

残差分裂自适应物理信息神经网络求解偏微分方程*

范昆昆¹, 张皓然², 岳煜铖¹, 袁冬芳²

(1. 内蒙古科技大学 自动化与电气工程学院, 内蒙古 包头 014010;

2. 内蒙古科技大学 理学院, 内蒙古 包头 014010)

摘要: 物理信息神经网络 (PINN) 损失函数之间的量级差异, 导致训练过程收敛缓慢, 有时甚至会在某些区域训练失败. 为解决这一挑战, 本文提出了一种融合残差分裂和权重自适应的 PINN 模型. 该方法通过将主导 PINN 训练过程的偏微分方程 (PDE) 残差项, 按照区域分解的方式分裂为多个独立子项, 并采用权重自适应加权策略, 自动调节各个子项之间的权重, 从而改善了 PINN 的收敛性. 该方法弥补了全局残差策略忽略和抹平局部特征的缺陷, 通过分裂子项的方式增加了对局部特征的关注, 改善了优化过程的效率, 从而提升了求解精度. 数值实验结果表明, 本文方法不仅在精度上超越了现有几种模型, 且达到了 2~3 个数量级的提升, 计算效率也表现出优越性能.

关键词: 物理信息神经网络; 自适应权重; 残差分裂; 区域分解

中图分类号: O241.82 **文献标志码:** A **DOI:** 10.21656/1000-0887.460018

Residual Splitting Adaptive Physics-Informed Neural Networks for Solving Partial Differential Equations

FAN Kunkun¹, ZHANG Haoran², YUE Yucheng¹, YUAN Dongfang²

(1. School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia 014010, P.R. China;

2. School of science, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia 014010, P.R. China)

Abstract: The magnitude difference between the loss functions of physical-informed neural networks (PINN) leads to a slow convergence of the training process and sometimes even training failure in some regions. To address this challenge, a physical-informed neural network model incorporating residual splitting and weight self-adaptation was proposed. The method improves the convergence of PINN by splitting the residual terms of PDE dominating the training process of PINN, into multiple independent components according to the domain decomposition, and adopts a self-adaptive weighting strategy to automatically adjust the weights among the components, thus promoting the convergence of PINN. This method makes up for the defects of the global residual strategy ignoring and smoothing out the local features, and increases the attention to the local features by splitting the subterms, which improves the efficiency of the optimization process, and thus enhances the solution

* 收稿日期: 2025-01-24; 修订日期: 2025-03-24

基金项目: 国家自然科学基金地区科学基金 (12261067; 12361088); 内蒙古自然科学基金 (2022MS01008); 内蒙古科技大学基础研究业务费专项资金 (2024QNJ052)

作者简介: 范昆昆 (1998—), 男, 硕士生 (E-mail: fankunkun914@163.com);

袁冬芳 (1985—), 女, 副教授, 硕士 (通信作者. E-mail: yuandf@imust.edu.cn).

引用格式: 范昆昆, 张皓然, 岳煜铖, 袁冬芳. 残差分裂自适应物理信息神经网络求解偏微分方程[J]. 应用数学和力学, 2026, 47(5): 655-667.

accuracy. Through numerical experiments, the results show that, the proposed method not only surpasses the existing models in terms of accuracy, but also achieves an improvement of 2-3 orders of magnitude with superior computational efficiency.

Key words: physics-informed neural network; adaptive weighting; residual splitting; domain decomposition

0 引言

偏微分方程(PDE)是用来描述电学^[1]、热量传输^[2]和流体运动^[3]等物理现象的重要数学工具,在解决自然科学问题与实际工程应用问题中充当着非常重要的角色.快速并且准确地求解 PDE 非常重要,常用的传统数值解法有:有限差分法、有限元法、有限体积法等^[4-6].但是传统方法为了处理间断解或高维 PDE 时,往往需要足够细的网格以及采用自适应算法等精巧的方法^[7-8],从而使得代码量非常庞大和需要大量的计算资源,以至于在计算复杂性和算力资源上花费昂贵的代价^[9].

物理信息神经网络(PINN)是融合物理信息和深度学习的新范式,是一种通用求解 PDE 的框架^[10].该方法将 PDE 的控制方程和定解条件直接嵌入到神经网络损失函数中,通过最小化损失函数逼近 PDE 的解.与传统的数值方法相比 PINN 具有 4 个优势:① 无网格求解,能够轻松处理不规则区域问题;② 可以克服维数灾难,求解具有高维状态特征的 PDE;③ 能够在同一框架中进行正向和逆向建模^[11];④ 便捷地融合观测数据.目前,PINN 已经被广泛应用于流体流动、系统生物学等领域^[12-13].尽管展现了广阔的应用前景并取得了显著的成功,但因为 PDE 的复杂性以及损失函数的非凸性为优化求解过程带来了巨大挑战.

PINN 的损失函数通常由多个部分组成,不同部分在量级上存在差异,优化过程中网络可能会过度关注较大损失项,而相对忽视较小损失项^[14].这种不平衡可能导致模型训练时对某些方面过分优化,而忽略了其他重要特征或条件,从而影响整体性能和泛化能力^[15].为改善 PINN 优化过程的效率,Jagtap 等^[16]提出一种区域分解的物理信息神经网络(XPINN)方法,它将问题域划分为多个子域,并在每个子域中使用单独的神经网络学习解方案,从而降低整体优化的难度.Wight 等^[17]证明了在训练神经网络时,损失函数各项的权重对收敛起着重要的作用,基于 PINN 使用了非自适应加权的算法增强对求解 Allen-Cahn 方程的能力.Wang 等^[18]提出了一种平衡各项损失函数量级的 PINN 框架(MMPINN),为损失项添加幂指数的形式来制约损失,提高模型在处理复杂 PDE 问题时的收敛速度以及准确性.Wang 等^[19]针对 PINN 训练过程中的梯度传播问题,提出了一种基于神经正切核(NTK)的方法.Elhamod 等^[20]提出 CoPhy-PGNN 方法用于对损失函数权重进行调节,可以让更复杂的损失项主导梯度下降并进行动态细化,在不同学习阶段平衡各损失函数.Wang 等^[21]在梯度流动动力学中由于刚性问题带来的困难,提出了一种学习率退火算法,该算法在训练过程中采用反向传播的梯度统计量以达到自适应平衡损失的贡献.Song 等^[22]针对时间尺度快速变化表现出的“刚性”特点,提出为每个损失函数不同数据点附加不同的权重,并且在每个训练周期利用多个神经网络之间的对抗训练动态更新每个权重.手动调节权重通常需要大量尝试和调试,很难准确找到最佳的权重组合,且容易因为权重选择不当导致训练过程收敛缓慢或陷入局部最优.

