N_\kappa}} \in \mathbb{C}^{N_\kappa \times N_\kappa} \tag{66}$$

with

$$b_{j,l}(\kappa) := \begin{cases} \omega_\kappa^{|j-l|}, & l \in \{1, N_\kappa\}, \\ 2\omega_\kappa^{|j-l|}, & l \in \{dq+1 : d \in \mathbb{N}_{p_\kappa - 1}\}, \\ 0, & \text{otherwise}, \end{cases} \quad \text{for} \quad j \in \mathbb{N}_{N_\kappa}.$$

Since $p_\kappa = \lceil \gamma \kappa^\beta \rceil$ where γ, β is chosen by the rule (64), the matrix $\boldsymbol{B}_\kappa$ is completely determined by four independent parameters κ, β, γ, q. However in this paper, we only concern the influence of wavenumber κ. Thus, notation $\boldsymbol{B}_\kappa$ indicates only its dependence on κ even though it depends on the other three parameters β, γ, q, and so do p_κ and N_κ.

We assume that the training data $\{(x_j, f(x_j)),\ j \in \mathbb{N}_{N_\kappa}\}$ are chosen as $x_j := -1 + 2(j-1)/(N_\kappa - 1)$, $j \in \mathbb{N}_{N_\kappa}$. Thus, for each $h \in C(I)$, the $[((\mathscr{I} - \lambda \mathscr{K}_{p_\kappa})h)(x_j) : j = 1, 2, \ldots, N_\kappa]^T$ has an equivalent matrix form. We show it in the next lemma.

Lemma 5.1. *For all $\kappa \geqslant 1$, $h \in C(I)$, there holds*

$$[((\mathscr{I} - \lambda \mathscr{K}_{p_\kappa})h)(x_j) : j = 1, 2, \ldots, N_\kappa]^T = \left(\boldsymbol{I}_\kappa - \frac{\lambda}{p_\kappa}\boldsymbol{B}_\kappa\right)\boldsymbol{v}_h, \tag{67}$$

where $\boldsymbol{v}_h := [h(x_j) : j \in \mathbb{N}_{N_\kappa}]$ and $\boldsymbol{I}_\kappa$ denotes the $N_\kappa \times N_\kappa$ identity matrix.

Proof. For an $h \in C(I)$, by the definition of operator $\mathscr{I} - \lambda \mathscr{K}_{p_\kappa}$, for each $j \in \mathbb{N}_{N_\kappa}$, $((\mathscr{I} - \lambda \mathscr{K}_{p_\kappa})h)(x_j)$ can be written as

$$H_j := h(x_j) - \frac{\lambda}{p_\kappa}\left(h(x_1)e^{i\kappa|x_1 - x_j|} + 2\sum_{l=1}^{p_\kappa - 1} h(x_{ql+1})e^{i\kappa|x_j - x_{ql+1}|} + h(x_{N_\kappa})e^{i\kappa|x_j - x_{N_\kappa}|}\right).$$

Using the definition (65) of ω_κ, we may further rewrite H_j as

$$H_j = h(x_j) - \frac{\lambda}{p_\kappa}\left(\omega_\kappa^{|1-j|}h(x_1) + 2\sum_{l=1}^{p_\kappa - 1} \omega_\kappa^{|j-ql-1|}h(x_{ql+1}) + \omega_\kappa^{|j-N_\kappa|}h(x_{N_\kappa})\right), \quad j \in \mathbb{N}_{N_\kappa}.$$

The right-hand side of this equation is seen to equal the j-th component of the vector $\left(\boldsymbol{I}_\kappa - \frac{\lambda}{p_\kappa}\boldsymbol{B}_\kappa\right)\boldsymbol{v}_h$. This proves the desired result. □

Lemma 5.1 relates the operator $\mathscr{I} - \lambda \mathscr{K}_{p_\kappa}$ with the matrix

$$\boldsymbol{M}_\kappa := \boldsymbol{I}_\kappa - \lambda \boldsymbol{B}_\kappa / p_\kappa \in \mathbb{C}^{N_\kappa \times N_\kappa}. \tag{68}$$

In (67), by choosing $h := \mathscr{T}\mathscr{N}_n\left(\{\boldsymbol{W}_l, \boldsymbol{b}_l\}_{l=1}^n; \cdot\right)$, we obtain that

$$[((\mathscr{I} - \lambda \mathscr{K}_{p_\kappa})\mathscr{T}\mathscr{N}_n\left(\{\boldsymbol{W}_l, \boldsymbol{b}_l\}_{l=1}^n; \cdot\right))(x_j) : j = 1, 2, \ldots, N_\kappa]^T = \boldsymbol{M}_\kappa \boldsymbol{v}_g\left(\{\boldsymbol{W}_l, \boldsymbol{b}_l\}_{l=1}^n\right),$$

where

$$\boldsymbol{v}_g\left(\{\boldsymbol{W}_l, \boldsymbol{b}_l\}_{l=1}^n\right) := [\mathscr{T}\mathscr{N}_n\left(\{\boldsymbol{W}_l, \boldsymbol{b}_l\}_{l=1}^n; x_l\right), l \in \mathbb{N}_{N_\kappa}] \in \mathbb{C}^{N_\kappa}.$$

Moreover, by definition (7) of $\tilde{e}$ and by letting $\boldsymbol{v}_f := [f(x_l) : l \in \mathbb{N}_{N_\kappa}] \in \mathbb{C}^{N_\kappa}$, it holds that

$$[\tilde{e}(x_j) : j = 1, 2, \ldots, N_\kappa]^T = \boldsymbol{v}_f - \boldsymbol{M}_\kappa \boldsymbol{v}_g\left(\{\boldsymbol{W}_l, \boldsymbol{b}_l\}_{l=1}^n\right).$$

Thus, the minimization problem (8) is equivalent to the following minimization problem

$$\arg\min_{\{\boldsymbol{W}_l, \boldsymbol{b}_l\}_{l=1}^n} \frac{1}{N_\kappa}\left\|\boldsymbol{M}_\kappa \boldsymbol{v}_g\left(\{\boldsymbol{W}_l, \boldsymbol{b}_l\}_{l=1}^n\right) - \boldsymbol{v}_f\right\|_{l^2}^2. \tag{69}$$

The optimization problem (69) may be solved by the Adam (adaptive moment estimation) optimization algorithm [13] with the initialization proposed in [11]. Upon finding the parameters $\{\tilde{\boldsymbol{W}}_j^*, \tilde{\boldsymbol{b}}_j^*\}_{j=1}^n$, the numerical solution $\tilde{y}^*$ of (2) and the related error $\tilde{e}^*$ are given by (9) and (10), respectively.

In the rest of this section, we will estimate the error $\|y-\tilde{y}^*\|_{N_\kappa}$ in terms of the loss error $\|\tilde{e}^*\|_{N_\kappa}$ and the quadrature error $\|(\mathscr{K}-\mathscr{K}_{p_\kappa})y\|_\infty$. We first estimate the error $\|(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})(y-\tilde{y}^*)\|_{N_\kappa}$.

Lemma 5.2. *For each $\kappa \geqslant 1$,*

$$\|(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})(y-\tilde{y}^*)\|_{N_\kappa} \leqslant \|\tilde{e}^*\|_{N_\kappa} + |\lambda|\|(\mathscr{K}-\mathscr{K}_{p_\kappa})y\|_\infty. \tag{70}$$

Proof. By the definition (10) of $\tilde{e}^*$, we have $\tilde{e}^* = f-(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})\tilde{y}^*$, which together with $f=(\mathscr{I}-\lambda\mathscr{K})y$ yields that

$$\tilde{e}^* = (\mathscr{I}-\lambda\mathscr{K})y-(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})\tilde{y}^*.$$

Thus, we obtain

$$(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})(y-\tilde{y}^*) = (\mathscr{I}-\lambda\mathscr{K})y-(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})\tilde{y}^*+\lambda(\mathscr{K}-\mathscr{K}_{p_\kappa})y = \tilde{e}^*+\lambda(\mathscr{K}-\mathscr{K}_{p_\kappa})y.$$

By the triangle inequality of the seminorm $\|\cdot\|_{N_\kappa}$, we find that

$$\|(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})(y-\tilde{y}^*)\|_{N_\kappa} \leqslant \|\tilde{e}^*\|_{N_\kappa} + |\lambda|\|(\mathscr{K}-\mathscr{K}_{p_\kappa})y\|_{N_\kappa}. \tag{71}$$

The definition of $\|\cdot\|_{N_\kappa}$ yields that

$$\|(\mathscr{K}-\mathscr{K}_{p_\kappa})y\|_{N_\kappa} = \sqrt{\frac{1}{N_\kappa}\sum_{j=1}^{N_\kappa}|(\mathscr{K}y-\mathscr{K}_{p_\kappa}y)(x_j)|^2} \leqslant \|(\mathscr{K}-\mathscr{K}_{p_\kappa})y\|_\infty.$$

Substituting this inequality into the right-hand side of (71), we obtain the desired result (70). □

Next, we provide a condition that ensures the invertibility of the matrix $\boldsymbol{M}_\kappa$.

Lemma 5.3. *If there exists a constant $C_\kappa > 0$ such that for any $h \in C(I)$,*

$$\|h\|_{N_\kappa} \leqslant C_\kappa\|(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})h\|_{N_\kappa}, \tag{72}$$

then the matrix $\boldsymbol{M}_\kappa$ is invertible.

