常微分方程与偏微分方程的自适应代理模型分层推断与闭合学习

Pengyu Zhang [*‡]、Arnaud Vadeboncoeur[*‡]、Alex Glyn-Davies[‡]和 Mark Girolami[‡, §]

[‡]剑桥大学工程系，*JJ Thomson* 大道 *7a*，剑桥，*CB3 0FA*，英国[§]艾伦·图灵研究所，尤斯顿路 *96* 号，伦敦，*NW1 2DB*，英国

arXiv:2603.03922v1 [cs.LG] 4 Mar 2026

摘要

逆问题旨在校准模型，使其与数据相符。逆问题在各种工程应用中发挥着关键作用，使从业者能够让模型与现实相一致。在许多应用中，工程师和科学家无法全面了解：i）系统的详细属性（例如材料属性、几何形状、初始条件等）；ii）描述所有相关动力学的完整定律（例如摩擦定律、复杂的阻尼现象和一般的非线性交互作用）。本文提出一种严谨的方法，利用来自多个彼此不同但相关的物理系统的数据，共同估计各系统各自的模型参数，并以基于机器学习的闭合模型形式学习共享的未知动力学。为稳健地推断每个系统的未知参数，我们采用分层贝叶斯框架，从而能够联合推断多个系统及其总体层面的统计特征。为学习闭合项，我们在问题的常微分方程/偏微分方程表述中嵌入神经网络，并采用边际似然最大估计。为实现这一框架，我们使用集成 Metropolis 调整朗之万算法（MALA），以进行稳定而高效的采样。为缓解求解逆问题时反复进行正向计算所造成的计算瓶颈，我们提出一种双层优化策略，在推断的同时训练代理前向模型。在此框架下，我们评估并比较了不同的代理架构，具体包括傅里叶神经算子（FNO）和参数化物理信息神经网络（PINNs）。

Keyw术语： 贝叶斯推断、闭合学习、代理建模、分层模型、偏微分方程/常微分方程 s

1 引言

由常微分方程（ODEs）和偏微分方程（PDEs）支配的物理系统的建模与理解，是现代科学与工程分析的基石。这类方程描述了物理过程在时间和空间上的演化，涉及流体与固体力学 [1, 2]，传热 [3]，电磁学 [4]，化学动力学[5]，以及大气和海洋动力学 [6]等多个领域。虽然这些方程通常由第一性原理推导而来，但在许多复杂的现实情境中，它们并不完备：控制方程可能包含未知参数，或者缺少未知的非线性闭合项。作为一个逆问题，从观测数据中准确确定未知参数和闭合项，对于模型校准以及实现物理动力学的准确预测和应用至关重要。尽管逆问题具有病态性，观测数据日益丰富，加之现代机器学习技术的发展，提升了人们识别复杂模式的能力，本文的发展

*通讯作者 {pz281, av537}@cam.ac.uk。

提高了直接从数据中发现未知参数或物理项、求解逆问题的能力。

在许多实际系统中，控制方程大多已知，只有部分动力学特性尚未得到充分理解。这与系统辨识或方程发现框架（例如非线性动力学稀疏辨识（SINDy）方法）形成对比；后者旨在从数据中恢复完整的控制方程 [7, 8, 9, 10], ，而我们的方法关注的是物理模型已有部分已知的情况。此类未建模组件的常见示例包括轴承系统中的摩擦力、机械装置中复杂的阻尼定律、流体动力学中的湍流模型，或由对流导致的非线性热耗散 [11, 12, 13]。在类似背景下，Rogers 等人 [14, 15]通过将缺失项建模为潜在恢复力，并使用高斯过程（GP） [16], ，研究机械系统中的未知非线性动力学；同样的方法也用于识别传热中的非线性对流效应 [17]。秉持类似的理念，我们的方法并非从头发现完整方程，而是利用已知物理规律，仅推断缺失的组件。

为提高推断质量，我们采用一种基于总体的建模形式，从属于同一物理系统族的多个系统中收集观测数据。这些系统以 $k = 1, ..., K$ 为索引，各自具有不同的参数 $\boldsymbol{\theta}^{(k)}$，但共享一个非线性闭合函数 $f(\cdot)$。底层物理模型可通过参数到解算子 $F^\dagger : \Theta \to U$ 表示，该算子将参数对 $(\boldsymbol{\theta}^{(k)}, f)$ 映射为相应的 k^{th} 系统状态，

$$F^\dagger(\boldsymbol{\theta}^{(k)}, f) = \mathbf{u}^{(k)}. \tag{1}$$

这种一般性建模形式涵盖了我们提出的框架能够处理的广泛物理系统，包括常微分方程和偏微分方程。对于每个系统 k，通过以下方式收集来自系统 k 的稀疏且含噪观测$\mathbf{y}^{(k)}$

$$\mathbf{y}^{(k)} = G^{(k)}(\boldsymbol{\theta}^{(k)}, f) + \boldsymbol{\xi}_\eta^{(k)}, \quad G^{(k)} = g^{(k)} \circ F^\dagger, \tag{2}$$

以及 $g : U \to \mathbb{R}^{d_y}$ 是一个解到数据映射，通常是可由观测矩阵 $\mathbf{H}^{(k)}$ 表示的线性映射。假设噪声项在不同系统之间相互独立，且 $\boldsymbol{\xi}_\eta^{(k)} \sim \mathcal{N}(\mathbf{0}, \sigma^{(k)}\mathbf{I})$ 独立同分布。给定观测数据 $\mathbf{y}^{(1:K)}$，所提出方法的目标是同时推断未知的系统特定参数 $\boldsymbol{\theta}^{(1:K)}$，并学习共享的闭合模型 f。

人们已开发出多种确定性方法，用于微分方程中的参数推断 [18, 19, 20]。尽管这些方法在恢复最优参数方面卓有成效，但通常侧重点估计，缺乏用于不确定性量化（UQ）的内在机制。然而，由于逆问题本身具有病态性，不确定性量化至关重要：我们有必要刻画可能参数的分布，而非依赖单一估计值。为实现不确定性量化，人们已广泛采用贝叶斯框架来推断物理系统中的未知参数 [21, 22, 23, 24, 25]。在这一框架中，逆问题被表述为概率问题：给定带有不确定性的观测数据、一个正向模型和一个先验分布，目标是恢复未知参数的完整后验分布。基于这一思路，我们采用分层贝叶斯模型 [26, 27]，联合推断多个相关系统的参数。这种分层表述既能捕捉系统特定变异性，也能刻画共享的总体层级结构，从而提高推断的稳定性，并促进系统间的知识迁移。

尽管由于系统参数 $\boldsymbol{\theta}^{(1:K)}$ 维度较低，采用贝叶斯方法推断它们是可行的，但未建模的非线性闭合项难以用少量参数表示，因此采用贝叶斯方法在计算上难以承受。为表示这种非线性，我们引入神经网络近似，并训练一个参数为

α的多层感知机（MLP），使其满足 $f^{\alpha} \approx f$，从而为非线性闭合项提供灵活的神经表示。

总之，我们提出的方法采用层次贝叶斯表述，并结合集成 Metropolis 调整朗之万算法（MALA）[28] 对系统参数进行概率采样$\boldsymbol{\theta}^{(1:K)}$，同时使用确定性神经网络学习未建模的非线性闭合项 f。我们在迭代方案中交替进行概率采样和闭合模型更新。这种混合方法兼顾了统计可解释性与计算效率。物理参数 $\boldsymbol{\theta}^{(1:K)}$ 维度较低，因此非常适合通过不确定性量化进行概率推断。相比之下，非线性闭合项具有高维且复杂的特点，采用确定性训练能够高效学习，避免在函数空间中采样所需的高昂成本。

此外，求解逆问题往往计算效率低下，因为需要反复评估正向模型，而这通常依赖计算成本高昂的数值求解器。对于非线性时变系统，这一过程尤其难以承受；特别是采用 MALA 等基于梯度的采样器时，还需要对数值求解器进行求导。为缓解这一瓶颈，近年来代理建模取得的进展已被广泛用于加速物理系统的仿真。具有代表性的方法包括高斯过程（GP）[16, 29, 30], ，以及基于深度学习的架构，例如深度神经网络（DNN） [31, 32],、物理信息神经网络（PINN） [18, 33], 、傅里叶神经算子（FNO） [34, 33], 、DeepONet [35, 36], 和 Transformer [37, 38]。神经代理模型能够显著加速前向仿真。因此，我们进一步提出了一种代理加速推断框架，通过针对逆问题联合训练可微神经代理模型，以近似正向算子。该代理模型能够快速、可微地近似正向求解器，从而降低计算成本。我们通过双层优化框架 [39], 同时处理正向代理模型训练和逆推断：低层训练正向代理模型，上层则使用代理模型替代计算成本高昂的数值求解器来处理逆问题。

1.1 相关工作

本节总结了代理模型加速逆问题方面的先前工作。代理模型已通过两种主要方式被纳入逆问题：确定性代理模型反演和概率性代理模型反演。物理信息神经网络和神经算子已被嵌入确定性逆问题表述中，以替代昂贵的偏微分方程求解器 [18, 40, 41, 42]。与我们提出的方法类似，Zhang等人[43] 提出了一种双层优化框架 BiLO，该框架联合学习正向算子并求解相关逆问题，但其表述仍完全是确定性的。在逆问题的贝叶斯公式中，物理信息神经网络和神经算子被用作可微的正向代理模型，以促进蒙特卡洛采样。例如，贝叶斯物理信息神经网络（贝叶斯物理信息神经网络）将网络权重和物理参数均视为随机变量，从而能够通过哈密顿蒙特卡洛 [44] 或随机梯度马尔可夫链蒙特卡洛（马尔可夫链蒙特卡洛）方法[45, 46, 47]进行后验推断。Zhang 等人 [48] 还利用马尔可夫链蒙特卡洛将其 BiLO 框架扩展用于逆问题中的不确定性量化。变分推断（变分推断）同样已被整合到这些架构中，Yang等人 [45] 也研究了物理信息神经网络的变分推断，而 Raj 等人 [49] 则在 DeepONet 框架中探索变分推断。更广泛地，Glyn-Davies 等人 [50] 全面概述了变分方法如何应用于正问题与反问题。与我们的方法密切相关，Akyildiz 等人和 Vadeboncoeur等人 [51, 52] 提出了一种双层优化框架，用于先验分布的联合推断以及代理模型的训练。

不同于现有方法通常聚焦于单系统逆问题，并假设控制常微分方程或偏微分方程的形式完全已知，我们提出的方法利用大量观测值，在多个相关系统之间进行层次贝叶斯推断，同时还学习

方程中未知的非线性闭合模型。

1.2 主要贡献

本工作的主要贡献总结如下：

C1. 联合概率参数估计与确定性闭合学习。我们提出了一种混合框架，可同时推断部分已知常微分方程和偏微分方程中的未知物理参数并学习非线性闭合项。物理参数通过层次贝叶斯表述进行概率推断，而闭合模型则使用多层感知机以确定性方式学习。

C2. 分层参数推断与闭合学习的迭代方案 **g.**
我们实施了一种训练策略，在后验采样与闭合模型更新之间交替进行。借助集成MALA，我们实现了对多个物理系统的稳健采样。这些样本还可有效近似优化闭合参数所需的梯度。

C3. 通过双层优化实现代理加速贝叶斯反演。
我们采用一种双层优化方案，其中可微神经前向代理模型与逆问题联合训练。这一由代理模型加速的框架通过用高效的代理模型替代昂贵的数值求解器，大幅降低了贝叶斯采样的计算成本。

C4. 代表性常微分方程与偏微分方程上的验证。
我们在三个不同的物理问题上验证所提出的框架：具有未知阻尼闭合项的非线性质量-弹簧-阻尼器系统（常微分方程）、具有未知扩散率闭合项的非线性达西流动系统（偏微分方程），以及广义伯格斯方程（偏微分方程）。

本文其余部分结构如下。第2 节概述了支撑我们方法的背景，重点介绍层次贝叶斯和集成MALA。第3 节详细介绍所提出的方法论。第4 节展示所提出方法在三个代表性常微分方程和偏微分方程系统上的性能。最后，第5节总结全文，并讨论潜在的未来研究方向。

2 背景

本节介绍支撑我们框架的两项理论基础：层次贝叶斯与集成MALA。具体而言，层次贝叶斯表述提供了联合建模所有系统的统计结构。集成MALA被用作层次模型上的后验抽样方法。

2.1 层次贝叶斯

许多应用涉及多个系统，这些系统在问题结构上具有内在联系。因此，与其将这些系统视为严格独立，不如采用层次联合概率模型。这样，各系统便能通过共同总体分布共享信息，从而有效捕捉它们潜在的依赖关系 [26]。一个经典例子是，同一家医院患者的恢复概率在统计上彼此相关，而非相互独立，因为患者所处的环境条件相似。对每位患者进行建模

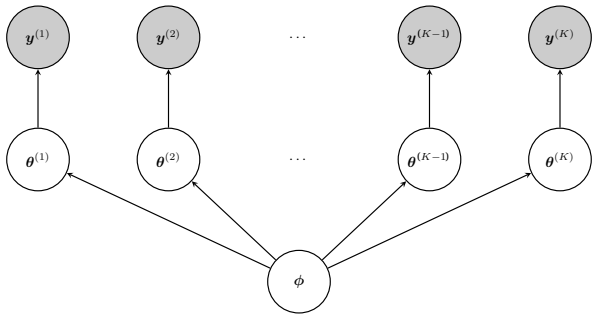


图1：分层贝叶斯模型展示了全局超参数 $\boldsymbol{\phi}$ 如何支配一组任务特定参数 $\boldsymbol{\theta}^{(1:K)}$，每个参数都会生成相应的数据集 $\boldsymbol{y}^{(k)}$。

因此，若将其孤立处理，可能会得出不稳定的估计，尤其是在个体数据稀少时。为解决这一问题，层次先验引入总体层超参数，用以描述整个群体的行为，并允许在个体之间共享信息。

我们的研究中所考虑的物理系统也适用类似的原理。我们观测到同一物理系统家族的多个实例，例如质量各不相同的多个质量-阻尼系统；它们都遵循相同的物理定律，但具有各自不同的未知参数 $\boldsymbol{\theta}^{(1:K)}$。尽管不同系统的参数有所差异，但它们并非完全独立——这些差异反映了同一总体内部的变化，例如由制造或环境条件造成的变化。为表示这种结构化变异，每个参数 $\boldsymbol{\theta}^{(k)}$ 都由同一总体生成，

$$\boldsymbol{\theta}^{(k)}|\boldsymbol{\phi} \sim p(\boldsymbol{\theta}^{(k)}|\boldsymbol{\phi}), \tag{3}$$

超参数 $\boldsymbol{\phi}$ 支配总体分布，该分布称为超先验，

$$\boldsymbol{\phi} \sim p(\boldsymbol{\phi}). \tag{4}$$

随后，通过每个 $\boldsymbol{\theta}^{(k)}$ 获得观测值 $\boldsymbol{y}^{(k)}$

$$\boldsymbol{y}^{(k)}|\boldsymbol{\theta}^{(k)}, \boldsymbol{\phi} \sim p(\boldsymbol{y}^{(k)}|\boldsymbol{\theta}^{(k)}, \boldsymbol{\phi}). \tag{5}$$

分层贝叶斯模型的结构如图 1 所示。

下面详细说明分层贝叶斯框架中一些关键分布的因子分解。联合先验分布可分解为超先验与各个条件先验的乘积，

$$p(\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi}) = p(\boldsymbol{\phi}) \prod_{k=1}^{K} p(\boldsymbol{\theta}^{(k)}|\boldsymbol{\phi}). \tag{6}$$

因此，在给定观测值$\mathbf{y}^{(1:K)}$的条件下，参数和超参数的联合后验分布表示为

$$p(\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi}|\boldsymbol{y}^{(1:K)}) = \frac{p(\boldsymbol{\phi}) \prod_{k=1}^{K} p(\mathbf{y}^{(k)}|\boldsymbol{\theta}^{(k)}) p(\boldsymbol{\theta}^{(k)}|\boldsymbol{\phi})}{p(\boldsymbol{y}^{(1:K)})}, \tag{7}$$

