arXiv:2302.12697v1 [physics.comp-ph] 24 Feb 2023

面向多任务和多尺度正问题与逆问题的贝叶斯物理信息神经网络自适应加权*

Sarah Perez[1], Suryanarayana Maddu[2,3,4,5,6], Ivo F. Sbalzarini[2,3,4,5] 和 Philippe Poncet[1]

[1]Pau大学及阿杜尔地区大学，E2S UPPA、CNRS、LMAP UMR CNRS-UPPA 5142，Pau，法国。[2]Technische Universität Dresden，计算机科学学院，Nöthnitzer Str. 46，01062 德累斯顿，德国。[3]马克斯·普朗克分子细胞生物学与遗传学研究所，Pfotenhauerstr. 108，01307 德累斯顿，德国

[4]德累斯顿系统生物学中心，Pfotenhauerstr. 108，01307 德累斯顿，德国。[5]可扩展数据分析与人工智能中心（ScaDS.AI），德累斯顿/莱比锡，德国。[6]计算生物学中心，Flatiron Institute，162 5th Avenue，纽约，NY 10010，美国。

2023年2月27日

摘要

本文提出了一种用于贝叶斯物理信息神经网络（BPINN）自动自适应加权的新方法，并证明该方法能够稳健地处理多目标和多尺度问题。贝叶斯物理信息神经网络是一种广泛应用于数据同化的框架，将不确定性量化（UQ）与偏微分方程（PDE）的约束相结合。贝叶斯物理信息神经网络目标分布项的相对权重与相应学习任务固有的不确定性直接相关。然而，这些权重通常需要事先手动设定，这可能导致病态行为、稳定性问题以及任务之间的冲突，因而阻碍了贝叶斯物理信息神经网络在具有多尺度动力学的反问题中的应用。

本文提出的加权策略通过考虑目标后验分布的多任务特性，自动调整权重。我们证明，该策略能够消除贝叶斯物理信息神经网络的失效模式，并高效探索最优帕累托前沿。这使贝叶斯物理信息神经网络训练的收敛性和稳定性得到改善，同时减少采样偏差。此外，确定的权重还包含有关任务不确定性的信息，反映数据中的噪声水平以及偏微分方程模型的适用性。

我们通过索博列夫训练中的数值实验展示这一点，并将其与解析最优基线进行比较，同时研究一个多尺度 Lotka–Volterra 逆问题。最后，我们将这一框架应用于图像修复任务和一个反问题，涉及复杂几何结构中不可压缩流的潜在场恢复。

关键词： 哈密顿蒙特卡洛、不确定性量化、多目标训练、自适应权重学习、人工智能、贝叶斯物理信息神经网络。

1 引言

直接数值模拟依赖于基于物理原理构建的恰当数学模型，以概念化现实世界的行为并帮助理解复杂现象。实验数据主要用于参数识别和事后模型验证。然而，许多现实世界应用的特点是缺乏具有预测能力的物理模型，生命科学领域尤为如此。因此，在这些应用中，数据驱动的物理模型推断已成为一种互补方法 [27]。在依赖数据同化和逆向建模的应用中，情况也同样如此，例如地球科学领域。这使数据驱动模型成为科学应用中理论驱动模型的互补手段。

*已提交预印本。

根据可用数据量，可以选择多种数据驱动建模策略。近期发表的一篇综述 [13] 介绍了数据驱动建模的最新进展及其尚存的挑战，重点关注多孔介质研究中的应用。该综述涵盖了多种方法，从利用稀疏回归进行模型推断 [8, 14, 45, 26]，（即从数据中推断偏微分方程（PDE）模型的符号结构），到基于观测到的动力学外推的无方程预测模型 [34, 53, 28]。因此，模型推断方法既适用于基于物理的场景，也适用于无方程场景。

一种广受欢迎的框架是物理信息神经网络（PINN），它结合了这两种情形 [40]。该框架将可能稀疏且含噪的数据与物理原理（如以数学方程表示的守恒定律）相结合。这些方程对神经网络进行正则化，同时根据数据推断网络权重$\theta \in \mathrm{R}^d$和未知的方程系数 $\Sigma \in \mathrm{R}^p$。因此，PINN 得以用作代理模型，例如应用于流体力学领域 [33, 49, 37]。总体而言，PINN 为纯粹的数据驱动方法提供了有效的替代方案，因为物理正则化可以弥补高保真度数据的不足 [22, 31]。

尽管物理信息神经网络（PINN）有效且用途广泛，但要正确使用它们并不容易，因为它们容易出现各种训练不稳定性。这是因为其训练过程本质上是针对参数集合 $\Theta = \{\theta, \Sigma\}$ 进行加权多目标优化。

$$\widehat{\Theta} = \arg\min_{\Theta} \sum_{k=0}^{K} \lambda_k \mathcal{L}_k(\Theta), \tag{1}$$

其中，损失函数的每一项 $\mathcal{L}_k(\Theta)$ 都对应一个不同的推断任务。对于典型的物理信息神经网络，这些任务包括：数据拟合、偏微分方程残差最小化、边界条件和初始条件匹配，以及其他物理约束，例如所学习场的无散性。要妥善训练这一多任务学习问题，关键在于正确设置各损失项的权重 λ_k [29]。权重设置不当可能导致优化偏置 [39], 、任务特定梯度消失 [46, 10], 或灾难性遗忘 [29]。然而，自动优化损失权重并非易事，尤其是在高度非线性和多尺度问题中。

如何调整物理信息神经网络的损失权重是一个广为人知的问题，目前已有若干潜在解决方案来平衡各项目标 [9, 55, 56, 29]。这为公正地优化不同任务并充分探索最优帕累托前沿提供了准则 [43]。尽管通过减少优化偏差提高了可靠性，但对于预测的可信度、噪声估计和模型适切性，仍有若干未解问题 [15, 31, 60]。这些问题表明，需要进行不确定性量化（UQ）以确保推断可信，并将物理信息神经网络扩展为贝叶斯推断形式的贝叶斯物理信息神经网络[59, 21]。然而，如何将行之有效的物理信息神经网络加权策略应用于贝叶斯物理信息神经网络，并将其与不确定性量化相结合，仍是一个开放问题。

贝叶斯物理信息神经网络通过基于马尔可夫链蒙特卡洛（MCMC）采样，为预测提供后验分布估计（也称为贝叶斯模型平均） [57] ，从而实现不确定性量化。贝叶斯物理信息神经网络最常用的马尔可夫链蒙特卡洛方案之一是哈密顿蒙特卡洛（HMC），它能为具有平滑（物理）动力学的高维问题提供特别高效的采样器 [4]。尽管已有研究表明，哈密顿蒙特卡洛对贝叶斯物理信息神经网络更高效，但其形式意味着势能与物理信息神经网络的代价函数相关。因此，物理信息神经网络的多目标损失会直接转化为贝叶斯物理信息神经网络采样中的多重势能。因此，它同样面临避免偏置并高效探索帕累托前沿的困难。

这通常会导致 HMC 在 BPINN 训练过程中无法正确探索帕累托前沿。对于纳入 UQ 的多任务和多尺度学习问题，高效探索高维帕累托前沿仍然颇具挑战，而贝叶斯情形下尚未解决这一问题。挑战源于多势能中的每一项在贝叶斯框架内都由参数加权，这些参数与尺度、噪声幅度以及不同学习任务固有的不确定性有关[10]。尽管这些权重被认为是采样过程中的关键参数，但它们大多依靠人工调节 [32, 21, 31], ；当真实不确定性未知时，这会引入隐式先验。恰当地设置这些参数既不容易，也缺乏计算效率，还可能导致估计产生偏置，或在无效调参中耗费大量计算资源。因此，恰当优化这些权重

对于确保 HMC 从帕累托前沿附近的后验分布中进行采样至关重要。这不仅是稳健训练 BPINN 的必要条件，也有助于提高 UQ 估计的可靠性。

为稳健地处理贝叶斯物理信息神经网络中的多任务不确定性量化推断，本文探讨以下尚待解决的问题：如何自动调整贝叶斯物理信息神经网络中的权重，以高效探索帕累托前沿并避免不确定性量化推断中的偏置？如何在不施加隐式先验的情况下，处理对噪声分布（同方差或异方差）及其幅度的敏感性？

我们首先分析贝叶斯物理信息神经网络可能出现的失效模式，这些问题在多尺度或多任务反问题中尤为常见。随后，我们提出一种改进的哈密顿蒙特卡洛采样器，称为自适应加权哈密顿蒙特卡洛（AW-HMC），通过在采样过程中平衡梯度方差来避免这一问题。我们证明，这种方法能够得到适于探索帕累托前沿的加权后验分布。基准测试结果表明，该策略能够降低采样偏差，并增强贝叶斯物理信息神经网络的鲁棒性。具体而言，该方法通过确保积分达到最佳状态，并避免采样过程中过度缩短时间步长，提高了训练过程中蛙跳步的稳定性。此外，它还能根据各项对噪声的敏感程度及其不同尺度，自动调整势能权重及相应的不确定性。这显著提升了不确定性量化的可靠性，减少了超参数调优（包括先验分布调优）的需求，也降低了对噪声水平或适当任务尺度缩放的先验知识的依赖。我们证明，该方法在收敛率和计算成本方面均能改善贝叶斯物理信息神经网络。此外，我们还提出一种新的预测质量指标，用于量化边缘化步骤中的收敛率。最后，我们展示了所提出的方法能够将贝叶斯物理信息神经网络应用于复杂现实世界问题中的多尺度、多任务贝叶斯推断，即使数据稀疏且含噪亦然。

本文其余部分安排如下：第 2 节回顾贝叶斯物理信息神经网络和哈密顿蒙特卡洛采样器的一般原理，并分析它们在索博列夫训练中的失效模式。第 3 节介绍所提出的、基于贝叶斯物理信息神经网络进行不确定性量化的自适应加权策略。我们在第 3.2 节和 B 节中，通过一个具有已知解析解的基准问题验证该策略。随后，我们在第 3.3 节中针对 Lotka–Volterra 逆问题，展示所提出的 AW-HMC 算法的有效性，重点关注动力学参数的多尺度推断。接着，我们在第 4 节中展示 AW-HMC 在一个现实流体动力学问题中的应用。具体而言，我们利用稀疏噪声数据成功修复重建不可压缩狭窄血流，并展示了与噪声水平和噪声敏感性相一致的不确定性量化估计。最后，第 4.2 节讨论复杂几何中的逆流动问题，即如何根据部分速度测量推断流动状态（雷诺数倒数）和潜在压力场。我们在第 5 节总结全文并讨论所得观察。

2 从不确定性量化到贝叶斯物理信息神经网络：概念与局限

数据驱动或黑箱代理模型在现实世界中的应用仍然是一项具有挑战性的任务。预测通常需要将先验物理知识与稀疏噪声数据相结合，而先验物理知识的可靠性可能受到质疑，稀疏噪声数据还存在测量不确定性。这些现实世界问题还受到非线性 [21],尺度 [11], 和刚性 [19] 问题的影响，可能显著降低常用方法的效率。因此，需要开发能够稳健应对这些问题的数据驱动建模策略。

与此同时，对构建贝叶斯推断的需求引发了研究界的一个问题：如何确保估计值具有可信区间。这对于量化基础物理模型和测量数据中的不确定性都十分重要，尽管在刚性、多尺度或多保真问题的背景下，这可能颇具挑战性。因此，将不确定性量化嵌入此前的数据驱动方法，对于有效处理现实世界应用至关重要。

2.1 HMC-BPINN 的概念与原理

贝叶斯物理信息神经网络日益普及， [59, 21, 20, 32, 24] 为将不确定性量化纳入物理信息神经网络标准并发挥其预测能力提供了机会。它采用一种有趣的贝叶斯框架，声称能够处理现实世界中的稀疏噪声数据

此外，它还能提高模型及其预测的可靠性。

BPINN 背后的基本思想是，将每个未知量，即神经网络和逆参数 Θ，视为具有特定分布的随机变量，而不是像 PINN 中那样将其视为单一参数。不同的采样策略都旨在通过边缘化过程探索后验分布 Θ

$$P(\Theta|\mathcal{D},\mathcal{M}) \propto P(\mathcal{D}|\Theta)P(\mathcal{M}|\Theta)P(\Theta) \tag{2}$$

给定一些测量数据 $\mathcal{D}$ 和一个假定模型 $\mathcal{M}$，而不是寻找满足优化问题（1）的最佳近似。后验分布表达式（2）由贝叶斯定理得到，基本上包括数据拟合似然项 $P(\mathcal{D}|\Theta)$、偏微分方程似然项 $P(\mathcal{M}|\Theta)$ 和联合先验分布 $P(\Theta)$。这些具体项将在各个应用及相应章节中逐一详细说明。随后，贝叶斯边缘化将参数 Θ 的分布转换为预测结果的后验分布，也称为贝叶斯模型平均（BMA）：

$$\underbrace{\mathcal{P}(y|x,\mathcal{D},\mathcal{M})}_{\text{predictive BMA distribution}} = \int \underbrace{\mathcal{P}(y|x,\Theta)}_{\text{prediction for }\Theta} \underbrace{\mathcal{P}(\Theta|\mathcal{D},\mathcal{M})}_{\text{posterior}} \mathrm{d}\Theta \tag{3}$$

其中，x 和 y 分别指神经网络的输入（例如空间和时间点）与输出（例如场预测）。在该方程中，由所有 Θ 参数采样（2）产生的不同预测，按照其后验概率加权并进行平均，从而为 BPINN 的输出提供内在的不确定性量化（UQ）。总体而言，BPINN 对参数 Θ 进行贝叶斯边缘化，由此形成关注量（QoI）的预测分布（3），即学习得到的场和逆参数。

针对深度神经网络中的贝叶斯推断，人们开发了多种方法，包括变分推断 [58, 23] 和马尔可夫链蒙特卡洛方法。基于哈密顿动力学的一种特殊 MCMC 采样器——哈密顿蒙特卡洛（HMC）——日益受到关注，因为它能够利用后验分布的几何性质处理高维问题。Betancourt 从概念层面阐释了 HMC 的高效性 [4]，并从理论上证明了链的遍历性和收敛性 [6, 25]。从数值角度来看，Yang 等人 [59] 指出，与变分推断版本相比，BPINNs-HMC 形式在正问题和逆问题上的表现更优。这使 HMC 无论在理论上还是数值上，都成为贝叶斯物理信息神经网络的一种高效马尔可夫链蒙特卡罗方案。

下面我们简要回顾经典 BPINNs-HMC 的基本原理，并指出其局限性，尤其是在多目标和多尺度问题中。

HMC 的基本思想是假设一个虚拟粒子，其连续的位置和动量（Θ，r）沿着无摩擦的负对数后验（NLP）几何上的哈密顿动力学演化。它需要引入辅助变量 r，将式（2）的采样嵌入到相空间中联合概率分布$\pi(\Theta, r)$的探索之中

$$\pi(\Theta, r) \sim \mathrm{e}^{-H(\Theta,r)}. \tag{4}$$

后者依赖于哈密顿量 $H(\Theta, r) = U(\Theta) + K(r)$ 的一种特定分解，其中势能和动能分别为 $U(\Theta)$ 和 $K(r)$，其选取方式满足

$$\pi(\Theta, r) \propto P(\Theta|\mathcal{D},\mathcal{M})\mathcal{N}(r|0,\mathbf{M}) \tag{5}$$

动量服从中心化多元高斯分布，其协方差矩阵（或质量矩阵）$\mathbf{M}$ 通常为缩放单位矩阵。因此，系统的哈密顿量为

$$H(\Theta, r) = U(\Theta) + \frac{1}{2} r^T \mathbf{M}^{-1} r \tag{6}$$

其中，势能与目标后验分布直接相关。该能量项通常表示为负对数后验 $U(\Theta) = -\ln P(\{\mathrm{v}5, \mathcal{M})$，由此形成第2.2节所述的多势能。这确保了 Θ 的边缘分布能够直接提供目标后验分布的样本

$$P(\Theta|\mathcal{D},\mathcal{M}) \sim \mathrm{e}^{-U(\theta)} \tag{7}$$

因为对联合分布 $\pi(\Theta,\ r)$ 的高效探索会直接转化为对目标分布的高效探索，正如 Betancourt [4] 所述。HMC 采样过程在确定性步之间交替进行；在确定性步中，我们根据哈密顿动力系统求解无摩擦粒子的运动路径

$$\begin{cases} \mathrm{d}\Theta = \mathbf{M}^{-1} r \, \mathrm{d}t \\ \mathrm{d}r = -\nabla U(\Theta)\, dt, \end{cases} \tag{8}$$

以及随机步，此时按照前文介绍的高斯分布对动量进行采样。由于哈密顿方程（8）在理论上能够保持系统的总能量，因此每个确定性步都被限制在特定能级上，而随机步则使我们能够在能级集合中扩散，从而在相空间中进行高效探索。在确定性步中，能级集合的这种理论守恒需要采用能够确保能量守恒的数值方案。

因此，通常使用辛积分器对哈密顿动力学（8）进行数值求解：Störmer-Verlet方法，也称为蛙跳法。然而，这些积分器并不能完全消除离散化误差；实际上，这些误差可能会破坏确定性迭代过程中的哈密顿量守恒。因此，最后还需在该过程中加入一个校正步骤，以减少数值积分中离散化误差引起的偏置：这便形成了基于哈密顿转移的Metropolis-Hastings准则。该接受准则倾向于通过拒绝导致概率转移发散的样本来保持能量。然而，确定性轨迹的探索仍对两个控制积分时间的特定超参数十分敏感：蛙跳法的步长 δt 和迭代次数L 。这些参数的调优可能颇具挑战，尤其是在后验分布存在病态区域或高曲率区域 [4], 、从而导致不稳定、性能不佳以及MCMC估计量有效性不足时。尽管采用了能够保持哈密顿性质的数值方案，传统的HMC-BPINN仍可能面临病态差异。

为抵消这些发散效应，人们致力于开发相应策略：在保持详细平衡条件的同时，自适应地设定轨迹长度 L [17] ，或者采用标准的自适应马尔可夫链蒙特卡洛方法，动态调整步长 δt [5]。在这方面，Hoffmann 和 Gelman 提出的 No-U-Turn采样器（NUTS）是最流行的自适应策略之一 [16]。尽管这类自适应方法或许能提供一种应对发散轨迹的替代方案，但发散轨迹仍表明MCMC估计存在显著偏置。这不禁引出一个问题：当面对多个势能项，且这些势能项在后验分布中导致截然不同的几何行为或尺度时，这种自适应方法是否有效？事实上，这种自适应策略主要是调整蛙跳参数，使最敏感的项满足能量守恒要求；这可能导致为其他势能项乃至整个后验分布选取的超参数并不理想。这反映出依赖调整蛙跳超参数的自适应策略存在局限性。

当这些相异的病态现象变得普遍时，Betancourt [4] 提出的另一种方法是对目标分布进行正则化；但在真实世界的应用中，这可能颇具难度，并会导致额外的调参。尽管如此，这为研究各项学习任务对目标分布整体行为的影响提供了绝佳机会，也为新颖的自适应加权策略铺平了道路。

在接下来的章节中，我们将重点关注真实世界多任务和多尺度范式所带来的挑战。我们将展示，当前的贝叶斯物理信息神经网络方法在这些情况下会遭遇重大失败，并利用基于这些发散概率转移的强大诊断方法，识别主要的病态现象。

2.2 多目标问题范式

关于物理信息神经网络中的多目标优化问题，针对直接问题或反问题而产生的目标后验分布 (2) 的采样，需要使用多势能项$U(\Theta)$。此外，在真实世界的应用中，我们必须处理稀疏且含噪的测量，其保真度也可能涵盖不同尺度：这就是包含多源数据[22, 31]的多保真问题。

~ ~为了一般性，我们引入一个时空域 $\Omega = \Omega \times \mathcal{T}$，其中 $\Omega \subset \mathrm{R}^n$ $n = 1,\ 2,\ 3$

并假设偏微分方程组具有以下形式：

$$\begin{cases} \mathcal{F}(u(t,x),\,\Sigma) &= 0, \qquad (t,x)\in\Omega \\ \mathcal{H}(u(t,x),\,\Sigma) &= 0, \qquad (t,x)\in\Omega \\ \mathcal{B}(u(t,x),\,\Sigma) &= 0, \qquad (t,x)\in\Omega^{\partial}:=\partial\widetilde{\Omega}\times\mathcal{T} \\ \mathcal{I}(u(t,x),\,\Sigma) &= 0, \qquad (t,x)\in\Omega^{I}:=\widetilde{\Omega}\times\mathcal{T}_0 \end{cases} \tag{9}$$

其中， u 是主未知量， $\mathcal{F}$ 是主要微分方程（例如 Navier–Stokes 方程）， $\mathcal{H}$是附加约束（例如不可压缩条件）， $\mathcal{B}$ 和 $\mathcal{I}$ 分别是边界条件和初始条件，而 Σ 是偏微分方程模型参数，可以是已知的，也可以是推断得到的。在定义域 $\Omega^u \subset \Omega$ 的某个子集中，还可能存在解场 u 的部分测量值。随后，对时空定义域的这种连续描述进行离散化，以便选取训练数据集，并将其用于贝叶斯物理信息神经网络采样。

