## 摘要

我们介绍了具有物理知识的神经网络-经过训练的神经网络，可以解决监督学习任务，同时尊重通用非线性偏微分方程描述的任何给定物理定律。在这项工作中，我们在解决两类主要问题的背景下介绍了我们的发展：数据驱动的解决方案和偏微分方程的数据驱动的发现。根据可用数据的性质和安排，我们设计了两种不同类型的算法，即连续时间模型和离散时间模型。第一种模型形成了一组新的数据有效的时空函数逼近器系列，而后一种类型则允许使用任意准确的隐式Runge-Kutta时间步长方案，且步数不受限制。提出的框架的有效性通过流体，量子力学，反应扩散系统以及非线性浅水波传播等一系列经典问题得到证明。

## 1、引入

随着可用数据和计算资源的爆炸性增长，机器学习和数据分析的最新进展已在包括图像识别[1]，认知科学[2]和基因组学[3]在内的不同科学学科上产生了变革性的成果。但是，在分析复杂的物理，生物或工程系统的过程中，数据获取的成本往往是高得令人望而却步，而且我们不可避免地面临根据部分信息得出结论和做出决策的挑战。在这种小数据体制下，绝大多数最新的机器学习技术（例如，深度/卷积/递归神经网络）缺乏鲁棒性，无法提供收敛保证。

乍一看，训练深度学习算法以从一些（可能是非常高维的）输入和输出数据对中准确识别非线性映射的任务看起来天真。关于我们的急救，在许多情况下，与物理和生物系统的建模有关，存在大量的先验知识，这些知识目前并未在现代机器学习实践中使用。假设是原则的物理定律支配着系统随时间变化的动态，或者是一些经过经验验证的规则或其他领域的专业知识，则该先验信息可以充当正则化代理，将允许的解决方案的空间约束到可管理的大小（例如，在不可压缩的流体动力学问题中，可以通过丢弃违反质量守恒原理的任何不现实的流动解决方案来解决）。反过来，将这种结构化信息编码为学习算法会导致算法看到的数据信息内容放大，从而即使只有几个训练示例可用，也能使其迅速转向正确的解决方案并获得良好的概括。

在最近的研究[4-6]中已经展示了利用结构化先验信息来构造数据有效型和物理知识型学习机的希望的第一眼。在那里，作者利用高斯过程回归[7]设计了针对给定线性算子的功能表示，并且能够准确地推导解并为数学物理学中的多个原型问题提供不确定性估计。 Raissi等人在随后的研究中提出了对非线性问题的扩展。 [8,9]在推理和系统识别的背景下。尽管高斯过程在编码先验信息方面具有灵活性和数学上的优雅，但是非线性问题的处理却带来了两个重要的局限性。首先，在[8,9]中，作者必须将任何非线性项在时间上进行局部线性化，从而限制了所提出方法在离散时域中的适用性，并损害了其在强非线性条件下的预测准确性。其次，高斯过程回归的贝叶斯性质要求某些先验的假设，这可能会限制模型的表示能力，并引起健壮性/脆性问题，特别是对于非线性问题[10]。

## 2、问题设置

在这项工作中，我们采用深层神经网络并利用其众所周知的功能作为通用函数逼近器[11]，采用了不同的方法。在这种情况下，我们可以直接解决非线性问题，而无需承诺任何先前的假设，线性化或局部时间步长。我们利用自动微分技术的最新发展[12]-科学计算中最有用但可能未得到充分利用的技术之一-来根据输入坐标和模型参数对神经网络进行微分，从而获得具有物理信息的神经网络。许多神经网络都被约束为尊重由支配观测数据的物理定律所产生的任何对称性，不变性或守恒原理，如一般的时间相关和非线性偏微分方程所模拟的那样。这种简单而强大的结构使我们能够解决计算科学中的各种问题，并引入了一种潜在的变革性技术，从而导致了新的数据有效和具有物理知识的学习机的发展，这些新的类是用于偏微分方程的数值解算器，例如以及用于模型反转和系统识别的新的数据驱动方法。

这项工作的总体目的是为建模和计算的新范式奠定基础，该范式通过数学物理学的长期发展来丰富深度学习。为此，我们的手稿分为两部分，目的是在两大类问题的背景下介绍我们的发展：数据驱动的解决方案和偏微分方程的数据驱动发现。随该手稿附带的所有代码和数据集都可以在GitHub上的https://github.com/maziarraissi/PINNs上找到。在整个工作中，我们一直在使用相对简单的深度前馈神经网络体系结构，该体系结构具有双曲正切激活函数并且没有其他正则化（例如L1 / L2罚分，下降等）。手稿中的每个数值示例都附带有关于我们采用的神经网络架构的详细讨论以及有关其训练过程的详细信息（例如优化子，学习率等）。最后，附录A和附录B中提供了一系列全面的系统研究，旨在证明所提出方法的性能。

在这项工作中，我们考虑一般形式的参数化和非线性偏微分方程

ut + N [u; λ] = 0, x ∈ , t ∈ [0, T ], （1）

其中，u（t，x）表示潜在（隐藏）解，N [·; λ]是由λ参数化的非线性算子，是RD的子集。该设置封装了数学物理学中的一系列问题，包括守恒定律，扩散过程，对流-扩散-反应系统和动力学方程。作为激励示例，一维伯格斯方程[13]对应于N [u; λ] =λ1uux-λ2uxx和λ=（λ1，λ2）。在此，下标表示时间或空间上的偏微分。给定系统的噪声测量结果，我们对解决两个不同的问题感兴趣。第一个问题是偏微分方程[4,8]的推论，滤波和平滑或数据驱动的解决方案，该问题指出：给定固定模型参数λ，关于未知隐藏状态u（t，x）可以说什么？系统？第二个问题是学习，系统识别或数据驱动的偏微分方程[5,9,14]的发现，其中指出：最能描述观测数据的参数λ是什么？

## 3、微分方程的驱动解

让我们首先集中讨论计算一般形式的偏微分方程的数据驱动解的问题（即上面概述的第一个问题）

ut + N [u] = 0, x ∈ , t ∈ [0, T ], (2)

其中u（t，x）表示潜在（隐藏）解，N [·]是非线性微分算子，并且是RD的子集。 在第3.1节和第3.2节中，我们提出了两种不同类型的算法，即连续时间模型和离散时间模型，并通过解决不同的基准问题突出了它们的性质和性能。 在研究的第二部分（请参见第4节）中，我们将注意力转移到偏微分方程的数据驱动发现问题上[5,9,14]。

### 3.1、连续时间模型

我们定义f（t，x）由方程（2）的左侧给出； 即

f := ut + N [u], (3)

然后通过深层神经网络逼近u（t，x）。 该假设与方程式（3）一起产生了一个物理知觉的神经网络f（t，x）。 该网络可以通过应用链式规则使用自动微分[12]来微分功能的组成而得出，并且具有与代表u（t，x）的网络相同的参数，尽管由于微分的作用而具有不同的激活函数。 运算符N。 可以通过最小化均方误差损失来学习神经网络u（t，x）和f（t，x）之间的共享参数



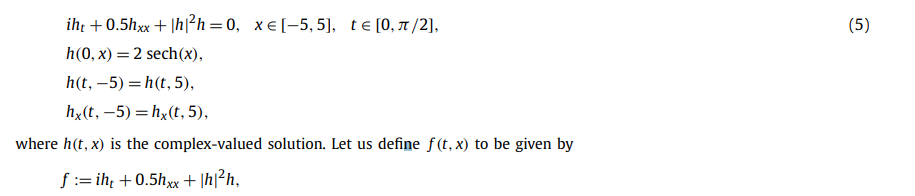
这里，{ti u，xi u，ui} Nu i = 1表示关于u（t，x）的初始和边界训练数据，{ti f，xi f} N fi = 1指定f（t， X）。损耗MSEu对应于初始数据和边界数据，而MSE f在有限的一组配置点处强制执行公式（2）施加的结构。尽管在以前的研究中已经探索了使用物理定律来约束神经网络的类似想法[15,16]，但在这里我们使用现代计算工具对其进行了重新研究，并将其应用于由时变非线性偏微分方程描述的更具挑战性的动力学问题。