这就是我们希望进行采样的分布。接下来，我们将讨论我们的方法所采用的采样方案。

2.2 集成MALA

Metropolis 调整朗之万算法（MALA）在标准 Metropolis-Hastings 采样的基础上，引入了基于朗之万扩散离散化、由梯度提供信息的提议。给定目标密度p，相应的朗之万随机微分方程为

$$\mathrm{d}X_t = \nabla \log p(X_t)\mathrm{d}t + \sqrt{2}\mathrm{d}W_t, \tag{8}$$

其中， W_t 是标准的 d维布朗运动，则随机过程 X_t 具有密度为 p的不变分布。经过 Euler-Maruyama 离散化后，我们得到

$$X_{\mathrm{t}+1} = X_{\mathrm{t}} + \gamma \nabla_{X_{\mathrm{t}}} \log p(X_{\mathrm{t}}) + \sqrt{2\gamma}\epsilon_{\mathrm{t}}, \quad \epsilon_{\mathrm{t}} \sim \mathcal{N}(0, I), \tag{9}$$

其中 γ 是步长，并且 MALA 还包含一个 Metropolis 接受-拒绝步骤。按照X （9）演化状态后，得到的样本渐近服从目标平稳分布。然而，单条朗之万链在高维或多峰后验分布中可能收敛缓慢；集成 MCMC 策略通过演化多条相互作用的链并整合基于集成的协方差信息来加速收敛。

集成方法最初是在集合卡尔曼滤波（EnKF）文献中发展起来的 [53, 54]，其中使用状态集成来近似系统状态底层分布的协方差。随后，这些思想通过集合卡尔曼反演（EKI）框架 [55]扩展到了逆问题。集合卡尔曼采样器（EKS） [56] 是 EKI 的一种带噪扩展，它融合了朗之万扩散的思想，同时避免了计算正向模型梯度的需要。其核心机制依赖于由多条朗之万链组成的相互作用系统，其中每条链按照预条件朗之万动力学演化，而经验集成协方差则充当预条件器。近期，类似的相互作用粒子框架被用于相互作用粒子朗之万算法（IPLA） [57] ，以解决最大边际似然估计问题；该方法最近已应用于偏微分方程反演场景 [58]。

基于这种相互作用链的视角，我们的集成MALA采用一组 M并行预条件朗之万链。每条链都会演化状态向量 X_{t}^m，其链索引为$m = 1$、.....、 M。预条件矩阵 $\mathbf{C}_{\mathrm{t}}$ 在迭代 t 时根据各链的经验协方差自适应计算，从而使采样动力学与后验分布的局部几何相匹配。这种预条件处理可显著提高采样稳定性和效率。每条预条件朗之万链

$$X_{\mathrm{t}+1}^m = X_{\mathrm{t}}^m + \gamma \mathbf{C}_{\mathrm{t}} \nabla_{X_{\mathrm{t}}^m} \log p(X_{\mathrm{t}}^m) + \sqrt{2\gamma \mathbf{C}_{\mathrm{t}}}\epsilon_{\mathrm{t}}^m, \tag{10}$$

而预条件矩阵 $\mathbf{C}_{\mathrm{t}}$ 定义为

$$\mathbf{C}_{\mathrm{t}} = \frac{1}{M}\sum_{m=1}^{M}(X_{\mathrm{t}}^m - \overline{X_{\mathrm{t}}}) \otimes (X_{\mathrm{t}}^m - \overline{X_{\mathrm{t}}}), \tag{11}$$

其中， $\overline{X_{\mathrm{t}}}$ 是第 t 次迭代时状态的经验均值。为纠正朗之万 SDE 时间离散化所引入的偏差，采用 Metropolis 接受-拒绝步骤。提议 $X_{\mathrm{t}+1}^m$以概率

$$a = \min\left\{\frac{p(X_{\mathrm{t}+1}^m)q(X_{\mathrm{t}}^m|X_{\mathrm{t}+1}^m)}{p(X_{\mathrm{t}}^m)q(X_{\mathrm{t}+1}^m|X_{\mathrm{t}}^m)}, 1\right\}, \tag{12}$$

提议分布 $q(\cdot|\cdot)$ 可由（10）推导得出，且为

$$q(X_{t+1}|X_t) \propto \exp\left(-\frac{1}{4\gamma}\left\|X_{t+1} - X_t - \gamma \mathbf{C}_t \nabla_{X_t} \log p(X_t)\right\|^2_{\mathbf{C}_t}\right). \tag{13}$$

下一节中，我们将讨论基于上述背景提出的方法论。

3 方法论

本节详细介绍如何联合优化分层参数推断与最大似然闭合。此外，我们还将介绍双层优化框架，通过联合训练代理模型来加速逆问题求解。最后，将详细讨论具体的代理模型结构及其相应的训练方法。

3.1 分层逆问题与最大似然闭合

我们首先介绍如何使用传统数值正向求解器求解逆问题，而不引入代理模型。本节对应于我们的贡献 C1 和C2。正向模型见（1），我们的目标是根据稀疏且含噪观测 $\boldsymbol{\theta}^{(1:K)}$和 $\boldsymbol{\phi}$，推断物理参数 f 以及非线性闭合项 $\mathbf{y}^{(1:K)}$。如第 2.1 节所述，不同物理系统的参数采用分层贝叶斯框架进行建模，并假设每个 $\boldsymbol{\theta}^{(k)}$ 均取自同一个高斯分布，

$$p(\boldsymbol{\theta}^{(k)}|\boldsymbol{\phi}) = \mathcal{N}(\boldsymbol{\theta}^{(k)}; \boldsymbol{\mu}_\phi, \boldsymbol{\tau}_\phi \mathbf{I}), \tag{14}$$

其中 $\boldsymbol{\phi} = \{\boldsymbol{\mu}_\phi, \boldsymbol{\tau}_\phi\}$ 以及超先验

$$p(\boldsymbol{\phi}) = p(\boldsymbol{\mu}_\phi)p(\boldsymbol{\tau}_\phi) = \prod_{i=1}^{P} \mathcal{N}(\mu_{\phi_i}; m_\mu, s_\mu)\mathcal{N}(\log(\tau_{\phi_i}); m_\tau, s_\tau), \tag{15}$$

其中，i 表示感兴趣的未知参数，每个系统中有 P 个未知参数。参数m_μ、s_μ、m_τ和 s_τ 是已知值，用于编码关于超先验分布的先验信念。对数形式的 $\boldsymbol{\tau}_\phi$ 对应于对数正态分布，并确保超先验中的τ_{ϕ_i}始终为正。

为了表示未知的非线性闭合项，我们采用一个神经网络 f^α，其可训练参数为 α。每个系统的观测值通过（2）获得，因此系统k 的观测值似然形式如下

$$p(\mathbf{y}^{(k)}|\boldsymbol{\theta}^{(k)}, \alpha) = \mathcal{N}(\mathbf{y}^{(k)}; G^{(k)}(\boldsymbol{\theta}^{(k)}, \alpha), \sigma^2 \mathbf{I}), \tag{16}$$

其中，$G^{(k)} = g^{(k)} \circ F^\dagger$表示将（$\boldsymbol{\theta}^{(k)}$，$\alpha$）映射到相应观测值的算子。为简化表示，此处将输入记为 α，表示 f^α。根据控制方程的结构，参数到解算子 $F^\dagger$ 可以是用于动力系统的蛙跳积分器或欧拉时间步进格式，也可以是用于非线性定常问题的高斯–牛顿求解器。

为了学习闭合模型 f^α，我们针对 α 最大化观测数据 $\mathbf{y}^{(1:K)}$的边际似然。然而，由于存在未知潜变量 $\boldsymbol{\theta}^{(1:K)}$ 和 $\boldsymbol{\phi}$，其对数

边际似然为

$$\begin{aligned}\log p(\mathbf{y}^{(1:K)}|\alpha) &= \log\left(\int p(\mathbf{y}^{(1:K)}|\boldsymbol{\phi},\alpha)p(\boldsymbol{\phi}|\alpha)d\boldsymbol{\phi}\right),\\ &= \log\left(\int\left[\prod_{k=1}^{K}\int p(\mathbf{y}^{(k)}|\boldsymbol{\theta}^{(k)},\alpha)p(\boldsymbol{\theta}^{(k)}|\boldsymbol{\phi})d\boldsymbol{\theta}^{(k)}\right]p(\boldsymbol{\phi}|\alpha)d\boldsymbol{\phi}\right),\end{aligned} \tag{17}$$

它在解析上和计算上都难以处理。因此，我们考虑参数估计问题中另一个有用的量，即对数边际似然相对于 α 的梯度。根据费舍尔恒等式 [59],

$$\nabla_\alpha \log p(\mathbf{y}^{(1:K)}|\alpha) = \mathbb{E}_{p(\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi}|\mathbf{y}^{(1:K)},\alpha)}\left[\nabla_\alpha \log p(\mathbf{y}^{(1:K)},\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi}|\alpha)\right]. \tag{18}$$

期望是相对于联合后验分布 $p(\boldsymbol{\theta}^{(1:K)},\ \boldsymbol{\phi}|\mathbf{y}^{(1:K)},\ \alpha)$ 求取的，该分布可以分解为

$$\begin{aligned}p(\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi}|\boldsymbol{y}^{(1:K)},\alpha) &= \frac{p(\mathbf{y}^{(1:K)},\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi}|\alpha)}{p(\boldsymbol{y}^{(1:K)}|\alpha)},\\ &= \frac{p(\boldsymbol{\phi})\prod_{k=1}^{K}p(\mathbf{y}^{(k)}|\boldsymbol{\theta}^{(k)},\alpha)p(\boldsymbol{\theta}^{(k)}|\boldsymbol{\phi})}{p(\boldsymbol{y}^{(1:K)}|\alpha)}.\end{aligned} \tag{19}$$

期望括号内的项， $p(\mathbf{y}^{(1:K)},\ \boldsymbol{\theta}^{(1:K)},\ \boldsymbol{\phi}|\alpha)$表示联合概率密度，是（19）中的分子。我们使用（19）中的样本来近似（18）中的期望。

为了通过（18）学习 α ，我们交替进行使用集成MALA对（19）的采样，以及对 α 的梯度下降。参数从 α_0 初始化，并维护由 M 条 MALA 链组成的集成，其状态向量从 $X_0^{1:M} = (\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi})_0^{1:M}$ 开始。在迭代 t 中，给定当前参数估计值 α_t，我们在所有链上执行一步朗之万采样。利用第 2.2 节详述的集成MALA过程，从当前后验分布 $p(\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi}|\boldsymbol{y}^{(1:K)},\alpha_\mathsf{t})$ 中抽取样本 $X_{\mathsf{t}+1}^{1:M}$。随后使用这些来自 M 条链的样本近似（18）中的梯度，得到 $\alpha_{\mathsf{t}+1}$，其中用于更新的梯度为

$$\nabla_\alpha \log p(\mathbf{y}^{(1:K)}|\alpha_\mathsf{t}) \approx \sum_{m=1}^{M}\nabla_\alpha \log p(\boldsymbol{y}^{(1:K)},(\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi})_{\mathsf{t}+1}^{m}|\alpha_\mathsf{t}). \tag{20}$$

为与标准深度学习框架兼容（在这些框架中，人们定义一个函数，并使用自动微分对其求导），我们将用于学习 α 的目标函数定义为

$$\mathcal{L}_{\mathsf{LML}}(\alpha) = -\sum_{m=1}^{M}\log p(\boldsymbol{y}^{(1:K)},(\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi})_{\mathsf{t}+1}^{m}|\alpha), \tag{21}$$

其中省略了不必要的常数。

训练损失 $\mathcal{L}_{\mathsf{LML}}$ 在经过 T 次迭代后收敛，我们便启动训练后采样阶段。固定神经网络参数 α_T 和预条件器 $\mathbf{C}_\mathsf{T}$后，我们推进 MALA 链，从由此得到的平稳后验分布 $p(\boldsymbol{\theta}^{(1:K)},\boldsymbol{\phi}|\boldsymbol{y}^{(1:K)},\alpha_\mathsf{T})$ 中进行采样。由此获得系统参数 $\boldsymbol{\theta}^{(1:K)}$ 和超参数 $\boldsymbol{\phi}$的最终后验样本，以及由 α_T定义的学习得到的闭合模型。

3.2 前向代理模型的联合训练

由于集成MALA需要大量前向模拟，且每个提议步骤都需要计算前向求解的梯度，使用传统数值求解器在计算上可能会变得难以承受。为解决这一局限，我们以一个将在推断方案中同步学习的代理模型取代数值正向求解器。这对应于贡献 C3。在这项工作中，我们将傅里叶神经算子和物理信息神经网络作为代理架构，但也可以使用其他可微模型。

我们将代理前向模型记为 F^β。与（1）中的数值正向求解器不同，代理模型无法直接将高维神经网络参数 α 作为输入。因此，F^β近似于在 $f^\alpha \approx f$处计算的数值求解器，记为

$$F^\dagger(\boldsymbol{\theta}^{(k)}, \alpha) \approx F^\beta(\boldsymbol{\theta}^{(k)}) = \hat{\mathbf{u}}^{(k)}. \tag{22}$$

我们期望，在给定观测数据的条件下，该近似在 $\boldsymbol{\theta}^{(k)}$ 的后验分布具有较高概率的区域内足够准确，其中 $k \in \{1, ..., K\}$，且 $(\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi}) \sim p\,(\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi}|\mathbf{y}^{(1:K)}, \alpha)$。

尽管 F^β 不直接以 α 作为输入，但 α 的影响会通过代理训练损失 $\mathcal{L}_{\text{Surrogate}}(\alpha, \beta)$ 间接体现，其形式将在第 3.2.2节讨论。在接下来的章节中，我们首先介绍如何通过双层优化框架将代理模型训练融入逆问题，从而联合学习代理模型和未知物理机制。随后，我们将介绍用于训练这些代理模型的不同损失函数形式。

3.2.1 双层更新方案

正问题与反问题的联合学习自然会形成一种双层优化结构，其中需要同时求解两个相互耦合的子问题。双层优化 [39] 指的是上层目标函数的解依赖于下层问题的最优解的问题。在本研究中，双层问题的形式为

$$\min_{\alpha} \ \mathcal{L}_{\mathsf{LML}}(\alpha, \beta^*(\alpha)) \qquad \text{s.t.} \qquad \beta^*(\alpha) = \arg\min_{\beta} \mathcal{L}_{\mathsf{Surrogate}}(\alpha, \beta), \tag{23}$$

其中，$\mathcal{L}_{\mathsf{LML}}$ 表示上层目标函数（最大化边际似然），$\mathcal{L}_{\mathsf{Surrogate}}$ 表示下层目标函数（训练代理模型）。这种表述要求两个层级之间进行迭代交互。在每一步中，首先根据当前的 α对下层变量 β 进行优化。随后，通过对上层目标函数求导来更新闭合参数 α，同时考虑最优代理模型 β^* 对 α的隐式依赖。该机制有效地将梯度信息从内部代理训练循环传递到外部闭合学习循环。双层结构源于这两个组成部分之间的相互依赖：

- **上层（闭合学习）：** 优化闭合参数 α 需要评估数据的边际似然。该评估依赖于代理前向模型 $F^{\beta^*(\alpha)}$ 将参数 $\boldsymbol{\theta}$ 映射到观测值 $\mathbf{y}$。

- **下层（代理训练）：** 训练由 β 参数化的代理模型需要了解控制方程，其中包括当前闭合模型 f^α 的形式。因此，最优代理参数 β^* 隐式地依赖于 α。

