无名稿件编号（由编辑填写）

自适应多级深度学习求解高度振荡的第二类弗雷德霍姆积分方程

Jie Jiang · Yuesheng Xu

arXiv:2601.04496v1 [math.NA] 8 Jan 2026

收稿日期：日期 / 接收日期：日期

摘要 本文研究了采用多级深度学习（MGDL）求解高度振荡的第二类弗雷德霍姆积分方程。我们对连续和离散 MGDL 模型进行了严格的误差分析，表明在求积误差足够小的情况下，离散模型能够保持其连续对应模型的收敛性和稳定性。我们确定深度神经网络训练误差是逼近误差的主要来源，由此提出一种新颖的自适应 MGDL 算法，根据训练性能选择网络等级。针对高度振荡情形（包括波数为500的情形）和奇异解开展的数值实验，验证了所提方法的精度、有效性和鲁棒性。

关键词 深度神经网络 · 振荡型弗雷德霍姆积分方程 ·自适应多级深度学习

数学主题分类（2020） 65R20

1 引言

本文研究了使用深度神经网络（DNNs）对第二类振荡Fredholm积分方程进行数值求解。这些方程广泛出现在各种物理和工程应用中，尤其是在波传播和散射问题中，例如电磁散射 [3,6]。尽管针对这类方程的经典数值方法已经

Jie Jiang 中山大学计算机科学与工程学院，中国广州 510275 电子邮件：jiangj73@mail2.sysu.edu.cn Yuesheng Xu 欧道明大学数学与统计系，美国弗吉尼亚州诺福克 23529 电子邮件：y1xu@odu.edu

确立 [2,3,5,14,15], ，但科学机器学习的最新进展推动了人们对探索基于深度神经网络的方法的兴趣。深度神经网络在求解偏微分方程（PDEs） [12,13], 、变分问题 [21], 以及近年来的积分方程 [9]方面展现出了良好的效果。

将深度神经网络应用于振荡积分方程的一大挑战，在于解具有高频特性，这是因为核 [4,14]中的振荡会传递到解中。标准深度神经网络存在众所周知的频谱偏差 [11],，这意味着它们倾向于先学习低频分量，却难以捕捉高频特征。因此，对于由高频模态主导的解，近似起来尤其困难；具有复指数核的振荡积分方程通常正是如此。

为克服这一局限， [16]提出了多级深度学习 （MGDL）框架。MGDL 不采用端到端方式训练深度网络，而是通过一系列浅层子网络逐级构建网络。每个层级都以前一层级的近似为基础，将其输出作为输入，并学习近似残差。较早的层级捕捉粗略的低频分量，而后续层级则接触更精细的高频残差——这实际上促使网络学习标准深度神经网络因频谱偏差而往往忽略的振荡特征。这种残差学习策略自然地将不同层级的频率成分解耦，使每个子网络能够专门处理不同的频谱带。因此，MGDL 能很好地适应振荡解的多尺度特性，并已在函数逼近 [17,16],偏微分方程 [18], 以及近期克服深度学习中的频谱偏差 [7]方面展现出有效性。

本文通过提出一种新颖的自适应算法，将 MGDL 框架扩展到振荡型弗雷德霍姆积分方程。该算法能够根据训练误差自动确定层级数量。尽管此前曾在 [9], 中以经验方式将 MGDL 应用于此类方程，但尚未建立理论基础。本文针对这一情形，对连续和离散 MGDL 模型进行了严格的数学分析。

标准深度学习方法的一大局限是需要在训练前指定网络深度。由于通常无法预知最优深度，性能不理想时往往需要调整深度，并从头开始重新训练网络。相比之下，MGDL 逐级增量构建网络，无需预先设定深度。这种增量构建方式自然引出了自适应多级深度学习（AMGDL）算法。这是一种此前工作中尚未研究的新方法，我们为其建立了严格的理论保证。

我们的主要贡献如下:

– 我们对连续和离散 MGDL 模型进行了严格的误差分析，证明当积分误差较小时，离散模型能够保持其连续对应模型的收敛性和稳定性。– 我们指出，深度神经网络训练误差是近似误差的主要来源，因此提出了一种根据训练表现进行自适应调整的方法。

– 我们提出了 AMGDL 算法，通过动态选择网络等级，在模型复杂度与解的精度之间取得平衡。
– 我们通过涉及高度振荡解和奇异解的数值实验，验证自适应策略的有效性。

本文其余部分安排如下。第 2 节介绍振荡Fredholm积分方程，并回顾基于深度神经网络的方法，包括连续和离散 MGDL 模型。第 3 节和第 4 节分别对连续和离散 MGDL 模型进行误差分析。第 5 节介绍 AMGDL 算法，并证明其理论保证。第 6 节通过数值实验展示 AMGDL 的精度和鲁棒性，包括在波数为 500 且存在奇异性的情况下，并验证本文确立的理论结果。第 7 节为结论。

2 深度学习模型

本节介绍所研究的振荡Fredholm积分方程，并描述为近似其解而构建的深度学习框架。

我们首先构造具有振荡核的弗雷德霍姆积分方程。设 $I := [-1, 1]$，并用 $C(I)$ 表示 I 上连续复值函数构成的空间，用 $C(I^2)$ 表示 I^2 上连续复值二元函数构成的空间。给定核函数 $K \in C(I^2)$ 和右端项 $f \in C(I)$，我们考虑以下振荡Fredholm积分方程

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

其中，$\kappa \geqslant 1$ 为波数，$y \in C(I)$为未知解。该核捕捉了波传播和散射问题中常见的振荡相互作用，例如声波、电磁波或量子波。指数项 $e^{i\kappa|s-t|}$ 表示以频率 κ 传播的一维波，其衰减或反射取决于区域的几何形状。我们尤其关注波数 $\kappa \geq 100$的情形。在此情况下，方程 (1) 被归类为高度振荡方程，这意味着解或核在整个区域内呈现快速振荡。这类问题对传统数值方法构成了重大挑战，因为为了保持精度和稳定性，通常需要极精细的离散化或专门的技术。

相应的积分算子 $\mathcal{K}$ 针对 $h \in C(I)$ 定义如下：

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

于是，弗雷德霍姆积分方程 (1) 可表示为如下算子形式

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

其中， $\mathcal{I}$ 表示 $C(I)$上的恒等算子。由于 $\mathcal{K}$ 是紧算子，当 1 不是 $\mathcal{K}$的特征值时，算子方程 (2) 在 $C(I)$中存在唯一解。已知 [14] 方程 (2) 的解通常表现出振荡行为，且振荡程度与波数成正比。本文旨在开发一种数值方法，通过利用多级深度学习框架，自适应地确定深度神经网络解的层数，从而有效处理此类高度振荡问题。

现在我们介绍用作积分方程近似解的深度神经网络。我们采用 [19,20] 中的记号。深度神经网络是由向量值函数复合而成的函数，每个向量值函数均通过对仿射变换应用激活函数来定义。具体而言，给定一个单变量激活函数 $\sigma:\mathbb{R}\to\mathbb{R}$，相应的向量值激活函数 $\boldsymbol{\sigma}:\mathbb{R}^d\to\mathbb{R}^d$ 定义为

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

对于向量值函数 f_j, $j = 1$，2，.....， n，若 f_j的值域包含于 f_{j+1}的区域内，则将它们依次复合记为

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

设 m_j, $j = 0$、1、.....、n为一个正整数序列，且 $m_0 := 1$， m_n:=2。网络的第一层和最后一层分别称为输入层和输出层，中间各层称为隐藏层。给定权重矩阵 $\mathbf{W}_j \in \mathbb{R}^{m_j \times m_{j-1}}$ 和偏置向量 $\mathbf{b}_j \in \mathbb{R}^{m_j}$,$j \in \mathbb{N}_n$，深度为 n 的深度神经网络定义为

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

其中， $\theta := \{\mathbf{W}_j,\ \mathbf{b}_j\}_{j=1}^n$ 表示所有可训练参数的集合。此外，最后一个隐藏层的输出称为深度神经网络的特征，其表达式为

$$\mathcal{F}_{n-1}(\{\mathbf{W}_j, \mathbf{b}_j\}_{j=1}^{n-1}; s) := \left(\bigodot_{j=1}^{n-1} \boldsymbol{\sigma}(\mathbf{W}_j \cdot + \mathbf{b}_j) \right)(s), \quad s \in I. \tag{4}$$

本文旨在开发一种自适应深度学习方法，利用形式为 (3) 的神经网络表示来逼近积分方程 (1) 的解。由于该解是复值的，我们引入算子 $\mathcal{T}$，将实值二维向量值函数映射为复值函数。具体而言，对于函数 $\mathbf{f}(s) = [f_1(s)\ f_2(s)]^T$，其满足 f_1， f_2 在 I 上为实值，我们定义

$$(\mathcal{T}\mathbf{f})(s) := f_1(s) + i f_2(s), \quad s \in I.$$

算子 $\mathcal{T}$ 将用于把实值深度神经网络输出映射为复函数。利用这一变换，我们将与积分方程 (2) 相关的残差（损失）函数定义为

$$e\left(\theta; s\right) := \left(\mathcal{I} - \mathcal{K}\right)\mathcal{T}\mathcal{N}_n(\theta; \cdot)(s) - f(s), \quad s \in I, \tag{5}$$

其中，$\theta := \{\mathbf{W}_j,\ \mathbf{b}_j\}_{j=1}^{n}$ 表示网络参数。最优参数 θ^* 通过求解以下优化问题得到:

$$\min_{\theta} \|e(\theta; \cdot)\|_2^2 . \tag{6}$$

该表述是深度神经网络训练问题的连续对应形式。由此得到的最小化问题通常是非凸的，且不保证存在全局极小值点。本文中，我们假设所有此类最小化问题都有解。

通过优化问题 (6) 求得 θ^* 后，函数

$$y^* := \mathcal{T}\mathcal{N}_n(\theta^*; \cdot)$$

可作为方程 (1) 解的深度神经网络近似。然而，由于深度神经网络的频谱偏差已得到充分证实，直接优化 (6) [11], 往往无法得到准确的近似，尤其是在解高度振荡时。为应对这一挑战，我们提出一种基于文献 [16] 中引入的MGDL框架的自适应学习策略。

我们首先介绍 MGDL 模型。固定一个正整数 $L \leq n$，并选取 L 个正整数 n_1、n_2、…、n_L，使得 $n - 1 = \sum_{l=1}^{L} n_l$。对于等级1，我们使用一个具有 n_1 层的神经网络。该层级的误差函数定义为

$$e_1\left(\{\mathbf{W}_j, \mathbf{b}_j\}_{j=1}^{n_1}; s\right) := \left(f - (\mathcal{I} - \mathcal{K})\mathcal{T}\mathcal{N}_{n_1}(\{\mathbf{W}_j, \mathbf{b}_j\}_{j=1}^{n_1}; \cdot)\right)(s), \quad s \in I. \tag{7}$$

最优参数 $\{\mathbf{W}_{1,\ j}^*,\ \mathbf{b}_{1,\ j}^*\}_{j=1}^{n_1}$ 通过求解优化问题获得

$$\min\left\{\left\|e_1\left(\{\mathbf{W}_j, \mathbf{b}_j\}_{j=1}^{n_1}; \cdot\right)\right\|_2^2 : \mathbf{W}_j \in \mathbb{R}^{m_{1,j} \times m_{1,j-1}}, \mathbf{b}_j \in \mathbb{R}^{m_{1,j}}, j \in \mathbb{N}_{n_1}\right\}, \tag{8}$$

其中 $m_{1,\ 0} = 1$ 和 $m_{1,\ n_1} = 2$。一旦获得参数 $\{\mathbf{W}_{1,\ j}^*,\ \mathbf{b}_{1,\ j}^*\}_{j=1}^{n_1}$，我们便将层级-1 特征函数定义为

$$\mathbf{g}_1(s) := \mathcal{F}_{n_1-1}\left(\{\mathbf{W}_{1,j}^*, \mathbf{b}_{1,j}^*\}_{j=1}^{n_1-1}; s\right), \quad s \in I,$$

其中 $\mathcal{F}_{n_1-1}$ 表示截至倒数第二层的特征映射。

层级-1 解分量为

$$\mathbf{f}_1(s) := \mathbf{W}_{1,2}^* \mathbf{g}_1(s) + \mathbf{b}_{1,2}^*, \quad s \in I,$$

且由此得到的最优误差函数为

$$e_1^*(s) := e_1\left(\{\mathbf{W}_{1,j}^*, \mathbf{b}_{1,j}^*\}_{j=1}^{n_1}; s\right) \in C(I), \quad s \in I. \tag{9}$$

假设已学习得到神经网络$\mathbf{g}_l:\ I\to\mathbb{R}^{m_l}$、$\mathbf{f}_l:\ I\to\mathbb{R}^2$，以及某一层级$l<L$对应的误差函数$e_l^*:\ I\to\mathbb{C}$。现在我们继续构建层级$l+1$。首先，将累积至层级$l$的近似解定义为

$$y_l^*:=\sum_{j=1}^{l}\mathcal{T}\mathbf{f}_j,\quad l\in\mathbb{N}_L. \tag{10}$$

层级$l+1$的误差函数定义为

$$e_{l+1}\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_{l+1}};s\right):=\left(f-(\mathcal{I}-\mathcal{K})(y_l^*+\mathcal{T}\mathcal{N}_2\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_{l+1}};\mathbf{g}_l(\cdot)\right)\right)(s),$$

对于$s\in I$。为了学习层级$l+1$的参数，我们求解以下优化问题，以获得极小化元$\{\mathbf{W}_{l+1,\ j}^*,\ \mathbf{b}_{l+1,\ j}^*\}_{j=1}^{n_{l+1}}$：

$$\min\left\{\left\|e_{l+1}\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_{l+1}};\cdot\right)\right\|_2^2:\right.$$
$$\left.\mathbf{W}_j\in\mathbb{R}^{m_{l+1,j}\times m_{l+1,j-1}},\mathbf{b}_j\in\mathbb{R}^{m_{l+1,j}},j\in\mathbb{N}_{n_{l+1}}\right\}, \tag{11}$$