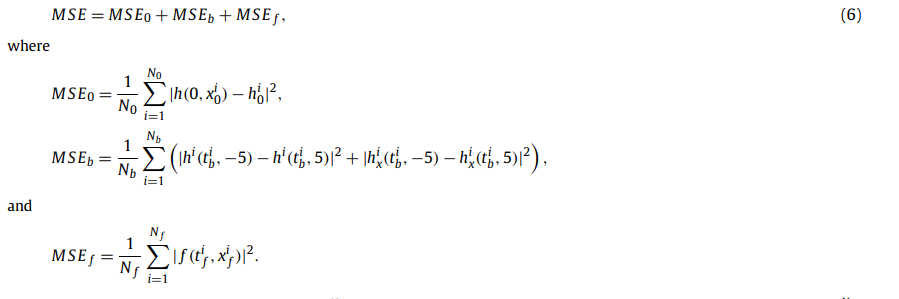
在这里，我们应该强调这条工作线与文献中详细阐述在计算物理学中使用机器学习的现有方法之间的重要区别。 Wang等人最近也使用了术语物理信息机器学习。 [17]在湍流建模的背景下。用于物理系统的预测建模的机器学习方法的其他示例包括[18-29]。所有这些方法都采用机器学习算法，例如支持向量机，随机森林，高斯过程以及前馈/卷积/递归神经网络，仅作为黑盒工具。如上所述，拟议的工作旨在通过重新研究针对基础差分算子定制的“自定义”激活和损失函数的构造来进一步向前迈进。这使我们能够通过理解和欣赏深度学习领域中自动差异化所扮演的关键角色来打开黑匣子。通常，自动微分，尤其是反向传播算法，目前是通过对模型的参数（例如权重和偏差）进行取导来训练深度模型的主要方法。在这里，我们使用深度学习社区使用的完全相同的自动微分技术，通过将其相对于其输入坐标（即空间和时间）的导数取导数，从而将物理信息通知神经网络，其中物理由偏微分方程描述。我们从经验上观察到，这种结构化方法引入了一种正则化机制，该机制使我们能够使用相对简单的前馈神经网络体系结构，并使用少量数据对其进行训练。这个简单想法的有效性可能与Lin，Tegmark和Rolnick [30]的言论有关，并提出了许多有趣的问题，需要在未来的研究中进行定量处理。为此，拟议的工作从Psichogios和Ungar [16]，Lagaris等人的早期贡献中得到启发。 [15]以及Kondor [31,32]，Hirn [33]和Mallat [34]的当代作品。

在所有与偏微分方程的数据驱动解有关的情况下，训练数据Nu的总数相对较小（几百到几千个点），我们选择使用L-BFGS优化所有损失函数，准牛顿，基于全批次梯度的优化算法[35]。对于较大的数据集，例如第4节中讨论的数据驱动模型发现示例，可以使用随机梯度下降及其现代变体[36,37]轻松采用计算效率更高的小批处理设置。尽管没有理论上的保证该程序收敛到全局最小值，但我们的经验证据表明，如果给定的偏微分方程是正定的，并且其解是唯一的，则在给定的条件下，我们的方法能够达到良好的预测精度充分表达的神经网络架构和足够数量的搭配点N f。这种普遍的观察与方程（4）的均方误差损失所导致的最终优化前景密切相关，并为与深度学习的最新理论发展同步的研究定义了一个开放性问题[38,39]。为此，我们将使用附录A和附录B中提供的一系列系统敏感性研究来测试所提出方法的稳健性。

#### 3.1.1、例子（Schrodinger equation）

此示例旨在强调我们的方法处理周期性边界条件，复值解以及支配偏微分方程中不同类型的非线性的能力。 一维非线性Schrödinger方程是一个经典的场方程，用于研究量子力学系统，包括光纤和/或波导中的非线性波传播，玻色-爱因斯坦凝聚物和等离子体波。 在光学中，非线性项源自给定材料的强度相关折射率。 同样，玻色-爱因斯坦凝聚物的非线性项是相互作用的N体系统的平均场相互作用的结果。 非线性Schrödinger方程以及周期边界条件由下式给出

然后在h（t，x）上先放置一个复数值神经网络。实际上，如果u表示h的实部，而v是虚部，则我们在h（t，x）= u（t，x）v（t，x）上放置一个多输出神经网络。 这将导致复值（多输出）物理信息神经网络f（t，x）。 神经网络的共享参数h（t，x）和f（t，x）可以通过最小化均方误差损失来学习



这里，{xi 0，hi 0} N0 i = 1表示初始数据，{ti b} Nb i = 1对应边界上的并置点，{ti f，xi f} N fi = 1表示并置f（t，x）上的点。因此，MSE0对应于初始数据上的损失，MSEb强制执行周期性边界条件，而MSE f惩罚在并置点上不满足的Schrödinger方程。

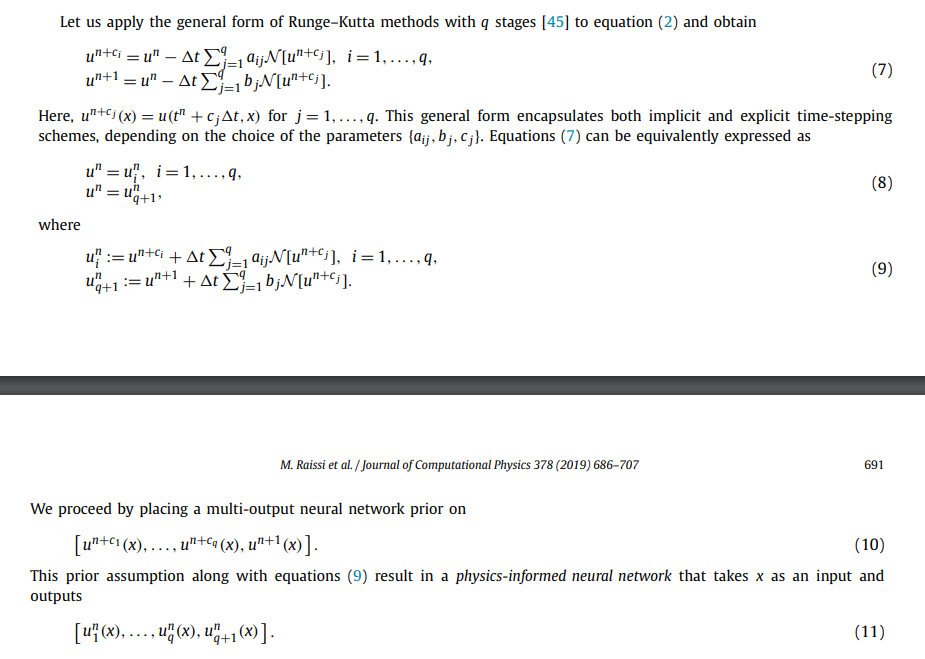
为了评估我们方法的准确性，我们使用常规的光谱方法模拟了公式（5），以创建高分辨率数据集。具体而言，从初始状态h（0，x）= 2 sech（x）开始，并假设周期性边界条件h（t，-5）= h（t，5）和hx（t，-5）= hx（t ，5），我们使用Chebfun软件包[40]结合了方程式（5），直到最后一个时间t =π/ 2为止，该软件包具有256个模式的频谱傅立叶离散化和具有时间-的四阶显式Runge- Kutta时间积分器步骤t ＝π/ 2·10-6。在我们的数据驱动设置下，我们观察到的只是在时间t = 0时对潜函数h（t，x）的测量值{xi 0，hi 0} N0 i = 1。特别是，训练集包括从完整的高分辨率数据集中随机解析的h（0，x）上的N0 = 50个数据点，以及用于执行周期边界的Nb = 50个随机采样的配置点{ti b} Nb i = 1。此外，我们假设N f = 20,000个随机采样的配置点，用于在解决方案域内实施公式（5）。所有随机采样的点位置都是使用填充拉丁的超立方体采样策略生成的[41]。

在这里，我们的目标是推断Schrödinger方程（5）的整个时空解h（t，x）。我们选择使用每层100个神经元和双曲正切激活函数的5层深度神经网络来联合表示潜函数h（t，x）= [u（t，x）v（t，x）]。通常，应该为神经网络提供足够的近似能力，以适应预期的u（t，x）复杂度。尽管可以采用诸如贝叶斯优化[42]之类的更系统的程序来微调神经网络的设计，但是在没有理论误差/收敛估计的情况下，神经体系结构/训练程序与神经网络的复杂性之间的相互作用。基本的微分方程仍然知之甚少。通过采用贝叶斯方法并监视预测后验分布的方差，可以找到一种评估预测解决方案准确性的可行途径，但这超出了当前工作的范围，将在以后的研究中进行研究。