我们首先定义训练数据集 $\mathcal{D}$，将其分解为 $\mathcal{D}=\mathcal{D}^{\Omega}\cup\mathcal{D}^{\partial}\cup\mathcal{D}^{I}\cup\mathcal{D}^{u}$，其中包括分别从集合 Ω、Ω^{∂}、Ω^{I} 和 Ω^{u} 中采样得到的离散含噪测量值。对于数据污染，我们假设 u 上的稀疏观测受到独立高斯噪声的影响，因此将 $\mathcal{D}^u$ 定义为

$$\mathcal{D}^u=\{(t_i,x_i,u_i),\quad \text{s.t}\quad (t_i,x_i)\in\Omega^u\quad\text{and}\quad u_i:=u(t_i,x_i)+\xi_u(t_i,x_i),\,i=1...N^u\} \tag{10}$$

其中，噪声 $\xi_u\sim\mathcal{N}(0,\ \sigma_u^2 I)$ 和标准差 σ_u 可在条件允许时根据传感器保真度估计。贝叶斯物理信息神经网络中的神经网络部分随后为 u 提供代理模型，对参数 $\Theta=\{\theta,\Sigma\}$ 的每个样本，该模型记为 u_Θ；这些参数的先验分布记为 $P(\Theta)$。后者同时考虑神经网络参数 θ 的先验（假设这些参数服从以零为中心的相互独立高斯分布）以及模型参数 Σ 的先验；在独立性条件下， $P(\Theta)=P(\theta)P(\Sigma)$。对于正问题，偏微分方程模型参数是预先给定的，此时先验分布可简化为 $P(\theta)$。当存在未知量的部分测量值，即 $\mathcal{D}^u$ 不是空集时（逆问题或图像修复问题即属此类），代理模型u_Θ 应在贝叶斯框架中满足数据拟合似然项。具体而言，需要在集合 $\mathcal{D}^u$上量化神经网络预测与训练数据之间的拟合程度：

$$P(\mathcal{D}^u|\Theta)\propto\prod_{i=1}^{N^u}\exp\left(\frac{-\left(u_\Theta(t_i,x_i)-u_i\right)^2}{2\sigma_u^2}\right). \tag{11}$$

同样，通过在集合 $\mathcal{D}^{\partial}$上施加边界条件来约束模型输出

$$\mathcal{D}^{\partial}=\left\{(t_i,x_i,\mathcal{B}(u_i)),\quad \text{s.t}\quad (t_i,x_i)\in\Omega^{\partial}\quad\text{and}\quad \mathcal{B}(u_i):=\mathcal{B}(u(t_i,x_i))+\xi_b(t_i,x_i),\,i=1...N^{\partial}\right\} \tag{12}$$

通过满足以下边界似然项

$$P(\mathcal{D}^{\partial}|\Theta)\propto\prod_{i=1}^{N^{\partial}}\exp\left(\frac{-\left(\mathcal{B}(u_\Theta(t_i,x_i))-\mathcal{B}(u_i)\right)^2}{2\sigma_b^2}\right). \tag{13}$$

边界条件项的噪声敏感性也通过独立高斯分布来表征，即 $\xi_b\sim\mathcal{N}(0,\ \sigma_b^2 I)$，其中标准差 σ_b需要估计。之所以对 ξ_u 和 ξ_b 加以区分，是因为无法保证数据损坏具有均匀性：事实上，在面对异方差噪声时，测量分布的方差可能因局部而异。地球科学领域便是如此，基于X射线显微断层成像图像的数据驱动建模需要特别关注边界噪声估计 ξ_b。这主要是由于成像伪影的局限（例如部分容积效应、边缘增强），它们往往会加剧材料界面处的模糊效应，从而影响对介质有效性质（如渗透率和微孔隙度）的量化 [35, 2]。初始条件也同样如此，其敏感性可能有所不同 ξ_i。在贝叶斯物理信息神经网络中，给定偏微分方程组 (9)，先前的数据拟合项还会补充物理原理，以此对神经网络的预测进行正则化。

关于偏微分方程似然项， $\mathcal{D}^{\Omega}$ 数据集定义为训练点，在这些点上，我们通过代理建模强制满足偏微分方程和额外的物理约束：

$$\mathcal{D}^{\Omega} = \left\{(t_i, x_i) \in \Omega, \quad \mathcal{F}(u_{\Theta}(t_i, x_i)) = \xi_f(t_i, x_i) \quad \text{and} \quad \mathcal{H}(u_{\Theta}(t_i, x_i)) = \xi_h(t_i, x_i),\, i = 1...N^{\Omega}\right\} \tag{14}$$

其中， ξ_f 和 ξ_h 表示两个方程中的模型不确定性，这些不确定性通常未知，且很容易导致物理模型设定错误。根据这些记号，正问题是在 $\mathcal{D}^u = \varnothing$ ，而 Σ 已知时，仅基于偏微分方程物理假设，对 Ω 上的场 u_{Θ} 进行直接预测。相反，反问题旨在结合偏微分方程模型以及关于预测场 u的部分含噪信息 $\mathcal{D}^u$ ，推断 Σ 。最后，图像修复问题利用这些部分测量来补充并恢复预测场中缺失的信息，同时结合基于偏微分方程的先验。

最后，根据贝叶斯法则，将 Θ (2)的目标后验分布分解为一系列多任务似然项——包括数据拟合和偏微分方程似然——以及先验：

$$P(\Theta|\mathcal{D}, \mathcal{M}) \propto P(\mathcal{D}^u|\Theta)P(\mathcal{D}^{\partial}|\Theta)P(\mathcal{D}^I|\Theta)P(\mathcal{D}^{\Omega}, \mathcal{F}|\Theta)P(\mathcal{D}^{\Omega}, \mathcal{H}|\Theta)P(\theta)P(\Sigma) \tag{15}$$

这使得哈密顿蒙特卡洛采样器对应的多势能为

$$\begin{aligned} U(\Theta) = &\frac{\|u_{\Theta} - u\|^2_{\mathcal{D}^u}}{2\sigma_u^2} + \frac{\|\mathcal{B}(u_{\Theta}) - \mathcal{B}(u)\|^2_{\mathcal{D}^{\partial}}}{2\sigma_b^2} + \frac{\|\mathcal{I}(u_{\Theta}) - \mathcal{I}(u)\|^2_{\mathcal{D}^I}}{2\sigma_i^2} \\ &+ \frac{\|\mathcal{F}(u_{\Theta})\|^2_{\mathcal{D}^{\Omega}}}{2\sigma_f^2} + \frac{\|\mathcal{H}(u_{\Theta})\|^2_{\mathcal{D}^{\Omega}}}{2\sigma_h^2} + \frac{\|\theta\|^2_{\mathbb{R}^d}}{2\sigma_{\theta}^2} + \frac{\|\Sigma - \mu_{\Sigma}\|^2_{\mathbb{R}^p}}{2\sigma_{\Sigma}^2} \end{aligned} \tag{16}$$

根据方程 (7)。记号 $\|\cdot\|$ 表示对数似然项所采用的均方根（RMS）范数（继承自开集 Ω 上的泛函 ^{2}-范数），或对数先验项所采用的通常欧几里得范数。此外，此处在一般框架下，根据先验假设 $P(\theta) \sim \mathcal{N}(0,\ \sigma_{\theta}^2 I_d)$ 和 $P(\Sigma) \sim \mathcal{N}(\mu_{\Sigma},\ \sigma_{\Sigma}^2 I_p)$ 写出多势能 (16)。我们注意到，在等价的约束优化问题中，对数先验项可视为一种 ^{2} 正则化。不过，合适地选择这些先验分布（因而恰当地调整参数 σ_{θ}、 μ_{Σ} 和 σ_{Σ} ）通常并非易事，而且耗时。总体而言，方程 (16) 表明，即使在简单的问题设置中，贝叶斯物理信息神经网络也可能遇到一个势能项，它与物理信息神经网络中的加权多目标损失十分相似，而后者的权重主要依靠人工调节。

因此，主要挑战在于对帕累托最优解附近进行采样，使贝叶斯物理信息神经网络能够提供高效且可靠的预测和不确定性量化。否则，风险在于所获得的样本会聚集在某个局部极小值附近，对应于多势能项中的一个，而以牺牲其他项为代价。

其次，尽管标准差 $\sigma_{\bullet}$ 是需要选择的关键参数，并且与固有任务中的不确定性相关，但大多数作者要么为其指定一个给定值，要么将其作为额外的超参数 [32, 50, 21]进行训练。这可能导致高度有偏的预测，尤其是在设置 PDE 残差标准差σ_f 和 σ_h 时，因为这会对模型适切性引入很强的先验。

最近，Psaros 等人 [38] 讨论了除其他事项外，在未知且异方差噪声分布的背景下，对其中一些参数（主要是数据拟合标准差）进行调整的推广方案。他们要么依赖离线学习 ，代价是需要一个预训练的生成对抗网络（GAN）；要么基于额外的参数训练进行权重的在线学习 。特别是，当考虑位置相关的方差时，这些额外参数的数量可能会急剧增加，正如 [38], 针对实际应用所建议的那样，因此会产生较高的计算成本。如何在不通过学习额外超参数增加计算复杂度的情况下，处理此类未知的（同方差或异方差）噪声分布，仍然是一个有待解决的问题。

最后，尽管总不确定性量化中已指出物理模型设定错误这一问题，但当在物理约束上假定存在误导性的模型不确定性时， [38] 尚未对此进行讨论 $\mathcal{F}$ $\mathcal{H}$。因此，如何避免通过手动调整通常未知或预先设定的超参数σ_f σ_h，对模型适切性引入强先验，仍是一项具有挑战性的任务。

鉴于此，我们希望在一个测试案例中检验通常采用的 BPINNs-HMC 方法（如第2.1节所述）的鲁棒性。该案例展示了使用索博列夫训练训练神经网络进行采样时，多目标和多尺度特性所引发的问题。

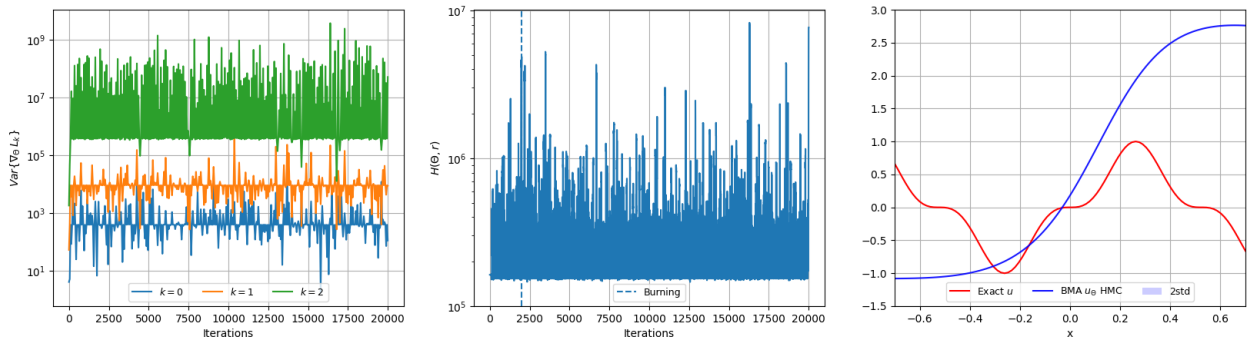

图 1：Sobolev 训练中，HMC 均匀加权在训练至二阶导数时的失效模式会导致哈密顿量不守恒（中间），并使最终函数近似效果极差（右侧）。这是因为有效梯度 $\nabla_\Theta L_k$分布（$k = 0$，1，2）的方差严重失衡；图中左侧展示了这些分布随 ($N_s \times L$) HMC 迭代次数的变化。

2.3 贝叶斯物理信息神经网络的索博列夫训练失效模式

索博列夫训练是多目标贝叶斯物理信息神经网络采样的一种特殊情况，由于涉及的尺度差异悬殊，它可能导致学习过程僵化 [29]。尽管如此，机器学习领域仍普遍采用这种方法：将目标函数导数的信息加入损失函数或其等效的势能中，以提高神经网络的预测效率和泛化能力 [12, 48, 54, 62]。

这种特殊训练为检验当前 BPINNs-HMC 方法对任务特定梯度消失 [29]这一失效模式的鲁棒性提供了基准。此外，它还提供了一个机会，可将其与索博列夫多目标优化中已知的解析 ε-最优权重进行基准比较 [29]。

这里在一个索博列夫回归任务上测试 BPINNs-HMC 采样，这意味着数据集仅限于 $\mathcal{D} = \mathcal{D}^u$ 包含函数及其导数测量值的数据 D_x^k, $k \geq 1$，最高阶数为 K，因此目标后验分布为

$$P(\Theta|\mathcal{D}) \propto \prod_{k=0}^{K} P(\mathcal{D}^u, D_x^k u|\Theta)P(\Theta) \tag{17}$$

因此，势能具有如下的一般形式

$$U(\Theta) = \sum_{k=0}^{K} \left[\frac{\lambda_k}{2\sigma_k^2} \|D_x^k u_\Theta - D_x^k u\|^2 \right] + \frac{\lambda_{K+1}}{2\sigma_{K+1}^2} \|\Theta\|^2 := \sum_{k=0}^{K+1} \lambda_k \mathcal{L}_k(\Theta) \tag{18}$$

其中，$L_k = \lambda_k \mathcal{L}_k$ 表示加权 k^{th} 目标项，λ_k 为一些用于定义该项的正加权参数（见第 3.1 节）。本节仅采用均匀加权策略，其中 $\lambda_k = 1, \forall k$，这对应于经典的 BPINNs-HMC 形式。为便于阅读，方程 (18) 汇总了神经网络和逆参数的对数先验项，并假设它们具有相同的先验分布。

我们首先介绍一种一维索博列夫训练，考虑最高至二阶导数，并采用定义在 $\Omega = [-0.7, 0.7]$ 上的测试函数 $u(x) = \sin^3(\omega x)$，其中 $\omega = 6$。我们使用 100 个训练点，将蛙跳参数$L = 100$ 和 $\delta t = 1\text{e}-3$ 分别设为迭代次数和时间步长，并进行 $N_s = 200$次采样迭代。此外，我们将测试限制在函数逼近问题上，因此后文的 Θ仅指神经网络参数。除非另有说明，以下所有 σ_k 均设为 1，因为在实践中，我们无法获取这些参数在导数或偏微分方程残差项中的取值；若有可用信息，我们只能获取数据场 u 上的观测噪声。

与 PINN 类似，这个采用均匀权重 λ_k 的测试案例会导致不同目标项之间的梯度方差失衡。具体而言，其中的高阶导数具有占主导地位的梯度方差，使其他任务的梯度趋于消失，并导致对后验分布的有偏探索。在图1（左）中，我们看到，与高阶导数对应的项 $\text{Var}\{\nabla_\Theta L_2\}$ 很快发展到比其他有效梯度方差高两个数量级。此外

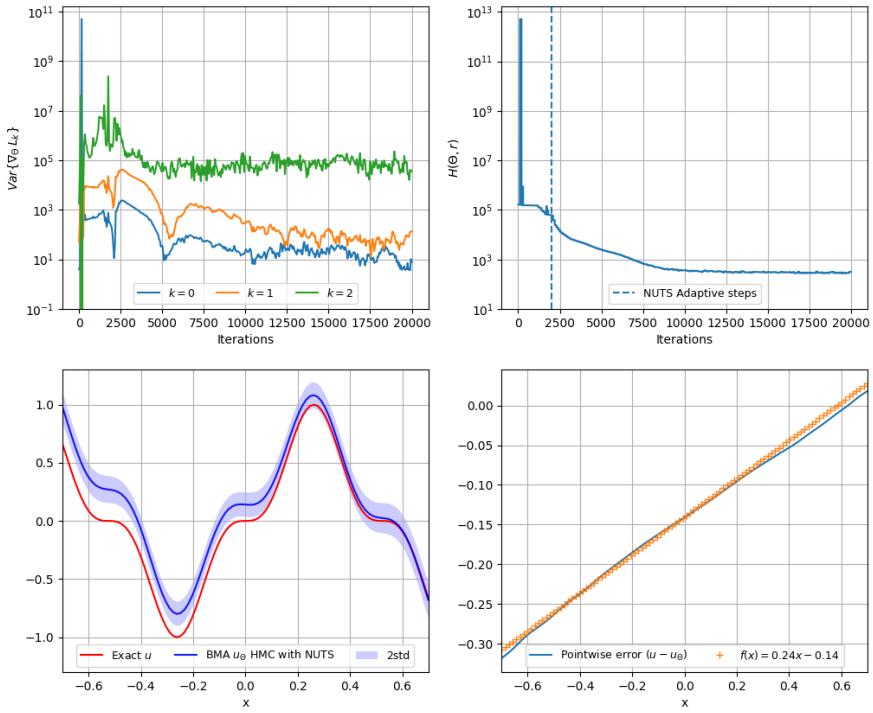

图2：索博列夫训练中，NUTS 步长自适应在处理二阶及更高阶导数时的失效模式：有效梯度 $\nabla_\Theta L_k$ 分布的方差（左上）和哈密顿演化（右上）分别显示了任务不平衡和对能级探索不足。信号近似（左下）和逐点误差（右下）凸显了 u_Θ的线性偏差。竖直虚线标示了 NUTS 采样器中的自适应步数。

，除了对帕累托前沿的低效探索之外，我们还面临由最高阶导数项引起的稳定性问题，导致哈密顿量沿蛙跳轨迹无法守恒（见图1中）。如第2.1节所述，经典 HMC 采用均匀加权时出现的此类发散病态，是所得估计量存在偏置的有力诊断依据，也会使人怀疑后者的有效性。

一种抵消这些影响的方法是减小时间步长 δt ，以平衡导数项的量级并提高贝叶斯物理信息神经网络-HMC 的 Metropolis-Hastings 接受率。然而，在蛙跳迭代中使用较小的时间步长，更容易产生病态的随机游走行为或有偏采样 [16, 4]。为此，我们尝试采用带步长自适应的 No-U-Turn采样器（NUTS）作为自适应策略，具体见 Hoffmann 和 Gelman 的算法5 [16] ，该算法由 Python 开源软件包 hamiltorch 实现 [11]。为遵循相同的假设，我们采用与之前完全相同的一组蛙跳参数，并设置 $N = 20$ 个自适应步，最终得到的自适应时间步长为 $\delta t = 1.29\text{e}{-4}$。在这种情况下，我们再次遇到无法有效探索帕累托前沿的情形，图2中有效梯度的方差证实了这一点。与信号本身相比，这使二阶导数的近似效果更好，也表明存在有偏采样：由于高阶导数项占主导，信号 u 只能确定到一个线性函数。这种线性偏差也显示在图2右下图中。这证实了 NUTS 的时间步长自适应更侧重于维持占主导地位的高阶导数项，而该项导致了刚性。

简而言之，即使是在均匀权重这一简单设置下进行简单的一维索博列夫训练，也会导致经典 BPINNs-HMC 方法严重失效，因为后验分布对产生不稳定性的高阶导数十分敏感。因此，哈密顿量守恒的这种偏离使采样方法无法运行。此外，能够确保哈密顿量守恒的替代方法也无济于事，因为它们要么无法有效探索各个能级，要么在多任务和多尺度采样中存在严重失衡。这表明，哈密顿马尔可夫链无法充分探索由该势能所决定的目标分布的帕累托前沿，而且严重

失衡的条件无法通过常规方法克服。

因此，我们旨在通过寻找合适的加权形式，制定一种策略，使不同任务之间达到平衡条件，而不受其尺度影响。无论是否需要处理 HMC 中有关蛙跳参数自适应设置的常见问题，这种方法都至关重要；它还有助于减少不稳定性，而无需不必要地减小时间步长。

3 贝叶斯物理信息神经网络中用于无偏不确定性量化的自适应加权策略

传统的贝叶斯物理信息神经网络形式在多目标推断方面存在局限，例如稳定性和刚性问题、任务之间的病态冲突，以及对帕累托最优前沿的有偏探索。仅靠自适应设置蛙跳参数（如 NUTS采样器所采用的方法）无法解决这些问题；手动调整标准差 σ 也无济于事，反而会增加计算成本或造成能量浪费。因此，我们研究另一种自适应方法，转而直接对目标分布进行正则化：该方法旨在通过自动调整关键 σ 参数来平衡任务加权。

3.1 逆狄利克雷自适应加权算法：AW-HMC

鉴于 HMC-BPINN方法的局限性（前文第 2 节已讨论），开发一种新的替代方案变得至关重要，尤其是在处理源自真实世界数据的复杂多目标问题时。该策略必须应对以下主要病态问题：1）确保对目标后验分布的帕累托前沿进行探索；2）管理不同项的尺度敏感性；3）控制哈密顿不稳定性。

无论这些病态问题如何，如何设定关键的 σ 参数仍然是一个问题，尤其是在数据中的噪声水平和对偏微分方程模型的信心都不是先验已知的情况下。尽管手动调整这些参数仍十分常见，我们可以依靠 λ 权重调整来隐式确定噪声和固有的任务不确定性，而不是引入可能导致误导性预测的、关于模型适切性的强先验。

为满足所有这些要求，我们采用一种基于逆狄利克雷的方法，该方法已证明在 PINNs框架中处理平衡训练和多尺度建模时行之有效 [29]。该方法依据损失项梯度的方差调整权重，这些方差可解释为高维多目标优化问题中相对于主要下降方向的训练不确定性。与传统训练相比，该策略还能显著提高收敛性，并避免特定任务的权重消失。