Proof. We establish the invertibility of the matrix $\boldsymbol{M}_\kappa$ by contradiction. Assume, to the contrary, that $\boldsymbol{M}_\kappa$ is not invertible. Then, there exists a nonzero vector $\boldsymbol{v} := [v_j,\, j\in\mathbb{N}_\kappa]^T \in \mathbb{C}^{N_\kappa}$ such that $\boldsymbol{M}_\kappa\boldsymbol{v}=\boldsymbol{0}$. By the Lagrange interpolation formula, we can construct a polynomial $h_{\boldsymbol{v}} \in C(I)$ such that $h_{\boldsymbol{v}}(x_j)=v_j$, $j\in\mathbb{N}_{N_\kappa}$. By letting $h := h_{\boldsymbol{v}}$ in inequality (72), we obtain that

$$\|h_{\boldsymbol{v}}\|_{N_\kappa} \leqslant C_\kappa\|(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})h_{\boldsymbol{v}}\|_{N_\kappa}. \tag{73}$$

Noting that

$$\|h_{\boldsymbol{v}}\|_{N_\kappa} = \|\boldsymbol{v}\|_{l_2}/\sqrt{N_\kappa} \quad\text{and}\quad \|(\mathscr{I}-\lambda\mathscr{K}_{p_\kappa})h_{\boldsymbol{v}}\|_{N_\kappa} = \|\boldsymbol{M}_\kappa\boldsymbol{v}\|_{l_2}\big/\sqrt{N_\kappa},$$

by (6) and (68) respectively, the inequality (73) implies that

$$\|\boldsymbol{v}\|_{l_2} \leqslant C_\kappa\,\|\boldsymbol{M}_\kappa\boldsymbol{v}\|_{l_2}\,. \tag{74}$$

However, note that the vector $\boldsymbol{v}$ is nonzero and $\boldsymbol{M}_\kappa\boldsymbol{v}=\boldsymbol{0}$, which contradicts inequality (74). This concludes that the matrix $\boldsymbol{M}_\kappa$ is invertible. □

Our next step is to identify the range of κ that ensures the invertibility of matrix $\boldsymbol{M}_\kappa$. Note that matrix $\boldsymbol{M}_\kappa$ is related to matrix $\boldsymbol{B}_\kappa$ by relation (68) and many columns of the matrix $\boldsymbol{B}_\kappa$ defined in (66) are zero. We denote the set of the indices of the zero columns of $\boldsymbol{B}_\kappa$ by $J := \mathbb{N}_{N_\kappa} \setminus \{ql+1 : l \in \mathbb{Z}_{p_{\kappa+1}}\}$. Hence, by relation (68), for any $j \in J$, the entries in the j-th column of matrix $\boldsymbol{M}_\kappa$ are all zeros but the j-th entry, which is 1. Combining this observation with the fact that a square matrix is invertible if and only if its determinant is non-zero, we will expand the determinant of matrix $\boldsymbol{M}_\kappa$ corresponding to the indices in set J, thereby obtaining the range of κ.

Specifically, assuming that j_1 is the largest index in set J, we expand the determinant $\det(\boldsymbol{M}_\kappa)$ by its j_1-th column and obtain that $\det(\boldsymbol{M}_\kappa) = \det(\boldsymbol{M}_{\kappa,1})$, where the matrix $\boldsymbol{M}_{\kappa,1} \in \mathbb{C}^{(N_\kappa-1)\times(N_\kappa-1)}$ is the submatrix of $\boldsymbol{M}_\kappa$ obtained by removing its j_1-th row and j_1-th column. We denote by J_1 the subset of J by removing j_1 from it, that is, $J_1 := J \setminus \{j_1\}$, and observe that all entries in the j-th column of matrix $\boldsymbol{M}_{\kappa,1}$ are zeros, but one, the j-th entry, which is 1. We denote by $\boldsymbol{M}_{\kappa,2}$ the submatrix of $\boldsymbol{M}_{\kappa,1}$ obtained by removing its j_2-th row and j_2-th column, where j_2 denotes the largest index in set J_2. Clearly, $\det(\boldsymbol{M}_{\kappa,2}) = \det(\boldsymbol{M}_{\kappa,1})$. Since $\#J = N_\kappa - p_\kappa - 1$, where $\#J$ denotes the cardinality of set J, we can repeat this process $N_\kappa - p_\kappa - 1$ times until obtaining $\boldsymbol{M}_{\kappa,N_\kappa-p_\kappa-1} \in \mathbb{C}^{(p_\kappa+1)\times(p_\kappa+1)}$, yielding

$$\det(\boldsymbol{M}_\kappa) = \det(\boldsymbol{M}_{\kappa,j}), \quad j = 1, 2, \ldots, N_\kappa - p_\kappa - 1. \tag{75}$$

Noting that matrix $\boldsymbol{M}_{\kappa,N_\kappa-p_\kappa-1} \in \mathbb{C}^{(p_\kappa+1)\times(p_\kappa+1)}$ is obtained by removing all the rows and columns of the indices in the set J, we have

$$(\boldsymbol{M}_{\kappa,N_\kappa-p_\kappa-1})_{(j,l)} = \begin{cases} 1-\lambda/p_\kappa, & j = l \in \{1, p_\kappa+1\}, \\ -\lambda\omega_\kappa^{q|j-l|}/p_\kappa, & l \in \{1, p_\kappa+1\}, \quad j \in \mathbb{N}_{p_\kappa+1} \setminus \{l\}, \\ 1-2\lambda/p_\kappa, & j = l \in \{2, 3, \ldots, p_\kappa\}, \\ -2\lambda\omega_\kappa^{q|j-l|}/p_\kappa, & l \in \{2, 3, \ldots, p_\kappa\}, \quad j \in \mathbb{N}_{p_\kappa+1} \setminus \{l\}. \end{cases}$$

According to (75), it suffices to consider the range of κ for $\det(\boldsymbol{M}_{\kappa,N_\kappa-p_\kappa-1}) \neq 0$. To this end, we perform elementary transformations on matrix $\boldsymbol{M}_{\kappa,N_\kappa-p_\kappa-1}$ to simplify it. Precisely, we multiply the first and $(p_\kappa+1)$-th columns of matrix $\boldsymbol{M}_{\kappa,N_\kappa-p_\kappa-1}$ by $-p_\kappa$ and the second column to the p_κ-th column by $-p_\kappa/2$ to obtain a new matrix. To describe the resultant matrix, for $d \in \mathbb{N}$ with $d \geqslant 4$, we denote by $\mathbb{C}[x]^{d\times d}$ the set of $d \times d$ matrices whose entries are polynomials of the variable $x \in \mathbb{C}$, and for $\lambda \in \mathbb{C}$, we define the matrix

$$\boldsymbol{A}_d(x;\lambda) := (a_{j,k}(x;\lambda))_{j,k\in\mathbb{N}_d} \in \mathbb{C}[x]^{d\times d},$$

where

$$a_{j,l}(x;\lambda) := \begin{cases} \lambda - (d-1), & j = l \in \{1, d\}, \\ \lambda - \frac{1}{2}(d-1), & j = l \in \{2, 3, \ldots, d-1\}, \\ \lambda x^{|j-l|}, & \text{otherwise}. \end{cases}$$

It can be verified that the matrix that results from matrix $\boldsymbol{M}_{\kappa,N_\kappa-p_\kappa-1}$ by the elementary transformations described above is $\boldsymbol{A}_{p_\kappa+1}(\omega_\kappa^q;\lambda)$. Noticing that $\omega_\kappa^q = e^{\frac{2\kappa}{p_\kappa}i}$ by the definition (65) of ω_κ, we conclude that the matrix $\boldsymbol{M}_\kappa$ is invertible if and only if $\det\left(\boldsymbol{A}_{p_\kappa+1}\left(e^{\frac{2\kappa}{p_\kappa}i};\lambda\right)\right) \neq 0$. We now introduce the set

$$S(\lambda) := \left\{\kappa \geqslant 1 : \det\left(\boldsymbol{A}_{p_\kappa+1}\left(e^{\frac{2\kappa}{p_\kappa}i};\lambda\right)\right) \neq 0\right\} \tag{76}$$

and summarize the discussion above in the following lemma.

Lemma 5.4. *Matrix $\boldsymbol{M}_\kappa \in \mathbb{C}^{N_\kappa \times N_\kappa}$ is invertible if and only if $\kappa \in S(\lambda)$.*

Lemma 5.4 indicates that matrix $\boldsymbol{M}_\kappa^{-1}$ exists if and only if $\kappa \in S(\lambda)$. Hence, the norm $\|\boldsymbol{M}_\kappa^{-1}\|_2$ is defined only for $\kappa \in S(\lambda)$. Is this too restricted? The next proposition reveals that the complementary set $[1, +\infty) \setminus S(\lambda)$ is at most countable. This in turn implies the set $S(\lambda)$ is equal to $[1, +\infty)$ almost everywhere, and thus, it is not too restricted.

Proposition 5.1. *For any $\lambda \in \mathbb{C}$, the set $[1, +\infty) \setminus S(\lambda)$ is at most countable.*

Proof. We prove this proposition by decomposing the set $[1, +\infty) \setminus S(\lambda)$. By the definition (76) of $S(\lambda)$, we know that

$$[1, +\infty) \setminus S(\lambda) = \left\{\kappa \geqslant 1 : \det\left(\boldsymbol{A}_{p_\kappa+1}\left(e^{\frac{2\kappa}{p_\kappa}i}; \lambda\right)\right) = 0\right\}. \tag{77}$$

Equation (77) reveals that set $[1, +\infty) \setminus S(\lambda)$ is determined by the zeros of polynomial $\det(\boldsymbol{A}_{p_\kappa+1}(x; \lambda))$, $x \in \mathbb{C}$. We decompose set $[1, +\infty) \setminus S(\lambda)$ according to the order, related to $p_\kappa + 1$, of the polynomial. Since $p_\kappa + 1 \geqslant 4$ by the definition (63) of p_κ, for $d \geqslant 4$, we define

$$S_d(\lambda) := \{x \in \mathbb{C} : \det(\boldsymbol{A}_d(x; \lambda)) = 0\}. \tag{78}$$

Thus, we may decompose the set $[1, +\infty) \setminus S(\lambda)$ as

$$[1, +\infty) \setminus S(\lambda) = \bigcup_{p=3}^{\infty} \left\{\kappa \geqslant 1 : e^{\frac{2\kappa}{p}i} \in S_{p+1}(\lambda), p = p_\kappa\right\}. \tag{79}$$

Meanwhile, for any $p \in \{3, 4, \dots\}$, with a direct computation, $\kappa \geqslant 1$ satisfies $p = p_\kappa = \lceil \gamma \kappa^\beta \rceil$ if and only if

$$\kappa \in I_p := \left(((p-1)/\gamma)^{1/\beta}, (p/\gamma)^{1/\beta}\right] \cap [1, +\infty).$$