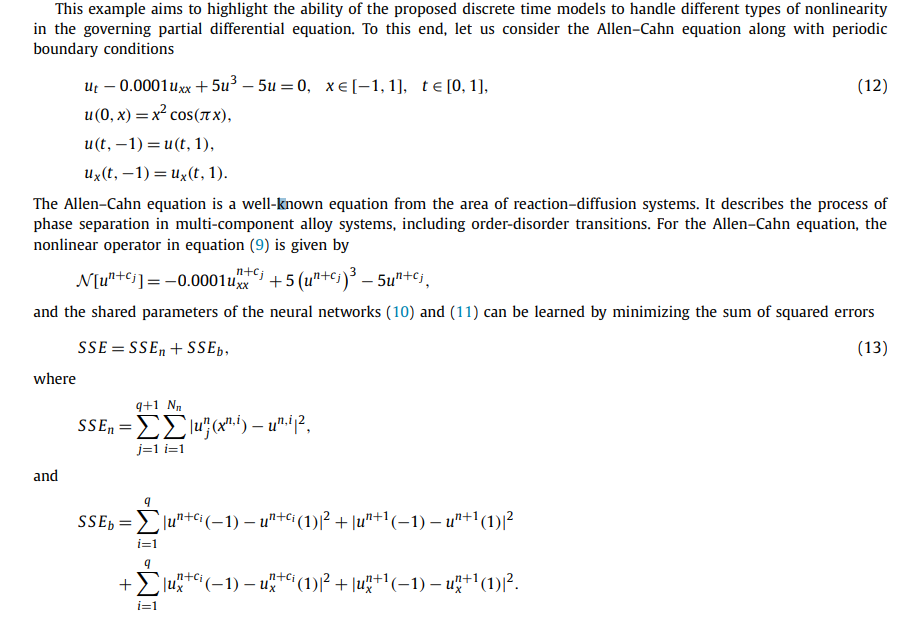
在此示例中，我们的设置旨在针对已知的过度拟合问题突出提出的方法的鲁棒性。具体而言，方程式（6）中的MSE f中的项用作惩罚不满足方程式（5）的解决方案的正则化机制。因此，物理信息神经网络的一个关键特性是可以使用小数据集对其进行有效训练。在物理系统研究中经常遇到的一种环境，对于这种环境而言，数据获取的成本可能会令人望而却步。图1总结了我们的实验结果。具体地，图1的顶部面板示出了预测的时空解| h（t，x）|的大小。 = u2（t，x）+ v2（t，x），以及初始训练数据和边界训练数据的位置。相对于该问题的测试数据验证了所得的预测误差，并在相对L2范数中测得的预测误差为1.97·10-3。在图1的底部面板中提供了对预测解的更详细的评估。特别是，我们给出了在不同时刻t = 0.59、0.79、0.98的精确解和预测解之间的比较。只需使用少量初始数据，即可通过物理知悉的神经网络准确捕获Schrödinger方程的复杂非线性行为。

到目前为止，考虑到的连续时间神经网络模型的一个潜在局限性在于，需要使用大量的搭配点N f，以便在整个时空域中实施物理信息约束。尽管这对于一两个空间维度上的问题不构成重大问题，但它可能会在更高维度的问题上引入严重的瓶颈，因为全局实施物理信息约束所需的配置点总数（即，在我们的情况下是部分约束）微分方程）将成倍增加。尽管可以使用稀疏网格或准蒙特卡洛采样方案在某种程度上解决此限制[43,44]，但在下一节中，我们将通过引入结构更复杂的神经网络提出另一种方法来避开对并置点的需求利用经典的Runge-Kutta时间步长方案[45]。

### 3.2、离散时间模型



#### 3.2.1、例子（Allen-Cahn equation）

这里，{xn，i，un，i} Nn i ＝ 1对应于时间步tn的数据。在经典数值分析中，由于显式方案的稳定性约束或隐式公式的计算复杂性约束，这些时间步长通常被限制为较小[45]。随着Rung-Kutta级数q总数的增加，这些限制变得更加严峻，并且，对于大多数具有实际意义的问题，人们需要采取数千至数百万个这样的步骤，直到解决方案解决到所需的最终时间为止。与经典方法形成鲜明对比的是，在这里我们可以采用隐含的Runge-Kutta方案，该方案具有任意数量的阶段，而实际上却很少花费额外的费用。1这使我们能够采取很大的时间步长，同时保持稳定性和高预测精度，因此允许我们只需一步就能解决整个时空解决方案。

在本例中，我们通过使用常规频谱方法模拟Allen-Cahn方程（12）生成了训练和测试数据集。具体而言，从初始条件u（0，x）= x2 cos（πx）开始，并假设周期性边界条件u（t，-1）= u（t，1）和ux（t，-1）= ux（t ，1），我们已经使用Chebfun软件包[40]结合了方程（12），直到最终时间t = 1.0，该软件包具有512个模式的频谱傅立叶离散化和带有时间步长的四阶显式Runge-Kutta时间积分器t ＝ 10-5。

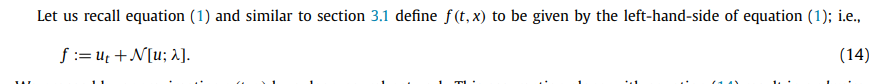
我们的训练数据集由Nn = 200个初始数据点组成，它们在时间t = 0.1时从精确解中随机进行二次采样，我们的目标是使用单个时长和大小来预测时间t = 0.9时的解t = 0.8。为此，我们采用了具有离散时间物理信息的神经网络，该网络具有4个隐藏层和每层200个神经元，而输出层则预测了101个感兴趣的量，对应于q = 100个Runge–Kutta阶段un + ci（x）， i = 1，...，q，最后时刻的解为un + 1（x）。此方案的理论误差估计可预测O（t2q）的时间误差累积[45]，在我们的情况下，这将转化为一种低于机器精度的误差方式，即t2q = 0.8200≈10-20。据我们所知，这是首次使用这种高阶隐式Runge-Kutta方案。值得注意的是，从t = 0.1处的平滑初始数据开始，我们可以在单个时间步中预测t = 0.9处几乎不连续的解，相对L2误差为6.99·10-3，如图2所示。完全归因于神经网络逼近u（t，x）的能力，以及平方误差损失之和允许对训练数据进行插值的程度。

控制我们离散时间算法性能的关键参数是Runge–Kutta阶段的总数q和时间步长t。正如我们在附录A和附录B中提供的系统研究中所证明的那样，低阶方法，例如q = 1对应于经典梯形规则，而q = 2对应于四阶高斯-勒根德式方法不能在较大的时间步长上保持其预测精度，因此必须采用多个较小的时间步长来制定解决方案。另一方面，将Runge–Kutta阶段数增加到32个甚至更多的能力使我们可以采取非常大的时间步长，并在不牺牲预测准确性的情况下一步有效地解决问题。而且，数值稳定性也不会受到影响，因为隐式的高斯-勒根德勒是唯一一个无论阶次如何都保持A稳定的时间步长方案家族，因此使它们非常适用于刚性问题[45]。对于实现如此简单的算法而言，这些特性是空前的，并且说明了我们离散时间方法的关键亮点之一。

## 4、数据驱动的微分方程的发现

在本研究的当前部分，我们将注意力转移到偏微分方程的数据驱动发现问题上[5,9,14]。 在第4.1节和第4.2节中，我们提出了两种截然不同的算法，即连续时间模型和离散时间模型，并通过各种规范问题突出了它们的性质和性能。

