arXiv:2603.00904v1 [math.NA] 1 Mar 2026

从频率偏差到谱平衡：物理信息神经网络的算子感知预条件器

Roy Y. He * Ying Liang † Hongkai Zhao ‡ Yimin Zhong §

2026年3月3日

摘要

当神经网络（NN）被用作求解偏微分方程（PDE）的非线性参数化表示时，其学习动力学往往呈现出与直接函数逼近任务不同的频率依赖特征。这是神经网络表示的频率偏差与底层微分算子的频率偏差相互平衡的结果。尽管许多常用神经网络在表示上偏向低频模态，但损失函数中的微分算子会放大高频分量，从而可能导致高频偏差。本文以二阶椭圆型偏微分方程为例，说明这两种因素如何相互竞争，并在不同情形下造成总体频率偏差。确定这种平衡后，设计计算策略来抵消由此产生的偏差，对于提高训练效率至关重要。我们提出一种简单的算子感知预条件策略：对残差应用辅助积分算子，以重新平衡优化景观和学习动力学。积分核函数可以是参考椭圆算子的格林函数或其近似，并且易于与常见的偏微分方程神经网络求解器集成。包括多尺度问题和变系数问题在内的大量实验表明，该方法能够恢复不同模态间更加平衡的学习动力学，并显著提升收敛性和准确度。

关键词。 频率偏差、预条件子、物理信息神经网络

1 引言

神经网络（NN）为近似函数提供了一种特殊的非线性参数化表示。关于神经网络，数学近似理论已取得了大量成果。其中一个主要主题是证明神经网络的通用逼近性质（UAP） [Cyb89，Bar02，DHP21，KL20，LPW+17，ES16，SYZ22]。然而，现有的数学结果并未解决以下关键问题：通过基于梯度的技术求解大规模非凸优化问题时，解的计算精度能达到多高、成本有多大。此类技术对参数化表示的条件性十分敏感，在实际应用中往往会成为计算瓶颈。

神经网络表示的一个常见特征是低频偏置。频率偏差指的是网络在训练过程中学习目标函数的低频分量比

*香港城市大学数学系，香港九龙塘
†美国北卡罗来纳州达勒姆杜克大学数学系
‡美国北卡罗来纳州达勒姆杜克大学数学系§美国阿拉巴马州奥本奥本大学数学与统计系

高频分量更快的倾向，这使得网络难以捕捉目标函数中的高频分量和精细特征，计算成本也更高。这一现 [RBA+19]象最早由相关研究系统地展示：研究者分析了神经网络输出的傅里叶谱，并指出，即使没有显式正则化，采用 ReLU 或双曲正切激活函数的标准架构也会优先学习较平滑的低频特征。频率偏差可以从神经正切核（NTK）理论的角度来理解：训练动力学在 NTK 的特征子空间上分解，每个子空间对应不同的核特征值 [JGH18，CFW+21]。与高频特征函数对应的 NTK 特征值会随频率呈多项式衰减，导致高频分量学习得更慢。尽管这种隐式正则化有助于提高平滑性，但当精细细节或多尺度特征对于准确近似至关重要时，它就会成为问题。一项明确的理论与数值分析 [ZZZZ25c] 表明，由于参数化全局激活函数之间存在很强的相关性，双层神经网络在表征和训练过程中都表现出病态性和频率偏差。

自然地，神经网络（NN）可用于表示/近似偏微分方程的解。许多研究都探讨了这一方法，包括物理信息神经网络（PINNs）[RPK19
RYK20], 深度 Ritz 方法（DRM） [EY18], 弱对抗网络（WAN） [ZBYZ20],等 [LJP+21，LKA+20，DRMM24]。人们已投入大量精力对这些方法进行理论论证。例如，文献 [SDK20], 证明了 PINNs 对线性二阶椭圆型和抛物型偏微分方程的一致性，而文献 [MM23] 则通过实例分析了 PINNs 在数据同化中的泛化误差。关于 DRM，也有类似的结果 [MZ19，DJL+22、LLW21、MZ22]。这些研究总体得出的结论是：在数据充足且边界条件处理得当的情况下，相关优化问题的极小化解会收敛到偏微分方程的解。

然而，这些理论保证往往难以直接转化为实践。尽管使用神经网络表示求解偏微分方程可能带来显著优势，例如无网格、易于实现、可扩展至高维，以及具备自适应性的潜力，但在实际中，非线性表示必须通过基于梯度的方法进行大规模非凸优化来求解。同样，病态性和频率偏差也可能给数值计算带来问题，尤其是对于解包含多个频率尺度或尖锐特征的偏微分方程。微分算子的存在还引入了额外因素，影响训练过程中的整体病态性和频率偏差。[WWP21] 将偏微分方程求解任务中的谱偏差重新表述为“NTK特征向量偏差”，并且还表明，加入傅里叶特征映射能够重塑网络的谱偏好，从而更好地学习高频分量。在我们早期的工作 [HLZZ25], 中，我们表明，对于两层神经网络，病态性和低频偏置在训练动力学中仍然占据主导地位，抵消了微分算子的相反作用。这一结果导致了效率和准确度方面的数值挑战，相较于有限元方法等具有成熟预条件子的线性表示尤为如此。

在这项工作中，我们将聚焦于多层神经网络。其情况与两层神经网络大不相同。首先，多层神经网络的表示能力和灵活性远胜于两层神经网络。特别是，其低频偏置会受到多种因素的影响，例如网络架构、激活函数、权重初始化，以及优化过程中梯度下降所遵循的轨迹。在近似任务的训练中， [BGG+20] 和 [TSM+20]的研究表明，输入编码方案和网络深度会显著影响训练过程中学到的频率成分。 [HSTX22], 的研究表明，激活函数的选择也起着至关重要的作用，而采用 ReLU 作为激活函数的网络则表现出与有限元理论相关的谱偏差。

替代激活函数（如分段线性B样条）则可以消除这种偏差。近期，[ZZZZ25c]中的理论进展表明，由于Gram矩阵的病态性，两层神经网络表现得像“低通滤波器”；后续研究 [ZZZZ25b、ZZZZ25a]，进一步探讨了这样的结论：采用多分量多层神经网络（MMNN）结构、正弦激活函数，并对第一层进行适当缩放的随机初始化后，神经网络能够有效地通过训练近似高频模态。此外，优化过程的性质（例如锐度的演化 [BD20] 以及穿越损失景观的路径）也会调节模态动力学。

使用多层神经网络表示求解偏微分方程时，损失函数中微分算子的参与（例如物理信息神经网络中的情形）会固有地引入高频放大/偏差，可能从根本上改变整个问题表述的病态性与频率偏差，以及学习动力学。具体而言，我们表明，设计良好的多层神经网络（例如采用正弦激活函数并进行适当初始缩放的傅里叶MMNN）具有灵活性、表达能力且频率偏差较弱，因此椭圆算子的病态性和高频偏差将主导整个问题表述及相应的学习动力学。相比之下，使用有限元方法求解椭圆型偏微分方程时，由于有限元表示不存在频率偏差，所得线性系统的病态性和高频偏差均源自微分算子。因此，使用简单迭代方法求解该线性系统时，低频误差衰减极其缓慢，即使对于光滑解，收敛速度也会很慢。这在实践中无法接受，因此需要设计有效的预条件策略。同样，基于底层微分算子会在问题表述中造成整体频率偏差这一观察，我们提出了一种算子感知预条件策略，相应地重新平衡谱权重和训练动力学。该预条件子是一个基于参考椭圆算子的线性算子，它会改变优化景观，有效地重塑损失函数以平衡各频率分量。该预条件子在计算损失之前变换残差，从而抵消微分算子造成的频率放大。它作为一种谱滤波器，在保留物理约束的同时削弱微分算子诱导的偏差，从而加速收敛并获得准确解。

已有多种方法被提出，用于缓解 PINNs 的病态性和频率偏差。Sobolev范数损失以包含导数信息的范数取代标准平方残差，从而有效地赋予高频分量更大的权重；例如，Sobolev物理信息神经网络 [SJHH21]以及关于 Sobolev 加速的分析 [LBY22]。近期，[DRBMdB23] 从算子条件的角度分析 PINN 训练，并提出显式预条件子，以降低复合神经正切核与偏微分算子的条件数。相关策略包括采用空间变化权函数来平衡损失的加权残差法、在训练过程中自适应地重新加权不同谱模态的学习型预条件子，以及基于残差的注意力机制 [ATSK24] ，后者会动态地强调残差较大的配点。这些策略通常需要额外的计算开销，例如计算高阶导数、求解辅助特征值问题，或进行大量超参数调优。与上述工作不同，我们的方法只需要一个简单参考算子的格林函数，该函数通常具有闭式表达，因此能够以简单、有效且可解释的方式，抵消训练过程中的频率偏差。

我们的贡献包括：

- 系统研究使用神经网络求解椭圆型偏微分方程时，神经网络表示与微分算子之间的频率偏差平衡，并展示微分算子如何

可能占据主导地位，导致不理想的高频偏差，并使训练过程收敛缓慢。

- 一种感知算子的预条件策略，通过在问题表述中重新加权频率模态，实现跨频率模态均衡而有效的学习。
- 全面的数值实验表明，我们的预条件策略能够显著提升效率和准确度，尤其适用于多尺度问题。

本文其余部分安排如下。第2节介绍问题表述，并为偏微分方程求解和函数逼近任务建立符号约定。第3 节详细阐述实验动机，展示频率偏差在函数逼近与偏微分方程求解中的不同表现，以及激活函数、网络深度和初始缩放等因素如何影响谱学习动力学。第4节提出我们感知算子的预条件器，并详细说明其如何改变梯度流以重新平衡频率模态。第5节介绍全面的数值实验，从简单的一维案例逐步扩展到具有挑战性的二维多尺度问题，展示我们预条件方法的有效性。第6节总结本文，并讨论本研究的意义和潜在扩展方向。

2 问题表述

我们以带狄利克雷边界条件的二阶椭圆方程为例。给定具有足够光滑边界的有界区域 $\Omega \subset \mathbb{R}^d$，考虑以下偏微分方程

$$\begin{cases} Lu = f & \text{in } \Omega \\ u = g & \text{on } \partial\Omega \end{cases}. \tag{1}$$

这里， L 是一个具有如下形式的二阶椭圆微分算子

$$Lu = -\sum_{i,j=1}^{d} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u}{\partial x_j}\right) + c(x)u, \tag{2}$$

也就是说，存在一个常数 $\alpha > 0$，使得

$$\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq \alpha|\xi|^2 \quad \forall x \in \Omega, \xi \in \mathbb{R}^d.$$

函数 u 是未知量，而 $f \in L^\infty(\Omega)$ $g \in L^\infty(\partial\Omega)$ 是已知的。

这里的主要思想是使用一个 NN，记作 u_θ，来表示解并学习参数 θ，以尽可能准确地近似 PDE 和边界条件。尽管使用 NN 表示具有一些潜在优势，例如无网格、易于实现、可扩展至高维，以及作为非线性表示所具备的自适应性这一关键潜力，但在实际应用中，这些参数必须通过使用基于梯度的方法进行大规模非凸优化来学习或训练。此外，由于采用了非线性表示，如何恰当地构造优化问题可能较为微妙，例如，对于椭圆型偏微分方程问题（1），应采用强形式还是弱形式，以及如何在区域内部施加 PDE 约束并同时满足边界条件。

在本文中，我们将研究一种最简单且最灵活的方法：使用最小二乘形式，通过平衡参数同时施加 PDE 和边界条件，该方法称为 PINN 模型。损失函数 $\mathcal{L}_{\mathrm{PINN}}(\theta)$ 的形式如下

$$\mathcal{L}_{\mathrm{PINN}}(\theta) = \mathcal{L}_{\mathrm{PINN},r}(\theta) + \lambda_b \mathcal{L}_b(\theta), \tag{3}$$

其中

$$\mathcal{L}_{\mathrm{PINN},r}(\theta) = \frac{1}{2}\int_\Omega |Lu_\theta(x) - f(x)|^2 dx = \frac{1}{2}\int_\Omega |L[u_\theta(x) - u(x)]|^2 dx, \tag{4}$$

$$\mathcal{L}_b(\theta) = \frac{1}{2}\int_{\partial\Omega} |u_\theta(x) - g(x)|^2 ds(x) = \frac{1}{2}\int_{\partial\Omega} |u_\theta(x) - u(x)|^2 ds(x). \tag{5}$$

损失函数由偏微分方程残差项$\mathcal{L}_{\mathrm{PINN}、\ r}$ 以及边界拟合项 $\mathcal{L}_b$组成，这等价于 H^2所定义的逼近误差的 L-范数平方，其中2见式（2）。这里 $\lambda_b > 0$ 是一个用于平衡两个损失分量的加权超参数。

作为对比，本研究中使用神经网络表示进行直接函数逼近时，最常用的最小二乘形式为

$$\mathcal{L}_{\mathrm{approx}}(\theta) = \frac{1}{2}\int_\Omega |u_\theta(x) - u(x)|^2 dx, \tag{6}$$

即逼近误差的 L^2 范数平方。

实际中，这些积分通过离散采样进行近似。例如，内部点$\{x_i^r\}_{i=1}^{N_r} \subset \Omega$ 和边界点 $\{x_i^b\}_{i=1}^{N_b} \subset \partial\Omega$ 分别从定义域及其边界上的均匀分布中随机抽取。随后在这些采样点处计算损失，以近似连续积分。优化通常采用基于梯度的方法，例如 Adam [KB15]或 L-BFGS，以求得

$$\theta^* = \arg\min_{\theta\in\mathbb{R}^p} \mathcal{L}_{\mathrm{F}}(\theta), \tag{7}$$

对于 F∈ {物理信息神经网络，近似}。

3 训练动力学

确定优化形式后，最具挑战性的计算任务是训练参数，使其达到令人满意的结果。为了在高维参数空间中求解这一大规模、非凸且缺乏明确解释或特殊结构的优化问题，训练过程通常依赖基于梯度的优化方法。一个关键问题是理解优化问题中潜在的病态性，因为这会导致基于梯度的方法效率低下并产生偏差。

我们从使用连续形式的梯度流开始。对于函数逼近问题（6），参数空间中的梯度流为

$$\frac{d\theta}{dt} = -\int_\Omega \frac{\partial u_\theta}{\partial\theta}(u_\theta - u)dx, \tag{8}$$

$$\frac{d\mathcal{L}_{\mathrm{approx}}(\theta)}{dt} = -\left\|\int_\Omega \frac{\partial u_\theta}{\partial\theta}(u_\theta - u)dx\right\|^2. \tag{9}$$

对于使用两层神经网络进行函数逼近，文献 [ZZZZ25c] 表明，参数化激活函数集合之间的强相关性会导致神经网络表示的病态性，并在训练和收敛过程中产生低频基。对于多层神经网络，由于其答案取决于更复杂的复合结构、激活函数的选择以及参数的初始缩放，因此情况更难以进行明确刻画和分析。例如，文献 [ZZZZ25b] 提出了一种使用正弦激活函数的多组件多层神经网络结构，称为 FMMNN [ZZZZ25a], ；研究表明，在适当的初始缩放下，该结构能够通过平滑复合实现强大的表示能力和灵活性，同时在训练过程中产生弱频率偏置。