Hence, (79) may be written as $[1, +\infty) \setminus S(\lambda) = \bigcup_{p=3}^{\infty} \left\{\kappa \in I_p : e^{\frac{2\kappa}{p}i} \in S_{p+1}(\lambda)\right\}$, that is,

$$[1, +\infty) \setminus S(\lambda) = \bigcup_{p=3}^{\infty} \bigcup_{z \in S_{p+1}(\lambda)} \left\{\kappa \in I_p : e^{\frac{2\kappa}{p}i} = z\right\}. \tag{80}$$

According to decomposition (80), it suffices to prove that

$$\#(S_{p+1}(\lambda)) < +\infty, \quad p \in \{3, 4, \dots\} \tag{81}$$

and

$$\#\left(\left\{\kappa \in I_p : e^{\frac{2\kappa}{p}i} = z\right\}\right) < +\infty, \quad z \in S_{p+1}(\lambda) \quad p \in \{3, 4, \dots\}. \tag{82}$$

We first establish (81). For each $p \geqslant 3$, $\lambda \in \mathbb{C}$, by the definition (78) of set $S_{p+1}(\lambda)$, the cardinality of the set corresponds to the number of zeros of the polynomial $\det(\boldsymbol{A}_{p+1}(x; \lambda))$, $x \in \mathbb{C}$, and this is certainly a finite number. As for (82), for each $p \in \{3, 4, \dots\}$, we observe that $e^{\frac{2\kappa}{p}i}$, $\kappa \geqslant 1$, as a function of κ has a period $p\pi$. Therefore, for each $p \in \{3, 4, \dots\}$ and $z \in S_{p+1}(\lambda)$, the number of solutions of equation $e^{\frac{2i\kappa}{p}} = z$ in the domain I_p is no larger than $((p/\gamma)^{1/\beta} - ((p-1)/\gamma)^{1/\beta})/(p\pi)$, which is clearly finite. This ensures that inequality (82) holds. □

Proposition 5.1 indicates that restriction $\kappa \in S(\lambda)$ is reasonable. We now return to the error estimate.

Lemma 5.5. *If $\kappa \in S(\lambda)$, then for any $h \in C(I)$,*

$$\|h\|_{N_\kappa} \leqslant \|\boldsymbol{M}_\kappa^{-1}\|_2 \|(\mathscr{I} - \lambda \mathscr{K}_{p_\kappa})h\|_{N_\kappa}.$$

Proof. Since $\kappa \in S(\lambda)$, Lemma 5.4 ensures that the inverse $\boldsymbol{M}_\kappa^{-1}$ exists. Hence, by the property of the l_2 norm, we derive for any $\boldsymbol{u} \in \mathbb{C}^{N_\kappa}$ that

$$\|\boldsymbol{M}_\kappa^{-1}\boldsymbol{u}\|_{l_2} \leqslant \|\boldsymbol{M}_\kappa^{-1}\|_2 \|\boldsymbol{u}\|_{l_2}. \tag{83}$$

Now, for any $h \in C(I)$, we define $\boldsymbol{v}_h := [h(x_j),\, j \in \mathbb{N}_\kappa]^T \in \mathbb{C}^{N_\kappa}$. Applying inequality (83) to $\boldsymbol{u} := \boldsymbol{M}_\kappa \boldsymbol{v}_h$ yields that

$$\|\boldsymbol{v}_h\|_{l_2} \leqslant \|\boldsymbol{M}_\kappa^{-1}\|_2 \|\boldsymbol{M}_\kappa \boldsymbol{v}_h\|_{l_2}.$$

This together with the equations $\|h\|_{N_\kappa} = \|\boldsymbol{v}_h\|_{l_2}/\sqrt{N_\kappa}$ and $\|(\mathscr{I} - \lambda \mathscr{K}_{p_\kappa})h\|_{N_\kappa} = \|\boldsymbol{M}_\kappa \boldsymbol{v}_h\|_{l_2}/\sqrt{N_\kappa}$ obtained from relation (6) and (68), respectively, we conclude the desired estimate. □

We next integrate Lemmas 5.2 and 5.5 to yield an estimate of the error $\|y - \tilde{y}^*\|_{N_\kappa}$.

Theorem 5.6. *If $\kappa \in S(\lambda)$, then*

$$\|y - \tilde{y}^*\|_{N_\kappa} \leqslant \|\boldsymbol{M}_\kappa^{-1}\|_2 \left(\|\tilde{e}^*\|_{N_\kappa} + |\lambda| \|(\mathscr{K} - \mathscr{K}_{p_\kappa})y\|_\infty\right). \tag{84}$$

Proof. Applying Lemma 5.5 with $h := y - \tilde{y}^*$ yields

$$\|y - \tilde{y}^*\|_{N_\kappa} \leqslant \|\boldsymbol{M}_\kappa^{-1}\|_2 \|(\mathscr{I} - \lambda \mathscr{K}_{p_\kappa})(y - \tilde{y}^*)\|_{N_\kappa}.$$

This together with inequality (70) in Lemma 5.2 leads to the desired estimate of this lemma. □

The coefficient $\|\boldsymbol{M}_\kappa^{-1}\|_2$ of the estimate (84) depends on κ. Figure 1 illustrates the dependence of $\|\boldsymbol{M}_\kappa^{-1}\|_2$ on κ for $\lambda = 1, 2, \ldots, 10, 100, 1000$. The figure indicates that $\|\boldsymbol{M}_\kappa^{-1}\|_2$ is either bounded above by a constant when $\lambda = 1, 2$ or increasing "linearly" as κ increases when $\lambda = 3, 4, \ldots, 10, 100, 1000$. In all the linear cases, the largest $\|\boldsymbol{M}_\kappa^{-1}\|_2$ value is 180 for $\lambda = 4$ at $\kappa = 600$. It is desirable to have $\|\boldsymbol{M}_\kappa^{-1}\|_2$ independent of κ, which occurs for $\lambda = 1, 2, 3$ in Figure 1. For this purpose, we make the following hypothesis.

Hypothesis 5.1. *There exists a $\kappa_0 \geqslant 1$ and a constant $C > 0$ such that $\|\boldsymbol{M}_\kappa^{-1}\|_2 \leqslant C$ for all κ in $S(\lambda) \cap [\kappa_0, +\infty)$.*

Under Hypothesis 5.1, Theorem 5.6 can be strengthened.

Theorem 5.7. *Suppose the solution $y \in H^m_{\kappa,0}(I)$ of the oscillatory Fredholm integral equation (2) has the form (11) and Hypothesis 5.1 holds. If parameters are chosen to satisfy*

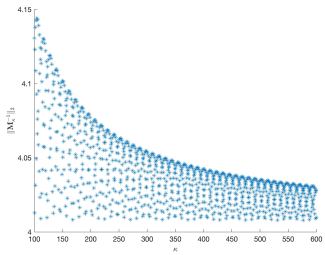
$$\Gamma \geqslant 0, \quad \beta \geqslant 1, \quad \gamma \geqslant \Gamma + 3, \quad q \in \mathbb{N},$$

and for any $\kappa \in S(\lambda) \cap [\kappa_0, \infty)$, let $p_\kappa := \lceil \gamma \kappa^\beta \rceil$, $N_\kappa := p_\kappa q + 1$, then

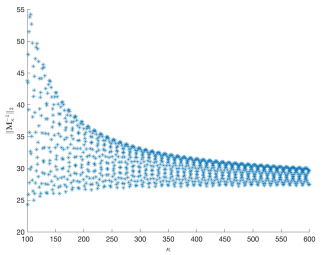
$$\|y - \tilde{y}^*\|_{N_\kappa} \leqslant C \left(\|\tilde{e}^*\|_{N_\kappa} + |\lambda| \|(\mathscr{K} - \mathscr{K}_{p_\kappa})y\|_\infty\right).$$

Proof. This theorem follows directly from Theorem 5.6 and Hypothesis 5.1. □

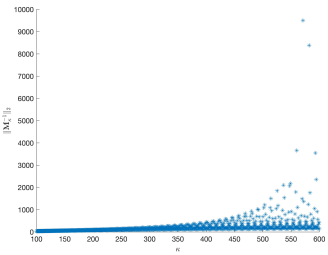
To close this section, we show, for λ satisfying $|\lambda| \in \left(0, \frac{1}{2}\right)$, how parameters β, γ, q may be chosen to ensure the validity of Hypothesis 5.1.



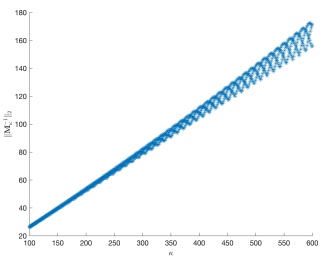
$(\lambda, \gamma, \beta, q) = (1, 6, 1, 1)$



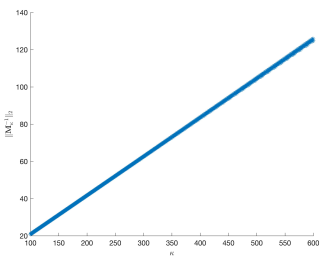
$(\lambda, \gamma, \beta, q) = (2, 6, 1, 1)$



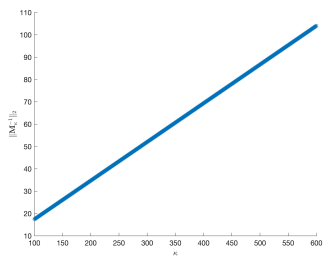
$(\lambda, \gamma, \beta, q) = (3, 6, 1, 1)$



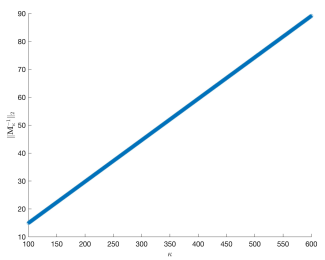
$(\lambda, \gamma, \beta, q) = (4, 6, 1, 1)$