其中$m_{l+1,\ 0}=m_{l,\ n_l-1}$和$m_{l+1,\ n_{l+1}}=2$。学习得到最优参数后，我们将层级$l+1$的特征表示定义为

$$\mathbf{g}_{l+1}(s):=\mathcal{F}_{n_{l+1}-1}\left(\left\{\mathbf{W}_{l+1,j}^*,\mathbf{b}_{l+1,j}^*\right\}_{j=1}^{n_{l+1}-1};\mathbf{g}_l(s)\right),\quad s\in I, \tag{12}$$

相应的解分量为

$$\mathbf{f}_{l+1}(s):=\mathbf{W}_{l+1,2}^*\mathbf{g}_{l+1}(s)+\mathbf{b}_{l+1.2}^*,\quad s\in I,$$

由此得到的最优误差为

$$e_{l+1}^*(s):=e_{l+1}\left(\left\{\mathbf{W}_{l+1,j}^*,\mathbf{b}_{l+1,j}^*\right\}_{j=1}^{n_{l+1}};s\right)\in C(I),\quad s\in I. \tag{13}$$

对每个层级$l<L$，都会重复这一过程。如果误差范数$\|e_l^*\|_2$低于预设容差，那么学习过程可以在某个层级$l<L$提前终止。下一节将对这一停止准则给出理论依据。

最后，解的多级深度神经网络近似定义为

$$y_L^*:=\sum_{l=1}^{L}\mathcal{T}\mathbf{f}_l.$$

在上述表述中，我们假设误差$\left\|e_l\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_l};\cdot\right)\right\|_2^2$是精确计算的，并将此框架称为连续多层级深度学习模型。在下一节中，我们将在连续 MGDL 框架下，建立误差$\|y-y_l^*\|_2$与训练误差e_l^*之间的上下界。

上述连续 MGDL 模型提供了理论基础，但无法直接应用于实践。在实际应用中，计算目标函数并训练模型需要进行数值积分，这会引入离散化。因此，自然地得到 MGDL 模型的离散版本。接下来，我们简要回顾先前工作中提出的离散 MGDL 框架 [9], ，该框架是实际实现的基础。

连续MGDL需要求解优化问题（8）和（11）。实现该方法需要计算

– 所涉及函数的 L_2-范数– 积分算子 $\mathcal{K}$。

为此，我们假设在 I 中选取了互不相同的点 $\{x_j\}_{j=1}^N$。函数 $g \in C（I）$ 的 L_2-范数 $\|g\|_2$ 可用离散半范数近似为

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

其中，$|\cdot|$表示复数的模。该半范数由 $\mathbb{C}^N$ 上的 ℓ_2-范数导出，但不是 $C（I）$ 上的真正范数，因为 $\|g\|_N = 0$ 并不一定意味着 $g = 0$。

对于任意 $g \in C（I）$，定义向量$\mathbf{v}_g := [g（x_j）]_{j=1}^N \in \mathbb{C}^N$。于是，离散半范数与标准 ℓ_2-范数之间的关系为

$$\|g\|_N = \frac{\|\mathbf{v}_g\|_{\ell_2}}{\sqrt{N}}. \tag{14}$$

我们假设已选定一个数值积分方案，其求积节点为 $\{s_j\}_{j=0}^{p_\kappa}$，用于构造离散算子 $\mathcal{K}_{p_\kappa}$，以近似连续积分算子 $\mathcal{K}$，其中 p_κ 是一个依赖于 κ的正整数。将方程 (2) 中的 $\mathcal{K}$ 替换为$\mathcal{K}_{p_\kappa}$，并在节点 $\{x_j\}_{j=1}^N$ 处对所得方程进行配置，便得到离散系统:

$$((\mathcal{I} - \mathcal{K}_{p_\kappa})y_h)(x_j) = f(x_j), \quad j = 1, 2, \ldots, N, \tag{15}$$

其中，y_h 是对真实解 y的近似。一般而言，求积节点与配置节点不必重合。然而，根据集体紧算子理论 [1], ，若 $\mathcal{K}_{p_\kappa}$ 逐点收敛于 $\mathcal{K}$，且 p_κ 足够大，那么只要求积节点与配置节点重合，对于任意给定的右端项 f，离散系统 (15) 都有唯一解。

利用离散 L_2 范数和离散积分算子 $\mathcal{K}_{p_\kappa}$，我们回顾在 [9] 中引入的单级学习模型。受离散系统 (15) 的启发，我们将相应的损失函数定义为

$$\tilde{e}\left(\{\mathbf{W}_j, \mathbf{b}_j\}_{j=1}^n; s\right) := \left(f - (\mathcal{I} - \mathcal{K}_{p_\kappa})\mathcal{T}\mathcal{N}_n(\{\mathbf{W}_j, \mathbf{b}_j\}_{j=1}^n; \cdot)\right)(s), \quad s \in I,$$

它是方程 (5) 中定义的连续损失函数的离散逼近。最优参数 $\{\mathbf{W}_j^*,\ \mathbf{b}_j^*\}_{j=1}^n$ 通过求解以下最小化问题得到

$$\min_{\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^n} \left\|\tilde{e}\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^n;\cdot\right)\right\|_N^2. \tag{16}$$

注意，在 (16) 中采用离散范数 $\|\cdot\|_N$ 会将优化问题限制在用于定义该范数的离散点上。因此，优化问题 (16) 是离散的。相应的数值解为

$$\tilde{y}^* := \mathcal{TN}_n(\{\tilde{\mathbf{W}}_j^*,\tilde{\mathbf{b}}_j^*\}_{j=1}^n;\cdot)$$

相应的误差函数定义为

$$\tilde{e}^*(s) := \tilde{e}\left(\{\tilde{\mathbf{W}}_j^*,\tilde{\mathbf{b}}_j^*\}_{j=1}^n;s\right),\quad s\in I.$$

现在，我们介绍用于数值求解方程 (1) 的离散 MGDL 模型。与连续 MGDL 模型一样，给定一个正整数序列 n_l，其中 $l=1,\ 2,\ \ldots,\ L$，满足 $n=\sum_{l=1}^L n_l$。基于此设定，该模型求解 L 个相互关联的最小化问题。

对于等级1，误差函数定义为

$$\tilde{e}_1\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_1};s\right) := \left(f-(\mathcal{I}-\mathcal{K}_{p_\kappa})\mathcal{TN}_{n_1}(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_1};\cdot)\right)(s)\in C(I), s\in I.$$

接着，我们求解以下优化问题:

$$\min\left\{\left\|\tilde{e}_1\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_1};\cdot\right)\right\|_N^2 : \mathbf{W}_j\in\mathbb{R}^{m_{1,j}\times m_{1,j-1}}, \mathbf{b}_j\in\mathbb{R}^{m_{1,j}}, j\in\mathbb{N}_{n_1}\right\}, \tag{17}$$

$\tilde{}$ $\tilde{}$以及 $m_{1,0}=1$ 和 $m_{1,n_1}=2$，由此得到最优参数 $\{\mathbf{W}_{1,j}^*,\mathbf{b}_{1,j}^*\}_{j=1}^{n_1}$。定义等级1的特征映射和解分量为

$$\tilde{\mathbf{g}}_1(s) := \mathcal{F}_{n_1-1}\left(\{\tilde{\mathbf{W}}_{1,j}^*,\tilde{\mathbf{b}}_{1,j}^*\}_{j=1}^{n_1-1};s\right),\quad s\in I,$$

and

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

分别为，其中 $\mathcal{F}_{n_1-1}$ 的定义同方程 (4)。相应的误差为

$$\tilde{e}_1^*(s) := \tilde{e}_1\left(\{\tilde{\mathbf{W}}_{1,j}^*,\tilde{\mathbf{b}}_{1,j}^*\}_{j=1}^{n_1};s\right)\in C(I),\quad s\in I. \tag{18}$$

假设已计算出层级 $l<L$ 的神经网络 $\tilde{}\mathbf{g}_l:\ I\to\mathbb{R}^{m_{l,n_l-1}}$, $\tilde{}\mathbf{f}_l:\ I\to\mathbb{R}^2$, and$\tilde{e}_l^*$: $I\to\mathbb{C}$。要计算下一个层级，定义层级 $l+1$ 的误差函数如下：对于 $s\in I$,

$$\tilde{e}_{l+1}\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_{l+1}};s\right) := \tilde{e}_l^*(s)-\left[(\mathcal{I}-\mathcal{K}_{p_\kappa})\mathcal{TN}_{n_{l+1}}\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_{l+1}};\mathbf{g}_l(\cdot)\right)\right](s). \tag{19}$$

接下来考虑优化问题

$$\min_{\substack{\mathbf{W}_j\in\mathbb{R}^{m_{l+1,j}\times m_{l+1,j-1}}\\ \mathbf{b}_j\in\mathbb{R}^{m_{l+1,j}},\, j\in\mathbb{N}_{n_{l+1}}}} \left\|\tilde{e}_{l+1}\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_{l+1}};\cdot\right)\right\|_N^2, \tag{20}$$

其中 $m_{l+1,\,0}=m_{l,\,n_l-1}$ 和 $m_{l+1,\,n_{l+1}}=2$。(20) 的解给出最优参数

$$\{\tilde{\mathbf{W}}_{l+1,j}^*,\tilde{\mathbf{b}}_{l+1,j}^*\}_{j=1}^{n_{l+1}}.$$

然后，将层级 $l+1$ 的特征和解分量定义为

$$\tilde{\mathbf{g}}_{l+1}(s):=\mathcal{F}_{n_{l+1}-1}\left(\left\{\tilde{\mathbf{W}}_{l+1,j}^*,\tilde{\mathbf{b}}_{l+1,j}^*\right\}_{j=1}^{n_{l+1}-1};\tilde{\mathbf{g}}_l(s)\right),\quad s\in I,$$

and

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

相应的最优误差定义为

$$\tilde{e}_{l+1}^*(s):=\tilde{e}_{l+1}\left(\left\{\tilde{\mathbf{W}}_{l+1,j}^*,\tilde{\mathbf{b}}_{l+1,j}^*\right\}_{j=1}^{n_{l+1}};s\right)\in C(I),\quad s\in I. \tag{21}$$

完成所有 L 个层级后， y 的最终离散多级深度神经网络近似为

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

3 连续 MGDL 模型的误差分析

本节将根据训练损失，推导连续 MGDL 模型误差的上界和下界。我们首先证明，最优误差的范数关于层级数量是非递增的。利用这一单调性，我们进一步证明，对于具有 L 个层级的连续 MGDL 模型，层级 $l+1$ 处的逼近误差被层级 l 处误差的某个常数倍所界定，对于所有 $l=1$，2、.....、$L-1$ 均成立。

我们注意到，神经网络的最后一层执行的是不带激活函数的仿射变换。由于仿射变换在解析上已得到充分理解，我们将这一最后变换单独分离出来，以便进行理论分析。

对于每个层级 $l=1$，2，.....，L，定义如下参数空间

$$\Theta_{l,1}:=\left\{\left(\{\mathbf{W}_j,\mathbf{b}_j\}_{j=1}^{n_l-1}\right):\mathbf{W}_j\in\mathbb{R}^{m_{l,j}\times m_{l,j-1}},\mathbf{b}_j\in\mathbb{R}^{m_{l,j}},j\in\mathbb{N}_{n_l-1}\right\}$$

and

$$\Theta_{l,2}:=\{(\mathbf{W},\mathbf{b}):\mathbf{W}\in\mathbb{R}^{m_{l,n_l}\times m_{l,n_l-1}},\mathbf{b}\in\mathbb{R}^{m_{l,n_l}}\}.$$

由于层级 l 的特征函数 $\mathbf{g}_l$ 固定，算子 $\mathcal{M}_l : \Theta_{l,\,2} \to C(I)$ 定义如下

$$\mathcal{M}_l((\mathbf{W},\mathbf{b})) := (\mathcal{I}-\mathcal{K})\mathcal{T}(\mathbf{W}\mathbf{g}_l(\cdot)+\mathbf{b}) \in C(I), \quad (\mathbf{W},\mathbf{b}) \in \Theta_{l,2}. \tag{22}$$

令

$$\mathcal{M}_l(\Theta_{l,2}) := \{\mathcal{M}_l((\mathbf{W},\mathbf{b})) : (\mathbf{W},\mathbf{b}) \in \Theta_{l,2}\}$$

并注意到 $\mathcal{M}_l(\Theta_{l,\,2})$是一个线性空间。

同时，对于任意 $l \in \mathbb{N}_L$，定义

$$z_l := \mathcal{M}_l\left(\left(\mathbf{W}^*_{l,n_l}, \mathbf{b}^*_{l,n_l}\right)\right) \in C(I).$$

显然， $z_l \in \mathcal{M}_l(\Theta_{l,\,2})$。在下面的分析中，我们将证明，对于任意$l = 1，2，\ldots，L-1$，函数 z_{l+1} 是误差 e^*_l 在线性空间 $\mathcal{M}_{l+1}(\Theta_{l+1,\,2})$ 中的唯一最佳逼近。为建立这一结果，我们首先给出两个辅助引理。为保证一般性，并确保其适用于连续 MGDL 模型和离散 MGDL 模型，我们在统一框架下给出证明。

以下引理可用于确定神经网络中除最后一层之外的参数。

引理 1 设 $\mathcal{B}_1$ 和 $\mathcal{B}_2$ 为两个给定集合， $F:\ \mathcal{B}_1 \times \mathcal{B}_2 \to [0,+\infty)$为目标函数。若 $(\theta^*_1,\theta^*_2) \in \mathcal{B}_1 \times \mathcal{B}_2$ 是该问题的局部极小值点

$$\min_{(\theta_1,\theta_2)\in\mathcal{B}_1\times\mathcal{B}_2} F(\theta_1,\theta_2), \tag{23}$$

则 $\theta^*_2 \in \mathcal{B}_2$ 是约化问题的局部极小值点

$$\min_{\theta_2\in\mathcal{B}_2} F(\theta^*_1,\theta_2). \tag{24}$$

证明 我们用反证法证明。假设 θ^*_2 不是 (24) 的局部极小值点。那么存在一个序列 $\{s_j\} \subset \mathcal{B}_2$，使得 $s_j \to \theta^*_2$ 且

$$F(\theta^*_1,s_j) < F(\theta^*_1,\theta^*_2) \quad \text{for all } j \in \mathbb{N}. \tag{25}$$