在求解偏微分方程（1）时，使用物理信息神经网络表述（3），我们有

$$\frac{d\theta}{dt} = -\int_\Omega [L\frac{\partial u_\theta}{\partial \theta}][L(u_\theta - u)]dx - \lambda_b \int_{\partial\Omega} \frac{\partial u_\theta}{\partial \theta}(u_\theta - u)ds(x),$$
$$\frac{d\mathcal{L}_{\text{PINN}}}{dt} = -\left\| \int_\Omega [L\frac{\partial u_\theta}{\partial \theta}][L(u_\theta - u)]dx + \lambda_b \int_{\partial\Omega} \frac{\partial u_\theta}{\partial \theta}(u_\theta - u)ds(x) \right\|^2.$$

我们看到，微分算子 L 在损失函数（3）中的引入会改变逼近误差中的频率权重，并带来高频偏差（就 H^2 范数而言）。因此，如上所述，它改变了训练动力学。然而，神经网络表示典型的低频偏置与微分算子的高频偏差之间存在微妙的竞争。一个引人深思且重要的问题是二者之间的平衡，以及在使用神经网络表示求解偏微分方程时，训练动力学的整体条件性与频率偏差。更重要的是，如何抵消频率偏差或对表述进行预条件化，对于训练效率和所得准确度都至关重要。

在最近的工作 [HLZZ25], 中表明，当使用两层神经网络求解椭圆型偏微分方程时，神经网络表示的低频偏置仍然占主导地位。因此，在训练过程中，解的低频或光滑分量会首先被捕获，而高频或精细特征的计算则更为昂贵且困难。由此得出结论：由于低频偏置，使用两层神经网络难以实现作为非线性表示的自适应性这一关键潜力，并且无法与有限元方法等发展成熟的线性方法表示相竞争。

本文将重点研究如何使用多层神经网络求解形式为（1）的椭圆型偏微分方程。具体而言，我们表明，对于设计合理、具有较强表示能力且频率偏差较温和的多层神经网络，微分算子的病态性和高频偏差将主导整体问题的构造，给基于梯度的训练过程带来极大困难，使其极其缓慢地捕捉解中的低频和平滑分量，这正如在有限元方法中使用未经预条件的迭代方法一样。为解决这一问题，我们提出了一种简单而有效的 **算子感知预条件** 策略，可显著提升效率和准确度。

下面，我们首先通过数值实验展示不同设置下的频率偏差行为。对于每个示例，我们跟踪训练迭代过程中解误差中不同频率模态系数的演变。随后，我们分析这些频率模态在神经网络表示 u_θ 中的学习动力学，并考察它们对激活函数、网络深度结构、初始缩放以及任务性质（例如函数逼近或偏微分方程求解）等因素的依赖性。本节所呈现的实验侧重于训练早期阶段（前 2,000 次迭代）的频率动力学。我们的目的并非分析平台期，也不是比较不同架构的最终精度，而是展示频率偏差的存在，以及频率偏差如何

影响训练过程中 L^2 误差的下降速度。此外，我们的实验采用过采样数据，也就是说，采样点足以充分解析目标函数。

3.1 函数逼近中的频率偏差

在示例1和示例2中，我们考察了激活函数、初始缩放和深度等不同因素对函数逼近中频率偏差的影响，并证明频率偏差降低通常与收敛性和准确度的提升相关。

示例1：本示例旨在考察激活函数的选择以及第一层斜率参数的初始缩放如何影响函数逼近中的频率偏差。

按照 [ZZZZ25c], ，我们采用首层初始缩放策略，以降低频率偏差并提高学习效率。标准初始化（例如 PyTorch 的默认初始化）后，将第一层激活函数 w 中的斜率权重 $\sigma(w(x-b))$ 按照与 $(n_1)^{\frac{1}{d}}$ 成正比的系数进行缩放，其中 n_1 是第一层的网络宽度， d 是输入维度。该缩放使第一层在神经元数量所能支持的范围内获得最大的频率范围，从而在训练开始时捕捉数据中多样的频率模态。后续层保持标准初始化，以避免层间放大导致的不稳定。

我们首先考虑宽度为 50、深度为 3 的标准全连接神经网络（FCNN），其包含 5,201 个可学习参数，用于近似目标函数

$$u_1(x) = \sin(2\pi x) + \sin(5\pi x) + \sin(9\pi x)$$

在区间 $[-1,1]$ 上的 L^2 范数，其中包含 2,000 个采样点。优化采用 Adam算法和循环学习率调度，学习率在基准值 $\eta_{min} = 1 \times 10^{-5}$ 和最大值 $\eta_{max} = 2 \times 10^{-3}$ 之间振荡。对上下限均应用指数衰减因子$\gamma = 0.999$ ，使整个训练过程中的学习率范围逐渐缩小。

为研究网络的谱学习行为，我们考察训练过程中的不同傅里叶模式误差。在本文中，标记为 m 的频率模态指误差在频率 m 处的傅里叶分量，频率以每单位长度的周期数衡量。形式为 $\sin(k\pi x)$ 的解析分量对应频率模态 $m = k/2$。因此，精确解 u_1 中的目标解分量分别对应频率模态 1、2.5 和 4.5。相应曲线展示了该傅里叶分量的幅值随训练迭代的变化。

结果如图1所示。我们观察到，在没有初始缩放的情况下，误差衰减较慢，且以正弦或双曲正切作为激活函数的表现不如以ReLU作为激活函数。这可以由 [ZZZZ25c] 中针对浅层神经网络的谱分析来解释，即一族随机参数化激活函数的Gram矩阵的谱衰减。在没有缩放时，光滑激活函数的谱衰减呈指数型，而ReLU的谱衰减呈多项式型。然而，通过对斜率参数进行适当缩放，这族随机参数化激活函数提供了更加多样、相关性更低的基，其Gram矩阵在与缩放成比例的一系列频率模态上具有较慢的谱衰减，因而能够捕捉底层数据的更多特征，如 [ZZZZ25c]所示。特别地，在第一层使用经过适当缩放的正弦最为高效，类似于使用一族多样的、相关性最低的全局函数——随机傅里叶基，如 [ZZZZ25a、TYZZ25]所示，本测试中也观察到了这一点。需要注意的是，ReLU及其幂次在缩放下是不变的。

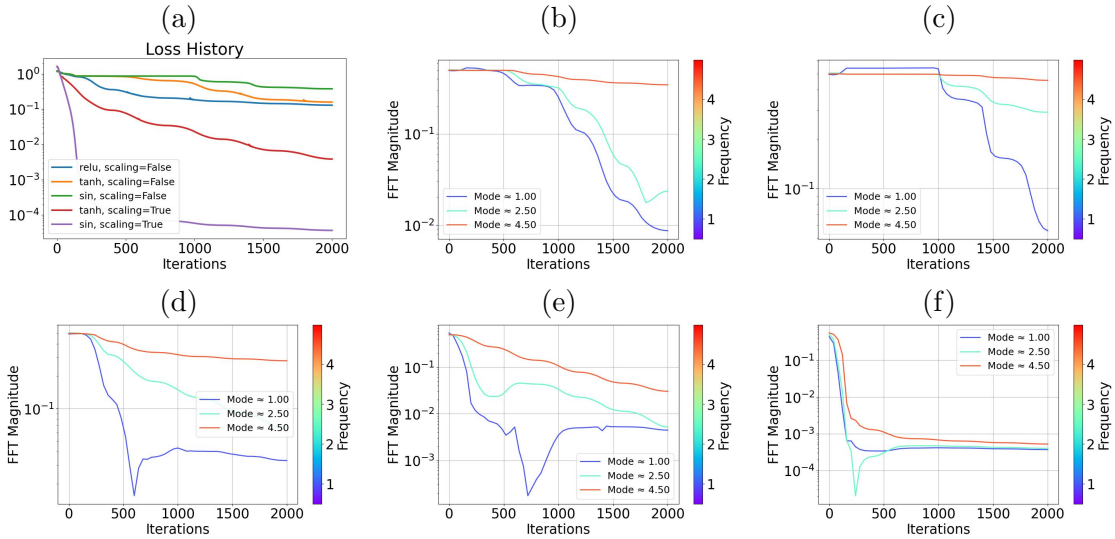


图1：示例1：使用不同激活函数并进行或不进行缩放时，函数逼近中的相对 L^2 误差和频率偏差。（a）训练过程中的相对 L^2 误差；（b）—（f）训练过程中不同模态的误差，分别对应激活函数双曲正切、正弦、ReLU、缩放后的双曲正切和缩放后的正弦。

至于频率偏差，未进行缩放的三个神经网络均表现出第一层偏置偏向低频；也就是说，较低频率模态中的误差在训练过程中衰减得更快。经过缩放后，采用正弦激活函数的神经网络对低频几乎没有偏差。这也使得训练过程更加高效。

示例2：第二项测试研究在其他条件相似的情况下，网络深度如何影响频率偏差。

我们使用相同的目标函数 u_1，并比较三种架构：浅层（2层）、4层和6层全连接神经网络，每种网络的宽度均为100，分别对应301、20,501和40,701个可学习参数。对于使用双曲正切或正弦激活函数的网络，均采用相同的首层缩放。模型使用带循环学习率调度的 Adam算法进行优化，学习率在基准值 $\eta_{min} = 1 \times 10^{-5}$ 和最大值 $\eta_{max} = 1 \times 10^{-2}$、$5 \times 10^{-3}$ 之间振荡；对于2层、4层和6层神经网络，最大值分别为 3×10^{-3}，指数衰减因子为 $\gamma = 0.999$。所有2层网络的总训练时间约为16秒，所有4层网络约为20秒，所有6层网络约为22秒。

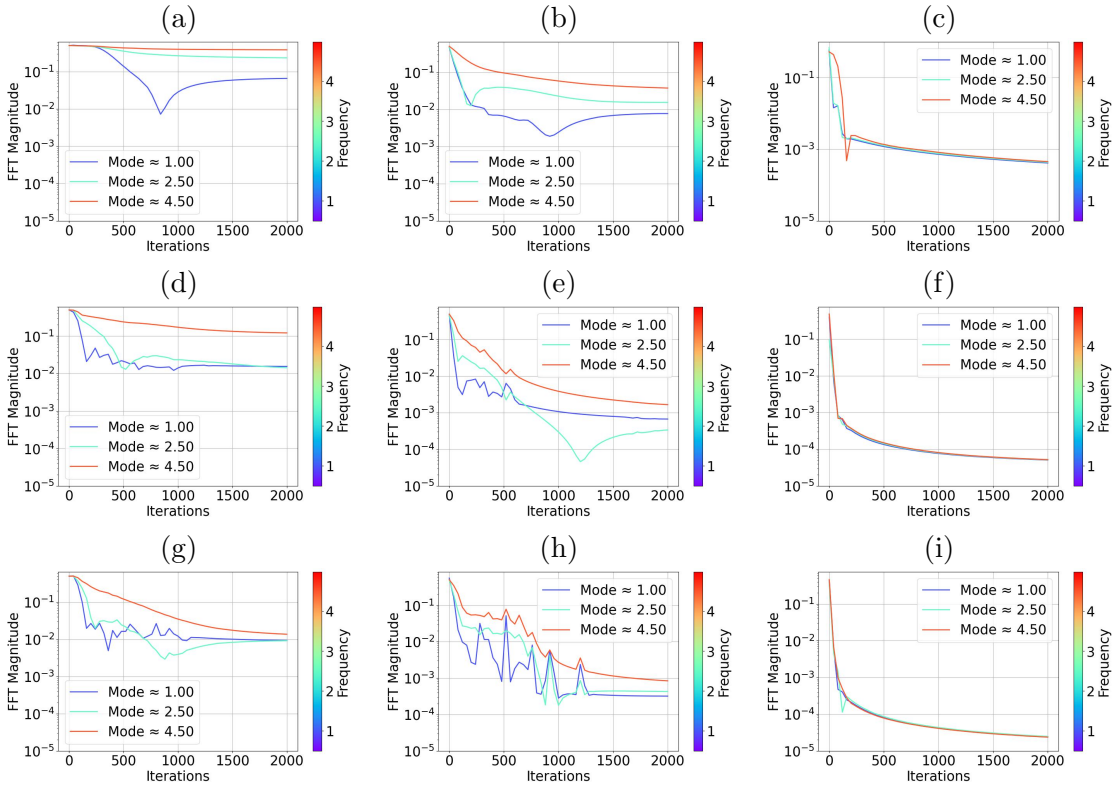


图2：示例2：使用 ReLU（左列）、双曲正切（中列）和正弦（右列）激活函数的 2 层（上排）、4 层（中排）和 6 层（下排）神经网络在不同模态下的误差。

随着层数增加，神经网络在训练早期表现出较低的频率偏差和更高的整体 L^2 准确度。对于 ReLU 激活函数（图 2(a)、2(d) 和 2(g)），层数更多的网络频率偏差显著降低。浅层网络明显偏向低频模态；也就是说，模式1的收敛速度远快于模式4.5。4层和6层网络的学习更加均衡，所有模态的误差均在2,000次迭代内降至约 10^{-2} 。对于采用缩放的双曲正切激活函数（图 2(b)、2(e) 和 2(h)），所有模型都表现出一定的频率偏差，但增加层数可使准确度提高约一个数量级（达到 10^{-3} ，而 10^{-2}），并使各模态的收敛更加均匀。对于采用缩放的正弦激活函数（图 2(c)、2(f) 和2(i)），浅层网络已表现良好，且频率偏差极小；深层网络则取得了出色结果，所有模态均收敛至 10^{-4} 误差或更低。

以下总结了上述函数逼近测试中的几项关键观察结果。

- 频率偏差越小，训练过程的效率和准确度越高。
- 频率偏差的程度取决于神经网络结构、激活函数和初始缩放。增加深度、采用适当的初始缩放以及使用正弦激活函数，都能减轻低频偏置，提高函数逼近效率。

3.2 求解偏微分方程时的频率偏差

现在，我们研究神经网络在训练求解偏微分方程时的频率偏差表现，以及这种表现与直接函数逼近中观察到的情况有何不同。我们以具有齐次狄利克雷边界条件的泊松方程为测试案例，研究不同设置下学习动力学的整体频率偏差。

求解偏微分方程时的训练动力学可能与函数逼近中的情况显著不同，这取决于微分算子与神经网络表征所具有的相反频率偏差之间的平衡。换言之，微分算子的存在会改变误差中不同频率分量的权重，并可能导致训练期间出现不同的频率偏差，从而在 L^2 误差的收敛模式上产生差异。研究表明，使用两层神经网络时 [HLZZ25]，即使解中包含高频分量，神经网络表征的低频偏置仍占主导地位，导致训练过程收敛缓慢且准确度较低。不过，我们将在下文表明，多层网络的情况可能相当微妙，也有所不同。多层神经网络的表达能力越强，其固有频率偏差就越弱；如上一节所示，这种偏差取决于网络深度、激活函数和初始缩放。此时，微分算子及其频率偏差在问题构建和训练过程中就越占主导地位。尽管如此，我们的测试表明，虽然频率偏差可能因设置而异，但总体频率偏差较弱仍能改善训练过程的收敛性和准确度。

