收稿日期：制作时添加 修订日期：制作时添加 接收日期：制作时添加DOI：xxx/xxxx

研究论文

物理信息引导的正交特征学习及其在求解偏微分方程中的应用

Qianxing Jia[1] | **Dong Wang**[1,2,3]

[1]香港中文大学（深圳）理工学院，广东，518172，中国

[2]深圳国际工业与应用数学中心，大数据研究院，广东，518172，中国

[3]深圳环湾区研究院，广东，518048，中国

通信作者
Dong Wang，香港中文大学（深圳）理工学院，中国广东省深圳市 518172。电子邮箱：wangdong@cuhk.edu.cn

arXiv:2602.03247v1 [math.NA] 3 Feb 2026

摘要

随机特征方法（RFM）通过从通用分布中初始化特征来构造近似空间，具有普适逼近性质，可用于求解一般偏微分方程。然而，这种标准初始化方式缺乏对底层物理定律和几何形状的认知，从而限制了逼近能力。本文提出物理驱动正交特征方法（PD-OFM），这是一种构造特征表示的框架：通过结合物理信息驱动目标与正交正则化对特征进行预训练，使其明确适配微分算子和计算域。这种预训练策略能够生成近乎正交的特征基。我们从理论和实证两方面证明，物理信息预训练能够提升所学习特征空间的逼近能力。将该方法用于求解亥姆霍兹方程、泊松方程、波动方程和纳维–斯托克斯方程时，其残差误差比可比方法低 2–3 个数量级。此外，正交正则化提升了可迁移性，使预训练特征能够针对同一偏微分方程，在不同源项和区域几何之间有效泛化。

关键词

随机特征方法，物理信息驱动神经网络，正交性

1 引言

偏微分方程的数值求解是众多学科的基础，包括传热、流体动力学、电磁学和金融建模。有限元法、有限差分法和谱方法等经典数值方法，对于边界条件明确定义的结构化区域十分有效。然而，这些方法受维数灾难的限制，而且在复杂几何中需要耗费大量人力进行网格划分。物理信息神经网络（PINN）[1,2,3,4,5] 已成为一种利用机器学习求解偏微分方程的替代框架。PINN通过训练神经网络，使其最小化控制方程的残差，从而近似偏微分方程的解。这种方法无需网格，适用于高维问题和不规则区域。尽管具有这些优势，物理信息神经网络的损失景观病态，导致随机梯度下降收敛缓慢，并且常陷入局部极小值[6]。物理信息神经网络会优化在随机采样的配置点处计算得到的偏微分方程残差均方。其他方法则采用不同的损失函数形式。例如，深度 Ritz 方法[7] 和变分物理信息神经网络[8,9] 等变分方法通过能量原理重新表述偏微分方程，通常能提高高维问题中的稳定性，并降低微分阶数，从而减少自动微分的计算成本[10]。另一些方法采用弱形式，例如弱对抗网络[11,12]，，已证明其在边界条件不光滑的问题中具有鲁棒性。在架构创新方面，多项研究探索了改进网络结构以提升准确性。Kolomogorov-Arnold 网络引入了可学习的单变量非线性函数，从而提高可解释性和近似准确性[13,14,15]。张量神经网络将网络结构设计为各维度上的单变量函数的组合，从而能够进行准确的基于求积的积分[16,17,18]。另一条独特的研究路线着重于将网络最后一层的优化与其余参数的优化解耦，从而把原本的非凸优化问题转化为关于最后一层系数的凸优化问题。将神经网络视为隐藏层特征的线性组合，

 © 2026年版权所有者名称

便可通过最小二乘法求解最后一层系数。这一思路催生了两大类方法：一类采用随机初始化，例如局部极限学习机[19,20,21]、随机特征方法[22,23,24]、随机神经网络[25,26]和 TransNet[27,28]；另一类采用经过训练的神经网络，包括子空间神经网络[29,30,31]、最小二乘VPINN[9]和 Deepoly[32]。相比随机初始化，训练的优势尚不明确。本文从理论和实证两方面证明，尽管训练特征的计算成本较高，但能带来更高的准确性。受经典谱方法的启发——这类方法得益于具有指数收敛率的正交基函数——本文提出一种称为 PD-OFM 的新框架，即物理驱动正交特征方法。我们的方法旨在学习彼此正交的算子自适应特征。在训练过程中，对特征偏离正交矩阵的程度施加惩罚，引导网络构建近似正交的特征空间。训练完成后，通过对学习到的特征所张成的空间进行最小二乘投影来求解偏微分方程。此外，本文引入投影误差，用以量化学习到的特征张成空间与微分算子特征空间之间的距离。在训练期间对正交性进行正则化，可以降低投影误差。一维泊松问题的实验表明，PD-OFM 的特征与解析特征函数高度近似，并取得了高精度。此外，可迁移性有助于降低重复求解偏微分方程的计算成本。在人工智能领域，冻结预训练特征空间，再在后续任务中进行特征提取或微调，是一种常见做法[33,34]。这一思路后来也成功应用于偏微分方程[27,28,35,36]和参数偏微分方程[37,38]的求解。此外，算子学习[39,40,41,42]具有可迁移性，能够将学习到的映射应用于新的输入函数或区域。所提出的 PD-OFM 也具有很强的可迁移性。预训练特征还可迁移用于求解源项不同或区域不同的同类偏微分方程。本文的主要贡献总结如下：

- 理论分析和数值实验表明，物理信息特征训练能够提升所学习特征空间的逼近质量，从而提高后续最小二乘求解过程的准确性。
- 本文引入正交性正则化损失，以确保神经网络倒数第二层特征矩阵生成一组近似正交的特征。
- 全面的数值实验将 PD-OFM 与随机特征方法及可迁移神经网络进行了比较，凸显了其在准确性、稳定性和可迁移性方面的优势。

本文其余部分安排如下。第 2.1–2.2 节回顾随机特征方法，并构建物理驱动正交特征方法的框架。第 2.3 节讨论理论见解，以及对比训练后的表示函数与随机表示函数的实验证据。第 2.4 节通过投影误差和有效秩等指标评估 PD-OFM。第 2.5 节探讨 PD-OFM 基在不同偏微分方程实例间的可迁移性。第 3 节给出数值评估结果，并与现有方法进行比较。

2 物理驱动的正交特征方法

2.1 随机特征方法

用于求解偏微分方程（PDE）的随机特征方法建立在以下基础上：(i) 用随机特征的线性组合表示偏微分方程的解，以及 (ii) 通过配点法施加控制方程。随机特征通常利用随机初始化的浅层神经网络构造。因此，该方法实现了经典数值技术与现代机器学习方法的自然融合。考虑如下形式的一般偏微分方程

$$\begin{cases} \mathcal{L}u(x) = f(x), & x \in \Omega, \\ \mathcal{B}u(x) = g(x), & x \in \partial\Omega, \end{cases}$$

其中 $x = (x_1, \cdots, x_d)^T$，且 Ω 是一个有界连通区域。随机特征逼近假设 s

$$u(x) = \sum_{i=1}^{N} c_i \, u_i(x),$$

并通过最小化惩罚最小二乘损失来确定系数 $\{c_i\}$：

$$\mathcal{L}(c) = \frac{1}{N_r}\sum_{i=1}^{N_r}\left|\mathcal{L}u(x_r^{(i)}) - f(x_r^{(i)})\right|^2 + \frac{\lambda_b}{N_b}\sum_{j=1}^{N_b}\left|\mathcal{B}u(x_b^{(j)}) - g(x_b^{(j)})\right|^2, \tag{1}$$

其中，$\{x_r^{(i)}\}$ 和 $\{x^{(j)}{}_b\}$ 分别为内部和边界配置点。尽管该损失函数的结构与物理信息神经网络（PINN）中的损失函数相同，但优化过程存在显著差异：所有随机特征 $u_i(x)$ 均固定不变，仅优化系数 c_i。因此，无需使用梯度下降，可以采用线性或非线性最小二乘等直接求解器：

$$c = \arg\min_c \mathcal{L}(c).$$

每个随机特征通常具有以下形式

$$u_i(x) = \sigma(w_i x + b_i),$$

其中，参数 (w_i, b_i) 根据标准神经网络初始化方案进行采样，例如均匀初始化、Xavier 初始化或 Kaiming 初始化[43,44]。这对应于一种仅输出层可训练的浅层神经网络。此外，偏微分方程解通常呈现显著的局部结构，而采用标准神经网络初始化方案的全局定义随机特征可能缺乏足够的逼近能力。为解决这一局限，随机特征方法通过单位分解（PoU）引入具有局部支撑的特征。为构造单位分解，对一组点 $\{x_n\}_{n=1}^{M_p} \subset \Omega$ 进行采样，作为单位分解中各分量的中心。对于每个n，构造归一化坐标：

$$\tilde{x}_n = \frac{1}{r_n}(x - x_n), \quad n = 1, \cdots, M_p,$$

其中r_n 是预先选定的。当$d = 1$时，单位分解函数为以下之一：

$$\psi_n^a(x) = \mathbb{I}_{-1\le\tilde{x}<1},$$

$$\psi_n^b(x) = \begin{cases} \dfrac{1+\sin(2\pi\tilde{x})}{2}, & -\dfrac{5}{4} \le \tilde{x} < -\dfrac{3}{4}, \\ 1, & -\dfrac{3}{4} \le \tilde{x} < \dfrac{3}{4}, \\ \dfrac{1-\sin(2\pi\tilde{x})}{2}, & \dfrac{3}{4} \le \tilde{x} < \dfrac{5}{4}, \\ 0, & \text{otherwise.} \end{cases}$$

高维单位分解通过一维单位分解函数的张量积构造而成 $\psi_n(x) = \Pi_{k=1}^d \psi_n(x_k)$。综上，偏微分方程的解可表示为：

$$u_M(x) = \sum_{n=1}^{M_p}\psi_n(x)\sum_{j=1}^{J_n} c_{nj}u_{nj}(\tilde{x}_j).$$

对于多尺度问题，近似同时包含全局分量和局部分量 s:

$$u_M(x) = \sum_{i=1}^{N} c_i^g u_i^g(x) + \sum_{n=1}^{M_p}\psi_n(x)\sum_{j=1}^{J_n} c_{nj}u_{nj}(\tilde{x}_j).$$

随机特征方法无疑是一种强大的工具，在求解一大类偏微分方程方面已取得显著成效，包括具有复杂几何、动力学表述和多尺度结构的偏微分方程。尽管具有诸多优势，它仍面临若干重要挑战。

首先，尽管随机特征方法将偏微分方程改写为线性系统，但所得矩阵往往具有严重的病态性，条件数甚至高达 10^{16}。这种病态性会显著降低迭代求解器的性能。尽管原始研究所考虑的线性系统规模相对较小，但降低

条件数对于将该方法扩展至更大规模的问题至关重要。Chen 和 Tan[22] 近期的研究引入了随机预条件器来解决这一问题，而 Shang 等人[26,45] 则采用 Petrov-Galerkin方法和间断Galerkin方法的表述，而非标准配置点方法，以提高稳定性。第二个挑战在于随机特征的初始化。Dong[19,20,46]最早提出利用神经网络初始化技术来初始化特征。此后，许多文献致力于研究特征初始化，包括随机特征方法。此外，TransNet[28] 提出了一种在物理域内分布更均匀的初始化方法。其他方法则通过特征预训练来适应不同的偏微分方程，包括子空间神经网络[31]，、自适应神经子空间方法[29]，和 Deepoly[32]。值得注意的是，这两个挑战并非相互排斥。在本研究中，我们提出了一种以物理信息驱动的方式对随机特征进行预训练的框架，并辅以正交正则化，从而同时提升特征质量并缓解病态性。