定义 $t_j :=(\theta^*_1,\ s_j) \in \mathcal{B}_1 \times \mathcal{B}_2$。于是 $t_j \to (\theta^*_1,\ \theta^*_2)$，并且由 (25) 可得，

$$F(t_j) < F(\theta^*_1,\theta^*_2) \quad \text{for all } j \in \mathbb{N},$$

这与假设 $(\theta^*_1,\ \theta^*_2)$ 是 (23) 的局部极小值点相矛盾。因此， θ^*_2 必定是 (24) 的局部极小值点。

下述引理表明，由于目标函数的凸性，参数空间 $\mathcal{B}$ 中半范数误差的任意局部极小值点都对应于函数子空间 $\mathcal{U}(\mathcal{B})$中的全局最佳逼近。该结果在一般的半范数空间框架下表述，并将随后应用于离散 MGDL 模型。

引理 2 设 $\mathcal{X}$ 是配备半范数 $|\cdot|$的线性空间，且$\mathcal{U}$： $\mathcal{B} \to \mathcal{X}$ 是从巴拿赫空间 $\mathcal{B}$ 到 $\mathcal{X}$的线性算子。假设$s \in \mathcal{X}$。若 $t^* \in \mathcal{B}$ 是该优化问题的局部极小值点

$$\min_{t \in \mathcal{B}} |s - \mathcal{U}t|, \tag{26}$$

那么， $\mathcal{U}t^* \in \mathcal{U}(\mathcal{B})$ 是相对于半范数 $|\cdot|$， s 在由线性子空间 $\mathcal{U}(\mathcal{B})$构成的集合中的最佳逼近。

证明 函数 $t \mapsto |s - \mathcal{U}t|$ 在 $\mathcal{B}$ 上是凸的，因为它是线性映射与半范数的复合。因此，任何局部极小值点也是全局极小值点。

由于 $t^* \in \mathcal{B}$ 是式 (26) 的局部极小值点，它也必定是全局极小值点。因此，

$$|s - \mathcal{U}t^*| \leq |s - \mathcal{U}t|, \quad \text{for all } t \in \mathcal{B}.$$

这意味着， $\mathcal{U}t^* \in \mathcal{U}(\mathcal{B})$使其到 s 的半范数距离在集合 $\mathcal{U}(\mathcal{B})$上达到极小，即它是相对于 $|\cdot|$， s在由线性子空间 $\mathcal{U}(\mathcal{B})$构成的集合中的最佳逼近。

以下关于我们模型的结论可直接由前面两个引理得出。

引理 3 对于任意 $l \in \mathbb{N}_{L-1}$， z_{l+1} 是从线性空间 $\mathcal{M}_{l+1}(\Theta_{l+1,\ 2})$ 到 e_l^* 的唯一最佳逼近。

证明 对于任意 $l \in \mathbb{N}_{L-1}$，我们对以下选择应用引理 1： $\mathcal{B}_1 := \Theta_{l+1,\ 1}$， $\mathcal{B}_2 := \Theta_{l+1,\ 2}$，

$$\theta_1^* := \left(\{\mathbf{W}_{l+1,j}^*, \mathbf{b}_{l+1,j}^*\}_{j=1}^{n_{l+1}-1}\right), \quad \theta_2^* := \left(\mathbf{W}_{l+1,n_{l+1}}^*, \mathbf{b}_{l+1,n_{l+1}}^*\right).$$

令 $F\colon \Theta_{l+1,\ 1} \times \Theta_{l+1,\ 2} \to [0, +\infty)$ 为由训练误差定义的目标函数:

$$F(\theta_1, \theta_2) := \|e_{l+1}(\theta_1 \times \theta_2; \cdot)\|_2 .$$

由于 $(\theta_1^*,\ \theta_2^*)$ 是 F 的局部极小值点，引理 1 表明 θ_2^* 是约化问题的局部极小值点

$$\min_{\theta \in \Theta_{l+1,2}} \|e_{l+1}(\theta_1^* \times \theta; \cdot)\|_2 .$$

根据 (11) 中 e_{l+1} 的定义，这等价于最小化

$$\min_{\theta \in \Theta_{l+1,2}} \left\|e_l^* - (\mathcal{I} - \mathcal{K})\mathcal{T}\mathcal{N}_{n_{l+1}}(\theta_1^* \times \theta; \cdot)\right\|_2 .$$

根据式 (22) 中算子 $\mathcal{M}_{l+1}$ 的定义，这可归结为以下优化问题

$$\min_{\theta \in \Theta_{l+1,2}} \|e_l^* - \mathcal{M}_{l+1}(\theta)\|_2 .$$

现在，使用以下设置应用引理 2：

$$\mathcal{X} := C(I), \quad |\cdot| := \|\cdot\|_2, \quad \mathcal{U} := \mathcal{M}_{l+1}, \quad s := e_l^*, \quad t^* := \theta_2^*.$$

那么， $\mathcal{M}_{l+1}(\theta_2^*) = z_{l+1}$ 是 $\mathcal{M}_{l+1}(\Theta_{l+1,\ 2}) \subset C(I)$ 中关于 L_2 范数的 e_l^* 最佳逼近。

最后，由于 L_2 范数的严格凸性，以及 $\mathcal{M}_{l+1}(\Theta_{l+1,\ 2})$ 是希尔伯特空间 $L_2(I)$ 的闭线性子空间，因此解是唯一的。

现在我们证明序列 $\|e_l^*\|_2, l = 1,\ 2,\ \ldots\ldots,\ L$ 是非递增的。这一结果将 [16], 的定理 3 从最初针对函数逼近的情形扩展到积分方程的数值解。此外，[16] 中的定理 3 考虑了一种 MGDL 架构，其中添加新层级时保留输出层。相比之下，本文提出的结果适用于一种 MGDL框架，其中添加新层级时移除前一层级的输出层，这种设置更贴近实际实现。

命题 1 对于每个 $l = 1,2,\ldots,L-1$，要么 $z_{l+1} = 0$，要么 $\|e_{l+1}^*\|_2 \lhd \|e_l^*\|_2$。

证明 对于每个 $l = 1,2,\ldots,L-1$，假设 $z_{l+1} \neq 0$。根据式 (13) 中 e_{l+1}^* 的定义，我们有

$$e_{l+1}^* = e_l^* - z_{l+1}.$$

由于 0 和 z_{l+1} 是 $\mathcal{M}_{l+1}(\Theta_{l+1,2})$ 中的不同元素，并且根据引理 3，z_{l+1} 是从 $\mathcal{M}_{l+1}(\Theta_{l+1,2})$ 中对 e_l^* 的唯一最佳逼近，因此

$$\|e_{l+1}^*\|_2 = \|e_l^* - z_{l+1}\|_2 < \|e_l^* - 0\|_2 = \|e_l^*\|_2.$$

证明至此完成。

如果 1 不是线性算子 $\mathcal{K}$ 的特征值，那么根据 Fredholm 备择定理，算子 $\mathcal{I}-\mathcal{K}$ 可逆，且其逆算子 $(\mathcal{I}-\mathcal{K})^{-1}$ 有界。这一事实使我们能够推导出一个定理，用以刻画连续层级间的误差行为。在陈述该定理之前，我们先给出以下引理，该引理描述了 $l = 1,2,\ldots,L$ 中 y_l^* 与 e_l^* 之间的关系。

引理4 对于每个 $l = 1$、2、……、L，均有

$$e_l^* = (\mathcal{I} - \mathcal{K})(y - y_l^*).$$

证明 我们通过对 l 进行归纳法来证明这一结果。根据式（9）中 e_1^* 的定义，并利用 $f = (\mathcal{I} - \mathcal{K})y$，可得

$$e_1^* = f - (\mathcal{I} - \mathcal{K})\mathcal{T}\mathbf{f}_1 = (\mathcal{I} - \mathcal{K})(y - \mathcal{T}\mathbf{f}_1). \tag{27}$$

现在，假设对于某个 $l=1$，2，.....，$L-1$，恒等式

$$e_l^* = (\mathcal{I}-\mathcal{K})\left(y-\sum_{j=1}^{l}\mathcal{T}\mathbf{f}_j\right)$$

成立。然后，根据式 (13) 中 e_{l+1}^* 的递归定义，我们有

$$e_{l+1}^* = e_l^* - (\mathcal{I}-\mathcal{K})\mathcal{T}\mathbf{f}_{l+1}.$$

将归纳假设代入右端项可得

$$e_{l+1}^* = (\mathcal{I}-\mathcal{K})\left(y-\sum_{j=1}^{l}\mathcal{T}\mathbf{f}_j\right) - (\mathcal{I}-\mathcal{K})\mathcal{T}\mathbf{f}_{l+1} = (\mathcal{I}-\mathcal{K})\left(y-\sum_{j=1}^{l+1}\mathcal{T}\mathbf{f}_j\right).$$

因此，根据归纳法，该公式对所有 $l=1$，2、.....、L 均成立。利用式 (10) 中 y_l^* 的定义，结论得证。

现在，我们陈述一个定理，用于量化连续多层级学习模型中的逼近误差。

定理 1 假设 1 不是线性算子 $\mathcal{K}$ 的特征值。那么，算子 $\mathcal{I}-\mathcal{K}$ 是可逆的，并且对于每个 $l=1$，2、.....、$L-1$，以下误差估计成立：

$$\|y-y_{l+1}^*\|_2 \leqslant c\|y-y_l^*\|_2,$$

其中常数 $c:=\|(\mathcal{I}-\mathcal{K})^{-1}\|\|\mathcal{I}-\mathcal{K}\|$。

证明 由于 1 不是 $\mathcal{K}$ 的特征值，算子 $\mathcal{I}-\mathcal{K}$ 可逆，且其逆算子 $(\mathcal{I}-\mathcal{K})^{-1}$ 有界。定义常数

$$c_1 := \|(\mathcal{I}-\mathcal{K})^{-1}\|^{-1} > 0 \text{ and } c_2 := \|\mathcal{I}-\mathcal{K}\| > 0.$$

那么，对于任意 $h\in C(I)$，以下不等式成立：

$$c_1\|h\|_2 \leqslant \|(\mathcal{I}-\mathcal{K})h\|_2 \leqslant c_2\|h\|_2.$$

将此不等式应用于 $h:=y-y_{l+1}^*$ 和 $h:=y-y_l^*$，得

$$c_1\|y-y_{l+1}^*\|_2 \leqslant \|(\mathcal{I}-\mathcal{K})(y-y_{l+1}^*)\|_2 \tag{28}$$

and

$$\|(\mathcal{I}-\mathcal{K})(y-y_l^*)\|_2 \leqslant c_2\|y-y_l^*\|_2. \tag{29}$$

根据引理4，我们可以将式（28）和式（29）改写为

$$c_1\|y-y_{l+1}^*\|_2 \leqslant \|e_{l+1}^*\|_2, \quad \|e_l^*\|_2 \leqslant c_2\|y-y_l^*\|_2. \tag{30}$$

由命题 1 可知，要么是 $e_{l+1}^* = e_l^*$（仅当 $z_{l+1}=0$ 时才会出现），要么是 $\|e_{l+1}^*\|_2 < \|e_l^*\|_2$。无论哪种情况，我们都有

$$\|e_{l+1}^*\|_2 \leqslant \|e_l^*\|_2.$$

将其与式 (30) 结合，可得

$$c_1\|y-y_{l+1}^*\|_2 \leqslant \|e_{l+1}^*\|_2, \leq \|e_l^*\|_2 \leqslant c_2\|y-y_l^*\|_2,$$

这意味着

$$\|y-y_{l+1}^*\|_2 \leqslant \frac{c_2}{c_1}\|y-y_l^*\|_2.$$

并令

$$c := c_2/c_1 = \|(\mathcal{I}-\mathcal{K})^{-1}\|\|\mathcal{I}-\mathcal{K}\|,$$

我们得到所需结果。

定理 1 给出了一个最坏情况稳定性界，表明 MGDL 网络每增加一个层级时的解误差都受到一个常数因子的控制。特别地，该定理保证随着新层级的加入，解误差不会无限增长，从而确保多等级细化过程的稳定性。该常数

$$c := \|(\mathcal{I}-\mathcal{K})^{-1}\|\,\|\mathcal{I}-\mathcal{K}\|$$

是算子 $\mathcal{I}-\mathcal{K}$ 的条件数，并满足 $c \geq 1$。虽然该界

$$\|y-y_{l+1}^*\|_2 \leq c\|y-y_l^*\|_2$$

虽然这本身并不能保证解误差严格单调递减，但它确保增加层级时始终保持稳定性；对于条件适中的算子，还能通过多等级细化有效降低误差。

我们的目标是逼近真实解 y，但在层级 $l \in \mathbb{N}$ 处的实际误差 $\|y-y_l^*\|_2$通常无法在训练过程中获取，因为它取决于未知的真实解 y。相比之下，训练误差 $\|e_l^*\|_2$衡量的是数据 f 与算子作用于网络输出后所得结果（$\mathcal{I}-\mathcal{K}$）之间的差异，可以完全根据模型计算得出。要评估模型性能并指导训练过程，理解这一可观测的训练误差与真实（但隐藏的）解误差之间的关系至关重要。以下定理对这种关系进行了严格刻画。

定理 2 如果1 不是线性算子 $\mathcal{K}$的特征值，那么对于每个$l = 1$、2、.....、 L，都有

$$\|\mathcal{I}-\mathcal{K}\|^{-1}\|e_l^*\|_2 \leqslant \|y-y_l^*\|_2 \leqslant \|(\mathcal{I}-\mathcal{K})^{-1}\|\|e_l^*\|_2.$$