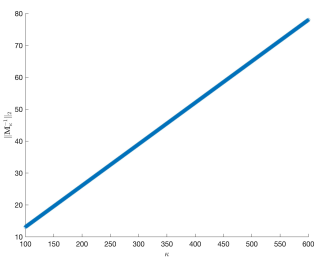
$(\lambda, \gamma, \beta, q) = (5, 6, 1, 1)$



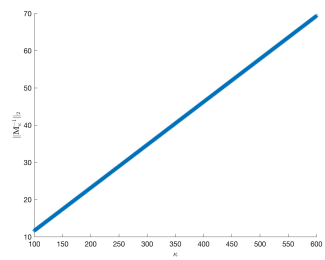
$(\lambda, \gamma, \beta, q) = (6, 6, 1, 1)$



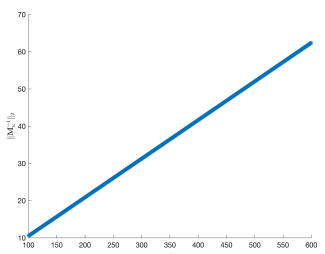
$(\lambda, \gamma, \beta, q) = (7, 6, 1, 1)$



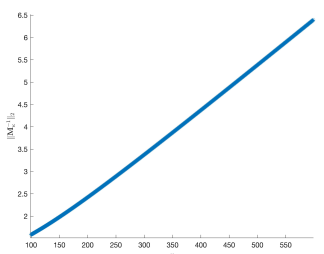
$(\lambda, \gamma, \beta, q) = (8, 6, 1, 1)$



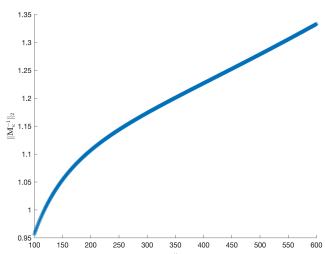
$(\lambda, \gamma, \beta, q) = (9, 6, 1, 1)$



$(\lambda, \gamma, \beta, q) = (10, 6, 1, 1)$



$(\lambda, \gamma, \beta, q) = (100, 6, 1, 1)$



$(\lambda, \gamma, \beta, q) = (1000, 6, 1, 1)$

Figure 1. Values of $\|\boldsymbol{M}_\kappa^{-1}\|_{N_\kappa}$ as a function of κ for different $(\lambda, \gamma, \beta, q)$.

Lemma 5.8. *For* λ *satisfying* $|\lambda| \in (0, \frac{1}{2})$, *if the parameters are chosen to satisfy*

$$\Gamma \geqslant 0, \quad \beta \geqslant 1, \quad \gamma > \max\left\{\Gamma + 3, \frac{4|\lambda|^2}{1-4|\lambda|^2}\right\}, \quad q \in \left[1, \frac{1}{4|\lambda|^2} - \frac{1}{\lceil \gamma \rceil}\right) \cap \mathbb{N}, \tag{85}$$

$p_\kappa := \lceil \gamma \kappa^\beta \rceil$ and $N_\kappa := p_\kappa q + 1$, then for any $\kappa \geqslant 1$, the matrix $\boldsymbol{M}_\kappa \in \mathbb{C}^{N_\kappa \times N_\kappa}$ defined in (68) is invertible and

$$\left\| \boldsymbol{M}_\kappa^{-1} \right\|_2 \leqslant \frac{1}{1-\eta}, \tag{86}$$

where

$$\eta := 2|\lambda| \sqrt{q + \frac{1}{\lceil \gamma \rceil}} \tag{87}$$

satisfies the condition $0 < \eta < 1$.

Proof. First, we show the existence of parameters Γ, β, γ, and q that satisfy condition (85). Clearly, we can find Γ, β, γ according to (85). By the choice of γ, we know that $\lceil \gamma \rceil > 4|\lambda|^2/(1-4|\lambda|^2)$, which implies that $1/(4|\lambda|^2) - 1/\lceil \gamma \rceil > 1$. Thus, we can find a q as in (85).

Next we estimate $\left\| \lambda \boldsymbol{B}_\kappa / p_\kappa \right\|_2$. Note that

$$\|\boldsymbol{B}_\kappa\|_2^2 = \rho(\boldsymbol{B}_\kappa^* \boldsymbol{B}_\kappa) \leqslant \|\boldsymbol{B}_\kappa^* \boldsymbol{B}_\kappa\|_1 \leqslant \|\boldsymbol{B}_\kappa^*\|_1 \|\boldsymbol{B}_\kappa\|_1 = \|\boldsymbol{B}_\kappa\|_\infty \|\boldsymbol{B}_\kappa\|_1,$$

where $\rho(\cdot)$ denotes the spectral radius of a matrix. This bound together with $\|\boldsymbol{B}_\kappa\|_1 = 2(qp_\kappa + 1)$ and $\|\boldsymbol{B}_\kappa\|_\infty = 2p_\kappa$ yields $\|\boldsymbol{B}_\kappa\|_2 \leqslant 2\sqrt{qp_\kappa^2 + p_\kappa}$, which implies

$$\left\| \frac{\lambda}{p_\kappa} \boldsymbol{B}_\kappa \right\|_2 \leqslant 2|\lambda| \sqrt{q + \frac{1}{p_\kappa}}. \tag{88}$$

Meanwhile, by the definition (63) of p_κ, we know that $p_\kappa = \lceil \gamma \kappa^\beta \rceil \geqslant \lceil \gamma \rceil$ for all $\kappa \geqslant 1$. Using it in inequality (88), we observe for all $\kappa \geqslant 1$ that

$$\left\| \frac{\lambda}{p_\kappa} \boldsymbol{B}_\kappa \right\|_2 \leqslant 2|\lambda| \sqrt{q + \frac{1}{\lceil \gamma \rceil}} = \eta, \tag{89}$$

whose right-hand side is independent of κ.

Furthermore, by the choice (85) of q, we notice that $q < 1/(4|\lambda|^2) - 1/\lceil \gamma \rceil$. Substituting it into the definition (87) of η yields that $0 < \eta < 1$. This estimation together with inequality (89), we conclude the matrix $\boldsymbol{M}_\kappa = \boldsymbol{I}_\kappa - \frac{\lambda}{p_\kappa} \boldsymbol{B}_\kappa$ is invertible and the inequality (86) holds. □

Lemma 5.8 conveys that for λ with $|\lambda| \in (0, \frac{1}{2})$, the parameters Γ, β, γ, and q can be chosen according to the rule (85) such that Hypothesis 5.1 holds with $\kappa_0 := 1$ and $C := \frac{1}{1-\eta}$, and in this case, there holds $S(\lambda) = [1, +\infty)$. The next theorem follows from Lemma 5.8 and Theorem 5.7.

Theorem 5.9. *Suppose λ satisfies $|\lambda| \in (0, \frac{1}{2})$ and the solution $y \in H^m_{\kappa,0}(I)$ of the oscillatory Fredholm integral equation (2) can be written as the form (11). If the parameters Γ, β, γ and q are chosen to satisfy (85), and $p_\kappa := \lceil \gamma \kappa^\beta \rceil$, $N_\kappa := p_\kappa q + 1$, then for all $\kappa \geqslant 1$,*

$$\|y - \tilde{y}^*\|_{N_\kappa} \leqslant \frac{1}{1-\eta} \left(\|\tilde{e}^*\|_{N_\kappa} + |\lambda| \|(\mathscr{K} - \mathscr{K}_{p_\kappa}) y\|_\infty \right),$$

where $\eta \in (0, 1)$, defined in (87), is independent of the wavenumber κ.

Combining Theorem 5.9 with Proposition 4.3 leads to the following corollary.

Corollary 5.1. *If the assumptions of Theorem 5.9 hold, then for all* $\kappa \geqslant 1$,

$$\|y - \tilde{y}^*\|_{N_\kappa} \leqslant \frac{1}{1-\eta}\left(\|\tilde{e}^*\|_{N_\kappa} + \frac{132\tau|\lambda|}{5\gamma\kappa^\beta} + \frac{81\tau|\lambda|(\Gamma+3)^m}{5\gamma^m\kappa^{m(\beta-1)}}\right),$$

where $\eta \in (0,1)$ *as defined in (87) is independent of the wavenumber* κ.

6. Multigrade learning model

The deep neural network model (8), which we will refer to as the single-grade learning model, has a computational issue: Solutions of the optimization problem (8) is often trapped in a local minimizer or even a saddle point due to too many layers used in DNNs. As a result, the loss error $\|\tilde{e}^*\|_{N_\kappa}$ may not be as small as we expect. In particular, in solving the oscillatory equation, as we will demonstrate in the next section, the single-grade learning model suffers from the spectral bias, that is, approximate solutions catch only low frequency components of the exact solution. To address this issue, following [25] we develop a multigrade learning model for numerical solutions of (1). The multigrade learning model introduced in [25] was motivated by the human education process which arranges learning in grades.