因此,Qiu 等^[23]基于梯度正则化技术^[21],提出了一种整体约束思想的自适应损失平衡策略,构建了适用于求解大尺寸比率动态耦合热力问题的 PINN.Yang 等^[24]提出了自适应任务分解物理信息神经网络(ATD-PINN),具体内容包含两个方面:① 在每次训练过程中采用自适应权重平衡策略;② 针对不同问题设计了差异化的采样策略.对于稳态问题,该方法在整个计算域内通过多分辨率数据点采样的分阶段叠加训练,逐步增加训练数据量;对于时变问题,采用时间维度的分域采样策略,通过时间域的渐进扩展实现计算效率优化.Xiang 等^[25]提出了一种名为自适应损失平衡物理信息神经网络(lbPINN)的方法,目的是通过构建 Gauss 概率模型并运用最大似然估计来动态调整各损失项的权重,从而实现损失函数的精细平衡,进而提升训练精度.尽管 lbPINN 在整体上实现了损失项间的平衡,但 PDE 残差内部各训练点之间的误差分布并不均匀,整体平衡策略抹平并忽略了某些局部特征,从而导致精度不足.McClenny 等^[26]提出了一种软注意力机制的自适应物理信息神经网络(SAPINN),为每个训练点引入了对应的参数可调的掩码函数作为权重.这种方法确保了训练过程中每个数据点的残差都得到了充分考虑,增强了局部敏感性,提高了对复杂解的学习能力.然

而,为每个训练点引入了独立的权重参数,显著增加了模型的参数数量,可能导致训练过程中的收敛困难和求解效率降低.庞大的可调参数不仅加重了优化难度,还可能引发过拟合问题,影响模型的泛化性能.

本研究针对全局权重调整方法和逐点权重调整方案表征的局限性,提出了一种融合残差分裂的权重自适应 PINN 框架.该框架通过引入残差分裂技术,增强对局部特征的捕捉能力和强化对奇异区域的表征能力.主要特征是将整体区域细分为多个子域,在每个子域上构造损失函数,实现了对整体残差项的分裂处理.然后,引入自适应权重机制,权重参数由各损失函数的量级决定,从而实现损失函数的精确调控.该方法解决了传统方法中残差分布不均匀,局部特征被抹平忽略的问题,通过精细化管理 PDE 的残差,优化模型性能,显著提升模型在处理复杂优化问题时的能力.此外,结合一种三段式的优化策略,不仅提高模型的收敛速度,而且能够增强逼近解的精度.数值实验部分对比分析了 PINN、lbPINN、SAPINN 及本文方法的相对 L2 误差和最大误差.结果表明,本文方法不仅在精度上超越了 lbPINN 和 SAPINN,实现了 2~3 个数量级的改进,而且在计算效率方面也展示了显著优势.

1 PINN 的基本原理

考虑一般形式的 PDE 如下:

$$\mathcal{N}[u](x) = f(x), \quad x \in \Omega \subset \mathbb{R}^d, \quad (1)$$

$$\mathcal{B}[u](x) = g(x), \quad x \in \partial\Omega, \quad (2)$$

其中, $\mathcal{N}[\cdot]$ 和 $\mathcal{B}[\cdot]$ 分别为一般非线性方程和边界微分算子, x 代表空间坐标, $u(x)$ 表示定义在 d 维域 Ω 内的解函数; $f(x)$ 为源函数, $g(x)$ 表示边界或初始条件,包括初始条件(I.C.)和边界条件(B.C.).含有时间变量 t 时,一般会将时间变量视为 x 的分量进行处理^[11].在这种情况下,初始条件可以被视作时空域上的一种特殊边界条件,确保了时间和空间边界的统一处理,从而简化了求解过程.

PINN 求解 PDE 过程,主要包括三个阶段:① 通过神经网络生成试探函数;② 构造目标损失函数,包括 PDE 残差项、边界条件项和标签数据项;③ 优化目标损失函数.利用具有 L 个隐藏层的全连接神经网络构造试探解函数 $\hat{u}_\theta(x)$,其中 θ 表示可训练参数,包含神经网络权重 W 和偏差 b ,这些参数是通过优化过程最小化损失函数获得的.

构造 PINN 方法的目标损失函数如下:

$$\mathcal{L}(x, \hat{u}, \theta) = \mathcal{L}_{\text{PDE}}(x, \hat{u}, \theta) + \mathcal{L}_{\text{BC}}(x, \hat{u}, \theta), \quad (3)$$

其中

$$\mathcal{L}_{\text{PDE}}(x, \hat{u}, \theta) = \frac{1}{N_\Omega} \sum_{i=1}^{N_\Omega} (\mathcal{N}[u_\theta](x_i) - f(x_i))^2, \quad x_i \in \Omega, \quad (4)$$

$$\mathcal{L}_{\text{BC}}(x, \hat{u}, \theta) = \frac{1}{N_{\partial\Omega}} \sum_{i=1}^{N_{\partial\Omega}} (\mathcal{B}[u_\theta](x_i) - g(x_i))^2, \quad x_i \in \partial\Omega, \quad (5)$$

式中, N_Ω 和 $N_{\partial\Omega}$ 分别表示在区域 Ω 内部和在边界 $\partial\Omega$ 上的采样点数量. \mathcal{L}_{PDE} 表示 PDE 残差构成的损失项, \mathcal{L}_{BC} 表示边界/初始条件的损失项.

通过使用适当的优化算法来最小化损失函数(3),可以获取最优参数 $\theta^* = \{W, b\}$.

$$\theta^* = \arg \min_{\theta} \mathcal{L}(x, \hat{u}, \theta), \quad (6)$$

$$\hat{u}(x) \approx \mathcal{N}(\theta^*; x). \quad (7)$$

所有关于输入变量 x 和参数 θ 的梯度均可以通过自动求导技术高效计算.这不仅简化了梯度计算的过程,还提高了优化的效率和精度,使得模型能够更快速、更准确地收敛到最优解.

2 残差分裂自适应 PINN

在 PINN 求 PDE 的框架中,损失函数通常由多个不同类型的损失项构成,如式(3)所示.训练过程中,这些损失在量级上往往存在显著差异:PDE 残差损失由于内部高度的复杂性可能占据主要地位,而边界/初始条件损失项量级较小,导致各项损失函数不平衡,难以确保模型的收敛.因此,训练过程中降低 PDE 残差的

同时,平衡各损失项之间的权重变得尤为重要.合理平衡各损失之间的量级差距,可以促进模型更稳定地收敛,提高解决方案的质量.

为了降低解决复杂 PDE 的整体难度,并确保损失函数能够更均匀地优化整个区域内的各部分.如图 1 所示,我们将原始区域 Ω 均匀分裂为互不相容的 n 个子域,记作: $\Omega = \Omega_1 + \Omega_2 + \dots + \Omega_n$.相应地,在整个区域上的采样点集合也被分为互不相关的 n 个子集:

$$\Omega = \bigcup_{i=1}^n \Omega_i, \quad i = 1, 2, \dots, n. \quad (8)$$

每个子域 Ω_i 上的 PDE 损失函数表示为

$$\mathcal{L}_{\text{PDE},i}(x, \hat{u}, \theta) = \frac{1}{N_{\Omega_i}} \sum_{i=1}^{N_{\Omega_i}} (\mathcal{N}[u_\theta](x_i) - f(x_i))^2, \quad x_i \in \Omega_i. \quad (9)$$

整体的损失函数表示为

$$\mathcal{L}(x, \hat{u}, \theta) = \sum_{i=1}^n \mathcal{L}_{\text{PDE},i}(x, \hat{u}, \theta) + \mathcal{L}_{\text{BC}}(x, \hat{u}, \theta). \quad (10)$$

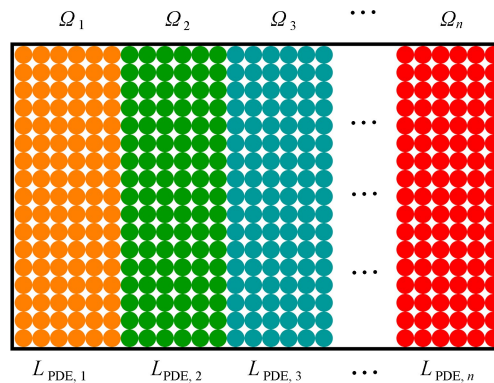


图 1 区域分裂后各区域表示情况

Fig. 1 The situation expressed by each region after regional division

尽管将 PDE 残差项根据几何分布分裂为多个子区域有助于降低训练难度,实现更精细化的局部优化,但各子损失函数组成部分的复杂性和梯度分布不均匀问题依然存在.这可能导致不同子域间的损失残差和边界/初始条件约束项之间的量级差异显著,在训练过程中过度关注高损失量级项损失,而忽视低损失量级项损失.这种不平衡现象不仅阻碍了损失函数的有效收敛,还可能限制模型对奇异值的捕捉能力.