2.2 采用正交正则化进行预训练

如图 1 所示，我们采用了一种改进的全连接神经网络架构。该网络通过一系列隐藏层，将输入物理坐标 $x = (x_1, x_2, \ldots, x_d) \in \Omega$ 映射至中间特征空间。网络包含一个中间特征空间。神经元数量（记为 m）被视为控制特征数量的超参数。网络产生两个不同的输出：特征空间层输出

$$U(x;\theta) = [u_1(x;\theta), u_2(x;\theta), \cdots, u_m(x;\theta)]^T \in \mathbb{R}^m$$

以及最终解 $u(x;\theta) = \sum_{i=1}^{m} c_i u_i(x;\theta)$，其中 c_i 是最后一层不带偏置的系数。该输出允许对特征空间层应用显式正则化。

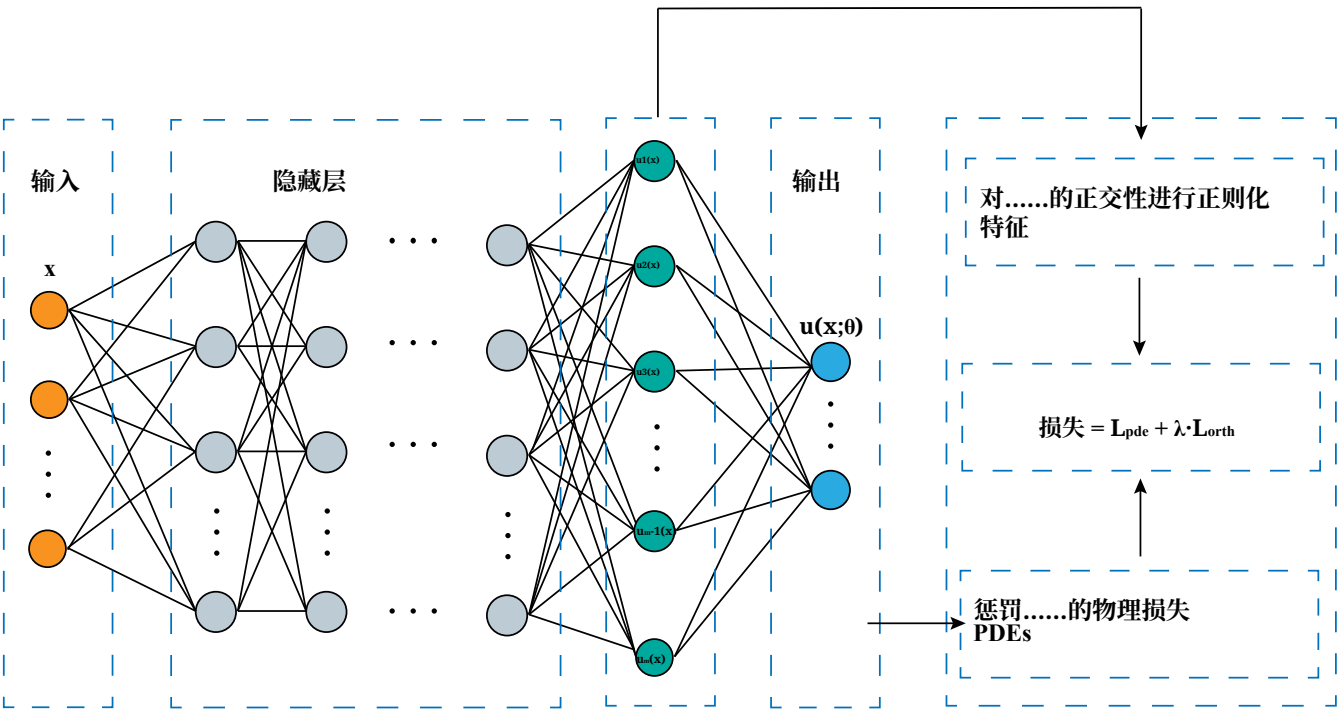


图 1 该网络有两个并行的输出分支：一个预测解 $u(x;\theta)$，另一个输出特征空间（绿色节点）。

预训练过程通过梯度下降，同时根据偏微分方程损失和正交性正则化最小化所有权重和偏置：

$$\begin{aligned}\mathcal{L}_{\text{PINN}} &= \frac{1}{N_r}\sum_{i=1}^{N_r}\left|\mathcal{L}u(x_r^{(i)};\theta)-f(x_r^{(i)})\right|^2+\frac{1}{N_b}\sum_{j=1}^{N_b}\left|\mathcal{B}u(x_b^{(j)};\theta)-g(x_b^{(j)})\right|^2\\ \mathcal{L}_{\text{orth}} &= \left\|U^\top U-I\right\|_F^2\\ \mathcal{L}_{\text{total}} &= \mathcal{L}_{\text{PINN}}+\lambda_{\text{orth}}\cdot\mathcal{L}_{\text{orth}}.\end{aligned} \tag{2}$$

算法 1 PD-OFM：物理驱动正交特征方法

要求：微分算子 $\mathcal{L}$、边界算子 $\mathcal{B}$、源函数 f、边界数据 g、配置点 $\{x_r^{(i)}\}$ 和 $\{x^{(j)}\ b\}$、容差 ε确保：近似解 u(x) 1：使用宽度为 m 的特征空间层初始化神经网络 $u(x;\theta)$ 2：重复 3：采样内部和边界配置点 4：计算总损失：$\mathcal{L}_{\text{total}}=\mathcal{L}_{\text{PDE}}+\lambda_{\text{orth}}\cdot\mathcal{L}_{\text{orth}}$5：通过梯度下降更新参数 ϕ 6：直至满足 $\mathcal{L}_{\text{total}}<\varepsilon$ 或收敛 7：固定特征空间层的输出 $U(x)=[u_1(x),\ldots,u_m(x)]^T$8：在所有配置点处计算 $\mathcal{L}u_j(x_r^{(i)})$ 和 $\mathcal{B}u_j(x^{(j)}\ b)$，构建系统矩阵 A 9：使用 $f(x_r^{(i)})$ 和 $g(x^{(j)}\ b)$ 组装右端项 b 10：求解最小二乘问题：$c^*=\arg\min_c\|Ac-b\|_2^2$11：返回近似解 $u(x)=U(x)^\top c^*$

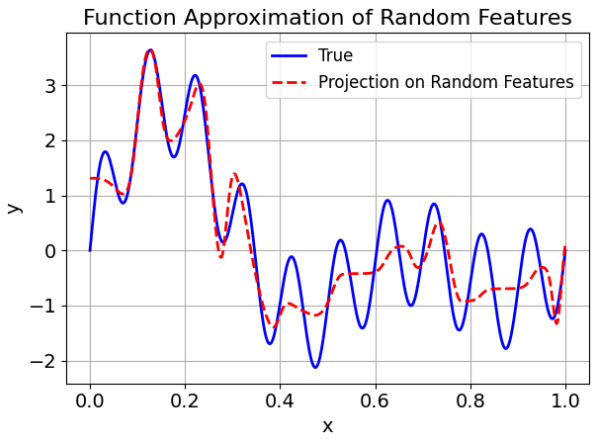
注*1*。当 $\lambda_{\text{orth}}=0$ 时，算法 1 将退化为 PD-FM（物理驱动特征方法），这等价于在物理信息神经网络训练后对最终线性系数进行重新优化。

PD-OFM 的动机源于随机初始化特征空间在逼近具有复杂频谱成分的函数时存在的内在局限性。特别是，当特征数量相对于目标函数的复杂度较少时，随机特征的逼近准确性会显著下降。为说明这一局限性，我们考虑定义在 $[0,1]$ 上的以下一维函数：

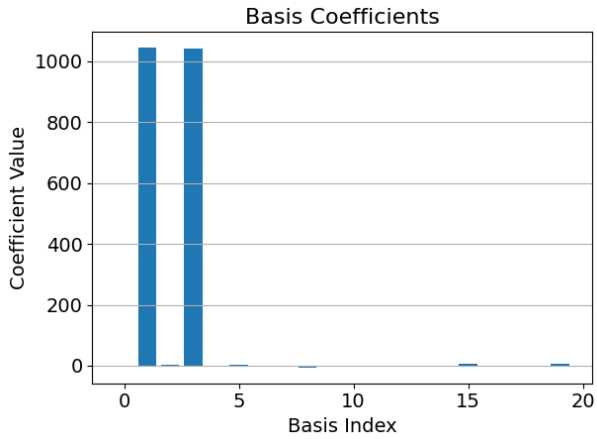
$$f(x)=\sin(2\pi x)+\sin(3\pi x)+\sin(4\pi x)+\sin(20\pi x),$$

该函数包含低频和高频成分的混合，因此可作为低维逼近空间的挑战性测试案例。为评估随机特征的内在逼近能力，我们在函数空间中通过直接最小二乘逼近，将 f 投影到一个包含 20 个特征的随机生成特征空间。如图2a所示，所得投影未能捕捉目标函数的基本结构。此外，图2b显示，最优系数呈现出高度不均匀的分布，其中一小部分特征主导了逼近过程。相比之下，当使用监督损失结合正交正则化对特征进行预训练时，会呈现出截然不同的行为。如图3所示，学习到的特征形成了类似正弦模态的振荡模式，表明训练过程诱导出了更具表达力且结构化程度更高的特征空间。相应地，逼近质量得到提升。如图4所示，与随机特征方法相比，PD-OFM 随特征数量增加，能够持续更快地降低相对 L^2 误差。

该示例表明，物理信息预训练结合正交正则化，能够更高效地利用有限宽度表示。



(a) 将f投影到随机特征空间。



(b) 与随机特征相关的系数分布。

图 2 随机特征的逼近行为。投影高度依赖少量特征，表明特征利用效率低下。

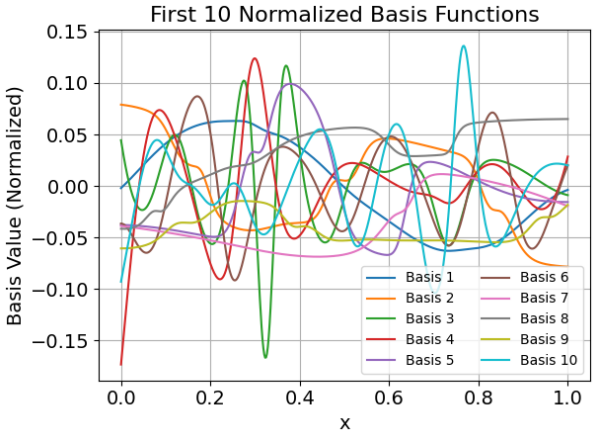


图 3 PD-OFM 学到的前 10 个归一化基函数。学到的特征仍然分布均匀且不退化，体现出有效的特征利用。

2.3 训练确实重要

前文各节通过静态比较考察了不同特征构造的逼近行为。现在，讨论转向动态视角，旨在探究逼近能力在训练过程中如何变化。前面的示例均被表述为监督逼近问题，并通过在函数空间中直接进行最小二乘拟合来评估逼近质量。然而，在求解偏微分方程时，逼近由基于算子的损失所支配，此时通过最小化微分方程的残差来求解，而非逐点匹配目标函数。这种表述方式的变化促使我们研究算子约束设置下的训练动态。作为一个典型案例，考虑一维亥姆霍兹方程：

$$\Delta u - k^2 u = f, \tag{3}$$

其中 $k = \sqrt{10}$，源项f由精确解推导得出

$$u(x) = \sin\left(3\pi x + \frac{3\pi}{20}\right)\cos\left(2\pi x + \frac{\pi}{10}\right) + 2.$$