为贝叶斯物理信息神经网络开发逆狄利克雷自适应加权算法，旨在将不同任务中的此类训练不确定性纳入贝叶斯框架，从而同时考虑噪声、模型适切性和任务敏感度，并确保对帕累托前沿进行探索。因此，我们试图确定正的加权参数 $\lambda_k, k = 0, \ldots\ldots, K$，使势能项的加权梯度 $\nabla_\Theta L_k = \lambda_k \nabla_\Theta \mathcal{L}_k$ 分布具有均衡的方差。我们提出，通过使梯度分布具有相同的方差

$$\gamma^2 := \mathrm{Var}\{\lambda_k \nabla_\Theta \mathcal{L}_k\} \simeq \min_{t=0,\ldots,K}(\mathrm{Var}\{\nabla_\Theta \mathcal{L}_t\}), \quad \forall k = 0, ..., K \tag{19}$$

采用基于逆狄利克雷的方法设置权重：

$$\lambda_k = \left(\frac{\min_{t=0,\ldots,K}(\mathrm{Var}\{\nabla_\Theta \mathcal{L}_t\})}{\mathrm{Var}\{\nabla_\Theta \mathcal{L}_k\}}\right)^{1/2} = \left(\frac{\gamma^2}{\mathrm{Var}\{\nabla_\Theta \mathcal{L}_k\}}\right)^{1/2} \tag{20}$$

使得

$$\lambda_k \mathcal{N}(\mu_k, \mathrm{Var}\{\nabla_\Theta \mathcal{L}_k\}) = \left(\frac{\gamma^2}{\mathrm{Var}\{\nabla_\Theta \mathcal{L}_k\}}\right)^{1/2} \mathcal{N}(\mu_k, \mathrm{Var}\{\nabla_\Theta \mathcal{L}_k\}) = \mathcal{N}(\mu_k, \gamma^2). \tag{21}$$

请注意，本文不讨论与先验 $P(\Theta)$ 相对应的 λ_{K+1} 情形，因为在等价的约束优化问题中，对数先验项更类似于一种 ^{2}-正则化，因此权重平衡方法主要关注势能中的对数似然项。事实上，采样应使我们能够高效探索对应于数据拟合项与不同基于偏微分方程的似然项之间平衡条件的帕累托前沿。相反，我们不希望依赖非信息先验来实现任务平衡，因此施加以下上界

$$\mathrm{Var}\{\lambda_{K+1}\nabla_{\Theta}\mathcal{L}_{K+1}\} \leq \gamma^2, \tag{22}$$

这可以通过设置 $\lambda_{K+1} \leq \sigma_{K+1}$ 来实现，这与关于先验 $P(\Theta) \sim \mathcal{N}(0,\ \sigma_{K+1}^2 I)$ 的假设有关。这源于以下观察:

$$\lambda_{K+1}\nabla_{\Theta}\mathcal{L}_{K+1}(\Theta^{t_\tau}) = \frac{\lambda_{K+1}}{\sigma_{K+1}^2}\Theta^{t_\tau} \quad \text{s.t} \quad \mathrm{Var}\{\lambda_{K+1}\nabla_{\Theta}\mathcal{L}_{K+1}(\Theta^{t_\tau})\} = \frac{\lambda_{K+1}^2}{\sigma_{K+1}^4}\mathrm{Var}\{\Theta^{t_\tau}\} \leqslant \frac{1}{\sigma_{K+1}^2}\mathrm{Var}\{\Theta^{t_\tau}\} \tag{23}$$

其中，Θ^{t_τ} 是在迭代 τ 中采样得到的参数集合。后一个上界还提供了一个离散程度指标，用于衡量 Θ 的后验方差与其先验分布之间的差异；给定 γ^2 后，可据此设定 σ_{K+1} 的值。

我们研究了用于应对 BPINNs-HMC 失效模式的在线方法，使得权重自适应策略（20）取决于采样迭代 τ。由此得到一种改进的哈密顿蒙特卡洛方法，称为自适应加权哈密顿蒙特卡洛（AW-HMC），详见算法1。该加权策略会持续进行，直至达到设定的自适应迭代次数 N；该次数可能不同于通常的预热步数 M。该方法假设 $N \leq M$，使我们能够得到适合探索帕累托前沿的加权后验分布。事实上，有限次自适应可保持链的遍历性和渐近收敛；同时，保持 $N \leq M$ 可确保后验分布始终从相同的加权势能中抽取。实践中，先验设定的预热阶段与自适应步数密切相关，令 $M = N$ 即可体现这一点。我们还引入记号 $H_{\lambda_\tau}(\Theta,\ r)$ 表示加权哈密顿量

$$H_{\lambda_\tau}(\Theta, r) = \sum_{k=0}^{K+1} \lambda_k(\tau)\mathcal{L}_k(\Theta) + \frac{1}{2}r^T\mathbf{M}^{-1}r \tag{24}$$

该概率定义了 Metropolis-Hastings接受准则中的新转移概率 .

基于梯度最小方差（20）对目标分布进行的当前平衡，可以解释为根据多势能中最可能或最不敏感的项调整权重。因此，它具有以下优势：促进物理信息神经网络向帕累托最优解收敛，同时增强输出不确定性量化的可靠性，其样本取自帕累托前沿。事实上，该加权策略通过自适应调整 λ 参数，自动提高最不可能任务的不确定性。这些观察结果源于对每个梯度方差上界的推导，详见 A，其中涉及预测误差和偏微分方程残差，以及表征每个任务平均梯度下降方向变异性的敏感性项。鉴于此，我们能够从基本且一般的角度，为联合方差 γ^2 提供一个上界，该上界在方程（48）中推导得到。

最后但同样重要的是，基于逆狄利克雷分布的自适应加权使我们不必不合理地缩短时间步长，因为时间步长不再需要满足所有严格的尺度要求来确保哈密顿量守恒。因此，该方法无需对蛙跳超参数 δt 和 L进行过度调优，就能实现采样。此外，由于该方法能够采用最优积分时间（无论从收敛率还是时间步长是否适合不同学习任务的角度来看），因此可以避免采样中出现病态的随机游走或发散行为。

首先，在不同复杂度的索博列夫训练基准上验证当前的 AW-HMC 算法，为与 ε-最优性结果进行比较提供依据。这也使我们能够为贝叶斯物理信息神经网络建立一种新的收敛诊断指标。随后，在各个章节中，我们将在更复杂的多任务和多尺度问题上检验该方法的鲁棒性和效率。

算法1： 自适应加权哈密顿蒙特卡洛（AW-HMC）

输入： 初始 Θ^{t_0}，N_s 样本数，L 蛙跳步数，δt 蛙跳步长，N 自适应迭代次数，M 预热步数以及 $\mathbf{M}$ 质量矩阵。

采样过程：
对于 $\tau = 1 \ldots N_s$ **do**
  采样 $r^{t_\tau - 1} \sim \mathcal{N}(0, \mathbf{M})$;
  设定 $(\Theta_0, r_0) \leftarrow (\Theta^{t_{\tau-1}}, r^{t_{\tau-1}})$;
  权重自适应：
  如果 $\tau \leq N$ **则**
    计算 $\lambda_k(\tau) = \left(\text{最小值}_{j=0,\ldots,K} (\text{Var}\{\nabla_\Theta \mathcal{L}_j(\Theta_0)\}) \right) \text{Var}\{\nabla_\Theta \mathcal{L}_k(\Theta_0)\}^{-1/2}$ $\forall k = 0, \ldots, K$ 与 $\lambda_{K+1}(\tau) = 1$;
  else
    $\lambda_k(\tau) = \lambda_k(\tau - 1)$ $\forall k = 0, \ldots\ldots, K$ 和 $\lambda_{K+1}(\tau) = 1$
  end
  蛙跳法：
  对于 $i = 0 \ldots L - 1$ **执行**
    $r_i \leftarrow r_i - \frac{\delta t}{2} \sum_{k=0}^{K+1} \lambda_k(\tau) \nabla_\Theta \mathcal{L}_k(\Theta_i)$;
    $\Theta_{i+1} \leftarrow \Theta_i + \delta t \mathbf{M}^{-1} r_i$;
    $r_{i+1} \leftarrow r_i - \frac{\delta t}{2} \sum_{k=0}^{K+1} \lambda_k(\tau) \nabla_\Theta \mathcal{L}_k(\Theta_{i+1})$;
  end
  梅特罗波利斯–黑斯廷斯算法：
  采样 $p \sim \mathcal{U}(0, 1)$;
  使用 (24) 计算 $\alpha = \min(1, \exp(H_{\lambda_\tau}(\Theta_0, r_0) - H_{\lambda_\tau}(\Theta_L, r_L))$;
  如果 $p \leqslant \alpha$ **则**
    $\Theta^{t_\tau} = \Theta_L$;
  else
    $\Theta^{t_\tau} = \Theta_0$;
  end
  燃烧后收集样本：$\{\Theta^{t_i}\}_{i=M}^{N_s}$
结束

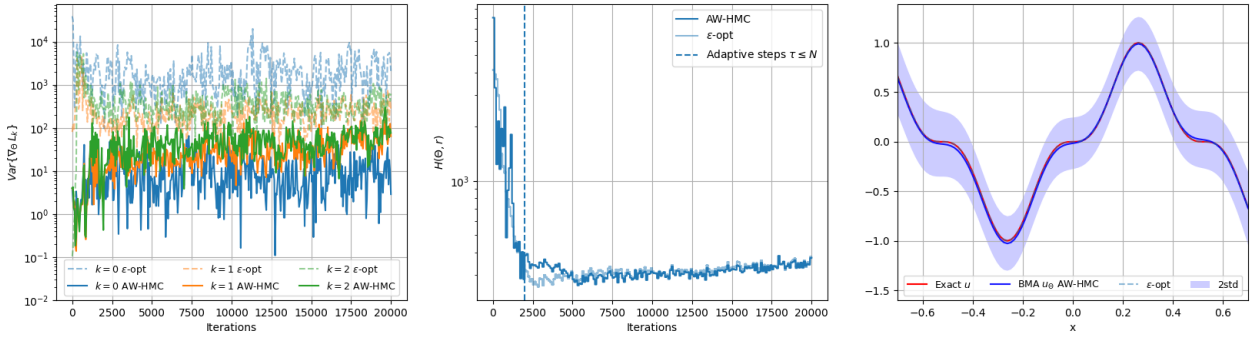

图 3：将自适应加权哈密顿蒙特卡洛（AW-HMC）应用于最高二阶导数的索博列夫训练，并与 ε-最优加权进行比较。平衡条件下各任务间的有效梯度分布方差 $\text{Var}\{\nabla_\Theta L_k\}$（左图）。采样过程中满足能量守恒的哈密顿演化（中图）。场预测结果，并比较 ε-最优结果与 AW-HMC策略（右图）。

3.2 Sobolev训练基准与收敛诊断

我们研究所提出的自动加权BPINN方法在若干应用中的表现，本节从索博列夫训练基准开始。首先，我们将这一新的自适应加权策略应用于第 2.3 节介绍的一维索博列夫训练，并采用完全相同的一组超参数。自适应步数设为 $N = 20$，与 NUTS 采样相同，以确保对不同方法进行公正比较。

此外，我们将预测结果与一个参考情形进行比较，其中权重 λ_k 设定为符合 ε最优解析解 [52], 的值；该解析解可根据方程 [29]针对索博列夫训练求得：

$$\lambda_k^{\varepsilon} = \frac{\prod_{j\neq k}\mathcal{I}_j}{\sum_{k=1}^{K}\prod_{j\neq k}\mathcal{I}_j} \quad \text{s.t. } \mathcal{I}_k = \int_{\Omega} |D_x^k(u)|^2 \,\mathrm{d}x. \tag{25}$$

我们以此 ε最优解为基准测试了我们的方法，并假设观测噪声为 $\xi_u \sim \mathcal{N}(0,\ \sigma_u^2 I)$，使得 $\sigma_k = \sigma_u$, $\forall k$ 且 $\sigma_u = 0.1$。如图 3（中）所示，两种方法的结果吻合良好，哈密顿量收敛至相同能级：事实上，自适应步之后，[2]最优方法与 AW-HMC 方法所得哈密顿量值之间的相对误差约为 $1\mathrm{e}{-4}$ 。与这一最优解相比，AW-HMC 方法在信号及其导数上的相对误差均约为 $1\mathrm{e}{-3}$。最后，我们还指出，如图 3（左）所示，梯度方差达到平衡，这与采用 ε最优解析权重时观察到的情况相同。因此，在梯度方差的平衡、对哈密顿量能级的探索以及总体 BMA 预测方面，AW-HMC 方法与 ε最优解的结果相近。

我们提出的新方法展现出不同任务间极为平衡的条件：在整个训练过程中，有效梯度分布方差 $\mathrm{Var}\{\nabla_\Theta L_k\}$ 始终处于相同的数量级，即使自适应步数有限也是如此。这意味着，权重自动调整后得到的后验分布非常适合收敛至帕累托前沿探索，从而提高采样效率。与第 2.3 节中表现出重大缺陷的传统方法相比，我们的方法有效避免了梯度方差严重失衡，进而减少了任务特异性偏差，显著改善了对此类多目标势能的边缘化。

为进一步展示该方法的鲁棒性，我们考虑此测试案例的三阶导数扩展。即使采用 NUTS 自适应策略调整时间步长（达到 $\delta t = 1.36\mathrm{e}{-7}$），此处仍会产生病态随机游走行为，使采样完全失效（图 4 顶行和图 5）。时间步长如此显著地减小，显然是由多任务学习中三阶导数项引起的刚性增强所致。事实上，由于这种刚性，哈密顿轨迹在确定性步期间更容易发散，因此需要较小的 δt 来补偿发散。为避免由此产生的病态随机游走，必须延长整体积分时间，但在蛙跳时间步长受到此类限制的情况下，这必然会导致计算成本过高。这凸显了 NUTS 在处理涉及不同尺度的刚性多任务采样时的主要局限。

相比之下，我们的方法克服了这些主要缺陷（见图 5），无需对δt施加额外约束，并如图 4（底行）所示，在不同任务之间提供了均衡的梯度方差。我们还将 AW-HMC 方法与基于 ε-最优性的解析权重进行比较，结果显示两种方法高度一致。此外，为了应对采样变异性所引起的 BPINNs 随机过程，我们使用神经网络参数和动量的不同初始化，对采样进行了多次重复。这使得我们能够得到沿自适应步的平均权重演化，如图 6 所示；其量级与解析的ε-最优权重相同。

除了对不同方法与解析解之间进行这些定性比较之外，我们随后引入了一个新的指标，用于量化预测的质量。这一指标在常用指标的基础上，增加了对边缘化过程中的采样收敛性的量化。AW-HMC 过程中预热步数之后收集的样本——即 $\{\Theta^{t_i}\}_{i=M}^{N_s}$ 的所有实例——首先用于确定贝叶斯模型平均估计，如方程（3）所定义。每个样本都会为由 Θ^{t_i}表征的神经网络提供一个预测 $P(y|x,\ \Theta^{t_i})$，并且理论上取自

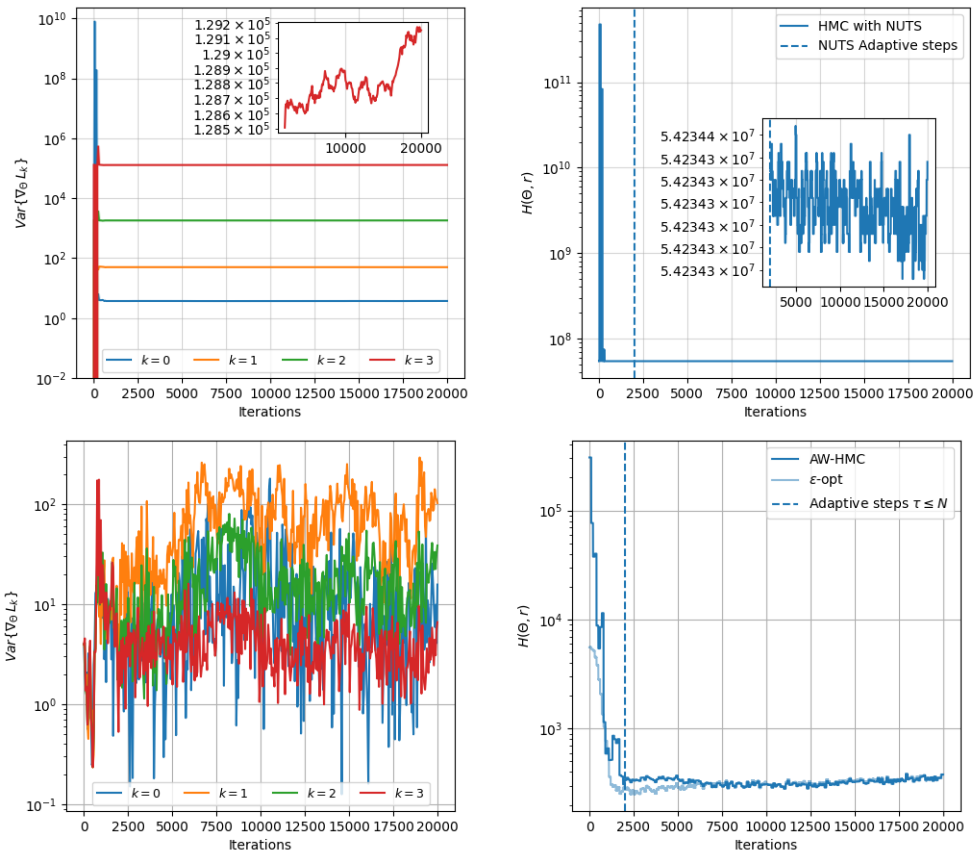

图4：一维索博列夫训练中截至三阶导数的有效梯度分布方差 $\mathrm{Var}\{\nabla_\Theta L_k\}$ 及哈密顿演化。NUTS公式（上）突出显示了学习任务之间的严重不平衡（左），以及探索能量等值集时的随机游走行为（右）。AW-HMC策略（下）及其与 ε-最优性的比较。

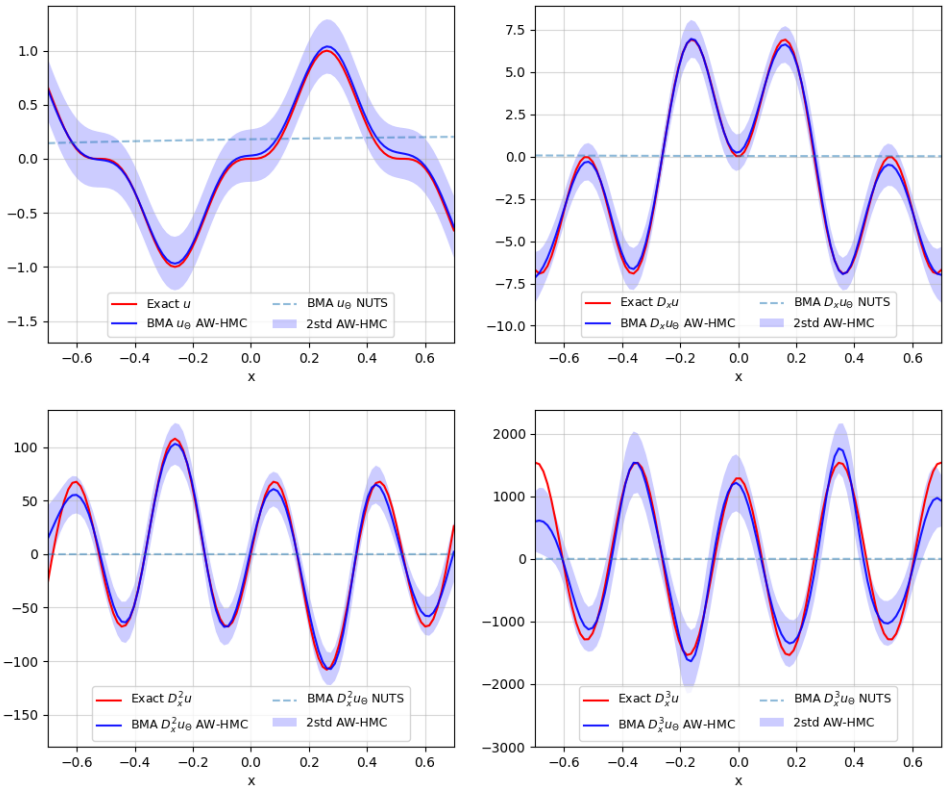

图 5：最高至三阶导数的一维索博列夫训练：AW-HMC 与 NUTS 方法下，函数及其导数的 BMA 预测对比。任务之间的不平衡以及 NUTS 的随机游走行为（见图 4）导致 BMA 预测效果不佳。AW-HMC 方法克服了这些影响，并显著改善了目标分布的采样。

