arXiv:2301.04793v1 [physics.comp-ph] 12 Jan 2023

用于动力偏微分方程的物理信息神经网络并非通过训练更深层网络，而是学习平流和时间变化

Siddharth Rout

不列颠哥伦比亚大学机械工程系，加拿大不列颠哥伦比亚省温哥华。

贡献作者：sidrout@mail.ubc.ca；

摘要

本文研究了物理信息神经网络（PINNs）的概念与技术，并指出其局限性，以提高其近似求解动力方程的效率。本文探讨了有望提升该技术稳健性的研究方向，以便求解偏微分方程。研究发现，物理信息神经网络在应对更强的平流和更长的时间跨度时可能会失效。此外，优化函数和约束设置也需要更加巧妙。对于许多问题，浅层网络已足够胜任，而功能强大的深层网络反而会失效。本文建议采用基于储备池计算的循环神经网络架构来求解动力学问题。

关键词：动力系统、混沌、物理信息神经网络、物理信息神经网络、Kuramoto–Sivashinsky 方程、常微分方程解、偏微分方程解、微分方程、循环神经网络、储备池计算

1 引言

微分方程的求解一直是世界上几乎所有领域的重要课题，例如金融、力学、气象学等。从牛顿和莱布尼茨的时代开始，求解微分方程就一直是推动世界发展的核心内容。并非所有微分方程都

能通过手算求解，并且会带来局限性，尤其是在方程中出现多个自变量或构成方程组时。然而，解决这些方程是当务之急，因此，随着函数运算和变换在概念上的局限，我们将关注点从精确解转向了近似解。于是，人们开始利用极限的基本原理、离散化和数值分析来求解方程[1]。常见的此类技术包括有限元法、有限体积法、有限差分近似等。这些方法具有普适性，这正是其优势所在。利用这些技术，我们几乎可以求解任意几何形状下的任何方程。然而，随着复杂性的增加，也会伴随一些问题，例如近似的准确程度和计算成本。离散化会产生一长串需要求解的简化近似方程。尽管我们知道如何求解这些方程，但实际求解仍需要借助计算机来完成繁重的数学计算。为了获得稳定且精确的解，这一过程会消耗大量时间和精力[1]。尽管我们能够计算这个世界上的许多事物，但时间、资源和金钱都很有限，而问题却可能无穷无尽。一类称为动力系统的微分方程，在时间超过某个范围后很难求解，这是有原因的。动力系统难以求解，是因为其解可能发生分岔，从而使系统变得混沌。在混沌系统中，初始条件或方程系数的微小变化，都可能导致截然不同的结果。这有时被称为“蝴蝶效应”。本项目旨在开发一种函数近似方法，以期替代计算成本高昂的动力系统求解器。一维库拉莫托-西瓦辛斯基方程已用于试验和研究[2, 3]。

函数近似或解析解以其轻量级特性而闻名[4]。这些技术可以消除全阶离散近似中明显存在的三种主要误差，即不稳定性、不准确性和偏移[5]。动力系统对不准确性尤为敏感，因为它们对初始条件的敏感性极为明显。函数近似的一项优势是能够进行校正和再现，此外还具有外推能力。在各类函数近似器中，神经网络一直是出色的通用近似器候选者[6, 7]。过去十年间，深度神经网络（本质上由多个神经网络层构成）凭借捕捉高维强非线性的能力，被用于各种复杂的回归问题。这些模型直接根据数据拟合输入到输出的映射[7]。还可以通过在输入域内的随机点上拟合后优化残差，使神经网络满足微分方程。这类模型称为物理信息神经网络，通常简称为物理信息神经网络（PINN）[4, 5, 8, 9]。

使用物理信息神经网络求解偏微分方程已得到科学界的普遍认可。与传统方法相比，这些方法有诸多优点。其主要优点是能够求解各种难以

求解的问题。此外，这些方法无需进行网格划分和离散化，而这有时是一项艰巨的任务。另一个优点是，与其他解析模型不同，它们无需依赖全阶解的数据来设定参数。然而，由于这些技术问世不久，在求解复杂方程方面还不够稳健，例如双曲方程、强非线性方程、强对流方程[5], 混沌动力系统、耦合方程组、激波方程等。