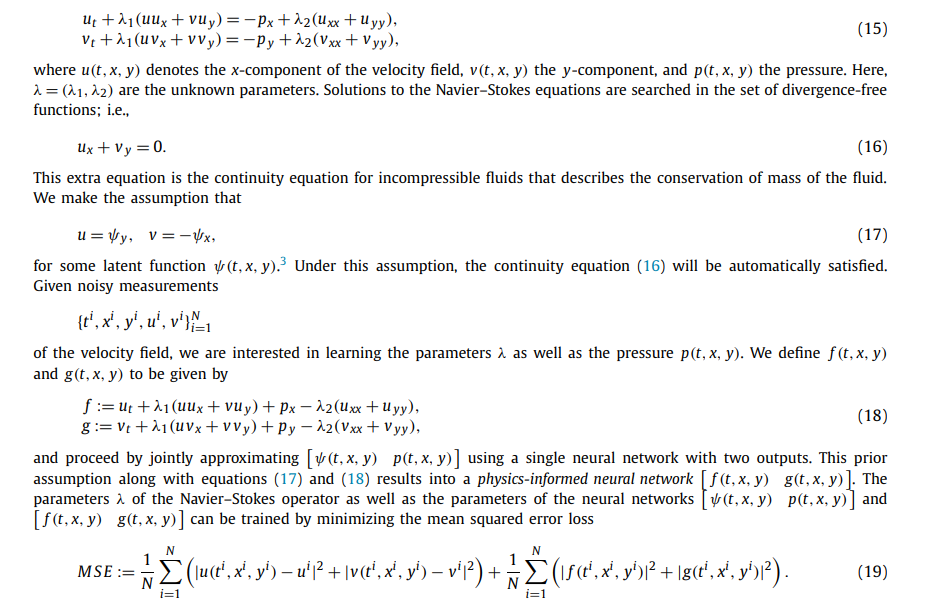
### 4.1、连续时间模型



我们通过深层神经网络逼近u（t，x）。 该假设与方程式（14）一起产生了一个物理知觉的神经网络f（t，x）。 该网络可以通过应用链式规则通过使用自动微分[12]来微分功能的组成而得出。 值得强调的是，微分算子λ的参数变成了物理信息神经网络f（t，x）的参数。

#### 4.1.1 例子（Navier-Stokes equation）

我们的下一个示例涉及到普遍存在的Navier–Stokes方程所描述的不可压缩流体流动的实际情况。 Navier–Stokes方程描述了许多科学和工程学兴趣现象的物理学。 它们可用于模拟天气，洋流，管道中的水流以及机翼周围的空气流。 完整和简化形式的Navier–Stokes方程有助于飞机和汽车的设计，血流研究，电站设计，污染物扩散分析以及许多其他应用。 让我们考虑二维明确表示的二维Navier–Stokes方程（2D）



在这里，我们考虑通过圆柱体的不可压缩流的原型问题。这是一个已知的问题，对于雷诺数Re =u∞D/ν的不同状态，它表现出丰富的动态行为和过渡。假设无量纲自由流速度u∞= 1，圆柱体直径D = 1，运动粘度ν= 0.01，则系统表现出周期性的稳态行为，其特征为在圆柱尾流中存在不对称涡流脱落模式，称为Kármán涡街[46]。

为了生成针对该问题的高分辨率数据集，我们采用了光谱/马力元素求解器NekTar [47]。具体来说，解决方案域通过由412个三角形元素组成的细分在空间中离散化，并且在每个元素内，解决方案均近似为十阶分层，半正交Jacobi多项式展开式的线性组合[47]。我们假设施加在左边界处的均匀自由流速度分布，施加在气缸下游25直径处的右边界处的零压力流出条件以及[−15，25]×的上下边界的周期性[-8、8]域。我们使用三阶刚稳定方案[47]积分方程（15），直到系统达到周期性稳态，如图3（a）所示。接下来，对应于此稳态解决方案的所得数据集的一小部分将用于模型训练，而其余数据将用于验证我们的预测。为简单起见，我们选择将采样限制在圆柱体下游的矩形区域中，如图3（a）所示。

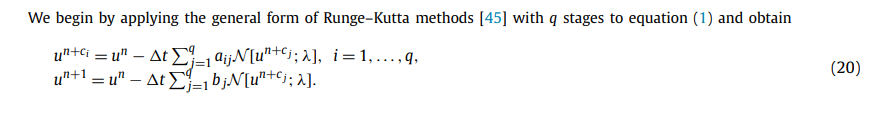
给定在流向u（t，x，y）和横向v（t，x，y）速度分量上的分散且可能有噪声的数据，我们的目标是识别未知参数λ1和λ2，并获得汽缸尾流中整个压力场p（t，x，y）的定性精确重建，根据定义只能确定为一个常数。为此，我们通过对完整的高分辨率数据集进行随机子采样来创建训练数据集。为了突出我们的方法从分散而稀缺的训练数据中学习的能力，我们选择N = 5,000，仅相当于总可用数据的1％，如图3（b）所示。还绘制了模型训练后预测速度分量u（t，x，y）和v（t，x，y）的代表性快照。这里使用的神经网络架构包括9层，每层20个神经元。

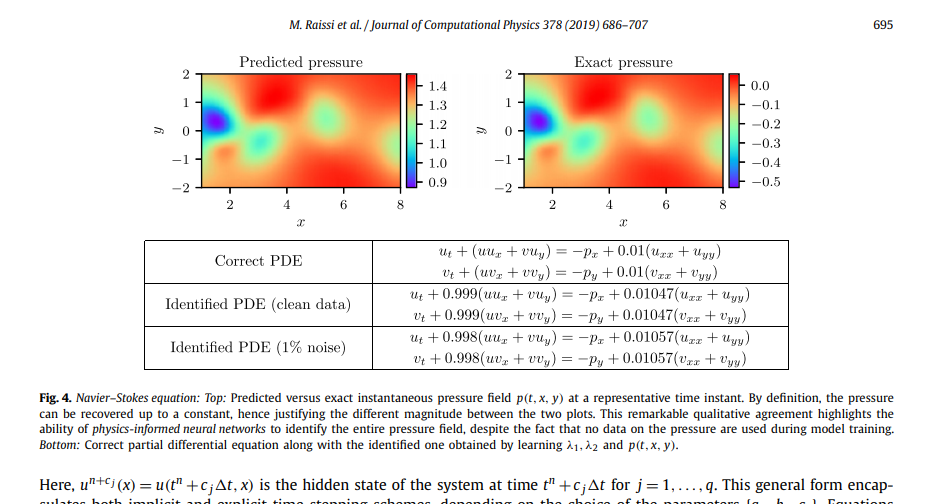
图4给出了本示例结果的摘要。我们观察到，即使训练数据被噪声破坏，具有物理信息的神经网络也能够以非常高的准确性正确识别未知参数λ1和λ2。具体地，对于无噪声训练数据的情况，估计λ1和λ2的误差分别为0.078％和4.67％。即使训练数据因1％不相关的高斯噪声而损坏，对于λ1和λ2分别返回0.17％和5.70％的误差，预测仍保持稳健。

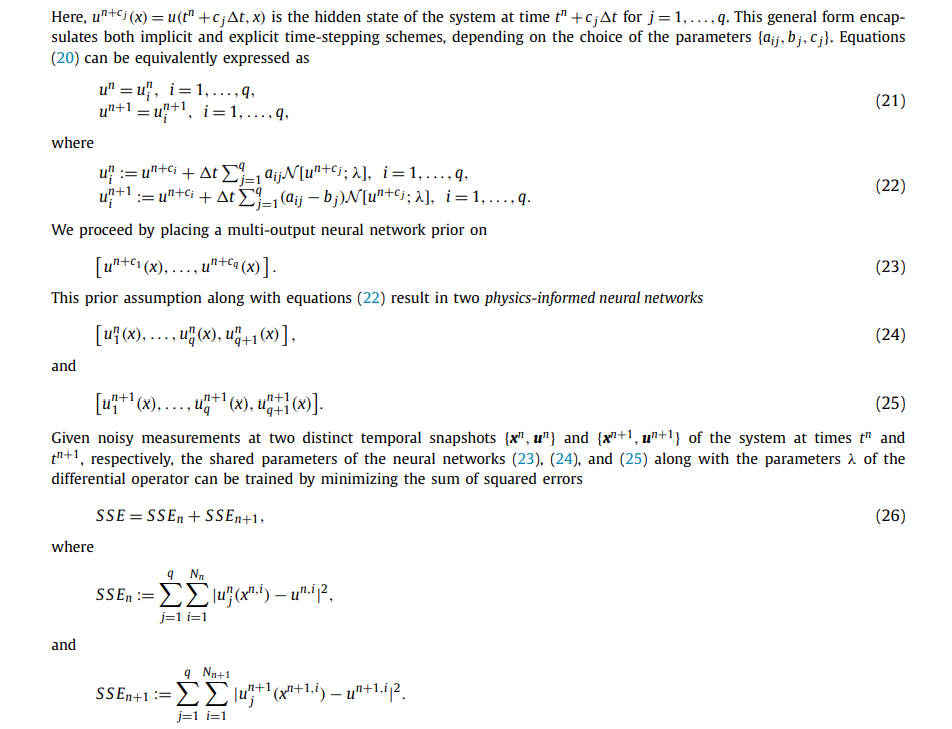
更为有趣的结果是由于在没有压力本身的任何训练数据的情况下，网络能够对整个压力场p（t，x，y）进行定性准确的预测。在图4中提供了与精确压力解决方案的直观比较，以代表压力快照。请注意，精确压力和预测压力之间的大小差异由不可压缩的Navier–Stokes系统的本质来证明是合理的，因为压力场只能确定到一个常数。通过利用基础物理原理从辅助测量中推断出连续不断的兴趣量的结果，是物理信息神经网络必须提供的增强功能的一个很好的例子，并突出了它们在解决高维逆问题中的潜力。