因此,必须引入自适应损失权重平衡系数,以动态调整各项损失,确保它们在训练过程中的有效性和量级一致性.添加自适应权重系数后的损失函数表达式为

$$\mathcal{L}(x, \hat{u}, \theta, \omega) = \sum_{i=1}^n \omega_{\text{PDE},i} \mathcal{L}_{\text{PDE},i}(x, \hat{u}, \theta) + \omega_{\text{BC}} \mathcal{L}_{\text{BC}}(x, \hat{u}, \theta), \quad (11)$$

其中,表达式中 $\omega_{\text{PDE},i}$ 和 ω_{BC} 为自适应权重系数,使用以下采用 Gauss 概率分布模型结合最大似然估计的方法来进行权重的动态调整^[25]:

$$-\ln p(u | \mathcal{N}(\theta; x)) \propto \frac{1}{2\varepsilon_d^2} \|u - \mathcal{N}(\theta; x)\|^2 + \ln \varepsilon_d = \frac{1}{2\varepsilon_d^2} \mathcal{L}(\theta; x) + \ln \varepsilon_d, \quad (12)$$

式中 ε_d 为可训练参数,与神经网络中的参数 θ 一同训练.

通过为每一项损失函数添加适当的平衡系数,可以显著改善训练中的损失失衡问题,实现更好的收敛效果,并增强模型对局部奇异值的捕捉能力.优化损失函数寻找网络参数 θ 是深度学习中的关键步骤,常用的优化算法包括 Adam 优化器和 L-BFGS 优化器^[27-28].Adam 优化器具有自适应学习率、偏置校正机制以及高效的计算性能,在处理大型数据集和高维空间时表现出更快的收敛速度,并对噪声具有良好的鲁棒性.然而,它

对于超参数较为敏感,可能收敛到次优解.L-BFGS 作为一种准 Newton 法,通过近似 Hesse 矩阵利用二阶导数信息,适用于更复杂的优化任务,提供更高的收敛精度.不过,L-BFGS 内存需求较高,不适合在线学习场景.为了提高优化效率,一种有效的策略是结合两种优化器的特性^[29]:在训练初期使用 Adam 优化器进行灵活探索,随后切换至 L-BFGS 优化器加速局部收敛并提升模型精度.这种分阶段的优化策略充分利用了两种优化器的优势,通常能够显著改善最终模型的性能.

本文提出的融合残差分裂自适应的 PINN 架构和三段式优化策略来改善损失函数的收敛性,提高逼近精度,具体框架如图 2 所示:

① 预训练阶段:基于标准 PINN 框架与 Adam 优化器进行初步训练以获得初始解.这一阶段的目标是快速收敛到一个较为合理的解空间,为后续更细致的训练奠定基础.

② 残差分裂阶段:引入残差分裂算法模型,继续使用 Adam 优化器调整各子项 PDE 残差的权重.该阶段的关键在于确保即使是残差量级较小的区域也不会被忽视,进而强化对复杂局部特征的捕捉能力,提升模型细节表达力.

③ 精细微调阶段:在完成残差分裂训练后,固定所求得的各项损失对应的权重,转而采用 L-BFGS 优化器进行最终的精细微调.此阶段旨在进一步优化模型参数,确保其在全局和局部层面均达到最佳性能.

与其他方式不同的是本文方法在整个训练过程只使用一个神经网络,然后将整个训练过程分为三个阶段并结合不同的优化器来实现改善损失函数收敛性的目的.

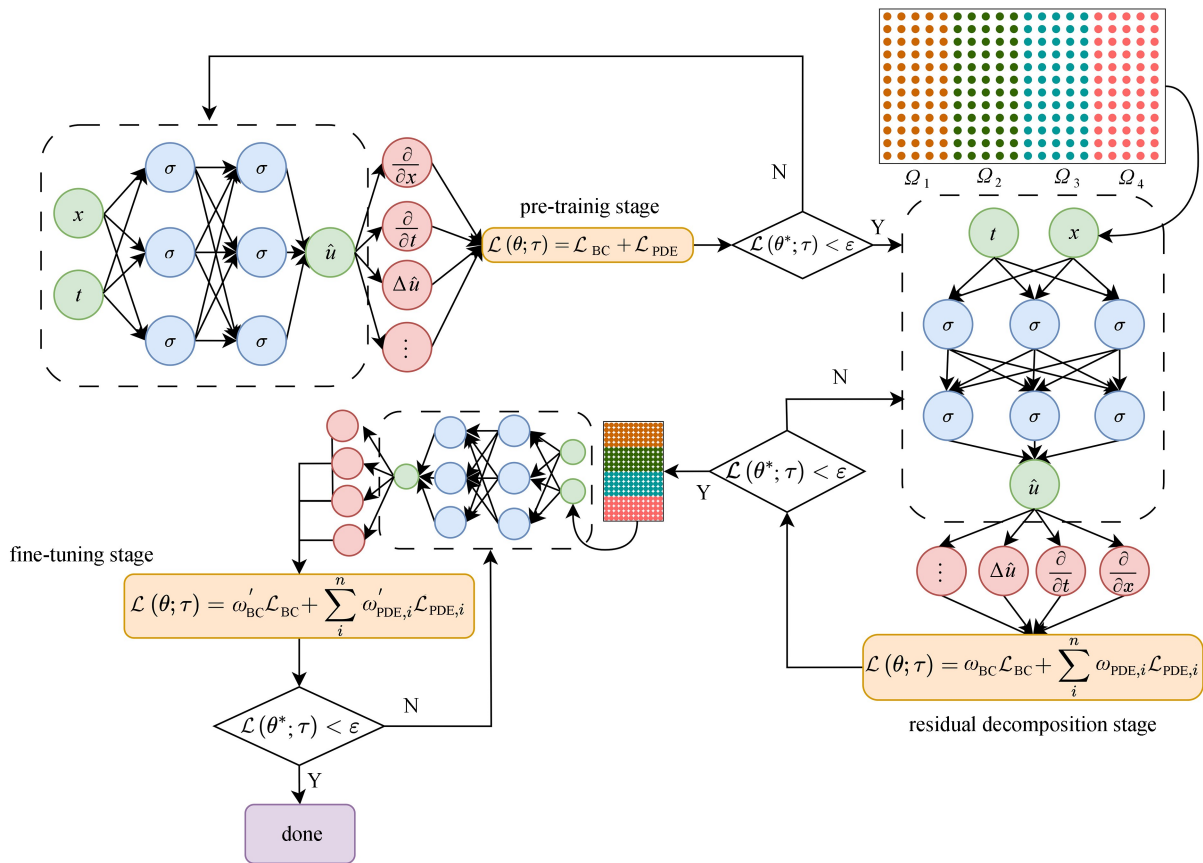


图 2 本文方法求解 PDE 的模型框架图

Fig. 2 The model frame diagram of solving PDE with this method

算法 1 残差分裂自适应 PINN

- ① 初始化网络参数 $\theta = \{W^0, b^0\}$, 初始数据点集 $\Omega = \{x_i\}$.
- ② 第一阶段:预训练
使用 Adam 优化器训练,得到网络参数 $\theta^{*1} = \{W^1, b^1\}$.