2 动力偏微分方程

世界上的活动大多发生在四维时空中，即三个空间维度和一个时间维度。每增加一个维度，复杂性便会增加一层。动力系统通常指描述系统状态随时间变化关系的函数。亨利·庞加莱是最早发现动力系统特殊行为的人。这些动力系统理论与研究复杂动力学行为密切相关，而这类行为通常以微分方程的形式表示，从而构成连续动力系统。在这一领域，主要关注吸引子、混沌、分形和分岔，它们定性地解释状态的长期行为。这有助于理解湍流、风暴、流体混合、环境变化、经济变化、行星运动等动力学事件的演变。

动力系统理论的实际应用包括寻找结构稳定性、李雅普诺夫时间、分岔点、位置跟踪和定量近似；这些因素以不同方式决定了某一特定时刻状态的可预测性。预测动力系统的行为是一项艰巨任务。在计算机问世之前，进行预测需要采用复杂的数学技术，而且这些技术针对的是特定类别的动力系统。这些方程有时是最难求解的微分方程之一。此外，考虑到上述其他因素，要准确预测这类系统并非易事。

3 案例选择

以下案例指出了清晰求解微分方程时面临的两个主要困难。相关概念将结合所提及方程的术语和框架进行说明。这两个方程是分析物理信息神经网络理论的理想示例。

3.1 一维稳态对流-扩散方程

下面的微分方程是描述一维稳态流动中平流与扩散共同作用现象的控制方程。

$$\alpha u_x = u_{xx}$$

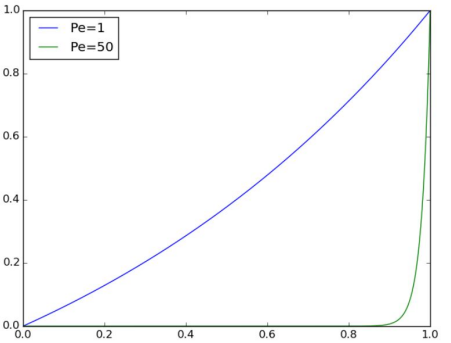

图 1 稳态对流-扩散方程的解

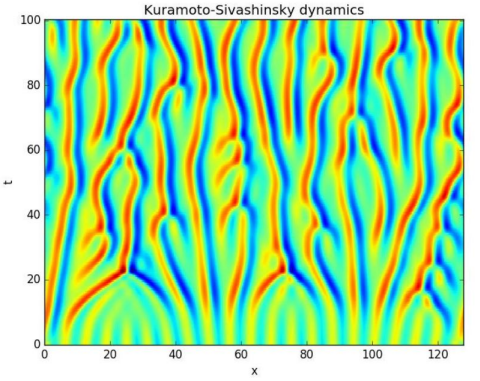

图 2 一维库拉莫托-西瓦辛斯基方程的解

请注意， α 是方程中平流项的权重。这意味着 α越大，平流效应就越显著，它会引入方向性特征，因此离散近似会变得更加困难。图 1 展示了对流主导程度不同时解的差异。佩克莱数越大，平流就越显著。图中比较了 Pe 为 1 和 50 时的解。因此，数值积分会出现快速增长的误差，使解变得不稳定且不准确[10]。因此，人们发展出一大类高阶方法来解决这一特定问题。

3.2 一维 Kuramoto–Sivashinsky方程

下方的方程是一维库拉莫托-西瓦辛斯基方程

$$u_t + \alpha u u_x + \beta u_{xx} + \gamma u_{xxxx} = 0$$

其线性形式如下：

$$u_t + \alpha u_x + \beta u_{xx} + \gamma u_{xxxx} = 0$$

该方程具有平流、扩散和耗散效应。它是解对初始条件极其敏感的方程之一[2, 3]。差分方程展开中的高阶项与误差随时间传播密切相关，因此对误差传播非常敏感。图 2 展示了使用新罕布什尔大学的 Mahatab Lak 等人开发的 Julia 代码求解一维 Kuramoto–Sivashinsky 方程的一个案例。

4 神经网络作为通用函数逼近器

George Cybenko 首先证明了使用具有 Sigmoid 激活函数的神经网络时，任意宽度的情形也成立 1989[6]。同年晚些时候，Hornik 等人证明了多层前馈网络是通用近似器[7]。多层人工神经网络由加权输入和构成，并将其传入双曲正切函数、Sigmoid函数等非线性（激活）函数中。这使其能够构成极具表达能力的高度非线性函数，并拥有大量可训练参数（权重和偏置），从而实现通用逼近。