证明 由于1不是 $\mathcal{K}$的特征值，因此算子 $\mathcal{I}-\mathcal{K}$ 可逆，且其逆算子$(\mathcal{I}-\mathcal{K})^{-1}$ 有界。因此，对于每个 $l \in \mathbb{N}_L$，我们得到不等式

$$\|(\mathcal{I}-\mathcal{K})^{-1}\|^{-1}\|y-y_l^*\|_2 \leqslant \|(\mathcal{I}-\mathcal{K})(y-y_l^*)\|_2 \leqslant \|\mathcal{I}-\mathcal{K}\|\|y-y_l^*\|_2.$$

应用引理4，根据其结论 $e_l^* = (\mathcal{I}-\mathcal{K})(y-y_l^*)$，我们将上述不等式改写为

$$\|(\mathcal{I}-\mathcal{K})^{-1}\|^{-1}\|y-y_l^*\|_2 \leqslant \|e_l^*\|_2 \leqslant \|\mathcal{I}-\mathcal{K}\|\|y-y_l^*\|_2,$$

从而得到所需的估计。

定理 2 建立了一个双侧界，表明真实解误差受到可观测训练误差的严格控制。因此，即使无法获得精确解，也可以可靠地监测该方法的收敛性。

改进的实际机制由训练误差 $\|e_l^*\|_2$ 的优化驱动。如命题 1 所示，在每个层级上成功完成优化，可保证训练误差序列单调不增。定理 2 随后通过根据训练误差对解误差进行上下界定，为训练误差与解误差建立了关键联系。因此，训练误差的降低会迫使包含解误差的可行区间向下移动，表明近似确实得到了改进。条件数 c刻画了这种传递的效率： c 越接近 1，训练误差的下降与解误差的降低之间的关联就越紧密。

总之，定理 1 保证了多等级细化在最坏情况下的稳定性，而训练误差的单调优化与定理 2 的双侧界共同构成了一种实用且有理论依据的机制，用于改进解的近似。

4 离散 MGDL 模型的误差分析

本节将建立离散 MGDL 模型的关键性质，这些性质与第 3 节讨论的连续模型的性质相对应。具体而言，我们将推导离散解逼近误差的上下界，并用训练损失和求积误差表示。主要结果根据训练误差（可在优化过程中计算）和求积误差（反映数值积分方案的精度），量化离散 MGDL 解与真实解之间的接近程度。

沿用连续情形中的类似方法，对于每个 $l = 1, 2, \ldots, L$，~我们固定层级为 l 的特征 ~$\mathbf{g}_l$，并定义算子 $\mathcal{M}_l : \Theta_{l,2} \to C(I)$ 如下：

$$\tilde{\mathcal{M}}_l((\mathbf{W}, \mathbf{b})) := (\mathcal{I} - \mathcal{K}_{p_\kappa})\mathcal{T}(\mathbf{W}\tilde{\mathbf{g}}_l(\cdot) + \mathbf{b}) \in C(I), \text{ for all } (\mathbf{W}, \mathbf{b}) \in \Theta_{l,2}. \quad (31)$$

~我们仿照连续情形，定义线性空间 $\mathcal{M}_l(\Theta_{l,2})$。此外，我们设

$$\tilde{z}_l := \tilde{\mathcal{M}}_l\left(\left(\tilde{\mathbf{W}}_{l,n_l}^*, \tilde{\mathbf{b}}_{l,n_l}^*\right)\right) \in \tilde{\mathcal{M}}_l(\Theta_{l,2}). \quad (32)$$

于是，直接类比连续 MGDL 模型的引理 3，我们立即得到以下引理。

引理 5 对于任意 $l = 1, 2, \ldots, L-1$, $\tilde{z}_{l+1}$ 是 $\tilde{e}_l^{*~}$ 在由 $\mathcal{M}_{l+1}(\Theta_{l+1,2})$ 构成的线性空间中的最佳逼近，意义上采用半范数 $\|\cdot\|_N$。

证明 固定任意 $l = 1, 2, \ldots\ldots, L-1$。我们采用以下设置应用引理 1：

$$\mathcal{B}_1 := \Theta_{l+1,1}, \quad \mathcal{B}_2 := \Theta_{l+1,2},$$

$$\theta_1^* := \left(\{\tilde{\mathbf{W}}_{l+1,j}^*, \tilde{\mathbf{b}}_{l+1,j}^*\}_{j=1}^{n_{l+1}-1}\right) \in \Theta_{l+1,1},$$

$$\theta_2^* := \left(\tilde{\mathbf{W}}_{l+1,n_{l+1}}^*, \tilde{\mathbf{b}}_{l+1,n_{l+1}}^*\right) \in \Theta_{l+1,2},$$

以及方程 (20) 中定义的目标函数 F。

由引理 1，(θ_1^*,θ_2^*)意味着θ_2^* 是函数 $\theta \mapsto \|\tilde{e}_{l+1}(\theta_1^* \times \theta;\ \cdot)\|_N^2$的局部极小值点。根据方程 (19) 中 $\tilde{e}_{l+1}$的定义，这一极小化等价于

$$\min_{\theta\in\Theta_{l+1,2}} \left\|\tilde{e}_l^* - \tilde{\mathcal{M}}_{l+1}(\theta)\right\|_N^2.$$

现在应用引理 2，其中:

$$\mathcal{X} := C(I), \quad |\cdot| := \|\cdot\|_N, \quad \mathcal{U} := \tilde{\mathcal{M}}_{l+1}, \quad \mathcal{B} := \Theta_{l+1,2}, \quad s := \tilde{e}_l^*, \quad t^* := \theta_2^*.$$

˜由引理 2，元素 $\tilde{z}_{l+1} := \mathcal{U}t^* = \mathcal{M}_{l+1}(\theta_2^*)$是线性空间 $\mathcal{M}_{l+1}(\Theta_{l+1,2})$中相对于半范数 $\|\cdot\|_N$对 $\tilde{e}_l^*$的最佳逼近˜。

与连续情形不同，引理 5 中定义的离散最佳逼近不一定唯一。尽管如此，我们仍可建立一个类似且至关重要的性质。为此，我们先给出以下引理，它本质上确立了离散半范数 $\|\cdot\|_N$的严格凸性。

引理 6 对于任意满足 $\|\phi\|_N = 1$、$\|\psi\|_N = 1$、$\|\phi-\psi\|_N \neq 0$ 的 ϕ、$\psi \in C(I)$，有

$$\|x\phi + (1-x)\psi\|_N < 1, \quad \textit{for any} \quad x \in (0,1).$$

证明 定义 $\mathbf{v}_\phi := [\phi(x_j)]_{j=1}^N$ 和 $\mathbf{v}_\psi := [\psi(x_j)]_{j=1}^N$ 于 $\mathbb{C}^N$ 中。根据 $\|\cdot\|_N$ 的定义及关系式（14），可得

$$\|\phi\|_N = \|\psi\|_N = 1 \quad \Rightarrow \quad \|\mathbf{v}_\phi\|_2 = \|\mathbf{v}_\psi\|_2 = \sqrt{N},$$

和 $\|\phi - \psi\|_N \neq 0 \Rightarrow \mathbf{v}_\phi \neq \mathbf{v}_\psi$。

$\mathbb{C}^N$ 中的 ℓ_2-范数是严格凸的，因此对于所有 $x \in$ （0，1），

$$\left\| x \cdot \frac{\mathbf{v}_\phi}{\sqrt{N}} + (1-x) \cdot \frac{\mathbf{v}_\psi}{\sqrt{N}} \right\|_2 < 1.$$

根据线性性和 $\|\cdot\|_N$ 的定义，我们进而有

$$\|x\phi + (1-x)\psi\|_N = \frac{\|x\mathbf{v}_\phi + (1-x)\mathbf{v}_\psi\|_2}{\sqrt{N}} < 1.$$

这就证明了该结论。

与最佳逼近唯一的连续模型不同，离散模型满足以下替代性质。

引理 7 对于任意 $l \in \mathbb{N}_{L-1}$，设 $\phi \in \mathcal{M}_{l+1}(\Theta_{l+1,2})$ 是关于半范数 $\|\cdot\|_N$，从线性空间 $\mathcal{M}_{l+1}(\Theta_{l+1,2})$ 到 $\tilde{e}_l^*$ 的一个最佳逼近。则

$$\|\phi - \tilde{z}_{l+1}\|_N = 0.$$

证明 定义 $d := \|\phi - \tilde{e}_l^*\|_N$。根据引理 5，我们还有 $\|\tilde{z}_{l+1} - \tilde{e}_l^*\|_N = d$，且对所有 $h \in \mathcal{M}_{l+1}(\Theta_{l+1,2})$，

$$\|h - \tilde{e}_l^*\|_N \geq d. \tag{33}$$

我们考虑两种情况：

情况 1：$d = 0$。那么，根据半范数 $\|\cdot\|_N$ 的三角不等式，我们得到

$$\|\phi - \tilde{z}_{l+1}\|_N \leq \|\phi - \tilde{e}_l^*\|_N + \|\tilde{e}_l^* - \tilde{z}_{l+1}\|_N = 2d = 0.$$

情况 2：$d \neq 0$。由于 ϕ 和 $\tilde{z}_{l+1}$ 都属于线性空间 $\mathcal{M}_{l+1}(\Theta_{l+1,2})$，它们的凸组合也属于该空间。对于任意 $\mu \in (0,1)$，定义

$$h := \mu\phi + (1-\mu)\tilde{z}_{l+1} \in \tilde{\mathcal{M}}_{l+1}(\Theta_{l+1,2}).$$

利用（33），我们有

$$\|h - \tilde{e}_l^*\|_N = \|\mu\phi + (1-\mu)\tilde{z}_{l+1} - \tilde{e}_l^*\|_N \geq d. \tag{34}$$

定义归一化向量：

$$\psi_1 := \frac{\phi - \tilde{e}_l^*}{d}, \quad \psi_2 := \frac{\tilde{z}_{l+1} - \tilde{e}_l^*}{d},$$

使得 $\|\psi_1\|_N = \|\psi_2\|_N = 1$。接下来我们观察到

$$\frac{1}{d}[(\mu\phi + (1-\mu)\tilde{z}_{l+1}) - \tilde{e}_l^*] = \mu\psi_1 + (1-\mu)\psi_2, \quad \text{for all} \quad \mu \in (0,1).$$

利用上述方程，不等式 (34) 化为：

$$\|\mu\psi_1 + (1-\mu)\psi_2\|_N \geq 1.$$

此外，由于 $d \neq 0$，我们有 $\|\psi_1 - \psi_2\|_N \neq 0$。这与半范数 $\|\cdot\|_N$ 的严格凸性（引理 6）矛盾，因此可得：

$$\|\mu\psi_1 + (1-\mu)\psi_2\|_N < 1, \quad \text{for all } \mu \in (0,1).$$

因此，所需结果得证。

至此，我们可以证明序列 $\|\tilde{e}_l^*\|_N$（其中 l =1，2，.....，L）是非递增的。

命题 2 对于每个 $l = 1$、2、.....、$L-1$，最优误差满足

$$\|\tilde{e}_{l+1}^*\|_N \leq \|\tilde{e}_l^*\|_N.$$

此外，当且仅当 $\|\tilde{z}_{l+1}\|_N = 0$ 时等号成立；也就是说，

$$\|\tilde{e}_{l+1}^*\|_N = \|\tilde{e}_l^*\|_N \quad \Longleftrightarrow \quad \|\tilde{z}_{l+1}\|_N = 0.$$

证明 固定 $l=1$，2，.....，$L-1$。根据式（21）中 $\tilde{e}_{l+1}^*$ 的定义以及式（32）中 $\tilde{z}_{l+1}$ 的定义，我们有

$$\tilde{e}_{l+1}^* = \tilde{e}_l^* - \tilde{z}_{l+1}.$$

根据引理 5，$\tilde{z}_{l+1}$ 是从空间 $\mathcal{M}_{l+1}(\Theta_{l+1,\ 2})$中对 $\tilde{e}_l^*$的最佳逼近。由于零函数属于该空间，我们得到

$$\|\tilde{e}_{l+1}^*\|_N = \|\tilde{e}_l^* - \tilde{z}_{l+1}\|_N \leq \|\tilde{e}_l^* - 0\|_N = \|\tilde{e}_l^*\|_N,$$

这证明了第一个断言。

为证明第二个结论，假设

$$\|\tilde{e}_{l+1}^*\|_N = \|\tilde{e}_l^*\|_N.$$

那么，根据上面的不等式，等号必须成立：

$$\|\tilde{e}_l^* - \tilde{z}_{l+1}\|_N = \|\tilde{e}_l^* - 0\|_N.$$

这意味着 $\tilde{z}_{l+1}$ 和 0 都是 $\mathcal{M}_{l+1}(\Theta_{l+1,\ 2})$中对 $\tilde{e}_l^*$的最佳逼近。根据引理 7，它们在半范数 $\|\cdot\|_N$下的差必须为零，因此我们有

$$\|\tilde{z}_{l+1}\|_N = \|\tilde{z}_{l+1} - 0\|_N = 0.$$

命题 2 将 [16],中最初针对函数逼近提出的定理 5 推广到数值求解积分方程的情形。命题 2 的第二部分进一步揭示，若 $\|\tilde{z}_{l+1}\|_N \neq 0$，则训练误差 $\|\tilde{e}_{l+1}^*\|_N$ 必然严格递减。换言之，层级 $l+1$对改进解有实质性贡献。反之，若 $\|\tilde{z}_{l+1}\|_N$ =0，则在逼近空间$\mathcal{M}_{l+1}(\Theta_{l+1,\ 2})$内无法进一步改进。

注 *1* 这提供了一种实用的诊断准则：可以通过判断新增层级是否 $\tilde{z}_{l+1}$ 非零来评估其有效性。如果改进微乎其微（即 $\|\tilde{z}_{l+1}\|_N \approx 0$），那么可能没有必要继续增加层级，并且可以提前停止训练。

接下来，我们将层级 l 下的离散 MGDL 逼近定义为

$$\tilde{y}_l^* := \sum_{j=1}^{l} \mathcal{T}\tilde{\mathbf{f}}_j \in C(I), \quad l = 1, 2, \ldots, L. \tag{35}$$