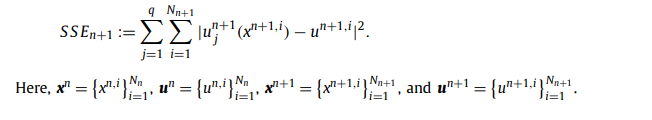
到目前为止，我们的方法假设了在整个时空域中分散数据的可用性。但是，在许多实际感兴趣的情况下，人们可能只能在不同的时刻观察系统。在下一节中，我们将介绍一种仅使用两个数据快照来解决数据驱动的发现问题的不同方法。我们将看到如何利用经典的Runge-Kutta时间步长方案，构建离散的物理物理学信息神经网络，即使数据快照之间的时间间隔非常大，该神经网络仍可以保持较高的预测精度。

### 4.2、离散时间模型



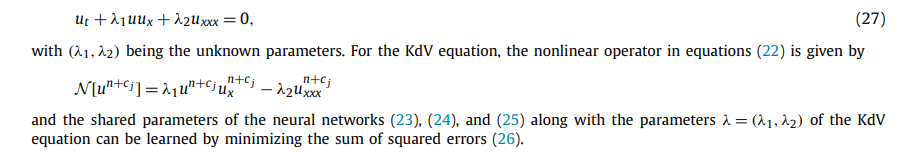






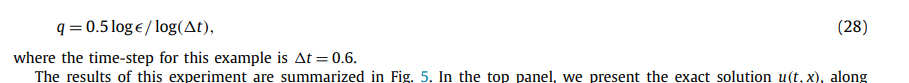
#### 4.2.1、例子（Korteweg-de Vries equation）

我们的最后一个例子旨在强调所提出框架处理涉及高阶导数的偏微分方程的能力。 在这里，我们考虑浅水表面波浪的数学模型。 Korteweg-de Vries（KdV）方程。 该方程式也可以看作是具有附加色散项的Burgers方程式。 KdV方程与物理问题有多种联系。 它描述了许多物理环境中一维长波的演化。 这样的物理环境包括具有弱非线性恢复力的浅水波，在密度分层的海洋中长的内部波，等离子体中的离子声波以及晶格上的声波。 此外，KdV方程是连续极限中费米－帕斯塔－乌拉姆问题[48]中弦的控制方程。 KdV方程为



可以通过使平方误差的总和最小化来学习公式（26）。

为了获得一组训练和测试数据，我们使用常规频谱方法进行了模拟（27）。具体来说，从初始条件u（0，x）= cos（πx）开始，并假设周期性边界条件，我们使用Chebfun软件包[40]将光谱（27）积分到最终时间t = 1.0，并使用傅立叶频谱具有512个模式的离散化和具有时间步长t = 10-6的四阶显式Runge-Kutta时间积分器。使用该数据集，然后在时间t n = 0.2和t n + 1 = 0.8时提取两个解决方案快照，并使用Nn = 199和Nn + 1 = 201对其进行随机子采样以生成训练数据集。然后，我们使用L-BFGS通过最小化方程式（26）的平方误差损失之和，来使用这些数据来训练离散时间的物理信息神经网络。这里使用的网络架构包括4个隐藏层，每层50个神经元，以及一个输出层，它在q个Runge-Kutta阶段预测解决方案，即un + cj（x），j = 1，...，q，其中，根据经验选择q，通过设置来产生机器精度的时间误差累积



该实验的结果总结在图5中。在顶部面板中，我们给出了精确的解决方案u（t，x），以及用于训练的两个数据快照的位置。 中间面板中提供了有关确切解决方案和培训数据的更详细的概述。 值得注意的是，方程（27）的复杂非线性动力学如何导致两个报告的快照之间的解决方案形式出现巨大差异。 尽管存在这些差异，并且两个训练快照之间存在较大的时间间隔，但无论训练数据是否被噪声破坏，我们的方法都能够正确识别未知参数。 具体来说，对于无噪声的训练数据，估计λ1和λ2的误差分别为0.023％和0.006％，而在噪声为1％的情况下，训练数据返回的误差为0.057％和0.017％。

## 5、总结

我们已经引入了物理信息神经网络，这是一类新的通用函数逼近器，它能够对控制给定数据集的任何基本物理定律进行编码，并且可以用偏微分方程来描述。在这项工作中，我们设计了数据驱动算法，以推导一般非线性偏微分方程的解，并构造出计算有效的物理信息替代模型。由此产生的方法展示了一系列针对计算科学中各种问题的有希望的结果，并为深层学习提供了强大的数学物理能力，为我们周围的世界建模，从而开启了道路。随着深度学习技术在方法论和算法开发方面的持续快速增长，我们相信这是及时的贡献，可以使广泛科学领域的从业者受益。可以轻松享受这些好处的特定应用程序包括但不限于物理过程的数据驱动预测，模型预测控制，多物理场/多尺度建模和仿真。

但是，我们必须注意，所提出的方法不应被视为求解偏微分方程（例如，有限元，谱方法等）的经典数值方法的替代。这样的方法在过去的50年中已经成熟，并且在许多情况下符合实践中要求的鲁棒性和计算效率标准。正如在3.2节中所提倡的那样，我们在这里传达的信息是，诸如Runge-Kutta时间步长方案之类的经典方法可以与深度神经网络和谐共存，并为构造结构化预测算法提供了宝贵的直觉。而且，后者的实现简单性极大地促进了新思想的快速开发和测试，从而有可能为数据驱动的科学计算开辟新时代的道路。

尽管提出了一系列令人鼓舞的结果，但读者也许会同意这项工作提出的问题多于答案。神经网络应该有多深/宽？真正需要多少数据？为什么算法会收敛到微分算子的参数的唯一值，即为什么算法不遭受微分算子的参数的局部最优？对于更深的体系结构和更高阶的差分运算符，网络是否会遭受梯度消失的困扰？使用不同的激活功能可以缓解这种情况吗？我们可以吗

改善初始化网络权重或标准化数据？均方误差和平方误差之和是否是适当的损失函数？为什么这些方法似乎对数据中的噪声如此健壮？我们如何量化与我们的预测相关的不确定性？在整个工作中，我们试图回答其中的一些问题，但是我们发现，对于一个方程式产生令人印象深刻的结果的特定设置可能会对另一个方程式失败。诚然，需要集体做更多的工作来奠定该领域的基础。

在更广泛的背景下，并在寻求这些问题的答案的过程中，我们相信这项工作提倡在机器学习和经典计算物理学之间取得富有成果的协同作用，这有可能丰富这两个领域并促成高影响力的发展。我们已经引入了物理信息神经网络，这是一类新的通用函数逼近器，它能够对控制给定数据集的任何基本物理定律进行编码，并且可以用偏微分方程来描述。在这项工作中，我们设计了数据驱动算法，以推导一般非线性偏微分方程的解，并构造出计算有效的物理信息替代模型。由此产生的方法展示了一系列针对计算科学中各种问题的有希望的结果，并为深层学习提供了强大的数学物理能力，为我们周围的世界建模，从而开启了道路。随着深度学习技术在方法论和算法开发方面的持续快速增长，我们相信这是及时的贡献，可以使广泛科学领域的从业者受益。可以轻松享受这些好处的特定应用程序包括但不限于物理过程的数据驱动预测，模型预测控制，多物理场/多尺度建模和仿真。