③ 第二阶段:残差分裂自适应平衡

将数据点集 Ω 分裂为 n 个子区间:

$$\Omega = \bigcup_{j=1}^n \Omega_j, \quad j = 1, 2, \dots, n.$$

④ 为每个子区间构造对应的损失函数:

$$\mathcal{L}_{\text{PDE},i}(x, \hat{u}, \theta^{*1}) = \frac{1}{N_{\Omega_i}} \sum_{i=1}^{N_{\Omega_i}} (\mathcal{N}[u_{\theta^{*1}}](x_i) - f(x_i))^2, \quad x_i \in \Omega_i, i = 1, 2, \dots, n.$$

⑤ 构造损失函数,为每项损失函数添加可自动调节的权重系数:

$$\mathcal{L}(x, \hat{u}, \theta^{*1}, \omega) = \sum_{i=1}^n \omega_{\text{PDE},i} \mathcal{L}_{\text{PDE},i}(x, \hat{u}, \theta^{*1}) + \omega_{\text{BC}} \mathcal{L}_{\text{BC}}(x, \hat{u}, \theta^{*1}).$$

⑥ 使用 Adam 优化器最小化损失函数 $\theta^{*2} = \arg \min_{\theta^{*1}} \mathcal{L}(x, \hat{u}, \theta^{*1}, \omega)$, 得到最优网络参数:

$$\theta^{*2} = \{W^2, b^2, \omega'_{\text{PDE},i}, \omega'_{\text{BC}}\}.$$

⑦ 第三阶段:微调阶段

将得到的各项权重系数固定,构造损失函数:

$$\mathcal{L}(x, \hat{u}, \theta^{*2}, \omega) = \sum_{i=1}^n \omega'_{\text{PDE},i} \mathcal{L}_{\text{PDE},i}(x, \hat{u}, \theta^{*2}) + \omega'_{\text{BC}} \mathcal{L}_{\text{BC}}(x, \hat{u}, \theta^{*2}).$$

⑧ 使用 L-BFGS 优化器最小化损失函数 $\theta^{*3} = \arg \min_{\theta^{*2}} \mathcal{L}(x, \hat{u}, \theta^{*2}, \omega)$, 得到最优网络参数:

$$\theta^{*3} = \{W^3, b^3\}.$$

⑨ 得到神经网络近似解: $\hat{u}(x) \approx \mathcal{N}(\theta^{*3}; x)$.

3 数值验证

本节针对二维 Poisson 方程、非线性 KdV 方程、平流扩散方程以及球面非线性对流扩散方程进行数值验证,对比本文方法与 PINN、lbPINN 及 SAPINN 四种方法的性能,通过比较各方法的损失函数误差 e_{loss} 、相对 L2 误差 e_2 、无穷范数误差 e_∞ , 进一步评估本文方法的稳定性.在训练过程中,所有模型均采用 sin 激活函数,学习率 $\eta:0.005$, 使用两个隐藏层、每层包含 20 个神经元的网络结构.迭代步数: PINN, 70 000 步; lbPINN, 70 000 步; SAPINN, Adam(40 000 步)+L-BFGS(30 000 步); 本文方法, Adam(10 000 步)+Adam(30 000 步)+L-BFGS(30 000 步). CPU: i5-12600kf; RAM: 32 GB; tensorflow 1. x 版本. 代码: <https://github.com/Fankunkun123/PINN>.

$$e_2 = \frac{\sqrt{\sum_{i=1}^N |\hat{u}(x^i) - u(x^i)|^2}}{\sqrt{\sum_{i=1}^N |u(x^i)|^2}}, \quad e_\infty = \max_{1 \leq i \leq N} |\hat{u}(x^i) - u(x^i)|, \quad (13)$$

式中的 $\hat{u}(x)$ 表示神经网络逼近解.

算例 1 考虑经典二维 Poisson 方程为

$$\begin{cases} \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = f(x, y), & 0 \leq x, y \leq 1, \\ u|_{x=0,1} = 0, \quad u|_{y=0,1} = 0, \end{cases} \quad (14)$$

其中, $f(x, y)$ 是已知的源项,在边界上满足齐次 Dirichlet 边界条件,方程的解析解为

$$u(x, y) = \sin(\omega y) (e^x - 1)(1 - x^2)(1 - y^2), \quad (15)$$

解析解中的频率参数 ω 能够调节解的空间变化率.在每次训练过程中采用时空域离散化训练策略,其边界初始条件配置 400 个采样点,区域内均匀选取 2 500 个采样点构成训练集.

表 1 展示了四种不同方法在不同频率下的损失函数误差、相对 L2 误差、最大误差以及训练所需时间.可以看出:本文提出的方法在将区域分裂为 3 份时,各种误差方面显著优于其他三种方法,而且无穷范数误差精度提高了 2~4 个数量级.本文方法在误差精度存在明显优势下,其训练时间与 PINN、lbPINN 基本持平,与 SAPINN 方法相比效率高近 7 倍.

表 2 展示了本文方法在不同残差分裂策略下误差对比结果.当不进行区域分裂的情况下,相当于是在自适应权重损失函数的基础上进行三段式训练.数据表明,在对整体 PDE 区域上施加自适应权重策略并不能有效降低训练难度,但通过 L-BFGS 优化器进行深度优化后,相较于 lbPINN 全过程使用 Adam 优化器训练的方案,其误差指标只有微弱的提升.当采用残差分裂策略时,通过将区域均匀分裂为若干子域,使得各子域损失函数能够更有效地捕捉局部特征,提升求解精度.值得注意的是,本文方法在 3~6 子域划分条件下均保持稳定的误差表现 E-3 量级,显现出不同分裂策略在训练精度和训练效率上都具有良好鲁棒性.随着子域数量增加自适应权重参数也随之增多,会导致优化变得困难,损失函数收敛效果变差,致使预测误差出现明显反弹现象.

表 1 算例 1 在不同频率下的误差、训练时间

Table 1 Errors and training times at different frequencies for example 1

method	ω	subdomain	e_{loss}	e_2	e_∞	training time/s
PINN	15	-	6.92E-2	7.64E-2	8.14E-2	190.63
	20	-	7.40E-1	3.00E-1	1.86E-1	189.02
	25	-	3.26	8.75E-1	8.13E-1	195.78
lbPINN	15	-	1.58E-2	1.55E-2	1.57E-2	180.67
	20	-	9.34E-3	1.82E-2	1.83E-2	199.56
	25	-	2.63E-2	9.41E-2	5.20E-2	210.10
SAPINN	15	-	6.12	3.76	1.78	1 554.46
	20	-	2.85E-1	4.20	2.89	1 499.22
	25	-	6.84E-2	6.59E-1	2.10E-1	1 514.71
our approach	15	(3,1)	7.34E-7	8.03E-4	7.84E-4	196.52
	20	(3,1)	2.49E-6	1.20E-3	8.41E-4	180.71
	25	(3,1)	1.17E-5	3.45E-3	2.29E-3	196.85

表 2 不同残差分裂策略的误差

Table 2 Errors of different residual splitting strategies

method	ω	subdomain	e_{loss}	e_2	e_∞	training time/s
our approach	25	(1,1)	1.57E-4	3.59E-2	2.02E-2	202.11
		(3,1)	1.17E-5	3.45E-3	2.29E-3	196.85
		(5,1)	8.84E-5	4.87E-3	5.34E-3	212.77
		(3,2)	1.07E-4	3.55E-3	1.19E-3	199.49
		(3,3)	4.72E-4	2.10E-2	1.90E-2	217.94

图 3 展示了在频率 $\omega = 25$ 的情况下,不同方法的损失函数收敛曲线.可以看出,本文方法在损失函数误差方面显著优于其他三种方法,不仅收敛速度更快,而且在三个阶段的训练过程中均表现出极小的波动,体现了出色的收敛性能.图 4 展示了算例 1 在 $\omega = 25$ 时,四种不同方法的预测解和最大误差.可以看出 PINN 与 SAPINN 在求解精度方面明显不足,其预测解与解析解之间的误差较大.本文方法比 lbPINN 表现出更为优越的性能,不仅提高了预测解的准确性,还增强了解析解的稳定性和鲁棒性.