以下引理给出了最优误差 $\tilde{e}_l^*$关于真实解 y 和离散逼近 $\tilde{y}_l^*$的表示。

引理 8 对于每个 $l \in \mathbb{N}_L$，最优误差满足

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

证明 由方程 (18) 中 $\tilde{e}_1^*$ 的定义可得

$$\tilde{e}_1^* = f - (\mathcal{I} - \mathcal{K}_{p_\kappa})\mathcal{T}\tilde{\mathbf{f}}_1 = (\mathcal{I} - \mathcal{K}_{p_\kappa})(y - \mathcal{T}\tilde{\mathbf{f}}_1) + (\mathcal{K}_{p_\kappa} - \mathcal{K})y. \tag{36}$$

对于 $l = 1, 2, \ldots, L-1$，由方程 (21) 中 $\tilde{e}_{l+1}^*$ 的递归定义可得

$$\tilde{e}_{l+1}^* = \tilde{e}_l^* - (\mathcal{I} - \mathcal{K}_{p_\kappa})\mathcal{T}\tilde{\mathbf{f}}_{l+1}.$$

递归应用这一结果并利用方程 (36)，可得

$$\tilde{e}_l^* = (\mathcal{I} - \mathcal{K}_{p_\kappa}) \left(y - \sum_{j=1}^{l} \mathcal{T}\tilde{\mathbf{f}}_j \right) + (\mathcal{K}_{p_\kappa} - \mathcal{K})y.$$

根据方程 (35) 中 $\tilde{y}_l^*$ 的定义，这可化简为所需表达式。

现在我们给出定理 1 的离散类比，回到离散系统 (15)。当求积节点与配置点重合时（此时为 $p_\kappa = N - 1$），并且 $\mathcal{K}_{p_\kappa}$ 逐点收敛于 $\mathcal{K}$ 时，对于充分大的 N，存在满足下式的可逆矩阵 $\mathbf{M}_\kappa \in \mathbb{C}^{N\times N}$

$$[((\mathcal{I} - \mathcal{K}_{p_\kappa})h)(x_j)]_{j=1}^N = \mathbf{M}_\kappa \mathbf{v}_h, \tag{37}$$

对于任意函数 $h \in C(I)$，其中 $\mathbf{v}_h := [h(x_j)]_{j=1}^N$。

因此，离散积分方程 (15) 化为线性系统

$$\mathbf{M}_\kappa \mathbf{v}_y = \mathbf{v}_f, \tag{38}$$

其中 $\mathbf{v}^f := [f(x_j)]_{j=1}^N$，且 $\mathbf{v}^y := [y(x_j)]_{j=1}^N$。

引理 9 对于任意 $h \in C(I)$，以下界成立：

$$\|\mathbf{M}_\kappa^{-1}\|_2^{-1}\|h\|_N \leqslant \|(\mathcal{I} - \mathcal{K}_{p_\kappa})h\|_N \leqslant \|\mathbf{M}_\kappa\|_2\|h\|_N.$$

证明 对于任意 $h \in C(I)$，令 $\mathbf{v}_h := [h(x_j) : j \in \mathbb{N}_N]^T \in \mathbb{C}^N$。根据关系式 (14) 和$\mathbf{M}_\kappa$ 的定义，可得

$$\|h\|_N = \frac{\|\mathbf{v}_h\|_{\ell_2}}{\sqrt{N}}, \qquad \|(\mathcal{I} - \mathcal{K}_{p_\kappa})h\|_N = \frac{\|\mathbf{M}_\kappa \mathbf{v}_h\|_{\ell_2}}{\sqrt{N}}. \tag{39}$$

利用谱范数的次乘性和可逆性，可得

$$\|\mathbf{M}_\kappa^{-1}\|_2^{-1}\|\mathbf{v}_h\|_{\ell_2} \leqslant \|\mathbf{M}_\kappa \mathbf{v}_h\|_{\ell_2} \leqslant \|\mathbf{M}_\kappa\|_2\|\mathbf{v}_h\|_{\ell_2}.$$

两边除以 $\sqrt{N}$ 并应用 (39)，即可得到所需的估计。

现在我们给出定理 1 的离散类比。

定理 3 对于 $l \in \mathbb{N}_{L-1}$，以下估计成立：

$$\|y - \tilde{y}_{l+1}^*\|_N \leq \mathrm{cond}(\mathbf{M}_\kappa)\|y - \tilde{y}_l^*\|_N + 2\|\mathbf{M}_\kappa^{-1}\|_2\|(\mathcal{K}_{p_\kappa} - \mathcal{K})y\|_N,$$

其中，$\mathrm{cond}(\mathbf{M}_\kappa)$ 表示矩阵 $\mathbf{M}_\kappa$ 的条件数。

证明 根据引理 8，对于所有 $l \in \mathbb{N}_{L-1}$，我们有

$$\tilde{e}_l^* = (\mathcal{I} - \mathcal{K}_{p_\kappa})(y - \tilde{y}_l^*) + (\mathcal{K}_{p_\kappa} - \mathcal{K})y,$$

$$\tilde{e}_{l+1}^* = (\mathcal{I} - \mathcal{K}_{p_\kappa})(y - \tilde{y}_{l+1}^*) + (\mathcal{K}_{p_\kappa} - \mathcal{K})y.$$

将其代入不等式 $\|\tilde{e}_{l+1}^*\|_N \leq \|\tilde{e}_l^*\|_N$ （由命题 2 得出），得到

$$\|(\mathcal{I} - \mathcal{K}_{p_\kappa})(y - \tilde{y}_{l+1}^*) + (\mathcal{K}_{p_\kappa} - \mathcal{K})y\|_N \leq \|(\mathcal{I} - \mathcal{K}_{p_\kappa})(y - \tilde{y}_l^*) + (\mathcal{K}_{p_\kappa} - \mathcal{K})y\|_N.$$

应用三角不等式可得

$$\|(\mathcal{I} - \mathcal{K}_{p_\kappa})(y - \tilde{y}_{l+1}^*)\|_N \leq \|(\mathcal{I} - \mathcal{K}_{p_\kappa})(y - \tilde{y}_l^*)\|_N + 2\|(\mathcal{K}_{p_\kappa} - \mathcal{K})y\|_N.$$

利用引理 9，根据矩阵 $\mathbf{M}_\kappa$对算子范数进行界定，可得

$$\|\mathbf{M}_\kappa^{-1}\|_2^{-1}\|y - \tilde{y}_{l+1}^*\|_N \leq \|\mathbf{M}_\kappa\|_2\|y - \tilde{y}_l^*\|_N + 2\|(\mathcal{K}_{p_\kappa} - \mathcal{K})y\|_N.$$

将两边乘以 $\|\mathbf{M}_\kappa^{-1}\|_2$ ，即可得到结果。

定理 3 表明，在离散（可实现）情形下，层级 $l+1$ 处的误差有界于层级 l处的误差乘以系统矩阵 $\mathbf{M}_\kappa$的条件数，再加上一个反映求积误差的项，即用 $\mathcal{K}_{p_\kappa}$逼近 $\mathcal{K}$ 的精度。这与连续情形相呼应，并证实离散化后，实际的 MGDL 算法仍保留其误差缩减特性。当$\mathbf{M}_\kappa$ 条件良好（通常如此） [2], 且求积误差较小时，随着每增加一个层级，逼近误差都会稳定下降。

在下一节中，我们将证明，求积误差

$$R_{p_\kappa} := \|(\mathcal{K}_{p_\kappa} - \mathcal{K})y\|_N \tag{40}$$

通过选择足够大的 p_κ，可以变得任意小。因此，在适当的条件下，离散 MGDL 模型的行为与其连续对应模型在定理 1 中的行为相似。

下面的定理表明，最小化训练误差会相应地减小解误差。

定理 4 对于所有 $l \in \mathbb{N}_L$，以下估计成立：

$$\|\mathbf{M}_\kappa\|_2^{-1}(\|\tilde{e}_l^*\|_N - R_{p_\kappa}) \leqslant \|y - \tilde{y}_l^*\|_N \leqslant \|\mathbf{M}_\kappa^{-1}\|_2(\|\tilde{e}_l^*\|_N + R_{p_\kappa}), \tag{41}$$

其中， $R_{p_\kappa}：=\|\left(\mathcal{K}_{p_\kappa}-\mathcal{K}\right) y\|_N$ 是定义于 （40）中的求积误差。

证明 我们首先证明（41）中的上界。根据引理 8，有

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

应用三角不等式可得

$$\|(\mathcal{I} - \mathcal{K}_{p_\kappa})(y - \tilde{y}_l^*)\|_N \leqslant \|\tilde{e}_l^*\|_N + R_{p_\kappa}.$$

利用引理 9，我们得到

$$\|y-\tilde{y}_l^*\|_N \le \|\mathbf{M}_\kappa^{-1}\|_2(\|\tilde{e}_l^*\|_N + R_{p_\kappa}),$$

这就证明了上界。

对于下界，根据引理 8 和三角不等式，同样有：

$$\|\tilde{e}_l^*\|_N \leqslant \|(\mathcal{I}-\mathcal{K}_{p_\kappa})(y-\tilde{y}_l^*)\|_N + R_{p_\kappa}.$$

再次应用引理 9，得到

$$\|\tilde{e}_l^*\|_N \leqslant \|\mathbf{M}_\kappa\|_2\|y-\tilde{y}_l^*\|_N + R_{p_\kappa},$$

整理后即可得到所需的下界：

$$\|y-\tilde{y}_l^*\|_N \ge \|\mathbf{M}_\kappa\|_2^{-1}(\|\tilde{e}_l^*\|_N - R_{p_\kappa}).$$

定理 4 为下一节介绍的自适应多级深度学习算法提供了关键的理论依据。它表明，解误差被训练误差和求积误差紧密地夹在上下界之间。因此，假设求积误差较小，最小化训练误差就能可靠地降低真实解误差。这证明了将训练误差用作解的精度的替代指标是合理的，而这对于自适应增加 MGDL 模型的层级数至关重要。

5 自适应 MGDL 算法

本节介绍自适应 MGDL（AMGDL）算法，用于在预设容差范围内近似求解该方程。每个层级由一个仅含单个隐藏层的浅层网络构成。与第 2 节不同，我们不预先设定层级数的上限 L；相反，AMGDL 会自动选择最优层级数。积分算子 $\mathcal{K}$ 采用复合梯形法则进行近似。我们建立了一项理论结果，证明 AMGDL 得出的解保证位于指定容差范围内，且计算时间随层级数线性增长。

我们首先回顾先前工作 [9], 中的离散振荡积分算子 $\mathcal{K}_{p_\kappa}$，该算子提供了 $\mathcal{K}$的数值近似。求积节点数定义为 $p_\kappa := \lceil \gamma\kappa^\beta \rceil$，其中参数 γ 和β 满足

$$\Gamma \geqslant 0, \quad \beta \geqslant 1, \quad \gamma \geqslant \Gamma + 3. \tag{42}$$

使用复合梯形法则，我们定义离散算子 $\mathcal{K}_{p_\kappa}$ 如下：对于任意 $F \in C(I)$ 和 $s \in I$，

$$\begin{aligned}(\mathcal{K}_{p_\kappa}F)(s) := \frac{h}{2}\Bigg[& F(s_0)K(s,s_0)e^{i\kappa|s-s_0|} + 2\sum_{j=0}^{p_\kappa} K(s,s_j)F(s_j)e^{i\kappa|s-s_j|} \\ & + K(s,s_{p_\kappa})F(s_{p_\kappa})e^{i\kappa|s-s_{p_\kappa}|}\Bigg],\end{aligned}$$

其中 $h := 2/p_\kappa$， $s_j := -1 + jh$，对于 $j = 0, 1, \ldots\ldots, p_\kappa$。训练数据$\{(x_j, f(x_j)) : j \in \mathbb{N}_{N_\kappa}\}$在均匀网格上的配置点处采样:

$$x_j := -1 + \frac{2(j-1)}{N_\kappa - 1}, \quad j \in \mathbb{N}_{N_\kappa}, \quad N_\kappa := p_\kappa q + 1, \quad \text{for a fixed } q \in \mathbb{N}.$$

在训练网格上，$(\mathcal{I} - \mathcal{K}_{p_\kappa})$的矩阵表示为 [9]:

$$\mathbf{M}_\kappa := \mathbf{I}_\kappa - \frac{\mathbf{B}_\kappa}{p_\kappa} \in \mathbb{C}^{N_\kappa \times N_\kappa}, \tag{43}$$

其中，$\mathbf{I}_\kappa$ 是单位矩阵，$\mathbf{B}_\kappa = (b_{j,l}(\kappa))_{j,l \in \mathbb{N}_{N_\kappa}}$ 定义为

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

其中 $\omega_\kappa := e^{\frac{2\kappa i}{qp_\kappa}}$。

接下来，我们给出 $\|(\mathcal{K}_{p_\kappa} - \mathcal{K})y\|_N$的误差界。根据 [14], , 存在函数 u_j、$j \in \mathbb{N}_3$，满足

$$|u_j^{(l)}(s)| \leqslant \tau, \quad \text{for all } \ s \in I, \ \ l \in \mathbb{Z}_{m+1}, \tag{44}$$

对于某个 $\tau > 0$，方程（2）的解 y 属于$H^m_{\kappa,0}(I)$。也就是说，

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

方程 (44) 结合核 K 的光滑性（特别是关于其第二个变量的光滑性），可知存在一个常数 $\tilde{\tau} > 0$，使得对每个 $t \in I$,

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

对于离散积分算子 $\mathcal{K}_{p_\kappa}$， [9] 中的命题 4.3 指出

$$\|\mathcal{K}y - \mathcal{K}_{p_\kappa} y\|_\infty \leqslant \mathcal{E}(\gamma, \beta, \kappa, m), \tag{45}$$

其中

$$\mathcal{E}(\gamma, \beta, \kappa, m) := \frac{132\tilde{\tau}}{5\gamma\kappa^\beta} + \frac{81\tilde{\tau}(\Gamma + 3)^m}{5\gamma^m \kappa^{m(\beta-1)}}. \tag{46}$$

由于

$$\|\mathcal{K}y - \mathcal{K}_{p_\kappa} y\|_{N_\kappa} \leqslant \|\mathcal{K}y - \mathcal{K}_{p_\kappa} y\|_\infty,$$