We now describe the multigrade learning model for numerical solutions of (1). Recalling the DNN that appears in minimization problem (8) has n layers, we choose L positive integers n_l, for $l = 1, 2, \ldots, L$, such that $n = \sum_{l=1}^{L} n_l$. Instead of solving one minimization problem (8) of n layers, we solve L intertwined minimization problems, which have n_l layers, for $l = 1, 2, \ldots, L$, respectively. For grade 1, we define the error function by

$$\tilde{e}_1\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^{n_1}; s\big) := \big(f - (\mathscr{I} - \lambda\mathscr{K}_{p_\kappa})\mathscr{T}\mathscr{N}_{n_1}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^{n_1}; \cdot\big)\big)(s) \in C(I), \quad s \in I$$

and find $\{\boldsymbol{W}_{1,j}^*, \boldsymbol{b}_{1,j}^*\}_{j=1}^{n_1}$ by solving the optimization problem

$$\min\Big\{\big\|\tilde{e}_1\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^{n_1}; \cdot\big)\big\|_{N_\kappa}^2 : \boldsymbol{W}_j \in \mathbb{R}^{m_{1,j}\times m_{1,j-1}}, \boldsymbol{b}_j \in \mathbb{R}^{m_{1,j}}, j \in \mathbb{N}_{n_1}\Big\},$$

with $m_{1,0} = 1$ and $m_{1,n_1} = 2$. Once the optimal parameters $\{\boldsymbol{W}_{1,j}^*, \boldsymbol{b}_{1,j}^*\}_{j=1}^{n_1}$ are learned, we obtain the feature of grade 1 as

$$\boldsymbol{g}_1(s) := \mathscr{F}_{n_1-1}\big(\{\boldsymbol{W}_{1,j}^*, \boldsymbol{b}_{1,j}^*\}_{j=1}^{n_1-1}; s\big), \quad s \in I,$$

where $\mathscr{F}_{n_1-1}$ is the feature of the DNN. Following the definition (3) of a standard DNN, we have

$$\mathscr{F}_{n-1}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^{n-1}; s\big) = \Big(\bigodot_{j=1}^{n-1} \boldsymbol{\sigma}(\boldsymbol{W}_j \cdot + \boldsymbol{b}_j)\Big)(s), \quad s \in I. \tag{90}$$

Thus, the approximate solution of grade 1 is

$$\boldsymbol{f}_1(s) := \boldsymbol{W}_{1,n_1}^* \boldsymbol{g}_1(s) + \boldsymbol{b}_{1,n_1}^*, \quad s \in I$$

with an error defined by

$$\tilde{e}_1^*(s) := \tilde{e}_1\big(\{\boldsymbol{W}_{1,j}^*, \boldsymbol{b}_{1,j}^*\}_{j=1}^{n_1}; s\big) \in C(I), \quad s \in I.$$

Suppose that the neural networks $\boldsymbol{g}_l : I \to \mathbb{R}^{m_{l,n_l-1}}$, $\boldsymbol{f}_l : I \to \mathbb{R}^2$, $\tilde{e}_l^* : I \to \mathbb{C}$ of grade $l < L$ have been learned and we will learn grade $l+1$. To this end, we define the error function of grade $l+1$ by

$$\tilde{e}_{l+1}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^{n_{l+1}}; s\big) := \tilde{e}_l^*(s) - \big[(\mathscr{I} - \lambda \mathscr{K}_{p_\kappa})\mathscr{T}\mathscr{N}_{n_{l+1}}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^{n_{l+1}}; \boldsymbol{g}_l(\cdot)\big)\big](s), \quad s \in I,$$

and find $\{\boldsymbol{W}_{l+1,j}^*, \boldsymbol{b}_{l+1,j}^*\}_{j=1}^{n_{l+1}}$ by solving the optimization problem

$$\min\Big\{\big\|\tilde{e}_{l+1}\big(\{\boldsymbol{W}_j, \boldsymbol{b}_j\}_{j=1}^{n_{l+1}}; \cdot\big)\big\|_{N_\kappa}^2 : \boldsymbol{W}_j \in \mathbb{R}^{m_{l+1,j} \times m_{l+1,j-1}}, \boldsymbol{b}_j \in \mathbb{R}^{m_{l+1,j}}, j \in \mathbb{N}_{n_{l+1}}\Big\},$$

with $m_{l+1,0} = m_{l,n_l-1}$ and $m_{l+1,n_{l+1}} = 2$. Note that when solving optimization problem 6, the parameters $\boldsymbol{W}_{\mu,j}^*$, $\boldsymbol{b}_{\mu,j}^*$, $j = 1, 2, \ldots, n_\mu$, $\mu = 1, 2, \ldots, l$, involved in $\boldsymbol{g}_l$ are fixed. Then we define the feature of grade $l+1$ by

$$\boldsymbol{g}_{l+1}(s) := \mathscr{F}_{n_{l+1}-1}\big(\{\boldsymbol{W}_{l+1,j}^*, \boldsymbol{b}_{l+1,j}^*\}_{j=1}^{n_{l+1}-1}; \boldsymbol{g}_l(s)\big), \quad s \in I, \tag{91}$$

and the solution component of grade $l+1$ by

$$\boldsymbol{f}_{l+1}(s) := \boldsymbol{W}_{l+1,n_{l+1}}^* \boldsymbol{g}_{l+1}(s) + \boldsymbol{b}_{l+1.n_{l+1}}^*, \quad s \in I.$$

Substituting (91) into the above equation, we see that $\boldsymbol{f}_{l+1}$ is actually the newly learned neural network stacked on the top of the feature layer learned in the previous grade. The optimal error of grade $l+1$ is defined by

$$\tilde{e}_{l+1}^*(s) := \tilde{e}_{l+1}\big(\{\boldsymbol{W}_{l+1,j}^*, \boldsymbol{b}_{l+1,j}^*\}_{j=1}^{n_{l+1}}; s\big) \in C(I), \quad s \in I.$$

We continue this process for $l < L$. The multigrade DNN approximation for the solution y is given by

$$\tilde{y}_L^* := \sum_{l=1}^{L} \mathscr{T}\boldsymbol{f}_l \in C(I).$$

We will show in the next section that the multigrade DNN solution $\tilde{y}_L^*$ is better than the single-grade DNN solution in catching the oscillation features of the exact solution of (1) and thus it has higher approximation accuracy.

7. Numerical experiments

This section is devoted to presentation of numerical experiments that assess and compare the performance of the proposed single-grade learning model and multigrade learning model with the traditional collocation method. Specifically, we focus on evaluating the methods' performance across varying wavenumbers and sample sizes. All the experiments presented in this section were performed on a Ubuntu Server 18.04 LTS 64bit equipped with Intel Xeon Platinum 8255C CPU @ 2.5GHz and NVIDIA Tesla T4 GPU.

In our experiments, we solved the oscillatory Fredholm integral equation (2) with $\lambda = 0.2$. For comparison purposes, we chose its exact solution as

$$y(s) := s + (3s^2 + 2s + 1)e^{i\kappa s} + (s+2)e^{-i\kappa s}, \quad s \in I,$$

which clearly satisfies the condition (11). The right-hand side f of the integral equation (2) is then calculated accordingly. We solved (2) with the right-hand side specified above using the single-grade,

multigrade models and the traditional collocation method, and compared the accuracy of these methods. The relative L_2 error defined by

$$\frac{\|y-\tilde{y}\|_2}{\|y\|_2} \approx \frac{1}{\|y\|_2}\left(\frac{2}{l}\left[|y(s_0)-\tilde{y}(s_0)|^2+2\sum_{j=1}^{l-1}|y(s_j)-\tilde{y}(s_j)|^2+|y(s_l)-\tilde{y}(s_l)|^2\right]\right)^{\frac{1}{2}},$$

was used to evaluate the accuracy of these methods, where y and $\tilde{y}$ denotes the exact solution and an approximate solution, respectively, $l := 20000$, $s_j := -1 + \frac{2j}{l}$, for $j \in \mathbb{Z}_{l+1}$, and $\|y\|_2$ was calculated analytically.

The wavenumber κ significantly influences the oscillatory behavior of the solution y and thus impacts the accuracy of its numerical approximations. To evaluate the numerical performance of our proposed model across varying oscillatory levels, we experimented with the κ values chosen from the set $\{100, 150, 200, 250, 300, 350, 400\}$. For the DNN method, we chose $\Gamma = 2$, $\gamma = 6$, $\beta = 1$, $q \in \{1, 2\}$ which satisfies the condition (85). Moreover, we set $p_\kappa := \lceil \gamma \kappa^\beta \rceil = \lceil 6\kappa \rceil$, and investigated the impact of the sample size on the approximation accuracy by considering $N_\kappa := 6\kappa + 1$ and $N_\kappa := 12\kappa + 1$, corresponding to the choices $q := 1$ and $q := 2$, respectively.

For the training model, we introduce a regularization [25] to the loss function of optimization problem (8) to address overfitting. Specifically, for the single-grade learning model, the training loss is defined as

$$training\_loss := \frac{1}{N_\kappa}\sum_{l=1}^{N_\kappa}\left|f(x_l)-\left((\mathscr{I}-\mathscr{K}_{p_\kappa})\mathscr{T}\mathscr{N}_n(\{\boldsymbol{W}_j,\boldsymbol{b}_j\}_{j=1}^n);\cdot\right)(x_l)\right|^2+\mu\sum_{j=1}^{n}\|\boldsymbol{W}_j\|_F^2,$$

where $\mu > 0$ is the regularization parameter, $\|\cdot\|_F$ is the Frobenius norm and x_l are training data points to be specified later. A sparse regularization was used in [28]. The training loss for the multigrade learning model is defined in a similar manner. All models of single-grade and multigrade were validated with the validation loss defined as

$$validation\_loss := \frac{1}{512}\sum_{l=1}^{512}\left|f(x_l')-((\mathscr{I}-\mathscr{K}_{p_\kappa})Y(x_l')\right|^2,$$

where $Y := \tilde{y}^*$ for the single-grade model and $Y := \tilde{y}_L^*$ for the multigrade model, and x_l' are validation data points to be specified.

Now, we specify the training and validation data.

Training data: For each chosen κ, we equidistantly choose N_κ points x_j from I and compute $f(x_j)$, for $j \in \mathbb{N}_{N_\kappa}$, where N_κ is either $6\kappa+1$ or $12\kappa+1$, and thus obtain the training data $\{(x_j, f(x_j))\}_{j=1}^{N_\kappa} \in I \times \mathbb{C}$.

Validation data: The validation set is given by $\{(x_j', f(x_j'))\}_{j=1}^{512} \in I \times \mathbb{C}$, where x_j', $j \in \mathbb{N}_{512}$, are uniformly distributed on the interval I. Note that N_κ is not equal to 512 for any chosen κ value.

In our experiments, the activation functions of hidden layers are all chosen to be the sin function for the single-grade and multigrade networks. We chose three single-grade network architectures SGL-1, SGL-2 and SGL-3 as shown in Table 1, where $[n]$ indicates a fully-connected layer with n neurons. Note that the network SGL-2 is an extension of the network SGL-1 with two additional hidden layers and SGL-3 is an extension of SGL-2 with four additional hidden layers.