算例 2 考虑用于描述浅水波的单向传播特性的 1+1 维非线性 KdV 方程:

$$u_t + 6uu_x + u_{xxx} = 0, \quad (x, t) \in (-30, 30) \times (0, 30), \tag{16}$$

其解析解为

$$u(x, t) = \frac{1}{2} \lambda \operatorname{sech}^2\left(\frac{1}{2} \sqrt{\lambda} (x - \lambda t + x_0)\right), \tag{17}$$

式中, $\lambda > 0$ 为波速, x_0 为任意常数(这里 $x_0 = 0$), sech^2 表示为一个孤立子波并且在传播过程中形状保持不变.在每次训练过程中采用时空域离散化训练策略,其边界初始化条件配置 600 个采样点,区域内均匀选取 11 250 个采样点构成训练集.

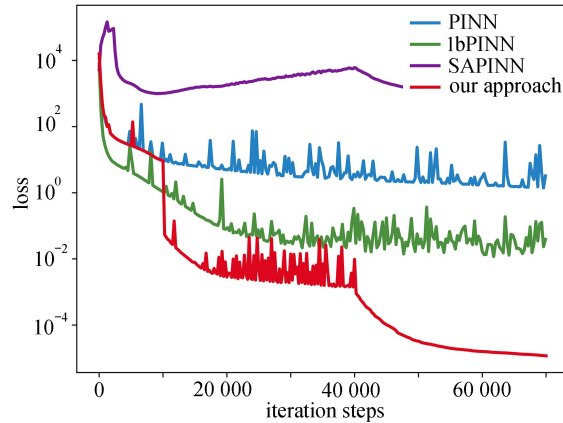


图3 不同方法的损失函数收敛曲线

Fig. 3 Loss function convergence curves for different methods

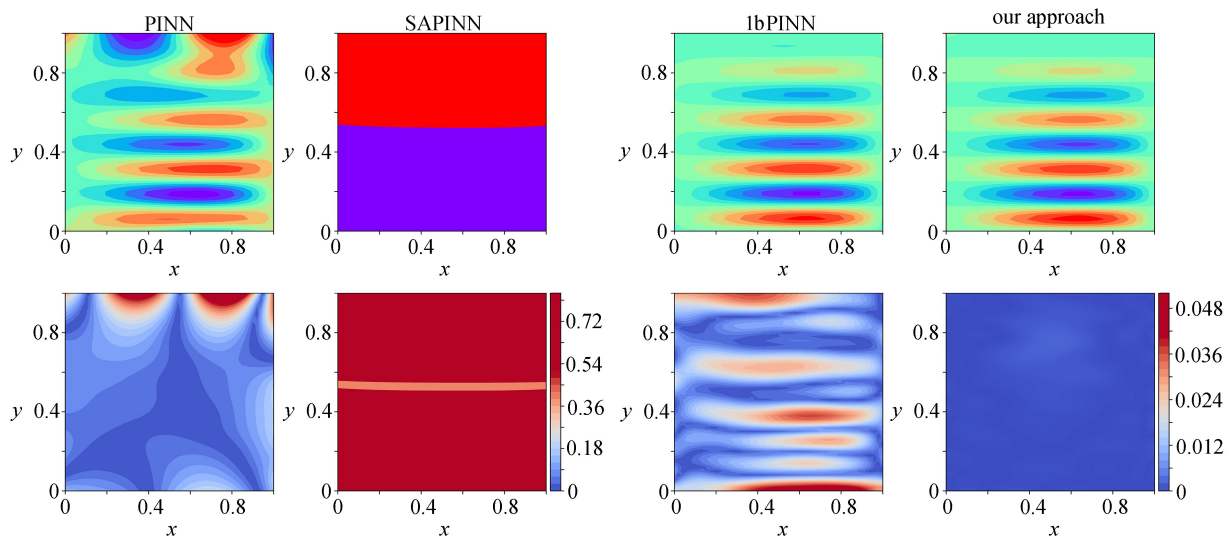


图4 四种不同方法的预测解和误差分布

Fig. 4 Predicted solutions and error distributions for 4 different methods

注 为了解释图中的颜色,读者可以参考本文的电子网页版本,后同。

表3对比了四种方法在求解不同的波速 λ 下KdV方程时的表现,包含损失函数误差、相对L2误差、最大误差以及训练时间.从表中数据可以看出,当波速 $\lambda = 0.2$ 时,PINN、lbPINN和SAPINN三种方法仍能在求解精度上保持一定的水平.但随着波速 λ 的增加,PINN、lbPINN和SAPINN三种方法的求解精度显著下降;而本文方法将区域分裂为5份时,求解精度相对于其他三种方法提升了3~4个数量级,尽管在训练速度方面与PINN和lbPINN相当,但相较于SAPINN,该方法的训练效率提高了近10倍.在提高求解精度的同时本文方法并未牺牲训练效率,在精度和效率两方面均展现出明显的优势.

图5展示了波速 $\lambda = 0.6$ 时四种不同方法损失函数收敛曲线.可以看出,尽管PINN、lbPINN和SAPINN三种方法的损失函数随着训练步数的增加有所下降,但它们并未达到理想的收敛效果.而本文方法在整个训练过程中表现出平滑且稳定的收敛性,显著优于其他三种方法.图6展示了KdV方程在 $\lambda = 0.6$ 时的预测解及对应的误差分布.可以看出,本文方法的预测解与解析解高度吻合.相较于其他三种方法,不仅能够准确捕捉KdV方程复杂的非线性动态特性,还能有效减少解析解中的波动和不稳定性,确保了模型在处理此类PDE时具有更高的精度和可靠性.