5 物理信息神经网络

将物理信息融入神经网络这一概念由 Lagaris 等人提出[4]。在 20 世纪 90 年代末，他们使用神经网络作为试探函数来求解微分方程，并通过在定义域内的多个点使用自动微分（AutoGrad，一种自动微分技术）来减小残差。通过手动修改神经网络函数，将边界约束强加于其中。2017 年，Raissi 等人采用更精确的自动微分和更深的网络来逼近更复杂的问题[8, 9]。他们工作的创新之处在于提出了降低残差的损失函数。他们没有通过修改试探函数来手动强加约束，而是将满足条件的数据点的均方误差作为约束项，与残差均方相加，使试探函数能够拟合边界和初始约束。这使该技术具有很强的通用性。几乎所有微分方程都可以采用这种技术来求解，他们将其命名为物理信息神经网络。

5.1 物理信息神经网络的进展

未知解 $\boldsymbol{u}(t,\ \boldsymbol{x})$ 由深度神经网络$\boldsymbol{u}_{\boldsymbol{\theta}}(t,\ \boldsymbol{x})$ 表示，其中 $\boldsymbol{\theta}$ 表示网络的所有可调参数（例如权重和偏置）。通过最小化以下损失函数，可以训练物理信息模型。

$$\mathcal{L}(\boldsymbol{\theta}) = \lambda_{ic}\mathcal{L}_{ic}(\boldsymbol{\theta}) + \lambda_{bc}\mathcal{L}_{bc}(\boldsymbol{\theta}) + \lambda_{r}\mathcal{L}_{r}(\boldsymbol{\theta}),$$

其中

这里 $\left\{\boldsymbol{x}_{ic}^{i}\right\}_{i=1}^{N_{ic}}$，$\left\{t_{bc}^{i}, \boldsymbol{x}_{bc}^{i}\right\}_{i=1}^{N_{bc}}$ 和 $\left\{t_{r}^{i}, \boldsymbol{x}_{r}^{i}\right\}_{i=1}^{N_{r}}$ 可以是固定网格的顶点，也可以是在每次梯度下降算法迭代中随机采样的点。超参数 $\{\lambda_{ic}, \lambda_{bc}, \lambda_{r}\}$ 允许为每个单独的损失项分配不同的学习率，从而在模型训练过程中平衡它们之间的相互作用[12, 13]。这些权重可以由用户指定，也可以在训练过程中自动调优。

5.2 物理信息神经网络相较于其他《神经网络》的优势

这项技术的主要优势在于，无需物理数据即可训练解析模型。此外，这项技术具有普适性：基于完全相同的概念，可以求解各种方程[4,5, 8, 9, 12, 13]。以往的模型需要根据耦合方程的数量、边界条件等调整学习函数，以施加约束。作为一种强大的近似函数，它可以同时处理三类主要误差，即不稳定性、不准确性和漂移误差。有限数值方法则分别处理这些问题[5]。这种技术非常适合用于开发稳健的高阶方法。神经网络不仅是近似器，更是一种智能近似器。因此，它可以根据局部物理特性，以不同方式运行，并呈现切换行为。尤其对于动力系统，随着时间推移，累积误差会迅速增大。物理信息神经网络是一种用于回归的优化技术，可以将实测物理数据点作为额外的损失项或正则化项，用于修正近似函数。

6 所选案例的实验

动力学问题的定性特征是具有很强的平移方差。因此，物理信息神经网络表现不佳的两个主要原因是对流主导和时间变化，下面将对此进行说明。

6.1 用于一维稳态对流-扩散问题的物理信息神经网络

佩克莱数是一个很好的无量纲参数，可用于衡量方程中平流相对于扩散的主导程度。它是对流输运速率与扩散输运速率之比。在我们的问题中，可以用对流项系数乘以定义域长度，再除以扩散项系数来量化这一比值，因此在我们的具体情形下，该值为 α。物理信息神经网络可以求解这个问题，但微分方程中的平流特性限制了其求解能力。无论我们将神经网络做得多深、多复杂，都无法求解佩克莱数高于约 8 的问题。图 3 展示了所述结果。对于佩克莱数较低的问题，可以看出没有必要使用更深、更宽的层。值得一提的是，传统数值方法无法处理该数值大于 2 的问题。因此，这些方案可用于形状