在函数逼近的情形下，我们此前考察了 ReLU 作为激活函数。对于物理信息神经网络，激活函数至少需要使用 ReLU^3，其二阶导数与 ReLU 成正比。因此，我们比较了在求解偏微分方程和函数逼近中均使用 ReLU^3 的情况。

示例3：本例采用宽度为 200、包含 4 层的全连接神经网络，并使用 4,000 个采样点。该网络有 81,001 个可学习参数。泊松方程的精确解 u_2 为

$$
\begin{aligned}
u_2(x) = 1000\,\chi(x)\big[&\sin\big(60(x+0.2)^2\big)\,e^{-80(x+0.2)^2} \\
&+ 0.8\,\sin\big(300x^2\big)\,e^{-50x^2} \\
&+ 0.6\,\sin\big(40(x-0.2)^2\big)\,e^{-60(x-0.2)^2}\big],
\end{aligned}
$$

其中，光滑截断函数 $\chi(x)$定义为：

$$
\chi(x) = \begin{cases} \exp\left(-\dfrac{1}{(x+0.6)^2(0.6-x)^2}\right), & \text{if } -0.6 < x < 0.6 \\ 0, & \text{otherwise.} \end{cases}
$$

比较使用 ReLU 进行近似^3（图 3（b））与使用 ReLU 的物理信息神经网络^3（图 3（e）），可以看到：1）神经网络表示具有相当显著的低频偏置；2）由于神经网络表示的低频偏置与微分算子占主导地位的高频偏置相互平衡，物理信息神经网络中的频率偏置发生了逆转。我们还观察到，在物理信息神经网络中，这两种相反的频率偏置几乎相互抵消，从而在训练中实现更好的整体频率平衡，误差衰减也比直接函数逼近更快，如图 L^2 范数所示，见图 3（d）。请注意，物理信息神经网络的相对 L^2 误差呈现振荡行为，因为它不同于物理信息神经网络的损失函数；后者等价于 H^2 误差。实际上，整体物理信息神经网络损失平均而言平滑下降，这表明这种振荡并非由学习率过大所致。

借助正弦激活函数和初始缩放，多层网络的表征能力显著增强，其极小的频率偏差便证明了这一点，如图 3（c）所示，并且取得了良好的

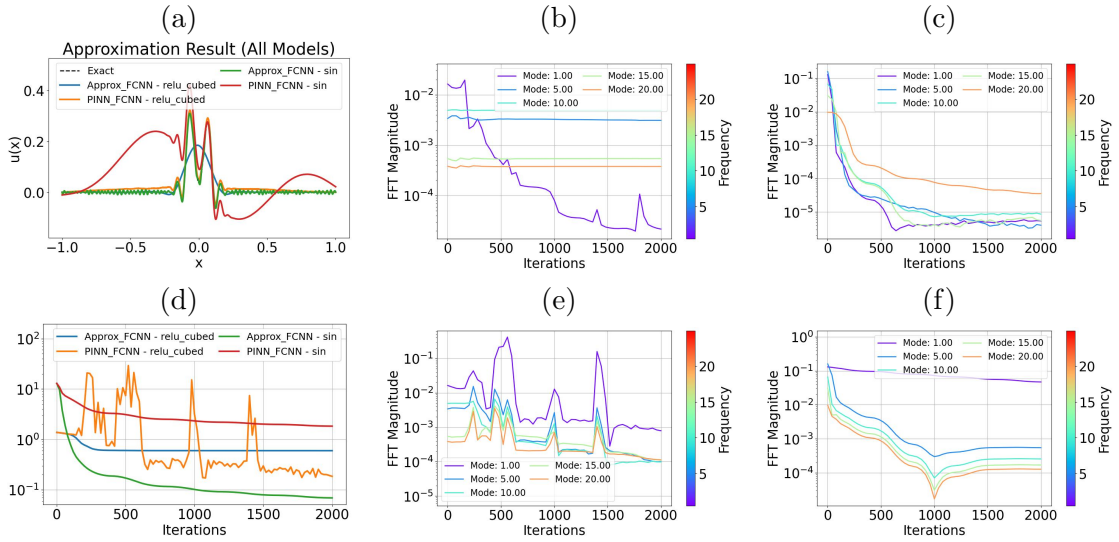


图3：示例3：训练性能与所选频率响应的比较。(a) 不同设置下训练 2,000 轮次后的神经网络输出；(b) 使用 ReLU^{3}激活函数进行近似；(c) 使用正弦激活函数进行近似；(d) 相对 L^2 误差历史；(e) 使用 ReLU^3 激活函数的物理信息神经网络；(f) 使用正弦激活函数的物理信息神经网络。

表1：不同设置下的频率偏差（示例3）。

Task	激活函数	宽度/深度	频率偏差	L^2 训练 2K 轮次后的误差
近似带来的内在偏差				
近似	ReLU^3	200 / 4	低频偏差	6.00E-01
近似	缩放正弦	200 / 4	**极小**	**6.91E-02**
PINN 求解偏微分方程中的频率偏差				
PINN	ReLU^3	200 / 4	轻度高频偏差	1.85E-01
PINN	缩放正弦	200 / 4	高频偏差	1.83E+00

近似结果，如图 3（d）所示。然而，这意味着物理信息神经网络模型中的微分算子占据主导地位，从而导致训练过程中出现显著的高频偏差，如图 3（f）所示，并使 L^2 误差的衰减速度相比直接近似更慢，如图3（d）所示。

表1 总结了这些观察结果。数值实验表明，最终训练性能与总体频率偏差相关，而总体频率偏差是网络表示与微分算子之间的一种平衡。

值得注意的是，我们经常观察到，物理信息神经网络损失持续下降的同时， L^2 误差却停滞不前甚至增大，这表明优化目标与目标误差指标之间存在不匹配。我们将这些现象归因于物理信息神经网络损失中微分算子所诱导的频率加权。PINNs 最小化的是残差范数（例如，$\|\Delta u_\theta - f\|_{L^2(\Omega)} = \|\Delta(u_\theta - u)\|_{L^2(\Omega)} \approx \|u_\theta - u\|_{H^2(\Omega)}$），而这并未与解误差 $\|u_\theta - u\|_{L^2(\Omega)}$ 直接对齐。在频域中，微分算子会放大较高频率模态，从而更加重视高频残差。这种算子诱导偏差可能抵消甚至逆转网络表示的频谱倾向，从而解释上述差异性表现。

4 面向物理信息神经网络的算子感知预条件器

基于我们在先前研究中获得的见解和观察，我们提出了一种用于 PINN 的预条件子，借助损失函数中的固定积分核，抵消微分算子诱导的频率偏差，并恢复直接函数逼近指标。

回顾一下，物理信息神经网络模型通过最小化损失函数进行训练：

$$\mathcal{L}_{\mathrm{PINN}}(\theta) = \mathcal{L}_{\mathrm{PINN},r}(\theta) + \lambda_b \mathcal{L}_b(\theta).$$

我们提出改为最小化以下损失函数：

$$\mathcal{L}_G(\theta) = \mathcal{L}_{G,r}(\theta) + \lambda_b \mathcal{L}_b(\theta), \tag{10}$$

其中

$$\begin{aligned} \mathcal{L}_{G,r}(\theta) &= \frac{1}{2}\int_\Omega \left| \int_\Omega G(x,y)\left[Lu_\theta(y) - f(y)\right] dy \right|^2 dx \\ &= \frac{1}{2}\int_\Omega \left| \int_\Omega G(x,y) L\left[u_\theta(y) - u(y)\right] dy \right|^2 dx \end{aligned} \tag{11}$$

和 $G(x,y)$ 表示参考椭圆算子 L_0 的格林函数，该算子与 L在谱意义上等价。实际应用中，我们选择 $G(x,y)$ 为拉普拉斯算子的格林函数，适用于简单区域上的齐次狄利克雷边界条件或自由空间情形；这些情形均有解析表达式。对于方程（ L 中的更复杂算子1），这种选择仍能有效地进行预条件，同时保持计算上的可处理性，数值实验将对此予以展示。

我们将积分算子 $\mathcal{G}f(x) := \int_\Omega G(x,y)f(y)dy$记为 $\mathcal{G}$。相应的连续梯度流为

$$\begin{aligned} \frac{d\theta}{dt} &= -\int_\Omega [\mathcal{G}L\frac{\partial u_\theta}{\partial \theta}][\mathcal{G}L(u_\theta - u)]dx - \lambda_b \int_{\partial\Omega} \frac{\partial u_\theta}{\partial \theta}(u_\theta - u)ds(x), \\ \frac{d\mathcal{L}_G}{dt} &= -\left\| \int_\Omega [\mathcal{G}L\frac{\partial u_\theta}{\partial \theta}][\mathcal{G}L(u_\theta - u)]dx + \lambda_b \int_{\partial\Omega} \frac{\partial u_\theta}{\partial \theta}(u_\theta - u)ds(x) \right\|^2. \end{aligned}$$

我们可以将 $\mathcal{G}L$ 视为经过预条件的微分算子。换言之，如果 $\mathcal{G} = L_0^{-1}$和、 L_0 与 L 在谱意义上等价，则 $\mathcal{L}_G(\theta) \approx \|u_\theta - u\|_{L^2(\Omega)}$。

4.1 预条件物理信息神经网络的实现

在实际实现中，可以进行若干修改，以提高预条件物理信息神经网络方法的效率和稳定性。

截断积分域。不同于进行全局积分，我们将格林函数预条件器的求和限制在每个点周围的局部邻域内。具体而言，对于点 $x \in\Omega$，积分（即离散形式下的求和）在一个球上进行，该球的半径等于从 x 到边界 $\partial\Omega$ 的距离。离散形式的预条件化损失函数中的项 $\mathcal{L}_{G,\ r}$ 可写为

$$\mathcal{L}_{G,r}(\theta) = \frac{1}{2N_r}\sum_{i=1}^{N_r}\left[\frac{1}{N_r}\sum_{j=1}^{M_i} G(x_i, y_{ij})\big(Lu_\theta(y_{ij}) - f(y_{ij})\big)\right]^2, \tag{12}$$

其中，$\{x_i\}_{i=1}^{N_r}$ 为内部配置点，$\{y_{ij}\}_{j=1}^{M_i} \subset B(x_i,\ r_i)$为球 $B(x_i,\ r_i)$ 内的采样点，该球以 x_i 为中心，半径为 $r_i = \mathrm{dist}(x_i,\ \partial\Omega)$。

截断的原因与好处如下。首先，若不进行截断，允许y_{ij} 遍及整个区域，会使靠近边界的点 x 的积分区域被区域边界部分截断。随着 x 接近边界，这种截断会变得越来越不对称，从而导致与真实格林函数的偏差不断增大，进而造成较差的预条件。其次，这种局部化降低了计算成本，并且最重要的是，仍能捕捉格林函数的奇异性，而该奇异性决定了微分算子的高频渐近行为，因而保持了良好的预条件效果。此外，这种局部化会随到边界的距离自适应调整，从而能够捕捉最具影响的边界条件。该截断的效果在第4.2小节中进行了详细分析。

公式（12）是连续积分算子的离散近似。具体而言，x_i 与y_{ij} 的分布决定了离散误差。在我们的测试中，我们对 x_i 和y_{ij}都在网格上选取配点。因此，对于底层网格能够良好分辨的频率模态，核函数的谱性质可以得到准确保持。我们指出，由于神经网络表示允许在任意点进行求值与求导，x_i和y_{ij} 可以自由采样。据此，人们有可能围绕每个 x_i 设计更高效、更准确的自适应采样。这种无网格表示对于复杂几何以及高维情形更为有益。

边界条件。 众所周知，为标准物理信息神经网络模型（3）设计损失函数时的一个挑战，是域内部的偏微分方程残差项 $\mathcal{L}_r$ 与边界误差项 $\mathcal{L}_b$之间的适当平衡。实际上，这种平衡依赖于解。以狄利克雷边界条件为例，$\mathcal{L}_r$ 是涉及解的二阶导数的逼近误差，而 $\mathcal{L}_b$ 是解在边界上的逼近误差。特别是，如果解具有多尺度特征，就很难选择常数λ_b 来实现适当的平衡。对于狄利克雷边界条件，这对预条件物理信息神经网络模型（10）而言是直接明了的，因为$\mathcal{L}_{G,\ r}(\theta) \approx \|u_\theta - u\|_{L^2(\Omega)}$，如果 $G = L_0^{-1}$且 L_0 与 L谱等价。因此，只要 λ_b 适当地平衡边界与内部的采样密度，这两项就能很好地匹配，并且 $\mathcal{L}_G(\theta) \approx \|u_\theta - u\|_{L^2(\Omega)}$，其中$\Omega = \Omega \cup \partial\Omega$。在本工作中，我们在所有数值实验中将预条件损失函数中的正则化参数设为1。我们还观察到，在施加预条件时，性能对该参数的选择并不敏感。

计算成本。 使用积分算子会引入额外的求和操作。该求和可通过快速多极子方法 (FMM) 加速 [GR87，YBZ04] 或通过层次半可分 (HHS) [CGP06，CGP06] 对核函数进行稀疏化。然 而，在实践中，与计算神经网络表示的导数以及每个训练步骤中的反向传播操作所需的计算成本相比，这项额外求和的成本微不足道。因此，正如数值测试所示，预条件物理信息神经网络与标准物理信息神经网络的计算成本相当。

我们进行了大量数值实验，从一维示例到二维多尺度问题，以展示应用预条件后收敛速度显著提升、准确度明显改善。此外，该预条件框架适用于任何网络结构，并能提升其性能。对于一维问题，我们使用全连接神经网络作为网络表示。对于更具挑战性的二维问题，我们采用 [ZZZZ25a、ZZZZ25b] 中提出的多分量多层神经网络（MMNN）架构来表示解 u_θ。通过在网络中引入平衡多分量结构，可以轻松地对全连接神经网络进行改造

——每层的分量数（称为秩）远小于网络宽度。与全连接神经网络相比，MMNN 能大幅减少训练参数，显著提高训练效率和准确度，尤其适用于具有复杂特征的函数。这类架构提升了高维问题中网络宽度和深度方面的可扩展性，使预条件物理信息神经网络更适用于大规模应用。我们将在例6(a)和6(b)中展示，对于相同的偏微分方程，采用 MMNN 架构可以用比全连接神经网络少一个数量级的可学习参数，达到同等准确度。对于所有网络结构，我们均使用正弦激活函数和初始缩放，以提升表示能力并降低神经网络表示的频率偏差。此外，使用正弦作为激活函数的网络可以按需进行任意次求导。结合这些神经网络设置与预条件策略，我们能够以接近直接函数逼近的效率和准确度求解偏微分方程。