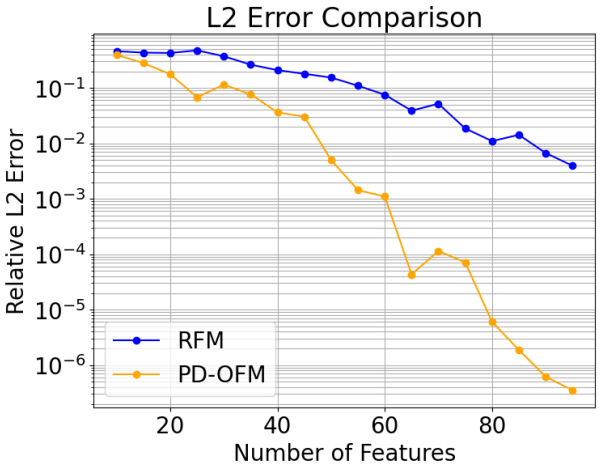


图 4 PD-OFM 与随机特征方法（RFM）的逼近准确性比较。随着特征数量增加，PD-OFM 的相对 L^2 误差持续更快地衰减，并在更大宽度下达到显著更高的准确性。

在整个实验中，采用固定的神经网络架构，由两个各含 100 个神经元的隐藏层组成。采用残差连接[47] 以及 $\tanh^3(\cdot)$ 激活函数[11]。为考察所观察到的训练行为的稳健性，选取了两种广泛使用的初始化策略：Xavier 初始化[44] 和 Kaiming 初始化[43]。

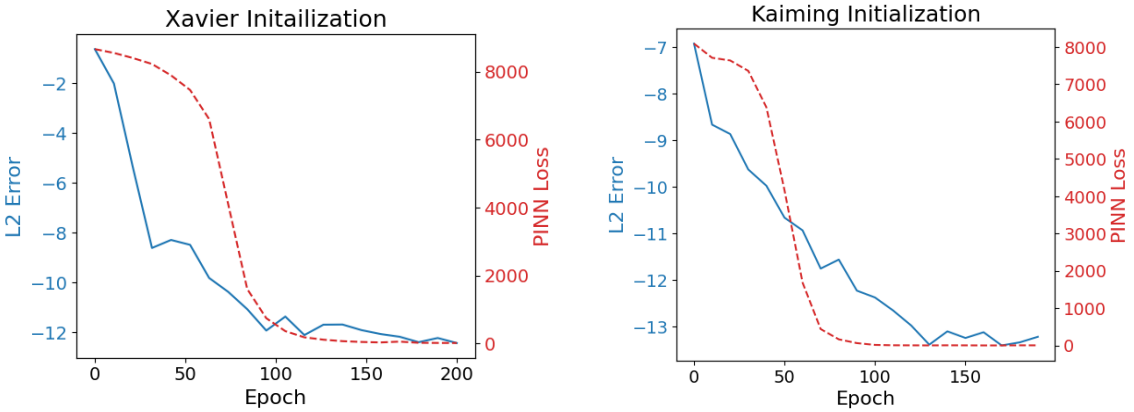


图 5 物理信息神经网络损失与最小二乘解误差之间的训练动态比较。

如图 5 所示，PINN 残差损失的演化与最小二乘解的准确性之间存在明显相关性。在训练早期，物理信息神经网络损失的下降伴随着最小二乘逼近误差的大幅降低。随着训练推进，物理信息神经网络损失趋于平台期，最小二乘解的进一步改进变得有限，逼近误差也同样趋于稳定。不同初始化策略下均观察到一致的行为。未经训练时，Xavier 初始化导致明显较差的逼近性能，而 Kaiming 初始化表现相对更好。然而，仅经过 100 个训练轮次后，两种初始化方案均获得显著改进，二者之间的差异大幅缩小。这一观察表明，训练过程会逐步优化特征空间，从而增强所学特征的逼近能力。

定理 1 设 $A : L^2(\Omega) \to L^2(\Omega)$ 是一个有界稳定线性算子，即存在一个常数 $\gamma > 0$ 使得

$$\|u\|_{L^2} \leq \gamma \|Au\|_{L^2}, \quad \forall u \in L^2(\Omega). \tag{4}$$

设 u_* 是方程 $Au = f$的精确解，且 u_{NN} 是神经网络近似。则，

$$\|u_* - u_{NN}\|_{L^2} \le \gamma \|Au_{NN} - f\|_{L^2}. \tag{5}$$

证明。 由于A 是稳定的，对于任意$v \in L^2(\Omega)$，都有 $\|v\|_{L^2} \le \gamma \|Av\|_{L^2}$。令$v = u_* - u_{NN}$，并利用$A$的线性性，可得

$$\|u_* - u_{NN}\|_{L^2} \le \gamma \|A(u_* - u_{NN})\|_{L^2} = \gamma \|f - Au_{NN}\|_{L^2}. \tag{6}$$

当物理信息神经网络损失足够小时，该界表明神经网络近似u_{NN} 与精确解u_* 之间的误差也很小。这说明网络学习到的特征空间能够很好地逼近该解。基于此特征空间，最小二乘后处理步骤可以提高解的准确性。

2.4 正交性为何重要：特征空间的结构性质

尽管较低的物理信息神经网络训练损失表明网络能够很好地近似特定问题实例的解，但这并不能保证底层特征空间的质量。要理解 PD-OFM 为何优于标准方法，我们必须考察所学习基函数的结构性质。在经典谱方法和 Galerkin 方法中，基的正交性对于最大限度地提高固定维数子空间的逼近能力，以及确保线性求解器的数值稳定性至关重要。受此启发，我们考察三个关键指标：有效秩、正交性和投影误差。我们的假设是，正交正则化会迫使神经网络张成更大的特征空间，从而防止特征坍缩，即多个神经元学习到冗余表示[48]。为量化这些性质，我们考虑基准一维泊松方程:

$$\begin{cases} -\dfrac{d^2u(x)}{dx^2} = 1, & x \in (-1, 1), \\ u(-1) = u(1) = 0. \end{cases} \tag{7}$$

区间 (–1, 1) 上狄利克雷拉普拉斯算子的解析特征函数， $\phi_k(x) = \sin\left(\frac{k\pi(x+1)}{2}\right)$，构成一个完备的正交基。这使我们能够直接衡量所学习的特征空间与最优谱基之间的差异。我们引入以下定义来量化特征空间的质量。

定义 1. 对于具有特征层输出 $\{u_j(x)\}_{j=1}^m$ 的神经网络 $u(x;\theta)$，其对应的格拉姆矩阵 $M_u \in \mathbb{R}^{m\times m}$ 定义为 $(M_u)_{ij} = \int_\Omega u_i(x)u_j(x)dx$。特征空间的 ϵ-有效秩定义为:

$$r_\epsilon(M_u) := \left|\{\lambda_k(M_u) : \lambda_k(M_u) > \epsilon\}\right|,$$

其中， λ_k 是 M_u 的特征值。

Yang 等人[49] 提出了有效秩这一指标，用作衡量基函数线性无关性的稳健度量，并过滤数值噪声。有效秩越高，意味着网络对其宽度的利用越高效。为衡量所学习的空间对真实物理规律的捕捉程度，我们定义**投影误差**。

定义 2。 投影误差用于衡量学习得到的特征$U(x)$表示真实特征函数 $\{\phi_k(x)\}$ 的能力。对于第 k 个特征函数，相对投影误差为:

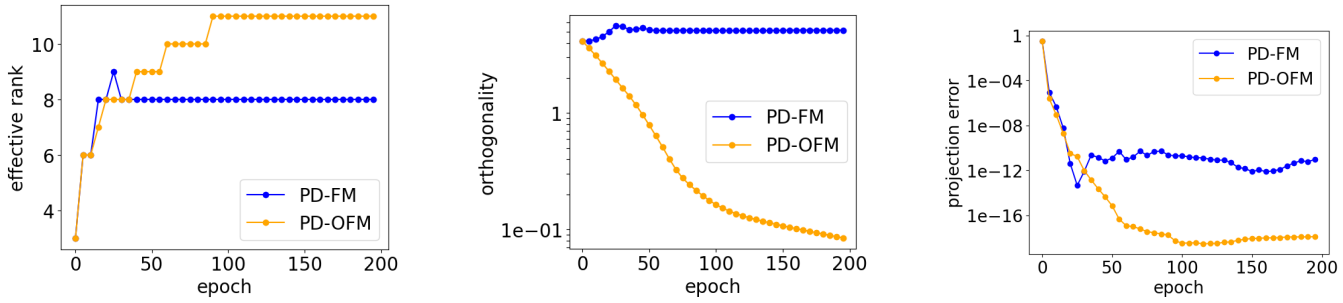
$$\epsilon_k^{\text{proj}} = \frac{\|\mathcal{P}_U\phi_k - \phi_k\|_2}{\|\phi_k\|_2}, \quad \text{where } \mathcal{P}_U\phi_k = U(U^\top U)^{-1}U^\top\phi_k \text{ is the orthogonal projection.}$$

平均值　　投影误差是在前 M_1 个主导特征函数上计算的:　$\epsilon^{\text{proj}} = \frac{1}{M_1}\sum_{k=1}^{M_1}\epsilon_k^{\text{proj}}$.

该指标源于 Kolmogorov$n$-宽度[50], 这一概念，其中前 k 个特征函数张成最优逼近子空间。我们在相同条件下，使用残差网络架构训练两个模型（分别采用和不采用正交正则化 $\lambda_{\text{orth}} = 0.1$），并对其进行比较。图 6展示了这些指标的变化过程:

- **有效秩（6a）**：正则化模型保持着显著更高的秩，这表明有更多神经元在提供独特信息。

- **正交性（6b）**：损失项有效地推动特征协方差趋近于单位阵。
- **投影误差（6c）**：至关重要的是，经过正则化的特征具有更低的投影误差，这意味着它们会自发地与微分算子的主导特征函数更好地对齐。



(a) 有效秩 (b) 正交性损失 (c) 投影误差

图 6 训练过程中，特征空间质量指标的变化。与 PD-FM（蓝色）相比，PD-OFM（橙色）学习到的基具有更高的秩、更强的正交性，并且与物理规律的对齐程度更高。

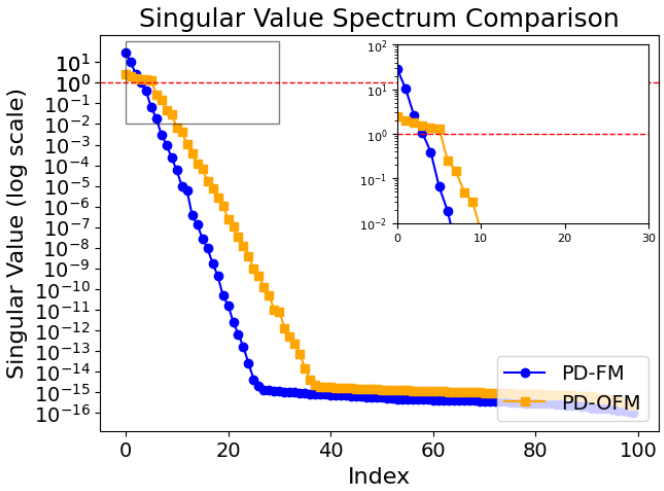


FIGURE 7 特征格拉姆矩阵的奇异值谱。PD-OFM 的衰减速度远慢于 PD-FM。

2.4.1 谱与条件性

图 7 展示了特征格拉姆矩阵的奇异值谱。一个关键观察是，PD-OFM 的奇异值衰减速度远慢于 PD-FM。虽然这并未使系统达到完美条件性（条件数 = 1）——因为逆微分算子的紧致性不可避免地会使特征值趋于零——但它显著减轻了病态性的严重程度。在衰减之前，谱在更多索引位置上保持平坦。这种较慢的衰减意味着，最小二乘步骤中的系统矩阵在数值上更加稳定。