根据 $\|\cdot\|_{N_\kappa}$ 的定义，可得

$$R_{p_\kappa} = \|\mathcal{K}y - \mathcal{K}_{p_\kappa} y\|_{N_\kappa} \leqslant \mathcal{E}(\gamma, \beta, \kappa, m), \tag{47}$$

其中 $R_{p_\kappa} := \|(\mathcal{K}_{p_\kappa} - \mathcal{K})y\|_N$ 是 (40) 中定义的求积误差。

通过分析 $\mathcal{E}(\gamma, \beta, \kappa, m)$ 的定义，可以看出，当 β 和 γ 足够大时，积分误差 R_{p_κ} 可以任意小。

此外，由定理 4 中的方程 (41) 可知，较小的积分误差可确保较小的训练误差 $\|e_l^*\|_N$ 直接转化为较小的整体解误差 $\|y-\tilde{y}_l^*\|_N$。

我们的目标是开发一种自适应多级方法，使解误差 $\|y-\tilde{y}_l^*\|_N$ 保持在预设容差 $\epsilon>0$范围内。然而，由于无法直接观测该误差，我们转而使用可直接计算的训练误差$\|e_l^*\|_N$。通过监测 $\|e_l^*\|_N$，我们可以为多级学习过程确定合适的停止准则。这种自适应策略能够提升实际应用效果，并确保算法高效收敛。自适应算法描述如下。

算法1 振荡Fredholm积分方程的自适应多级深度学习（含误差容限） 要求：弗雷德霍姆积分方程 (2)、误差容限 $\epsilon>0$、最大层级数 $L\in\mathbb{N}$、满足条件 (42) 和 $q\in\mathbb{N}$ 的参数 Γ,β,γ。过程：1：初始化：设置层级索引 $l=1$。定义参数为 $\{\mathbf{W}_j^l,\mathbf{b}_j^l\}_{j=1}^2$ 的网络 $\mathcal{N}_2$。2：求解最小化问题 eq. (17)，得到 $\tilde{\mathbf{f}}_l$，误差为 $\tilde{e}_l^*$。3：当 $l<L$ 且 $\|\tilde{e}_l^*\|_N>\dfrac{\epsilon}{2\|\mathbf{M}_\kappa^{-1}\|_2}$ 时执行 4：增加层级：$l\leftarrow l+1$。5：初始化参数为 $\{\mathbf{W}_j^l,\mathbf{b}_j^l\}_{j=1}^2$ 的网络 $\mathcal{N}_2$。6：求解最小化问题 (20)，得到 $\tilde{\mathbf{f}}_l$，误差为 $\|\tilde{e}_l^*\|_N$。7：结束循环。8：返回：$\tilde{y}_l^*:=\sum_{j=1}^l\mathcal{T}\tilde{\mathbf{f}}_j$。

在该算法中，数字 L 与传统 SGL 模型中的含义不同；在传统 SGL 模型中，它表示一个参数，使得 $\sum_{j=1}^L n_j$ 给出网络中隐藏层的总数。此处，L 则表示允许的最大层级数；在收敛缓慢的情况下，它也作为停止准则的一部分。

请注意，算法1中的每个层级都对应一个只有单个隐藏层的浅层网络 $\mathcal{N}_2$；也就是说，对于所有 j，都有 $n_j:=2$。在 [8], 中，基于 [10], 的研究表明，当 MGDL 的每个层级都由一个单隐藏层 ReLU 网络构成时，该方法可归结为求解一系列凸子问题，每个层级对应一个子问题。同样地，我们可以证明，采用 ReLU 激活函数时，优化问题 (17) 和 (20) 等价于凸优化问题。这揭示了 AMGDL 的另一项重要优势：所得凸问题可以通过梯度下降高效求解，并保证收敛性。

下面的定理证明，算法1生成的近似解满足规定的误差容限。

定理 5 如果离散矩阵 $\mathbf{M}_\kappa$ 可逆，并且假设自适应多层次深度学习算法在迭代 l^* 时终止，使得

$$\|\tilde{e}_{l^*}^*\|_N\le\frac{\epsilon}{2\|\mathbf{M}_\kappa^{-1}\|_2} \tag{48}$$

对于给定的容差 $\epsilon > 0$。如果求积误差界满足

$$\mathcal{E}(\gamma, \beta, \kappa, m) \leq \|\tilde{e}_{l^*}^*\|_N,$$

则所得近似解 $\tilde{y}_{l^*}^*$ 满足

$$\|y - \tilde{y}_{l^*}^*\|_N \leq \epsilon.$$

证明 由定理 4，真实解 y 与近似解 $\tilde{y}_{l^*}^*$ 之间的误差有界于

$$\|y - \tilde{y}_{l^*}^*\|_N \leq \|\mathbf{M}_\kappa^{-1}\|_2 \left(\|\tilde{e}_{l^*}^*\|_N + \mathcal{E}(\gamma, \beta, \kappa, m)\right).$$

由于 $\mathcal{E}(\gamma, \beta, \kappa, m) \leq \|\tilde{e}_{l^*}^*\|_N$，因此

$$\|y - \tilde{y}_{l^*}^*\|_N \leq 2\|\mathbf{M}_\kappa^{-1}\|_2 \|\tilde{e}_{l^*}^*\|_N.$$

最后，由停止准则（48）可得

$$\|y - \tilde{y}_{l^*}^*\|_N \leq \epsilon,$$

由此完成证明。

注。 在算法1的实际实现中，可以用一种更简单、更方便的准则，替代基于预先设定的误差容限的停止规则。具体而言，命题 2 表明，训练误差序列 $\|\tilde{e}_l^*\|_N$ 随层级索引 $l \in \mathbb{N}_L$ 单调递减。这种单调性意味着，如果两个连续层级之间的训练误差没有下降，即，

$$\|\tilde{e}_l^*\|_N = \|\tilde{e}_{l-1}^*\|_N,$$

那么进一步增加层级不太可能改善近似。因此，满足这一条件时即可终止算法。这种替代停止规则在实际应用中特别有用，因为它无需估计或预先设定逆矩阵 $\|\mathbf{M}_\kappa^{-1}\|_2$ 的范数和误差容限 ϵ。

最后，我们证明 AMGDL˜算法的计算成本随层级数量线性增长。设$\mathbf{f}_l$ 是一个包含 l 个层级、使用算法1训练的深度神经网络，并假设所训练网络的各层级˜的宽度均由一个正常数 ρ 有界。记 $\mathcal{T}(\mathbf{f}_l)$ ˜为$\mathbf{f}_l$的总训练时间，并令 $c(p)$ 表示求解宽度为 p 的网络的优化问题所需的计算时间。

˜**定理6** 假设$\mathbf{f}_l$ 是一个具有 l 个层级、通过算法*1*训练的深度神经网络，各层级的宽度均由正的常数 ρ有界，并令 $c(p)$ 表示求解宽度为 p的网络的优化问题所需的计算时间。那么总训练时间满足

$$\mathcal{T}(\tilde{\mathbf{f}}_l) \leq c(\rho) l. \tag{49}$$

证明 总训练时间是各层级所用时间之和：

$$\mathcal{T}(\tilde{\mathbf{f}}_l) = \sum_{j=1}^{l} c(p_j).$$

由于层级 j（$j = 1$，2，.....，l）的宽度 p_j 有一个正的常数上界 ρ，因此

$$c(p_j) \le c(\rho), \quad j = 1, 2, \ldots, l.$$

因此，

$$\mathcal{T}(\tilde{\mathbf{f}}_l) \le \sum_{j=1}^{l} c(\rho) = c(\rho) l,$$

这就证明了该结果。

定理6中的估计（49）得到了下一节所给数值示例的验证。

6 数值实验

本节给出两个数值示例，以说明所提出算法的有效性。第一个示例评估其学习具有多个振荡尺度的解的能力，第二个示例则考察其对不仅包含多个振荡尺度、还具有奇异性的解的处理效果。在这两个示例中，解所涉及的最高波数均为 $\kappa = 500$。对于每个示例，我们都使用等宽网络和变宽网络评估性能。

文献 [9] 对基于深度神经网络的方法与传统分片多项式配置法在求解振荡积分方程方面进行了广泛比较。结果表明，MGDL 始终优于分片线性配置法和二次多项式配置法，而这两种方法是数值分析中的标准工具。因此，我们不再重复这一比较。本文的数值实验转而侧重于将 MGDL 与单级深度学习（SGDL）模型进行比较，并验证此前建立的理论结果。

所有实验均在 Ubuntu Server 18.04 LTS（64 位）系统上进行，设备配备 Intel Xeon Platinum 8255C CPU@ 2.5GHz 和 NVIDIA A4000 GPU。

我们比较 SGDL 模型和 AMGDL 模型在给定右端项下求解弗雷德霍姆积分方程 (2) 的性能。比较基于相对 L_2 误差，其定义为

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

其中， y 和 $\tilde{y}$ 分别表示精确解和近似解。这里，我们设定 $l := 20{,}000$，并将评估点定义为 $s_j := -1 + \frac{2j}{l}$，其中$j \in \mathbb{Z}_{l+1}$。

SGDLand 和 AMGDL 模型分别按照第 2 节和第 5 节所述的步骤构建。对于 SGDL 模型，训练损失定义为

$$\frac{1}{N_\kappa} \sum_{l=1}^{N_\kappa} \left| f(x_l) - ((\mathcal{I} - \mathcal{K}_{p_\kappa}) \mathcal{T} \mathcal{N}_n(\{\mathbf{W}_j, \mathbf{b}_j\}_{j=1}^n); \cdot)(x_l) \right|^2,$$

其中， x_l 为稍后指定的训练数据点。

为评估泛化性能，所有模型均采用如下定义的验证损失进行评估：

$$\frac{1}{2048} \sum_{l=1}^{2048} \left| f(x'_l) - (\mathcal{I} - \mathcal{K}_{p_\kappa}) Y(x'_l) \right|^2,$$

其中，对于 SGL， $Y := \tilde{y}^*$ ；对于 AMGDL， $Y := \tilde{y}^*_L$ ； x'_l 是一组独立选取的验证点。在算法1的实现中，该验证损失取代训练误差，作为停止准则。唯一的区别在于评估点，这有助于减少过拟合，并提高自适应选择过程的鲁棒性。

现在介绍用于生成训练数据集和验证数据集的步骤。

训练数据： 我们通过在区间 I 中选取 N_κ 个等距点 $\{x_j\}_{j=1}^{N_\kappa}$ 来生成训练数据集。对于每个点 x_j，我们计算相应的函数值 $f(x_j)$，从而得到训练数据集

$$\{(x_j, f(x_j))\}_{j=1}^{N_\kappa} \subset I \times \mathbb{C}.$$

验证数据： 验证数据集由区间 I 上均匀分布的 2,048 个点 $\{x'_j\}_{j=1}^{2048}$ 组成。对于每个 x'_j，我们计算 $f(x'_j)$，从而得到验证数据集

$$\{(x'_j, f(x'_j))\}_{j=1}^{2048} \subset I \times \mathbb{C}.$$

在所有实验中，SGDL 和 AMGDL 模型的隐藏层所使用的激活函数均统一选择为正弦函数。

示例 1： 本示例旨在评估所提出方法学习包含多个振荡尺度的解的有效性，其中最高波数为 $\kappa = 500$。

我们考虑核函数为.....的振荡Fredholm积分方程 (2)。

$$K(s,t) = \cos(s(t+1)), \quad s, t \in I$$

以及 $\kappa = 500$ 的振荡Fredholm积分方程 (2)。这里，光滑核 K 不可分离。精确解选为

$$\begin{aligned} y(s) := {} & e^s + (\sin(s) + s + 1)e^{100is} + (\cos(s) + s^3)e^{-150is} + |s|e^{-200is} \\ & + s^3 e^{250is} + \cosh(s)e^{300is} + s^2 e^{-350is} + \sinh(s)e^{400is} \\ & + (s^3 + \sin(s))e^{450is} + |s|^3 e^{-500is}, \quad s \in I. \end{aligned}$$

尽管函数 y 并不严格属于空间 $H^m_\kappa,\,0(I)$——这是由于其各组成部分以不同频率振荡——但这一设计引入了更丰富的多尺度特征，使学习任务更具挑战性。尽管如此，与 $H^m_\kappa,\,0(I)$ 的偏差并不会显著影响模型误差的理论分析。式（2）中的右端函数 f 根据该精确解相应计算得到。

对于深度神经网络的实现，我们根据条件（42）设置了 $\Gamma = 2$、$\gamma = 8$、$\beta = 1$ 和 $q = 1$。因此，我们得到 $p_\kappa :=\lceil \gamma\kappa^\beta \rceil = 4$，000。

AMGDL 算法支持灵活的神经网络设计，从而实现各种架构变体。在这些变体中，我们探讨了两种具有代表性的配置：具有统一宽度的网络和具有可变宽度的网络。**等宽网络：** 在此配置中，我们将最大层级数设为 12，最小层级数设为 3。对于每个层级 $j \in \mathbb{N}_{12}$，神经网络架构定义如下：

$$\text{Grade } j\text{:} \quad [1] \to (j-1) \times [256]_F \to [256] \to [2],$$

其中（$j-1$）$\times[256]_F$表示每层包含 256 个神经元的 $j-1$ 隐藏层，其参数在之前的层级中固定（即不更新）。因此，每个后续层级都会向网络中新增一个宽度为 256 的可训练层，并在前一层级的结构基础上逐步构建。

现在我们介绍 AMGDL 的训练和超参数调优策略。每个模型都使用 Adam 优化器训练 4,000 个训练轮次，采用指数衰减学习率，从选定的初始值开始，逐渐降低至 10^{-7}。初始学习率从$\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$中选择，批次大小从 $\{128, 256, 512\}$中选择。根据三次独立运行的平均最佳验证性能，确定最优超参数组合。依据这一准则，最终选择初始学习率 10^{-2} 和批次大小 128。

各层级的验证误差见表 1，并在图 1 中进行了可视化。如观察所见，验证误差从等级1到六年级单调递减，但在七年级时有所增加。因此，根据自适应算法的停止准则，训练在六年级时终止。