4.2 预条件化损失项的有效范数

在本小节中，我们分析预条件化损失的连续形式 $\mathcal{L}_{G,\ r}$ 见（12），并证明对于任意足够小的 $\varepsilon > 0$，采用截断积分域时，该项弱于 $\|L(u_\theta - u)\|^2_{H^{-1+\varepsilon}(\Omega)}$。这实质上意味着，损失函数从 $L(u_\theta - u)$ 中至少恢复了 $(1-\varepsilon)$ 阶导数。此外，如果我们能够对 $L(u_\theta - u)$ 的傅里叶变换衰减获得一致估计，则损失函数有可能恢复 $3/2$ 阶导数。

引理 4.1。令 $\Phi_d(r) = \log(r)$（$d = 2$ 时），且 $\Phi_d(r) = \sqrt[2]{\frac{1}{\pi}}\frac{1}{r}$（$d = 3$ 时）。定义 $\mathcal{A}_d(\rho, \ell)$、$\ell \in [0,1]$ 如下：

$$\begin{aligned}\mathcal{A}_d(\rho,\ell) &:= \rho^{-\frac{d-2}{2}} \int_0^\ell \Phi_d(r) J_{\frac{d-2}{2}}(\rho r) r^{\frac{d-2}{2}} r dr \\ &= \begin{cases} \frac{1}{\rho^2}\left(1 - J_0(\rho\ell) - \log(\ell)\rho\ell J_1(\rho\ell)\right), & d = 2, \\ \frac{1}{\rho^2}(1 - \cos(\rho\ell)), & d = 3, \end{cases}\end{aligned}$$

其中，J_ν 是 ν 阶第一类贝塞尔函数。则存在一个绝对常数 $C_{\mathcal{A}} > 0$，使得

$$|\mathcal{A}_d(\rho,\ell)| \le \begin{cases} C_{\mathcal{A}} \min(1, \rho^{-3/2}), & d = 2 \\ C_{\mathcal{A}} \min(1, \rho^{-2}), & d = 3. \end{cases}$$

对所有 $\rho \ge 0$。

证明。当 $d = 2$ 时，对于 $x \in [0,1]$，$J_0(x) = 1 + \mathcal{O}(x^2)$ 与 $J_1(x) = \mathcal{O}(x)$。对于 $x > 1$，我们使用 $k = 0$ 的渐近近似，1，见 [Kra14] 即

$$J_k(x) = \sqrt{\frac{2}{\pi x}} \cos(x - \frac{k\pi}{2} - \frac{\pi}{4}) + \mathcal{O}\left(\frac{|k^2 - \frac{1}{4}|}{x^{3/2}}\right),$$

由此，令 $x = \rho\ell$ 即可得到所需估计。注意 $\ell^2|\log \ell| \le \ell|\log \ell|$ 在 $\ell \in [0,1]$ 上一致有界。

当 $d = 3$ 时，我们有 $1-\cos(\rho\ell) = \mathcal{O}((\rho\ell)^2)$，当 $\rho\ell \in [0,1]$ 时；若 $\rho\ell > 1$，我们使用 $|1-\cos(\rho\ell)| = \mathcal{O}(1)$。 □

定理 4.2。设 $\Omega = B(0,\ 1)$ 为 $\mathbb{R}^d$, $d = 2,3$ 中的单位圆盘，且 $\ell(x) = \mathrm{dist}(x, \partial\Omega) = 1 - |x|$。记 χ 为单位圆盘的特征函数，并

$$G_d(x) = \begin{cases} -\frac{\log|x|}{2\pi}, & d = 2, \\ -\frac{1}{4\pi|x|}, & d = 3, \end{cases}$$

为自由空间中拉普拉斯算子的格林函数。定义

$$\mathcal{T}f(x) := \int_{\mathbb{R}^2} \chi\left(\frac{x-y}{\ell(x)}\right) G(x-y) f(y) dy.$$

1. 设 $f \in H^s(\Omega)$，则对任意 $\varepsilon > 0$，存在常数 C_ε 使得

$$\|\mathcal{T}(I-\Delta)^s f\|_{L^2(\Omega)} \le C_\varepsilon \|f\|_{H^{s-(1-\varepsilon)}(\Omega)},$$

2. 设 $\widehat{f}(\xi)$ 为 f 的傅里叶变换。假设

$$\widehat{f}(\xi) = \begin{cases} \displaystyle\sum_{k=-\infty}^{\infty} c_k(|\xi|) e^{ik\arg(\xi)}, & d = 2, \\ \displaystyle\sum_{k=0}^{\infty} \sum_{l=-k}^{k} c_{k,l}(|\xi|) Y_{kl}\left(\frac{\xi}{|\xi|}\right), & d = 3, \end{cases}$$

满足该条件。

$$|c_\alpha(|\xi|)| \le D\gamma(\alpha)(1+|\xi|^2)^{-\frac{s}{2}-\delta-\frac{d-2}{2}},$$
$$|c_\alpha(|\xi|)| \le \gamma(\alpha) \left(\sum_\beta |c_\beta(|\xi|)|^2\right)^{1/2},$$

其中 α、β 表示 $d=2$ 或 $d=3$ 中的多重下标，Y_{kl} 为球谐函数，D 为常数，$\delta > 0$，且 $\sum_\beta |\gamma(\beta)|^2 < \infty$。则存在 $p \in (2,\infty)$ 及常数 $C' > 0$ 使得

$$\|\mathcal{T}(I-\Delta)^s f\|_{L^2(\Omega)} \le C' \|f\|_{B,s,p}^{(p-2)/p} \|f\|_{H^{s-3/2}(\Omega)}^{2/p},$$

其中范数

$$\|f\|_{B,s,p} := \left(\int_0^\infty \left| (1+|\xi|^2)^{s-(2-\frac{d}{2})} \sum_\alpha |c_\alpha(\xi)|^2 \right|^{\frac{p-2}{2(p-1)}} d|\xi| \right)^{\frac{p-1}{p-2}}.$$

证明。设 $\phi \in L^2(\Omega)$，则

$$\begin{aligned}
&\int_\Omega \left| \int_{\mathbb{R}^2} \chi\left(\frac{x-y}{\ell(x)}\right) G_d(x-y)\phi(y) dy \right|^2 dx \\
&= \int_\Omega \left| \int_{\mathbb{R}^2} \chi\left(\frac{x-y}{\ell(x)}\right) G_d(x-y) \int_{\mathbb{R}^2} \widehat{\phi}(\xi) e^{i\pi\xi\cdot y} dy \right|^2 dx \\
&= \int_\Omega \left| \int_{\mathbb{R}^2} \widehat{\phi}(\xi) \left(\int_{\mathbb{R}^2} \chi\left(\frac{x-y}{\ell(x)}\right) G_d(x-y) e^{i\xi\cdot y} dy \right) d\xi \right|^2 dx \\
&= \int_\Omega \left| \int_{\mathbb{R}^2} \widehat{\phi}(\xi) e^{i\xi\cdot x} \mathcal{A}_d(|\xi|, \ell(x)) d\xi \right|^2 dx \\
&= \int_{\mathbb{R}^2} \int_{\mathbb{R}^2} \mathcal{H}_d(\xi,\zeta) \widehat{\phi}(\xi) \overline{\widehat{\phi}(\zeta)} d\xi d\zeta.
\end{aligned}$$

核函数

$$\begin{aligned}\mathcal{H}_d(\xi,\zeta) &= \int_\Omega \mathcal{A}_d(|\xi|,\ell(x))\mathcal{A}_d(|\zeta|,\ell(x))e^{i(\xi-\zeta)\cdot x}dx \\ &= \begin{cases} 2\pi\displaystyle\int_0^1 \mathcal{A}_d(|\xi|,1-r)\mathcal{A}_d(|\zeta|,1-r)J_0(|\xi-\zeta|r)rdr, & d=2, \\ (2\pi)^{3/2}\displaystyle\int_0^1 \mathcal{A}_d(|\xi|,1-r)\mathcal{A}_d(|\zeta|,1-r)\frac{J_{1/2}(|\zeta-\xi|r)}{\sqrt{|\zeta-\xi|r}}r^2dr, & d=3, \end{cases}\end{aligned} \tag{13}$$

其中 J_ν 表示阶数为 ν 的第一类贝塞尔函数。接下来，我们利用 Graf 加法公式和根德鲍尔加法公式重写（13）

$$\begin{aligned} J_0(|\zeta-\xi|r) &= \sum_{k=-\infty}^{\infty} J_k(|\xi|r)J_k(|\zeta|r)C_k^{(0)}(\cos\theta), \\ \frac{J_{1/2}(|\zeta-\xi|r)}{\sqrt{|\zeta-\xi|r}} &= \sqrt{\frac{\pi}{2}}\sum_{k=0}^{\infty}(2k+1)J_{k+1/2}(|\xi|r)J_{k+1/2}(|\zeta|r)C_k^{(1/2)}(\cos\theta), \end{aligned}$$

其中 θ 是 ξ 与 ζ之间的角， $C_k^{(0)}$ 和 $C_k^{(1/2)}$ 是根德鲍尔多项式。于是我们得到 $d=2,\ 3$ 的分解：

$$\mathcal{H}_d(\xi,\zeta) = \begin{cases} 2\pi\displaystyle\sum_{k=-\infty}^{\infty}\mathcal{H}_{d,k}(|\xi|,|\zeta|)C_k^{(0)}(\cos\theta), & d=2, \\ 2\pi^2\displaystyle\sum_{k=0}^{\infty}(2k+1)\mathcal{H}_{d,k}(|\xi|,|\zeta|)C_k^{(\frac{1}{2})}(\cos\theta), & d=3, \end{cases}$$

$$\mathcal{H}_{d,k}(|\xi|,|\zeta|) = \begin{cases} \displaystyle\int_0^1 \mathcal{A}_d(|\xi|,1-r)\mathcal{A}_d(|\zeta|,1-r)J_k(|\xi|r)J_k(|\zeta|r)rdr, & d=2, \\ \displaystyle\int_0^1 \mathcal{A}_d(|\xi|,1-r)\mathcal{A}_d(|\zeta|,1-r)J_{k+1/2}(|\xi|r)J_{k+1/2}(|\zeta|r)r^2dr, & d=3. \end{cases}$$

根据引理 4.1， $\mathcal{H}_{dk}(|\xi|,|\zeta|)$ 对所有 ξ、ζ 和 k 都一致有界。对于 $|\nu|\le\frac{1}{2}$，已知 $|J_\nu(\rho r)|\le\sqrt{\frac{2}{\pi\rho r}}$ [Sze39, Theorem 7.31.2]。因此，存在一个绝对常数 C，使得 $|\mathcal{H}_{d,0}(|\xi|,|\zeta|)|\le\frac{C}{|\xi|^{(d+2)/2}|\zeta|^{(d+2)/2}}$ 对于 $d=2,3$ 成立。对于 $|\nu|\ge 1$，我们采用 [Kra14] $|J_\nu(\rho r)|\le\sqrt{\frac{2}{\pi}}|\rho^2r^2-\omega_\nu^2|^{-1/4}$, $\omega_\nu^2=\nu^2-\frac{1}{4}$ 中的估计。然后，取 $\omega=\omega_k$ 或 $\omega=\omega_{k+1/2}$，可得估计

$$\begin{aligned} |\mathcal{H}_{d,k}(|\xi|,|\zeta|)| &\le \frac{C_{\mathcal{A}}^2}{|\xi|^{(d+2)/2}|\zeta|^{(d+2)/2}}\int_0^1 \frac{r}{|r^2-(\frac{\omega}{|\xi|})^2|^{1/4}|r^2-(\frac{\omega}{|\zeta|})^2|^{1/4}}dr \\ &\le \frac{C_{\mathcal{A}}^2}{|\xi|^{(d+2)/2}|\zeta|^{(d+2)/2}}\int_0^1 \frac{\sqrt{r}}{|r-\frac{\omega}{|\xi|}|^{1/4}|r-\frac{\omega}{|\zeta|}|^{1/4}}\frac{\sqrt{r}}{|r+\frac{\omega}{|\xi|}|^{1/4}|r+\frac{\omega}{|\zeta|}|^{1/4}}dr \\ &\le \frac{C_{\mathcal{A}}^2}{|\xi|^{(d+2)/2}|\zeta|^{(d+2)/2}}\int_0^1 \frac{\sqrt{r}}{|r-\frac{\omega}{|\xi|}|^{1/4}|r-\frac{\omega}{|\zeta|}|^{1/4}}dr \\ &\le \frac{C}{|\xi|^{(d+2)/2}|\zeta|^{(d+2)/2}}. \end{aligned}$$

因此，结合$\mathcal{H}_{d,\ k}$的一致有界性，我们得到以下界

$$|\mathcal{H}_{d,k}(|\xi|,|\zeta|)| \leq \frac{C'}{(1+|\xi|^2)^{(d+2)/4}(1+|\zeta|^2)^{(d+2)/4}}$$

其中，$C'>0$为某个绝对常数。对于任意 $f \in H^s(\Omega)$，取 $\phi=(I-\Delta)^{s/2}f \in L^2(\Omega)$，则

$$\widehat{\phi}(\xi)=\begin{cases}\displaystyle\sum_{k=-\infty}^{\infty}(1+|\xi|^2)^{s/2}c_k(|\xi|)e^{ik\arg(\xi)}, & d=2,\\ \displaystyle\sum_{k=0}^{\infty}\sum_{l=-k}^{k}(1+|\xi|^2)^{s/2}c_{k,l}(|\xi|)Y_{kl}\left(\frac{\xi}{|\xi|}\right), & d=3,\end{cases}$$

其中，Y_{kl} 表示 $\mathbb{S}^{d-1}$上的球谐函数。然后，记 $\Theta(|\xi|,|\zeta|)=(1+|\xi|^2)(1+|\zeta|^2)$，应用柯西-施瓦茨不等式，

$$\begin{aligned}
&\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}\mathcal{H}_d(\xi,\zeta)\widehat{\phi}(\xi)\overline{\widehat{\phi}(\zeta)}d\xi d\zeta\\
&\asymp\begin{cases}\displaystyle\sum_{k=-\infty}^{\infty}\iint_{\mathbb{R}^2_+}\mathcal{H}_{d,k}(|\xi|,|\zeta|)\Theta(|\xi|,|\zeta|)^{s/2}c_k(|\xi|)\overline{c_k(|\zeta|)}|\zeta|^{d-1}|\xi|^{d-1}d|\zeta|d|\xi|, & d=2\\ \displaystyle\sum_{k=0}^{\infty}\sum_{l=-k}^{k}\iint_{\mathbb{R}^2_+}\mathcal{H}_{d,k}(|\xi|,|\zeta|)\Theta(|\xi|,|\zeta|)^{s/2}c_{k,l}(|\xi|)\overline{c_{k,l}(|\zeta|)}|\zeta|^{d-1}|\xi|^{d-1}d|\zeta|d|\xi|, & d=3\end{cases}\\
&\leq C'\sum_{\alpha}\left(\int_0^{\infty}(1+|\xi|^2)^{\frac{s}{2}-\frac{d+2}{4}}|c_\alpha(|\xi|)||\xi|^{d-1}d|\xi|\right)^2\\
&\leq C'\sum_{\alpha}\left(\int_0^{\infty}\frac{1}{(1+|\xi|^2)^{\frac{d}{2}+\varepsilon}}|\xi|^{d-1}d|\xi|\right)\left(\int_0^{\infty}(1+|\xi|^2)^{s-1+\varepsilon}|c_\alpha(|\xi|)|^2|\xi|^{d-1}d|\xi|\right)\\
&\asymp\|f\|^2_{H^{s-(1-\varepsilon)}},
\end{aligned}$$