We use the three grades for the multigrade learning model corresponding to SGL-1, SGL-2 and SGL-3, respectively. Their network architectures are described below:

Methods	Network Structure
SGL-1	$[1] \to [256] \to [256] \to [2]$
SGL-2	$[1] \to [256] \to [256] \to [128] \to [128] \to [2]$
SGL-3	$[1] \to [256] \to [256] \to [128] \to [128] \to [64] \to [64] \to [32] \to [32] \to [2]$

Table 1. Network structure of single-grade learning model

		$\kappa = 100$		$\kappa = 150$		$\kappa = 200$		$\kappa = 250$		$\kappa = 300$		$\kappa = 350$		$\kappa = 400$	
		bs	μ	bs	μ	bs	μ	bs	μ	bs	μ	bs	μ	bs	μ
$N = 6\kappa + 1$	SGL-1	64	0	128	10^{-6}	64	0	128	10^{-6}	64	10^{-5}	256	10^{-6}	256	10^{-6}
	SGL-2	128	10^{-5}	128	10^{-5}	128	10^{-6}	128	10^{-6}	128	10^{-4}	128	0	64	10^{-4}
	SGL-3	128	10^{-4}	64	10^{-4}	128	10^{-4}	128	10^{-4}	128	10^{-4}	128	10^{-4}	128	10^{-4}
	MGDL	64	10^{-6}	64	10^{-6}	128	10^{-6}	128	10^{-6}	128	10^{-6}	128	10^{-6}	128	10^{-6}
$N = 12\kappa + 1$	SGL-1	128	10^{-6}	128	0	256	10^{-6}	256	0	256	0	128	10^{-6}	256	10^{-6}
	SGL-2	128	0	256	10^{-6}	256	10^{-6}	256	10^{-6}	256	0	256	0	256	10^{-6}
	SGL-3	128	10^{-4}	256	10^{-4}	256	10^{-4}	256	10^{-4}	256	10^{-4}	256	10^{-4}	256	10^{-4}
	MGDL	64	0	64	0	128	0	128	0	128	0	128	0	128	0

Table 2. Batch size (bs) and regularization parameter (μ) for SGL-1, SGL-2, SGL-3 and MGDL.

Grade 1 : $[1] \to [256] \to [256] \to [2]$.
Grade 2 : $[1] \to [256]_F \to [256]_F \to [128] \to [128] \to [2]$.
Grade 3 : $[1] \to [256]_F \to [256]_F \to [128]_F \to [128]_F \to [64] \to [64] \to [32] \to [32] \to [2]$.

Here, $[n]_F$ indicates a layer having its parameters trained in the previous grades and remained fixed in training of the current grade.

We now describe the training and the tuning strategies for the single-grade models and the associated multigrade model. For the single-grade models, we used $3,500$ epochs for training. While for the multigrade model, we utilized 500, $1,000$ and $2,000$ epochs for grades 1, 2 and 3, respectively. For all training processes, we uniformly set the initial learning rate 10^{-2} and have it exponentially decay to the final learning rate 10^{-7}. We used regularization parameters 0, 10^{-6}, 10^{-5}, 10^{-4} and batch sizes $64, 128, 256$, and chose the best pair of the hyper-parameters in the sense that with it the model produces the minimum validation error over five independent experiments for each pair of the parameters. We list in Table 2 the best pair of the hyper-parameters found for all models.

Table 2 shows that MGDL incorporates implicit regularization. For the case $N := 6\kappa + 1$, the regularization parameters for MGDL are all 1e-6 very small, and for the case $N := 12\kappa + 1$, all regularization parameters for MGDL turn out to be zero, indicating that the MGDL model has a built-in regularization feature and additional regularization may not be needed. While the regularization parameters for the deepest model SGL-3 are all 1e-4, hinting that it requires regularization.

We compared performance of the DNN models to that of the traditional collocation method using continuous piecewise linear functions and continuous piecewise quadratic functions as bases. For details of the collocation method, the readers are referred to [1; 5]. We now describe the collocation method using the continuous piecewise polynomial of degree d as the basis for $d = 1, 2$. For each chosen κ, let N_κ be equal to $6\kappa + 1$ or $12\kappa + 1$. For each $l \in \mathbb{N}_{N_\kappa}$, the basis function $\phi_l \in C(I)$ is defined to be a

		$\kappa = 100$	$\kappa = 150$	$\kappa = 200$	$\kappa = 250$	$\kappa = 300$	$\kappa = 350$	$\kappa = 400$
$N_\kappa = 6\kappa+1$	CM1	$1.16\cdot10^{-2}$	$1.15\cdot10^{-2}$	$1.15\cdot10^{-2}$	$1.15\cdot10^{-2}$	$1.15\cdot10^{-2}$	$1.15\cdot10^{-2}$	$1.15\cdot10^{-2}$
	CM2	$1.67\cdot10^{-3}$	$1.67\cdot10^{-3}$	$1.67\cdot10^{-3}$	$1.66\cdot10^{-3}$	$1.66\cdot10^{-3}$	$1.66\cdot10^{-3}$	$1.66\cdot10^{-3}$
	SGL-01	$1.11\cdot10^{-2}$	$8.82\cdot10^{-1}$	$9.82\cdot10^{-1}$	$9.83\cdot10^{-1}$	$9.01\cdot10^{-1}$	$9.48\cdot10^{-1}$	$9.83\cdot10^{-1}$
	SGL-02	$1.49\cdot10^{-3}$	$1.29\cdot10^{-3}$	$1.67\cdot10^{-3}$	$1.94\cdot10^{-3}$	$3.75\cdot10^{-3}$	$3.36\cdot10^{-3}$	$5.54\cdot10^{-3}$
	SGL-03	$8.11\cdot10^{-3}$	$6.91\cdot10^{-3}$	$6.80\cdot10^{-3}$	$6.61\cdot10^{-3}$	$4.92\cdot10^{-3}$	$5.92\cdot10^{-3}$	$5.38\cdot10^{-3}$
	MGDL	$\mathbf{4.23\cdot10^{-4}}$	$\mathbf{3.09\cdot10^{-4}}$	$\mathbf{3.87\cdot10^{-4}}$	$\mathbf{3.20\cdot10^{-4}}$	$\mathbf{3.05\cdot10^{-4}}$	$\mathbf{3.53\cdot10^{-4}}$	$\mathbf{3.91\cdot10^{-4}}$
$N_\kappa = 12\kappa+1$	CM1	$2.94\cdot10^{-3}$	$2.91\cdot10^{-3}$	$2.89\cdot10^{-3}$	$2.89\cdot10^{-3}$	$2.89\cdot10^{-3}$	$2.89\cdot10^{-3}$	$2.90\cdot10^{-3}$
	CM2	$2.10\cdot10^{-4}$	$2.10\cdot10^{-4}$	$2.10\cdot10^{-4}$	$2.09\cdot10^{-4}$	$2.09\cdot10^{-4}$	$2.09\cdot10^{-4}$	$2.08\cdot10^{-4}$
	SGL-01	$1.22\cdot10^{-1}$	$3.17\cdot10^{-1}$	$9.81\cdot10^{-1}$	$9.82\cdot10^{-1}$	$9.82\cdot10^{-1}$	$9.83\cdot10^{-1}$	$9.84\cdot10^{-1}$
	SGL-02	$7.36\cdot10^{-4}$	$5.93\cdot10^{-4}$	$6.17\cdot10^{-4}$	$7.01\cdot10^{-4}$	$7.73\cdot10^{-4}$	$7.71\cdot10^{-4}$	$8.05\cdot10^{-4}$
	SGL-03	$3.91\cdot10^{-3}$	$3.43\cdot10^{-3}$	$3.36\cdot10^{-3}$	$3.57\cdot10^{-3}$	$3.43\cdot10^{-3}$	$3.71\cdot10^{-3}$	$4.20\cdot10^{-3}$
	MGDL	$\mathbf{1.30\cdot10^{-4}}$	$\mathbf{8.20\cdot10^{-5}}$	$\mathbf{7.81\cdot10^{-5}}$	$\mathbf{5.85\cdot10^{-5}}$	$\mathbf{4.32\cdot10^{-5}}$	$\mathbf{4.54\cdot10^{-5}}$	$\mathbf{4.54\cdot10^{-5}}$

Table 3. Relative error for CM1, CM2, SGL-1, SGL-2, SGL-2 and MGDL.