	训练损失	验证损失	RE	时间（秒）
等级1	3.73e+0	3.26e+0	8.86e-1	371
等级2	1.51e+0	1.70e+0	6.53e-1	397
三年级	2.15e-7	2.77e-7	2.70e-4	389
四年级	5.88e-9	8.24e-9	8.91e-5	394
五年级	3.04e-9	6.45e-9	**8.67e-5**	392
六年级	2.88e-9	**6.39e-9**	8.69e-5	406
七年级	2.84e-9	6.84e-9	8.96e-5	418

表 1 采用等宽网络的 AMGDL 在最优超参数下不同层级的性能指标。

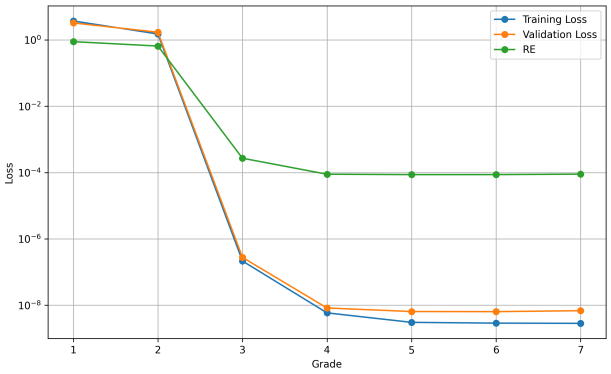


图1 采用等宽网络的 AMGDL 的训练损失和验证损失。

图2展示了从频域角度观察时，等宽的 AMGDL 模型如何逐层提升近似效果。

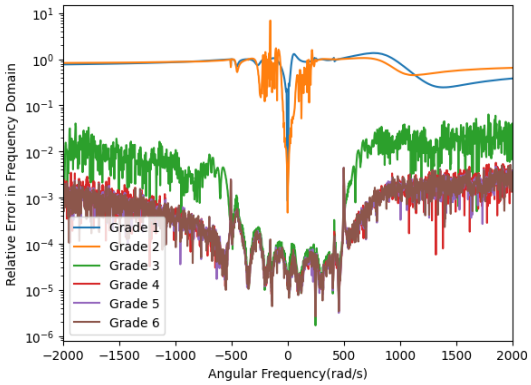


图1 采用等宽网络的 AMGDL 在 1–7 年级的频域相对误差。

我们比较 AMGDL 模型与 SGDL 模型的性能。对于每个 $j \in \mathbb{N}_{10}$，SGDL-j 模型采用的网络架构如下：

$$[1] \to j \times [256] \to [2],$$

这对应于 AMGDL 模型中宽度相同的层级-j 网络。

训练时，每个 SGDL 模型均以指数衰减的学习率训练 4,000 轮，学习率从初始值逐渐降至 10^{-7}。我们测试了集合 $\{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$ 中的初始学习率，以及 $\{128, 256, 512\}$ 中的批次大小。根据五次独立运行中取得的最低验证误差，选出最佳超参数组合。SGDL 模型的最优超参数及对应误差汇总于表 2。

模型	批次大小	初始学习率	训练损失	验证损失	RE
SGDL-1	512	1e-2	3.24e-0	3.25e-0	8.86e-1
SGDL-2	128	1e-2	1.42e-0	1.50e-0	6.12e-1
SGDL-3	256	1e-2	1.14e-6	1.43e-6	5.85e-4
SGDL-4	256	1e-2	4.77e-7	**4.39e-7**	**3.37e-4**
SGDL-5	512	1e-2	8.19e-7	7.07e-7	4.40e-4
SGDL-6	128	1e-3	1.27e-6	1.48e-6	6.30e-4
SGDL-7	128	1e-3	8.02e-7	7.38e-7	4.51e-4
SGDL-8	128	1e-3	4.34e-7	4.86e-7	3.64e-4
SGDL-9	128	1e-3	4.13e-7	8.00e-7	5.46e-4
SGDL-10	256	1e-3	7.65e-7	7.39e-7	4.57e-4
SGDL-11	256	1e-3	4.81e-7	6.87e-7	5.17e-4
SGDL-12	256	1e-3	5.46e-7	1.00e-6	7.11e-4

表 2 采用等宽网络的 SGDL 模型的最优超参数及其对应误差。

变宽网络： 现在展示使用变宽网络的 AMGDL 算法示例。与等宽情形相同，我们将层级数的最大值设为 12，最小值设为 3。层级 j 的网络架构定义如下：

$$\text{Grade } j: \ [1] \to [a_1]_F \to [a_2]_F \to \cdots \to [a_{j-1}]_F \to [a_j] \to [2], \ \text{for } j \in \mathbb{N}_{12},$$

其中 $a_j := 200 + 100\lceil j/2 \rceil$，对于 $j \in \mathbb{N}_{12}$。这里，$[a_k]_F$ 表示宽度为 a_k 的层从先前的层级固定下来（即冻结）。因此，在每个后续层级中，都会在先前固定的层之上添加一个宽度增加的新层。

变化宽度 AMGDL 模型的训练与超参数调优策略与等宽情形相同。评估多种配置后，发现最优参数为批次大小 256，初始学习率为 10^{-2}。各层级的验证误差列于表3，并在图3中以可视化形式呈现。如观察所示，验证误差从等级1到六年级单调递减，但在七年级时有所上升。因此，算法在六年级处终止。

	训练损失	验证损失	RE	时间（秒）
等级1	2.99e+0	3.26e+0	8.86e-1	176
等级2	2.23e+0	1.98e+0	7.03e-1	182
三年级	1.11e-5	1.09e-5	1.64e-3	206
四年级	9.45e-9	1.21e-8	8.76e-5	210
五年级	6.58e-9	7.58e-9	7.84e-5	217
六年级	2.70e-9	**4.63e-9**	**7.48e-5**	243
七年级	2.90e-9	1.87e-7	3.26e-4	296

表3 使用等宽网络和最优超参数时，AMGDL 在不同层级下的验证误差。

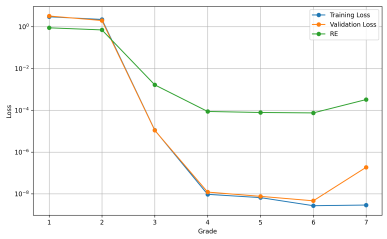


图3 使用变宽网络时 AMGDL 的训练损失和验证损失。

接下来，我们说明 MGDL 模型如何在频域中逐层提升近似质量。设 $(\mathcal{F}v)(z)$ 表示向量 v 的快速傅里叶变换（FFT），其中 z 为角频率。为评估近似解 Y 的精度，我们计算其在每个频率 $z \in \{\pi j - 10001\pi$ 处的相对误差 $j \in \mathbb{N}_{20001}\}$，如下所示:

$$\frac{\left|\mathcal{F}([y(s_j)]_{j=1}^{20001})(z) - \mathcal{F}([Y(s_j)]_{j=1}^{20001})(z)\right|}{\left|\mathcal{F}([y(s_j)]_{j=1}^{20001})(z)\right|},$$

其中， y 为精确解， $s_j := -1 + {}^{2}{}_{(j-1)\ 20000}$ 表示对于 $j \in \mathbb{N}_{20001}$， $Y = \tilde{y}_l^*$ 表示在 L层级 AMGDL 模型的第 l 层级获得的近似解。

图4从频域的角度展示了不同宽度的 AMGDL 算法如何逐层级改善近似效果。

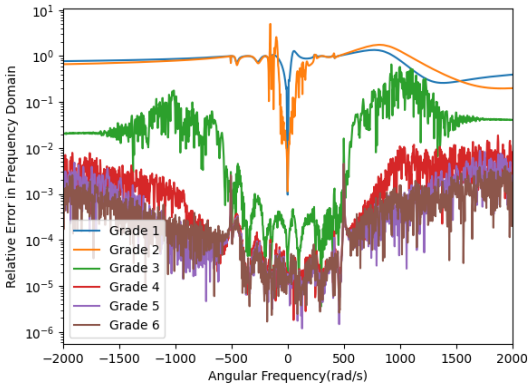


图4 变宽网络下 AMGDL 的频域相对误差：层级 1–8。

我们还比较了 AMGDL 模型和 SGDL 模型的性能。对于每个 $j \in \mathbb{N}_{10}$, SGDL-j 模型采用如下网络架构:

$$[1] \to [a_1] \to [a_2] \to \cdots \to [a_j] \to [2],$$

其中宽度为 $a_k := 200 + 100\lceil k/2 \rceil$, 对于 $k = 1$, ..., j, 与变宽 AMGDL 模型所使用的层宽度相匹配。

训练和超参数调优策略与等宽SGDL模型所采用的策略相同。SGDL模型的最优超参数及相应误差汇总于表4中。

模型	批次大小	初始学习率	训练损失	验证损失	RE
SGDL-1	512	1e-2	3.29e-0	3.25e-0	8.86e-1
SGDL-2	256	1e-1	7.59e-6	7.52e-6	1.37e-3
SGDL-3	256	1e-2	4.73e-7	5.87e-7	3.73e-4
SGDL-4	512	1e-2	6.60e-7	6.25e-7	4.15e-4
SGDL-5	512	1e-2	4.72e-7	6.00e-7	4.00e-4
SGDL-6	128	1e-3	7.66e-7	5.23e-7	3.89e-4
SGDL-7	128	1e-3	2.98e-7	**2.27e-7**	**2.64e-4**
SGDL-8	256	1e-3	5.01e-7	4.32e-7	3.47e-4
SGDL-9	256	1e-3	2.78e-7	2.67e-7	2.99e-4
SGDL-10	256	1e-3	1.50e-7	3.17e-7	3.87e-4
SGDL-11	512	1e-3	6.39e-7	5.67e-7	3.96e-4
SGDL-12	512	1e-3	3.41e-7	3.46e-7	3.12e-4

表4 使用变宽网络的 SGDL 模型的最优超参数及对应误差。

数值实验验证了前几节建立的理论结果。具体而言，图1和图3表明，在数值求积误差可忽略的前提下，如定理 4 所述，相对误差在一个常数因子的范围内受训练误差控制。此外，最后

几列的表1和表3证实，计算时间随层级数量线性增长，这与定理6一致。

最后，我们在表5中汇总了本次实验中所有模型的相对误差。如表中所示，无论是等宽网络架构还是变宽网络架构，AMGDL 的精度始终比 SGDL 高约一个数量级。此外，在 AMGDL 和 SGDL 两种框架中，变宽网络的表现都略优于相应的等宽网络。在评估的所有模型中，我们提出的 V-AMGDL算法误差最低，整体表现最佳。

模型	E-AMGDL	E-SGL	V-AMGDL	V-SGL
RE	8.69e-5	3.37e-4	**7.48e-5**	2.64e-4

表5 不同模型的相对误差（相对误差）汇总。“E-”和“V-”分别表示具有等宽和变宽网络架构的模型。

示例2： 本示例旨在评估 AMGDL 算法学习具有多重振荡尺度、包含奇异性且最高波数为 $\kappa = 500$ 的解时的性能。

我们考虑具有常核 $K(s,\ t) = 0.45,\ s,\ t \in I$ 和 $\kappa = 500$ 的振荡 Fredholm积分方程 (2)。这里选择常核，是为了与前一个使用不可分核的示例形成对比。精确解为

$$y(s) := \operatorname{sign}(s)\,|s|^{\epsilon}\ln(|s|)\big[\sin(|s|) + s^3 e^{100is} + \cosh(s)e^{300is} + s^2 e^{-350is} + \sinh(s)e^{400is} + (s^3 + \sin(s))e^{450is} + |s|^3 e^{-500is}\big], \quad s \in I \setminus \{0\}, \tag{50}$$

其中 $\epsilon > 0$，且对于 $s > 0$，$\text{sign}(s) := 1$；对于 $s < 0$，$\text{sign}(s) : -1$。为保证连续性，我们定义 $y(0) := 0$，使得 y 在区间 I 上为连续函数。随后据此计算积分方程 (2) 对应的右端项 f。

在以下实验中，我们考虑

$$\epsilon \in \{1, 0.8, 0.6, 0.4, 0.2, 0.1\}.$$

随着 ϵ 减小，解变得越来越奇异，数值近似也面临更大的挑战。对于基于深度神经网络的方法，我们按照条件 (42) 设置 $\Gamma = 2$、$\gamma = 10$、$\beta = 1$ 和 $q = 1$。我们还固定

$$p_{500} := \lceil \gamma \cdot 500^{\beta} \rceil = 5,000.$$

对于采用等宽网络的 AMGDL 模型及相应的 SGDL 模型，不同 ϵ 值对应的相对误差分别列于表 6 和表 7 中。结果也在图 5 中进行了可视化。为保持

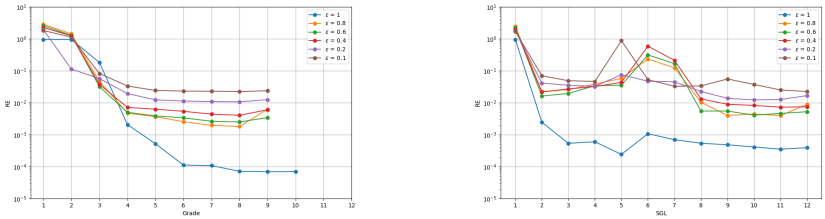
	$\epsilon = 1$	$\epsilon = 0.8$	$\epsilon = 0.6$	$\epsilon = 0.4$	$\epsilon = 0.2$	$\epsilon = 0.1$
等级1	9.56e-1	2.87e-0	2.58e-0	2.26e-0	1.94e-0	1.82e-0
等级2	9.55e-1	1.42e-0	1.28e-0	1.24e-0	1.13e-1	1.13e-0
三年级	1.82e-1	4.69e-2	3.20e-2	3.77e-2	5.71e-2	8.22e-2
四年级	2.08e-3	4.73e-3	5.01e-3	7.14e-3	1.94e-2	3.34e-2
五年级	5.21e-4	3.66e-3	3.89e-3	6.29e-3	1.23e-2	2.43e-2
六年级	1.12e-4	2.55e-3	3.36e-3	5.39e-3	1.13e-2	2.32e-2
七年级	1.08e-4	1.96e-3	2.64e-3	4.41e-3	1.09e-2	2.29e-2
八年级	7.17e-5	**1.80e-3**	**2.50e-3**	**4.05e-3**	**1.07e-2**	**2.23e-2**
九年级	**6.92e-5**	6.28e-3	3.43e-3	6.02e-3	1.25e-2	2.38e-2
10年级	6.98e-5	-	-	-	-	-
最佳相对误差	6.92e-5	1.80e-3	2.50e-3	4.05e-3	1.07e-2	2.23e-2