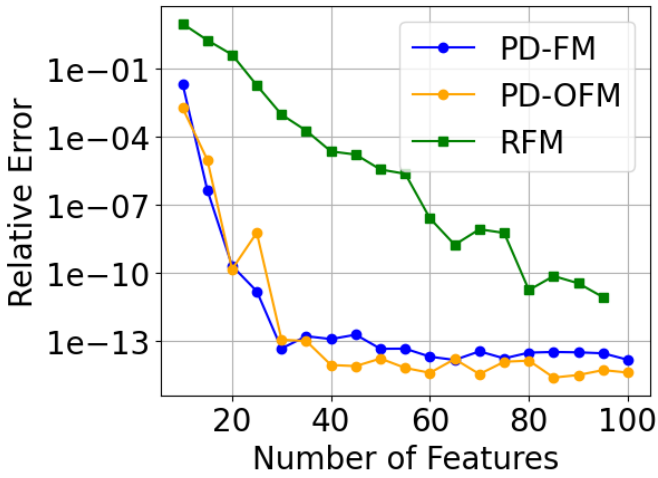


图 8 相对L^2误差与特征数量的关系。比较结果揭示了两项优势：(1) 训练（PD-FM/OFM）优于随机初始化（RFM）；(2) 正交性（PD-OFM）进一步提升了相较于标准训练（PD-FM）的效率。

2.4.2 近似效率

这些结构特性对最终准确性的影响如图 8 所示，我们分析了相对L^2误差随特征数量的变化。该结果揭示了两个不同层面的改进：

1. **训练与随机初始化**：PD-FM 和 PD-OFM 的表现均显著优于 RFM。这证实了物理信息训练能够成功地使特征适应特定算子，而随机特征可能缺少必要的频率分量。2. **正交性与标准训练**：对于相同数量的特征，PD-OFM 始终比 PD-FM 实现更低的误差。这验证了正交性能够减少冗余；在 PD-OFM 中，每增加一个特征都会提升近似能力，而标准训练则因特征相关性而遭遇收益递减。

备注 *2.* 计算复杂度。正交正则化的开销可以忽略不计。设网络有L层，每层宽度为m，批量大小为n。计算正则化项涉及矩阵乘法，其代价为 $\mathcal{O}(nm^2)$。相比之下，物理损失需要通过反向模式自动微分计算拉普拉斯算子。对于d维问题，其主要项的计算复杂度为 $\mathcal{O}(d \cdot nLm^2)$。由于通常$d$、$L \geq 2$，偏微分方程残差的梯度计算占主导，因此相对而言，正交性约束的计算开销几乎为零。

2.5 通过特征在不同几何形状间的可迁移性提高效率

上一节的分析表明，正交正则化能够促使学习到的特征空间与微分算子的谱结构对齐。因此，预训练特征近似于算子的内在特征空间，而不是对某个特定解过拟合。这一性质对于**计算效率**至关重要：它使我们能够将特征学习阶段与求解阶段分离。一旦学得一组算子自适应特征，便可将其作为通用基，重复用于求解一大类具有相同控制算子的问题，从而大幅降低后续实例求解的边际成本。为了验证这一效率，我们提出了一种迁移学习框架：先在简单的典型区域上预训练特征，再将其迁移到源项各异、尤其是几何形状不同的问题中。考虑定义在标准正方形区域 $\Omega_{base} = [-1, 1]^2$ 上的二维泊松方程：

$$-\frac{\partial^2 u}{\partial x^2} - \frac{\partial^2 u}{\partial y^2} = f, \quad (x, y) \in \Omega_{base}, \quad u|_{\partial\Omega_{base}} = 0. \tag{8}$$

通过在该基准问题上训练 PD-OFM，可得到一个特征集 $\{u_i(x,y)\}_{i=1}^{m}$。随后，这些特征将被冻结并迁移到新的实例中进行求解。

2.5.1 跨源项迁移

我们首先评估特征对于源项 $f(x, y)$变化的稳健性。我们构建了一个包含 1000 个随机光滑解 $\{u^{(j)}\}_{j=1}^{1000}$的测试数据集，这些解由具有随机协方差结构的K 个高斯核的线性组合生成：

$$u^{(j)}(x,y)=\sum_{k=1}^{K} c_k^{(j)} \exp\left(-\frac{1}{2}\mathbf{x}_k^{\top}(\Sigma_k^{(j)})^{-1}\mathbf{x}_k\right), \quad \mathbf{x}_k=\begin{bmatrix} x-\mu_{x,k}^{(j)} \\ y-\mu_{y,k}^{(j)} \end{bmatrix}. \tag{9}$$

系数采样为 $c^{(j)}{}_k \sim \mathcal{N}(0,1)$。中心向量 $\boldsymbol{\mu}^{(j)}{}_k$和协方差矩阵 $\Sigma^{(j)}{}_k$的构造如下：

$$\boldsymbol{\mu}_k^{(j)}=\begin{bmatrix} \mu_{x,k}^{(j)} \\ \mu_{y,k}^{(j)} \end{bmatrix}, \quad \text{with} \quad \mu_{x,k}^{(j)}, \mu_{y,k}^{(j)} \sim \mathcal{U}[-1,1], \tag{10}$$

$$\Sigma_k^{(j)}=\begin{bmatrix} \sigma_{x,k}^2 & \rho_k\sigma_{x,k}\sigma_{y,k} \\ \rho_k\sigma_{x,k}\sigma_{y,k} & \sigma_{y,k}^2 \end{bmatrix}. \tag{11}$$

其中，形状参数均匀采样为 $\sigma_{x,k}, \sigma_{y,k} \sim \mathcal{U}[0.1, 0.5]$，相关系数为 $\rho_k \sim \mathcal{U}[-0.5, 0.5]$。相应的源项$f_j$ 通过解析方式导出。

2.5.2 跨几何迁移

一个更具挑战性的情形是将特征迁移到不同的几何域。由于不同的几何形状会施加不同的边界条件，并对应不同的谱基，因此这一测试旨在评估在超集域 Ω_{base} 上学习到的特征，是否通常能够捕捉到在子集上表示解所需的频率分量。我们将基于**正方形**区域训练的特征迁移到**圆环**区域和**L形**区域。对于这两种情形，第j个实例的近似解通过计算成本较低的最小二乘拟合进行重构：

$$\hat{u}^{(j)}(x,y)=\sum_{i=1}^{m} c_{j,i}u_i(x,y), \tag{12}$$

其中，只重新计算线性系数$c_{j,\ i}$ 使其满足新实例的偏微分方程和边界约束。逼近质量通过相对L^2 误差衡量：

$$\mathrm{Err}_j=\frac{\|\hat{u}^{(j)}-u^{(j)}\|_{L^2}}{\|u^{(j)}\|_{L^2}}. \tag{13}$$

表 1 总结了平均相对误差。在所有区域中，PD-OFM 的误差都比基线 PD-FM 低 1 至 2 个数量级。值得注意的是，即使不微调神经网络权重，这些特征也能很好地泛化到 L形和圆环区域。这证实了正交正则化能够生成稳健的算子自适应基，支持高效求解多几何问题，相较于每次几何变化都需要重新训练的方法具有显著优势。

表 1 效率验证：源项和域迁移下的平均相对L^2 误差。与 PD-FM 相比，PD-OFM 在正方形区域上训练的特征具有显著更强的泛化能力。

源项	模型	正方形	L形	环形
$2\pi^2\sin(\pi x)\sin(\pi y)$	PD-OFM	2.35×10^{-8}	1.13×10^{-9}	2.92×10^{-9}
	PD-FM	2.92×10^{-7}	1.06×10^{-8}	2.80×10^{-7}
$(1-x^2)(1-y^2)$	PD-OFM	2.35×10^{-8}	8.32×10^{-8}	5.84×10^{-10}
	PD-FM	2.12×10^{-7}	2.15×10^{-6}	9.02×10^{-8}

与需要在多个物理实例上联合训练的单次迁移学习方法不同，PD-OFM 能够基于单个 偏微分方程实例构建可迁移的基。不同网络深度和宽度的详细比较见附录 B（表 B1–B3）。

3 数值实验

为进一步展示 PD-OFM 的性能，针对不同的偏微分方程开展了一系列数值实验，包括线性、非线性、时变和多维问题。测试案例还包括从简单区间到环域和 L 形区域等复杂几何的区域。目标是评估 PD-OFM 的准确性和效率，并将其与包括随机特征方法和 TransNet 在内的几种现有方法进行比较。在一组均匀采样的测试点上计算了两种标准误差指标：相对 L^2 误差和最大逐点L^∞ 误差：

$$\begin{aligned} \|e\|_{L^2} &= \sqrt{\frac{\sum_{i=1}^{N} |u_\theta(x_i) - u^*(x_i)|^2}{\sum_{i=1}^{N} |u^*(x_i)|^2}}, \\ \|e\|_{L^\infty} &= \max_{1\le i\le N} |u_\theta(x_i) - u^*(x_i)|. \end{aligned} \tag{14}$$

同时还记录了最小二乘步骤中遇到的系统矩阵的有效秩和条件数，这些指标反映了特征空间的表达能力和数值稳定性。

3.1 标准问题基准测试：亥姆霍兹方程

首先考虑一维亥姆霍兹方程：

$$\begin{cases} \dfrac{d^2u}{dx^2} - \lambda u(x) = f(x), & x \in (0,2), \\ u(0) = h_1, \quad u(2) = h_2. \end{cases} \tag{15}$$

$\lambda = 10$ 并选择 $f(x)$，使得解析解为

$$u^*(x) = \sin\left(3\pi x + \frac{3\pi}{20}\right)\cos\left(2\pi x + \frac{\pi}{10}\right) + 2. \tag{16}$$

所有模型都将解构造为 100 个特征的线性组合，这对应于最终隐藏层的宽度或随机特征的数量。对于 PD-OFM，采用带有 $\tanh^3$ 激活函数的全连接神经网络，并使用 Adam 优化器进行训练，初始学习率为 0.001，最多训练 1000 步。如果总损失低于初始损失的 $\varepsilon = 10^{-3}$，则提前终止训练。对于 PD-OFM，正交正则化强度设为 $\lambda_{\text{orth}} = 1$。每次迭代时，以 1000 为批量大小随机采样内部配置点，并使用两个固定点在区域端点处施加狄利克雷边界条件。测试集包含在该区域内等间距采样的 2000 个点。结果如表 2 和图 9 所示，其中所提出的方法达到了 10^{-13} 准确性。

方法	最小二乘残差	相对L^2 误差
PD-OFM	6.85×10^{-10}	3.98×10^{-13}
PD-FM	5.83×10^{-7}	1.38×10^{-9}
RFM	1.01×10^{-5}	9.64×10^{-9}
TransNet	1.00×10^{-2}	1.43×10^{-3}