表 3 在不同波速情况下四种方法误差、时间对比
Table 3 Comparison of errors and times of 4 methods under different wave speeds

method	λ	subdomain	e_{loss}	e_2	e_∞	training time/s
PINN	0.2	-	2.07E-8	3.78E-2	5.81E-3	109.72
	0.4	-	1.12E-6	3.90E-1	1.14E-1	111.06
	0.6	-	2.25E-8	5.34E-2	3.20E-2	114.39
lbPINN	0.2	-	1.20E-11	3.01E-2	4.08E-3	117.08
	0.4	-	2.19E-11	5.35E-2	2.04E-2	123.77
	0.6	-	6.66E-11	2.72E-1	1.36E-1	146.07
SAPINN	0.2	-	9.39E-6	1.80E-2	3.37E-3	981.95
	0.4	-	1.60E-6	8.82E-3	2.76E-3	962.94
	0.6	-	1.02E-5	8.93E-1	2.96E-1	1019.10
our approach	0.2	(5,1)	5.74E-16	1.52E-5	1.54E-6	103.90
	0.4	(5,1)	1.34E-13	9.01E-5	2.94E-5	101.15
	0.6	(5,1)	8.18E-14	9.10E-5	4.28E-5	103.82

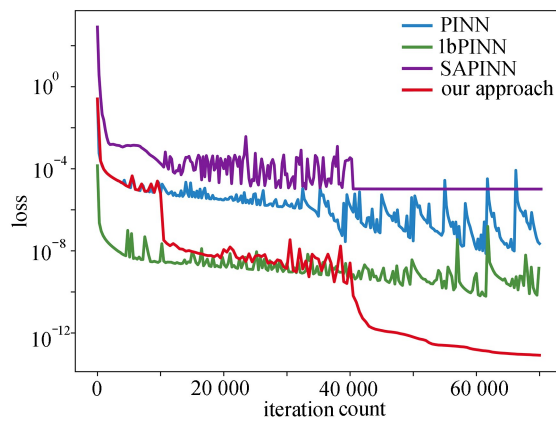


图 5 不同方法损失函数收敛曲线

Fig. 5 Convergence curves of loss functions for different methods

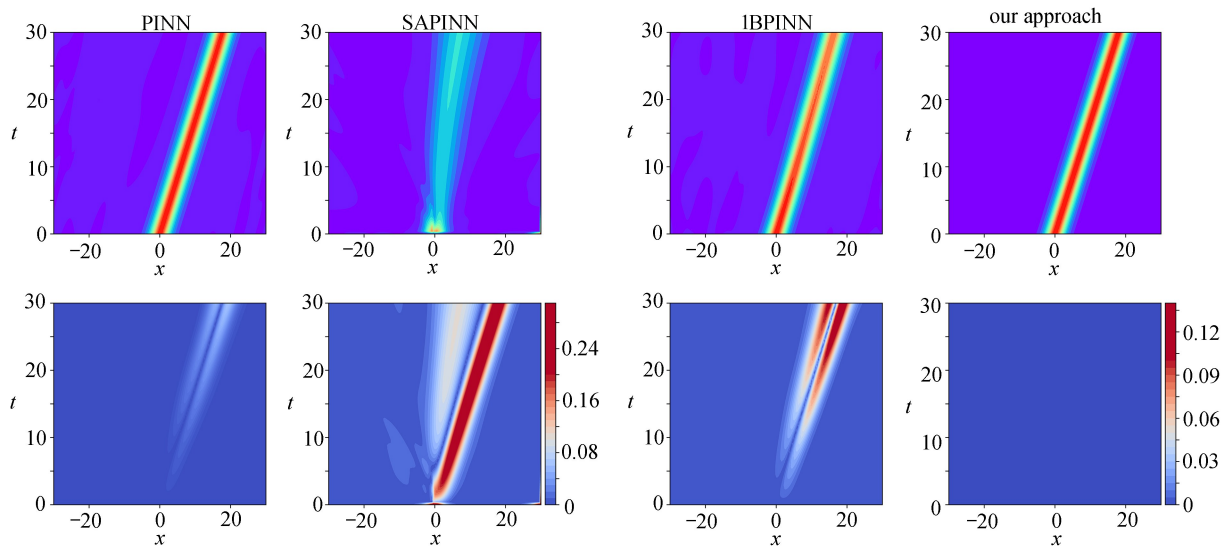


图 6 四种不同方法的预测解和误差分布

Fig. 6 Predicted solutions and error distributions for 4 different methods

算例 3 考虑广泛应用于流体力学、大气科学等领域的 1+1 维平流方程:

$$\frac{\partial u}{\partial t} + \lambda \frac{\partial u}{\partial x} = 0, \quad (x, t) \in (a, b) \times (0, T), \quad (18)$$

满足以下解析解:

$$u(x, t) = \sin(\pi(x - \lambda t)), \quad (19)$$

其中, λ 为波速, 表示沿 x 轴的扩散速率, $a = -1, b = 1, T = 1$. 在每次训练过程中采用时空域离散化训练策略, 其边界初始化条件配置 300 个采样点, 区域内均匀选取 2 500 个采样点构成训练集. 表 4 对比了四种方法在不同扩散速率条件下的性能, 包含了损失函数误差、相对 L2 误差、最大误差以及训练时间. 可以看出, 在区域被划分为四个子区域时, 本文方法在精度上相较于其他三种方法有着明显的提升, 而采用全局数据点添加自适应权重的方法在求解该方程时, 却无法有效捕捉解的形状特征. 尽管 CPU 计算成本与 PINN 方法相当, 但本文方法在收敛性和精度方面展现出显著的优势.

表 4 四种方法在不同扩散速率下的各种误差与训练时间

Table 4 Various errors vs. training times for the 4 methods at different diffusion rates

method	λ	subdomain	e_{loss}	e_2	e_∞	training time/s
PINN	8	-	3.78E-5	1.02E-2	1.86E-2	109.72
	10	-	9.84E-3	7.80E-2	1.38E-1	111.06
	12	-	1.63E-3	3.11E-2	5.27E-2	114.39
lbPINN	8	-	7.39E-3	1.12E-2	2.22E-2	117.08
	10	-	1.76E-3	4.61E-2	8.38E-2	123.77
	12	-	5.50E-3	1.16E-1	1.83E-1	146.07
SAPINN	8	-	4.51E-1	1.36	1.99	778.91
	10	-	4.51E-1	1.36	1.99	773.56
	12	-	4.51E-1	1.36	1.99	775.83
our approach	8	(2,2)	4.67E-10	7.75E-4	1.61E-3	103.90
	10	(2,2)	9.29E-9	3.76E-3	5.50E-3	101.15
	12	(2,2)	5.09E-10	9.31E-4	1.58E-3	103.82

图 7 展示了波速 $\lambda = 12$ 时的损失函数收敛情况. 从图中看出, 尽管 PINN、lbPINN 和 SAPINN 损失函数均有所下降, 但它们与解析解之间的误差显著. 相比之下, 本文方法在不同的扩散速率下表现出色: 其损失函数能够稳定且持续下降, 无穷范数误差维持在 E-3 量级, 并且这是在保持高效率训练的前提下实现的. 这表明, 通过采用残差分裂和分阶段训练策略, 该方法在求解 PDE 方面具有显著优势, 不仅提高了求解精度, 还确保了计算效率.

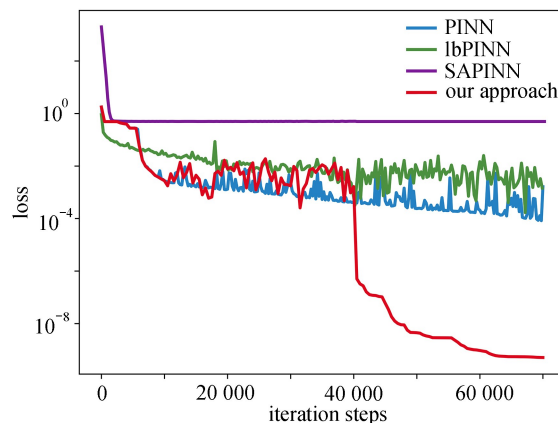


图 7 不同方法的损失函数误差收敛图

Fig. 7 Error convergence curves of the loss functions for different methods

图 8 则比较了不同方法在相同条件下的预测解和相对于解析解的无穷范数误差. 从图中可以看出 PINN 和 lbPINN 方法可以大致重现解析解的趋势, 但 SAPINN 则未能有效捕捉到解的关键特征. 相较于其他三种

方法,本文方法不仅预测解与解析解更为贴近,而且误差分布也表现出更高的均匀性.结果凸显了残差分裂技术在提高求解精度方面的显著优势.