其中，我们用 α 表示 $d=2$ 和 $d=3$求和中的所有可能下标。对于第二部分，在最后一步中，我们采用 Hölder 不等式，而不是柯西-施瓦茨不等式。利用定理中 $|c_k(|\xi|)|$ 的受控衰减率，可得

$$\begin{aligned}
&\sum_{\alpha}\left(\int_0^{\infty}(1+|\xi|^2)^{\frac{s}{2}-\frac{d+2}{4}}|c_\alpha(|\xi|)||\xi|^{d-1}d|\xi|\right)^2\\
&\leq\sum_{\alpha}\left(\int_0^{\infty}\left|(1+|\xi|^2)^{\frac{s-1}{2}}|c_\alpha(\xi)|\right|^{\frac{p-2}{p-1}}d|\xi|\right)^{2(p-1)/p}\left(\int_0^{\infty}(1+|\xi|^2)^{s-3/2}|c_\alpha(\xi)|^2|\xi|^{d-1}d|\xi|\right)^{2/p}\\
&\leq C''\left(\int_0^{\infty}\left|(1+|\xi|^2)^{s-(2-\frac{d}{2})}\sum_{\alpha}|c_\alpha(|\xi|)|^2\right|^{\frac{p-2}{2(p-1)}}d|\xi|\right)^{2(p-1)/p}\|f\|^{4/p}_{H^{s-3/2}(\Omega)}\\
&=C''\|f\|^{2(p-2)/p}_{B,s,p}\|f\|^{4/p}_{H^{s-3/2}(\Omega)},
\end{aligned}$$

其中 $C''=\sum_{k=-\infty}^{\infty}\gamma_k^2<\infty$，当 p 足够大时，$\|f\|_{B,\ s,\ p}$ 是有限的。 □

5 数值实验

本节中，我们在一系列问题设置下考察预条件物理信息神经网络的性能，以评估所提出的预条件策略的有效性。示例4–5考虑一维泊松方程。示例6(a)–(b)考察具有变系数的一维椭圆方程。示例7和8分别研究具有齐次和非齐次狄利克雷边界条件的二维泊松方程，而示例9则研究具有变系数的二维椭圆方程。示例10考察三维泊松方程。对于所有示例，我们均采用全批量训练，并将激活函数设为正弦，第一层的权重按 $(n_1)^{\frac{1}{d}}$ 成比例缩放，其中 n_1 是第一层的网络宽度， d 是输入维度。

在实现预条件损失形式（10）时，对于定义在区域 $\Omega = (-1,1)$ 上的所有一维示例，我们通过选择 $\mathcal{L}_{G,r}$ 为形式（11）中的函数，并令 G 为满足齐次狄利克雷边界条件的拉普拉斯算子的格林函数，其解析表达式为：

$$G(x,y) = \begin{cases} \frac{1}{2}(1-y)(1+x), & \text{if } x \le y \\ \frac{1}{2}(1-x)(1+y), & \text{if } x > y. \end{cases}$$

对于定义在不同形状区域上的二维（$d=2$）和三维（$d=3$）问题，我们采用（12）中的截断积分形式，并在 G 预条件形式中选取函数，使其为自由空间中泊松方程的格林函数，具体如下：

$$G(x,y) = \begin{cases} -\frac{1}{2\pi}\log(|x-y|), & d=2, \\ -\frac{1}{4\pi|x-y|}, & d=3. \end{cases}$$

在下面的示例中，我们使用预条件物理信息神经网络求解偏微分方程，并与未进行预条件的物理信息神经网络及函数逼近进行比较，绘制 L^2 误差，并展示误差中所选频率模态的动态变化。

示例4： 我们求解具有齐次狄利克雷边界条件的泊松方程，其中强迫项 f 经过选取，使得 $u_1(x) = \sin(2\pi x) + \sin(5\pi x) + \sin(9\pi x)$ 为精确解。我们使用宽度为 50、深度为 3 的全连接神经网络来表示该解，对应 5,251 个可学习参数。模型使用 2,000 个配点进行训练。

图 4 展示了标准物理信息神经网络、预条件物理信息神经网络（图例中标记为 $ker\text{-}nelGPINN\ FCNN$ ）与直接函数逼近之间的比较。我们确实观察到，预条件物理信息神经网络的频率模态动态与函数逼近的非常相似。其误差 L^2 衰减率也远优于未经预条件的物理信息神经网络表述，并且与直接近似相当。直接函数逼近的训练时间为16秒，标准物理信息神经网络和预条件物理信息神经网络的训练时间分别为24秒和26秒。尽管额外开销略有增加，预条件策略仍有效缓解了标准物理信息神经网络表述中由微分算子引起的高频偏差，恢复了不同频率模态之间的平衡，并提高了收敛性和训练精度。

示例5： 我们求解具有齐次狄利克雷边界条件的泊松方程，其精确解是示例3中的局部化函数 u_2 ，其中包含精细特征和高频模式。我们使用宽度为200、深度为3的全连接神经网络，并保持与示例4相同的训练配置。

在此示例中，图中展示了选定频率模态 1。0、5。0、10。0、15。0 和 20。0 的误差。 5所示的范围比示例4更广。由于局部化函数 u_2

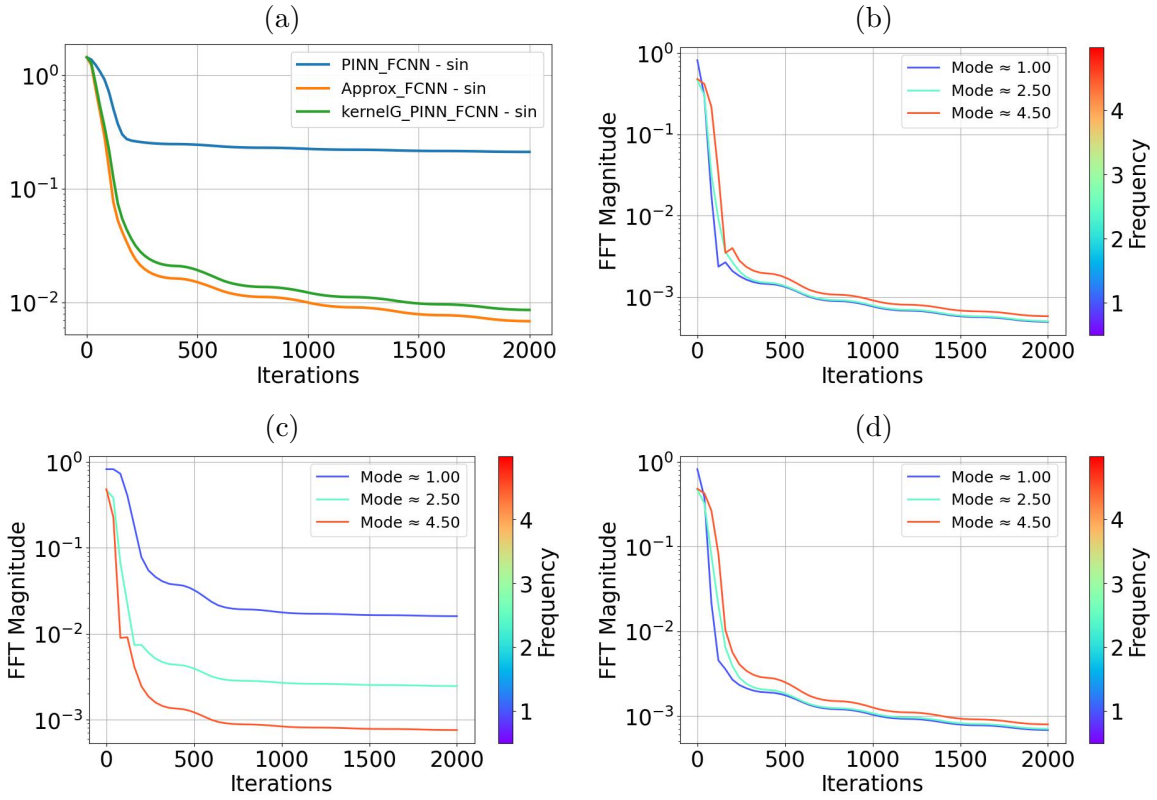


图4：示例4：不同模型训练过程中 L^2 误差和选定频率模态误差的比较。(a) 所有模型的训练损失；(b) 函数逼近；(c) 物理信息神经网络；(d) 预条件物理信息神经网络。

包含相当多的高频分量，因此使用相对浅层的神经网络进行函数逼近，仍会表现出轻微的低频偏置，并且需要更多迭代次数才能捕捉频率模态20，如图 5(b)所示。然而，图 5(c)显示，对于标准物理信息神经网络，由于损失函数对频率模态重新加权并更加重视高频分量，微分算子的存在彻底改变了情况。因此，这会导致强烈的高频偏差，并使训练中平滑分量的收敛速度变慢，这在实际应用中不可接受（类似于对椭圆型偏微分方程使用有限元方法却不采用预条件子）。相比之下，预条件物理信息神经网络呈现出的频率模态动力学与直接函数逼近几乎完全相同，因此训练效率和准确度都显著提高。直接函数逼近的训练时间为32秒，标准物理信息神经网络和预条件物理信息神经网络的训练时间则与之相当（分别为38秒和41秒）。这些测试再次证明了我们提出的预条件策略的有效性。

Exa例6(a)：我们考虑一个具有快速振荡变系数的椭圆型偏微分方程 t:

$$\begin{cases} \nabla \cdot (k(x)\nabla u) = -1 & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}, \tag{14}$$

其中

$$k(x) = \frac{2}{2 + A\sin\left(\frac{2\pi x}{\varepsilon}\right)} + 2 + A\cos\left(\frac{2\pi x}{\varepsilon}\right). \tag{15}$$

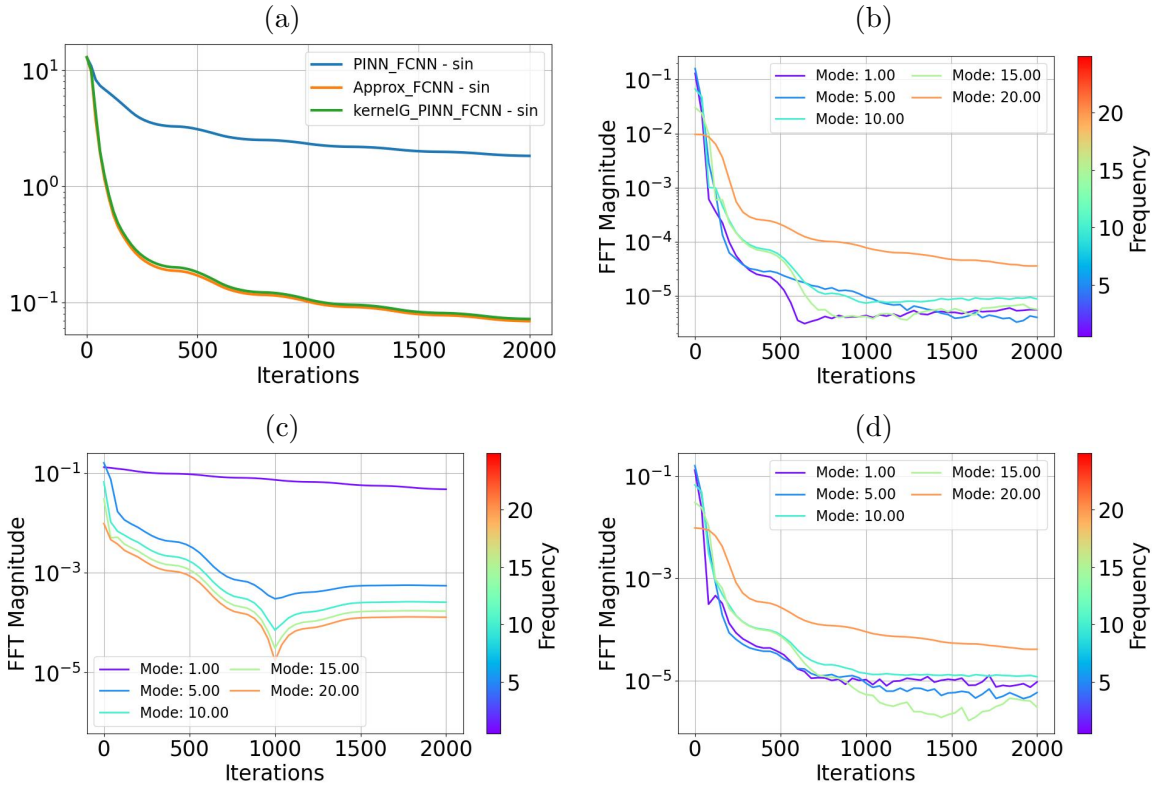


图5：示例5：不同模型训练过程中 L^2 误差及所选频率模态误差的对比。(a) 所有模型的训练损失；(b) 函数逼近；(c) PINN；(d) 预条件物理信息神经网络。

在此测试中，我们取 $A = 1.8$ 和 $\varepsilon = 0.1$，由此构造出一个周期为 ε 的快速振荡扩散系数，使解同时具有缓慢的全局变化和快速的局部振荡，这是一个经典的均匀化问题。我们使用宽度为 100、深度为 3 的全连接神经网络来表示该解，对应 20,501 个可学习参数。

由于解具有多尺度特性，这个例子对 PINN 尤具挑战性。我们观察到，标准物理信息神经网络存在显著的高频偏差，且难以捕捉低频模态，导致相对 L^2 误差很大；即使经过 20,000 个训练轮次，误差仍然衰减得非常缓慢。尽管参考微分算子 Δ与方程中的精确微分算子不同，预条件子仍能有效缓解频率偏差，模态动力学6(d) 显示其高频偏差显著减轻。频率偏差的降低也使整体准确度显著提升。具体而言，采用全连接神经网络的标准物理信息神经网络，其相对 L^2 误差为 9.61×10^{-1}，而预条件物理信息神经网络将误差降至 1.80×10^{-2}。作为对比，采用全连接神经网络的直接函数逼近模型，其相对 L^2 误差为 1.88×10^{-3}。

例6(b)：在本例中，我们评估宽度为 50、秩为 20、深度为 3 的傅里叶MMNN [ZZZZ25a]（FMMNN）用于函数逼近，或用于求解与例6(a)中相同的变系数方程。

观察到与例6(a)中类似的现象。MMNN 架构仅用显著更少的可学习参数便达到了相同的准确度：其参数量为 2,091 个，而全连接神经网络为 20,501 个，约减少了十倍。尽管