表 2 一维亥姆霍兹方程上各方法的比较。误差在 2000 个均匀间隔测试点上进行评估。

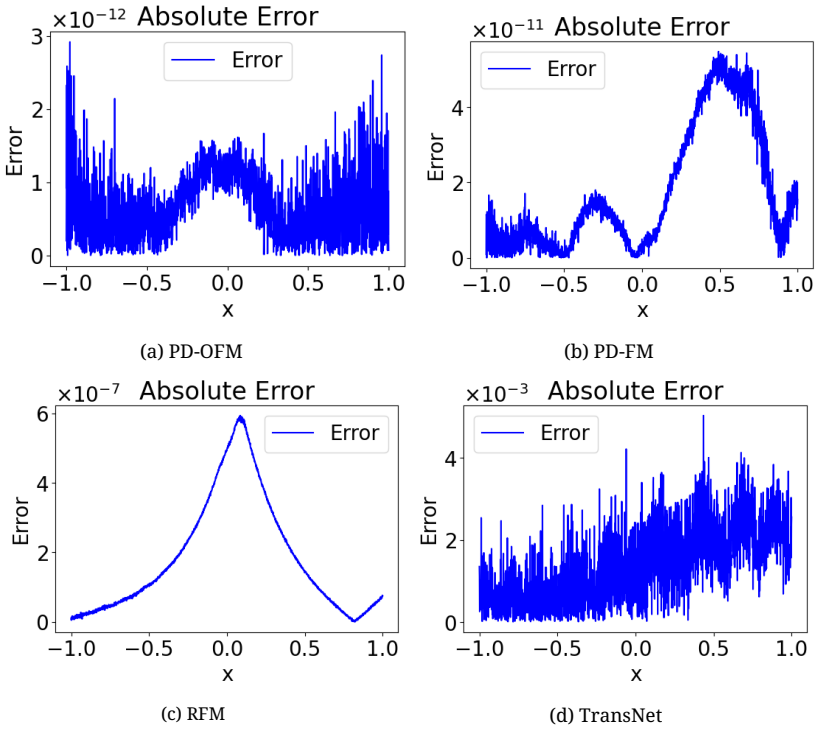


图 9 不同方法在一维亥姆霍兹方程上的逐点绝对误差。

3.2 复杂几何评估：不规则区域上的泊松方程

考虑二维泊松方程：

$$\begin{cases} -\Delta u(x,y) = f(x,y), & (x,y) \in \Omega, \\ u|_{\partial\Omega} = g(x,y). \end{cases} \tag{17}$$

为评估各方法对区域几何形状的鲁棒性，我们考虑了三个具有代表性的案例：

- **正方形区域**： $\Omega_1 = [-1,1]^2$
- **L形区域**： $\Omega_2 = [-1,1]^2 \setminus [0,1] \times [0,1]$
- **环形区域**： $\Omega_3 = \{(x,y) \in \mathbb{R}^2 \mid 0.25 \leq \sqrt{x^2+y^2} \leq 1\}$

所有模型均使用 500 个特征。对于 PD-OFM，神经网络由两个宽度为 500 的隐藏层组成，采用 $\tanh^3$激活函数，并使用 Adam 优化器进行训练，初始学习率为 0.001，最多训练 1000 步。当总损失低于 $\varepsilon = 10^{-3}$ 时，训练终止。正交正则化强度设为 $\lambda_{\text{orth}} = 0.01$。每种情况下，每次迭代采样 1024 个内部点和 128 个边界点。在区域内 2500 个均匀间隔测试点上评估误差。

如表3和图10所示，PD-OFM在所有测试区域中始终取得最高准确性。有趣的是，各种方法的最终预测都呈现出不同的误差模式。在PD-OFM中，误差高度局部化，并集中在单个空间点上。这可能是因为正交性约束促使表示在全局上保持平衡，却忽略了细微的局部特征。相比之下，随机特征方法尽管采用了光滑单位分解，但在子域之间的界面处误差最大。这可能是由于最小二乘形式中内部约束与边界约束存在固有的不对称性，导致边界强制未得到足够重视。

表 3 不同区域上四种方法求解泊松方程的比较。

区域	方法	最小二乘残差	相对L^2 误差
正方形区域	PD-OFM	2.87×10^{-7}	2.81×10^{-9}
	PD-FM	7.27×10^{-7}	9.62×10^{-9}
	RFM	7.83×10^{-5}	1.73×10^{-5}
	TransNet	1.47×10^{-5}	1.43×10^{-7}
L形区域	PD-OFM	4.09×10^{-8}	6.48×10^{-11}
	PD-FM	5.98×10^{-7}	5.82×10^{-9}
	RFM	2.88×10^{-6}	2.48×10^{-6}
	TransNet	1.04×10^{-6}	5.16×10^{-9}
环形区域	PD-OFM	5.43×10^{-9}	1.29×10^{-11}
	PD-FM	8.98×10^{-9}	4.13×10^{-11}
	RFM	4.93×10^{-6}	4.50×10^{-8}
	TransNet	4.30×10^{-7}	2.23×10^{-9}

3.3 向更高维度的可扩展性：单位立方体上的三维泊松方程

考虑区域 $\Omega = [0,1]^3$ 上的三维泊松方程：

$$\begin{cases} -\Delta u(x,y,z) = f(x,y,z), & (x,y,z) \in \Omega, \\ u(x,y,z) = 0, & (x,y,z) \in \partial\Omega. \end{cases} \tag{18}$$

精确解如下：

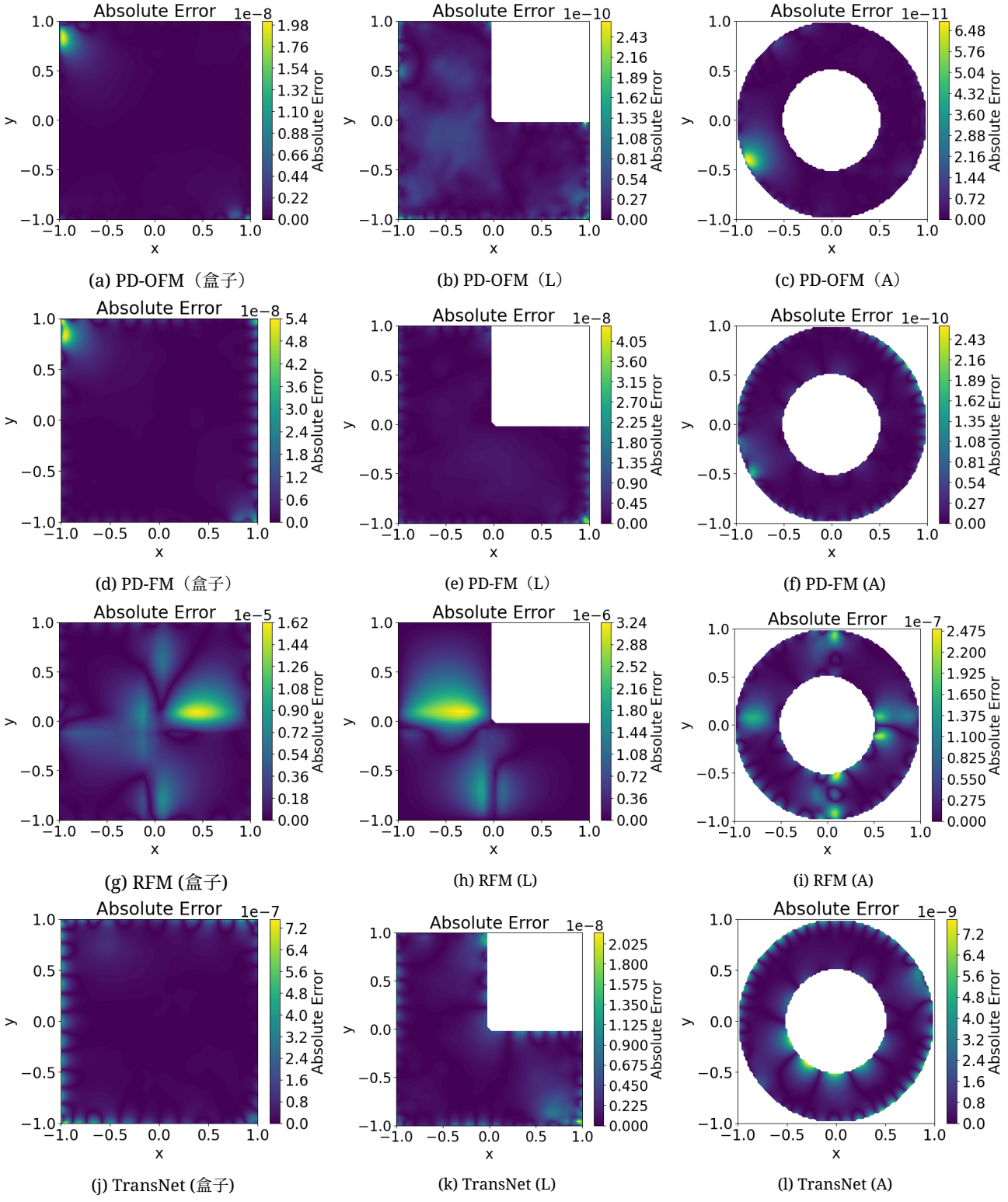
$$u(x,y,z) = \sin(\pi x)\sin(\pi y)\sin(\pi z), \tag{19}$$

相应的源项通过解析方式计算。网络架构由两个隐藏层组成，每层有 600 个神经元，并带有残差连接。特征空间层宽度设为 600。正交正则化强度为 $\lambda_{\text{orth}} = 0.01$。共使用 2048 个内部配置点和 6,00 个边界配置点。

TABLE 4 三维泊松方程上的方法比较。误差在 2700 个均匀间隔测试点上评估，并沿 z轴取平均值。

方法	最小二乘残差	相对L^2 误差
PD-OFM	6.89×10^{-5}	1.64×10^{-5}
PD-FM	6.37×10^{-5}	1.07×10^{-4}
RFM	1.01×10^{-2}	1.38×10^{-1}
TransNet	1.42×10^{-4}	2.03×10^{-4}

表 4 报告了三维泊松方程上的定量比较。PD-OFM 在所有方法中取得了最低的最小二乘残差和相对 L^2 误差，且有效秩适中。尽管 RFM 因单位分解而获得最高秩，但所学特征未能捕捉偏微分方程的谱结构，这由其显著更高的残差和投影误差所表明。图 11 中可视化的逐点绝对误差证实，尽管 RFM 特征的秩很大，仍无法准确逼近解。TransNet 方法的误差较 RFM 有所改善，但仍不及 PD-OFM。