Architecture		Valid for:
Layers	Neurons per Layer	
1	2	Pe < 3.8
2	2	Pe < 4.8
3	2	Pe < 6.3
3	3	Pe < 6.5
3	10	Pe < 7.6

图 3 用于求解一维稳态对流扩散问题的优化后的架构。

Models			Loss				
C	Neurons	Optimizer	Sigmoid	Tanh	SoftPlus	Log Sigmoid	ArcTan
0.5	2	Adam	7.75615e-04	1.47605e-05	4.77046e-03	7.31412e-03	1.11537e-04
1	2		1.55154e-03	2.19862e-04	4.46882e-03	1.46074e-02	4.01917e-04
0.5	2	L-BFGS-B	1.28998e-03	1.27516e-06	1.49661e-02	1.79618e-02	6.72869e-04
1	2		1.70795e-02	1.91109e-04	1.93503e-02	1.93258e-02	1.08187e-03
0.5	4		2.55921e-06	1.59368e-05	2.43783e-03	2.39689e-03	9.42070e-05
1	4		2.81592e-06	3.33567e-04	4.54190e-03	2.79268e-03	4.18945e-04
0.5	8		9.95310e-05	5.13360e-06	1.79042e-02	2.69044e-03	2.04379e-05
1	8		3.38175e-04	1.34256e-05	1.61106e-02	9.27748e-03	1.62906e-04
1.5	8		1.92506e-03	7.63881e-05	1.70217e-02	8.23745e-03	7.14311e-04
2	8		3.10389e-04	2.35750e-04	1.77785e-02	7.77693e-03	1.16140e-04

图 4 求解一维稳态对流扩散问题时，采用不同激活函数得到的最小损失，其中 C 为对流项系数。

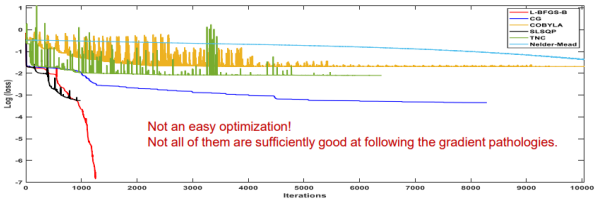

图 5 求解一维稳态对流扩散问题时，采用不同优化器的损失变化趋势

能够在保持相近精度的同时使用更大的网格。本节中的工作和图表摘自 2019 年的论文《使用物理信息机器学习求解计算流体力学问题中的数值近似》[5]。

开展参数分析是为了更深入地了解性能和有效性。研究了非线性函数以及损失优化算法变化所带来的影响。图4以表格形式记录了损失值。在各种非线性函数中，双曲正切函数和正切函数的表现始终优于 Sigmoid函数。不过，双曲正切函数明显胜出。图5展示了不同优化器下损失值随迭代次数的变化趋势。神经网络使用多种优化器进行训练，其中 L-BFGS-B 和 SL-SQP 的表现优于其他专用优化技术。BFGS 的表现尤为出色。主要原因可能是这些优化器属于二阶方法，在优化多目标函数时比其他方法表现更好。不过，可以看出，像 Adam 这样的一阶方法即使未能收敛，表现也还算不错

。一个值得注意的现象是，配置点处的偏微分方程残差与约束的拟合损失明显不在同一量级，这与常规的数据驱动神经网络学习并不相同。此外，梯度病态并不平滑，因此其他优化器难以应对。引导式优化（如爬山法）在少数情况下有所帮助。

6.2 一维库拉莫托-西瓦辛斯基的物理信息神经网络

近期有研究表明，物理信息神经网络能够求解一维库拉莫托-西瓦辛斯基方程；该方程是可能演变为混沌的标准动力学方程之一。不过，只有在平流、扩散和耗散项的系数处于一个很小的交集范围内时，这种方法才有效。如前例所示，平流占主导地位会使物理信息神经网络的优化更加困难。此处的问题不仅具有动力学特性，而且是非线性的。它包含不同阶数的空间导数，因此求解难度较大，尤其是在系统进入混沌状态时。