但是，我们必须注意，所提出的方法不应被视为求解偏微分方程（例如，有限元，谱方法等）的经典数值方法的替代。这样的方法在过去的50年中已经成熟，并且在许多情况下符合实践中要求的鲁棒性和计算效率标准。正如在3.2节中所提倡的那样，我们在这里传达的信息是，诸如Runge-Kutta时间步长方案之类的经典方法可以与深度神经网络和谐共存，并为构造结构化预测算法提供了宝贵的直觉。而且，后者的实现简单性极大地促进了新思想的快速开发和测试，从而有可能为数据驱动的科学计算开辟新时代的道路。

尽管提出了一系列令人鼓舞的结果，但读者也许会同意这项工作提出的问题多于答案。神经网络应该有多深/宽？真正需要多少数据？为什么算法会收敛到微分算子的参数的唯一值，即为什么算法不遭受微分算子的参数的局部最优？对于更深的体系结构和更高阶的差分运算符，网络是否会遭受梯度消失的困扰？使用不同的激活功能可以缓解这种情况吗？我们可以吗

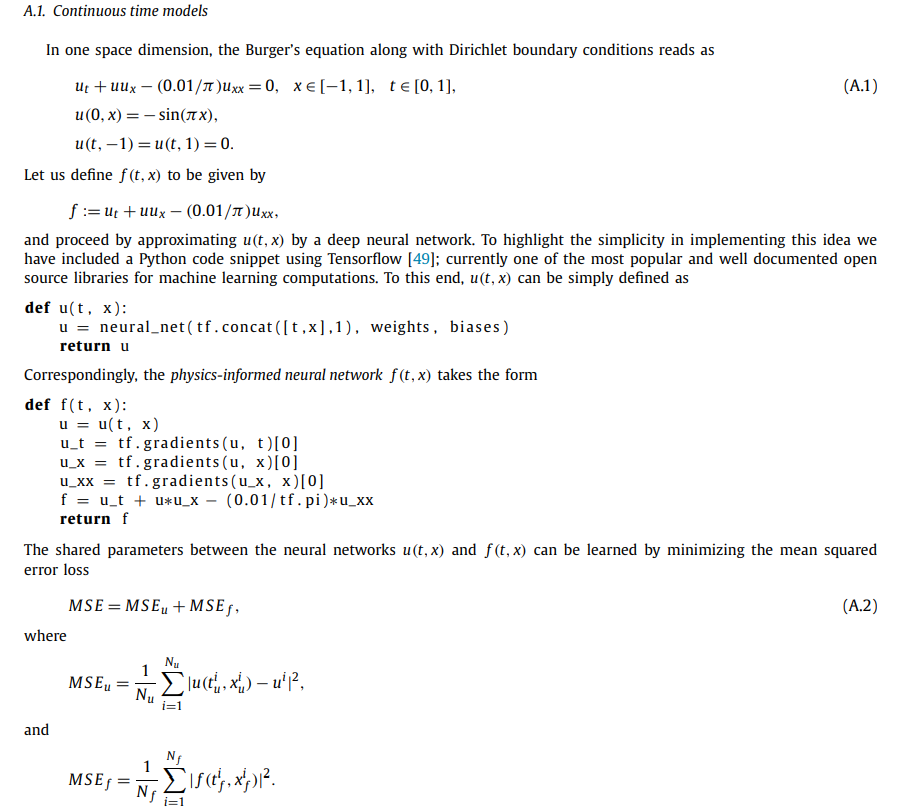
改善初始化网络权重或标准化数据？均方误差和平方误差之和是否是适当的损失函数？为什么这些方法似乎对数据中的噪声如此健壮？我们如何量化与我们的预测相关的不确定性？在整个工作中，我们试图回答其中的一些问题，但是我们发现，对于一个方程式产生令人印象深刻的结果的特定设置可能会对另一个方程式失败。诚然，需要集体做更多的工作来奠定该领域的基础。

在更广泛的背景下，并在寻求这些问题的答案的过程中，我们相信这项工作提倡在机器学习和经典计算物理学之间取得富有成果的协同作用，这有可能丰富这两个领域并促成高影响力的发展。

## 附录A :偏微分方程的数据驱动解

本附录与主要手稿一起进行，包含一系列系统研究，旨在证明所提出算法在解决与数据驱动的偏微分方程有关的问题方面的性能。 在整个文档中，我们将使用Burgers方程作为典范示例。

### A.1 连续时间模型



在这里，{ti u，xi u，ui} Nu i = 1表示关于u（t，x）的初始和边界训练数据，{ti f，xi f} N fi = 1表示f（t， X）。损耗MSEu对应于初始数据和边界数据，而MSE f在有限的并置点集合处强制执行公式（A.1）施加的结构。尽管在先前的研究中已经探索了使用物理定律来约束神经网络的类似想法[15,16]，但在这里我们使用现代计算工具对其进行了重新研究，并将其应用于由时变非线性偏微分方程描述的更具挑战性的动力学问题。

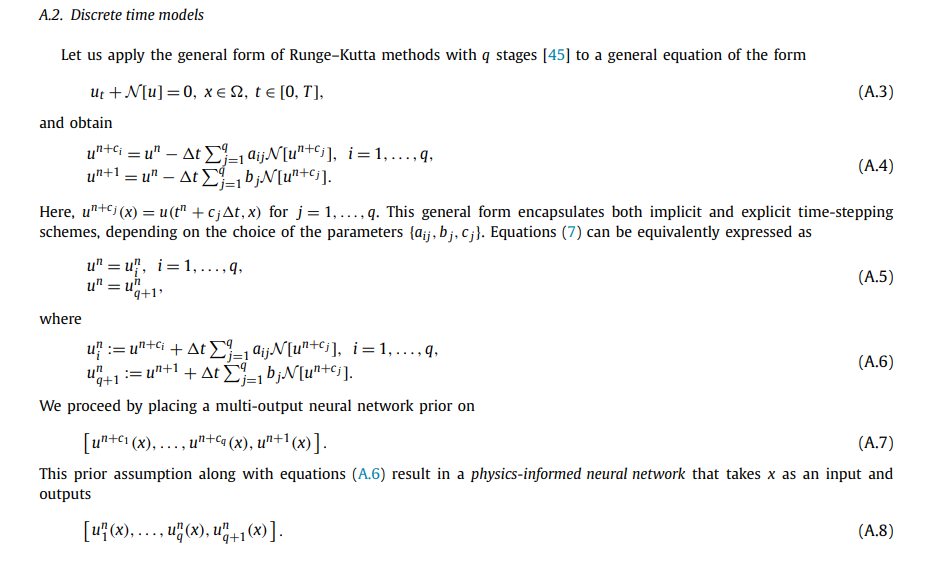
Burgers方程通常被认为是双曲守恒定律（ν→0）的原型。请注意，如果我们要“制造”该方程式的“精确”解，我们将选择一个解u（t，x）（例如e-t sin（πx）），并获得相应的右侧f（t， x）通过区分。保证得到的u（t，x）和f（t，x）满足Burgers方程，并通过构造保留所有相关的不变性。在我们的工作中，我们用神经网络u（t，x; W，b）代替u（t，x），并通过自动微分获得物理知觉的神经网络f（t，x; W，b）。因此，无论权重W和偏向b参数的选择如何，结果对u（t，x; W，b）和f（t，x; W，b）都必须满足Burgers方程。因此，在此“先验”级别上，即在我们根据给定的数据集训练网络之前，我们的模型应通过构造精确地保留连续性和动量方程。在训练过程中，给定数据集ti，xi，ui和tj，xj，fj，然后我们尝试找到“正确”参数W ∗和b ∗，以便我们尽可能地拟合两个观测数据和微分方程残差在此过程中，尽管很小，但残差将不完全为零，因此我们的近似值将在残差损失的精度范围内节省质量和动量。在经典的Galerkin有限元方法中观察到了相似的行为，而已知唯一在这种情况下具有精确守恒性质的数值方法是不连续的Galerkin和有限体积。