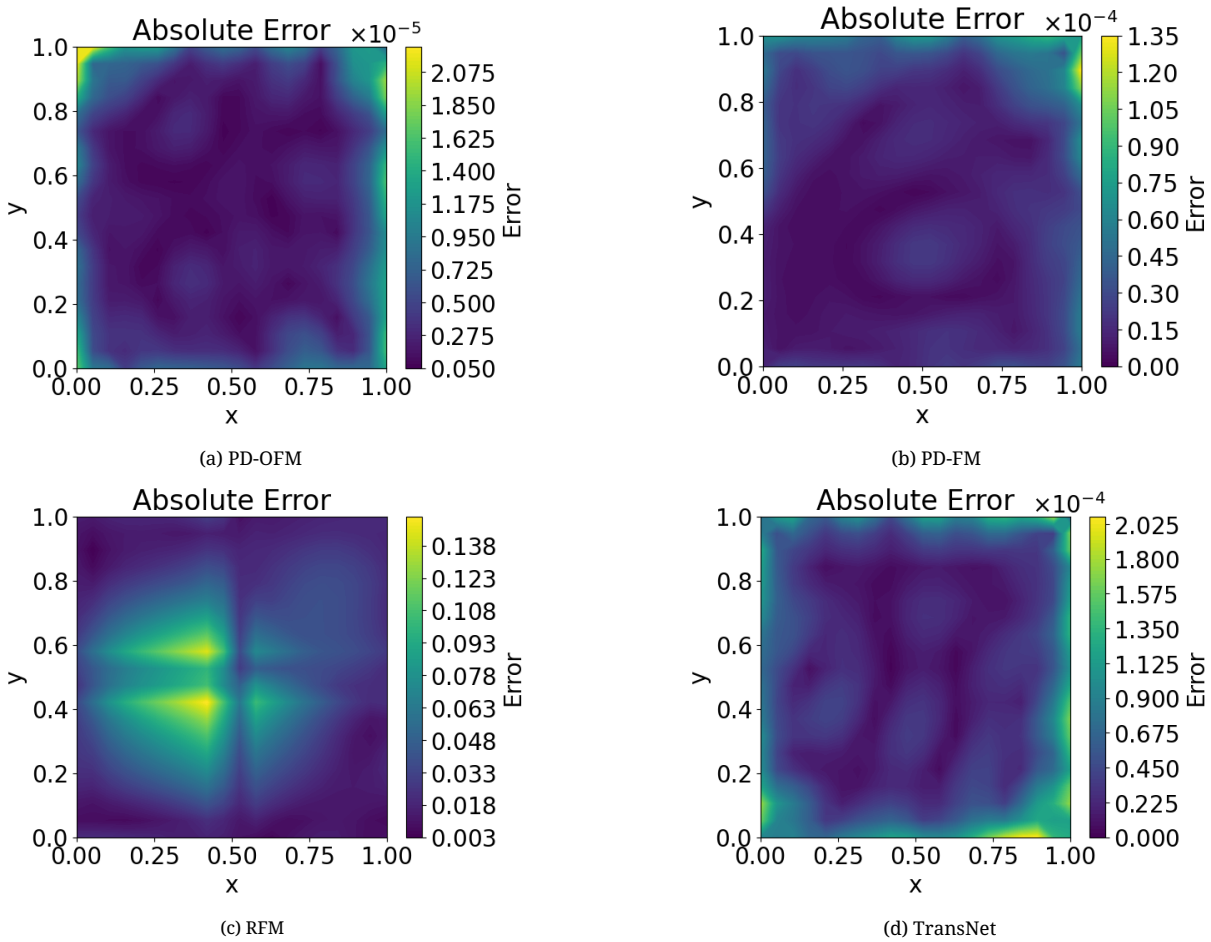
(a) PD-OFM（盒子） (b) PD-OFM（L） (c) PD-OFM（A）

(d) PD-FM（盒子） (e) PD-FM（L） (f) PD-FM (A)

(g) RFM (盒子) (h) RFM (L) (i) RFM (A)

(j) TransNet (盒子) (k) TransNet (L) (l) TransNet (A)

图 10不同方法（行）在三个区域（列）上的二维泊松方程逐点绝对误差：盒子、L形和圆环。



(a) PD-OFM (b) PD-FM

(c) RFM (d) TransNet

图 11 三维泊松方程不同方法的逐点绝对误差。

3.4 探索多种边界条件：波动方程

在第三个测试案例中，考虑随时间变化的一维波动方程：

$$\begin{cases} \dfrac{\partial^2 u}{\partial t^2} = c\dfrac{\partial^2 u}{\partial x^2}, & x \in [0,1],\ t \in [0,2], \\ u(x,0) = \sin(4\pi x), \\ \left.\dfrac{\partial u}{\partial t}\right|_{t=0} = 0, \\ u(0,t) = u(1,t), \end{cases} \tag{20}$$

其中 $c = \frac{1}{16\pi^2}$。该方程的精确解为

$$u(x,t) = \frac{1}{2}(\sin(4\pi x + t) - \sin(4\pi x - t)). \tag{21}$$

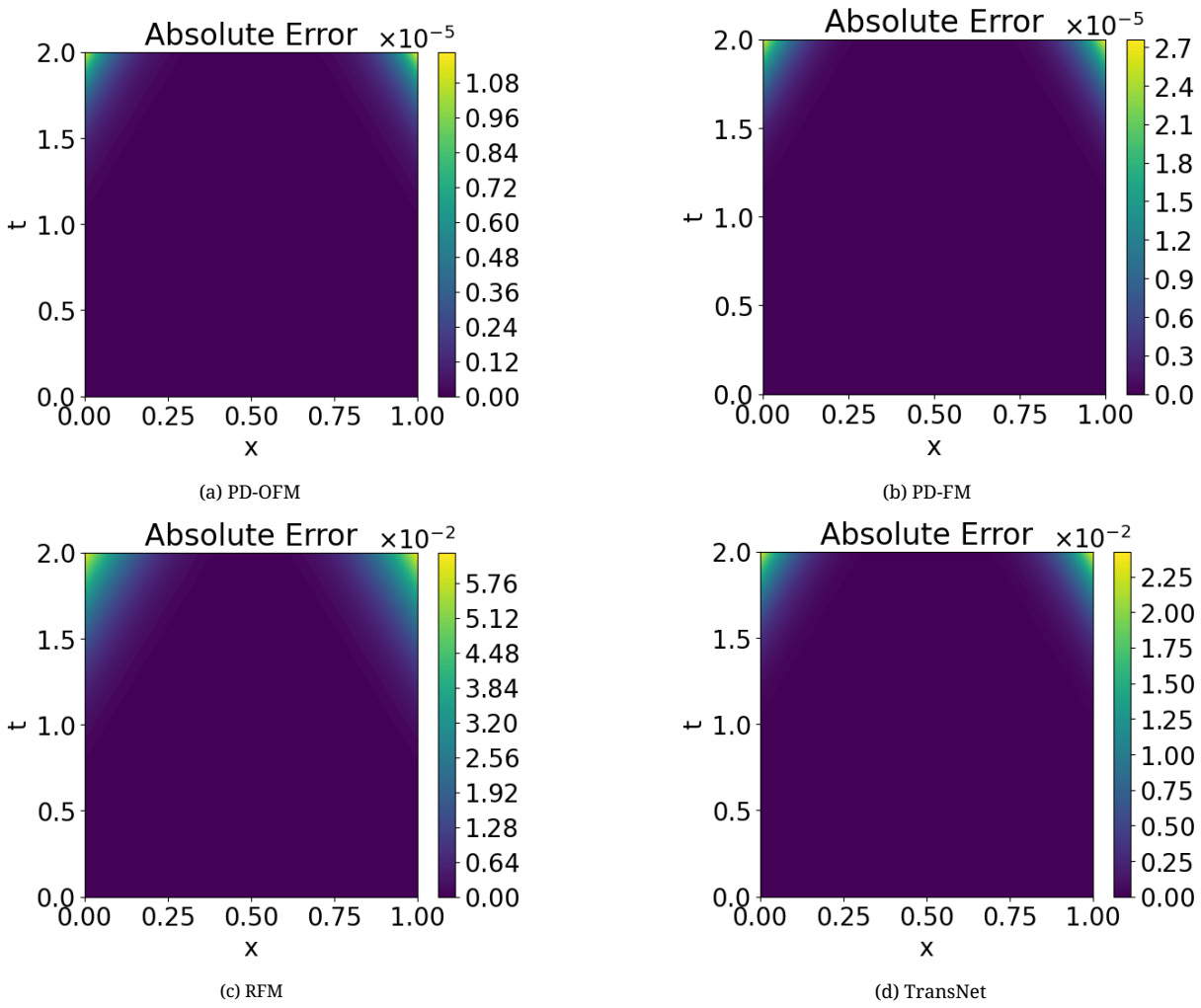
该方程不仅具有通常的初始条件（初始位移和初始速度），还具有周期边界条件。我们的方法灵活，可适应不同类型的边界条件。具体而言，周期边界条件在以下点处强制施加

$$\{(0,t_1),\ldots,(0,t_n),\ (1,t_1),\ldots,(1,t_n)\},$$

这些点位于边界$x = 0$和$x = 1$上，并共享同一组时间值$\{t_1, \ldots, t_n\}$。结果如表5和图12所示。

方法	最小二乘残差	相对L^2误差
PD-OFM	6.45×10^{-11}	1.99×10^{-6}
PD-FM	1.25×10^{-11}	4.77×10^{-6}
RFM	1.09×10^{-6}	1.62×10^{-2}
TransNet	2.75×10^{-9}	4.41×10^{-3}

TABLE 5 一维波动方程上各方法的比较。误差在 2500 个均匀间隔测试点上进行评估



(a) PD-OFM

(b) PD-FM

(c) RFM

(d) TransNet

FIGURE 12 一维波动方程上不同方法的逐点绝对误差。

如图 12 所示，与前面的示例相比，所有方法产生的误差都更大。值得注意的是，将周期边界条件替换为狄利克雷边界条件后，绝对误差降至 1×10^{-10} 以下。这表明，周期边界条件与其他约束结合时会造成不平衡。尽管如此，在所有测试方法中，PD-OFM 的误差最低。

3.5 非线性偏微分方程：稳态纳维–斯托克斯方程

考虑稳态不可压缩纳维–斯托克斯方程：

$$\begin{cases} \boldsymbol{u} \cdot \nabla \boldsymbol{u} + \nabla p - \nu \Delta \boldsymbol{u} = 0, \\ \nabla \cdot \boldsymbol{u} = 0, \end{cases} \tag{22}$$

其中，$\boldsymbol{u} = (v_1, v_2)$ 为速度场，p 为压力，ν 表示运动黏度。雷诺数定义为$Re = 1/\nu$。

作为基准，采用经典的 Kovasznay 流，它为二维稳态纳维–斯托克斯方程提供了精确解析解：

$$v_1(x_1, x_2) = 1 - e^{\lambda x_1} \cos(2\pi x_2), \tag{23}$$

$$v_2(x_1, x_2) = \frac{\lambda}{2\pi} e^{\lambda x_1} \sin(2\pi x_2), \tag{24}$$

$$p(x_1, x_2) = \frac{1}{2} \left(1 - e^{2\lambda x_1}\right), \tag{25}$$

其中

$$\lambda = \frac{1}{2\nu} - \sqrt{\frac{1}{4\nu^2} + 4\pi^2}. \tag{26}$$

在本实验中，Re 等于 40，对应于 $\nu = 0.025$。计算域为 $\Omega = [-1, 1]^2$, ，并通过将精确解限制在边界 $\partial\Omega$ 上来施加狄利克雷边界条件。为处理纳维–斯托克斯方程的非线性，采用 Picard 迭代。在第k次迭代中，使用前一步的速度场$\boldsymbol{u}_{\mathrm{NN}}^{k-1}$ 对对流项进行线性化。第k 次迭代中最小化的损失函数为：

$$\mathcal{L}^{(k)} = \boldsymbol{u}_{\mathrm{NN}}^{k-1} \cdot \nabla \boldsymbol{u}_{\mathrm{NN}}^{k} + \nabla p_{\mathrm{NN}}^{k} - \nu \Delta \boldsymbol{u}_{\mathrm{NN}}^{k}. \tag{27}$$