Wang等人提出的CausalPINN被认为是当前最先进的物理信息神经网络[13]。他们正确指出，正如Rout等人所述，残差与约束的尺度不同会导致多目标优化问题。[5] 他们通过将各项损失除以各自的累积损失进行归一化，为每个损失项设计了权重。他们是首批成功求解一维库拉莫托-西瓦辛斯基方程的研究者。我们可以利用他们提供的开源代码验证其模型。图6展示了CausalPINN在图示案例中的解随时间的变化情况。可以看出，物理信息神经网络现在能够求解复杂的动力学问题。初始正弦曲线如预期般平滑卷曲，同时满足各项约束。初始曲线呈规整的正弦函数形态，边界则始终保持在零基线上。不过，在对该方程进行常规研究时，通常会采用所有系数均为1的典型方程。即使投入相当于运行一天的计算时间，当前最先进的物理信息神经网络仍无法完成优化。记录到的总损失为33.263，其中约束损失为0.0016。这表明，尽管物理信息神经网络能够满足约束，但拟合该偏微分方程仍然困难。残差损失为1808.506，其在损失函数中的权重几乎小到无法显示。这清楚地表明了多目标优化问题的存在。

7 观察与结论

根据实验和分析，可以提出几点看法。对于配置点和拟合点处的残差平方损失项，简单相加或按均值加权后相加，再直接用于约束，两者的尺度和数量级各不相同。这是一个主要问题，因为它可能导致非帕累托最优解。还注意到，损失函数有时会陷入局部最优，而且参数超空间中的梯度路径异常棘手且不均匀[5, 12]。因此，随机一阶优化器可以奏效；否则，适用于约束优化的高阶优化器（如序列二次规划和 BFGS）能够发挥作用，而其他方法则不行[5]。更好的损失表示

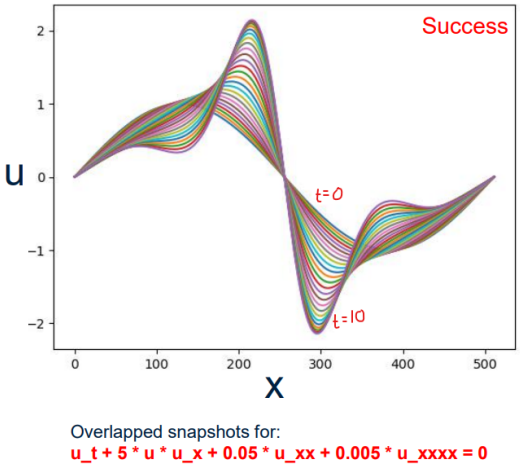

图 6 一维库拉莫托-西瓦辛斯基方程中速度（U）沿 x 轴随时间变化的叠加快照。

方式，例如采用合适的损失函数或自适应加权损失，可能会有所帮助。否则，可以通过修改网络架构或试探函数来强制满足约束，正如 Isaac Lagaris 等人所述[4]。具体而言，对于动力系统问题，循环神经网络（RNN）可能比简单的深度网络更适合[14]。Eldad Haber 等人给出了具体的解释：循环神经网络可以表示为微分方程，因此符合相关理论，能够更好地学习动力系统的微分方程[15, 16]。尤其是在混沌系统中，储备池计算已被证明表现更佳[11]。储备池计算器是一类循环神经网络，其中间节点随机排列并相互连接[11, 14]。它们具有随机循环连接。中间节点杂乱交织，但通过线性方式连接到输出层。这种交织的架构使反向传播变得困难，因此为方便起见，只有最后一层的权重是可训练的。可训练的输出层使有效的非线性网络相对于可训练参数呈线性，从而在保留强非线性的同时便于训练。显然，也可以将循环神经网络与物理信息神经网络的损失定义结合起来，以解决湍流问题和进行极端事件预测[17]。

最终，通过识别导致性能不佳的两个主要原因，我们可以解释误差，并明确求解偏微分方程动力系统的正确路径。这两个原因是对流主导和时间变化，已通过案例研究加以识别。我们可以得出结论，架构中并非总是需要深层且庞大的网络层。关键在于如何构建优化问题，以及问题的可优化性。必须妥善处理梯度病态。更深的层能够捕捉高维空间中极强的非线性