接下来，我们详细介绍本方法论的完整训练方案。每次迭代包含三个阶段：一次采样步骤、在 β 上进行 N 优化步骤，以及在 α上进行一次更新步骤。在第 t 次迭代中，

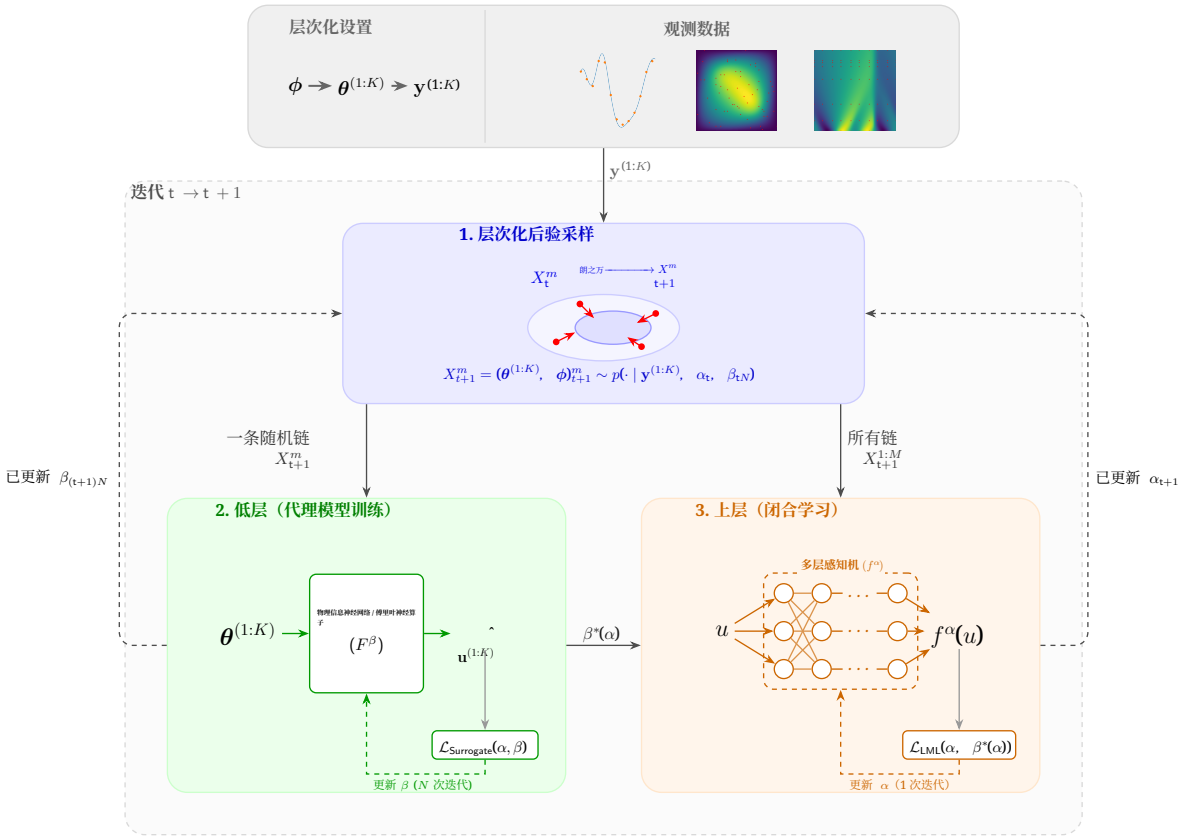


图 2： **所提出框架的概览**。 该算法在三个主要阶段之间交替进行：**(1) 层次化后验采样** 使用集成MALA更新参数链，**(2) 下层代理模型训练** 优化代理模型 F^{β}，其参数化形式为 β，历经 N 次迭代，其中损失函数 $\mathcal{L}_{\text{Surrogate}}$ 可采用式（27）、（28）和（29）中的形式；**(3) 上层闭合学习** 优化闭合模型 f^{α}，其参数化形式为 α，以发现未知的非线性动力学，其中损失函数 $\mathcal{L}_{\text{LML}}$ 遵循式（26）。

首先使用集成MALA 采样器执行采样步骤，从而获得

$$(\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi})_{\mathrm{t+1}}^{1:M} \sim p(\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi} | \boldsymbol{y}^{(1:K)}, \alpha_{\mathrm{t}}, \beta_{\mathrm{t}N}). \tag{24}$$

请注意，此处使用高效的代理模型计算朗之万采样中的梯度和接受概率，从而避免调用计算开销巨大的数值求解器。此外，采样阶段之前会阻断梯度传播，因此，对 α 和 β 的优化不涉及对 MALA采样轨迹进行微分。

随后， N 执行下层优化步骤以更新 β。为降低计算成本，损失仅使用一个随机MALA链的样本进行评估（$\boldsymbol{\theta}^{(1:K)}$， $\boldsymbol{\phi}$）^{p} $_{\mathrm{t+1}}$，其中链索引 p 从均匀分布$\{1, \ldots, M\}$中抽取。 N对 β执行梯度下降步骤以最小化$\mathcal{L}_{\text{Surrogate}}$（$\alpha$， β），其具体形式将在第3.2.2节讨论。

最后，在上层，闭合参数 α 在代理参数固定于其优化状态时进行更新 $\beta^{*}(\alpha) = \beta_{(\mathrm{t}+1)N}$。目标函数与第3.1节中所述的类似。然而，如前所述，代理模型不会显式地将 α 作为输入。因此，需要对某些表示形式进行更改。式（16）中的原始似然分布可重新表述为

$$p(\mathbf{y}^{(k)} | \boldsymbol{\theta}^{(k)}, \beta^{*}(\alpha)) = \mathcal{N}(\mathbf{y}^{(k)}; g^{(k)} \circ F^{\beta^{*}(\alpha)}(\boldsymbol{\theta}^{(k)}), \sigma^{2}\mathbf{I}). \tag{25}$$

因此，式（ α 21）中的损失函数变为

$$\mathcal{L}_{\mathsf{LML}}(\alpha, \beta^*(\alpha)) = -\sum_{m=1}^{M} \log p(\boldsymbol{y}^{(1:K)}, (\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi})_{\mathrm{t+1}}^{m} | \beta^*(\alpha)). \tag{26}$$

架构示意图如图2所示。完整算法，包括后验采样以及闭合项和代理模型的双层优化，见算法1。

3.2.2 代理模型训练

本文考虑两种代理架构。第一种是傅里叶神经算子（FNO），它通过在傅里叶域中学习全局卷积核来逼近函数空间之间的映射。对于每个系统，FNO 以空间坐标以及可能的时间坐标和物理参数 $\boldsymbol{\theta}^{(k)}$作为输入，并输出完整的预测解场 $\hat{\mathbf{u}}^{(k)}$。我们使用两类损失函数训练 FNO。第一类是监督损失，其中将 $F^{\beta}(\boldsymbol{\theta}^{(k)})$ 与数值求解器计算得到的状态进行匹配，损失函数等于

$$\mathcal{L}_{\text{FNO Super}}(\alpha, \beta) = \frac{1}{K} \sum_{k=1}^{K} \left\| F^{\beta}(\boldsymbol{\theta}^{(k)}) - F^{\dagger}(\boldsymbol{\theta}^{(k)}, \alpha) \right\|^2, \tag{27}$$

其中， $\boldsymbol{\theta}^{(k)}$ 是所选 MALA 链（$\boldsymbol{\theta}^{(1:K)}$， $\boldsymbol{\phi}$）$^{p}{}_{\mathrm{t+1}}$中的一个元素。第二种方法是纯粹的基于物理的损失，其中使用有限差分计算常微分方程/偏微分方程残差，以确保与控制方程的物理一致性。损失函数表示为

$$\mathcal{L}_{\text{FNO Phy}}(\alpha, \beta) = \frac{1}{K} \sum_{k=1}^{K} \left\| \mathcal{R}(\boldsymbol{\theta}^{(k)}, \alpha, \beta) \right\|^2, \tag{28}$$

其中 $\mathcal{R}$ 通过对代理模型的输出进行有限差分来计算物理残差 $F^{\beta}(\boldsymbol{\theta}^{(k)})$。在某些偏微分方程情形下，也会采用弱形式残差，我们将在第 4 节详细讨论。

我们还采用物理信息神经网络作为另一种代理架构。与仅将时空坐标作为输入的标准物理信息神经网络不同，我们的公式将物理参数 $\boldsymbol{\theta}^{(k)}$ 加入输入。这样，物理信息神经网络便能够表示整个动力系统族，而非单个实例，从而近似多个系统中从 $\boldsymbol{\theta}^{(k)}$ 到 $\mathbf{u}^{(k)}$ 的映射。训练完全基于物理信息目标函数进行，损失函数遵循传统物理信息神经网络的结构，由在配点处计算的控制常微分方程/偏微分方程的物理残差，以及用于约束初始条件和边界条件的惩罚项组成。损失函数为

$$\mathcal{L}_{\text{PINNs}}(\alpha, \beta) = \frac{1}{K} \sum_{k=1}^{K} \left(\mathcal{L}_{\text{Physics}}(\boldsymbol{\theta}^{(k)}, \alpha) + \mathcal{L}_{\text{BC/IC}}(\boldsymbol{\theta}^{(k)}) \right), \tag{29}$$

其中， $\mathcal{L}_{\text{Physics}}$ 根据给定方程计算，并通过自动微分求得。并且

$$\mathcal{L}_{\text{BC/IC}}(\boldsymbol{\theta}^{(k)}) = \|F^{\beta}|_{\partial\Omega}(\boldsymbol{\theta}^{(k)}) - \mathbf{g}_{\text{BC}}^{(k)}\|^2 + \|F^{\beta}|_{t=0}(\boldsymbol{\theta}^{(k)}) - \mathbf{g}_{\text{IC}}^{(k)}\|^2, \tag{30}$$

其中$\mathbf{g}^{(k)}$ 边界条件 和 $\mathbf{g}^{(k)}$ 初始条件 分别表示边界条件和初始条件。

总之，基于 FNO 和基于 PINN 的代理模型能够高效地近似多个系统中的正向算子 $F^{\dagger}$ ，同时提供基于梯度的方法所需的可微性

算法 1 通过双层优化框架进行联合逆推断和代理模型训练 要求：观测值 $\mathbf{y}^{(1:K)}$；初始参数 α_0, β_0；MALA 链的数量 M；初始 MALA 链 $X_0^{1:M}$；上层迭代次数 T；低层迭代次数 N；MALA步长 γ；闭合模型步长 η_α；代理模型步长 η_β；1: 对于 t $= 0, 1, \ldots,$ T $- 1$ 执行 // 使用集成MALA进行后验采样 2: 对于 $m = 1, \ldots, M$ 执行 3: 通过一次 MALA 步更新链 m：$X_{t+1}^m = \text{MALA}(X_t^m; \beta_{tN}, \gamma)$，其中 $X_t^m = (\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi})_t^m$。4: 结束循环 // 低层：代理模型更新 5: 从均匀分布$\{1, \ldots, M\}$中选择一个随机 MALA 链索引 $p \sim$， 6: 对于 $n = 0, \ldots, N-1$ 执行 7: 计算 $\mathcal{L}_{\text{Surrogate}}(\alpha, \beta) = \mathcal{L}_{\text{FNO}}$ Super/FNO Phy/PINNs(α, β)，使用 $\left(\boldsymbol{\theta}^{(1:K)}\right)^p$ t$+1$。8: 更新代理参数：$\beta_{tN+n+1} = \beta_{tN+n} - \eta_\beta \nabla_\beta \mathcal{L}_{\text{Surrogate}}(\alpha, \beta)$。9: 结束循环 // 上层：闭合模型更新 10: 计算 $\mathcal{L}_{\text{LML}}(\alpha, \beta^*(\alpha)) = -\sum_{m=1}^M \log\, p(\boldsymbol{y}^{(1:K)}, (\boldsymbol{\theta}^{(1:K)}, \boldsymbol{\phi})_{t+1}^m | \beta^*(\alpha))$，其中 $\beta^*(\alpha) = \beta_{(t+1)N}$。11: 更新闭合参数：$\alpha_{t+1} = \alpha_t - \eta_\alpha \nabla_\alpha \mathcal{L}_{\text{LML}}(\alpha, \beta^*(\alpha))$。12: 结束循环 13: 返回后验样本 $X^{(1:M)}$ T、闭合参数 α_T、代理参数 β_{TN}。

在集成MALA框架中进行采样。值得注意的是，这些代理模型并非预先离线训练，而是与逆问题在线联合优化。这种耦合训练策略确保代理模型能够在 MALA 采样器探索的参数区域内自适应地得到优化。

4 数值实验

本节针对三类微分方程验证所提出的方法论：非线性质量-阻尼常微分方程、非线性二维达西流以及广义伯格斯方程，这些方程分别对应贡献C4。

4.1 非线性质量-阻尼系统

我们的第一个实验研究一个非线性质量-阻尼系统。在此设置中，每个系统的刚度、初始位置和速度均视为未知物理参数，而非线性阻尼定律则作为未知闭合模型。

4.1.1 问题表述

非线性质量-阻尼系统定义如下：

$$\begin{aligned} \ddot{u}(t) + f(\dot{u}(t)) + k\,u(t) &= 10\sin(t), \quad t \in [0, 8], \\ u(0) &= u_0, \\ \dot{u}(0) &= v_0. \end{aligned} \tag{31}$$

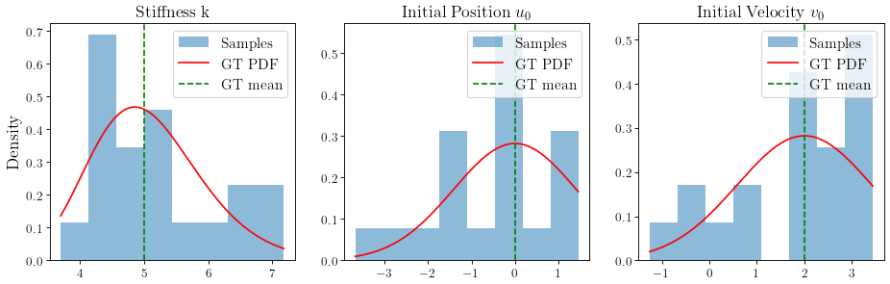


图3：**实验1中 K = 20 个质量-阻尼系统的真实物理参数分布**。每个系统的参数向量 $\boldsymbol{\theta} = \{\log(k)\ u_0,\ v_0\}$ 均从超参数为 $\boldsymbol{\mu}_\phi = \{\log(5.0)\,0.0,\ 2.0\}$ 和 $\boldsymbol{\tau}_\phi = \{0.03, 2.0, 2.0\}$ 的层次先验中采样得到。直方图显示了20个系统中的参数分布，叠加曲线则表示层次先验所蕴含的相应真实概率密度函数。绿色虚线表示真实总体均值。直方图与真实曲线看似存在偏差，这是因为仅由20个系统构成的有限总体无法充分代表完整的底层概率分布。

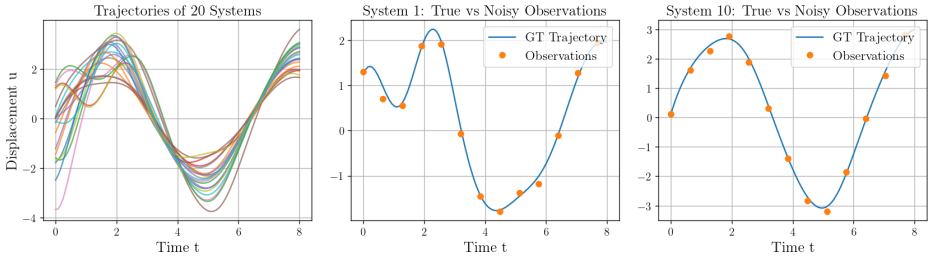


图4：**20个非线性质量-阻尼系统的位移轨迹 $\mathbf{u}(\mathbf{t})$**。这些系统由方程（31）描述，并采用蛙跳积分器求解，所用参数因系统而异，见图3。右侧两个面板展示了两个代表性系统通过观测算子生成的相应稀疏且含噪观测 $g^{(k)}$。