表6 AMGDL 使用等宽网络在不同层级和奇异性级别。

表述简洁，本例仅报告相对误差，受篇幅限制省略训练损失和验证损失。

	$\epsilon = 1$	$\epsilon = 0.8$	$\epsilon = 0.6$	$\epsilon = 0.4$	$\epsilon = 0.2$	$\epsilon = 0.1$
SGDL-1	9.56e-1	2.47e-0	2.23e-0	1.98e-0	1.78e-0	1.71e-0
SGDL-2	2.46e-3	2.24e-2	1.62e-2	2.17e-2	4.11e-2	6.99e-2
SGDL-3	5.38e-4	2.58e-2	1.95e-2	2.70e-2	3.52e-2	4.89e-2
SGDL-4	6.01e-4	3.83e-2	3.42e-2	3.22e-2	3.27e-2	4.65e-2
SGDL-5	2.42e-4	5.79e-2	3.55e-2	4.33e-2	7.52e-2	8.78e-1
SGDL-6	1.07e-3	2.34e-1	3.14e-1	5.90e-1	4.73e-2	5.26e-2
SGDL-7	7.00e-4	1.24e-1	1.68e-1	2.12e-1	4.51e-2	3.27e-2
SGDL-8	5.40e-4	1.04e-2	5.51e-3	1.32e-2	2.26e-2	3.36e-2
SGDL-9	4.84e-4	3.99e-3	5.50e-3	8.96e-3	1.37e-2	5.55e-2
SGDL-10	4.12e-4	4.44e-3	**4.10e-3**	8.25e-3	**1.23e-2**	3.75e-2
SGDL-11	**3.53e-4**	**3.98e-3**	4.63e-3	**7.28e-3**	1.26e-2	2.51e-2
SGDL-12	3.93e-4	8.85e-3	5.23e-3	7.43e-3	1.67e-2	**2.25e-2**
最佳相对误差	3.53e-4	3.98e-3	4.10e-3	7.28e-3	1.23e-2	2.25e-2

表7 使用等宽网络的 SGDL模型在不同奇异性级别下的相对误差。



（a）AMGDL (b) SGDL

图5 不同奇异性级别下，使用等宽网络的 AMGDL 和 SGDL模型的相对误差。

对于变宽网络，AMGDL 及其对应 SGDL 模型在不同奇异性层级下的相对误差 ϵ 分别列于表8和表9中，并在图6中进行了可视化。

	$\epsilon = 1$	$\epsilon = 0.8$	$\epsilon = 0.6$	$\epsilon = 0.4$	$\epsilon = 0.2$	$\epsilon = 0.1$
等级1	9.56e-1	2.87e-0	2.58e-0	2.26e-0	1.95e-0	1.82e-0
等级2	9.47e-1	1.35e-0	1.27e-0	1.15e-0	1.11e-0	1.13e-0
三年级	9.59e-3	1.73e-2	1.25e-2	1.73e-2	4.30e-2	1.10e-1
四年级	2.82e-4	4.75e-3	3.72e-3	6.05e-3	1.75e-2	4.30e-2
五年级	1.08e-4	2.92e-3	2.74e-3	4.86e-3	1.37e-2	3.00e-2
六年级	4.76e-5	2.10e-3	2.27e-3	3.89e-3	1.15e-2	2.40e-2
七年级	**3.62e-5**	1.41e-3	2.09e-3	3.44e-3	1.14e-2	**2.20e-2**
八年级	4.35e-5	**1.03e-3**	1.76e-3	3.14e-3	**1.13e-2**	3.06-2
九年级	-	1.93e-2	**1.48e-3**	2.87e-3	1.25e-2	-
10年级	-	-	6.20e-2	2.60e-3	-	-
11年级	-	-	-	**2.32e-3**	-	-
12年级	-	-	-	1.53e-2	-	-
Best	3.62e-5	1.27e-3	1.48e-3	2.32e-3	1.13e-2	2.20e-2

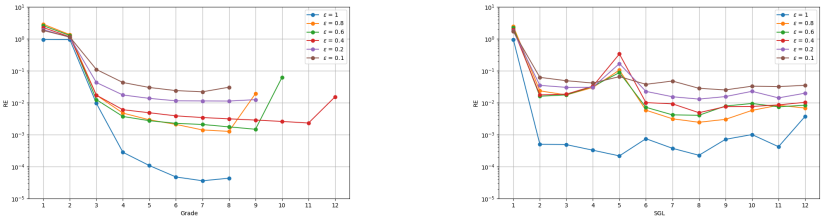
表8 不同层级和奇异性级别下，采用变宽网络的 AMGDL 的相对误差。

	$\epsilon = 1$	$\epsilon = 0.8$	$\epsilon = 0.6$	$\epsilon = 0.4$	$\epsilon = 0.2$	$\epsilon = 0.1$
SGDL-1	9.56e-1	2.47e-0	2.25e-0	1.95e-0	1.75e-0	1.71e-0
SGDL-2	5.03e-4	2.36e-2	1.60e-2	1.76e-2	3.50e-2	6.23e-2
SGDL-3	4.89e-4	1.75e-2	1.76e-2	1.86e-2	3.03e-2	4.89e-2
SGDL-4	3.27e-4	3.01e-2	3.25e-2	3.30e-2	3.05e-2	4.14e-2
SGDL-5	**2.17e-4**	1.05e-1	8.79e-2	3.36e-1	1.68e-1	6.57e-2
SGDL-6	7.50e-4	5.85e-3	7.20e-3	1.01e-2	2.27e-2	3.77e-2
SGDL-7	3.75e-4	3.15e-3	4.18e-3	9.25e-3	1.54e-2	4.78e-2
SGDL-8	2.27e-4	**2.45e-3**	**4.05e-3**	**4.88e-3**	**1.30e-2**	2.83e-2
SGDL-9	7.16e-4	3.02e-3	7.90e-3	7.58e-3	1.56e-2	**2.52e-2**
SGDL-10	1.01e-3	5.77e-3	9.49e-3	7.64e-3	2.28e-2	3.30e-2
SGDL-11	4.21e-4	8.36e-3	7.47e-3	8.57e-3	1.42e-2	3.19e-2
SGDL-12	3.72e-3	6.83e-3	8.36e-3	1.03e-2	1.99e-2	3.52e-2
Best	2.17e-4	2.45e-3	4.05e-3	4.88e-3	1.30e-2	2.52e-2

表9 使用变宽网络的 SGDL 模型在不同奇异性级别下的相对误差。

表6和表8均表明，所提出的算法在不同程度的奇异性下都能保持稳健的自适应性。其一大优势是，无需预先设定神经网络的深度；停止准则会自动确定网络深度。这克服了传统深度神经网络的一项常见局限：对于给定的解，通常无法事先知道最佳网络深度。

最后，我们在表10中总结了本实验评估的所有模型的相对误差。如表中所示，所有深度学习模型——包括 SGDL 和 AMGDL——在面对各种奇异性时都能将误差控制在较低水平。然而，AMGDL 模型始终优于其单等级对应模型，尤其是在奇异性增强（即 ϵ 值减小）时。



(a) AMGDL (b) SGDL

图6 AMGDL 和 SGDL 模型采用变宽网络时在不同奇异性级别下的相对误差。

此外，采用变宽网络的 AMGDL 模型误差略低于采用等宽架构的模型，表明其对解的奇异性具有更强的鲁棒性。总体而言，所提出的 V-AMGDL 方法性能最佳，尤其适用于高度奇异问题。

本节的数值结果表明，无论光滑核 K 是可分离还是不可分离，AMGDL 模型在等宽网络和变宽网络上，以及对于具有不同类型奇异性的解，均始终优于 SGDL 模型。此外，在 AMGDL框架中，变宽网络的性能优于相应的等宽网络。

模型	E-AMGDL	E-SGL（最佳）	V-AMGDL	V-SGL（最佳）
$\epsilon = 1$	6.92e-5	3.53e-4	**3.62e-5**	2.17e-4
$\epsilon = 0.8$	1.80e-3	3.98e-3	**1.03e-3**	2.45e-3
$\epsilon = 0.6$	2.50e-3	4.10e-3	**1.48e-3**	4.05e-3
$\epsilon = 0.4$	4.05e-3	7.28e-3	**2.32e-3**	4.88e-3
$\epsilon = 0.2$	**1.07e-2**	1.23e-2	1.13e-2	1.30e-2
$\epsilon = 0.1$	2.23e-2	2.25e-2	**2.20e-2**	2.52e-2

表10 不同 ϵ 取值下各模型相对误差（RE）的汇总。“E-”和“V-”分别表示采用等宽网络和变宽网络的模型。“(最佳)”表示 12 个 SGDL 模型中的最佳表现。

7 结论

本文针对求解高度振荡的弗雷德霍姆积分方程，系统开展了多级深度学习（MGDL）的理论与算法研究。尽管已有研究已将 MGDL 经验性地应用于此类问题，但本文首次针对这一背景下的连续 MGDL 模型和离散 MGDL 模型提供了严格的误差分析。一个关键的理论贡献是证明了：在数值求积足够精确的条件下，离散模型继承了连续模型的收敛性与稳定性。

基于上述分析，我们开发了一种新颖的自适应 MGDL 算法，可根据训练误差动态选择网络等级的数量——

这一自适应性在理论上得到了证明：我们的结果表明，近似误差的主要来源是深度神经网络的训练过程，而非离散化或数值求积。所提出的方法逐级递增地构建网络，无需预先指定网络深度——这正是标准深度学习方法的一项显著局限。

数值实验证实了自适应 MGDL 算法的有效性和鲁棒性，尤其是在涉及高频振荡和奇异解的复杂情形下。与传统的单层级网络相比，自适应 MGDL 在精度和泛化能力方面表现更优。

总而言之，本研究提出了一种具有坚实理论基础、可扩展且自适应的算子方程深度学习框架，展现了 MGDL 在处理多尺度和振荡现象方面的独特优势。

致谢： Y.Xu 是雪城大学数学系荣休教授。所有通信请联系 Y. Xu。

经费声明： Y.Xu 获得美国国家科学基金会项目 DMS-2208386 的部分资助。

数据可用性： 如有合理需求，可提供数据。

声明： 作者声明不存在利益冲突。

参考文献

1. Anselone, P.M.：《集体紧算子逼近理论及其在积分方程中的应用》。Prentice-Hall，恩格尔伍德克利夫斯，新泽西州（1971） 2. Atkinson, K.E.：《第二类积分方程的数值解》。剑桥大学出版社，剑桥（1997） 3. Brunner, H., Iserles, A., Nørsett, S.P.：一类高振荡弗雷德霍姆积分算子的谱问题。IMA数值分析杂志**30**(1), 108–130（2010） 4. Brunner, H., Ma, Y., Xu, Y.：高振荡核的沃尔泰拉积分方程与积分微分方程解的振荡。《积分方程与应用杂志》**27**(4), 455–487（2015） 5. Chen, Z., Micchelli, C.A., Xu, Y.：弗雷德霍姆积分方程的多尺度方法，第28卷。剑桥大学出版社，剑桥（2015） 6. Colton, D.L., Kress, R.：《声学与电磁反散射理论》，第93卷。施普林格，纽约（1998） 7. Fang, R., Xu, Y.：通过多级深度学习解决深度神经网络的频谱偏差问题。神经信息处理系统进展**37**, 114122–114146（2024） 8. Fang, R., Xu, Y.：多级深度学习的计算优势：收敛性分析与性能洞察。arXiv预印本 arXiv:2507.20351（2025） 9. Jiang, J., Xu, Y.：振荡弗雷德霍姆积分方程的深度神经网络解。《积分方程与应用杂志》**36**(1), 23–55（2024） 10. Pilanci, M., Ergen, T.：神经网络是凸正则项：两层网络的精确多项式时间凸优化形式。国际机器学习大会，第7695–7705页（PMLR，2020）

11. Rahaman, N., Baratin, A., Arpit, D., Draxler, F., Lin, M., Hamprecht, F.A., Bengio, Y., Courville, A.：论神经网络的频谱偏差。第36届国际机器学习大会论文集，长滩，加利福尼亚州，PMLR 97（2019） 12. Raissi, M.：深度隐藏物理模型：非线性偏微分方程的深度学习。《机器学习研究杂志》**19**(1), 932–955（2018） 13. Raissi, M., Perdikaris, P., Karniadakis, G.E.：物理信息神经网络：用于求解涉及非线性偏微分方程的正问题与反问题的深度学习框架。《计算物理学杂志》**378**, 686–707（2019） 14. Wang, Y., Xu, Y.：振荡核第二类弗雷德霍姆积分方程的保振荡伽辽金方法。arXiv预印本 arXiv:1507.01156（2015） 15. Xiang, S., Zhang, Q.：高振荡弱奇异核弗雷德霍姆积分方程的渐近分析。《应用数学与计算》**456**, 128112（2023） 16. Xu, Y.：多级深度学习。《应用数学与计算通讯》，第1–52页（2025） 17. Xu, Y.：深度神经网络的逐次仿射学习。《分析与应用》**23**, 1425–1467（2025） 18. Xu, Y., Zeng, T.：用于偏微分方程的多级深度学习及其在伯格斯方程中的应用。arXiv预印本 arXiv:2309.07401（2023） 19. Xu, Y., Zhang, H.：深度卷积神经网络的收敛性。《神经网络》**153**, 553–563（2022） 20. Xu, Y., Zhang, H.：深度ReLU网络的收敛性。《神经计算》**571**, 127174（2024） 21. Yu, B.等：深度Ritz方法：一种用于求解变分问题的深度学习数值算法。《数学与统计学通讯》**6**(1), 1–12（2018）