经过 20 次 Picard 迭代后，不同方法中u、v、p 的逐点绝对误差如图 13 所示。

4 讨论与未来工作

本节将对所提出的 PD-OFM 进行批判性评估，重点讨论其计算影响，以及正交性与数值性能之间的权衡。

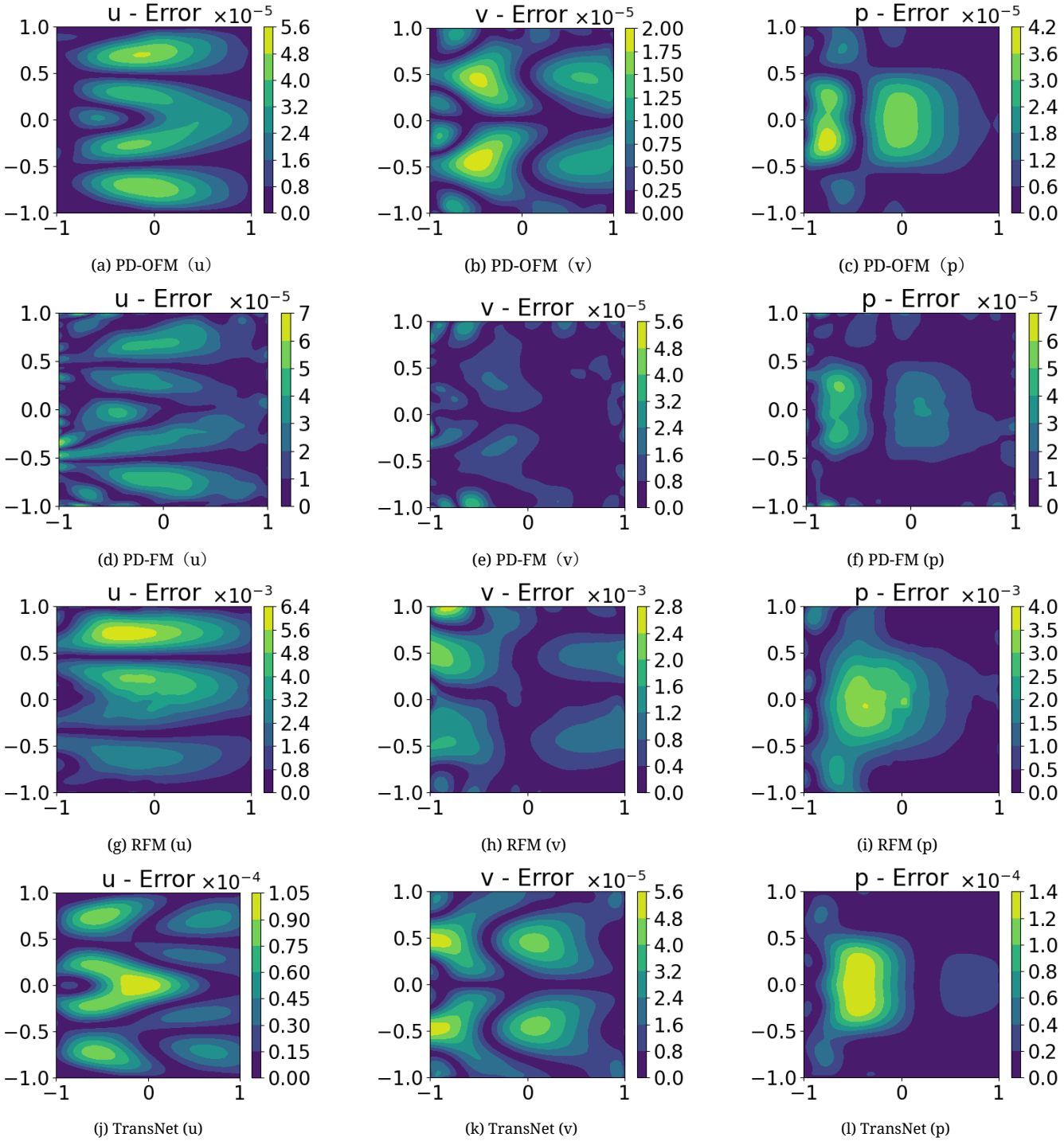
4.1 效率

• **正交性正则化**：与计算 PINN 损失相比，计算正交性损失的成本几乎可以忽略不计，因为该成本与偏微分方程的阶数和神经网络的深度无关。
• **可迁移性**：尽管 RFM 受益于零成本初始化，但 PD-OFM 的预训练成本是一项战略性投入：学习到的近似正交特征表现出很强的可迁移性。这使得已发现的特征能够重复使用，将范式从随机初始化转向自动化特征工程。

4.2 准确性

• **准确性理论基础中的空白**：尽管在我们的实验中施加正交性始终能提高数值精度，但严格的理论联系仍是一个悬而未决的问题。我们的结果表明，正则化扩大了特征空间的有效秩（如第2.4节中的指标所示），但这种正则化如何降低逼近误差的理论机制仍需进一步研究。

• **病态性**：尽管 PD-OFM 减缓了奇异值的衰减，所得系统矩阵仍可能表现出一定程度的病态性。通过基于梯度的训练完全消除特征相关性仍然困难。



(a) PD-OFM（u） (b) PD-OFM（v） (c) PD-OFM（p）

(d) PD-FM（u） (e) PD-FM（v） (f) PD-FM (p)

(g) RFM (u) (h) RFM (v) (i) RFM (p)

(j) TransNet (u) (k) TransNet (v) (l) TransNet (p)

图cuntegn **E 13** 不同方法（各行）下稳态纳维–斯托克斯方程中u、v、p 的逐点绝对误差 n.

5 结论

在本工作中，我们提出了物理驱动正交特征方法，这是一种将正交性正则化融入物理信息神经网络以构建算子自适应正交特征的新框架。PD-OFM 能够发现具有更强表示能力与更好可迁移性的特征。针对一系列偏微分方程（包括亥姆霍兹方程、泊松方程、波动方程和纳维-斯托克斯方程）的大量数值实验一致表明，PD-OFM 显著优于现有方法，例如

例如随机特征方法[23]和 TransNet[28]。总体而言，PD-OFM 通过利用正交神经表示，弥合了经典谱方法与现代基于机器学习的求解器之间的差距，为偏微分方程应用提供了稳健的特征学习框架。

数据可用性声明

支持本研究结论的数据和源代码已在 GitHub 仓库中公开，网址为 https://github.com/kinjonfire/Physics-Driven-Discovery-of-Orthogonal-Features。代码实现了 PD-OFM 框架，并包含用于复现论文中数值实验的脚本。

基金资助

D. Wang 获得了国家自然科学基金委员会（项目编号：12422116）、广东省基础与应用基础研究基金（项目编号：2023A1515012199）、深圳市科技创新计划（项目编号：JCYJ20220530143803007、RCYX20221008092843046）以及河套深港科技创新合作区项目（项目编号：HZQSWS-KCCYB-2024016）的部分资助。

参考文献

1. Han J, Jentzen A, E W. 使用深度学习求解高维偏微分方程。《美国国家科学院院刊》2018年;115(34):8505-8510. doi: 10.1073/pnas.1718942115 2. Raissi M, Perdikaris P, Karniadakis G. 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逆问题的深度学习框架。《计算物理学杂志》2019;378:686-707. doi: 10.1016/j.jcp.2018.10.045 3. Sirignano J, Spiliopoulos K. DGM：一种用于求解偏微分方程的深度学习算法。《计算物理学杂志》2018;375:1339-1364. doi: 10.1016/j.jcp.2018.08.029 4. Sukumar N, Srivastava A. 在物理信息深度神经网络中利用距离函数精确施加边界条件。《计算力学与工程方法》2022年;389:114333. doi: 10.1016/j.cma.2021.114333 5. Wang S, Yu X, Perdikaris P. PINN何时以及为何训练失败：神经切核视角。《计算物理学杂志》2022年;449:110768. doi: https://doi.org/10.1016/j.jcp.2021.110768 6. Rathore P, Lei W, Frangella Z, Lu L, Udell M. PINN训练中的挑战：损失景观视角。载于：International Machine Learning Society (IMLS). JMLR.org 2024:1715:1–1715:33 7. Yu B, others . 深度 Ritz 方法：一种用于求解变分问题的深度学习数值算法。数学与统计通讯2018年;6(1):1–12. doi: 10.1007/s40304-018-0127-z 8. Kharazmi E, Zhang Z, Karniadakis GE. hp-VPINNs：采用区域分解的变分物理信息神经网络。《计算力学与工程方法》2021年;374:113547. doi: 10.1016/j.cma.2020.113547 9. Uriarte C, Bastidas M, Pardo D, Taylor JM, Rojas S. 使用最小二乘优化变分物理信息神经网络。《计算机与应用数学》2025年;185:76-93. doi: 10.1016/j.camwa.2025.02.022 10. Baydin AG, Pearlmutter BA, Radul AA, Siskind JM. 机器学习中的自动微分：综述。机器学习研究杂志2018年;18(153):1–43. 11. Bao G, Wang D, Zou B. WANCO：用于约束优化问题的弱对抗网络。arXiv预印本 arXiv:2407.03647; 2024年。 12. Zang Y, Bao G, Ye X, Zhou H. 用于高维偏微分方程的弱对抗网络。《计算物理学杂志》2020年;411:109409. doi: 10.1016/j.jcp.2020.109409 13. Jacob B, Howard AA, Stinis P. SPIKANs：可分离物理信息 Kolmogorov-Arnold 网络。arXiv预印本 arXiv:2411.06286; 2024年。 14. Liu Z, Wang Y, Vaidya S, et al. KAN：Kolmogorov-Arnold 网络。arXiv预印本 arXiv:2404.19756; 2025年。 15. Toscano J, Oommen V, Varghese A, others . 从 PINN 到 PIKAN：物理信息机器学习的最新进展。科学与工程中的机器学习与计算2025年;1:15. doi: 10.1007/s44379-025-00015-1 16. Wang T, Hu Z, Kawaguchi K, Zhang Z, Karniadakis GE. 用于高维 Fokker–Planck 方程的张量神经网络。神经网络2025年;185:107165. doi: https://doi.org/10.1016/j.neunet.2025.107165 17. Wang Y, Lin Z, Liao Y, others . 使用张量神经网络和后验误差估计子求解高维偏微分方程。科学计算杂志2024年;101(67). doi: 10.1007/s10915-024-02700-4