其中，u 表示质量块的位移，非线性阻尼 $f(\dot{u}) = 0.08\dot{u}^3 + 0.08\dot{u}$ 在所有系统中共享。刚度 k、初始位置和速度 u_0, v_0 因系统而异。为确保刚度 k 始终为正，待推断的参数设为 $\boldsymbol{\theta} = \{\log(k)\ u_0, v_0\}$。总体的真值超参数为 $\boldsymbol{\mu}_\phi = \{\log(5.0)\ 0.0, 2.0\}$ 和 $\boldsymbol{\tau}_\phi = \{0.03, 2.0, 2.0\}$。按照（14），真实参数 $\boldsymbol{\theta}$ 由超先验生成，生成系统的参数直方图如图 3 所示。

4.1.2 求解器

用于求解（31）的数值求解器是蛙跳积分器，这是一种常用于振荡系统的二阶格式。使用时间步长 $\Delta t = 0.08s$ 的蛙跳求解器，图 4 展示了针对 $K = 20$ 个系统生成的位移轨迹，这些系统对应于图 3 所示的参数值。此外，还进一步展示了两个代表性系统及其对应的稀疏且含噪的测量值 $\mathbf{y}$，这些测量值每 8 步获取一次。在第一个实验中，观测算子 $g^{(k)}$ 在各系统间相同，观测噪声设为 $\sigma = 0.15$。

4.1.3 代理模型实现细节

我们评估了质量—阻尼器系统的四个模型。第一个基线模型使用蛙跳积分器直接求解逆问题，不引入任何代理模型，如第3.1节所示。为了避免在贝叶斯推断过程中使用迭代求解器，随后我们构建了两种不同的代理模型——傅里叶神经算子和物理信息神经网络，并通过双层方案将其与逆问题结合进行训练。

如第3.2.2节所述，傅里叶神经算子代理模型采用两种可选的损失形式进行训练：纯基于物理的残差损失和监督损失。在质量—阻尼器系统中，傅里叶神经算子的输入为参数向量 $\boldsymbol{\theta}^{(k)}$ 和时间坐标$\mathbf{t}$，输出位移$\mathbf{w}$ 和速度$\mathbf{v}$。同时生成这两个量能够通过有限差分获得更准确的时间导数。公式（ $\mathcal{L}_{\text{FNO Phy}}$ 中的基于物理的损失由两个残差项求和得到：28）

$$
\begin{aligned}
\mathcal{R} &= \mathcal{R}_1 + \mathcal{R}_2, \\
\mathcal{R}_1 &= \|\dot{\mathbf{w}} - \mathbf{v}\|^2, \\
\mathcal{R}_2 &= \|\dot{\mathbf{v}} - (10\sin(\mathbf{t}) - f^{\alpha}(\mathbf{v}) - k\mathbf{w})\|^2,
\end{aligned} \tag{32}
$$

其中时间导数 $\dot{\mathbf{w}}$ 和 $\dot{\mathbf{v}}$ 采用中心有限差分进行近似。式（ $\mathcal{L}_{\text{FNO Super}}$27）中定义的监督式傅里叶神经算子损失用于惩罚傅里叶神经算子输出与蛙跳积分器计算出的参考轨迹之间的差异。

基于 PINN 的代理模型也将 $\boldsymbol{\theta}^{(k)}$ 和 t 作为输入，并仅预测位移。所需的一阶和二阶导数通过自动微分获得，使 PINN 能够在式（31）中施加控制方程。有关网络结构和训练的更多实现细节见附录A。

在接下来的章节中，我们将评估本文所考虑的四种模型的性能：一种使用**数值正向求解器求解逆问题**的基线方法，以及三种由代理模型加速的方法。在这些方法中，正向求解器分别被**采用监督式损失或基于物理的损失训练的傅里叶神经算子**或**基于 PINN 的代理模型**所取代。

4.1.4 结果

我们将评估分为三个主要部分：四种模型的性能比较、层次贝叶斯设定与非层次贝叶斯设定的比较，以及计算效率分析。

逆问题和正问题上的性能比较。

我们根据三个关键标准比较这四个模型：（i）推断物理参数的准确性；（ii）估计非线性闭合模型的准确性；以及（iii）代理模型近似正向动力学的质量。实验涵盖了 K的各种取值范围，从 5 到 100 不等。然而，对于 $K > 30$的情况，采用数值求解器作为前向映射的方法未予实现。这一限制是由于内存溢出造成的，因为求解器的多次时间迭代会在按照 MALA 的要求计算梯度时消耗过多内存。

图 5 总结了四个模型在估计闭合模型以及使用学习得到的代理模型预测系统轨迹方面的性能，其中 K 设为 20。由于本实验中的闭合项 f 是在速度 $\dot{u}$上进行评估的，因此我们还绘制了从真实轨迹中提取的速度分布。该分布揭示了真实数据中存在的速度范围，并相应地解释了为什么学习得到的闭合模型在这一范围内表现出更高的准确性。定量结果，包括闭合项估计的均方误差（MSE）（在速度范围 $\dot{u} \in [-5, 5]$上进行评估，因为绝大多数真实速度数据都落在此区间内）以及代理模型在所有系统上的预测均方误差，均列于表 1中。

根据图 5 和表 1，数值求解器和监督式傅里叶神经算子在闭合模型预测中表现出很高的准确性，能够有效捕捉三次非线性。物理信息神经网络也取得了相对准确的结果。相比之下，基于物理的傅里叶神经算子表现不佳，未能提供

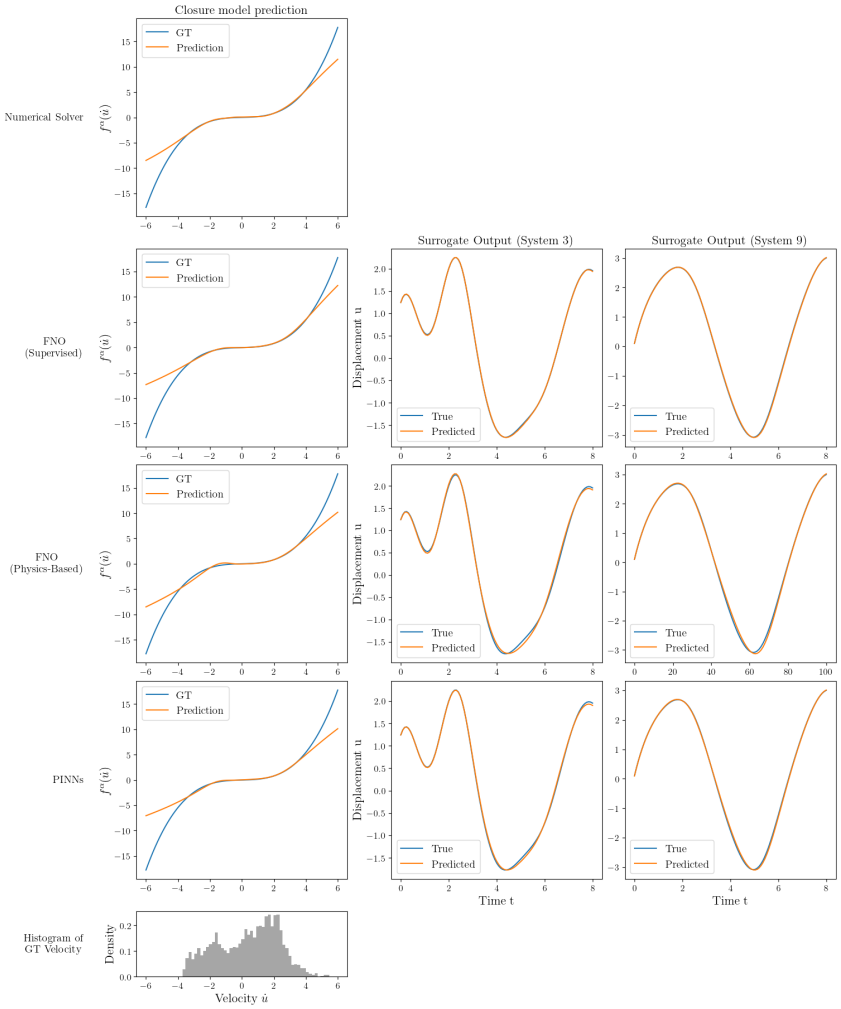


图 5：**不同模型在K = 20 下的闭合模型估计和代理性能比较，实验1。** 各行分别对应数值求解器、监督式傅里叶神经算子、基于物理的傅里叶神经算子和物理信息神经网络。**左列：**在速度范围 $\dot{u} \in [-6, 6]$ 内，比较学习得到的闭合项 $f^{\alpha}(\dot{u})$（橙色）与真值 $f(\dot{u})$（蓝色）。**中间列 & 右列：**两个选定系统（系统3和系统 9）的位移代理预测 $u(t)$。**底部面板：**真值数据集中速度值的直方图。

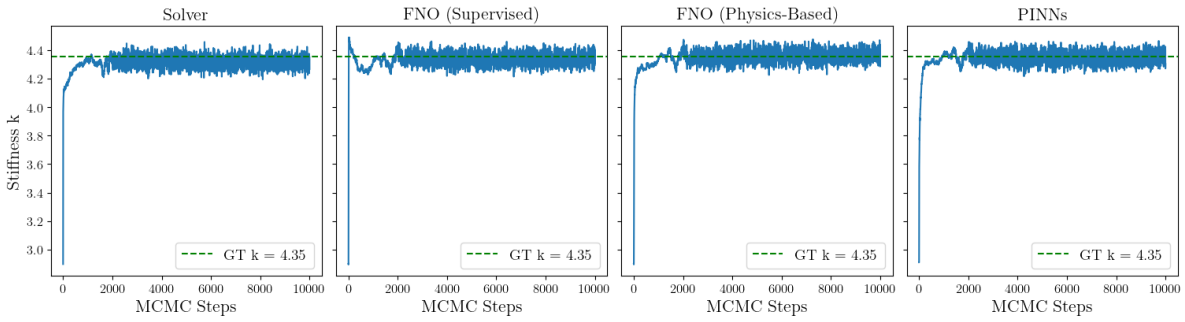


图 6：**轨迹图展示了实验1中 K = 20 条件下，系统 9 刚度k 的一个随机选取的 MALA 链在四种模型中的演变。**图中展示了相对于真值参数值（绿色虚线）的收敛行为。

准确　　预测，在 $K = 5$、10、30 处观察到显著较大的误差，如表 1　　.

表1：**实验1中，四种模型在系统数量（$\mathbf{K} \in \{5, 10, 20, 30, 50, 100\}$）不同时的定量比较。** 该表通过后验均值相对于真实参数值的均方误差和**覆盖率**评估**参数推断**性能；覆盖率定义为落在后验标准差两倍范围内的真实参数所占百分比。此外，**闭合均方误差** 衡量学习得到的非线性闭合项 $f^{\alpha}(\bar{u})$ 在 $\bar{u} \in [-5, 5]$ 上的均方误差，而**代理均方误差** 则衡量代理模型的轨迹预测相对于真实轨迹的准确性。

模型	系统数量	参数推断		闭合均方误差	代理均方误差
		平均均方误差	覆盖率		
求解器	$K = 5$	1.44×10^{-2}	100%	1.05	-
	$K = 10$	2.41×10^{-2}	93.3%	0.48	-
	$K = 20$	2.70×10^{-2}	96.7%	0.50	-
	$K = 30$	3.02×10^{-2}	91.1%	0.45	-
FNO（监督式）	$K = 5$	4.85×10^{-2}	93.3%	1.05	1.53×10^{-3}
	$K = 10$	1.74×10^{-2}	96.7%	0.84	8.60×10^{-4}
	$K = 20$	3.13×10^{-2}	93。3%	0.88	8.44×10^{-4}
	$K = 30$	3.42×10^{-2}	93。3%	0.88	1.24×10^{-3}
	$K = 50$	3.56×10^{-2}	85.3%	0.58	9.37×10^{-4}
	$K = 100$	4.75×10^{-2}	79。7%	0.69	9.68×10^{-4}
FNO（基于物理的）	$K = 5$	1.51×10^{-1}	73.3%	4.08	1.43×10^{-2}
	$K = 10$	1.27×10^{-1}	60.0%	2.16	1.62×10^{-2}
	$K = 20$	6.77×10^{-2}	86.7%	1.38	4.58×10^{-3}
	$K = 30$	2.62×10^{-1}	73.3%	1.54	1.44×10^{-2}
	$K = 50$	6.86×10^{-2}	78。7%	1.30	6.31×10^{-3}
	$K = 100$	7.38×10^{-2}	63。3%	1.26	7.78×10^{-3}
物理信息神经网络	$K = 5$	4.19×10^{-2}	86.7%	1.18	3.80×10^{-3}
	$K = 10$	2.44×10^{-2}	93。3%	1.15	1.66×10^{-3}
	$K = 20$	6.37×10^{-2}	90.0%	1.27	3.97×10^{-3}
	$K = 30$	5.59×10^{-2}	92。2%	1.15	2.64×10^{-3}
	$K = 50$	5.06×10^{-2}	81.3%	0.79	2.29×10^{-3}
	$K = 100$	5.77×10^{-2}	70.0%	1.02	3.30×10^{-3}

此外，图 5 表明，学习得到的闭合模型的准确性取决于数据的可用性：在训练数据密度较高的范围内，其性能更优。关于代理模型训练，监督式傅里叶神经算子表现出很高的准确性，在所有设置下始终取得最低的均方误差。物理信息神经网络排名第二，表现具有竞争力，其误差处于 10^{-3}范围内，略高于监督式傅里叶神经算子。相比之下，基于物理的傅里叶神经算子准确性最低，误差显著恶化至 10^{-2} 范围，约比其他方法高一个数量级。

接下来，我们评估四种模型在参数推断方面的性能。图 6 展示了从 M 条刚度为 k 的 MALA 链中随机选取的一条链在系统 9 中的演化过程，说明每种模型生成的链如何收敛到各自的后验分布。如第 3.1 节所述，采样过程包括两个阶段：前 9,000 个样本是在逆问题和正问题的联合训练期间生成的，随后通过集成MALA，在固定 α、β 和预条件器 C的情况下，额外收集了 1,000 个样本。图 7 进一步展示了每种模型最后 1,000 个步骤的后验直方图，以及系统 9 的全部三个参数所对应的近似概率密度函数。表 1 还总结了四种模型在参数推断方面的定量比较。该表报告了估计的后验均值与真实参数之间的均方误差，该误差在所有系统和所有推断参数上取平均。表中还报告了估计后验分布的覆盖率，其定义为真实参数值落在两个后验标准差范围内的百分比。这些统计量共同评估了参数推断的准确性以及后验不确定性量化的质量。此外，在实验过程中，我们注意到，省略偏差校正 Metropolis 步骤或基于集成的预条件器中的任意一个，都会导致不收敛的朗之万动力学，表明这些组件对于获得稳健且稳定的算法发挥着关键作用。