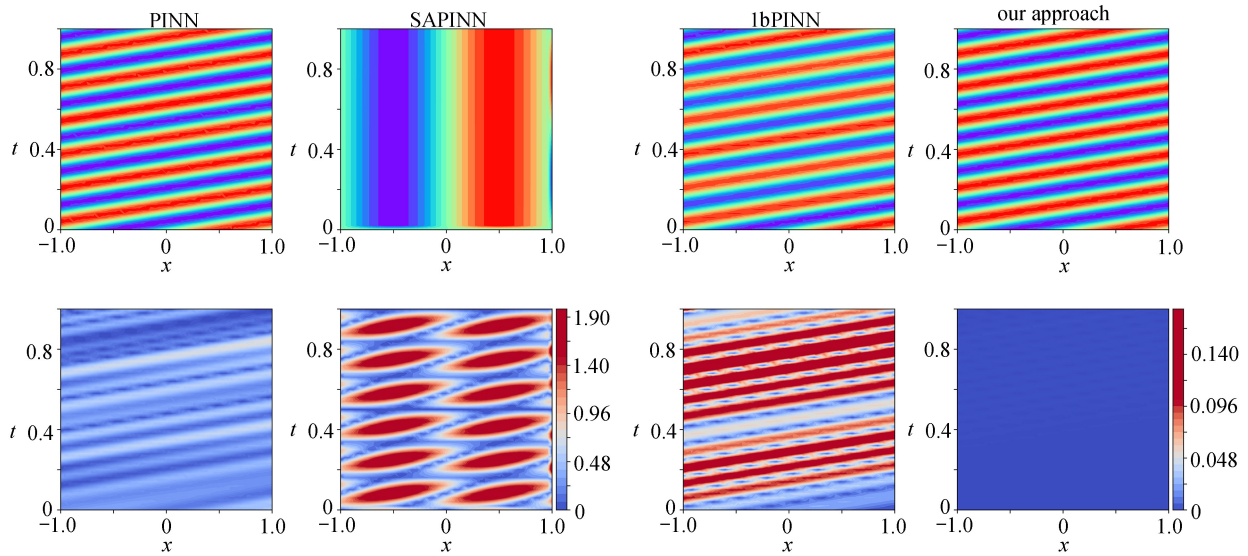


图 8 四种方法的预测解和误差分布
Fig. 8 Predicted solutions and error distributions for the 4 methods

算例 4 考虑 3+1 维的球面非线性对流扩散方程:

$$\frac{\partial u}{\partial t} = (a\Delta_s - \mathbf{v} \cdot \nabla_s + c)u(x, y, z, t) + g(u) + f, \quad 0 \leq x, y, z \leq 1, 0 \leq t \leq T. \quad (20)$$

其方程的解析解为

$$u = \sin(x + y + z) \cos(t), \quad (21)$$

其中, $g(u) = u^2, a = 1, \mathbf{v} = [0, 0, 1]^T, c = 0, f$ 为源函数^[30].在每次训练过程中采用时空域离散化训练策略,其边界初始化条件配置 500 个采样点,区域内均匀选取 2 000 个采样点构成训练集.为评估残差分裂策略对模型性能的影响,分别将时间域 T 均分为 2, 4, 8 个子域,重点考察按照时间维度划分对求解精度的影响.表 5 统计了在不同时间域尺度 ($T = 1 \text{ s}, 2 \text{ s}$) 下的相对 L2 误差和最大误差.可以看出,在时间域 ($0 \leq t \leq 2 \text{ s}$) 条件下,随着时间子域划分数量的增加,最大误差 e_∞ 呈现递减趋势,且数值解保持较好的稳定性.在未使用监督学习点的情况下,与文献[30]的结果相比,求解精度有 1~2 个数量级的提升,凸显了本方法在减少数据依赖性和提升求解精度上具有一定的优势.图 9 展示了本方法在时间域 $[0, 2 \text{ s}]$ 内采用不同的时间域分裂策略,预测解 $u(x, y, z, t)$ 在 $z = 0, t = 2 \text{ s}$ 时刻误差分布.结果表明,随着子域划分数量增加,其预测解误差呈现递减趋势,且在不同子域划分策略下误差值波动幅度较小.通过合理增加子域划分,能够有效提升模型对复杂时空问题的逼近精度.

表 5 不同时间区域的不同分裂策略误差对比

Table 5 Comparison of different splitting strategies' errors in different time domain

method	$T = 1 \text{ s}$				$T = 2 \text{ s}$			
	subdomain	e_2	e_∞	training time/s	subdomain	e_2	e_∞	training time/s
PINN	-	7.92E-4	8.24E-3	223.56	-	6.92E-4	3.49E-3	205.89
lbPINN	-	6.54E-4	6.85E-3	210.66	-	5.86E-4	3.79E-3	189.19
our approach	1	0.67E-5	9.66E-4	198.67	1	9.18E-4	1.10E-3	188.98
	2	8.86E-6	7.29E-5	158.82	2	8.64E-5	5.44E-4	191.56
	4	1.15E-5	1.38E-4	165.73	4	8.18E-6	4.44E-5	229.71
	8	1.23E-5	6.60E-5	248.29	8	1.22E-5	8.09E-5	252.92

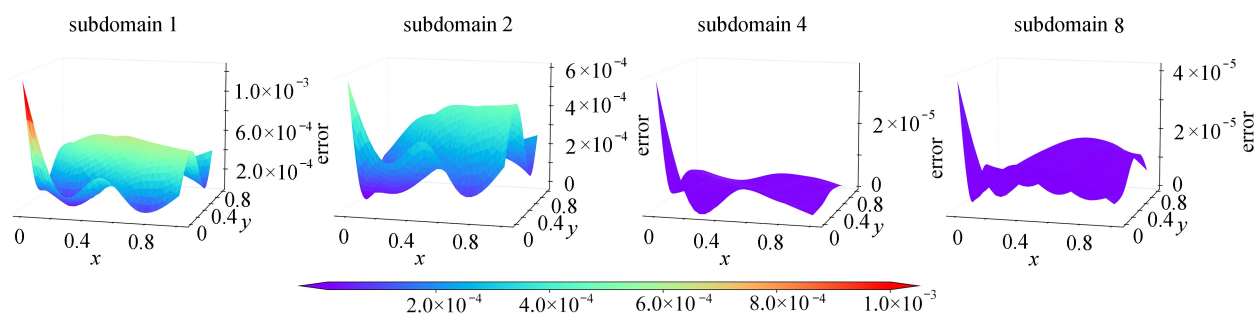


图9 不同区域分裂策略下预测解 $u(x, y, 0, 2)$ 的误差对比

Fig. 9 Comparison of prediction errors of solution $u(x, y, 0, 2)$ under different regional division strategies

4 结 论

使用 PINN 求解 PDE 时,经常会遇到 PDE 残差项的量级过大而导致损失函数整体不平衡,这将严重影响损失函数的收敛性.本文提出残差分裂自适应权重策略.其核心思想是将复杂的 PDE 残差损失拆分为多个相对独立的部分,将一个复杂的大问题转换为若干个小而集中的子问题来处理.这样做不仅使每个子损失函数能够更好地捕捉到那些在全局损失中容易被忽略的局部特性,还降低了优化的难度,并减少了全局损失掩盖局部特征的可能性.此外,为提高优化效率,在训练过程中划分为具有三个阶段的混合训练结构,通过分阶段逐步优化,每个阶段都针对特定的目标进行调整和改进,最终使得整个优化过程更加高效且精准,为解决实际应用中的优化难题提供了强有力的支持,进一步提升了模型的精确度与鲁棒性.然而,本文的思路是基于所求问题的解梯度分布不均匀导致的训练不稳定情况,因此对于此类问题相对比较有效.当问题解存在高频振荡或多尺度等特征时,采用区域分解的办法很难对此类特征进行捕捉,需要设计更精妙的算法.