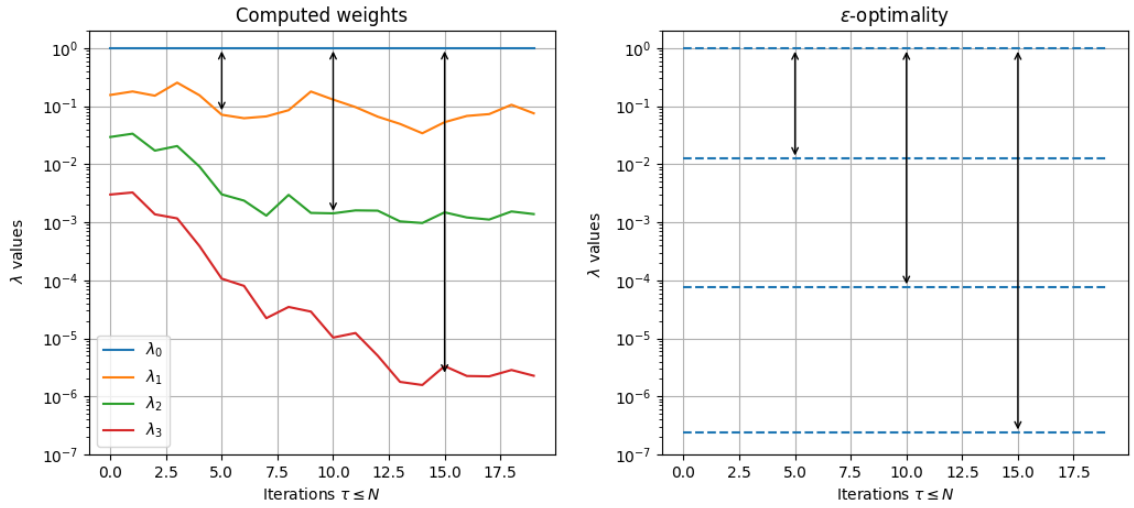

图6：左侧为自适应步（$\tau \leq N$）中 λ_k 权重的演化，以及与最高达三阶导数的索博列夫训练所对应的解析 ε最优权重的比较。AW-HMC 算法多次重复运行中平均权重的演化。这是由神经网络参数和动量的不同初始化引起的，以考虑样本变异性。相对权重的数量级 $\lambda_0/\lambda_i,\ \ i = 1...3$ 由双向箭头表示。

来自后验分布 $P(\Theta|\mathcal{D},\ \mathcal{M})$，使得贝叶斯模型平均通常可近似为 [57]:

$$P(y|x,\mathcal{D},\mathcal{M}) \simeq \frac{1}{N_s - M}\sum_{i=M}^{N_s} P(y|x,\Theta^{t_i}) \quad \text{with} \quad \Theta^{t_i} \sim P(\Theta|\mathcal{D},\mathcal{M}). \tag{26}$$

在索博列夫训练中，我们将函数本身及其所有导数 $y = \left\{D_x^k u_\Theta,\ k = 0...K\right\}$ 的预测视为神经网络输出，因此可以根据方程（26）计算相对于每个输出定义的 BMA误差:

$$\text{BMA-E}^k = \frac{\left\|P\left(D_x^k u_\Theta \,|\, x,\mathcal{D},\mathcal{M}\right) - D_x^k u\right\|^2}{\|D_x^k u\|^2}, \qquad \forall k = 0...K \tag{27}$$

其中，此处使用的记号 $\|\cdot\|$ 指函数 ^{2}-范数。基于前述定义，为了将边缘化过程中 BMA 的收敛纳入考量，我们引入一种名为累积（相对）BMA误差的新诊断指标，定义如下:

$$\text{BMA-CE}^k(\tau) = \frac{\left\|\dfrac{1}{\tau - N}\displaystyle\sum_{i=N}^{\tau} P\left(D_x^k u_\Theta \,|\, x,\Theta^{t_i}\right) - D_x^k u\right\|^2}{\|D_x^k u\|^2}, \qquad \forall k = 0...K \tag{28}$$

取决于算法1中自适应步之后的采样迭代次数 $\tau > N$。这些公式可直接推广至更一般框架下的所有神经网络输出，并以收敛率衡量采样效率。图7展示了索博列夫训练三阶扩展的累积 BMA误差，突显了 AW-HMC采样器在各个函数任务上的收敛情况（左图）。相较之下，对于病态的 HMC 和 NUTS 方案，这些量几乎保持不变，分别是由于大量拒绝和随机游走行为所致（见图7右图）。

最后，我们将这一索博列夫训练测试案例扩展到若干二维基准测试，研究函数复杂度和训练点数量对贝叶斯模型平均误差的影响。附录 B 提供了这些基准问题及训练设置的详细信息；结果表明，AW-HMC 算法提升了贝叶斯物理信息神经网络的鲁棒性和效率。

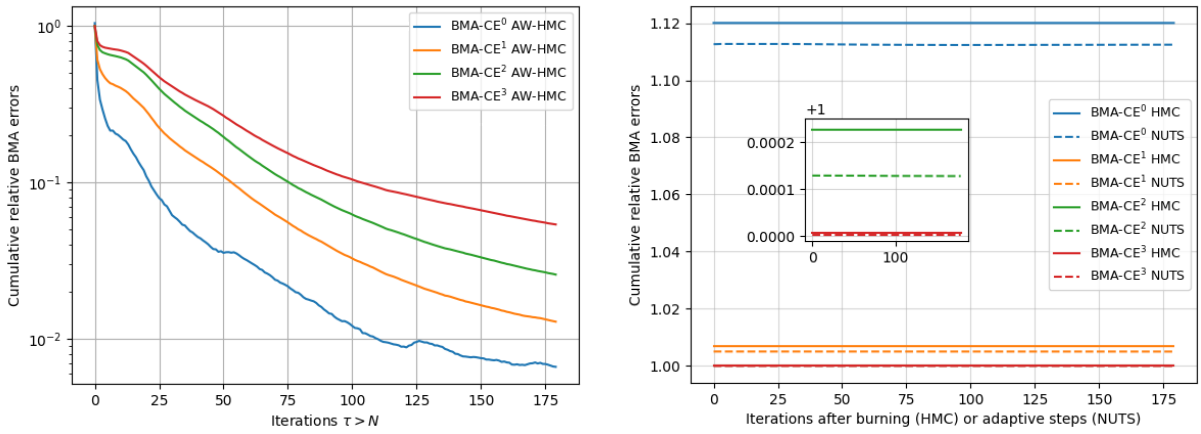

图 7：根据方程 (28) 计算的索博列夫训练累积相对贝叶斯模型平均误差，适用于三阶导数，贯穿整个采样迭代过程 $\tau > N$。比较 AW-HMC 策略（左）与经典 HMC 和 NUTS 形式（右）。在病态情形下，这些量几乎保持不变，原因可能是大量拒绝或病态随机游走，这凸显了常见 BPINNs-HMC 形式（右）的不收敛问题。

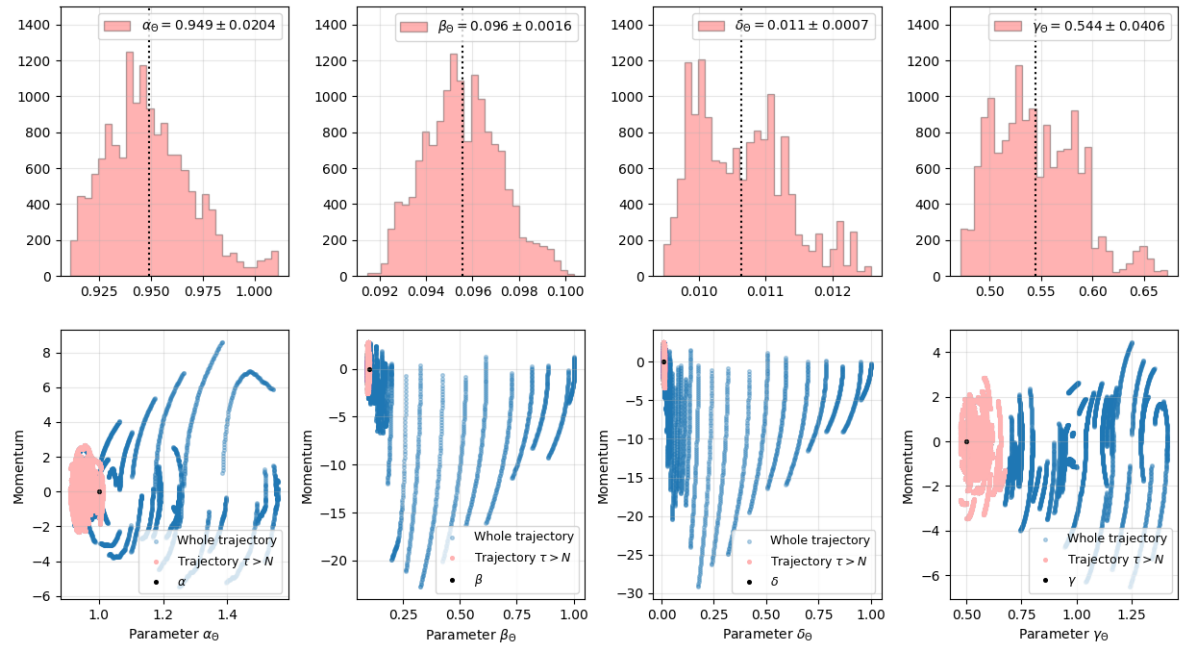

图 8：洛特卡-沃尔泰拉多尺度推断：逆参数 α_Θ、 β_Θ、 δ_Θ 和 γ_Θ 的边缘后验分布直方图（上）。采样过程中参数轨迹的相图（下），展现了自适应步期间（蓝色）向各自众数收敛，以及自适应步之后（红色）对众数邻域的有效探索。真实参数分别为 $\alpha = 1$、 $\beta = 0.1$、 $\delta = 0.01$ 和 $\gamma = 0.5$，构成了一个具有不同尺度的反问题。

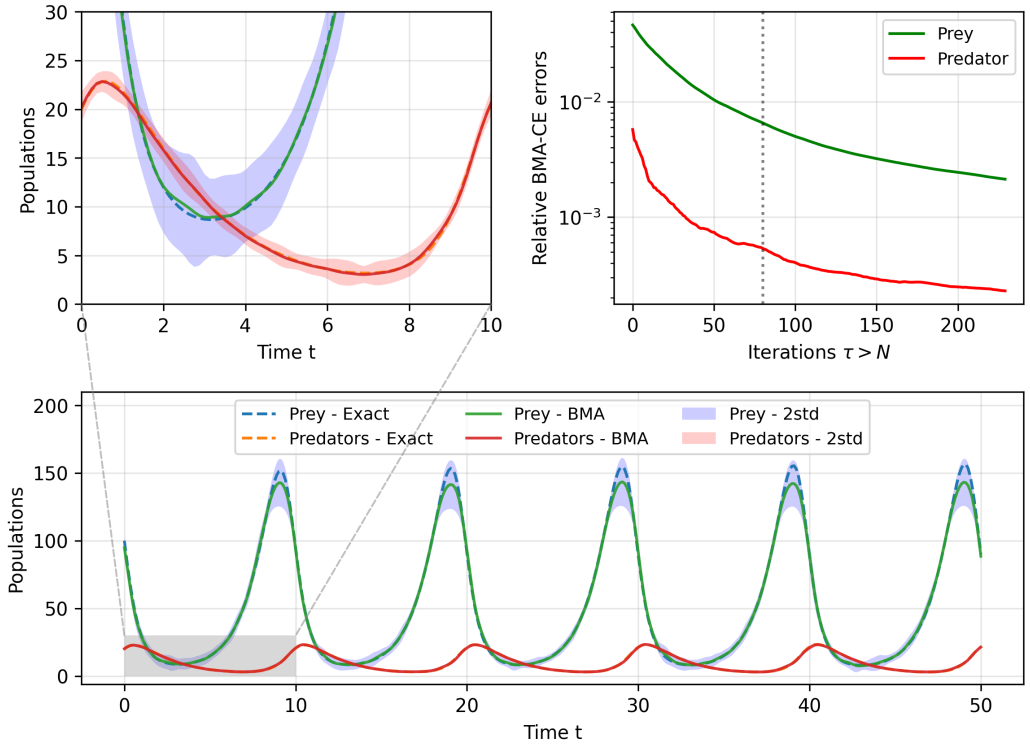

图 9：Lotka–Volterra多尺度推断：沿物理时间变化的双物种种群的BMA预测及其不确定性，见左下图和左上图。按照方程 (28) 定义的神经网络输出 $y = (u_\Theta,\ v_\Theta)$ 的相对BMA-CE误差，在整个采样迭代过程中绘制于右上图。点状竖线标记了在序贯训练中引入常微分方程似然项的时刻。

3.3 一个多尺度Lotka–Volterra逆问题

我们展示了AW-HMC在多尺度动力学逆问题中的应用，以量化尺度变化的影响。正如Linka等人 [21] 指出的那样，对尺度的敏感性可能会阻碍经典贝叶斯物理信息神经网络的性能，尤其是在考虑非线性动力系统时。因此，现实世界问题所导致的多尺度特性和刚性，以及其中普遍存在的任务特定梯度消失现象，为量化当前方法的鲁棒性提供了一个有趣的基准。

在此背景下，我们考虑一个参数数量级差异显著的Lotka–Volterra动力系统，其由以下常微分方程（ODE）系统定义：

$$\begin{cases} \dfrac{\mathrm{d}u}{\mathrm{d}t} = \alpha u - \beta uv, \quad t \in \Omega \\ \dfrac{\mathrm{d}v}{\mathrm{d}t} = \delta uv - \gamma v, \quad t \in \Omega \\ u(0) = u_0,\ v(0) = v_0 \end{cases} \tag{29}$$

该模型刻画了捕食者-猎物物种随时间的演化。符号 $u(t)$ 和 $v(t)$ 分别表示给定时刻 t 的猎物种群规模和捕食者种群规模，而参数 α，β，δ，$\gamma \geq 0$则控制种群动力学，代表种群增长和衰减速率。随后，我们将初始种群规模设为 $u_0 = 100$ 和 $v_0 = 20$，并将以下参数 $\alpha = 1$，$\beta = 0.1$，$\delta = 0.01$ 和 $\gamma = 0.5$ 特意选为具有不同数量级的数值。由此构建了一个基于现实世界动力学且涉及不同尺度的逆问题基准。

首先，通过在均匀时间网格 $\Omega = [0, 50]$ 上求解常微分方程系统（29），以400个点的精细分辨率对观测数据进行数值生成。随后对数据进行随机采样，以便在不同采样器的训练阶段仅使用其中的一半。数据集 $\mathcal{D}$ 随后包含在200个不同时间点对 u 和 v 的这些部分测量，可能还加入了一些噪声，同时保留相同的配点以满足常微分方程约束。在本节中，我们聚焦于一个反问题，即推断

λ_k / 序列步骤	λ_0	λ_1	λ_2	λ_3
数据拟合（步骤 1）	3.83e−2	1	—	—
数据拟合 + 常微分方程任务（步骤 2）	4。87e−2	1	9.16e−3	1。16e−1

$\widetilde{\sigma}_k$ / 序列步骤	$\widetilde{\sigma}_0$	$\widetilde{\sigma}_1$	$\widetilde{\sigma}_2$	$\widetilde{\sigma}_3$
数据拟合（步骤 1）	5.109	1	—	—
数据拟合 + 常微分方程任务（步骤 2）	4.531	1	10.45	2.936

表 1：在洛特卡-沃尔泰拉多尺度逆问题中，每个序列步骤（上方各行）经过自适应步后得到的权重参数 λ_k。由权重调整产生的有效标准差 $\widetilde{\sigma}_k$，按 $\widetilde{\sigma}_k = \sqrt{1/\lambda_k}$ 计算，分别对应每个序列步骤（下方各行）。这凸显了与猎物物种相关任务的不确定性有所增加。序列步骤的划分详见第3.3节。

未知模型参数 $\Sigma = \{\alpha, \beta, \delta, \gamma\}$，同时从这些测量数据中恢复原始精细分辨率下的完整物种演化。

鉴于两个种群之间明显的尺度差异，我们将任务划分为捕食者–猎物两组，使得每个场 u 和 v 都满足一个数据拟合似然项和一个常微分方程残差似然项。我们还假设 Σ 服从对数正态先验分布，以确保逆参数为正；Yang 等人 [59] 已证明，此类先验有助于改善推断；此外，我们为神经网络参数 θ设定相互独立的正态分布。不过在实际操作中，我们通过引入 $\Sigma = e^{\widetilde{\Sigma}} := \left\{e^{\widetilde{\alpha}},\ e^{\widetilde{\beta}},\ e^{\widetilde{\delta}},\ e^{\widetilde{\gamma}}\right\}$ 对每个逆参数进行变量变换，从而推断 $\widetilde{\Sigma}$，并同样假设其服从正态先验分布~。对于这个测试用例，我们采用弱信息先验，尤其是对 Σ，因为我们希望我们的方法能够凭借任务的无偏自动加权处理多尺度推断。因此，我们假设神经网络参数和逆参数都服从相同的先验分布，即 $\Theta \sim \mathcal{N}(0,\ \sigma_\Theta^2 I_{p+d})$，其中 $\Theta = \left\{\theta,\ \widetilde{\Sigma}\right\}$。

在这些假设下，我们可以定义相应哈密顿系统的多势能：

$$
\begin{aligned}
U(\Theta) = \frac{\lambda_0}{2\sigma_0^2}\left\|u_\Theta - u\right\|_{\mathcal{D}}^2 + \frac{\lambda_1}{2\sigma_1^2}\left\|v_\Theta - v\right\|_{\mathcal{D}}^2 + \frac{\lambda_2}{2\sigma_2^2}\left\|\frac{\mathrm{d}u_\Theta}{\mathrm{d}t} - \alpha_\Theta u_\Theta + \beta_\Theta u_\Theta v_\Theta\right\|_{\mathcal{D}}^2 \\
+ \frac{\lambda_3}{2\sigma_3^2}\left\|\frac{\mathrm{d}v_\Theta}{\mathrm{d}t} - \delta_\Theta u_\Theta v_\Theta + \gamma_\Theta v_\Theta\right\|_{\mathcal{D}}^2 + \frac{1}{2\sigma_\Theta^2}\|\Theta\|_{\mathbb{R}^{p+d}}^2
\end{aligned}
\tag{30}
$$

~其中，推断得到的逆参数定义为 $\Sigma_\Theta = e^{\Sigma_\Theta}$，并且我们还将所有 $\sigma_\bullet$ 设为 1，因为我们不希望对任务和模型不确定性施加强先验。正如前文第2.2节所述，对于最后一项，这些范数分别为均方根和欧几里得范数。假设参数的先验服从具有较大标准差 $\sigma_\Theta = 10$ 的高斯分布，因为稍微弥散的分布会为 Θ 参数带来弱信息先验。这也确保满足非信息先验的约束（22）。

对于此类逆向建模，采样通过序贯训练分解为两个步骤。这意味着：1）采用 AW-HMC策略对神经网络参数进行采样，主要针对数据拟合似然项（设置 $\lambda_2 = \lambda_3 = 0$）。2）随后，我们引入式（30）中的 ODE 残差任务，使用 AW-HMC 算法，并以步骤 1）得到的初始神经网络参数 θ^{t_0}来估计缺失的逆参数。最终，BMA 预测和不确定性量化依赖于整个采样过程。在两步序贯训练中，首先将自适应迭代次数和采样迭代次数分别设为 $N = 20$ 和 $N_s = 100$，然后设为 $N = 50$ 和 $N_s = 200$；步骤 1）和步骤 2）的时间步长所对应的蛙跳参数分别为$L = 100$、$\delta t = 5\mathrm{e}{-4}$ 和 $2\mathrm{e}{-4}$。神经网络本身

由 4 层组成，每层有 32 个神经元。考虑到 Lotka-Volterra系统解的周期性，我们使用正弦激活函数。

对于此类反问题，经典BPINNs-HMC算法会遭遇大量拒绝，因为哈密顿轨迹无法守恒，从而导致无效采样（图17）。即使采用针对时间步长的自适应策略，也难以应对多尺度动力学，并且必须将 δt 值大幅降低才能获得一定的稳定性，详见C。蛙跳时间步长受到此类限制，自然会导致算法无法收敛至帕累托前沿，且在弱信息先验的条件下，逆参数的推断效果不佳（见C中的图18和图19）。事实上，Linka等人 [21] 在学习COVID-19动力学时也遇到了同样的问题，并在（ [21]第4.3节中）对逆参数采用了对数正态先验分布，而这些分布本身已依赖于恰当的尺度设定。对恰当尺度设定的这种需求会显著影响推断，因为它需要先验知识，从而使采样产生偏置。

相反，我们假设先验分布与尺度无关，并表明我们的方法只需极少调参且无需降低 δt，就能准确恢复所有 Σ 参数，并预测物种演化。恢复不同尺度不再需要关于反问题参数缩放的先验知识，即可收敛至各自的众数。图8所示结果同时展示了每个推断所得逆参数的边缘后验分布 Σ_Θ 及其在探索相空间分布 π时的轨迹Θ，r。对于后者，我们绘出了完整的采样轨迹：它们在自适应步期间收敛至各自的众数，最终如 $\tau > N$的最终轨迹所示，在众数附近进行采样。这证实了AW-HMC能够快速识别该反问题中不同的众数，并处理此类多尺度动力学。

为了量化参数识别的有效性，我们还测量了参数推断中的相对误差 $\Sigma_\Theta = \{\alpha_\Theta, \beta_\Theta, \delta_\Theta, \gamma_\Theta\}$

$$E_{\Sigma_\Theta} = \frac{|\Sigma_\Theta - \Sigma|}{\Sigma} \tag{31}$$

其中，预测由 $\Sigma_\Theta = \dfrac{1}{N_s - N} \sum_{i=N}^{N_s} e^{\widetilde{\Sigma_\Theta}^{t_i}}$ 给出；我们表明，四个逆参数的相对误差均在 $5\mathrm{e}{-2}$ 左右。