总体而言，数值求解器和监督式傅里叶神经算子提供了最稳健的参数推断，

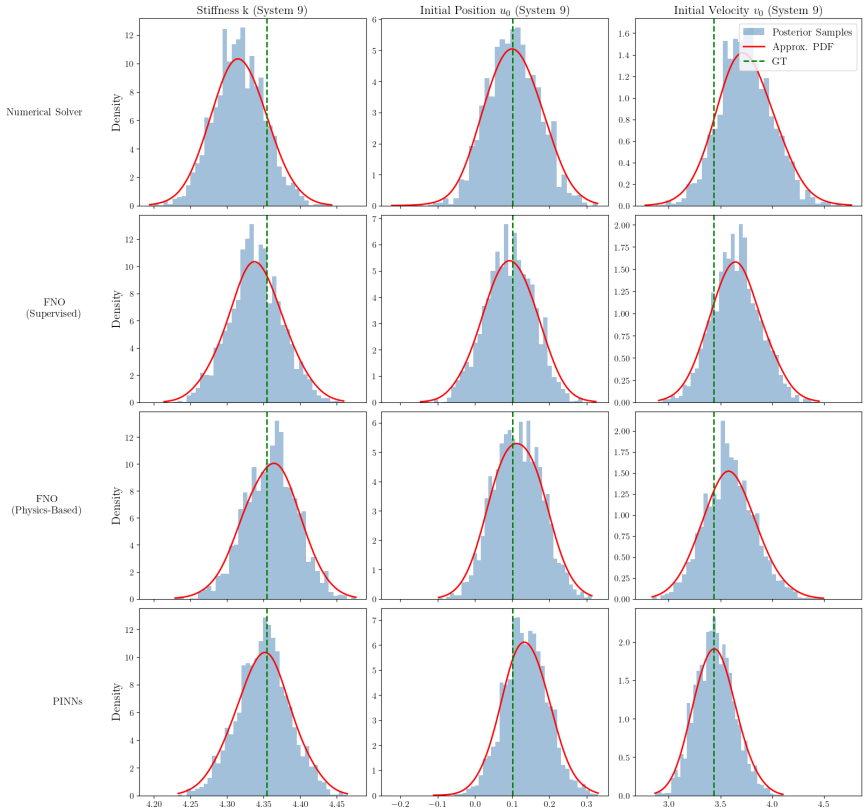


图 7：**实验1中，系统 9 的三个参数在四个模型下的后验直方图及相应的近似概率密度函数，其中 $\mathbf{K} = 20$** 。每个面板展示了收敛后从一条随机 MALA 链的最后 1,000 个样本得到的边际后验分布。蓝色柱状图表示后验样本，红色曲线表示近似密度，绿色虚线标出真值参数值。

其均方误差最低，且覆盖率始终较高。物理信息神经网络表现不俗，证明其是可靠的基于物理的替代方案，准确率仅略低。相比之下，基于物理的傅里叶神经算子的后验分布校准不佳；其均方误差显著偏高，覆盖率也较低（在 $K = 10$ 时降至 60%），这表明估计误差显著，真实值经常落在预测后验分布之外。

此外， K 表 1 中对不同系统数量的比较显示出明显趋势。随着可用系统数量增加，闭合模型和代理预测的准确性通常都会提高，这可能是由于训练数据量增加所致。相反，参数推断的准确性往往会下降。这种下降归因于在更高维后验分布上进行推断本身具有难度。

层次贝叶斯与非层次贝叶斯

接下来，我们进行消融研究，通过与非层次替代方案进行比较，探究层次贝叶斯表述的影响。在非层次设置中，每个参数都独立处理，且不推断总体层超参数 ϕ。我们使用物理信息神经网络作为代理前向模型，对两种表述进行测试；该模型在双层框架中与逆问题联合训练。表 2报告了这两种方法的性能，所用指标与表 1相同。

层次贝叶斯的一个基本机制是能够借助总体的信息，从而获得准确的参数估计并降低估计不确定性。对于后验

表 2： **实验1中使用物理信息神经网络作为代理模型训练的分层与非分层框架的比较。** 报告的指标与表1相同。

方法	数量系统	参数推断		闭合均方误差	代理均方误差
		平均均方误差	覆盖率		
分层	$K = 5$	4.19×10^{-2}	86.7%	1.18	3.80×10^{-3}
	$K = 10$	2.44×10^{-2}	93。3%	1.15	1.66×10^{-3}
	$K = 20$	6.37×10^{-2}	90。0%	1.27	3.97×10^{-3}
	$K = 30$	5.59×10^{-2}	92。2%	1.15	2.64×10^{-3}
	$K = 50$	5.06×10^{-2}	81。3%	0.79	2.29×10^{-3}
	$K = 100$	5.77×10^{-2}	70。0%	1.02	3.30×10^{-3}
Non-分层	$K = 5$	4.79×10^{-1}	26.7%	36.89	2.25×10^{-1}
	$K = 10$	2.55×10^{-2}	83。3%	1.52	2.42×10^{-3}
	$K = 20$	6.02×10^{-2}	86.7%	1.44	5.41×10^{-3}
	$K = 30$	9.56×10^{-2}	73。3%	1.25	9.76×10^{-3}
	$K = 50$	6.74×10^{-2}	70。0%	0.99	3.68×10^{-3}
	$K = 100$	1.01×10^{-1}	55.7%	1.51	5.13×10^{-3}

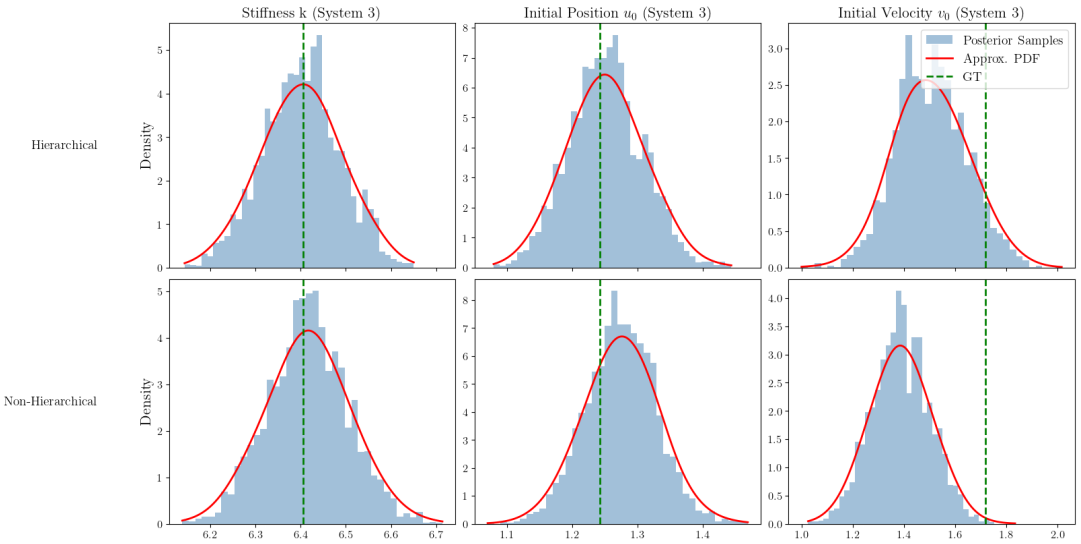


图8：**使用基于 PINN 的代理模型，对比实验1中分层与非分层贝叶斯框架下的后验参数分布$\mathbf{K = 20}$。** 该图比较了系统3中三个参数的不确定性量化性能。**上排（分层）**和**下排（非分层）**显示了不同的后验密度。真实值以绿色虚线表示，并叠加在后验样本（蓝色直方图）及其 PDF 近似值（红色曲线）上。

图 8所示的直方图中，层次模型的后验标准差在视觉上似乎仅略有降低，其估计不确定性与非层次替代方案的十分相似。然而，表 2中的定量结果显示，层次设置在准确性方面具有显著优势。此外，训练期间的观测结果表明，非层次模型的收敛速度明显慢于层次方法，这一点可从图 9中的 MALA 轨迹图看出。

此外，分层表述还能估计总体层超参数 $\boldsymbol{\phi}$，而这些参数在非分层设置中无法获得。图10 展示了推断的超先验分量的 MALA 轨迹。推断得到的超先验均值为 $\boldsymbol{\mu}_\phi = \{5.08, -0.36, 1.70\}$ 和 $\boldsymbol{\tau}_\phi = \{0.030, 1.64, 2.08\}$，真实值则为 $\boldsymbol{\mu}_\phi = \{5.0, 0.0, 2.0\}$ 和 $\boldsymbol{\tau}_\phi = \{0.03, 2.0, 2.0\}$。重要的是，这些学习得到的总体统计量可作为信息丰富的数据驱动先验，用于对新系统进行后续推断，从而无需依赖任意的先验选择，并提升

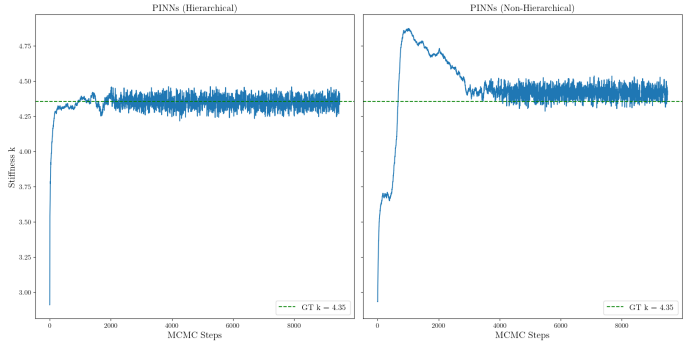


图9：**实验1中分层和非分层贝叶斯框架下系统 9 的刚度 k 的后验样本 MALA 链**。图中比较了分层模型与非分层模型之间的收敛率。

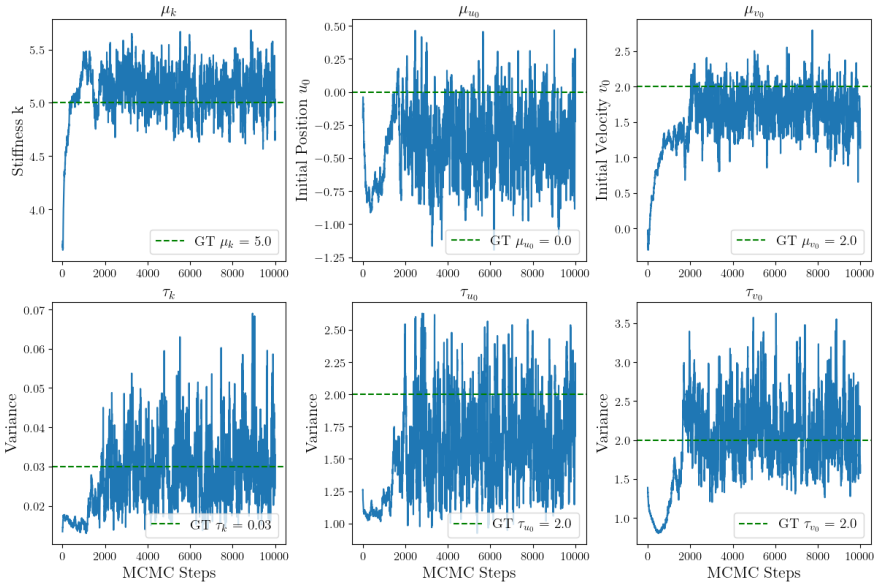


图10：**实验1中使用基于 PINN 的分层贝叶斯框架推断超参数的后验样本 MALA 链**。各面板展示了 $\boldsymbol{\mu}_\phi$ 和 $\boldsymbol{\tau}_\phi$的 MALA 链。顶行对应超先验均值（μ_k、 μ_{u_0}、 μ_{v_0}），底行展示方差（τ_k、 τ_{u_0}、 τ_{v_0}）。绿色虚线表示真实值。

表3：**实验1中不同系统数量（$\mathbf{K} \in \{5, 10, 20, 30, 50, 100\}$）下各模型的计算效率比较**。 所报告的数值表示在双层框架内完成一个完整训练轮次所需的平均时间（以秒为单位）。

样本	求解器	FNO（监督式）	FNO（基于物理的）	物理信息神经网络	物理信息神经网络（非分层）
$K = 5$	0.140秒	0.549秒	0.099秒	0.038秒	0.040秒
$K = 10$	0.171秒	0.615秒	0.104秒	0.038秒	0.042秒
$K = 20$	0.201秒	0.712秒	0.124秒	0.040秒	0.045秒
$K = 30$	0.238秒	0.763秒	0.157秒	0.041秒	0.046秒
$K = 50$	-	0.842秒	0.232秒	0.048秒	0.058秒
$K = 100$	-	1.158秒	0.489秒	0.091秒	0.091秒

整体推理鲁棒性。

不同算法的效率

最后，我们评估不同框架的计算效率。表3报告了平均

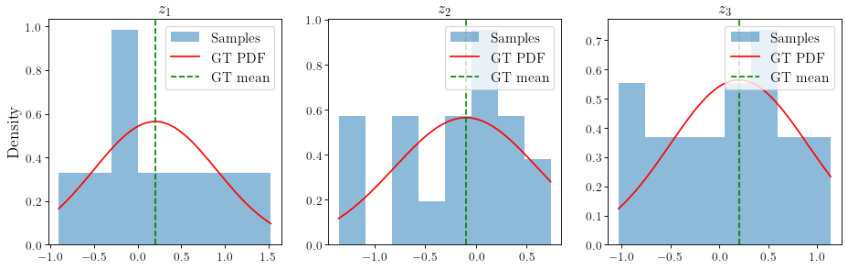


图11：**实验2中20个达西流系统的真实物理参数分布**。每个系统的参数向量 $\boldsymbol{\theta} = \{z_1, z_2, z_3\}$ 均从超参数为 $\boldsymbol{\mu}_\phi = \{0.2, -0.1, 0.2\}$ 和 $\boldsymbol{\tau}_\phi = \{0.5, 0.5, 0.5\}$的层次先验中采样。

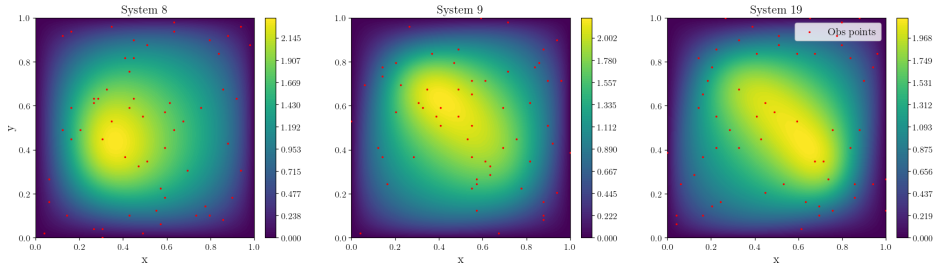


图12：**使用不动点数值求解器生成的代表性达西流场真值**。图中展示了数据集中的 $K = 20$ 个系统里选出的三个系统（系统8、9和19）的流场。红色标记表示观测数据的空间位置。

完成所有模型一个完整训练轮次所需的时间。对于使用数值求解器作为前向映射的基线模型，报告的计算时间包括更新闭合模型参数和执行集成MALA步骤所需的时间；对于其他方法，记录的时间则包括整个双层迭代过程，包括闭合更新、MALA推断步骤以及代理模型的训练步骤。

尽管监督式傅里叶神经算子在此前的指标中表现出更高的精度，但其计算成本最高。这是因为在线训练期间计算监督损失时，需要依赖昂贵的数值求解器来生成所需的真实参考轨迹。基于物理的傅里叶神经算子对此有所改进，其运行时间低于采用数值求解器的基线方法。相比之下，物理信息神经网络是最高效的架构，相较于求解器显著加快了运行速度。尤为重要的是，物理信息神经网络在系统数量方面展现出更强的可扩展性K。如表3所示，随着K增加，求解器和傅里叶神经算子模型的运行时间显著增加，而物理信息神经网络的计算成本基本保持不变。这表明，物理信息神经网络尤其适用于涉及大规模总体的高维分层推断任务，能够在保持计算可行性的同时，不会过度牺牲速度。

4.2 非线性达西流

接下来，我们考虑一个二维非线性达西流问题。在此设定中，空间变化的渗透率场由两部分构成：一项为空间项，由投影到已知基函数上的三个未知参数定义；另一项为依赖于解场的未知非线性闭合模型。

4.2.1 问题表述

$\Omega = [0,1]^2$ 上的二维非线性达西流形式为

$$
\begin{aligned}
-\nabla(a(u,\mathbf{x})\nabla u) &= s(\mathbf{x}), \qquad \mathbf{x} \in \Omega, \\
u(\mathbf{x}) &= 0, \qquad \mathbf{x} \in \partial\Omega,
\end{aligned} \tag{33}
$$