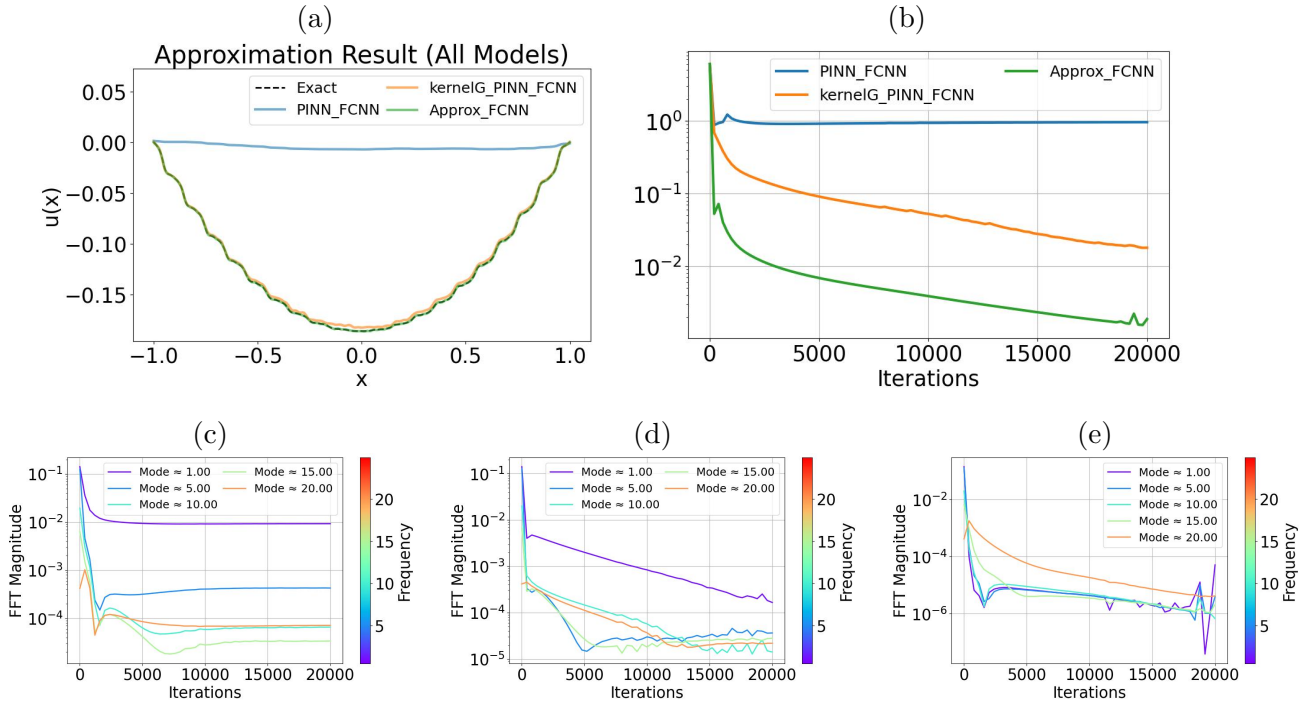


图6：例6(a)：(a) 精确解及使用全连接神经网络得到的数值结果；(b) L^2 相对误差历史；物理信息神经网络 (c)、预条件物理信息神经网络 (d) 和近似 (e) 中选定模态的误差。

模型规模大幅缩小，定性的趋势仍然相同。具体而言，采用 MMNN 的标准物理信息神经网络具有较大的相对误差 L^2，为 9.78×10^{-1}；而经过预条件的物理信息神经网络将误差降低至 3.07×10^{-2}。作为对比，基于 MMNN 的近似模型的相对误差 L^2 为 4.16×10^{-3}。这表明，无论采用全连接神经网络还是 MMNN，架构创新与预条件都是提升物理信息神经网络性能的互补策略。

如例6所示，MMNN 架构能够以显著更少的可学习参数达到相当的准确度。后续所有二维数值实验均采用 MMNN 架构。

在二维情形下，通过对误差进行二维离散傅里叶变换来分析频率偏差。每个傅里叶系数对应一对频率模态 $(k_x,\ k_y)$，分别沿 x 和 y 方向。为可视化频率相关学习动态，我们跟踪一组选定的傅里叶模态 $(k_x,\ k_y)$ $=(1,\ 0),\ (2,\ 0),\ (3,\ 0),\ (4,\ 0)$，它们对应沿 x 方向的振荡。监测这些模态在训练迭代过程中的误差，可直接与一维频率分析进行比较。同时，还将绘制空间误差，以整体展示近似质量并揭示学习得到的解中的误差模式。结合选定模态的动力学与空间误差可视化，可以从互补角度深入了解二维情形下的频率偏差。

示例 e 7: 我们求解 Ω 中的泊松方程 $\Delta u=-f=[-1,1]\times[-1,1]$，其边界条件为齐次 us

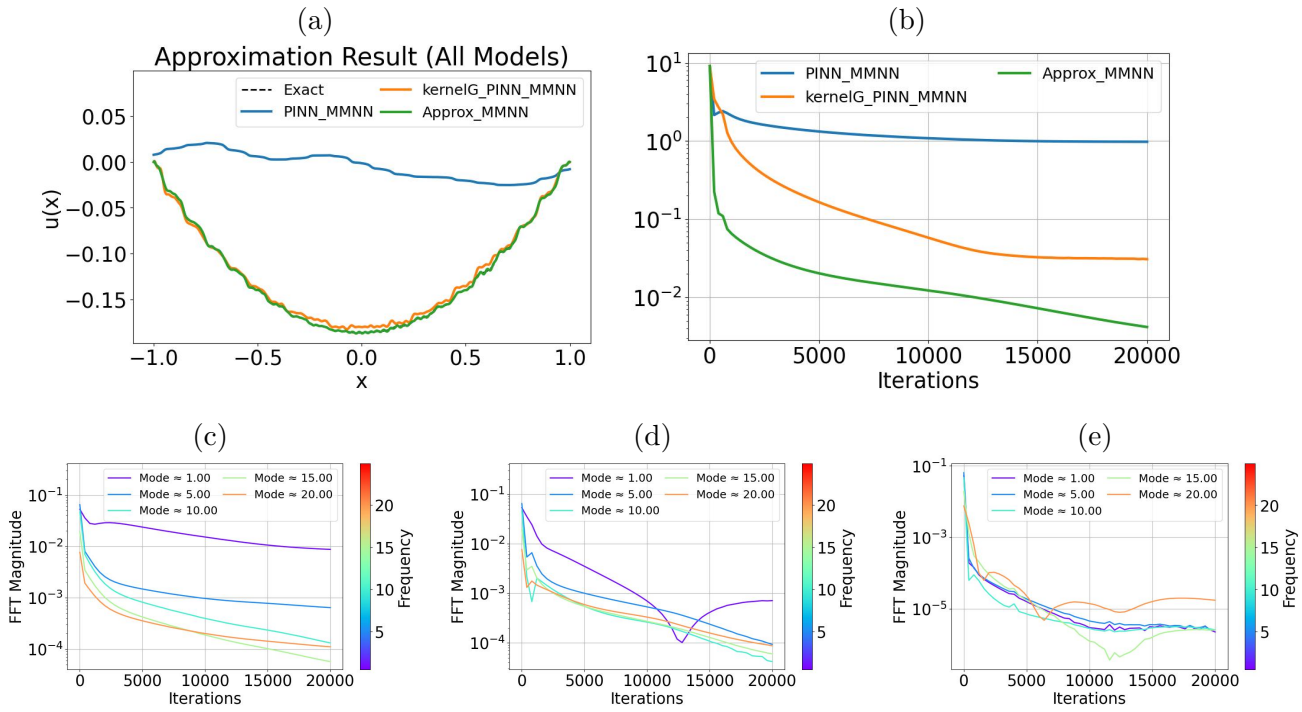


图7：例6(b)：(a) 使用 MMNN 得到的精确解和数值结果；(b) L^2 相对误差历史；PINN (c)、预条件物理信息神经网络 (d) 和近似 (e) 在选定模态上的误差。

狄利克雷边界条件 边界条件，源函数 f 的选取使得精确解 is

$$\begin{aligned} u_3(x) = 1.0 \cdot \phi(x; c^{(1)}, 0.8) \\ + 0.7 \cdot \phi(x; c^{(2)}, 0.2) \\ + 0.5 \cdot \phi(x; c^{(3)}, 0.1) \\ + 0.3 \cdot \phi(x; c^{(4)}, 0.1), \end{aligned}$$

其中

$$c^{(1)} = (0.0, 0.0), \quad c^{(2)} = (0.6, 0.6), \quad c^{(3)} = (-0.5, 0.3), \quad c^{(4)} = (0.2, -0.7),$$

且 ϕ 是光滑紧支撑鼓包函数

$$\phi(x; c, s) = \begin{cases} \exp\left(-\dfrac{1}{1 - \dfrac{|x-c|^2}{s^2}} \right), & \text{if } |x-c| < s, \\ 0, & \text{otherwise.} \end{cases}$$

这构造出了一个包含四个局部鼓包、振幅和尺度各不相同的解，如图 8(a) 所示。原点处最大的鼓包尺度为 0.8，而最小的鼓包尺度为 0.1，形成了多尺度挑战。我们采用宽度为 200、秩为 20、深度为 3 的 MMNN 架构，对应 8,241 个可学习参数，并使用 10,000 个配点进行训练。

图 8(b) 显示了 L^2 在三个任务的训练过程中，误差衰减情况存在显著差异。直接逼近函数时，MMNN 的收敛速度最快，在

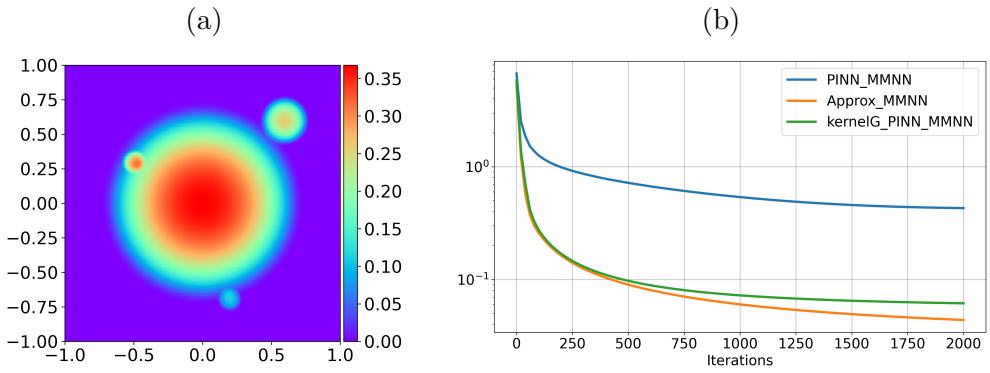


图8：示例7：(a) 精确解；(b) 不同模型在训练过程中的相对 L^2 误差。

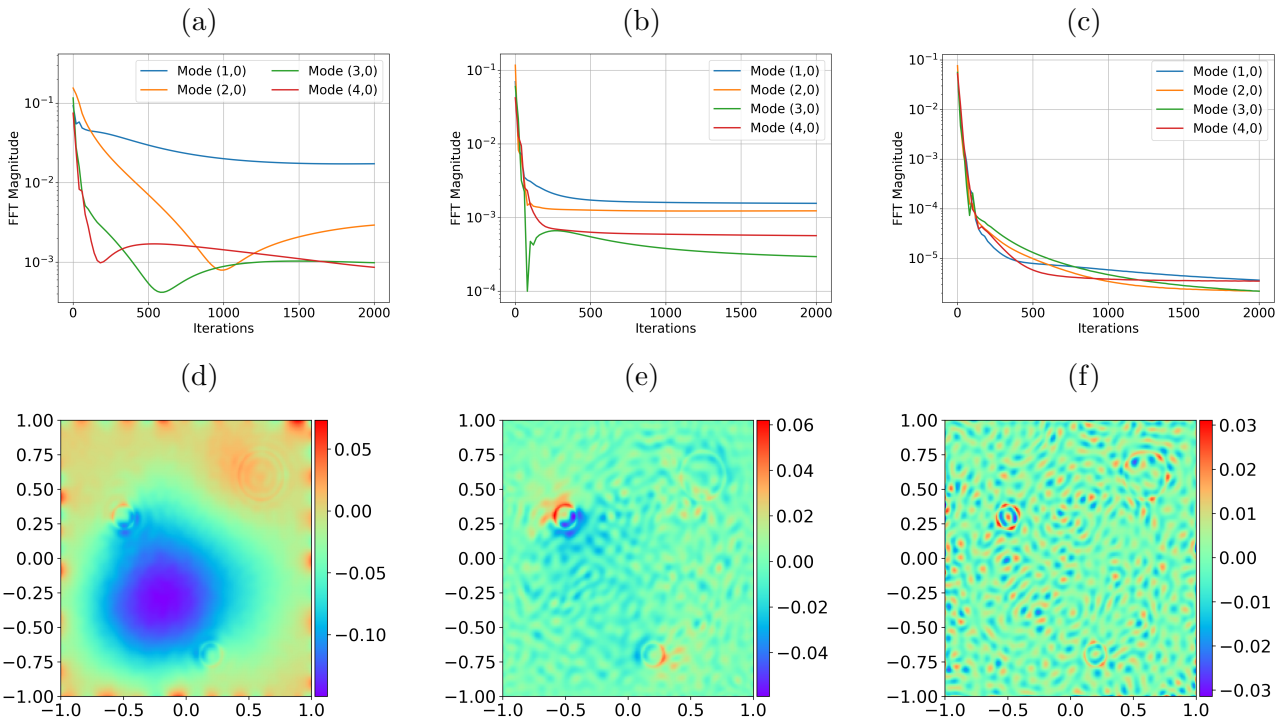


图9：示例7：PINN (a)、预条件物理信息神经网络 (b) 和近似 (c) 在选定模态中的误差；PINN (d)、预条件物理信息神经网络 (e) 和近似 (f) 的空间误差分布。

相对 L^2 误差为 4.38×10^{-2}，在 2,000 个训练轮次内达到这一结果。相比之下，标准物理信息神经网络的收敛速度慢得多，并且停滞在较大的误差水平，相对 L^2 误差为 4.30×10^{-1}。直接近似、预条件物理信息神经网络和标准物理信息神经网络的训练时间分别为 217 秒、259 秒和 256 秒。与标准物理信息神经网络相比，预条件物理信息神经网络的性能显著提升，将相对 L^2 误差降至 6.15×10^{-2}，与直接函数逼近的结果相当。这些结果及其内在机制可从图9(a)–(c) 所示的频率偏差中看出。图9(c) 显示，FMMNN 的频率偏差很小，且在所有模态上均具有出色的收敛性，这与其整体性能高度相关

中的 L^2 误差。由于 FMMNN 的表征具有弱频率偏置，在标准物理信息神经网络模型中引入微分算子会导致强烈的高频偏差，如图 9(a) 所示，其中最低频率模态的误差衰减非常缓慢。相比之下，预条件物理信息神经网络（图 9(b)）在所有模态上都呈现出均匀得多的收敛情况，与函数逼近任务类似，再次验证了我们的预条件策略的有效性。图 9(d)–(f) 所示的相应空间误差分布进一步印证了这些现象：标准物理信息神经网络的误差主要集中在最大的凸起处，而预条件物理信息神经网络和直接函数逼近产生的误差分布更加随机化，即集中在高频模态中。