图 9 展示了物种种群的预测演化，该预测是对神经网络输出 $y = (u_\Theta,\ v_\Theta)$ 进行贝叶斯模型平均（BMA）得到的，并从定性和定量两方面与精确解进行了比较。为此，我们计算了两种物种的相对 BMA累积误差，以突出采样的收敛情况，结果见图 9 右上方。与纯粹基于数据的采样相比，我们发现，在两步序贯训练中加入常微分方程残差似然项，能够改善预测的收敛性。

该测试案例还揭示，猎物种群的演化具有更高的不确定性，其有效标准差约为其他种群的四倍（见表 1）。与势能（30）中的 λ_1 和 λ_3 相比，自适应步结束时 λ_0 和 λ_2 的值较小，这凸显了这些特定任务的不确定性更高。因此，AW-HMC策略能够自适应地调整 λ 参数的权重，并基于其梯度方差对任务不确定性进行内在刻画。

4 在计算流体力学中的应用：狭窄血流

我们通过一个流体力学领域的实际问题，具体而言，通过研究非对称几何中的不可压缩狭窄流修复重建与逆问题，来展示第 3.1 节所述方法的应用。目标是在数据存在噪声且稀疏的情况下，展现当前 AW-HMC 算法在更复杂的二维几何和非线性偏微分方程动力学中的泛化能力与性能。

测量数据通过在离散位置对完全解析的计算流体力学（CFD）解进行随机采样生成。按照 [7] 中的详细介绍，采用基于离散化校正粒子强度交换（DCPSE）方法的无网格求解器，对非对称狭窄血管几何中的血管流进行直接数值模拟。

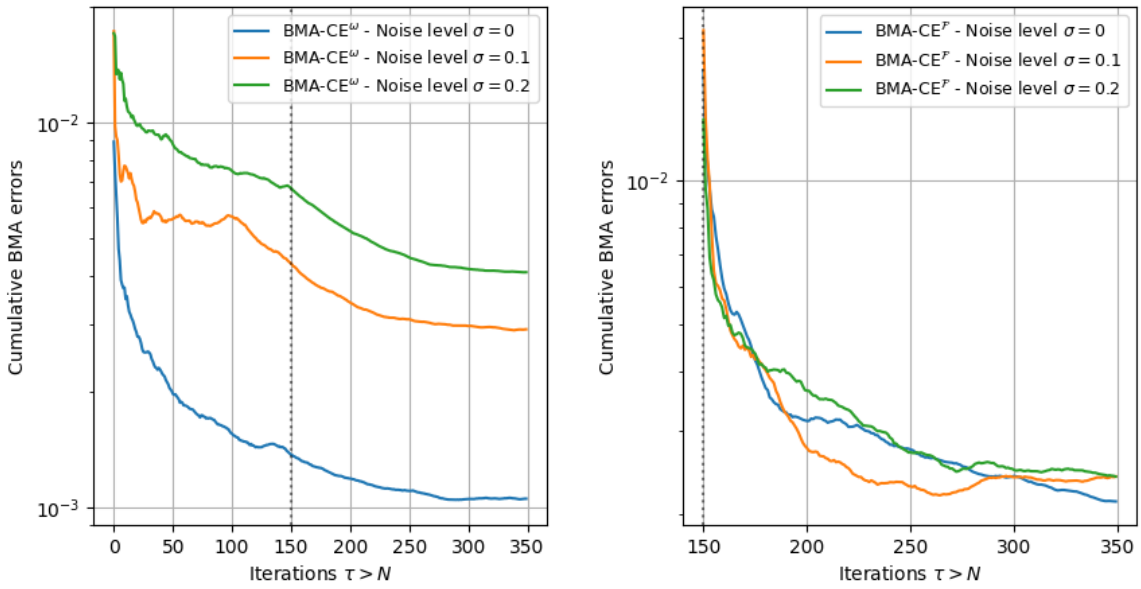

图 10：涡量物理信息修复问题：整个采样迭代过程中、不同噪声水平下的贝叶斯模型平均累积误差诊断指标，定义见式 (35)。左图为涡量场预测的贝叶斯模型平均累积误差，右图为满足式 (33) 的偏微分方程残差 $\mathcal{F}$ 的贝叶斯模型平均累积误差。竖直虚线标示序贯训练中引入偏微分方程约束的时刻。

4.1 稀疏且含噪数据下的图像修复问题

在磁共振成像（MRI）或计算机断层扫描（CT）医学成像中，修复问题日益受到关注，因为深度学习方法有望减少伪影并恢复缺失信息 [30, 3, 51]。尽管通常的修复框架仅在图像处理中纳入测量数据，Zheng 等人通过将潜在物理规律作为间接测量纳入其中，研究了该问题的物理信息版本 [63]。本节处于相同背景下——其目标是基于稀疏且含噪测量推断完整的流场重建，同时在一些互补的配点上施加偏微分方程约束。

这里以二维速度 $\mathbf{u} = (u, v)$ 和涡量 ω 形式，写出狭窄流的控制方程，满足如下不可压稳态纳维–斯托克斯方程：

$$(\mathbf{u} \cdot \nabla)\omega = Re^{-1}\Delta\omega, \quad \text{in } \Omega \tag{32}$$

或等价地

$$u\frac{\partial \omega}{\partial x} + v\frac{\partial \omega}{\partial y} = \frac{1}{Re}\Delta\omega, \quad \text{in } \Omega \tag{33}$$

其中，Re 表示无量纲雷诺数，ω 为涡量场 $\omega = \dfrac{\partial v}{\partial x} - \dfrac{\partial u}{\partial y}$，不可压缩条件保证 $\nabla \cdot \mathbf{u} = 0$。我们考虑二维狭窄空间域 $\Omega \subset [0, 10] \times [0, 1]$，并假设两类不同的边界条件：1）狭窄段上下壁面，记为$\partial\Omega_1$, ，施加无滑移条件，即 $\mathbf{u}_{\partial\Omega_1} = 0$ 且 $\omega = (\nabla \times \mathbf{u})_{\partial\Omega_1}$；2）入口和出口边界，记为 $\partial\Omega_2$, ，沿主流方向的速度分别满足给定的抛物线型分布和诺伊曼条件。这些边界条件详见 DC PSE 文章 [7]第 4.4 节。我们还先假设雷诺数已知，并根据计算流体力学模拟将其设为 $Re = 200$，因此这里需要推断的参数集合 Θ 仅限于神经网络权重和偏置。

测量数据集 $\mathcal{D}$ 由含噪涡量数据组成，分别定义在集合 $\partial\Omega_1$ 和 $\partial\Omega_2$，上的 $\mathcal{D}^{\partial_1}$ 和 $\mathcal{D}^{\partial_2}$，如式 (12) 所示；此外，还包括覆盖范围不足的 1282 个内部配点 $\mathcal{D}^{\omega}$。

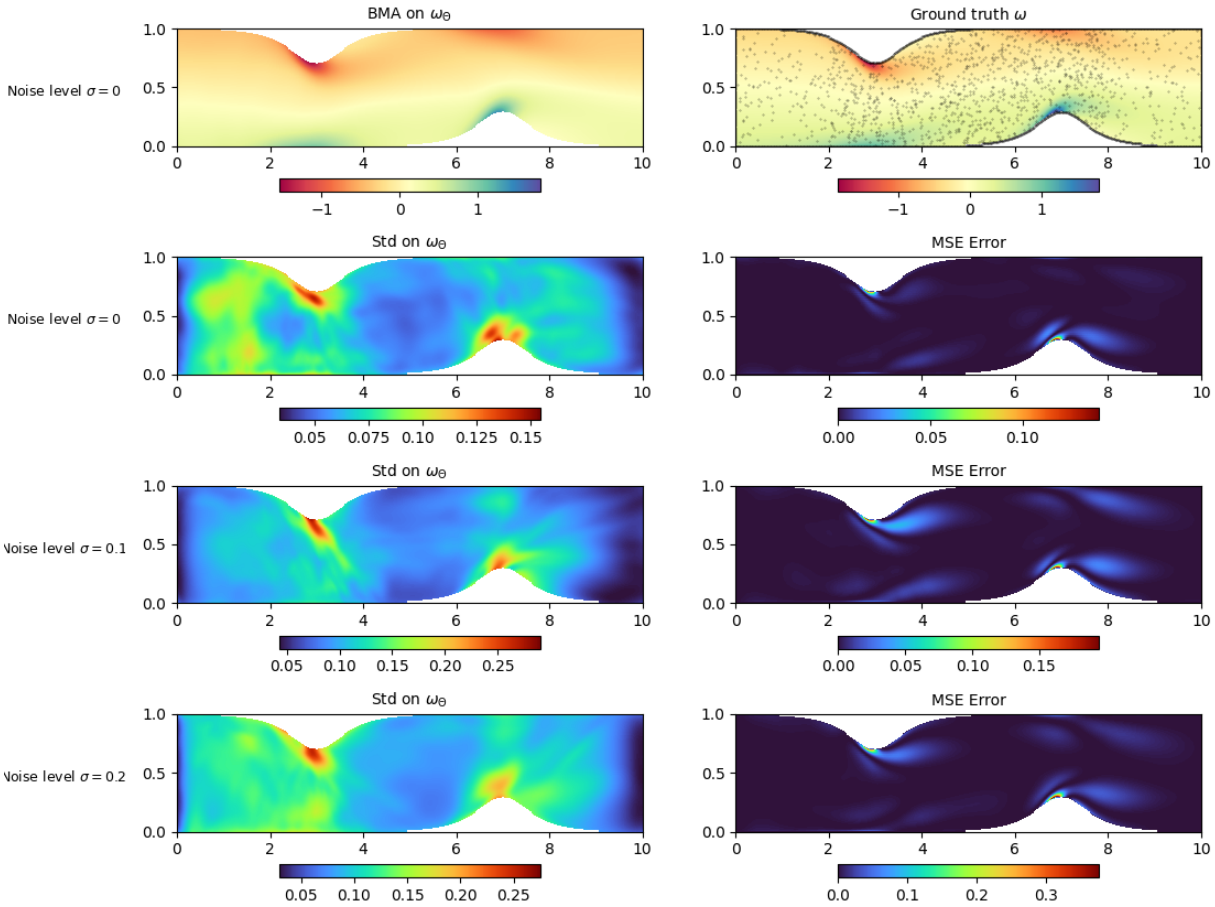

图 11：物理信息引导的图像修复问题：非对称狭窄中无噪声情况下涡量场的 BMA 预测 ω_Θ 与真实解 ω（上图）的对比。精确场上的黑点对应于数据集 $\mathcal{D}$的训练测量值。下方各行展示了不同噪声水平（σ =0，0.1，0.2）下预测涡量场 ω_Θ 的不确定性标准差（Std）和均方误差（MSE）的对比。

噪声水平 \ λ_k	λ_0	λ_1	λ_2	λ_3
$\sigma = 0$	1	0.46	0.88	0.51
$\sigma = 0.1$	1	0.19	0.35	0.29
$\sigma = 0.2$	1	0.12	0.39	0.27

噪声水平 \ $\widetilde{\sigma}_k$	$\widetilde{\sigma}_0$	$\widetilde{\sigma}_1$	$\widetilde{\sigma}_2$	$\widetilde{\sigma}_3$
$\sigma = 0$	1	1.47	1.07	1.40
$\sigma = 0.1$	1	2.29	1.69	1.86
$\sigma = 0.2$	1	2.89	1.60	1.92

表2：物理信息引导的图像修复问题中各个噪声水平下的最终 λ_k 权重参数（上方各行）。权重调整后得到的有效 $\widetilde{\sigma}_k$ 标准差，并按 $\widetilde{\sigma}_k = \sqrt{1/\lambda_k}$计算各噪声水平下的有效标准差（下方各行）。这体现了有效标准差对噪声幅度的整体适应性，以及各任务对噪声水平的敏感性。尤其是，与 λ_1 相关的壁面边界条件具有最高的噪声敏感性。

Ω上完整涡量场重建所需全部数据的2%。最后，我们将 $\mathcal{D}^{\Omega}$ 数据集定义为6408个内部点，占整个重建数据场的6%；在物理约束修复公式中，我们要求满足偏微分方程（33）。多势能随后定义为:

$$
\begin{aligned}
U(\Theta) = \frac{\lambda_0}{2\sigma_0^2}\left\|\omega_\Theta - \omega\right\|_{\mathcal{D}^\omega}^2 + \frac{\lambda_1}{2\sigma_1^2}\left\|\omega_\Theta - \omega_{|\partial\Omega_1}\right\|_{\mathcal{D}^{\partial_1}}^2 + \frac{\lambda_2}{2\sigma_2^2}\left\|\omega_\Theta - \omega_{|\partial\Omega_2}\right\|_{\mathcal{D}^{\partial_2}}^2 \\
+ \frac{\lambda_3}{2\sigma_3^2}\left\|u_n\frac{\partial\omega_\Theta}{\partial x} + v_n\frac{\partial\omega_\Theta}{\partial y} - \frac{1}{Re}\Delta\omega_\Theta\right\|_{\mathcal{D}^\Omega}^2 + \frac{1}{2\sigma_\Theta^2}\|\Theta\|_{R_p}^2
\end{aligned}
\tag{34}
$$

其中， u_n 和 v_n 是在 $\mathcal{D}^{\Omega}$ 集合上对速度场的含噪评估值。随后，我们采用序贯训练，在第二个采样阶段加入偏微分方程残差似然项，因此AW-HMC参数先设为 $N = 50$ 和 $N_s = 200$，再设为 $N = 50$ 和 $N_s = 250$；蛙跳路径长度为 $L = 150$，时间步长为 $\delta t = 5\mathrm{e}{-4}$。与先前的基准测试一样，我们将所有 $\sigma_\bullet$ 设为1，并假设神经网络参数的先验服从标准差为 $\sigma_\Theta = 10$ 的中心正态分布。该神经网络由4层组成，每层有32个神经元，并采用双曲正切激活函数。速度场和涡量的计算流体力学解（$\mathbf{u}$ ω）均受到加性高斯噪声的污染，使得 $\bullet_n = \bullet + \sigma\xi$，其中 $\xi \sim \mathcal{N}(0,\ \psi^2)$是一个向量，其各元素相互独立且服从均值为零、方差为 $\psi^2 = \mathrm{Var}\{\bullet\}$的同分布高斯随机数， σ 表示添加噪声的水平。

在这个物理信息引导的图像修复问题中，我们通过将贝叶斯模型平均收敛的概念扩展到偏微分方程残差，研究噪声水平 σ 对涡量场的 BMA预测以及物理约束的影响。因此，我们计算场 ω 和偏微分方程约束的 BMA-CE诊断，依据的是

$$
\begin{aligned}
\text{BMA-CE}^{\omega}(\tau) &= \left\| \frac{1}{\tau - N}\sum_{i=N}^{\tau} P\left(\omega_\Theta \,|\, x, \Theta^{t_i}\right) - \omega \right\|^2 \\
\text{BMA-CE}^{\mathcal{F}}(\tau) &= \left\| \frac{1}{\tau - N}\sum_{i=N}^{\tau} P\left(\mathcal{F}(\omega_\Theta) \,|\, x, \Theta^{t_i}\right) \right\|^2
\end{aligned}
\tag{35}
$$

其中 $\mathcal{F}(\omega_\Theta)$ 为对方程 (33) 中偏微分方程的求值。不同噪声水平下的对比曲线如图10所示，结果表明采样收敛后，最终的 BMA误差分别约为 $1.05\mathrm{e}{-3}$、$2.9\mathrm{e}{-3}$ 和 $4.08\mathrm{e}{-3}$，对应的噪声水平为 $\sigma = 0$，0.1 和 0.2。此外，我们还看到

偏微分方程残差约束的收敛与噪声水平无关，所有情况下最终的 BMA误差均约为 2e−3 。

为了补充对修复公式在恢复整个涡量场及其不确定性方面的性能量化，我们还采用了 Yao等人 [61]定义的预测区间覆盖概率（PICP）指标。该指标是后验近似的一种质量指标，用于评估真实观测值中落在预测区间95%范围内的比例，定义如下：

$$PICP = \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}_{(\omega_{\Theta}^{l})_i \le \omega_i \le (\omega_{\Theta}^{h})_i} \tag{36}$$

其中，ω_{Θ}^{l} 和ω_{Θ}^{h} 分别是涡量预测分布的第2.5和第97.5百分位数。此处，符号 N 指预测数据集中的观测总数，换言之，即计算域Ω的网格分辨率。在我们的应用中，该PICP指标表明，无论噪声水平如何，神经网络输出 ω_{Θ}的后验分布都覆盖了超过99%的涡量真实观测值。

我们还期望对 λ_k 采用自加权自适应，能够捕捉相对于 σ取值的噪声敏感性，以及各任务对噪声水平的内在敏感性，而无需对噪声水平估计施加任何先验信息。这是我们方法的关键所在，因为我们有意将式（34）中的 σ_k 与噪声幅度解耦，并依靠自加权策略量化二者相关的不确定性。相反，在应用中处理含噪测量数据时，研究人员通常假设各传感器的保真度已知，并据此设置标准差 σ_k 。也可以将其定义为需要推断的额外可学习参数。后一种做法通常会在在线学习中带来额外的计算成本，或者需要在离线学习中采用作为预训练手段的其他神经网络形式 [38]。相比之下，AW-HMC 方法的优势在于，其计算成本与经典 BPINNs-HMC 相近。此外，AW-HMC 通过着重探索帕累托前沿并采用最优积分时间，改善了收敛性。因此，它可以减少总体采样需求，使其成为一种在计算成本方面具有竞争力的策略。

图11所示结果证明了 AW-HMC 方法具有抗噪能力，并凸显了针对噪声和任务的敏感性考量（见表2）。我们首先注意到，随着噪声水平变化，lambda 值会自动调整，噪声增大时整体不确定性也随之增强。我们还观察到，不同任务对噪声的敏感性不同，因此不确定性的调整方式也各异。事实上，对比图11中涡量场的局部标准差可以发现，壁面边界条件对噪声最为敏感，因此这些区域的不确定性会自动增大。入口和出口边界则不太敏感，其不确定性随噪声水平变化的幅度较小。总之，这一应用表明，我们提出的新型自适应方法能够根据任务对噪声的内在敏感性自动调整权重及相应的不确定性，并使不确定性适应噪声幅度本身。

4.2 反问题：参数估计与潜在场恢复

作为第二个计算流体力学应用，我们考虑非对称且陡峭的狭窄几何中的多目标流动反问题。该问题旨在利用我们的自适应加权策略，同时估计流动状态参数并恢复隐藏场。这类考虑受到现实应用的启发，使用不完整或受污染的测量数据来尝试推导额外信息，而这些信息仍难以或无法直接获取。

Raissi 等人着重研究物理和生物医学问题，利用物理信息引导的深度学习[41, 42]，从流动可视化中提取隐藏的流体力学关注量。作者仅依赖满足不可压缩 Navier–Stokes 方程的被动标量浓度测量数据，推断外流和内流中的速度场与压力场。

在这一方 面，我们聚焦于狭窄流的速度 $\mathbf{u} = (u, v)$ 和压力 p 表述

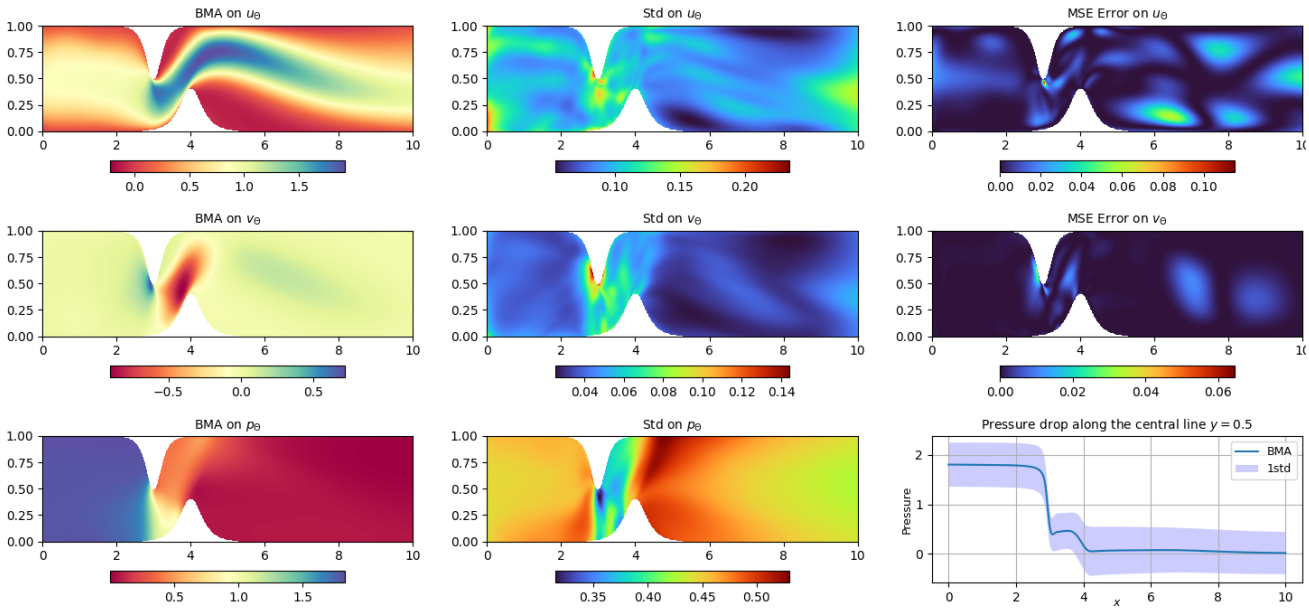

图12：计算流体力学逆问题：顶部显示非对称狭窄中的速度场 $\mathbf{u}_\Theta = (u_\Theta,\ v_\Theta)$ 的 BMA预测及其不确定性标准差（Std）和均方误差（MSE）。底部显示推断得到的潜在压力场的贝叶斯模型平均及其不确定性，并绘出了沿中心线 $y = 0.5$ 的压力演化，平均压降为 1。78。