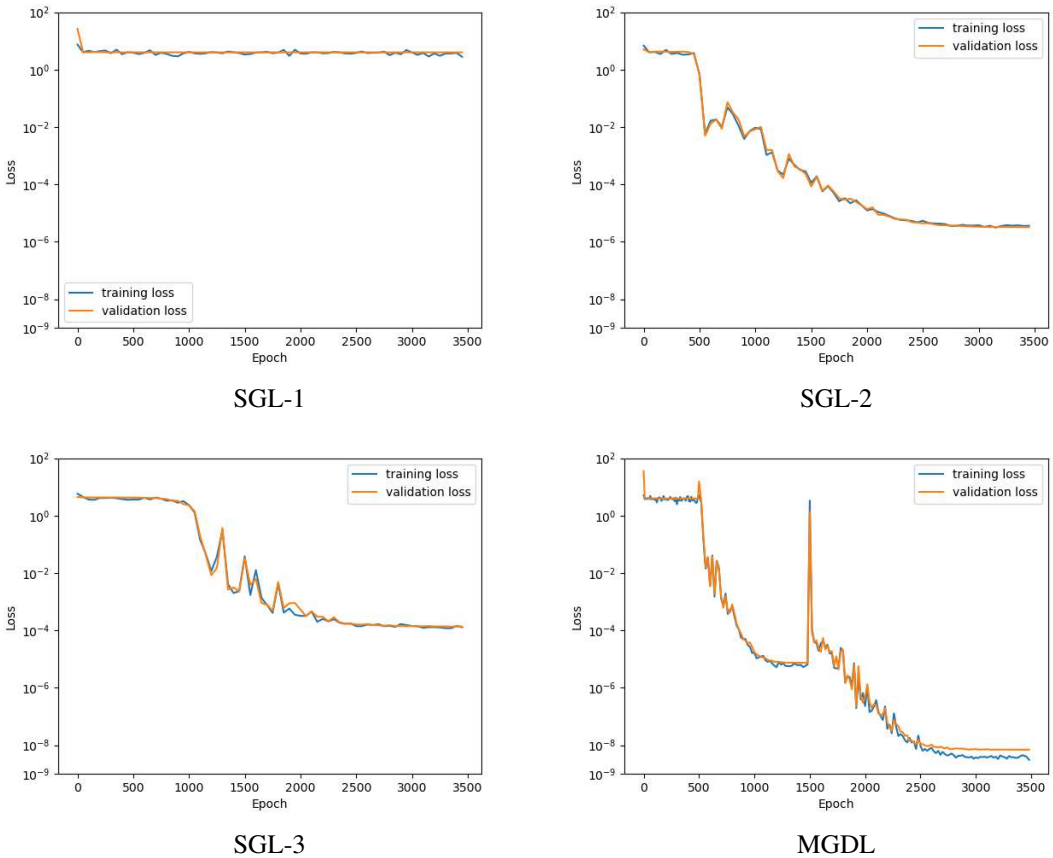
polynomial of degree d in the interval $[x_{jd+1}, x_{(j+1)d+1}]$ for each $j \in \mathbb{Z}_{(N_\kappa-1)/d}$ and satisfies $\phi_l(x_j) = \delta_{j,l}$ for any $j \in \mathbb{N}_{N_\kappa}$, where $\delta_{j,l} := 1$, for $j = l$, and 0 otherwise, for $j, l \in \mathbb{N}_{N_\kappa}$. The collocation method for solving the integral equation (2) with the described bases leads to the algebraic system

$$\boldsymbol{G}\boldsymbol{t} = \boldsymbol{f}, \text{ where } \boldsymbol{G} := [((\mathscr{I} - \lambda\mathscr{K})\phi_l)(x_j) : j, l \in \mathbb{N}_{N_\kappa}], \ \boldsymbol{t} := [t_j : j \in \mathbb{N}_{N_\kappa}]^\top, \ \boldsymbol{f} := [f(x_j) : j \in \mathbb{N}_{N_\kappa}]^\top.$$

By solving the above system, we obtain the coefficients $\boldsymbol{t}^* := [t_j^* : j \in \mathbb{N}_{N_\kappa}]^\top$, which gives rise to the collocation solution $\hat{y}_d := \sum_{j=1}^{N_\kappa} t_j^* \phi_j$. In our discussion to follow, we use CM1 and CM2 for the collocation method with piecewise polynomials of degree $d = 1$ and $d = 2$, respectively.

Relative errors of approximate solutions for all methods are summarized in Table 3 for wavenumbers $\kappa := 100, 150, 200, 250, 300, 350, 400$ and sample sizes $N_\kappa := 6\kappa + 1, 12\kappa + 1$. Comparing the three single-grade models SGL-1, SGL-2 and SGL-3, and the multigrade model MGDL with the two collocation methods, we find that MGDL significantly outperforms all other methods for all wavenumbers and all sample sizes. Among the three single-grade models, SGL-2 performs the best. Moreover, SGL-2 outperforms CM1 for all wavenumbers and all sample sizes and has slight larger errors than CM2 for most wavenumbers and sample sizes. However, model SGL-3, which is deeper than SGL-2, performs worse than SGL-2 for all cases, against the expectation that as the depth of the network increases, the expressive power of the neural network should improve. This may be due to the reason that as the neural network becomes deeper, the resulting optimization problem is more difficult to solve, leading to decline in the overall accuracy of the model. Moreover, as it will be shown later in this section, the single-grade deep learning model may suffer from the spectrum bias phenomenon when it is applied to solve an oscillatory integral equation. This serves as motivation for the development of multigrade learning model.

We further compare the performance of the single-grade learning model and multigrade learning model for the case when $\kappa = 350$ and $N_\kappa = 12\kappa + 1$. Specifically, we tabulated in Table 4 the training loss, the validation loss and the relative error for models SGL-1, SGL-2, SGL-3 and MGDL. We also plot in Figure 2 the training loss and the validation loss against the number of epochs. Furthermore, we display in Figure 3 the absolute error in the time domain between the exact solution and the approximate solutions



SGL-1 SGL-2

SGL-3 MGDL

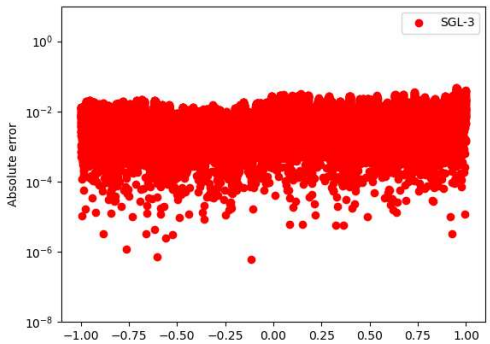
Figure 2. Training loss and validation loss for SGL-1, SGL-2, SGL-3 and MGDL for the case $\kappa = 350$, $N_\kappa = 12\kappa + 1$.

	SGL-1	SGL-2	SGL-3	Grade 1	Grade 2	Grade 3
Training loss	3.83	$2.51 \cdot 10^{-6}$	$7.19 \cdot 10^{-5}$	3.61	$6.31 \cdot 10^{-6}$	$3.34 \cdot 10^{-9}$
Validation loss	3.98	$2.89 \cdot 10^{-6}$	$6.83 \cdot 10^{-5}$	3.93	$8.87 \cdot 10^{-6}$	$6.88 \cdot 10^{-9}$
Relative error	$9.83 \cdot 10^{-1}$	$7.71 \cdot 10^{-4}$	$3.71 \cdot 10^{-3}$	$9.81 \cdot 10^{-1}$	$1.19 \cdot 10^{-3}$	$4.54 \cdot 10^{-5}$

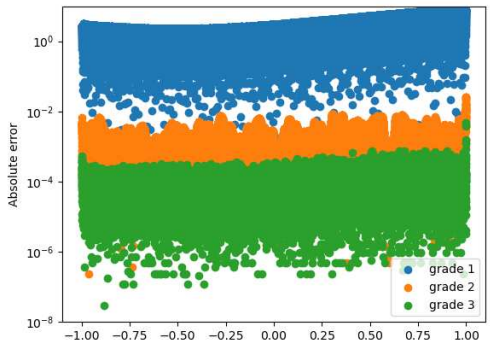
Table 4. Training loss, validation loss and relative error of the solution for SGL-1, SGL-2, SGL-3 and MGDL for the case $\kappa = 350$, $N_\kappa = 12\kappa + 1$.

generated by SGL-3 and grades of MGDL whose network architecture matches that of SGL-3. It is clear that MGDL generates a more accurate approximate solution than SGL-3.

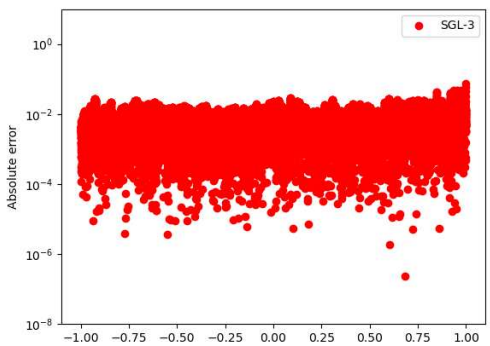
Figure 4 plots different solution components generated by different grades of the MGDL model. It illustrates that the MGDL model can effectively extract the intrinsic multiscale information hidden in the oscillatory solution of the integral equation. This well explains why the MGDL model can overcome the spectral bias from which single-grade learning models suffer.



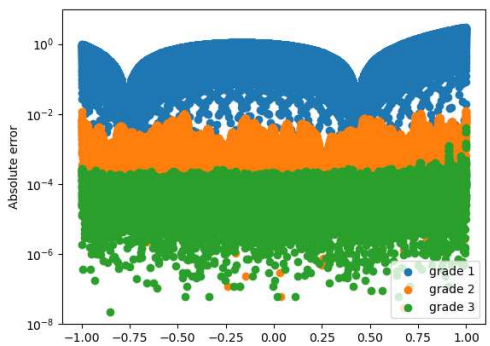
The real part for SGL-3



The real part for MGDL



The imaginary part for SGL-3



The imaginary part for MGDL

Figure 3. Absolute errors of the approximate solutions of SGL-3 and MGDL at $s_j := -1 + j/10000$, $j \in \mathbb{Z}_{20001}$ for the case $\kappa := 350$, $N_\kappa := 12\kappa + 1$.