其中 $s(\mathbf{x}) = 2\pi^2 \sin(\pi x)\sin(\pi y)$， $u(\mathbf{x})$表示解场， $a(u,\mathbf{x})$是非线性、空间变化的渗透率场，定义为

$$
a(u,\mathbf{x}) = \exp(z(\mathbf{x}))\,\boldsymbol{\sigma}(f(u)), \tag{34}
$$

其中 $\boldsymbol{\sigma}(\cdot)$是Sigmoid函数，未知非线性闭合模型由 $f(u) = u^2/2$给出。空间项 $z(\mathbf{x})$采用三项基展开表示

$$
z(\mathbf{x}) = \sum_{i=1}^{3} z_i \phi_i(\mathbf{x}), \tag{35}
$$

with

$$
\phi_i(\mathbf{x}) = \sin(m_i \pi x)\sin(n_i \pi y), \tag{36}
$$

以及 $\{m_1, m_2, m_3\} = \{2, 2, 1\}$ 和 $\{n_1, n_2, n_3\} = \{2, 1, 2\}$。

对 $z(\mathbf{x})$ 施加指数变换，以保证渗透率场为正。系数 $\boldsymbol{\theta} = \{z_1, z_2, z_3\}$ 是待推断的未知参数。与上一实验相同，我们对这些参数施加层次先验，其真实总体超参数为$\boldsymbol{\mu}_\phi = \{0.2, -0.1, 0.2\}$ 和 $\boldsymbol{\tau}_\phi = \{0.5, 0.5, 0.5\}$。生成的参数 $\boldsymbol{\theta}$ 来自真实超先验，针对$K = 20$ 个系统如图11所示。

4.2.2 求解器

为求解（33）中的非线性达西流问题，我们采用有限元法（FEM） [60]。通过推导弱形式，并将问题投影到由区域三角剖分定义的有限维子空间上，可得以下非线性代数方程组：

$$
\mathbf{A}(\mathbf{u})\mathbf{u} = \mathbf{b}, \tag{37}
$$

其中，$\mathbf{u}$ 是未知节点系数向量，$\mathbf{b}$是载荷向量，而$\mathbf{A}(\mathbf{u})$ 是刚度矩阵。关键在于，由于渗透率场 $a(u,\mathbf{x})$ 取决于解，刚度矩阵也依赖于状态，因此该系统是非线性的。弱形式的详细推导以及矩阵元素的明确表达式见附录B。

为处理这种非线性，我们采用不动点迭代求解器。从初始猜测$\mathbf{u}_0$, 出发，该方法通过迭代求解以下关于$\mathbf{u}^{k+1}$的线性化系统，

$$
\mathbf{A}(\mathbf{u}_k)\mathbf{u}_{k+1} = \mathbf{b}. \tag{38}
$$

$\mathbf{A}(\mathbf{u}_k)$ 使用上一$\mathbf{u}_k$ 次迭代得到的已知解构造，因此该系统是线性的。重复此过程，直至残差低于预设容差。

三　图中展示了使用不动点求解器生成的 20 个系统中的1个系统　2

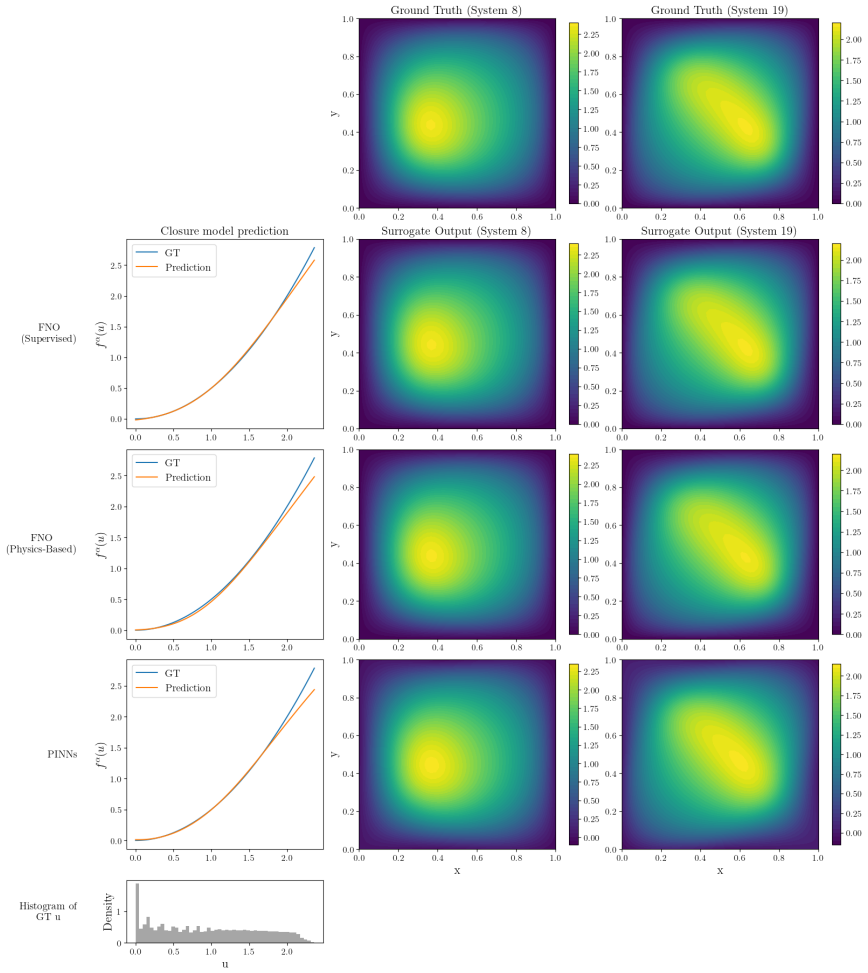


图 13：**实验2中不同模型在 $\mathbf{K = 20}$ 下的闭合模型估计与代理性能比较**。各行分别对应由不动点求解器生成的真值场，以及监督式傅里叶神经算子、基于物理的傅里叶神经算子和物理信息神经网络的结果。**左列：**学习得到的闭合模型 $f^{\alpha}(u)$（橙色）与真值 $f(u)$（蓝色）的比较。**中间**和**右列：**两个系统（系统8和系统19）的代理预测。**底部面板：**真值数据集中 u 的直方图。

并使用 JAXOpt 提供的不动点求解器软件包[1] 来生成数据并进行后续训练。此外，从每个系统中选取 60 个观测点，并加入观测噪声 $\sigma = 0.1$。在本实验中，每个系统的观测算子 $g^{(k)}$ 各不相同。

4.2.3 代理模型实现细节

在本实验中，由于 MALA 框架中的迭代不动点求解器计算成本过高，因此省略了采用数值求解器作为正向模型的基准方法。关于计算效率的详细讨论见第 4.2.4 节。

我们采用与第一个实验相同的三个代理模型架构，不同之处在于，为了适用于这个稳态偏微分方程问题，输入现在是空间坐标 x 和 y，而不是时间 t。对于代理模型的在线训练，监督式傅里叶神经算子的损失函数是代理预测与不动点求解器所得参考解之间的差异。

[1] https://jaxopt.github.io/stable/_autosummary/jaxopt.FixedPointIteration.html

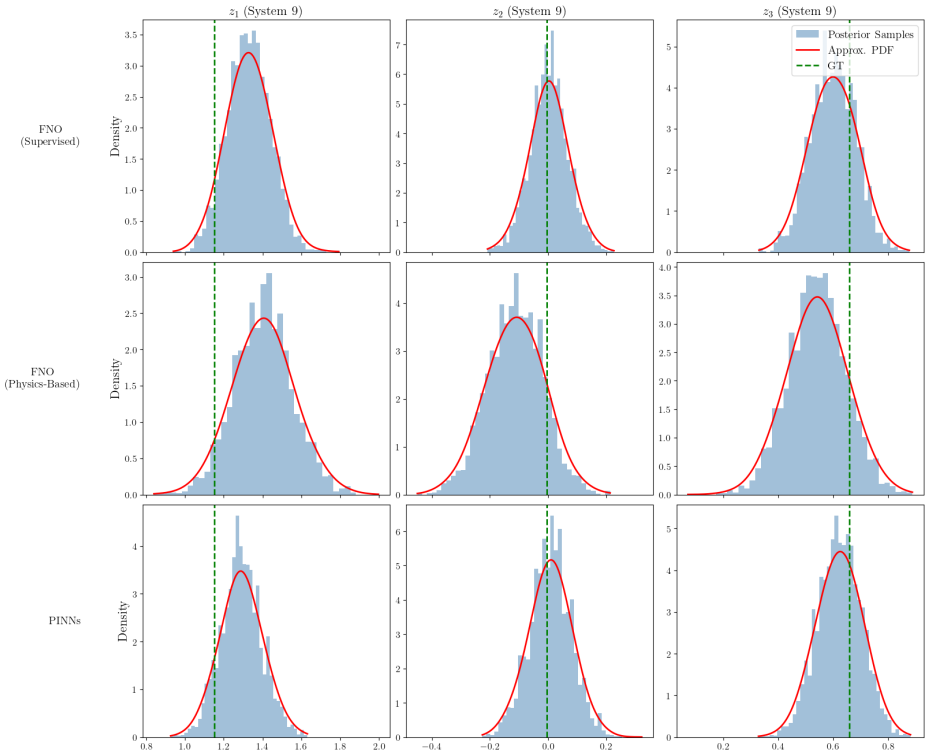


图14：**实验2中，系统 9 的三个参数在 K = 20 下、三种模型对应的后验直方图及近似后验密度函数**。每个面板展示一条随机 MALA 链收敛后最后 1,000 个样本得到的边际后验分布。

对于基于物理的傅里叶神经算子，我们采用式（28）中定义的弱形式残差损失，其中

$$\mathcal{R} = \mathbf{A}(\mathbf{u})\mathbf{u} - \mathbf{b}, \tag{39}$$

与（37）中的离散系统一致。最后，利用自动微分直接根据强形式方程（33）计算 PINN 损失。更多实现细节见附录 A。

4.2.4 结果

与第一次实验一样，我们的评估遵循相同的三部分结构：比较四种模型的性能，对比层次贝叶斯设定与非层次贝叶斯设定，以及分析计算效率。

逆问题和正问题的性能比较

为评估闭合模型和代理模型的性能，图 13 可视化了全部三种架构在 $K = 20$ 处的估计闭合模型和代理模型输出。表4详细给出了针对不同系统数量的相应定量分析。监督式傅里叶神经算子始终优于物理信息模型，尤其是其代理均方误差通常比其他方法低一个数量级，并伴随高度准确的闭合项重建。相反，基于物理的傅里叶神经算子和物理信息神经网络均表现出更高的均方误差，表明在此二维设置中仅依赖物理信息残差来训练代理模型不如直接监督稳定。这种性能下降在 $K = 10$ 处最为严重，此时观测到的系统数量有限，可能无法为基于物理的优化收敛提供足够约束。此外，与傅里叶神经算子不同

表 4： **实验2中不同系统数量（K ∈ {10, 20, 30}）下三个模型的定量比较。** 该表评估了**参数推断、闭合均方误差和代理均方误差**，所使用的指标与表1中的相同。闭合均方误差在范围 $u \in [0, 2.5]$上进行评估。

模型	系统数量 系统	参数推断		闭合均方误差	代理均方误差
		平均均方误差	覆盖率		
FNO（监督式）	$K = 10$	3.39×10^{-2}	93.3%	5.01×10^{-3}	8.92×10^{-4}
	$K = 20$	1.43×10^{-2}	93.3%	3.95×10^{-3}	3.48×10^{-5}
	$K = 30$	2.37×10^{-2}	90。0%	3.28×10^{-3}	3.57×10^{-5}
FNO（基于物理的）	$K = 10$	1.17×10^{-1}	56.7%	4.28×10^{-2}	7.20×10^{-3}
	$K = 20$	2.47×10^{-2}	91.7%	1.49×10^{-2}	6.35×10^{-5}
	$K = 30$	3.00×10^{-2}	88。9%	9.27×10^{-3}	8.52×10^{-4}
物理信息神经网络	$K = 10$	4.89×10^{-2}	83。3%	2.62×10^{-1}	1.09×10^{-3}
	$K = 20$	1.53×10^{-2}	93.3%	8.31×10^{-3}	1.56×10^{-4}
	$K = 30$	2.56×10^{-2}	90。0%	1.11×10^{-2}	9.33×10^{-4}

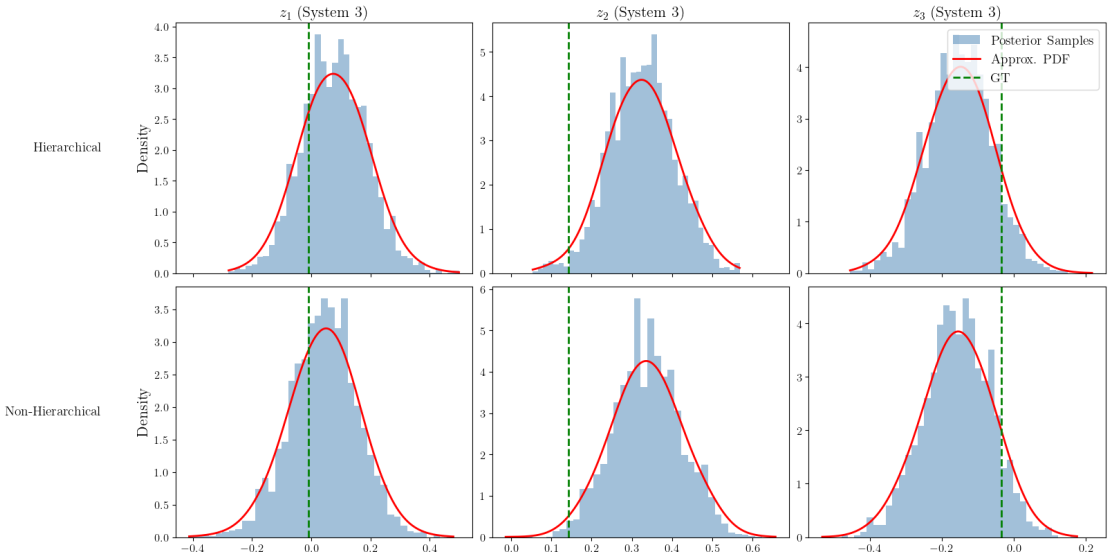


图15： **在实验2中，使用 K = 20 的基于 PINN 的代理模型比较分层与非分层贝叶斯框架的后验参数分布。** 该图比较了两种层次化设置下，系统3中三个参数的不确定性量化性能。

表5： **实验2中使用物理信息神经网络作为代理模型训练的分层与非分层方法比较。** 报告的指标与表1相同。

方法	数量 系统	参数推断		闭合均方误差	代理均方误差
		平均均方误差	覆盖率		
分层	$K = 10$	4.89×10^{-2}	83.3%	2.62×10^{-1}	1.09×10^{-3}
	$K = 20$	1.53×10^{-2}	93.3%	8.31×10^{-3}	1.56×10^{-4}
	$K = 30$	2.56×10^{-2}	90。0%	1.11×10^{-2}	9.33×10^{-4}
Non-分层	$K = 10$	6.21×10^{-2}	80.0%	5.33×10^{-1}	1.69×10^{-3}
	$K = 20$	1.46×10^{-2}	93.3%	9.62×10^{-3}	1.79×10^{-4}
	$K = 30$	3.73×10^{-2}	82。2%	1.52×10^{-2}	1.69×10^{-3}

架构能够自然编码边界结构，物理信息神经网络则如（29）所述将边界条件作为软约束施加。如图 13 所示，这会导致代理模型输出在区域边界处出现误差。

关于 ding参数推断，图 14 展示了最终 1,000 个后验样本的直方图 fo r

表6：**实验2中不同系统数量（$\mathbf{K} \in \{10, 20, 30\}$）下所考虑模型的计算效率比较。**

样本	求解器	FNO（监督式）	FNO（基于物理的）	物理信息神经网络	物理信息神经网络（非分层）
$K = 10$	$4.158s$	$3.121s$	$0.686s$	$0.132s$	$0.134s$
$K = 20$	$10.374s$	$3.837s$	$1.119s$	$0.222s$	$0.223s$
$K = 30$	$17.962s$	$4.962s$	$1.934s$	$0.378s$	$0.381s$