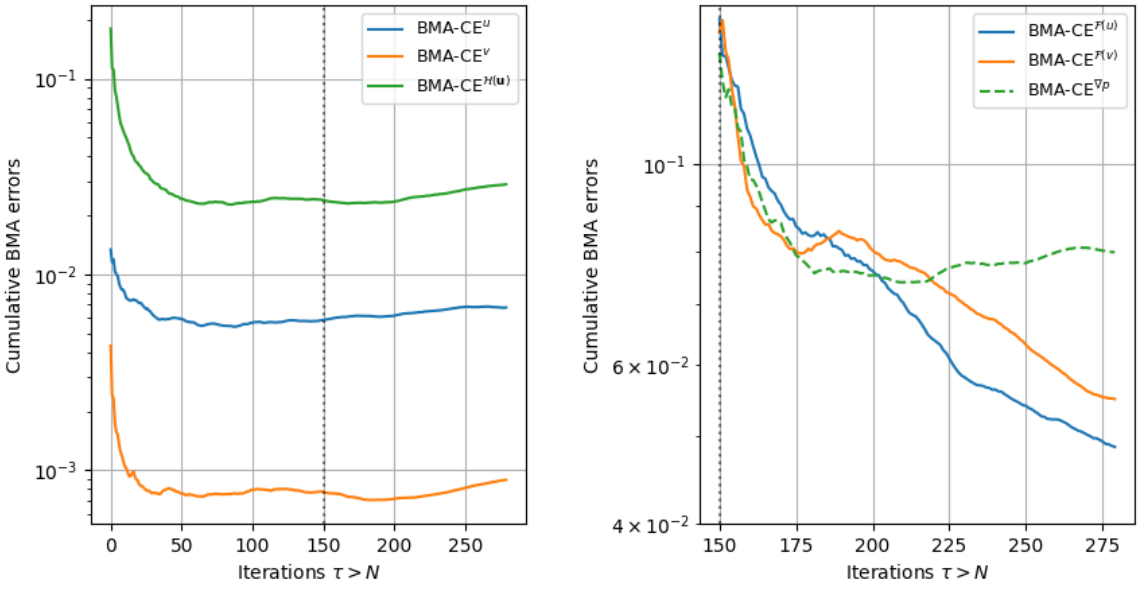

图13：计算流体力学逆问题：左侧为速度场分量 $\mathbf{u} = (u,\ v)$ 和无散条件 $\mathcal{H}(\mathbf{u})$ 在各采样迭代中的贝叶斯模型平均累积误差；右侧为偏微分方程残差 $\mathcal{F}(u)$ 和 $\mathcal{F}(v)$。点线表示压力梯度范数 BMA-CE误差的后验检验，其定义见方程 (40)。

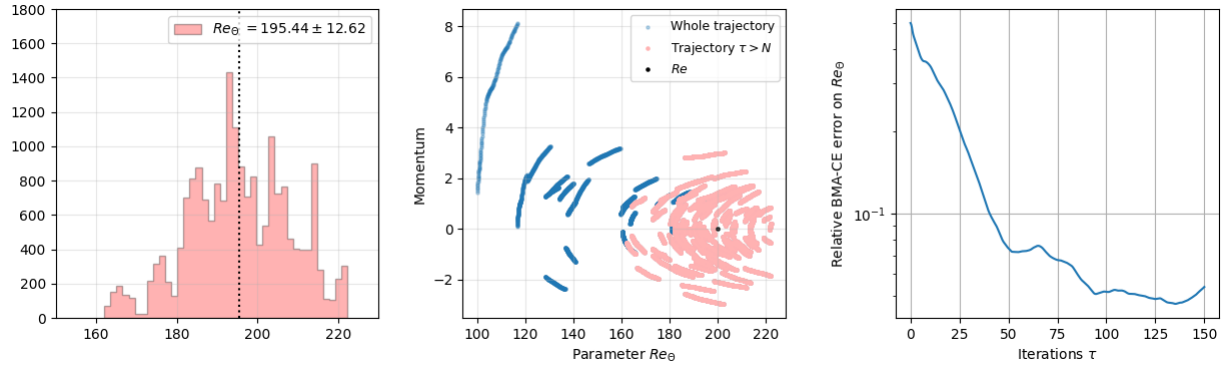

图14：CFD逆问题：从左到右，逆雷诺参数的边缘后验分布直方图、其在整个采样过程中的轨迹相图，以及使用式（39）定义的绝对相对范数计算得到的BMA-CE误差。在序贯训练第二步的所有 τ 次采样迭代中， Re_Θ 上的相对BMA-CE误差。

动力学使得连续性和动量控制的稳态方程可写为：

$$\begin{cases} (\mathbf{u}\cdot\nabla)\mathbf{u} = -\nabla p + Re^{-1}\Delta\mathbf{u}, & \text{in } \Omega \\ \nabla\cdot\mathbf{u} = 0, & \text{in } \Omega \end{cases} \tag{37}$$

在狭窄域 $\Omega \subset [0,10]\times[0,1]$上施加不可压缩条件。我们在壁面界面上施加黏附边界条件，使得 $\mathbf{u}_{\partial\Omega_1} = 0$，并分别施加以下入口/出口边界条件：

$$\begin{aligned} u = 4y - 4y^2,\ v = 0 \quad \forall (x,y) \in \{0\}\times[0,1] \\ \frac{\partial u}{\partial x} = 0,\ v = 0 \quad \forall (x,y) \in \{10\}\times[0,1]. \end{aligned} \tag{38}$$

直接数值模拟采用 DC-PSE公式 [7] ，并将雷诺数设为 $Re = 200$，与上一节相同。利用该模拟在 Ω 上以较高分辨率生成观测数据。随后， $\mathcal{D}$ 数据集由对 $\mathbf{u}$ 进行随机采样得到的部分测量组成，共包含9559个训练点，仅占整个目标分辨率的不到3%。同时采用相同的配点施加偏微分方程约束，记为 $\mathcal{F}(\mathbf{u}) := (\mathcal{F}(u)\ \mathcal{F}(v))$，以及无散条件 $\mathcal{H}(\mathbf{u})$。

最后，我们通过推断流动状态来建立反问题，并将雷诺数视为未知模型参数 $\Sigma = \{Re\}$。同时，我们处理多任务问题，根据速度场的部分测量值和流体流动动力学假设恢复潜在压力。特别是，在采样过程中不断调整压力场预测，使其梯度满足控制方程（37）。正如纳维-斯托克斯方程的性质通常所示，压力并非唯一确定；鉴于该场缺少精确的边界条件，因此压力只能确定到一个常数。随后，在图12中，我们在原始的较精细分辨率下恢复了各关注量（即速度和压力）的预测。与第3.3节一样，我们为物理参数选取对数正态先验分布，为神经网络参数选取相互独立的正态分布，并采用序贯训练方法，在第二个采样阶段纳入偏微分方程约束。

首先，通过计算速度场分量、偏微分方程约束以及按与方程 (35) 相同方式表示的不可压缩条件的 BMA-CE诊断，来验证推断。结果见图 13，显示各项均收敛至最终的 BMA误差，其尺度分别约为 BMA-CE$^u(N_s) = 6.4\text{e}{-3}$、BMA-CE$^v(N_s) = 1\text{e}{-3}$、BMA-CE$^{\mathcal{F}(\mathbf{u})}(N_s) = (4.2\text{e}{-2},\ 4.7\text{e}{-2})$ 和 BMA-CE$^{\mathcal{H}(\mathbf{u})}(N_s) = 2.9\text{e}{-2}$。随后，在图 12 中将速度场的贝叶斯模型平均预测与真实观测值进行比较，得到局部均方误差（MSE）；这些误差均包含在其不确定性范围内，且误差较大的区域表现出更高的标准差。PICP指标还可用于估计，后验分布对$\mathbf{u}_\Theta$的预测恢复了超过 95% 的速度场真实值。

对于逆参数，我们基于相对L1范数计算了 BMA累积误差，定义如下：

$$\text{BMA-CE}^{Re}(\tau) = \frac{\left| \frac{1}{\tau} \sum_{i=1}^{\tau} Re_{\Theta^{t_i}} - Re \right|}{|Re|}, \qquad \forall \tau = 1 ... N_s \tag{39}$$

其中， $Re_{\Theta^{t_i}}$ 表示由参数Θ^{t_i}表征的样本对应的雷诺数预测值。图14显示，该相对误差逐渐收敛，在采样结束时残差达到 5.4e$-$2。我们还在此展示了 Re_{Θ} 边缘后验分布的直方图及其在相空间中的轨迹，说明其在自适应步 $\tau < N$期间逐渐收敛至众数。事实上，我们的方法根据测量数据 $\mathcal{D}$推断出雷诺数估计值，该值与精确值一致，且预测区间为 $Re_{\Theta} \in [182.82, 208.06]$。

图12展示了推断得到的潜在压力场贝叶斯模型平均（仅确定到一个常数）及其不确定性；该结果能够捕捉到因狭窄几何形状陡峭而产生的显著压降，平均估计值为 1.78。事实上，Sun等指出，在对称几何形状中，随着狭窄几何形状变窄，此类压降会呈现非线性 [49], ，这与我们在非对称情形下得到的结果一致。由于本应用中的压力真实值未知，我们通过对压力梯度进行后验检验，补充验证反问题。为此，我们给出了压力恢复的PICP估计，其梯度范数对应的数值约为91%；此外，还引入以下关于压力BMA-CE误差的后验诊断：

$$\text{BMA-CE}^{\nabla p}(\tau) = \left\| \frac{1}{\tau - N} \sum_{i=N}^{\tau} \left| P\left(\nabla p_{\Theta} \,|\, x, \Theta^{t_i} \right) \right| - |\nabla p| \right\|^2 \tag{40}$$

其中， $|\cdot|$ 表示向量范数， ∇p 表示根据方程（37）计算的精确压力梯度。图 13 中以点线绘出了采样迭代过程中的结果，其残差误差达到 7.6e$-$2。这表明，真实值与采用自适应加权策略得到的预测压力梯度吻合良好。

总体而言，目前的 AW-HMC 方法依靠多任务采样，通过对速度场进行部分测量来识别流动状态，从而处理具有潜在场恢复任务且满足非线性物理偏微分方程约束的复杂流动反问题。

5 结语

贝叶斯物理信息神经网络（BPINN）近期已成为一种颇具前景的数据同化深度学习框架，也是用于不确定性量化（UQ）的宝贵工具 [59]。这使得我们能够在贝叶斯推断框架中，通过马尔可夫链蒙特卡洛（MCMC）采样，将物理信息神经网络（PINN）的预测能力与不确定性量化相结合。借助这一方法，即使在数据稀疏且含噪、并受物理模型约束的情况下，也能量化预测结果的置信度；这对于复杂系统中的应用尤其具有吸引力。为此，哈密顿蒙特卡洛因能够高效探索高维目标分布，已成为一种强大的 MCMC 采样器 [4]。借助该方法，贝叶斯物理信息神经网络将物理信息神经网络的应用拓展到了贝叶斯不确定性量化设定。

正如我们在此所展示的，贝叶斯物理信息神经网络与物理信息神经网络具有相似的失效模式：多目标代价函数在贝叶斯物理信息神经网络中转化为多势能采样问题。这带来了与标准物理信息神经网络中相同的困难，即如何平衡推断任务并高效探索帕累托前沿[43]。我们在一个索博列夫训练基准中对此进行了说明，该基准容易出现刚性、尺度差异悬殊以及任务特定梯度消失等问题。我们强调，贝叶斯物理信息神经网络对势能中 λ 权重的选择十分敏感，这可能导致有偏预测或无效采样。因此，标准加权策略在多尺度问题和多任务推断中似乎效率不高；与此同时，以可复现且可靠的方式手动调节权重也被证明难以持续。近期提出的替代方法 [38] 需要额外的超参数调优，或使用 GAN 对权重进行预训练，代价是增加计算复杂度。此外，以往的方法主要

侧重于测量噪声估计，而未纳入物理模型错误设定这一同样关键的考虑因素，尤其是在不确定性量化建模为目标时。

因此，稳健的自动加权策略对于将贝叶斯物理信息神经网络应用于多尺度和多任务（逆）问题，以及提高不确定性量化估计的可靠性至关重要。因此，我们提出了 AW-HMC 贝叶斯物理信息神经网络公式，为标准贝叶斯物理信息神经网络提供了一种即插即用的自动自适应加权策略。正如所展示的基准测试所示，AW-HMC 能有效处理多势能采样、能量守恒不稳定性、尺度差异以及数据中的噪声。

我们已经证明，所提出的策略能够确保加权后验分布适于探索帕累托前沿，并通过基于逆狄利克雷加权 [29]适当调整 λ 权重，实现均衡采样。因此，这些权重可直接解释为训练不确定性，其衡量依据是各任务训练梯度的方差。由此得到的权重会根据模型进行调整，从而使贝叶斯物理信息神经网络哈密顿蒙特卡洛采样中的多势能对模型最不敏感。由于采样聚焦于帕累托前沿，这一方法提升了收敛性、鲁棒性和不确定性量化的可靠性。这使贝叶斯物理信息神经网络能够有效且高效地处理多任务不确定性量化。

所提出的方法在计算效率方面也优于以往方法，因为它不需要额外的超参数或网络层。这也确保了蛙跳训练中的最优积分时间和收敛性。该方法避免了时间步长趋近于零或变得非常小，从而规避了无转向采样（NUTS）中常见的问题：为了避免具有 HMC 不稳定性特征的病态发散轨迹，时间步长往往会变得极小。本文方法改善了这一情况，因为时间步长不再需要满足所有严格的尺度要求来确保能量守恒。因此，它缩短了总体积分时间并减少了所需的样本数量，同时兼顾了计算效率和对采样不稳定性的鲁棒性。

我们的结果还表明，AW-HMC 能够减少采样中的偏置，因为它可以根据每个项对噪声或固有尺度的敏感性，自动调整 λ 参数以及相应的不确定性估计。在经典方法中，手动权重调节所引入的偏置和隐式先验会阻碍这一点。事实上，我们证明了，在一个多尺度问题中，即使针对尺度采用完全独立的先验分布，本文方法也能有效捕获不同数量级的逆参数。以往的方法需要通过对这些参数施加先验分布来处理这一问题，而这些先验分布本身已经依赖于适当的尺度。否则，经典的贝叶斯物理信息神经网络公式很容易失效。所提出的自适应加权策略彻底避免了这些问题，在多尺度逆问题中表现出显著更好的性能。

我们已在现实世界的应用中展示了这一点，例如计算流体力学（CFD）中非对称二维几何结构内不可压缩流的研究。我们展示了使用 AW-HMC 贝叶斯物理信息神经网络进行 CFD 修复，并研究了噪声对多势能的影响。这凸显了当前方法对于含噪测量的鲁棒性，同时也体现了其自动调整 λ 值、准确估计噪声水平本身的能力。在这一意义上，我们证明了随着噪声增加，不确定性也会增强，而无需对噪声水平本身设置任何先验，并且能够捕捉不同任务对噪声的内在敏感性。总体而言，这为自动处理由未知异方差噪声分布测量结果产生的多保真问题提供了一种有效替代方案。

综上所述，当前结果表明，贝叶斯物理信息神经网络是科学数据同化的一种很有前景的方法。它们如今有潜力有效处理多尺度和多任务推断问题，将不确定性量化与物理先验相结合，并应对稀疏噪声数据问题。在所有这些方面，所提出的方法都能确保高效的帕累托前沿探索，正确缩放多尺度刚性动力学，并从数据中推导出无偏的不确定性信息。我们的方法对噪声分布、不同问题尺度以及权重仅作出极少假设，同时具有较高的计算效率。这将贝叶斯物理信息神经网络的应用扩展到了更复杂的现实世界问题，而这些问题此前并不容易直接处理。

我们预计，多孔介质研究、系统生物学和地球科学等领域尤其能从这些改进中受益。在这些领域，贝叶斯物理信息神经网络如今为数据驱动建模带来了良好前景。它们有望支持并推动直接从样本图像中提取和预测形态几何结构[36, 47], 、材料性质的上尺度和粗粒化 [1] 以及物理性质 [44] 。然而，从不完善的图像中捕捉这些特征仍然颇具挑战，且通常存在不确定性，例如不可避免的成像伪影所造成的不确定性。这就需要开发

基于均质化的方法 [18]，以衔接不同尺度并量化这些不确定性 [35]，或者利用数据同化来弥补图像中部分缺失的信息。本文提出的结合 AW-HMC 的贝叶斯物理信息神经网络方法为此提供了一种潜在解决方案。

A 逆狄利克雷加权方差的上界

第 3.1 节提出的逆狄利克雷自适应加权哈密顿蒙特卡洛算法，可确保在整个采样过程中，多势能项的梯度分布保持均衡，如下文的联合方差所示：

$$\gamma^2 := \mathrm{Var}\{\lambda_k \nabla_\Theta \mathcal{L}_k\} \simeq \min_{t=0,\dots,K}(\mathrm{Var}\{\nabla_\Theta \mathcal{L}_t\}), \quad \forall k = 0, ..., K. \tag{41}$$

本节通过一个一般情形说明，γ^2 存在上界，并受可靠性准则控制；该准则取决于预测误差或偏微分方程残差、其均值相对于 Θ 的变异离散程度，以及 $\sigma_\bullet$ 值的设定。

首先，这表明必须恰当设定 σ 参数，以避免任务梯度分布出现有偏且不均衡的情况，因为这些参数会关键性地、且任意地影响梯度分布的控制。这也凸显出手动调整 σ 值可能极其敏感，在实践中难以实现。因此，在本文展示的所有应用中，我们选择对这些参数进行统一设定，而依靠 λ 自动调整，以确保帕累托前沿得到高效探索等。由此可见，这些标准差参数对每个梯度分布施加了严格约束（相对于 Θ 而言），进而影响每项任务的不确定性。

为简便起见，我们采用双任务采样，其中一个数据拟合项来自场 u，另一个是偏微分方程约束，记作 $\mathcal{F}$，因此数据集分解为 $\mathcal{D} = \mathcal{D}^u \cup \mathcal{D}^\Omega$，符号沿用第2.2节中的定义。由此，多势能可简化为：

$$U(\Theta) = \frac{\lambda_0}{2\sigma_0^2} \left\| u_\Theta - u \right\|^2_{\mathcal{D}^u} + \frac{\lambda_1}{2\sigma_1^2} \left\| \mathcal{F}(u_\Theta) \right\|^2_{\mathcal{D}^\Omega} + \frac{1}{2\sigma_\Theta^2} \left\| \Theta \right\|^2 := \sum_{k=0}^{K+1} \lambda_k \mathcal{L}_k(\Theta) \tag{42}$$

为便于演示，我们保留 σ 符号，并将 Θ 限制为神经网络参数，即使在反问题范式下，以下结论仍然成立。需要说明的是，用于训练 $\mathcal{D}^u$ 的测量数据可能与施加偏微分方程约束 $\mathcal{D}^\Omega$ 的配点不同，其相应数量分别记为 N^u和 N^Ω。根据第2.2节中的符号，两个任务的势能梯度分别为：

$$\begin{aligned} \frac{\partial \mathcal{L}_0}{\partial \Theta_j}(\Theta) &= \frac{1}{\sigma_0^2 N^u} \sum_{i=0}^{N^u} \Big(u_\Theta(x_i) - u_i \Big) \frac{\partial u_\Theta}{\partial \Theta_j}(x_i) \\ \frac{\partial \mathcal{L}_1}{\partial \Theta_j}(\Theta) &= \frac{1}{\sigma_1^2 N^\Omega} \sum_{i=0}^{N^\Omega} \mathcal{F}\Big(u_\Theta(x_i) \Big) \frac{\partial \mathcal{F}(u_\Theta)}{\partial \Theta_j}(x_i) \end{aligned} \tag{43}$$

对于 $\Theta \in \mathrm{R}^p$，因此我们可以根据这些梯度分解方差 $\mathrm{Var}_\Theta\big[\nabla_\Theta \mathcal{L}_k\big]$，$k = 0$，$1$。为此，我们首先计算它们关于 Θ 的均值，分别得到

$$\mathbb{E}_\Theta\big[\nabla_\Theta \mathcal{L}_0\big] = \frac{1}{N^p} \sum_{j=0}^{N^p} \frac{\partial \mathcal{L}_0}{\partial \Theta_j}(\Theta) = \frac{1}{\sigma_0^2 N^u} \sum_{i=0}^{N^u} \Big(u_\Theta(x_i) - u_i \Big) \mathbb{E}_\Theta\big[\nabla_\Theta u_\Theta\big](x_i) = \frac{1}{\sigma_0^2} \mathbb{E}_{\mathcal{D}^u}\big[(u_\Theta - u) \mathbb{E}_\Theta\big[\nabla_\Theta u_\Theta\big] \big] \tag{44}$$

and

$$\mathbb{E}_\Theta\big[\nabla_\Theta \mathcal{L}_1\big] = \frac{1}{\sigma_1^2} \mathbb{E}_{\mathcal{D}^\Omega}\big[\mathcal{F}(u_\Theta) \mathbb{E}_\Theta\big[\nabla_\Theta \mathcal{F}(u_\Theta)\big] \big] \tag{45}$$