本例表明，当使用具有良好高频模态及精细特征表征能力（即弱频率偏置）的、经过精心设计的神经网络来求解椭圆型偏微分方程时，适当的预条件至关重要。否则，使用基于梯度的训练，即使要捕捉一个平滑的解也会十分困难或代价高昂。

示例8： 在本示例中，我们求解单位圆盘 $\Omega = B(0,\ 1) \in \mathbb{R}^2$ 上的泊松方程 $\Delta u = -f$，并采用非齐次狄利克雷边界条件。精确解为

$$u_4(x) = 0.3 \cdot \sin\left(4\pi x_1 + 2\pi x_1 x_2\right) + 0.05 \cdot \sin\left(12\pi x_2 + 4\pi x_1 x_2\right),$$

其中 $x = (x_1,\ x_2)$。我们使用宽度为 200、秩为 40、深度为 3 的 MMNN 架构，对应 16,281 个可学习参数，并使用 7,854 个采样点进行训练。

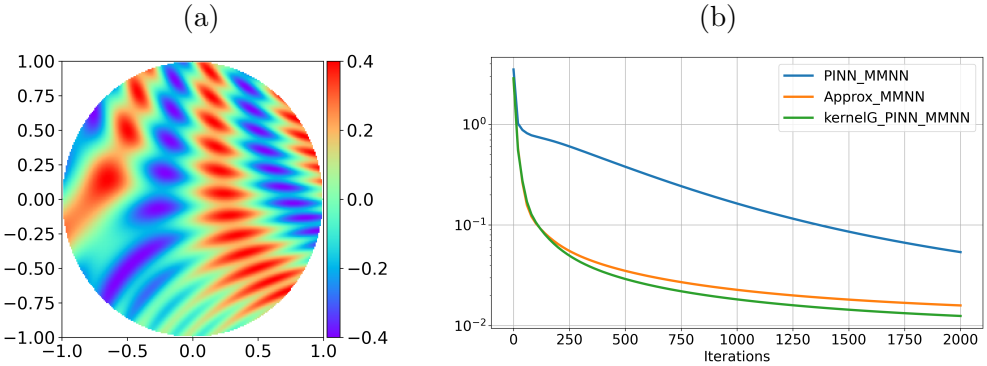


图10：E　示例8：(a) 精确解；(b) 不同模式下训练过程中的相对 L^2 误差　ls.

结果见图 10 和图11，其行为和模式与示例7相似。神经网络表示便于处理区域几何。直接近似、预条件物理信息神经网络和标准物理信息神经网络的训练时间分别为 231 秒、253 秒和 250 秒。

示例　9: 我们考虑如下二维多尺度变系数问题　ws.

$$\begin{cases} \nabla \cdot (k(x)\nabla u) = -1 & \text{in } \Omega = (-1,1)^2 \\ u = 0 & \text{on } \partial\Omega \end{cases},$$

其中 $x = (x_1,\ x_2)$，扩散系数定义为

$$k(x) = \frac{2 + A\sin\left(\dfrac{2\pi x_1}{\varepsilon}\right)}{2 + A\cos\left(\dfrac{2\pi x_2}{\varepsilon}\right)} + \frac{2 + A\sin\left(\dfrac{2\pi x_2}{\varepsilon}\right)}{2 + A\sin\left(\dfrac{2\pi x_1}{\varepsilon}\right)},$$

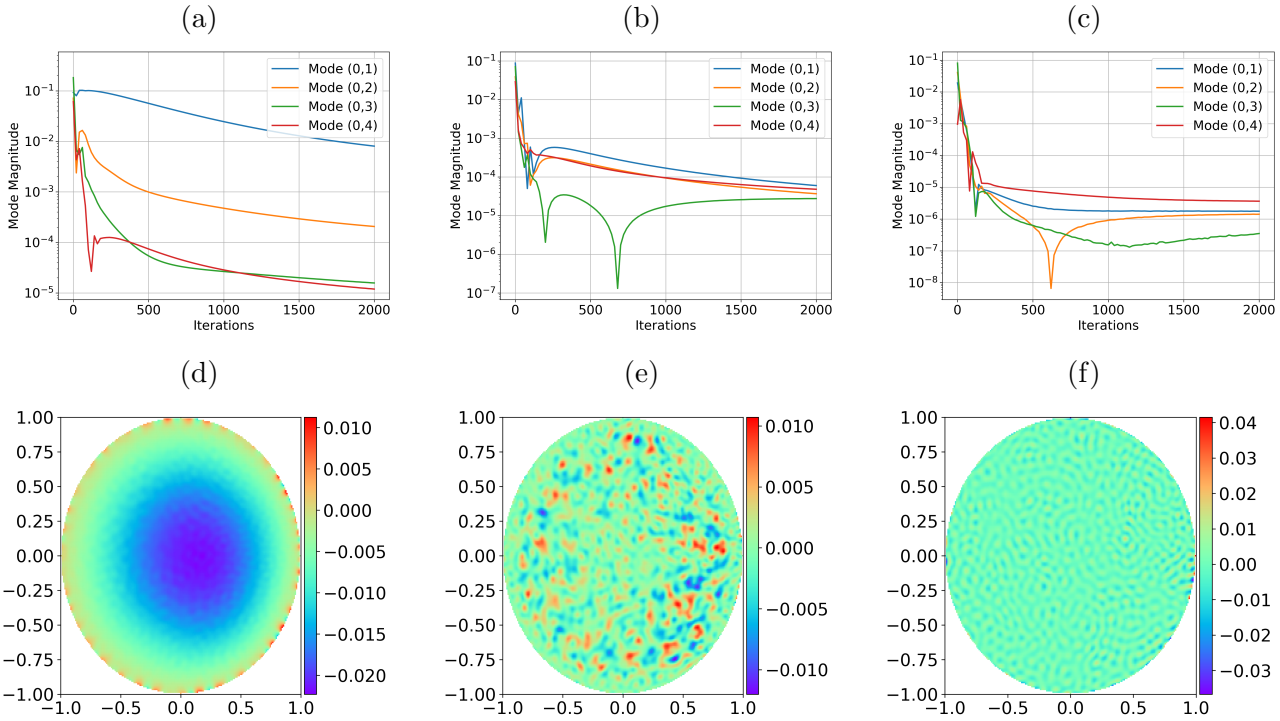


图11：示例8：物理信息神经网络（a）、预条件物理信息神经网络（b）和近似（c）在选定模态中的误差；物理信息神经网络（d）、预条件物理信息神经网络（e）和近似（f）的空间误差分布。

其中 $A = 1.8$ 且 $\varepsilon = 0.3$。这一选择使扩散系数在两个方向上都出现快速振荡。偏微分方程解（图12（a））呈现出平滑的宏观轮廓，并受到微观非均质性的调制。我们采用宽度为 200、秩为 40、深度为 3 的 FMMNN 架构，对应 16,281 个可学习参数，并使用 10,000 个采样点进行训练。

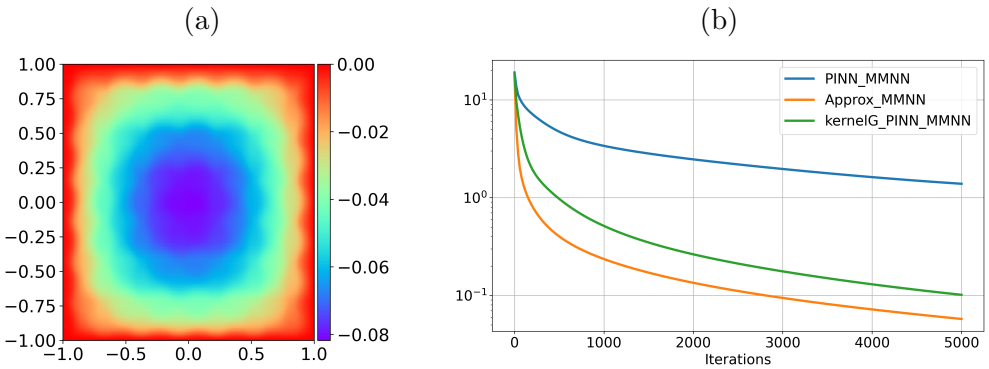


图12：示例9：（a）精确解；（b）不同模型训练过程中的相对 L^2 误差

这是一个二维多尺度特征方程的典型示例。未经预条件时，PINN 无法捕捉解；经过 5,000 个训练轮次后，相对 L^2 误差仍为 1.39。然而，预条件物理信息神经网络的准确度要高得多，相对 L^2 误差为 1.02×10^{-1}，尽管其准确度仍不及相对 L^2 误差为

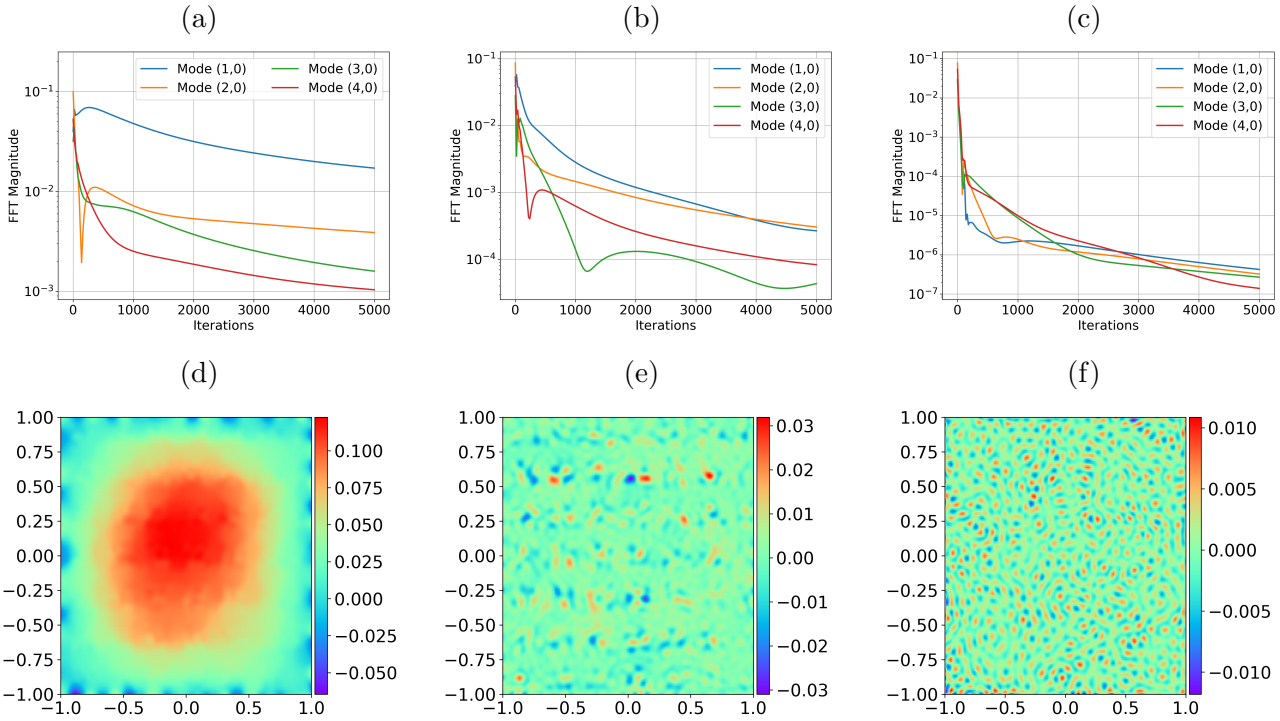


图13：示例9：PINN（a）、预条件物理信息神经网络（b）和近似（c）的选定模态误差；PINN（d）、预条件物理信息神经网络（e）和近似（f）的空间误差分布

5.75×10^{-2}。虽然所用的拉普拉斯方程自由空间格林函数并非该变系数算子的精确格林函数，但微分算子造成的频率偏差通过近似预条件得到了较好的抵消，这一点从误差模态动力学以及图13中的误差图均可看出。此示例表明，即使对于预条件子仅为近似的困难多尺度变系数问题，预条件仍然有效。

示例10： 我们求解单位球 $\Omega = B(0,1) \subset \mathbb{R}^3$ 内的泊松方程 $\Delta u = -f$，其边界条件为齐次狄利克雷边界条件，源函数 f 的选取使得精确解为

$$\begin{aligned} u(x) = 1.0 \cdot \phi(x; c^{(1)}, 0.8) \\ + 0.7 \cdot \phi(x; c^{(2)}, 0.4) \\ + 0.5 \cdot \phi(x; c^{(3)}, 0.3), \end{aligned}$$

其中

$$c^{(1)} = (0.0, 0.0, 0.0), \quad c^{(2)} = (0.5, 0.5, 0.1), \quad c^{(3)} = (-0.4, 0.3, -0.1),$$

且 ϕ 是光滑紧支撑鼓包函数

$$\phi(x; c, s) = \begin{cases} \exp\left(-\dfrac{1}{1 - \dfrac{|x-c|^2}{s^2}} \right), & \text{if } |x-c| < s, \\ 0, & \text{otherwise.} \end{cases}$$

我们采用宽度为 200、秩为 40、深度为 3 的 MMNN 架构，共包含 16,281 个可学习参数，并使用 65,450 个采样点进行训练。

在三维情形下，我们利用基于球贝塞尔函数和球谐函数的谱分解来分析频率偏差。我们考虑如下形式的基函数

$$j_\ell(z_{\ell n} r)\, Y_{\ell m}(\theta, \varphi),$$

它们在单位球内构成一个正交系。这里，j_ℓ 表示 ℓ 阶球贝塞尔函数，$Y_{\ell m}$ 表示 ℓ 次、m 阶球谐函数，$r = |x|$，而 $z_{\ell n}$ 是 j_ℓ 的第 n 个正零点。每个模态由三元组 (ℓ, m, n) 索引，其中 ℓ 控制角频率，m 指定角向，n 决定径向频率。为了研究与频率相关的学习行为，我们跟踪一组选定的模态 (ℓ, m, n)，以量化不同空间频率的学习情况。此外，我们还将空间误差可视化于截面 $z = 0$ 上。结果见图 14 和图 15，呈现出与二维情形相似的行为和模式。模态动力学和空间误差可视化结果证实，在三维情形下，近似预条件显著减轻了微分算子导致的频率偏差。