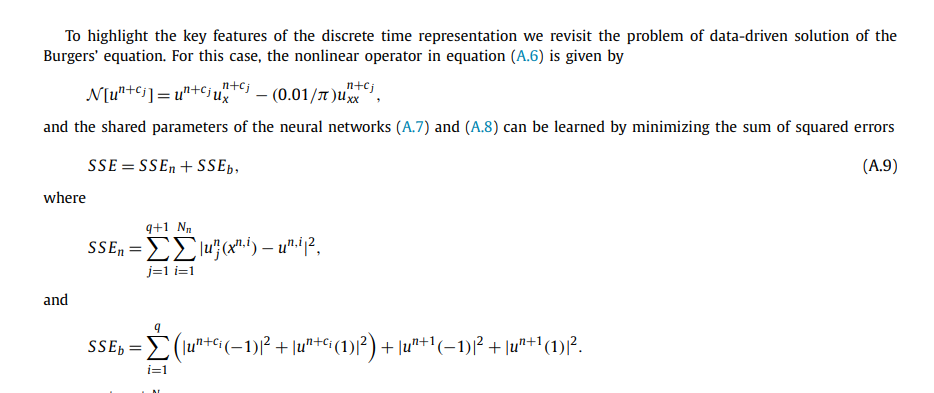
在这项工作中考虑的所有基准中，训练数据的总数Nu相对较小（几百到几千个点），我们选择使用准牛顿全批次L-BFGS优化所有损失函数基于梯度的优化算法[35]。对于较大的数据集，可以使用随机梯度下降及其现代变体[36,37]轻松采用计算效率更高的小批量设置。尽管没有理论上的保证该过程收敛于全局最小值，但我们的经验证据表明，如果给定的偏微分方程是正定的并且其解是唯一的，则在给定的条件下，我们的方法能够达到良好的预测精度充分表达的神经网络架构和足够数量的搭配点N f。这项一般性观察与方程（4）的均方误差损失所导致的最终优化前景密切相关，并为与深度学习的最新理论发展同步的研究提出了一个开放性问题[38,39]。在这里，我们将使用一系列系统的敏感性研究来检验所提出方法的鲁棒性，这些研究将结合下面的数值结果。

图A.6总结了我们对Burgers方程的数据驱动解的结果。具体来说，给定一组Nu = 100随机分布的初始和边界数据，我们通过使用（A.的均方误差损失）训练9层深度神经网络的所有3021个参数来学习潜在解u（t，x）。 2）。每个隐藏层包含20个神经元和一个双曲正切激活函数。图A.6的顶部面板显示了预测的时空解u（t，x），以及初始和边界训练数据的位置。我们必须强调，与任何经典的求解偏微分方程的数值方法不同，该预测无需任何时空域离散即可获得。该问题的确切解决方案可以通过分析获得[13]，并且在相对L2-范数中测得的预测误差为6.7·10-4。请注意，该误差比我们以前使用高斯过程对偏微分方程的数据驱动解进行的工作中报道的误差低两个数量级[8]。图A.6的底部面板提供了对预测解决方案的更详细评估。特别是，我们提出了在不同时刻t = 0.25、0.50、0.75的精确解和预测解之间的比较。仅使用少量的初始和边界数据，具有物理学知识的神经网络就可以准确地捕获Burgers方程的复杂非线性行为，从而导致在t = 0.4附近形成清晰的内部层。众所周知，后者很难用经典的数值方法精确地求解，并且需要费力的时空离散方程式（A.1）。

为了进一步分析我们方法的性能，我们进行了以下系统研究，以量化其对于不同数量的训练和搭配点以及不同神经网络体系结构的预测准确性。在表A.1中，我们报告了在保持9层网络体系结构固定的同时，不同数量的初始和边界训练数据Nu和不同数量的配置点N f所导致的相对L2误差。在给定足够数量的搭配点N f的情况下，总体趋势显示，随着训练数据Nu总数的增加，预测准确性也会提高。这一观察结果凸显了物理信息神经网络的关键优势：通过搭配点N f对基本物理定律的结构进行编码，可以获得更准确，数据效率更高的学习算法。5最后，表A.2显示了对于不同数量的隐藏层和每层不同数量的神经元，得出的相对L2分别保持训练和搭配点的总数分别固定为Nu = 100和N f = 10,000。正如预期的那样，我们观察到随着层和神经元数量的增加（因此，神经网络能够近似更复杂的功能），预测准确性也会提高。

### A.2 离散时间模型





这里，{xn，i，un，i} N n i ＝ 1对应于时刻t n的数据。现在，Runge-Kutta方案使我们能够按顺序推断潜在解u（t，x）。从时间tn的初始数据{xn，i，un，i} Nn i = 1和域边界x = -1和x = 1处的数据开始，我们可以使用上述损失函数（A.9）来训练（A.7），（A.8）的网络，并预测在时间t n + 1处的解。然后，Runge-Kutta时步方案将这一预测用作下一步的初始数据，并再次进行训练并预测u（t n + 2，x），u（tn + 3，x）等。一次踩。

图A.7给出了将该过程应用于Burgers方程的结果。为了说明起见，我们从t = 0.1时的一组Nn = 250个初始数据开始，并采用由隐含的Runge-Kutta方案诱导的具有物理学信息的神经网络，该神经网络具有500个阶段，以预测在t = 0.9时的解。一小步。该方案的理论误差估计可预测O（t2q）的时间误差累积[45]，在我们的情况下，这将转化为一种低于机器精度的误差方式，即t2q = 0.81000≈10-97。据我们所知，这是首次使用这种高阶隐式Runge-Kutta方案。值得注意的是，从t = 0.1处的平滑初始数据开始，我们可以在单个时间步中预测t = 0.9处几乎不连续的解，相对L2误差为8.2·10-4。该误差比[8]中报道的误差低两个数量级，这完全归因于神经网络逼近u（t，x）的能力，以及平方误差损失之和允许的程度训练数据的插值。这里使用的网络体系结构由4层组成，每个隐藏层中有50个神经元。

表A.3提供了详细的系统研究以量化不同网络体系结构的影响。通过将Runge–Kutta阶段的数量固定为q = 500，将时间步长固定为t = 0.8，我们改变了隐藏层的数量和每层神经元的数量，并监控了由此产生的相对L2误差。时间t = 0.9时的预测解。显然，随着神经网络容量的增加，预测准确性也会提高。

控制我们离散时间算法性能的关键参数是Runge–Kutta阶段的总数q和时间步长t。在表A.4中，我们总结了一项广泛的系统研究的结果，在该研究中，我们将网络体系结构固定为每层具有50个神经元的4个隐藏层，并更改了Runge–Kutta阶段q的数量和时间步长t。具体来说，当时间步长很大时，我们将看到阶段数少的案例无法产生准确的结果。例如，对应于经典梯形法则的情况q = 1，对应于四阶高斯-勒根德方法的情况q = 2，对于大于0.2的时间步长无法保持其预测准确性，因此必须采取解决方案具有多个小尺寸的时间步长。另一方面，将Runge–Kutta阶段数增加到32个甚至更多的能力使我们可以采取非常大的时间步长，并在不牺牲预测准确性的情况下一步有效地解决问题。而且，数值稳定性也不会受到影响，因为隐式的高斯-勒根德勒是唯一一个无论阶次如何都保持A稳定的时间步长方案家族，因此使它们非常适用于刚性问题[45]。对于实现如此简单的算法而言，这些特性是空前的，并且说明了我们离散时间方法的关键亮点之一。

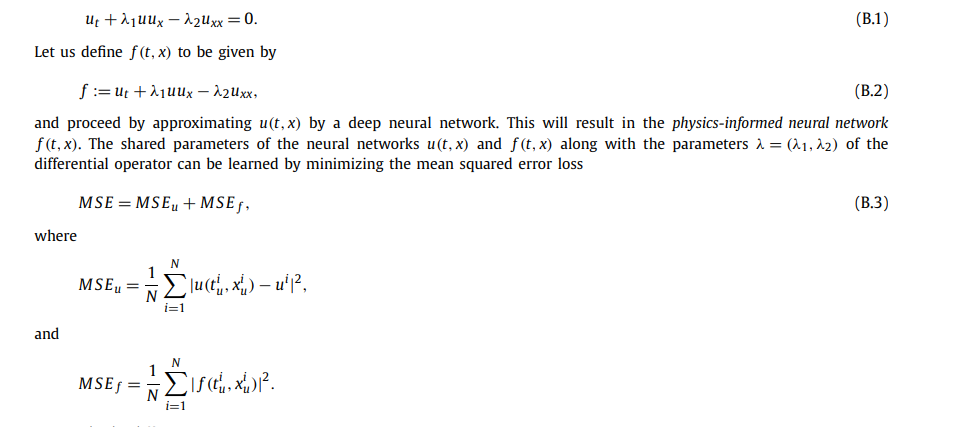
最后，在表A.5中，我们提供了系统的研究，可以在改变输入数据的空间分辨率时量化预测解决方案的准确性。不出所料，增加训练数据的总数可以提高预测精度。