当 $\mathcal{F}$为线性时，采用特殊配置 $\nabla_\Theta \mathcal{F}(u_\Theta) = \mathcal{F}(\nabla_\Theta u_\Theta)$。最后，我们可以将其扩展到

方差计算如下：

$$\begin{aligned}
\mathrm{Var}_{\Theta}\left[\nabla_{\Theta}\mathcal{L}_0\right] &= \frac{1}{N^p}\sum_{j=0}^{N^p}\left(\frac{\partial\mathcal{L}_0}{\partial\Theta_j} - \mathbb{E}_{\Theta}\left[\nabla_{\Theta}\mathcal{L}_0\right]\right)^2 \\
&= \frac{1}{N^p(N^u\sigma_0^2)^2}\sum_{j=0}^{N^p}\left[\sum_{i=0}^{N^u}\Big(u_{\Theta}(x_i)-u_i\Big)\left(\frac{\partial u_{\Theta}}{\partial\Theta_j}(x_i) - \mathbb{E}_{\Theta}\left[\nabla_{\Theta}u_{\Theta}\right](x_i)\right)\right]^2 \\
&= \frac{1}{(N^u\sigma_0^2)^2}\sum_{i=0}^{N^u}\sum_{k=0}^{N^u}\Big(u_{\Theta}(x_i)-u_i\Big)\Big(u_{\Theta}(x_k)-u_k\Big)\mathrm{Cov}_{\Theta}\left[\nabla_{\Theta}u_{\Theta}(x_i),\nabla_{\Theta}u_{\Theta}(x_k)\right] \\
&\leqslant \frac{1}{\sigma_0^4}\|u_{\Theta}-u\|^2_{\infty,\mathcal{D}^u}\mathrm{Cov}_{\Theta}\left[\frac{1}{N^u}\sum_{i=0}^{N^u}\nabla_{\Theta}u_{\Theta}(x_i),\frac{1}{N^u}\sum_{k=0}^{N^u}\nabla_{\Theta}u_{\Theta}(x_k)\right] \\
&= \frac{1}{\sigma_0^4}\|u_{\Theta}-u\|^2_{\infty,\mathcal{D}^u}\mathrm{Var}_{\Theta}\left[\mathbb{E}_{\mathcal{D}^u}\left[\nabla_{\Theta}u_{\Theta}\right]\right]
\end{aligned} \tag{46}$$

它为数据拟合项的梯度方差提供了一个上界。然后，我们以相同的方式得到偏微分方程约束界：

$$\mathrm{Var}_{\Theta}\left[\nabla_{\Theta}\mathcal{L}_1\right] \leqslant \frac{1}{\sigma_1^4}\|\mathcal{F}(u_{\Theta})\|^2_{\infty,\mathcal{D}^{\Omega}}\mathrm{Var}_{\Theta}\left[\mathbb{E}_{\mathcal{D}^{\Omega}}\left[\nabla_{\Theta}\mathcal{F}(u_{\Theta})\right]\right]. \tag{47}$$

符号$\|\cdot\|_{\infty,\mathcal{D}^{\bullet}}$此处表示由 $\mathcal{D}^{\bullet}$训练点构成的空间域上的离散 ℓ^{∞} 范数，而 $\mathrm{E}_{\mathcal{D}^{\bullet}}$ 表示相应数据集上的空间均值。因此，各任务的梯度方差由交叉复数分量 $\mathrm{Var}_{\Theta}\mathrm{E}_{\mathcal{D}^{\bullet}}$ 控制，该分量可解释为敏感性项，用于评估梯度下降方向相对于 Θ 的离散程度，并对空间取平均。最后，由于 $\sigma_{\bullet}$ 值均设为 1 以避免有偏采样，这意味着 λ值的计算方式可确保梯度分布的联合方差有界：

$$\gamma^2 \leqslant \min\left\{\|u_{\Theta}-u\|^2_{\infty,\mathcal{D}^u}\mathrm{Var}_{\Theta}\left[\mathbb{E}_{\mathcal{D}^u}\left[\nabla_{\Theta}u_{\Theta}\right]\right],\ \|\mathcal{F}(u_{\Theta})\|^2_{\infty,\mathcal{D}^{\Omega}}\mathrm{Var}_{\Theta}\left[\mathbb{E}_{\mathcal{D}^{\Omega}}\left[\nabla_{\Theta}\mathcal{F}(u_{\Theta})\right]\right]\right\} \tag{48}$$

这突显了权重会根据最可能的任务进行调整这一事实，从而提高不确定性量化的可靠性。上述计算可直接扩展到更复杂的多势能项，用于直接和逆向现实世界问题，至此我们的分析告一段落。

B 二维索博列夫训练基准

我们将第3.2节中使用的索博列夫基准扩展到二维训练，并采用梯度和拉普拉斯算子，使目标泛函具有以下形式：

$$u(x,y) = \sum_{i=1}^{N_{rep}} A_x^i\cos\left(2\pi L^{-1}l_x^i x + \phi_x^i\right)A_y^i\sin\left(2\pi L^{-1}l_y^i y + \phi_y^i\right) \tag{49}$$

这使我们能够处理各种形状复杂度和尖锐界面，以及高阶导数引入的刚性。我们将定义域大小设为 $L = 2\pi$，重复次数设为 $N_{rep} = 5$；参数 A_x 和 A_y 分别从区间 $[-2,2]$中独立均匀采样， ϕ_x和 ϕ_y则从 $[0,2\pi]$中独立均匀采样。为了处理多种形状复杂度，我们考虑一系列参数l，使局部长度尺度 l_x和 l_y从集合 $\{1,2,...,l\}$中随机采样。二维空间域 $[0,2\pi]^2$ 由分辨率为 256×256的均匀网格覆盖，并结合随机选取的训练点。随后，我们通过设置 l的不同取值，研究函数复杂度以及训练点数量对采用AW-HMC方法得到的贝叶斯模型平均的影响。

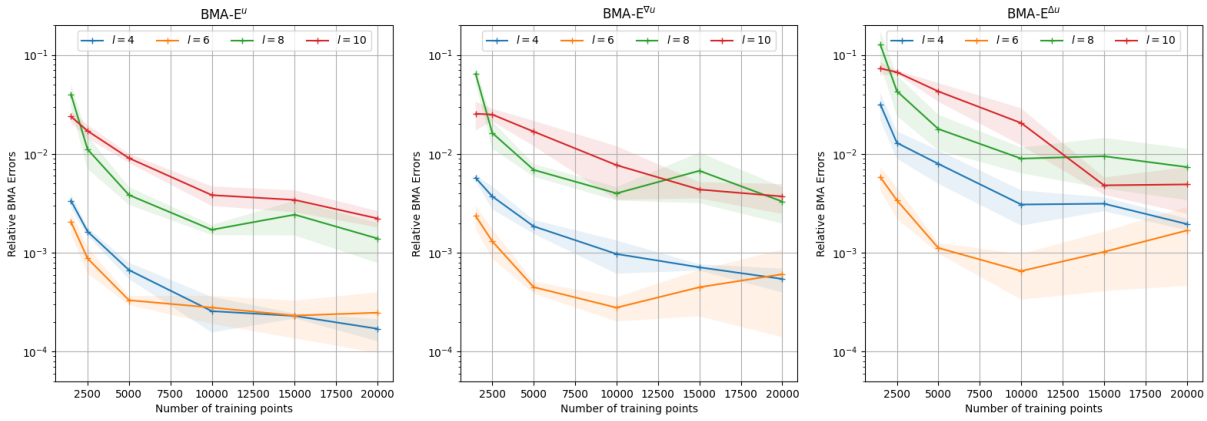

图 15：二维索博列夫训练基准：根据式（27）定义的相对 BMA 误差与训练点数量的关系图，展示了由 l 的不同取值所引起的各种形状复杂度。训练点数量逐步增加，直至达到整个数据集的约 30%，即 20000 个训练点。

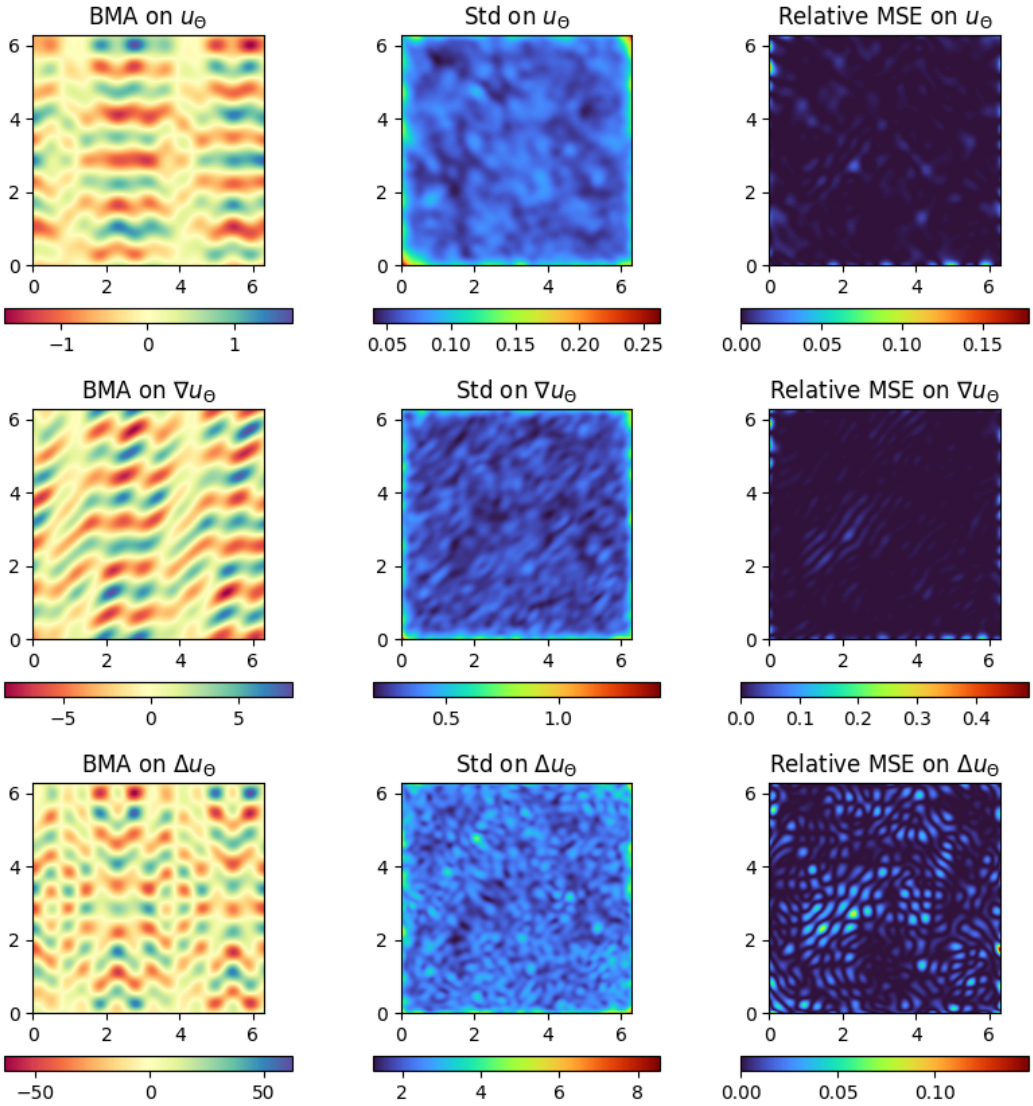

图 16：二维索博列夫训练基准：针对多目标势能的每一项，分别局部展示 BMA 预测、预测标准差和相对 BMA 误差。我们研究了一个有限情形 $l = 8$，其中训练点约占 15%。在整个定义域上取平均的全局相对 BMA 误差，其量级分别约为 $1.69e-3$、$3.42e-3$ 和 $5.6e-3$。

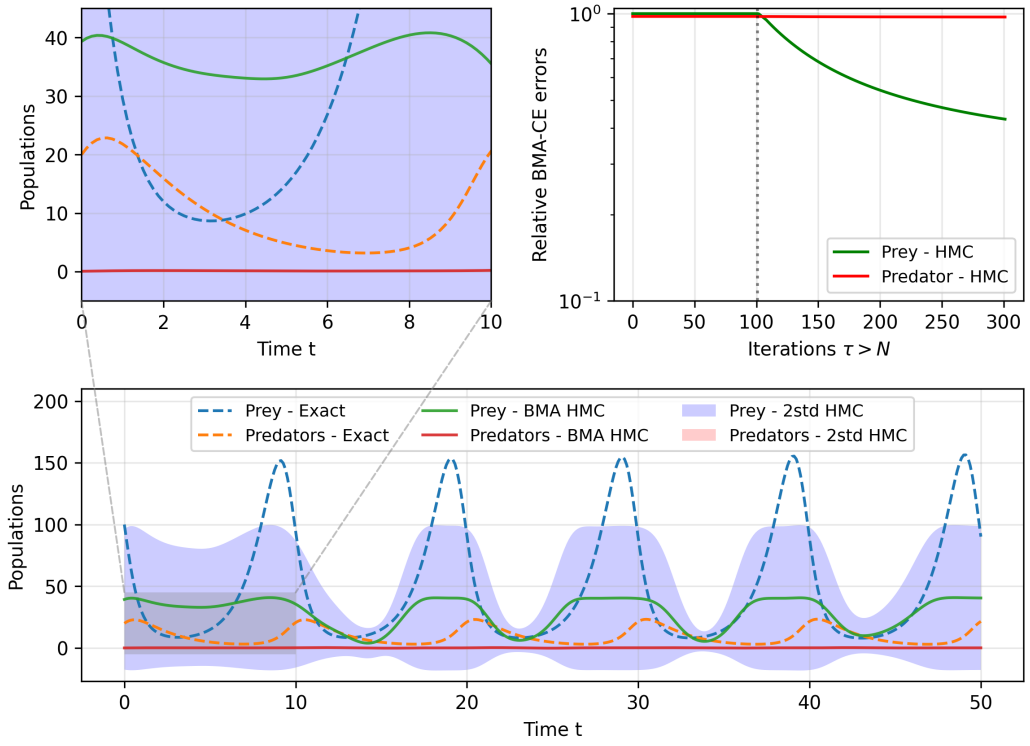

图 17：在第3.3节定义的 Lotka–Volterra 多尺度反问题上，经典 HMC 采用均匀加权时的失效模式。双物种种群随物理时间变化的 BMA 预测及其不确定性——左下图和左上图。整个采样迭代过程中相对 BMA-CE 误差的变化表明该方法未能收敛——右上图。

图 15 展示了整个基准设置的结果。随着训练点数量增加，结果呈现收敛趋势，且这一趋势与 l 值无关，尽管随着形状复杂度增加，相对 BMA 误差的上限也会升高。这些相对 BMA 误差根据方程 (27) 计算，是多次索博列夫采样重复实验的平均结果，各实验同时并行运行。事实上，为了应对采样变异性本身可能引起的随机过程，我们进行了多次实现，每次均从神经网络 Θ^{t_0} 和动量 r^{t_0} 参数的不同初始化开始，从而得到不同的采样实现。我们还可以考虑这些采样变异性，计算多次重复实验的标准差，如图 15 中的彩色带所示。

我们还在图16中分别展示了多势泛函各项的贝叶斯模型平均预测、基于整个采样过程中预测标准差的不确定性，以及在 $l = 8$ 情况下的相对贝叶斯模型平均误差；此处使用了10000个训练点，这些点在整个定义域内随机采样。结果表明，误差较大的边界壁附近不确定性更高，也凸显了我们的方法能够同时捕捉量级各异的复杂形状场。我们还要强调，这种二维索博列夫训练基准此前无法通过经典BPINNs-HMC形式实现。

C 常用方法在Lotka–Volterra逆问题上的失效

我们考虑第3.3节中介绍的Lotka–Volterra逆问题，以研究多尺度动力学对常用方法（即采用均匀加权的哈密顿蒙特卡洛和NUTS）的影响。采样参数和蛙跳参数相应地设为AW-HMC测试案例中的参数，其中 N 分别表示哈密顿蒙特卡洛和NUTS形式的预热步数与自适应步数。因此，我们比较不同采样器，并假设：1）它们的时间复杂度相同；2）我们不对逆参数缩放施加任何信息性先验。事实上，第一个条件表明，不同

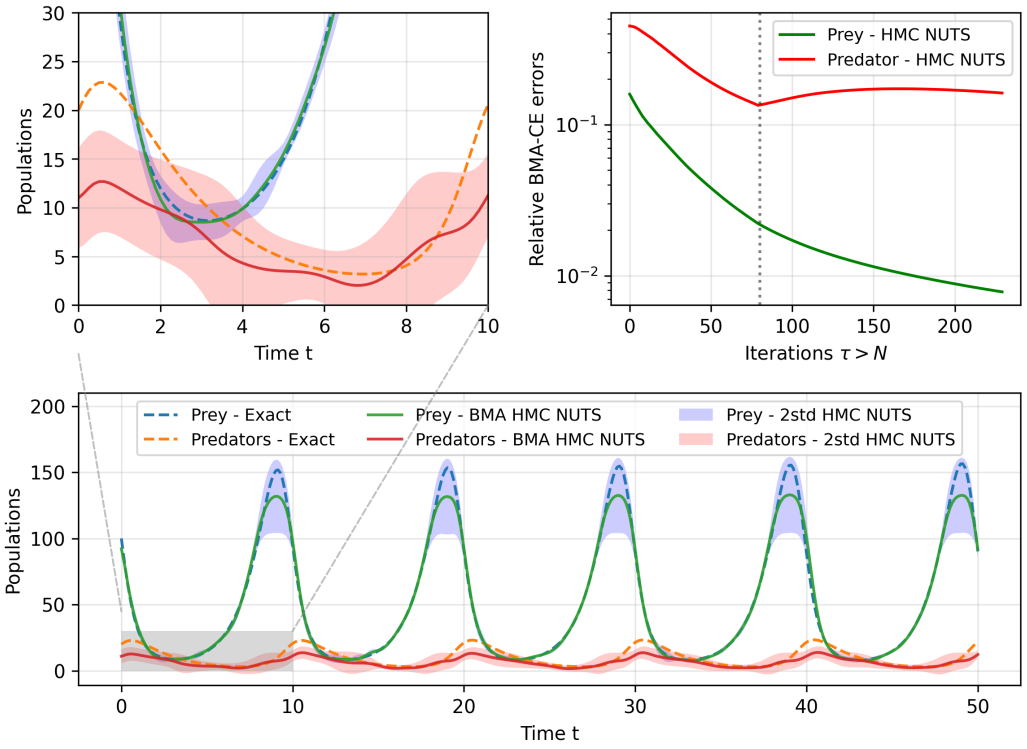

图18：在第3.3节定义的洛特卡-沃尔泰拉多尺度反问题中，采用 NUTS 自适应的 HMC 失效模式。双物种种群随物理时间变化的 BMA 预测及其不确定性——左侧上、下图。整个采样迭代过程中相对 BMA-CE 误差的变化，显示任务之间存在不平衡，且对猎物种群的自适应具有优先性——右上图。

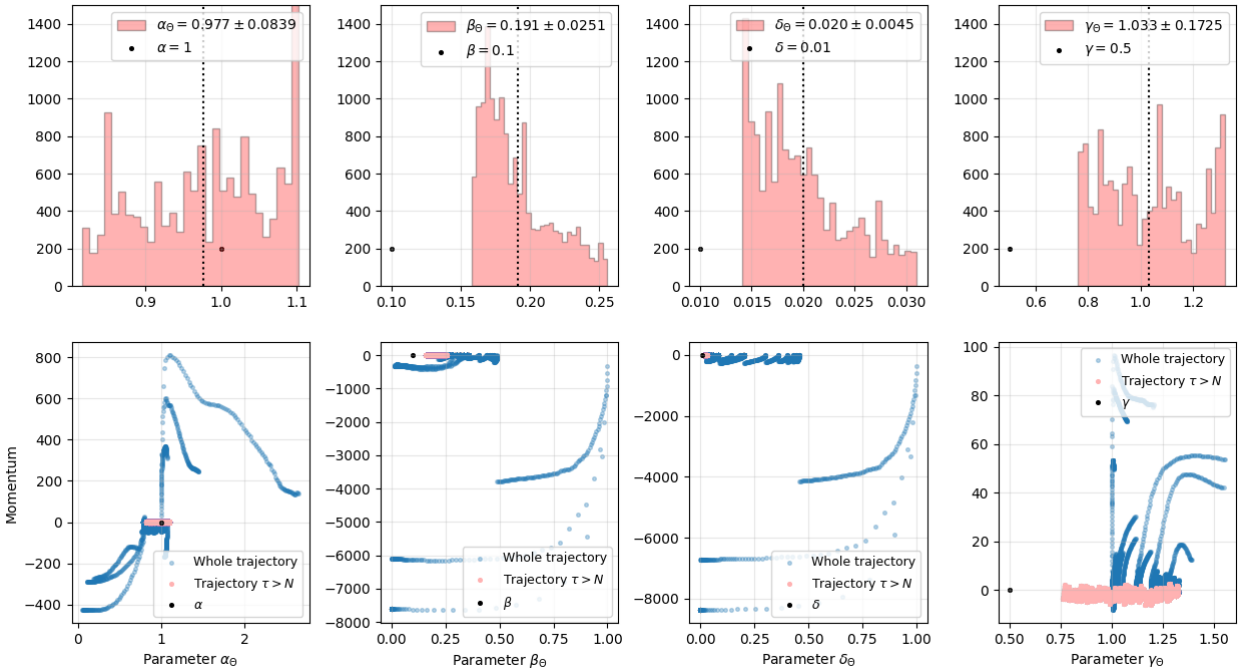

图19：洛特卡-沃尔泰拉多尺度推断中采用 NUTS 的 HMC 失效模式：逆参数边缘后验分布的直方图（上）以及其轨迹在整个采样过程中的相图（下）。图18中的有偏预测妨碍了对逆参数的正确推断，导致更新后的参数出现随机游走式病态行为。