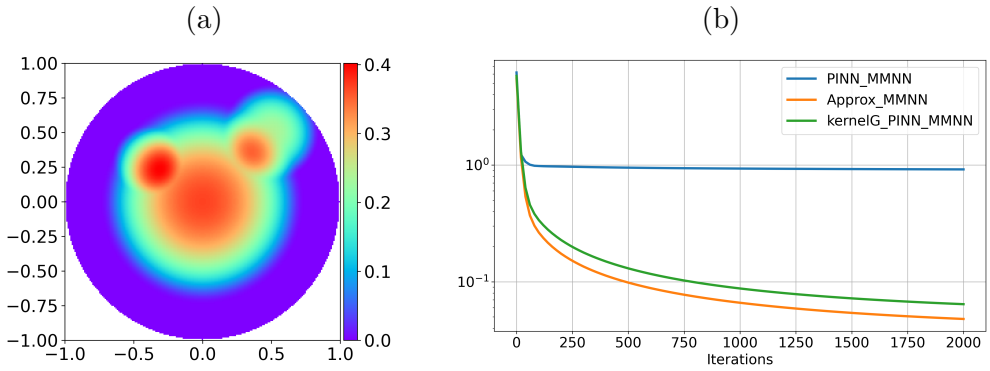


图14：示例10：(a) $z = 0$处精确解的截面；(b) 不同模型在训练过程中的相对 L^2 误差。

6 结论

在本文中，我们研究了神经网络表征中的谱偏差与微分算子之间的竞争与平衡，讨论了神经网络在训练以求解椭圆型偏微分方程时的情况。如 [HLZZ25],所示，对于多层神经网络，情况可能大不相同：经过适当设计后，它们具有强大的表示能力和较弱的频率偏置。本文表明，微分算子固有的高频偏差可能主导整体公式，从而导致训练效率低下，即使对于光滑解也是如此。

为了解决这一问题，我们提出了一种算子感知预条件策略，明确抵消微分算子的频率偏差，显著提升训练过程的效率和准确度。大量数值实验验证了该策略的有效性。

本研究的一个重要结论是，在实践中，理解谱行为至关重要，因为它与神经网络表征及其训练

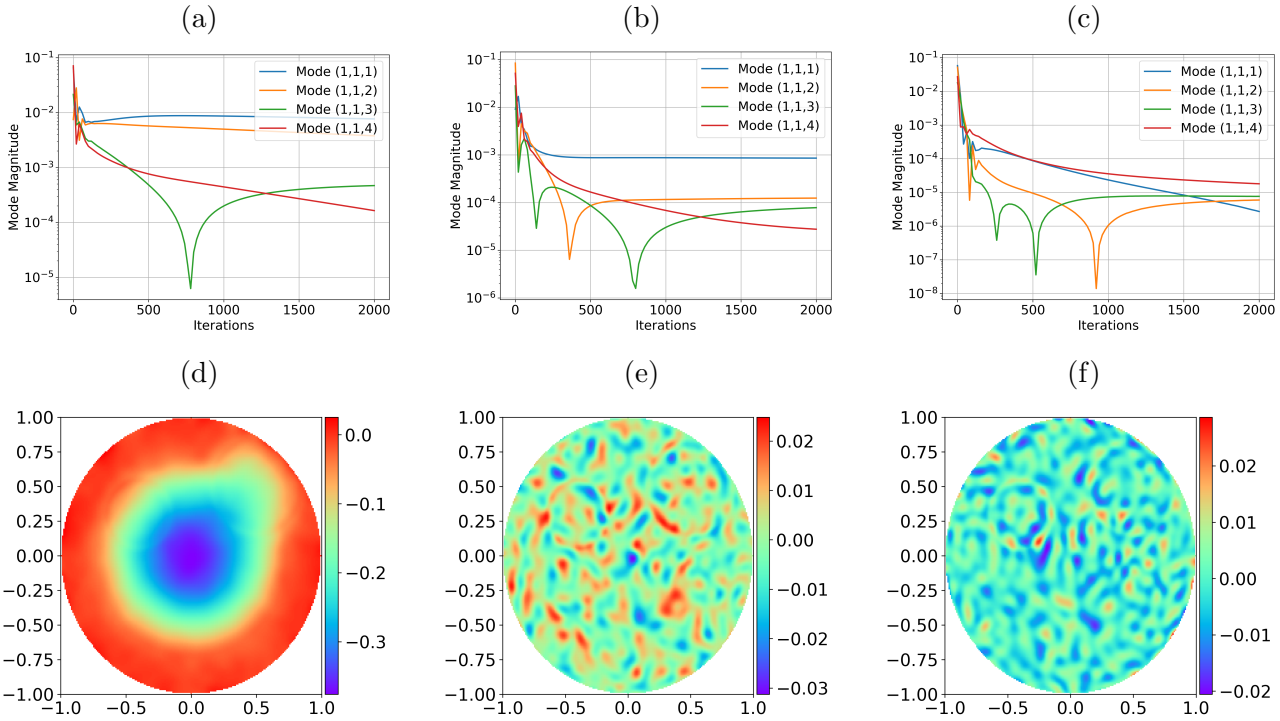


图15：示例10：PINN (a)、预条件物理信息神经网络 (b) 和近似 (c) 在选定模态中的误差；PINN (d)、预条件物理信息神经网络 (e) 和近似 (f) 的空间误差分布。

过程的效率和准确度密切相关。本文提出的算子感知视角为问题依赖型方法提供了有效思路，也引出了关于其他形式的有趣问题，例如深度 Ritz 方法以及其他变分型或基于残差的神经求解器。此外，在算子学习等更复杂的问题中，多个相互竞争的损失项（如内部残差、边界误差和算子评估）可能表现出不同的谱行为，从而导致复杂的训练动力学。设计能够实现整体谱平衡的有效预条件策略至关重要。

致谢

Hongkai Zhao 部分获得美国国家科学基金会通过项目 DMS- 2309551 提供的资助。Yimin Zhong 部分获得美国国家科学基金会通过项目 DMS-2309530 提供的资助。Roy Y. He 部分获得国家自然科学基金委员会项目 12501594、香港研究资助局与法国驻香港总领事馆共同资助的法国—香港联合研究计划 (F-CityU101/24)、香港城市大学提供的 StUp - CityU 7200779，以及香港研究资助局新进学者计划资助 21309625 的支持。

参考文献

[ATSK24] Sokratis J Anagnostopoulos、Juan Diego Toscano、Nikolaos Stergiopulos 和 George Em Karniadakis。物理信息神经网络中的基于残差的注意力——

网络。《计算力学与工程方法》，421:116805，2024年。[Bar02] Andrew R Barron。S形函数叠加的通用逼近界。IEEE信息论汇刊，39(3):930–945，2002年。[BD20] David GT Barrett 和 Benoit Dherin。隐式梯度正则化。arXiv预印本 arXiv:2009.11162，2020年。[BGG+20] Ronen Basri、Meirav Galun、Amnon Geifman、David Jacobs、Yoni Kasten 和 Shira Kritchman。非均匀密度输入下神经网络的频率偏差。载于国际机器学习大会，685–694页。PMLR，2020年。[CFW+21] Yuan Cao、Zhiying Fang、Yue Wu、Ding-Xuan Zhou 和 Quanquan Gu。走向理解深度学习的谱偏差。载于第三十届国际人工智能联合会议（IJCAI 2021），2205–2211页。国际人工智能联合会议，2021年。[CGP06] Shivkumar Chandrasekaran、Ming Gu 和 Timothy Pals。用于结构化矩阵分层半可分（HSS）表征的快速稳定求解器。SIAM矩阵分析与应用期刊，28(3):817–839，2006年。[Cyb89] George Cybenko。通过S形函数叠加进行近似。《控制、信号与系统数学》，2:183–192，1989年。[DHP21] Ronald DeVore、Boris Hanin 和 Guergana Petrova。神经网络近似。《数值学报》，30:327–444，2021年。[DJL+22] Chenguang Duan、Yuling Jiao、Yanming Lai、Dingwei Li、Jerry Zhijian Yang 等。深度 Ritz 方法的收敛速率分析。《计算物理通讯》，31(4):1020–1048，2022年。[DRBMdB23] Tim De Ryck、Florent Bonnet、Siddhartha Mishra 和 Emmanuel de B´ezenac。从算子预条件视角看物理信息机器学习中的训练。arXiv预印本 arXiv:2310.05801，2023年。[DRMM24] Tim De Ryck、Siddhartha Mishra 和 Roberto Molinaro。wpinns：用于逼近双曲守恒律熵解的弱物理信息神经网络。SIAM数值分析期刊，62(2):811–841，2024年。[ES16] Ronen Eldan 和 Ohad Shamir。深度对前馈神经网络的作用。载于第29届学习理论年度会议（COLT），机器学习研究论文集第49卷，907–940页。PMLR，2016年。[EY18] Weinan E 和 Bing Yu。深度 Ritz 方法：一种用于求解变分问题的基于深度学习的数值算法。《数学与统计学通讯》，6(1):1–12，2018年。[GR87] Leslie Greengard 和 Vladimir Rokhlin。粒子模拟的快速算法。《计算物理学杂志》，73(2):325–348，1987年。[HLZZ25] Roy Y He、Ying Liang、Hongkai Zhao 和 Yiming Zhon。使用浅层神经网络求解偏微分方程时可以期待什么？arXiv预印本 arXiv:2510.27658，2025年。

[HSTX22] Qingguo Hong、Jonathan W Siegel、Qinyang Tan 和 Jinchao Xu。论激活-函数对神经网络谱偏差的依赖性。*arXiv* 预印本 *arXiv:2208.04924*，2022年。

[JGH18] Arthur Jacot、Franck Gabriel 和 Cl´ement Hongler。神经正切核：收敛-收敛性与神经网络的泛化。神经信息处理系统进展系统，31，2018。

[KB15] Diederik P. Kingma 和 Jimmy Ba。Adam：一种随机优化方法。载于 Yoshua Bengio 和 Yann LeCun，编，第三届国际学习会议表征，*ICLR 2015*，圣迭戈，*CA*，美国，*2015*年*5*月*7*日至*9*日，会议专题论文集，2015年。

[KL20] Patrick Kidger 和 Terry Lyons。深窄网络的通用逼近——论文。在第*33*届学习理论会议（*COLT*）论文集，第125卷，载于机器学习研究论文集，第2306–2327页。PMLR，2020年。

[Kra14] Ilia Krasikov。带显式误差项的贝塞尔函数和艾里函数近似误差项。伦敦数学会计算与数学期刊，17(1)：209–225，2014年。

[LBY22] Yiping Lu、Jose Blanchet 和 Lexing Ying。Sobolev 加速与统计通过梯度下降学习椭圆方程的最优性。神经信息处理系统进展，35:33233–33247，2022年。

[LJP+21] Lu Lu、Pengzhan Jin、Guofei Pang、Zhongqiang Zhang 和 George Em Karniadakis。基于算子的通用逼近定理，通过 DeepONet 学习非线性算子算子定理。《自然·机器智能》，3(3):218–229，2021。

[LKA+20] Zongyi Li、Nikola Kovachki、Kamyar Azizzadenesheli、Burigede Liu、Kaushik Bhattacharya、Andrew Stuart 和 Anima Anandkumar。傅里叶神经算子用于参数化偏微分方程的傅里叶神经算子。*arXiv*预印本 *arXiv:2010.08895*，2020年。

[LLW21] Yulong Lu、Jianfeng Lu 和 Min Wang。深度求解高维椭圆偏微分方程的Ritz方法。载于学习理论会议，第3196–3241页。PMLR，2021。

[LPW+17] Zhou Lu、Hongming Pu、Feicheng Wang、Zhiqiang Hu和Liwei Wang。表达能力：从宽度看神经网络。载于神经领域进展信息处理系统（*NeurIPS*），2017。

[MM23] Siddhartha Mishra和Roberto Molinaro。关于泛化误差的估计用于逼近偏微分方程的物理信息神经网络。《*IMA*数值-分析杂志，43(1):1–43，2023。

[MZ19] Johannes M¨uller和马里乌斯·蔡因霍费尔。重新审视深度 Ritz。*arXiv* 预印本 *arXiv:1912.03937* ，2019年。

[MZ22] Johannes M¨uller 和马里乌斯·蔡因霍费尔。深度 Ritz 方法的误差估计带边界惩罚。在数学与科学机器学习，第215–230。PMLR ，2022年。

[RBA+19] Nasim Rahaman、Aristide Baratin、Devansh Arpit、Felix Draxler、Min Lin、Fred A. Hamprecht、Yoshua Bengio 和 Aaron Courville。论神经网络的谱偏差。2019年。[RPK19] Maziar Raissi、Paris Perdikaris 和 George E Karniadakis。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和反问题的深度学习框架。《计算物理学杂志》，378:686–707，2019年。[RYK20] Maziar Raissi、Alireza Yazdani 和 George Em Karniadakis。隐藏的流体力学：从流动可视化中学习速度场和压力场。《科学》，367(6481):1026–1030，2020年。[SDK20] Yeonjong Shin、Jerome Darbon 和 George Em Karniadakis。论物理信息神经网络的收敛与泛化。*arXiv*电子预印本，页码 arXiv–2004，2020年。[SJHH21] Hwijae Son、Jin Woo Jang、Woo Jin Han 和 Hyung Ju Hwang。面向物理信息神经网络的索博列夫训练。*arXiv*预印本 *arXiv:2101.08932*，2021。[SYZ22] Zuowei Shen、Haizhao Yang 和 Shijun Zhang。深度网络逼近：使用固定数量的神经元实现任意精度。《机器学习研究杂志》，23(276):1–60，2022年。[Sze39] Gabor Szeg。正交多项式，第23卷。美国数学会，1939年。[TSM+20] Matthew Tancik、Pratul Srinivasan、Ben Mildenhall、Sara Fridovich-Keil、Nithin Raghavan、Utkarsh Singhal、Ravi Ramamoorthi、Jonathan Barron 和 Ren Ng。傅里叶特征使网络能够在低维域中学习高频函数。神经信息处理系统进展，33:7537–7547，2020年。[TYZZ25] Tao Tang、Jiang Yang、Yuxiang Zhao 和 Quanhui Zhu。基于秩的结构化首层初始化预训练技术，以加速训练过程。*arXiv*预印本 *arXiv:2507.11962*，2025年。[WWP21] Sifan Wang、Hanwen Wang 和 Paris Perdikaris。论傅里叶特征网络的特征向量偏差：从回归到使用物理信息神经网络求解多尺度偏微分方程。计算力学与工程中的计算方法，384:113938，2021。[YBZ04] Lexing Ying、George Biros 和 Denis Zorin。二维和三维中的核无关自适应快速多极算法。《计算物理学杂志》，196(2):591–626，2004。[ZBYZ20] Yaohua Zang、Gang Bao、Xiaojing Ye 和 Haomin Zhou。面向高维偏微分方程的弱对抗网络。《计算物理学杂志》，411:109409，2020年。[ZZZZ25a] Shijun Zhang、Hongkai Zhao、Yimin Zhong 和 Haomin Zhou。傅里叶多分量多层神经网络：释放高频潜力。*arXiv*预印本 *arXiv:2502.18959*，2025年。

[ZZZZ25b] Shijun Zhang、Hongkai Zhao、Yimin Zhong 和 Haomin Zhou。结构化且均衡的多分量多层神经网络。*SIAM*科学计算杂志，47(5):C1059–C1090，2025年。

[ZZZZ25c] Shijun Zhang、Hongkai Zhao、Yimin Zhong和Haomin Zhou。浅层网络为何难以近似并学习高频。信息与推断：*IMA*期刊，14(3):iaaf022，2025年。