## 附录B :偏微分方程的数据驱动发现

本附录与主要手稿一起进行，包含一系列系统研究，旨在证明所提出算法的性能与数据驱动的偏微分方程的发现有关。 在整个文档中，我们将使用Burgers方程作为典范示例

### B.1 连续时间模型

作为第一个示例，让我们考虑伯格斯方程。 这个方程出现在应用数学的各个领域，包括流体力学，非线性声学，气体动力学和交通流量[13]。 它是一个基本的偏微分方程，可以通过除去压力梯度项从速度场的Navier–Stokes方程导出。 对于较小的粘度参数，Burgers方程会导致形成冲击，众所周知，这很难通过经典的数值方法解决。 在一个空间维度中，等式为

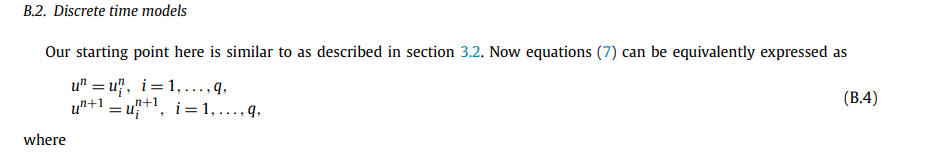


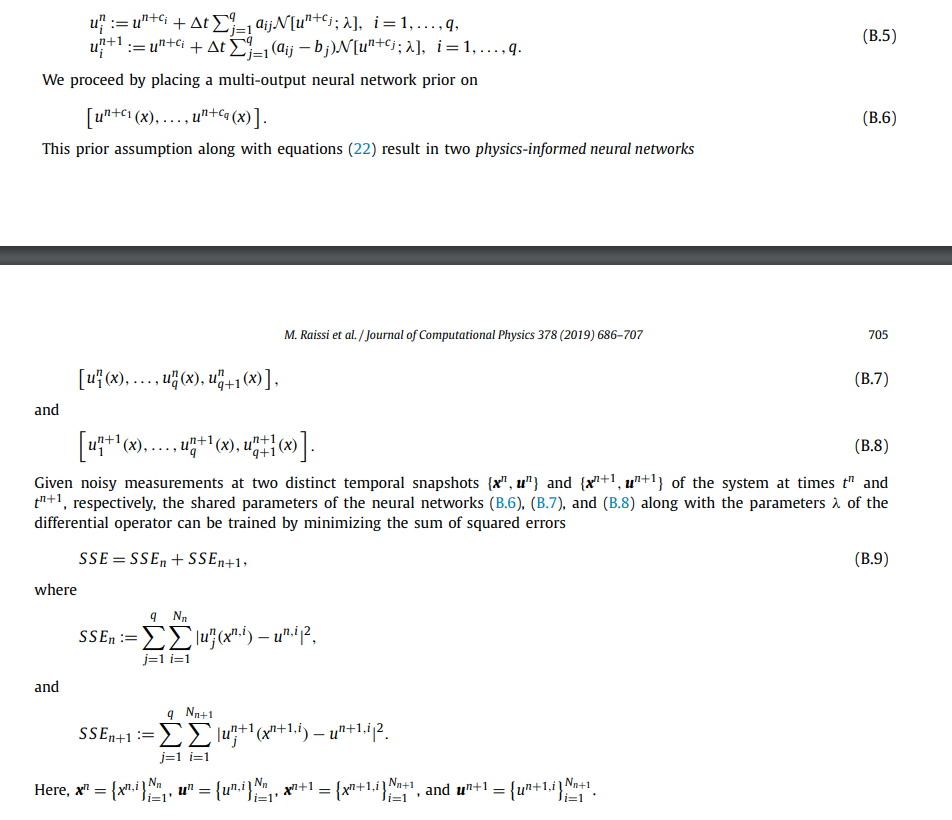
这里，{ti u，xi u，ui} Ni ＝ 1表示关于u（t，x）的训练数据。损耗MSEu对应于u（t，x）上的训练数据，而MSE f在有限的并置点集合处强制执行由公式（B.1）施加的结构，该配置点的数量和位置与训练相同数据。

为了说明我们方法的有效性，我们创建了一个训练数据集，方法是根据与λ1= 1.0和λ2= 0.01 /π对应的精确解在整个时空域中随机生成N = 2,000个点。训练点的位置显示在图B.8的顶部面板中。然后，使用LBFGS优化器通过最小化（B.3）的均方误差损失，将此数据集用于训练每个隐藏层具有20个神经元的9层深度神经网络。[35]训练后，对网络进行校准以预测整个解决方案u（t，x），以及定义基本动态的未知参数λ=（λ1，λ2）。在图B.8的中下部显示了对物理信息神经网络的预测准确性的视觉评估。即使在分散的训练数据被1％不相关的噪声破坏的情况下，该网络也能够以极高的精度识别出基础的偏微分方程。

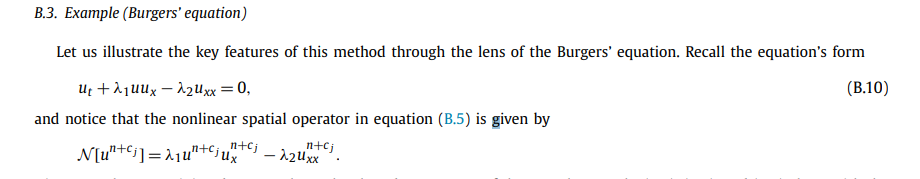
为了进一步检查我们算法的性能，我们对训练数据的总数，噪声破坏水平和神经网络体系结构进行了系统的研究。结果总结在表B.6和B.7中。此处的主要观察结果是，相对于数据中的噪声水平，所提出的方法似乎非常健壮，即使对于高达10％的噪声破坏，也能提供合理的识别精度。这种增强的鲁棒性似乎大大优于先前在[9]中报道的使用高斯过程回归的竞争方法以及依赖稀疏回归的方法，这些方法需要相对干净的数据才能准确计算数值梯度[50]。我们还观察到表B.6和B.7中的某些可变性和非单调趋势，因为网络体系结构和训练点总数发生了变化。这种可变性可能归因于与方程本身以及特定神经网络设置有关的不同因素，并引起一系列需要进一步研究的问题，如本文的结论部分所述。

### B.2 离散时间模型





### B.3 例子（Burgers’s equation）



仅给出两个训练数据快照，神经网络（B.6），（B.7）和（B.8）的共享参数以及Burgers方程的参数λ=（λ1，λ2）可以是 通过最小化公式（B.9）中的平方误差总和来学习。 在这里，我们通过分别在时刻t n = 0.1和tn + 1 = 0.9时随机采样精确解，创建了一个包含Nn = 199和Nn + 1 = 201个空间点的训练数据集。 训练数据显示在图B.9的顶部和中间面板中。 这里使用的神经网络架构由4个隐藏层组成，每个都有50个神经元，而凭经验选择Runge–Kutta阶段的数量以通过设置来产生机器精度的时间误差累积



该示例的时间步长为t = 0.8。图B.9的底部面板针对无噪声数据以及具有1％高斯不相关噪声破坏的噪声数据，总结了已识别的参数λ=（λ1，λ2）。对于这两种情况，尽管用于训练的两个数据快照相距很远，并且可能描述不同的机制，但所提出的算法仍能够以显着的准确性学习正确的参数值λ1= 1.0和λ2= 0.01 /π。基本动力。

进行了敏感性分析，以量化关于训练快照t，训练数据中的噪声水平以及基于物理学的神经网络体系结构之间的差距的预测准确性。如表B.8所示，所提出的算法对于t和噪声破坏水平都非常鲁棒，并且针对未知参数返回合理的估计值。这种鲁棒性主要归因于底层隐式Runge-Kutta方案的灵活性，它允许任意数量的级数，从而允许数据快照在时间上相距很远，同时又不影响方程式非线性动力学的准确性（ B.10）解决。这是我们用于识别问题的离散时间公式的关键亮点，将其与竞争方法区分开[9,50]。最后，表B.9给出了所识别参数的百分比误差，证明了我们相对于底层神经网络架构的估计的鲁棒性。尽管总体上取得了积极的结果，但表B.8和表B.9中观察到的可变性仍然无法解释，并且自然地引起了本文结论部分中提出的一系列问题。

## 附录C:补充材料

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jcp.2018.10.045.