To close this section, we show how the MGDL model improves approximation grade-by-grade looking from the frequency domain. To this end, we denote by $(\mathscr{F}v)(z)$ the fast Fourier transform of a vector v, where z denotes the frequency variable. We then use

$$\frac{|\mathscr{F}([y(s_j) : j \in \mathbb{N}_{20001}]^T)(z) - \mathscr{F}([Y(s_j) : j \in \mathbb{N}_{20001}]^T)(z)|}{|\mathscr{F}([y(s_j) : j \in \mathbb{N}_{20001}]^T)(z)|}$$

to compute the relative error of an approximate solution Y at the frequency $z \in \{0.5j - 5000.5 : j \in \mathbb{N}_{20001}\}$, where y denotes the exact solution and $s_j := -1 + \frac{2j-2}{20000}$, for $j \in \mathbb{N}_{20001}$. Here, $Y = \tilde{y}^*$ for the single-grade learning model and $Y = Y_l := \sum_{d=1}^{l} \mathscr{T}\boldsymbol{f}_d$ for $l \in \mathbb{N}_L$ for the l-th grade solution of the L-grade learning model. In Figure 5, we plot the relative errors between the frequency of the exact solution and that of approximate solutions for SGL-3 and MGDL. Figure 5 demonstrates that the single-grade learning model favors low frequency components. When the frequency greater than 100, the single-grade learning model gives large errors, which are larger than the errors for the grade 2 solution Y_2. On the contrary, errors of the MGDL model reduce grade-by-grade across all frequency levels and for frequencies higher than 100,

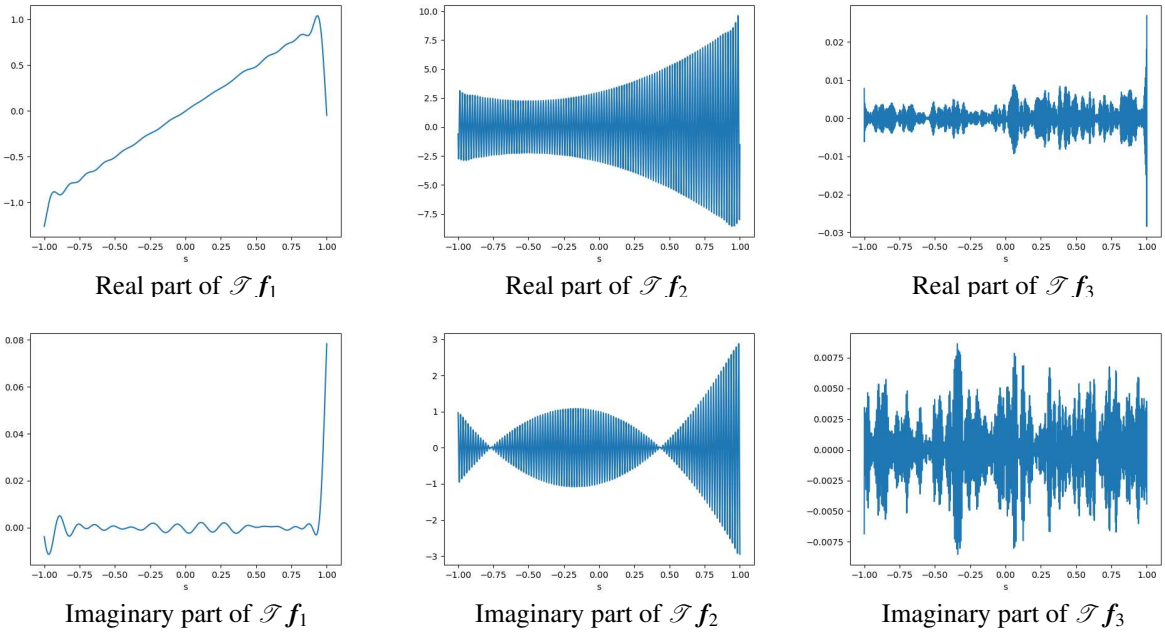


Figure 4. Grade components of MGDL for the case $\kappa = 350$, $N_\kappa = 12\kappa + 1$.

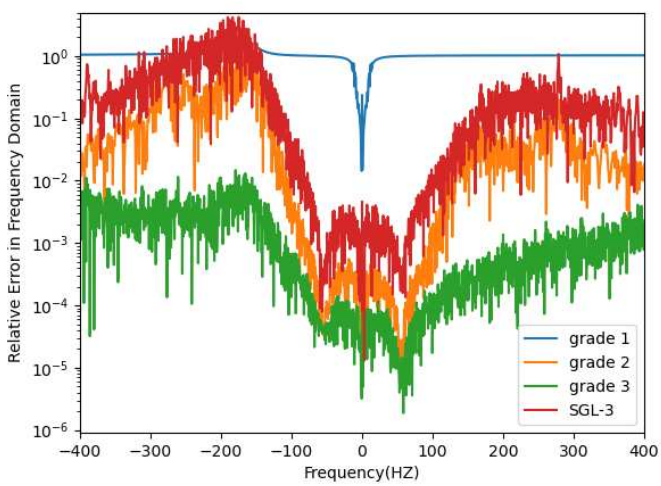


Figure 5. Relative errors of the approximate solutions generated by SGL-3 and grades 1, 2, 3 of MGDL in the frequency domain for the case $\kappa = 350$, $N_\kappa = 12\kappa + 1$.

the errors for the grade 3 solution Y_3 reduce remarkably. This pinpoints that the higher grade can capture the high-frequency component of the solution and thus, the MGDL model can overcome the spectral bias from which the single-grade learning suffers.

8. Concluding remarks

We developed a DNN method for the numerical solution of the oscillatory Fredholm integral equation of the second. The proposed methodology includes two major components: the numerical quadrature scheme that tailors to computing oscillatory integrals in the context of DNNs and the multigrade deep learning model that aims at overcoming the spectral bias issue of neural networks. We established the error of the single-grade neural network approximate solution of the equation by showing that it is bounded by the training loss plus the quadrature error. We demonstrated by numerical examples that the multigrade deep learning model is effective in extracting multiscale information of the oscillatory solution and overcoming the spectral bias issue from which the traditional single-grade learning model suffers.

Acknowledgement

Y. Xu is supported in part by the US National Science Foundation under grant DMS-2208386 and by the US National Institutes of Health under grant R21CA263876. Send all correspondence to Y. Xu.

References

[1] K. E. Atkinson, *The numerical solution of integral equations of the second kind*, Cambridge Monographs on Applied and Computational Mathematics **4**, Cambridge University Press, Cambridge, 1997.

[2] J. P. Boyd, "Asymptotic coefficients of Hermite function series", *J. Comput. Phys.* **54**:3 (1984), 382–410.

[3] H. Brunner, A. Iserles, and S. P. Nø rsett, "The spectral problem for a class of highly oscillatory Fredholm integral operators", *IMA J. Numer. Anal.* **30**:1 (2010), 108–130.

[4] H. Brunner, Y. Ma, and Y. Xu, "The oscillation of solutions of Volterra integral and integro-differential equations with highly oscillatory kernels", *J. Integral Equations Appl.* **27**:4 (2015), 455–487.

[5] Z. Chen, C. A. Micchelli, and Y. Xu, *Multiscale methods for Fredholm integral equations*, Cambridge Monographs on Applied and Computational Mathematics **28**, Cambridge University Press, Cambridge, 2015.

[6] D. Colton and R. Kress, *Inverse acoustic and electromagnetic scattering theory*, 2nd ed., Applied Mathematical Sciences **93**, Springer, 1998.

[7] P. J. Davies and D. B. Duncan, "Stability and convergence of collocation schemes for retarded potential integral equations", *SIAM J. Numer. Anal.* **42**:3 (2004), 1167–1188.

[8] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*, Dover Publications, Mineola, NY, 2007.

[9] W. E and B. Yu, "The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems", *Commun. Math. Stat.* **6**:1 (2018), 1–12.

[10] E. A. Flinn, "A modification of Filon's method of numerical integration", *J. Assoc. Comput. Mach.* **7** (1960), 181–184.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification", pp. 1026–1034 in *Proceedings of the IEEE International Conference on Computer Vision*, 2015.

[12] A. Iserles and S. P. Nø rsett, "Efficient quadrature of highly oscillatory integrals using derivatives", *Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **461**:2057 (2005), 1383–1399.

[13] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization", 2014. arXiv 1412.6980

[14] D. Levin, "Analysis of a collocation method for integrating rapidly oscillatory functions", *J. Comput. Appl. Math.* **78**:1 (1997), 131–138.

[15] Y. Ma and Y. Xu, "Computing highly oscillatory integrals", *Math. Comp.* **87**:309 (2018), 309–345.

[16] Y. Ma and Y. Xu, "Computing integrals involved the Gaussian function with a small standard deviation", *J. Sci. Comput.* **78**:3 (2019), 1744–1767.

[17] S. Olver, "GMRES for the differentiation operator", *SIAM J. Numer. Anal.* **47**:5 (2009), 3359–3373.

[18] S. Olver, "Fast, numerically stable computation of oscillatory integrals with stationary points", *BIT* **50**:1 (2010), 149–171.

[19] N. Rahaman, A. Baratin, D. Arpit, F. Draxler, M. Lin, F. A. Hamprecht, Y. Bengio, and A. Courville, "On the spectral bias of neural networks", in *Proceedings of the* 36 *th International Conference on Machine Learning*, *Long Beach*, *California*, *PMLR* 97, 2019.

[20] M. Raissi, "Deep hidden physics models: deep learning of nonlinear partial differential equations", *J. Mach. Learn. Res.* **19** (2018), 932–955.

[21] M. Raissi, P. Perdikaris, and G. E. Karniadakis, "Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations", *J. Comput. Phys.* **378** (2019), 686–707.

[22] H. Wang and S. Xiang, "Asymptotic expansion and Filon-type methods for a Volterra integral equation with a highly oscillatory kernel", *IMA J. Numer. Anal.* **31**:2 (2011), 469–490.

[23] Y. Wang and Y. Xu, "Oscillation Preserving Galerkin Methods for Fredholm Integral Equations of the Second Kind with Oscillatory Kernels", 2015. arXiv 1507.01156

[24] S. Xiang and Q. Zhang, "Asymptotics on the Fredholm integral equation with a highly oscillatory and weakly singular kernel", *Appl. Math. Comput.* **456** (2023), art. id. 128112.

[25] Y. Xu, "Multi-Grade Deep Learning", 2023. arXiv 2302.00150

[26] Y. Xu, "Successive affine learning for deep neural networks", 2023. arXiv 2305.07996

[27] Y. Xu and T. Zeng, "Multi-grade deep learning for partial differential equations with applications to the Burgers equation", 2023. arXiv 2309.07401

[28] Y. Xu and T. Zeng, "Sparse deep neural network for nonlinear partial differential equations", *Numer. Math. Theory Methods Appl.* **16**:1 (2023), 58–78.

[29] Y. Xu and H. Zhang, "Convergence of deep convolutional neural networks", *Neural Networks* **153** (2022), 553–563.

[30] Y. Xu and H. Zhang, "Convergence of deep ReLU networks", *Neurocomputing* **571** (2024), art. id. 127174.

JIE JIANG: jiangj73@mail2.sysu.edu.cn
School of Computer Science and Engineering, Sun Yat-sen University, Guangzhou, 510275, China

YUESHENG XU: y1xu@odu.edu
Department of Mathematics & Statistics, Old Dominion University, Norfolk, VA 23529, United States