系统 9 中的全部三个参数。同样，参数推断的准确性与覆盖率的定量分析见表 4。结果清楚地表明，监督式傅里叶神经算子是最准确且最稳健的方法。其优势在有限数据场景中尤为关键（$K = 10$）。在此设定下，纯粹的物理信息方法无法可靠地识别参数，表现出较高的平均均方误差和较差的覆盖概率（56.7% 和 83.3%）。相比之下，监督式傅里叶神经算子在该场景中保持了较高的准确性和可靠的不确定性量化，实现了较低的均方误差以及 93.3%。

分层与非分层贝叶斯 **es**

同时也对层次贝叶斯表述与非层次贝叶斯表述进行了比较。同样，物理信息神经网络被用作代理架构。图15 展示了系统3中三个参数的后验分布，而表 5 则详细给出了定量指标。

从视觉上看，图 15 表明非层次模型会产生更宽的后验分布，反映出更高的不确定性。相比之下，层次框架对参数估计表现出更高的置信度。通过将各个参数条件于共享总体分布，该模型利用组级信息有效约束了特定系统的方差。在定量方面，表 5中，层次设置在不同系统数量 K下，于逆推断和代理模型训练两方面均持续优于非层次表述。

当 $K = 20$ 为 $\boldsymbol{\mu}_\phi = \{0.37, -0.09, -0.01\}$且 $\boldsymbol{\tau}_\phi = \{0.35, 0.29, 0.28\}$时，由层次设置推断得到的超先验信息¯ 真值分别为 $\boldsymbol{\mu}_\phi = \{0.2, -0.1, 0.2\}$ 和 $\boldsymbol{\mu}_\phi = \{0.5, 0.5, 0.5\}$。在超先验方差 $\boldsymbol{\tau}_\phi$中观察到明显偏差。这一差异在统计上可归因于有限的样本量（$K = 20$），因为方差的准确估计本质上需要更多被观测的系统。

不同算法的效率

最后，我们评估所提出框架的计算效率，结果汇总于表6中。与实验1中的方法论一致，所报告的运行时间衡量的是完成一个训练轮次的平均时长，其中包括逆推断步骤（集成MALA和闭合更新）以及代理模型训练步骤。值得注意的是，在该非线性偏微分方程问题中，使用传统数值求解器进行正向评估的基线方法表现出显著的计算成本（例如，在17.96秒的每轮训练耗时， $K = 30$）。

相比之下，所有基于代理的方法都显著提升了计算速度。物理信息神经网络再次展现出最高的计算效率，运行时间短且可扩展性出色。然而，考虑到本实验中纯基于物理的方法表现欠佳，监督式傅里叶神经算子提供了最均衡的解决方案。尽管其计算需求高于物理信息神经网络，但与数值求解器相比，仍能显著缩短运行时间，同时达到稳健开展参数与闭合发现所需的更高精度。

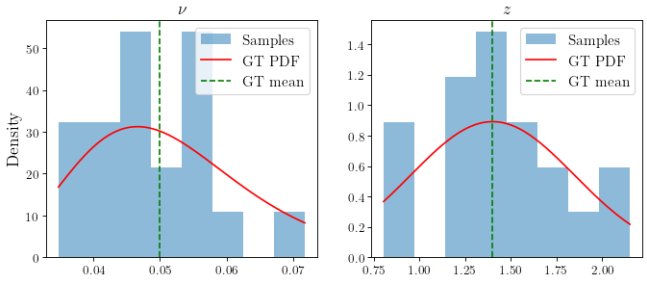


图16：**实验 3 中 20 个 Burgers系统的真实物理参数分布**。每个系统的参数向量 $\boldsymbol{\theta} = \{\log(\nu)\ z\}$ 均从超参数为 $\boldsymbol{\mu}_\phi = \{\log(0.05)\ 1.4\}$ 和 $\boldsymbol{\tau}_\phi = \{0.07, 0.2\}$ 的层次先验中采样。

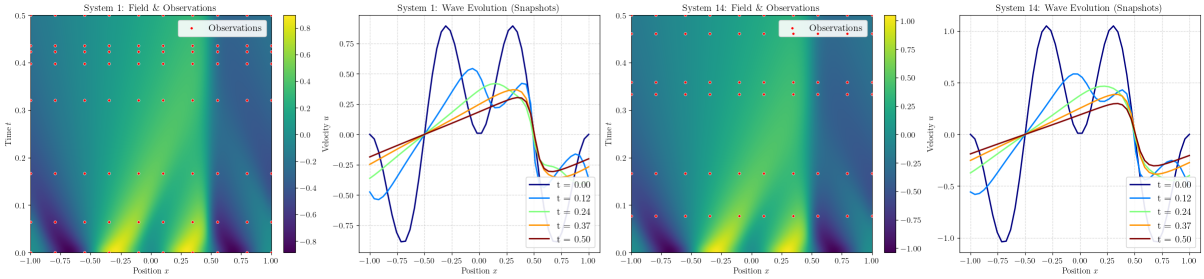


图17：**使用数值求解器生成的代表性真实 Burgers场**。这些图展示了从包含 $K = 20$ 个系统的数据集中选出的两个系统（系统 1 和 14）的场。红色标记表示观测数据的空间位置。

4.3 广义伯格斯方程

最后一个实验评估广义伯格斯方程。在此设置中，需要推断的未知参数为黏度和初始条件振幅，而对流项则被视为未知闭合项。

4.3.1 问题表述

广义伯格斯方程的形式为

$$\begin{aligned} \frac{\partial u}{\partial t} + f(u)\frac{\partial u}{\partial x} &= \nu \frac{\partial^2 u}{\partial x^2}, \quad x \in [-1, 1],\ t \in [0, 0.5], \\ u(x, 0) &= z \sin(2\pi x) \sin(\pi x), \\ u(-1, t) &= u(1, t), \end{aligned} \tag{40}$$

其中采用周期边界条件。 $u(x, t)$ 表示速度解场。 $f(u)$ 是 u 的非线性函数，其形式为 $f(u) = 7(\sigma(3u) - 0.5)$，其中 $\sigma(\cdot)$ 是 Sigmoid函数。在此情形下，黏度 ν 和初始条件中的 z 都是未知参数。由于 ν 必须为正，因此待推断的参数设为 $\boldsymbol{\theta} = \{\log(\nu)\ z\}$。总体的真实超先验为 $\boldsymbol{\mu}_\phi = \{\log(0.05)\ 1.4\}$ 和 $\boldsymbol{\tau}_\phi = \{0.07, 0.2\}$。生成的 $K = 20$ 个系统的参数如图16所示。

4.3.2 求解器

为数值求解（40），我们采用显式四阶 Runge–Kutta（RK4）方法进行时间积分，时间步长为 $\Delta t = 0.00125s$，空间分辨率为 $\Delta x = 0.04$。为减少存储

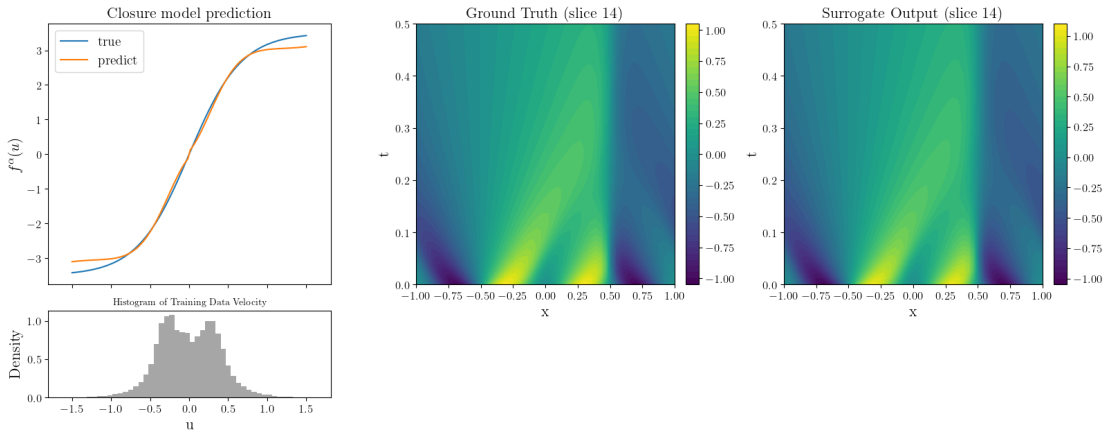


图 18：**实验 3 中，监督式傅里叶神经算子在 K = 20 条件下的闭合模型估计与代理性能。左：** 提供真实速度数据直方图时的闭合模型估计。**中间 & 右：** 系统 14 解场的真值与代理预测。

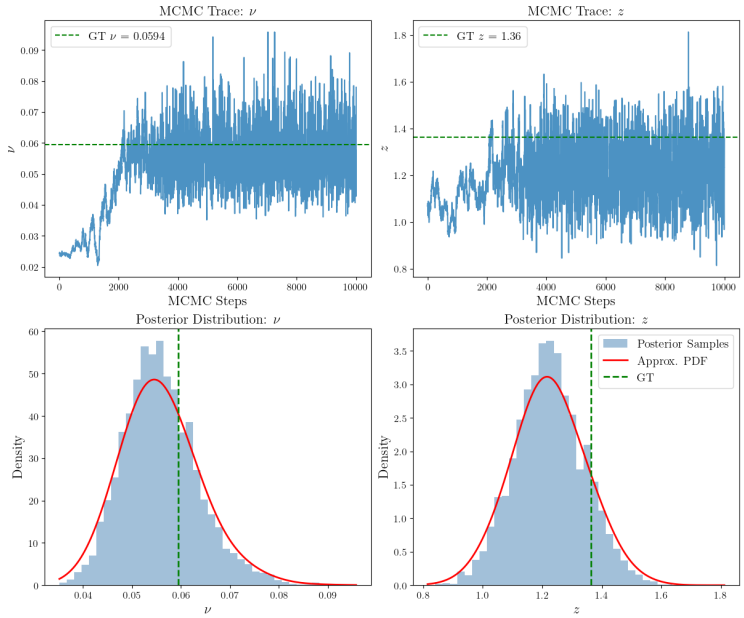


图 19：**实验 3 中系统 14 参数的 MALA 轨迹与后验直方图。上排：** 参数 ν 和 z的 MALA 轨迹。**下排：** MALA 轨迹最后 1,000 个样本的后验直方图及其对应的概率密度函数近似。

开销，我们每隔 10 个时间步记录一次数据，对解进行降采样，从而在整个时间域内得到共 40 个离散快照。值得注意的是，模拟此类时变系统所需的严格数值设置，凸显了使用代理模型进行前向传播的一项主要优势。显式数值求解器需要采用极小的时间步长以保持数值稳定性，因此必须进行大量的连续迭代。在逆向推断阶段，为计算梯度而对这些冗长的求解器迭代进行反向传播，计算成本高得难以承受。相比之下，代理模型无需密集的迭代时间步进，大幅加快了逆问题中的梯度计算。

在本实验中，也采用系统特定的观测算子 $g^{(k)}$ 。测量值记录在固定的空间位置，但时间步长是随机抽取的，每个系统的时间观测总数在 5 到 10 之间。图17展示了解场、波的

演化过程，以及系统 1 和 14 对应的观测位置。所有测量值均受到 $\sigma = 0.2$ 水平的观测噪声影响。

4.3.3 结果

本案例仅对监督式训练的傅里叶神经算子进行了实现，适用于 $K = 20$ 个系统，更多实现细节见附录 A。

估计得到的闭合模型和代理模型输出见图18。在 $u \in [-1.5, 1.5]$ 处评估的闭合模型估计均方误差为 2.90×10^{-2}，所有系统的代理模型输出均方误差为 1.49×10^{-4}。一个选定的系统 14 的参数推断结果见图 19，图中显示了两个未知参数的 MALA 轨迹和后验直方图。在 $K = 20$ 个系统中，估计后验均值与真实参数之间的均方误差为 1.64×10^{-2}，估计后验分布在 2 个标准差范围内的覆盖率为 95%。超先验后验样本的均值为 $\boldsymbol{\mu}_\phi = \{\log(0.051)\ 1.42\}$ 和 $\boldsymbol{\tau}_\phi = \{0.029, 0.176\}$，与 $\boldsymbol{\mu}_\phi = \{\log(0.05)\ 1.4\}$ 和 $\boldsymbol{\tau}_\phi = \{0.07, 0.2\}$ 相比。

5 结论与未来工作

本文提出了一个全面的双层优化框架，用于同时求解微分方程的正问题与反问题。我们的方法专门针对由部分已知物理定律所支配的情形，从稀疏观测值中推断未知标量参数和非线性闭合项。该方法将推断任务解耦，结合用于概率参数估计的分层贝叶斯框架，以及针对未知动力学的确定性最大似然闭合学习。通过采用带协方差预条件的集成MALA，我们能够高效地从后验中采样，并利用集成链生成的样本促进准确的闭合学习。

为降低重复正向计算的成本，我们在优化循环中集成了联合训练的代理模型。针对常微分方程和偏微分方程系统开展的实验结果验证了所提框架的有效性，并揭示了不同代理架构之间的关键权衡。在低维常微分方程场景中，物理信息神经网络兼顾了准确性与效率，表现出较强的竞争力。对于非线性达西流和广义伯格斯方程等复杂问题，代理建模对于确保计算可行性至关重要。在此情形下，与基于物理的方法相比，监督式傅里叶神经算子表现出更强的稳健性和稳定性，尤其是在数据有限的场景中。

总之，本研究通过有效结合概率推断、神经闭合学习和代理模型训练，建立了一种灵活的途径，可基于稀疏观测进行参数推断和闭合学习。未来的研究方向之一是引入卡尔曼滤波等技术，以促进动力系统中的在线联合状态与参数推断。

A 实验细节

所有实验均使用 JAX 实现，并在 NVIDIA GeForce RTX 3090 GPU 上运行。本节将详细介绍所提出框架的超参数设置和架构规格。下文列出了所有实验中神经网络架构、MALA 采样器及优化方案的实现细节。

我们采用三种不同的网络架构：用于代理模型的傅里叶神经算子和物理信息神经网络 F^β（监督式训练和基于物理的训练采用相同的架构），以及用于闭合模型的多层感知机f^α。所有模型均使用 Adam 优化器进行训练 [61]。表 7 和 8 概述了网络结构

和实验1的优化超参数；表 9 和 10 给出了实验2的相应设置，表 11 和 12 则对应实验 3。

表 7：实验1中代理模型和闭合模型的网络架构。

模型	超参数	规格
FNO	提升维度（宽度）	32
	模态数量	16
	FNO 层数	4
	输入维度	4
	输出维度	2
	激活函数	SiLU
物理信息神经网络	隐藏层	5
	每层隐藏神经元数	64
	输入维度	4
	输出维度	1
	激活函数	Tanh
多层感知机（闭合模型）	隐藏层	4
	每层隐藏神经元数	64
	输入维度	1
	输出维度	1
	激活函数	SiLU

表8：实验1中优化和训练的超参数设置。

超参数	FNO	物理信息神经网络
MALA链数量（M）	200	200
MALA步长（γ）	0.010	0.018
低层迭代次数（N）	20	20
闭合学习率（η_α）	5×10^{-4}	1×10^{-3}
代理学习率（η_β）	5×10^{-4}	1×10^{-3}

表9：实验2中代理模型和闭合模型的网络架构。

模型	超参数	规格
FNO	提升维度（宽度）	32
	模态数量	16
	FNO 层数	4
	输入维度	5
	输出维度	1
	激活函数	SiLU
物理信息神经网络	隐藏层	5
	每层隐藏神经元数	64
	输入维度	5
	输出维度	1
	激活函数	Tanh
多层感知机（闭合模型）	隐藏层	4
	每层隐藏神经元数	64
	输入维度	1
	输出维度	1
	激活函数	SiLU

表10：实验2中优化和训练的超参数设置。

超参数	FNO	物理信息神经网络
MALA链数量（M）	160	160
MALA步长（γ）	0.018	0.020
低层迭代次数（N）	18	20
代理学习率（η_β）	5×10^{-4}	5×10^{-4}
闭合学习率（η_α）	5×10^{-4}	5×10^{-4}