以及强耦合方程组。此外，针对时间变化这一特征，我们应使用循环神经网络，尤其是储备池网络；这类网络实际上轻量级，却表现更好。“物理信息循环神经网络（PIRNN）是求解动力系统和混沌问题的正确路径。”

参考文献

[1]J. Strikwerda，《前言》，《有限差分格式与偏微分方程》第二版，第 i–xii 页，2004 年 1 月，doi: 10.1137/1.9780898717938.fm。[2]Y. Kuramoto，《反应系统中的扩散诱导混沌》，《理论物理进展增刊》，第 346–367 页，1978 年，doi: 10.1143/ptps.64.346。[3] G. I. Sivashinsky，《层流火焰中流体动力学不稳定性的非线性分析——I. 基本方程的推导》，《宇航学报》，第 11–12 期，第 1177–1206 页，1977 年 11 月，doi: 10.1016/0094-5765(77)90096-0。[4] I. E. Lagaris、A. Likas 和 D. I. Fotiadis，《用于求解常微分方程和偏微分方程的人工神经网络》，《IEEE 神经网络汇刊》，第 5 期，第 987–1000 页，1998 年，doi: 10.1109/72.712178。[5]S. Rout、V. Dwivedi 和 B. Srinivasan，《使用物理信息机器学习对计算流体力学问题进行数值近似》，arXiv，2021 年 11 月，doi: 10.48550/arXiv.2111.02987。硕士论文，2019 年，印度理工学院马德拉斯分校。[6] G. Cybenko，《通过 S 形函数的叠加进行逼近》，《控制、信号与系统数学》，第 4 期，第 303–314 页，1989 年 12 月，doi: 10.1007/bf02551274。[7]K. Hornik、M. Stinchcombe 和 H. White，《使用多层前馈网络对未知映射及其导数进行通用逼近》，《神经网络》，第 5 期，第 551–560 页，1990 年 1 月，doi: 10.1016/0893-6080(90)90005-6。[8] M. Raissi、P. Perdikaris 和 G. E. Karniadakis，《物理信息神经网络：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》，第 686–707 页，2019 年 2 月，doi: 10.1016/j.jcp.2018.10.045。[9] M. Raissi、P. Perdikaris 和 G. Karniadsakis，《物理信息深度学习（第一部分）：非线性偏微分方程的数据驱动解》，arXiv，2017 年 11 月，doi: 10.48550/arXiv.1711.10561。

[10] S. Patankar, 《数值传热与流体流动》（《力学与热科学计算方法》）。CRC出版社，1980 年。[11] D. J. Gauthier、E. Bollt、A. Griffith 和 W. A. S. Barbosa, 《新一代储备池计算》, 《自然·通讯》, 第 1 期，2021 年 9 月，doi: 10.1038/s41467-021-25801-2。[12] S. Wang、Y. Teng 和 P. Perdikaris, 《理解并缓解物理信息神经网络中的梯度流病态问题》, 《SIAM科学计算杂志》, 第 5 期，第 A3055–A3081 页，2021 年 1 月，doi: 10.1137/20m1318043。[13] S. Wang、S. Sankaran 和 P. Perdikaris, 《训练物理信息神经网络，只需遵循因果性》, arXiv.org，2022 年 3 月，doi: 10.48550/arXiv.2203.07404。[14] A. Chattopadhyay、P. Hassanzadeh 和 D. Subramanian, 《使用机器学习方法对多尺度 Lorenz96 混沌系统进行数据驱动预测：储备池计算、人工神经网络和长短期记忆网络》, 《地球物理学中的非线性过程》, 第 3 期，第 373–389 页，2020 年 7 月，doi: 10.5194/npg-27-373-2020。[15] E. Haber 和 L. Ruthotto, 《深度神经网络的稳定架构》, 《逆问题》, 第 1 期，第 014004 页，2017 年 12 月，doi: 10.1088/1361-6420/aa9a90。[16]B. Chang、L. Meng、E. Haber、F. Tung 和 D. Begert, 《动力系统视角下的多级残差网络》, arXiv.org, 2018 年。https://arxiv.org/abs/1710.10348。[17] N. A. K. Doan、W. Polifke 和 L. Magri, 《物理信息回声状态网络》, 《计算科学杂志》, 第 101237 页，2020 年 11 月，doi: 10.1016/j.jocs.2020.101237。