参考文献 (References):

- [1] ZHANG S, GU W, ZHANG X P, et al. Dynamic modeling and simulation of integrated electricity and gas systems[J]. *IEEE Transactions on Smart Grid*, 2023, **14**(2): 1011-1026.
- [2] AVALOS G, LASIECKA I, TRIGGIANI R. Heat-wave interaction in 2/3 dimensions: optimal rational decay rate [J]. *Journal of Mathematical Analysis and Applications*, 2016, **437**(2): 782-815.
- [3] ARQUB O A. Numerical solutions for the Robin time-fractional partial differential equations of heat and fluid flows based on the reproducing kernel algorithm[J]. *International Journal of Numerical Methods for Heat & Fluid Flow*, 2018, **28**(4): 828-856.
- [4] ZHANG Y. A finite difference method for fractional partial differential equation[J]. *Applied Mathematics and Computation*, 2009, **215**(2): 524-529.
- [5] QIN X Q, MA Y C, ZHANG Y. Two-grid method for characteristics finite-element solution of 2D nonlinear convection-dominated diffusion problem[J]. *Applied Mathematics and Mechanics*, 2005, **26**(11): 1506-1514.
- [6] EYMARD R, GALLOUËT T, HERBIN R. *Handbook of Numerical Analysis: Finite Volume Methods*[M]. Elsevier, 2000, **7**: 713-1018.
- [7] NOCHETTO R H, SIEBERT K G, VEESER A. Theory of adaptive finite element methods: an introduction [C]//*Multiscale, Nonlinear and Adaptive Approximation*. Berlin, Heidelberg: Springer, 2009: 409-542.
- [8] RASP S, PRITCHARD M S, GENTINE P. Deep learning to represent subgrid processes in climate models[J]. *Proceedings of the National Academy of Sciences*, 2018, **115**(39): 9684-9689.
- [9] LENG K, NISSEN-MEYER T, VAN DRIEL M, et al. AxiSEM3D: broad-band seismic wavefields in 3-D global earth models with undulating discontinuities[J]. *Geophysical Journal International*, 2019, **217**(3): 2125-2146.
- [10] RAISSI M, PERDIKARIS P, KARNIADAKIS G E. Physics-informed neural networks: a deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations[J]. *Journal of Computational Physics*, 2019, **378**: 686-707.

- [11] LU L, MENG X, MAO Z, et al. DeepXDE: a deep learning library for solving differential equations[J]. *SIAM Review*, 2021, **63**(1): 208-228.
- [12] SUN L, GAO H, PAN S, et al. Surrogate modeling for fluid flows based on physics-constrained deep learning without simulation data[J]. *Computer Methods in Applied Mechanics and Engineering*, 2020, **361**: 112732.
- [13] ARZANI A, WANG J X, D'SOUZA R M. Uncovering near-wall blood flow from sparse data with physics-informed neural networks[J]. *Physics of Fluids*, 2021, **33**(7): 071905.
- [14] WANG Y, BAI J, LIN Z, et al. Artificial intelligence for partial differential equations in computational mechanics: a review[PP/OL]. arXiv(2024-11-23)[2025-03-24]. <https://arxiv.org/abs/2410.19843v2>.
- [15] WU H, LUO H, MA Y, et al. RoPINN: region optimized physics-informed neural networks[PP/OL]. arXiv(2024-10-23)[2025-03-24]. <https://arxiv.org/abs/2405.14369v3>.
- [16] JAGTAP A D, KARNIADAKIS G E. Extended physics-informed neural networks (XPINNs): a generalized space-time domain decomposition based deep learning framework for nonlinear partial differential equations[J]. *Communications in Computational Physics*, 2020, **28**(5): 2002-2041.
- [17] WIGHT C L, ZHAO J. Solving Allen-cahn and cahn-Hilliard equations using the adaptive physics informed neural networks[J]. *Communications in Computational Physics*, 2021, **29**(3).
- [18] WANG Y, YAO Y, GUO J, et al. A practical PINN framework for multi-scale problems with multi-magnitude loss terms[J]. *Journal of Computational Physics*, 2024, **510**: 113112.
- [19] WANG S, YU X, PERDIKARIS P. When and why PINNs fail to train: a neural tangent kernel perspective[J]. *Journal of Computational Physics*, 2022, **449**: 110768.
- [20] ELHAMOD M, BU J, SINGH C, et al. CoPhy-PGNN: learning physics-guided neural networks with competing loss functions for solving eigenvalue problems[J]. *ACM Transactions on Intelligent Systems and Technology*, 2022, **13**(6): 1-23.
- [21] WANG S, TENG Y, PERDIKARIS P. Understanding and mitigating gradient flow pathologies in physics-informed neural networks[J]. *SIAM Journal on Scientific Computing*, 2021, **43**(5): A3055-A3081.
- [22] SONG Y, WANG H, YANG H, et al. Loss-attentional physics-informed neural networks[J]. *Journal of Computational Physics*, 2024, **501**: 112781.
- [23] QIU L, WANG Y, GU Y, et al. Adaptive physics-informed neural networks for dynamic coupled thermo-mechanical problems in large-size-ratio functionally graded materials[J]. *Applied Mathematical Modelling*, 2025, **140**: 115906.
- [24] YANG J, LIU X, DIAO Y, et al. Adaptive task decomposition physics-informed neural networks[J]. *Computer Methods in Applied Mechanics and Engineering*, 2024, **418**: 116561.
- [25] XIANG Z, PENG W, LIU X, et al. Self-adaptive loss balanced physics-informed neural networks[J]. *Neurocomputing*, 2022, **496**: 11-34.
- [26] MCCLENNY L, BRAGA-NETO U. Self-adaptive physics-informed neural networks using a soft attention mechanism[PP/OL]. arXiv(2024-06-18)[2025-03-24]. <https://arxiv.org/abs/2009.04544v5>.
- [27] KINGMA D P, BA J. Adam: a method for stochastic optimization[PP/OL]. arXiv(2017-01-30)[2025-03-24]. <https://arxiv.org/abs/1412.6980v9>.
- [28] BYRD R H, LU P, NOCEDAL J, et al. A limited memory algorithm for bound constrained optimization[J]. *SIAM Journal on Scientific Computing*, 1995, **16**(5): 1190-1208.
- [29] WANG Y, LAI C Y. Multi-stage neural networks: function approximator of machine precision[J]. *Journal of Computational Physics*, 2024, **504**: 112865.
- [30] 闵建, 傅卓佳, 郭远. 课程-迁移学习物理信息神经网络用于曲面长时间对流扩散行为模拟[J]. 应用数学和力学, 2024, **45**(9): 1212-1223. (MIN Jian, FU Zhuojia, GUO Yuan. Curriculum-transfer-learning-based physics-informed neural networks for simulating long-term-evolution convection-diffusion behaviors on curved surfaces[J]. *Applied Mathematics and Mechanics*, 2024, **45**(9): 1212-1223. (in Chinese))