表11：实验 3 中代理模型和闭合模型的网络架构。

模型	超参数	规格
FNO	提升维度（宽度）	32
	模态数量	16
	FNO 层数	4
	输入维度	4
	输出维度	1
	激活函数	SiLU
多层感知机（闭合模型）	隐藏层	2
	每层隐藏神经元数	64
	输入维度	1
	输出维度	1
	激活函数	SiLU

表12：实验 3 中优化和训练的超参数设置。

超参数	FNO
MALA 链数量 (M)	120
MALA步长 (γ)	0.018
低层迭代次数 (N)	18
代理学习率 (η_β)	5×10^{-4}
闭合学习率 (η_α)	5×10^{-4}

B 弱形式与有限元法离散化

本节将详细介绍第4.2.2节中非线性达西流问题的弱形式及离散化。

我们首先推导式（33）中偏微分方程的弱形式。假设满足齐次狄利克雷边界条件，我们将试验与检验函数空间定义为 $V = H_0^1(\Omega)$。将控制方程乘以任意检验函数 $v \in V$ 并在区域Ω 上积分，得到

$$\int_\Omega -\nabla \cdot (a(u,\mathbf{x})\nabla u)v \, d\mathbf{x} = \int_\Omega s(\mathbf{x})v \, d\mathbf{x}. \tag{41}$$

对左侧项应用分部积分后，我们得到

$$\int_\Omega a(u,\mathbf{x})\nabla u \cdot \nabla v \, d\mathbf{x} - \int_{\partial\Omega} a(u,\mathbf{x})(\nabla u \cdot \mathbf{n})v \, ds = \int_\Omega s(\mathbf{x})v \, d\mathbf{x}. \tag{42}$$

由于 $v \in H_0^1(\Omega)$ 在边界 $\partial\Omega$ 上为零，边界积分也为零。因此，弱形式表述如下：求满足以下条件的 $u \in V$

$$\int_\Omega a(u,\mathbf{x})\nabla u \cdot \nabla v \; d\mathbf{x} = \int_\Omega s(\mathbf{x})v \; d\mathbf{x}, \quad \forall v \in V. \tag{43}$$

为了对解进行数值近似，我们采用有限元法 [60]，将问题限制在有限维子空间 $V_h \subset V$ 上。我们对区域 Ω 进行三角剖分，并定义一组基函数 $\{\psi_j\}_{j=1}^{Q}$，其中 Q 表示自由度总数。离散解$u_h \in V_h$ 表示为这些基函数的线性组合，

$$u_h(\mathbf{x}) = \sum_{j=1}^{Q} u_j \psi_j(\mathbf{x}), \tag{44}$$

以及$\mathbf{u} = [u_1, u_2, \ldots, u_Q]^T$表示未知节点系数向量。将此展开式代入弱形式（43），并应用伽辽金法（即，$v = \psi_i$对 $i = 1$，...，Q），得到以下非线性方程组，

$$\sum_{j=1}^{Q} u_j \left(\int_\Omega a(u_h,\mathbf{x})\nabla\psi_j \cdot \nabla\psi_i \, d\mathbf{x} \right) = \int_\Omega s(\mathbf{x})\psi_i \, d\mathbf{x}, \quad \text{for } i = 1, \ldots, Q. \tag{45}$$

该方程组可用紧凑的矩阵记号表示为

$$\mathbf{A}(\mathbf{u})\mathbf{u} = \mathbf{b}, \tag{46}$$

其中，$\mathbf{b} \in \mathbb{R}^N$是载荷向量，其分量为$b_i = \int_\Omega s(\mathbf{x})\psi_i \, d\mathbf{x}$，而$\mathbf{A}(\mathbf{u}) \in \mathbb{R}^{N\times N}$是刚度矩阵，其各元素取决于解$\mathbf{u}$，

$$\mathbf{A}_{ij}(\mathbf{u}) = \int_\Omega a(u_h,\mathbf{x})\nabla\psi_j \cdot \nabla\psi_i \, d\mathbf{x}. \tag{47}$$

致谢

我们感谢 Tim J. Rogers 就系统辨识展开的富有见地的讨论。AV 获得英国工程与自然科学研究委员会 ROSEHIPS 项目资助 [EP/W005816/1]。AGD 获得英国工程与自然科学研究委员会 UbOHT 计划资助 [EP/X037770/1]。MG 获得英国皇家工程院 Re-

search Chair 以及英国工程与自然科学研究委员会的资助 [EP/X037770/1, EP/Y028805/1, EP/W005816/1, EP/V056441/1 和 EP/V056522/1]。

代码可用性

本研究使用的源代码可在 GitHub 仓库中获取，
https://github.com/ypzpy/Hierarchical-Inference-and-Closure-Learning.git。

参考文献

[1] S. P. Timoshenko, J. N. Goodier, 弹性理论, 麦格劳-希尔, 纽约, 1970.

[2] G. K. Batchelor, 流体动力学导论, 剑桥大学出版社, 剑桥, 2000.[3] H. S. Carslaw, J. C. Jaeger, 固体中的热传导, 牛津大学出版社, 牛津, 1959.[4] J. D. Jackson, 经典电动力学, 第3版, 威利, 纽约, 1999.[5] J. D. Murray, 数学生物学 I：导论, 施普林格, 纽约, 2002.[6] G.K. Vallis, 大气与海洋流体动力学, 第2版, 剑桥大学出版社, 剑桥, 2017.

[7] S. L. Brunton, J. L. Proctor, J.N. Kutz, 通过非线性动力系统稀疏辨识从数据中发现控制方程, 《美国国家科学院院刊》113(15) (2016) 3932–3937。

[8] S. H. Rudy, S. L. Brunton, J. L. Proctor, J.N. Kutz, 基于数据发现偏微分方程, 《科学进展》3(4) (2017) e1602614。

[9] D.A. Messenger, D. M. Bortz, 用于偏微分方程的弱形式 SINDy, 《计算物理学杂志》443 (2021) 110525。

[10] M.D.Champneys, T. J. Rogers, BINDy：利用可逆跳跃马尔可夫链蒙特卡罗进行非线性动力学的贝叶斯辨识, 《英国皇家学会会刊 A：数学、物理与工程科学》481 (2319) (2025)。[11] K. Duraisamy, G. Iaccarino, H. Xiao, 数据时代的湍流建模, 《流体力学年评》51 (1) (2019) 357–377。[12] S. Safari, J. L. Monsalve, 非线性组合结构的数据驱动结构辨识：带螺栓连接的结构, 《机械系统与信号处理》195 (2023) 110296。[13] J. Sirignano, J. F. MacArt, 用于钝体绕流大涡模拟的深度学习闭合模型, 《流体力学杂志》966 (2023) A26。[14] T. Rogers, K. Worden, E. Cross, 关于高斯过程潜在力模型在输入—状态—参数联合估计中的应用：兼论贝叶斯运行状态辨识, 《机械系统与信号处理》140 (2020) 106580。[15] T. J. Rogers, T. Friis, 一种用于非线性系统辨识的潜在恢复力方法, 《机械系统与信号处理》180

[16] C. E. Rasmussen，《机器学习暑期学校》中的机器学习高斯过程，施普林格，2003年，第63–71页。[17] W. M. Kouw、C. Gruijthuijsen、L. Blanken、E. Evers、T. Rogers，传热动力学中非线性对流效应的贝叶斯灰箱辨识，载于：2024年IEEE控制技术与应用会议（CCTA），IEEE，2024年，第382–387页。[18] M. Raissi、P. Perdikaris、G. E. Karniadakis，物理信息神经网络：一种用于求解涉及非线性偏微分方程的正问题与反问题的深度学习框架，《计算物理学杂志》378（2019）686–707。[19] S. Arridge、P. Maass、O. Öktem、C.-B. Schönlieb，使用数据驱动模型求解逆问题，《数值学报》28（2019）1–174。[20] C. Aarset、M. Holler、T. T. N. Nguyen，非线性时变偏微分方程中的学习辅助参数辨识，《应用数学与优化》88（3）（2023）76。[21] A. M. Stuart，逆问题：贝叶斯视角，《数值学报》19（2010）451–559。[22] S. L. Cotter、M. Dashti、A. M. Stuart，偏微分方程贝叶斯逆问题的近似，《SIAM数值分析杂志》48（1）（2010）322–345。[23]T. Bui-Thanh、O. Ghattas、J. Martin、G. Stadler，无限维贝叶斯逆问题的计算框架第一部分：线性化情形及其在全球地震反演中的应用，《SIAM Journal on Scientific Computing》35（6）（2013）A2494–A2523。[24] N. Petra、J. Martin、G. Stadler、O. Ghattas，无限维贝叶斯逆问题的计算框架第二部分：随机牛顿马尔可夫链蒙特卡洛及其在冰盖流动逆问题中的应用，《SIAM Journal on Scientific Computing》36（4）（2014）A1525–A1555。[25] T. Bui-Thanh、M. Girolami，使用黎曼流形哈密顿蒙特卡洛求解大规模偏微分方程约束的贝叶斯逆问题，《逆问题》30（11）（2014）114014。[26] A. Gelman、J. B. Carlin、H. S. Stern、D. B. Rubin，《贝叶斯数据分析》，Chapman and Hall/CRC，1995年。[27] S. Chib、E. Greenberg，《层次贝叶斯模型》，载于：《新帕尔格雷夫经济学词典》，施普林格，2018年，第5830–5838页。[28] G. O. Roberts、R. L. Tweedie，朗之万分布及其离散近似的指数收敛，《伯努利》2（4）（1996）341–363。[29] A. Stuart、A. Teckentrup，贝叶斯后验分布高斯过程近似的后验一致性，《计算数学》87（310）（2018）721–753。[30] T. Bai、A. L. Teckentrup、K. C. Zygalakis，与线性偏微分方程相关的贝叶斯逆问题的高斯过程，《统计与计算》34（4）（2024）139。[31] T. Deveney、E. Mueller、T. Shardlow，使用深度代理方法实现偏微分方程和积分方程模型中的高效贝叶斯反演，arXiv预印本 arXiv:1910.01547（2019）。[32] L. Yan、T. Zhou，基于深度神经网络的大规模贝叶斯逆问题自适应代理建模，arXiv预印本 arXiv:1911.08926（2019）。

[33] Z. Li, H. Zheng, N. Kovachki, D. Jin, H. Chen, B. Liu, K. Azizzadenesheli, A. Anandkumar, 用于学习偏微分方程的物理信息神经算子, ACM/IMS 数据科学期刊 1 (3) (2024) 1–27.[34] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar, 参数化偏微分方程的傅里叶神经算子, 见：国际学习表征会议（ICLR）, 2021.[35] L. Lu, P. Jin, G. Pang, Z. Zhang, G. E. Karniadakis, 基于算子普适逼近定理通过 DeepONet 学习非线性算子, 《自然·机器智能》 3 (3) (2021) 218–229.[36] S. Wang, H. Wang, P. Perdikaris, 利用物理信息 DeepONet 学习参数化偏微分方程的解算子, 《科学进展》 7 (40) (2021) eabi8605.[37] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention Is All You Need, 见：第31届国际神经信息处理系统会议论文集, 2017, 第6000–6010页.[38] B. Holzschuh, Q. Liu, G. Kohl, N. Thuerey, PDE-transformer：面向物理模拟的高效通用 Transformer, 见：第42届国际机器学习会议论文集, 2025.[39] A. Sinha, P. Malo, K. Deb, 双层优化综述：从经典方法到进化方法及其应用, IEEE 进化计算汇刊 22 (2) (2017) 276–295.[40] A. Behroozi, C. Shen, D. Kifer, 面向参数化微分方程正问题与反问题的灵敏度约束傅里叶神经算子, 见：国际学习表征会议（ICLR）, 2025.[41] S. W. Cho, H. Son, 用于求解偏微分方程反问题的物理信息深度逆算子网络, 见：国际学习表征会议（ICLR）, 2025.[42] L. Lu, R. Pestourie, S. G. Johnson, G. Romano, 用于高效学习偏微分方程的多保真深度神经算子及其在纳米尺度热输运快速逆向设计中的应用, 《物理评论研究》 4 (2) (2022) 023210.[43] R. Z. Zhang, C. E. Miles, X. Xie, J. S. Lowengrub, BiLO：面向偏微分方程反问题的双层局部算子学习, 《计算物理学杂志》 (2026) 114679.[44] R. M. Neal 等, 使用哈密顿动力学的马尔可夫链蒙特卡洛, 《马尔可夫链蒙特卡罗手册》 2 (11) (2011) 2.[45] L. Yang, X. Meng, G. E. Karniadakis, B-PINNs：面向含噪数据下偏微分方程正问题与反问题的贝叶斯物理信息神经网络, 《计算物理学杂志》 425 (2021) 109913.[46] G. Lin, Y. Wang, Z. Zhang, 基于贝叶斯物理信息神经网络、面向逆问题和正问题的多方差副本交换随机梯度马尔可夫链蒙特卡罗, 《计算物理学杂志》 460 (2022) 111173.[47] Y. Li, Y. Wang, L. Yan, 基于物理信息神经网络的贝叶斯逆问题代理建模, 《计算物理学杂志》 475 (2023) 111841.

[48] R. Z. Zhang, C. E. Miles, X. Xie, J. S. Lowengrub, Bayesian BiLO：面向贝叶斯偏微分方程逆问题高效不确定性量化的双层局部算子学习与低秩适配, arXiv 预印本 arXiv:2507.17019 (2026).[49] A. Raj, C. E. Gudumotou, S. Bun, K. Srinivasa, A. Sarshar, 用于偏微分方程贝叶斯参数估计的深度算子网络, arXiv 预印本 arXiv:2501.10684 (2025).[50] A. Glyn-Davies, A. Vadeboncoeur, O. D. Akyildiz, I. Kazlauskaite, M. Girolami, 物理信息深度生成建模的变分推断入门, Philosophical Transactions A 383 (2299) (2025) 20240324.[51] O. D. Akyildiz, M. Girolami, A. M. Stuart, A. Vadeboncoeur, 基于间接数据的高效先验校准, SIAM Journal on Scientific Computing 47 (4) (2025) C932–C958.[52] A. Vadeboncoeur, M. Girolami, A. M. Stuart, 总体逆问题中的高效反卷积, arXiv 预印本 arXiv:2505.19841 (2025).[53] G. Evensen, 数据同化：集合卡尔曼滤波, 施普林格, 2009.[54] A. Stuart, K. Zygalakis, 数据同化：数学导论, 施普林格, 纽约, NY, 美国, 2014.[55] M. A. Iglesias, K. J. Law, A. M. Stuart, 面向逆问题的集合卡尔曼方法, 《逆问题》 29 (4) (2013) 045001.[56] A. Garbuno-Inigo, F. Hoffmann, W. Li, A. M. Stuart, 相互作用朗之万扩散：梯度结构与集合卡尔曼采样器, SIAM Journal on Applied Dynamical Systems 19 (1) (2020) 412–441.[57] O. D. Akyildiz, F. R. Crucinio, M. Girolami, T. Johnston, S. Sabanis, 用于最大边际似然估计的相互作用粒子朗之万算法, ESAIM: Probability and Statistics 29 (2025) 243–280.[58] A. Glyn-Davies, C. Duffin, I. Kazlauskaite, M. Girolami, Ö. D. Akyildiz, 基于相互作用粒子朗之万动力学的统计有限元, SIAM/ASA Journal on Uncertainty Quantification 13 (3) (2025) 1200–1227.[59] R. Douc, E. Moulines, D. Stoffer, 非线性时间序列：理论、方法及应用（含 R 示例）, CRC Press, 2014.[60] A. Ern, J.-L. Guermond, 有限元的理论与实践, 第159卷, 施普林格, 2004.[61] D. P. Kingma, Adam：一种随机优化方法, arXiv 预印本 arXiv:1412.6980 (2025).