18. Wang Y, Xie H, Jin P. 张量神经网络及其数值积分。《计算数学杂志》2024年;42(6):1714–1742. doi: 10.4208/jcm.2307-m2022-0233 19. Dong S, Li Z. 用于求解线性与非线性偏微分方程的局部极限学习机和区域分解。《计算力学与工程方法》2021年;387:114129. doi: 10.1016/j.cma.2021.114129 20. Dong S, Li Z. 一种用于预训练极限学习机随机系数的改进批量内在可塑性方法。《计算物理学杂志》2021年;445:110585. doi: https://doi.org/10.1016/j.jcp.2021.110585 21. Dong S, Yang J. 极限学习机超参数的计算：算法及其在计算偏微分方程中的应用，并与经典有限元和高阶有限元进行比较。《计算物理学杂志》2022年;463:111290. doi: https://doi.org/10.1016/j.jcp.2022.111290 22. Tan L, Chen J. 用于随机特征方法的高精度随机预条件迭代方法。《计算与应用数学杂志》2026年;481:117255. doi: https://doi.org/10.1016/j.cam.2025.117255 23. Chen J, Chi X, E W, Yang Z. 融合用于求解偏微分方程的传统算法与基于机器学习的算法：随机特征方法。《机器学习杂志》2022年;1(3):268–298. doi: 10.4208/jml.220726 24. Liu Y, McCalla SG, Schaeffer H. 用于学习相互作用动力系统的随机特征模型。《皇家学会会刊*A*辑：数学、物理与工程科学》2023年;479:20220835. doi: 10.1098/rspa.2022.0835 25. De Ryck T, Mishra S, Shang Y, Wang F. 随机神经网络的逼近理论及其在求解高维偏微分方程中的应用。arXiv预印本 arXiv:2501.12145; 2025年。 26. Shang Y, Wang F. 利用 Petrov–Galerkin 方法求解线弹性和纳维–斯托克斯方程的随机神经网络。《工程力学杂志》2024年;150(4). doi: 10.1061/JENMDT.EMENG-7463 27. Lu T, Ju L, Zhu L. 一种结合区域分解、用于求解椭圆界面问题的多重可迁移神经网络方法。《计算物理学杂志》2025年;530:113902. doi: https://doi.org/10.1016/j.jcp.2025.113902 28. Zhang D, Li Y, Ying S. Trans-Net：一种基于时间区域分解、用于求解偏微分方程的可迁移预训练神经网络。《计算物理通讯》2024年;299:109130. doi: 10.1016/j.cpc.2024.109130 29. Lin Z, Wang Y, Xie H. 用于高精度求解偏微分方程的自适应神经网络子空间方法。arXiv预印本 arXiv:2412.02586; 2024年。 30. Song H, Sheng Z, Wang D, Lv J. 用于无界区域声散射问题的交替优化SNN方法。arXiv预印本 arXiv:2504.16523; 2025年。 31. Xu Z, Sheng Z. 基于神经网络求解偏微分方程的子空间方法。《计算机与应用数学》2025年;195:109-138. doi: https://doi.org/10.1016/j.camwa.2025.07.001 32. Liu L, Yong H. DeePoly：一种用于函数逼近和求解偏微分方程的高阶精度科学机器学习框架。*arXiv*预印本 *arXiv:2409.15818*2025年。doi: 10.13140/RG.2.2.24649.25446 33. Bengio Y. 用于无监督学习和迁移学习的表示的深度学习。载于：. 27. International Machine Learning Society (IMLS). PMLR 2012年。 34. Pan SJ, Yang Q. 迁移学习综述。《*IEEE*知识与数据工程汇刊》2010年;22(10):1345-1359. doi: 10.1109/TKDE.2009.191 35. Chakraborty S. 基于迁移学习的多保真物理信息深度神经网络。《计算物理学杂志》2021年;426:109942. doi: 10.1016/j.jcp.2020.109942 36. Desai S, Mattheakis M, Joy H, Protopapas P, Roberts SJ. 物理信息神经网络的单次迁移学习。载于：International Machine Learning Society (IMLS). 2022年。 37. Chen Y, Koohy S. GPT-PINN：面向参数化偏微分方程非侵入式元学习的生成式预训练物理信息神经网络。《有限元分析与设计》2024年;228:104047. doi: https://doi.org/10.1016/j.finel.2023.104047 38. Chen Y, Ji Y, Narayan A, Xu Z. TGPT-PINN：利用变换后的 GPT-PINN 进行非线性模型降阶。《计算力学与工程方法》2024年;430:117198. doi: https://doi.org/10.1016/j.cma.2024.117198 39. He J, Liu X, Xu J. MgNO：通过多重网格高效参数化线性算子。载于：国际学习表征会议。2024年。 40. Li Z, Kovachki NB, Azizzadenesheli K, et al. 用于参数偏微分方程的傅里叶神经算子。载于：国际学习表征会议。2021年。 41. Li Z, Zheng H, Kovachki N, et al. 用于学习偏微分方程的物理信息神经算子。《*ACM/IMS*数据科学杂志》2024年;1(3). doi: 10.1145/3648506

42. Tripura T, Chakraborty S. 用于求解计算力学问题中参数偏微分方程的小波神经算子。《计算力学与工程方法》。2023年;404:115783. doi: 10.1016/j.cma.2022.115783

43. He K, Zhang X, Ren S, Sun J. 深入研究整流器：在 ImageNet图像分类 上超越人类水平的性能。载于：IEEE计算机学会。2015年:1026–1034

44. Glorot X, Be ngio Y. 理解训练深度前馈神经网络的难点。载于：PMLR。2010年:249–256。

45. Sun J, Dong S, Wang F. 基于间断伽辽金方法求解偏微分方程的局部随机神经网络。《计算与应用数学杂志》。2024年;445:115830. doi: https://doi.org/10.1016/j.cam.2024.115830

46. Dong S, Yang J. 使用人工神经网络和变量投影法数值近似偏微分方程。《计算力学与工程方法》。 2022年;398:115284. doi: https://doi.org/10.1016/j.cma.2022.115284

47. He K, Zhang X, Ren S, Sun J. 用于图像识别的深度残差学习。载于：计算机视觉与模式识别。201 6

48. Zhou H, Qixuan Z, Luo T, Zhang Y, Xu ZQ. 探索神经网络在初始训练阶段的凝聚现象。载于：Koyejo S, Mohamed S, Agarwal A, Belgrave D, Cho K, Oh A.，编。《神经信息处理系统进展》。35. Curran Associates, Inc. 2022年:2184–2196.

49. Yang J, Zhao Y, Zhu Q. 有效秩与阶梯现象：对神经网络训练动态的新见解。arXiv预印本 arXiv:2412.05144；2025年。

50. Pinkus A. 逼近理论中的 *N* 宽度。7. Springer Science & Business Media，2012年。

本文引用格式： Jia Q. 和 Wang D.（2026年），物理信息引导的正交特征学习及其在求解偏微分方程中的应用，《偏微分方程数值方法》，2026年；00：1–20。

附录

A 使用变分损失求解不同区域上的二维泊松方程

本附录研究使用变分损失而非传统物理信息神经网络损失，在三类区域——盒子、圆环和 L 形区域——上求解二维泊松方程的效果。方程为：

$$-\Delta u(x,y) = f(x,y), \quad (x,y) \in \Omega \tag{A1}$$

边界条件为零狄利克雷边界条件。采用变分损失[7]：

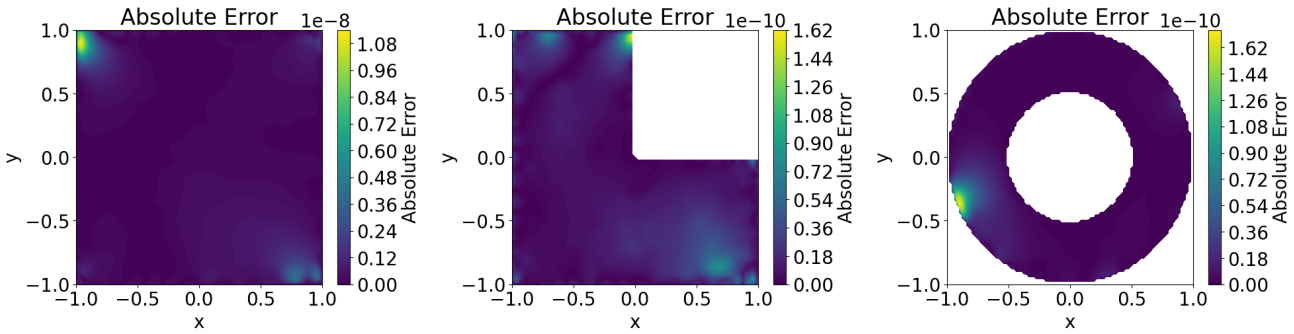
$$L_{\text{var}} = \frac{1}{2}\int_{\Omega} |\nabla u(x,y)|^2 \, dx\, dy - \int_{\Omega} f(x,y)u(x,y)\, dx\, dy. \tag{A2}$$

在实际应用中，变分损失可近似表示为：

$$L_{\text{var}} \approx \frac{1}{2N_q}\sum_{i=1}^{N_q} |\nabla u(x_i,y_i)|^2 - \frac{1}{N_q}\sum_{i=1}^{N_q} f(x_i,y_i)u(x_i,y_i) \tag{A3}$$

其中 N_q 是在区域中采样的求积点数量。我们采用如下神经网络结构：两层隐藏层，每层包含 500 个神经元，并使用 $\tanh$[3] 激活函数。训练采用 Adam 优化器。所提出的框架支持灵活设计偏微分方程损失。损失函数不局限于经典的物理信息神经网络残差损失（强形式）。变分形式、加权残差法、弱形式或其他定制损失函数也可以在该结构中应用。这种灵活性为基于神经网络的偏微分方程求解器的未来发展提供了广阔空间。

B PD-OFM 与 PD-FM 的可迁移性比较



(a) 盒形区域 (b) L形区域 (c) 环形区域

图 A1 不同区域上变分损失解的逐点绝对误差。

表 B1 实验 1：盒形区域迁移问题中的误差与宽度、深度的关系。

宽度	深度 = 1		深度 = 2		深度 = 3	
	PD-OFM	PD-FM	PD-OFM	PD-FM	PD-OFM	PD-FM
100	5.08×10^{-2}	5.13×10^{-2}	3.13×10^{-2}	5.66×10^{-2}	4.40×10^{-2}	6.87×10^{-2}
200	1.08×10^{-3}	4.51×10^{-3}	4.94×10^{-4}	1.13×10^{-3}	1.13×10^{-3}	1.87×10^{-3}
300	2.46×10^{-4}	2.49×10^{-3}	2.47×10^{-5}	3.06×10^{-5}	6.64×10^{-5}	4.96×10^{-5}
400	8.13×10^{-4}	1.98×10^{-3}	6.34×10^{-7}	1.01×10^{-6}	7.66×10^{-6}	7.75×10^{-7}
500	2.41×10^{-4}	6.28×10^{-4}	2.35×10^{-8}	2.12×10^{-7}	3.40×10^{-7}	7.66×10^{-8}
600	7.85×10^{-4}	1.17×10^{-3}	2.73×10^{-8}	1.24×10^{-7}	3.16×10^{-8}	1.10×10^{-8}

T AB L E B2 实验 2：L 形域迁移问题中的误差与宽度和深度的关系。

宽度	深度 = 1		深度 = 2		深度 = 3	
	PD-OFM	PD-FM	PD-OFM	PD-FM	PD-OFM	PD-FM
100	8.64×10^{-2}	1.12×10^{-1}	6.76×10^{-2}	7.61×10^{-2}	8.53×10^{-2}	8.87×10^{-2}
200	2.83×10^{-3}	1.09×10^{-2}	1.07×10^{-3}	1.98×10^{-3}	1.91×10^{-3}	2.58×10^{-3}
300	1.34×10^{-3}	4.39×10^{-3}	4.86×10^{-5}	6.97×10^{-5}	1.19×10^{-4}	7.99×10^{-5}
400	1.29×10^{-3}	4.33×10^{-3}	1.83×10^{-6}	3.74×10^{-6}	1.17×10^{-5}	1.95×10^{-6}
500	1.13×10^{-3}	3.19×10^{-3}	8.32×10^{-8}	2.15×10^{-6}	5.87×10^{-7}	4.47×10^{-7}
600	1.29×10^{-3}	2.52×10^{-3}	2.03×10^{-8}	9.30×10^{-7}	3.92×10^{-8}	1.39×10^{-7}

表 B3 实验 3：环域迁移问题中的误差与宽度、深度的关系。

宽度	深度 = 1		深度 = 2		深度 = 3	
	PD-OFM	PD-FM	PD-OFM	PD-FM	PD-OFM	PD-FM
100	1.87×10^{-2}	2.41×10^{-2}	8.02×10^{-3}	1.29×10^{-2}	1.42×10^{-2}	2.11×10^{-2}
200	2.12×10^{-4}	1.51×10^{-3}	4.96×10^{-5}	1.39×10^{-4}	1.38×10^{-4}	2.82×10^{-4}
300	7.41×10^{-5}	6.34×10^{-4}	1.43×10^{-6}	2.04×10^{-6}	3.87×10^{-6}	3.65×10^{-6}
400	7.28×10^{-5}	4.60×10^{-4}	1.99×10^{-8}	7.12×10^{-8}	2.49×10^{-7}	2.93×10^{-8}
500	8.63×10^{-5}	3.80×10^{-4}	5.84×10^{-10}	9.02×10^{-8}	7.78×10^{-9}	2.41×10^{-8}
600	8.08×10^{-5}	2.91×10^{-4}	5.62×10^{-10}	3.19×10^{-8}	3.56×10^{-10}	4.11×10^{-9}