蛙跳参数可能会改善这些传统方法的推断效果。然而，这意味着蛙跳时间步长 δt 会明显减小，从而减缓对能级的探索。因此，这些方法要获得合适的预测结果，要么需要通过增加 L 来延长积分时间，要么需要使用大量样本。需要提醒的是，除了后验分布采样效率低下这一问题之外，对多势能权重的不当选择还可能使采样器产生偏置，使其偏离帕累托前沿的探索。第二个假设源于这样一种意愿：在不预先了解不同尺度的情况下，处理多尺度动力学中的 UQ。这是由 Linka等人 [21] 提出的一项论断引发的，该论断指出，对尺度的敏感性会破坏 BPINNs-HMC 的性能。最后，我们还考虑了序贯训练，以便为不同方法之间的比较提供适当的基础。

结果显示，采用均匀加权的经典 HMC 未能收敛（图17——右上），并且任务之间也存在严重的不平衡。相对 BMA-CE 误差有效地刻画了捕食者种群相对于猎物种群极差的收敛性，这直接导致双物种种群的 BMA 预测效率低下（图17——下方和左上）。这种失效模式本质上是由于样本被大量拒绝（接受率低于1%），其原因在于蛙跳步过程中哈密顿轨迹未能保持守恒。因此，这证实了 BPINNs-HMC 范式在面对多尺度动力学引起的稳定性问题时缺乏鲁棒性。

NUTS 替代方法在这一多尺度反问题上同样难以收敛，且预测效果不足，尤其是对捕食者种群的预测。其原因并非大量样本被拒绝，而是时间步长大幅缩小，至自适应步结束时分别降至 $\delta t = 8.26\mathrm{e}{-5}$ 和 $2.81\mathrm{e}{-5}$——与 AW-HMC 相比，时间步长几乎缩小了十倍。相对 BMA-CE 误差（图 18 右上）表明，这种时间步长自适应有助于捕食者种群的收敛，因为捕食者种群似乎是最敏感的任务。这里的敏感性是指，势能相对于 Θ 的微小变化会对哈密顿能量守恒施加最强的约束。然而，这种时间步长自适应对捕食者种群并不理想，甚至会导致神经网络在序贯训练过程中出现低效遗忘。这使得种群演化预测产生误导（见图 18 下方和左上方），也使逆参数推断失败（图 19）。逆参数轨迹的相图展示了 NUTS 采样器在充分识别不同尺度所产生的众数方面面临的困难。总的来说，NUTS 采样器难以收敛至帕累托前沿，且在弱信息先验下对逆参数的推断具有误导性，这是因为它无法捕捉多尺度行为。

D 计算流体力学反问题中多势能的表征

第 4.2 节中定义的 CFD逆问题涉及恢复潜在压力场 p_Θ，以及基于速度场的部分测量结果确定流态参数——即雷诺数 Re_Θ。用于 AW-HMC采样的训练数据集 $\mathcal{D}$ 首先被分解为 9559 个随机采样的 $\mathbf{u}$ 测量值，分别定义了内部点集 $\mathcal{D}^{\mathbf{u}}$ 和边界点集 $\mathcal{D}^{\partial}$。相同的配点定义了 $\mathcal{D}^{\Omega}$，我们在此施加偏微分方程约束和无散条件。本问题所考虑的严重狭窄几何会在壁面界面处产生尖锐梯度。为确保对潜在压力和逆参数进行一致的推断，必须充分捕捉这些梯度。因此，我们在训练数据中补充了一些速度一阶导数的部分测量值。这使我们能够确保偏微分方程约束 (37) 中的对流项与速度数据一致，从而推断出相应的压力场。

因此，多势能写作：

$$
\begin{aligned}
U(\Theta) = & \frac{\lambda_0}{2\sigma_0^2}\left\|u_\Theta - u\right\|_{\mathcal{D}^{\mathbf{u}}}^2 + \frac{\lambda_1}{2\sigma_1^2}\left\|v_\Theta - v\right\|_{\mathcal{D}^{\mathbf{u}}}^2 + \frac{\lambda_2}{2\sigma_2^2}\left\|u_\Theta - u\right\|_{\mathcal{D}^{\partial}}^2 + \frac{\lambda_3}{2\sigma_3^2}\left\|v_\Theta - v\right\|_{\mathcal{D}^{\partial}}^2 \\
& + \frac{\lambda_4}{2\sigma_4^2}\left\|\partial_x u_\Theta - \partial_x u\right\|_{\mathcal{D}^{\mathbf{u}}}^2 + \frac{\lambda_5}{2\sigma_5^2}\left\|\partial_x v_\Theta - \partial_x v\right\|_{\mathcal{D}^{\mathbf{u}}}^2 + \frac{\lambda_6}{2\sigma_6^2}\left\|\partial_x u_\Theta - \partial_x u\right\|_{\mathcal{D}^{\partial}}^2 \\
& + \frac{\lambda_7}{2\sigma_7^2}\left\|\partial_x v_\Theta - \partial_x v\right\|_{\mathcal{D}^{\partial}}^2 + \frac{\lambda_8}{2\sigma_8^2}\left\|Re_\Theta^{-1}\Delta u_\Theta - (u_\Theta \partial_x u_\Theta + v_\Theta \partial_y u_\Theta) - \partial_x p_\Theta\right\|_{\mathcal{D}^{\Omega}}^2 \\
& + \frac{\lambda_9}{2\sigma_9^2}\left\|Re_\Theta^{-1}\Delta v_\Theta - (u_\Theta \partial_x v_\Theta + v_\Theta \partial_y v_\Theta) - \partial_y p_\Theta\right\|_{\mathcal{D}^{\Omega}}^2 \\
& + \frac{\lambda_{10}}{2\sigma_{10}^2}\left\|\nabla \cdot u_\Theta\right\|_{\mathcal{D}^{\Omega}}^2 + \frac{\lambda_{11}}{2\sigma_{11}^2}\left\|\partial_y u_\Theta - \partial_y u\right\|_{\mathcal{D}^{\mathbf{u}}}^2 + \frac{\lambda_{12}}{2\sigma_{12}^2}\left\|\partial_y u_\Theta - \partial_y u\right\|_{\mathcal{D}^{\partial}}^2 + \frac{1}{2\sigma_\Theta^2}\|\Theta\|_{R_{p+1}}^2
\end{aligned}
\tag{50}
$$

其中 记号 $\|\cdot\|$ 表示 $\mathcal{D}^{\bullet}$ 上的均方根范数，或 R^{p+} 上通常的欧几里得范数 [1].

参考文献

[1] Naif Alqahtani、Fatimah Alzubaidi、Ryan T. Armstrong、Pawel Swietojanski 和 Peyman Mostaghimi。从二维 X 射线图像预测多孔介质性质的机器学习方法。《石油科学与工程杂志》，184:106514，2020年。[2] Naif J. Alqahtani、Yufu Niu、Ying Da Wang、Traiwit Chung、Zakhar Lanetc、Aleksandr Zhuravljov 等。利用深度学习对碳酸盐岩 X 射线图像进行超分辨率分割。《多孔介质中的传输》，143(2):497–525，2022年。[3] Karim Armanious、Vijeth Kumar、Sherif Abdulatif、Tobias Hepp、Sergios Gatidis 和 Bin Yang。ipA-MedGAN：医学成像中任意区域的修复重建。载于*2020*年 *IEEE* 图像处理国际会议（*ICIP*），第3005–3009页，2020年。[4] Michael Betancourt。哈密顿蒙特卡洛的概念性介绍。*arXiv*，2018年。arXiv:1701.02434。[5] Michael Betancourt、Simon Byrne 和 Mark Girolami。优化哈密顿蒙特卡洛的积分器步长。*arXiv*，2014年。arXiv:1411.6669。[6] Michael Betancourt、Simon Byrne、Sam Livingstone 和 Mark Girolami。哈密顿蒙特卡洛的几何基础。《伯努利》，23(4):2257–2298，2017年。[7] George C. Bourantas、Bevan L. Cheeseman、Rajesh Ramaswamy 和 Ivo F. Sbalzarini。在欧拉无网格配置法中使用 DC PSE 算子离散化，可提高其在复杂几何结构中的鲁棒性。《计算机与流体》，136:285–300，2016年。[8] Steven L. Brunton、Joshua L. Proctor 和 J. Nathan Kutz。通过非线性动力系统的稀疏辨识，从数据中发现控制方程。《美国国家科学院院刊》，113(15):3932–3937，2016年。[9] Zhao Chen、Vijay Badrinarayanan、Chen-Yu Lee 和 Andrew Rabinovich。GradNorm：深度多任务网络中用于自适应损失平衡的梯度归一化。载于第*35*届国际机器学习会议论文集，第794–803页。PMLR，2018年。[10] Roberto Cipolla、Yarin Gal 和 Alex Kendall。利用不确定性对场景几何与语义任务的损失进行加权的多任务学习。载于*2018*年 *IEEE/CVF* 计算机视觉与模式识别会议，第7482–7491页，2018年。[11] Adam D. Cobb 和 Brian Jalaian。采用对称分裂扩展贝叶斯神经网络的哈密顿蒙特卡洛推断。载于第三十七届人工智能不确定性会议论文集，第675–685页。PMLR，2021年。[12] Wojciech M. Czarnecki、Simon Osindero、Max Jaderberg、Grzegorz Swirszcz 和 Razvan Pascanu。神经网络的索博列夫训练。载于《神经信息处理系统进展》，第30卷。Curran Associates, Inc.，2017年。[13] Marta D'Elia、Hang Deng、Cedric Fraces、Krishna Garikipati、Lori Graham-Brady、Amanda Howard 等。非均质多孔材料中的机器学习。*arXiv*，2022年。arXiv:2202.04137。[14] Salar Fattahi 和 Somayeh Sojoudi。数据驱动的稀疏系统辨识。载于*2018*年第*56*届*Allerton*通信、控制与计算年度会议（*Allerton*），第462–469页。IEEE，2018年。

[15] Alex Graves。神经网络的实用变分推断。载于《神经信息处理系统进展》，第24卷。Curran Associates, Inc.，2011年。[16] Matthew D. Hoffman 和 Andrew Gelman。No-U-Turn采样器：自适应设定哈密顿蒙特卡洛中的路径长度。《机器学习研究杂志》，15(1):1593–1623，2014年。[17] Matthew D. Hoffman、Alexey Radul 和 Pavel Sountsov。用于设定哈密顿蒙特卡洛轨迹长度的自适应马尔可夫链蒙特卡洛方案。载于第24届人工智能与统计国际会议论文集，第3907–3915页。PMLR，2021年。[18] Laurène Hume 和 Philippe Poncet。用于高黏度三维流动的速度－涡量方法及其在数字岩石物理中的应用。计算物理学杂志，425:109910，2021年。[19] Weiqi Ji、Weilun Qiu、Zhiyu Shi、Shaowu Pan 和 Sili Deng。Stiff-PINN：用于刚性化学动力学的物理信息神经网络。《物理化学杂志A辑》，125(36):8098–8106，2021年。[20] Guang Lin、Yating Wang 和 Zecheng Zhang。通过贝叶斯物理信息神经网络处理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洛。计算物理学杂志，460:111173，2022年。[21] Kevin Linka、Amelie Schäfer、Xuhui Meng、Zongren Zou、George Em Karniadakis 和 Ellen Kuhl。用于真实世界非线性动力系统的贝叶斯物理信息神经网络。《计算力学与工程方法》，第115346页，2022年。[22] Dehao Liu 和 Yan Wang。多保真物理约束神经网络及其在材料建模中的应用。《机械设计杂志》，141(12)，2019年。[23] Qiang Liu 和 Dilin Wang。Stein变分梯度下降：一种通用贝叶斯推断算法。载于《神经信息处理系统进展》，第29卷。Curran Associates, Inc.，2016年。[24] Xu Liu、Wen Yao、Wei Peng 和 Weien Zhou。用于含噪数据下偏微分方程正问题和逆问题的贝叶斯物理信息极限学习机。arXiv，2022年。arXiv:2205.06948。[25] Samuel Livingstone、Michael Betancourt、Simon Byrne 和 Mark Girolami。论哈密顿蒙特卡洛的几何遍历性。《伯努利》，25(4A):3109–3138，2019年。[26] Suryanarayana Maddu、Bevan L. Cheeseman、Christian L. Müller 和 Ivo F. Sbalzarini。利用组稀疏性从数据中学习物理一致的微分方程模型。《物理评论E辑》，103:042310，2021年。[27] Suryanarayana Maddu、Bevan L. Cheeseman、Ivo F. Sbalzarini 和 Christian L. Müller。稳定性选择有助于从有限含噪数据中稳健地学习微分方程。《皇家学会会刊A辑》，478(2262):20210916，2022年。[28] Suryanarayana Maddu、Dominik Sturm、Bevan L. Cheeseman、Christian L. Müller 和 Ivo F. Sbalzarini。STENCIL-NET：偏微分方程的数据驱动解自适应离散化。arXiv，2021年。arXiv:2101.06182。[29] Suryanarayana Maddu、Dominik Sturm、Christian L Müller 和 Ivo F Sbalzarini。逆狄利克雷加权可实现物理信息神经网络的可靠训练。《机器学习：科学与技术》，3(1):015026，2022年。[30]José V. Manjón、José E. Romero、Roberto Vivo-Hernando、Gregorio Rubio、Fernando Aparici、Maria de La Iglesia-Vaya 等。利用深度学习进行盲式MRI脑病灶修复重建。载于《医学成像中的仿真与合成》，第41–49页。Springer国际出版公司，2020年。[31] Xuhui Meng、Hessam Babaee 和 George Em Karniadakis。多保真贝叶斯神经网络：算法与应用。计算物理学杂志，438:110361，2021年。[32] Joseph P. Molnar 和 Samuel J. Grauer。利用贝叶斯物理信息神经网络进行流场层析成像与不确定性量化。《测量科学与技术》，33(6):065305，2022年。[33] Jan Oldenburg、Finja Borowski、Alper Öner、Klaus-Peter Schmitz 和 Michael Stiehm。用于求解纳维–斯托克斯方程的几何感知物理信息神经网络代理模型（GAPINN）。《工程科学高级建模与仿真》，9(1):8，2022年。[34] Jaideep Pathak、Brian Hunt、Michelle Girvan、Zhixin Lu 和 Edward Ott。利用储备池计算方法，从数据中对大型时空混沌系统进行无模型预测。《物理评论快报》，120:024102，2018年1月。[35] Sarah Perez、Peter Moonen 和 Philippe Poncet。孔隙空间未解析形态特征导致的计算渗透率偏差。《多孔介质中的传输》，141(1):151–184，2022年。[36] Johan Phan、Leonardo C. Ruspini 和 Frank Lindseth。基于深度学习的三维数字岩石自动分割工具。《科学报告》，11(1):19123，2021年。[37] Philippe Poncet、Roland Hildebrand、Georges-Henri Cottet 和 Petros Koumoutsakos。针对圆柱体绕流最优减阻的空间分布式控制。《流体力学杂志》，599:111–120，2008年。

[38] Apostolos F. Psaros、Xuhui Meng、Zongren Zou、Ling Guo 和 George Em Karniadakis. 科学机器学习中的不确定性量化：方法、指标与比较. 计算物理学杂志, 477:111902, 2023年.[39] Nasim Rahaman、Aristide Baratin、Devansh Arpit、Felix Draxler、Min Lin、Fred A. Hamprecht 等. 关于神经网络的谱偏置. 载于第36届国际机器学习会议论文集，第5301–5310页. PMLR, 2019年.[40] Maziar Raissi、Paris Perdikaris 和 George Em Karniadakis. 物理信息神经网络：用于求解涉及非线性偏微分方程的正问题和逆问题的深度学习框架. 计算物理学杂志, 378:686–707, 2019年.[41] Maziar Raissi、Alireza Yazdani 和 George Em Karniadakis. Hidden Fluid Mechanics：一种基于 Navier–Stokes 的深度学习框架，用于同化流动可视化数据. arXiv, 2018年. arXiv:1808.04327.[42] Maziar Raissi、Alireza Yazdani 和 George Em Karniadakis. 隐藏流体力学：从流动可视化中学习速度场和压力场.《科学》, 367(6481):1026–1030, 2020年.[43] Franz M. Rohrhofer、Stefan Posch 和 Bernhard C. Geiger. 物理信息神经网络的帕累托前沿. arXiv, 2021年. arXiv:2105.00862.[44] Javier E. Santos、Ying Yin、Honggeun Jo、Wen Pan、Qinjun Kang、Hari S. Viswanathan 等. 应用于复杂三维多孔介质中流体流动的计算高效多尺度神经网络.《多孔介质中的传输》, 140(1):241–272, 2021年.[45] Hayden Schaeffer. 通过数据发现和稀疏优化学习偏微分方程.《英国皇家学会学报A辑：数学、物理与工程科学》, 473(2197):20160446, 2017.[46] Ozan Sener 和 Vladlen Koltun. 多任务学习即多目标优化. 载于《神经信息处理系统进展》第31卷. Curran Associates, Inc., 2018.[47] Reza Shams、Mohsen Masihi、Ramin Bozorgmehry Boozarjomehry 和 Martin J. Blunt. 结合统计方法与条件生成对抗神经网络方法重建三维多孔介质（ST-CGAN）.《水资源进展》, 158, 2021年.[48] Hwijae Son、Jin Woo Jang、Woo Jin Han 和 Hyung Ju Hwang. 面向物理信息神经网络的索博列夫训练. arXiv, 2021年. arXiv:2101.08932.[49] Luning Sun、Han Gao、Shaowu Pan 和 Jian-Xun Wang. 一种不依赖模拟数据、基于物理约束深度学习的流体流动代理建模方法.《计算力学与工程方法》, 361:112732, 2020年.[50] Luning Sun 和 Jian-Xun Wang. 面向稀疏噪声数据流场重建的物理约束贝叶斯神经网络.《理论与应用力学快报》, 10(3):161–169, 2020年.[51] Minh-Trieu Tran、Soo-Hyung Kim、Hyung-Jeong Yang 和 Guee-Sang Lee. 基于器官边界感知的医学图像修复多任务学习.《应用科学》, 11(9), 2021年.[52] Remco van der Meer、Cornelis W. Oosterlee 和 Anastasia Borovykh. 使用神经网络求解偏微分方程的最优加权损失函数.《计算与应用数学杂志》, 405:113887, 2022年.[53] Pantelis R. Vlachas、Jaideep Pathak、Brian R. Hunt、Themistoklis P. Sapsis、Michelle Girvan、Edward Ott 等. 用于预测复杂时空动力学的循环神经网络反向传播算法与储备池计算.《神经网络》, 126:191–217, 2020年.[54] Nikolaos N. Vlassis 和 WaiChing Sun. 热力学信息神经网络的索博列夫训练：用于水平集硬化的可解释弹塑性模型.《计算力学与工程方法》, 377:113695, 2021年.[55] Sifan Wang、Yujun Teng 和 Paris 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》, 43(5):A3055–A3081, 2021年.[56] Sifan Wang、Xinling Yu 和 Paris Perdikaris. PINN何时以及为何训练失败：神经切线核视角. 计算物理学杂志, 449:110768, 2022年.[57] Andrew G Wilson 和 Pavel Izmailov. 贝叶斯深度学习与泛化的概率视角. 载于《神经信息处理系统进展》第33卷，第4697–4708页. Curran Associates, Inc., 2020年.[58] Anqi Wu、Sebastian Nowozin、Edward Meeds、Richard E. Turner、José Miguel Hernández-Lobato 和 Alexander L. Gaunt. 鲁棒贝叶斯神经网络的确定性变分推断. 载于第7届国际学习表示会议. OpenReview.net, 2019年.[59] Liu Yang、Xuhui Meng 和 George Em Karniadakis. B-PINNs：利用含噪数据求解偏微分方程正问题和逆问题的贝叶斯物理信息神经网络. 计算物理学杂志, 425:109913, 2021年。

[60] Yibo Yang 和 Paris Perdikaris. 物理信息神经网络中的对抗式不确定性量化。计算物理学杂志, 394:136–152, 2019年.[61] Jiayu Yao、Weiwei Pan、Soumya Shubhra Ghosh 和 Finale Doshi-Velez. 贝叶斯神经网络推断的不确定性量化质量. *arXiv*, 2019年. arXiv:1906.09686.[62] Wentao Yuan、Qingtian Zhu、Xiangyue Liu、Yikang Ding、Haotian Zhang 和 Chi Zhang. 利用近似图像导数对隐式神经表示进行索博列夫训练. 载于计算机视觉 – *ECCV 2022*, 第13675卷, 第72–88页. 施普林格·自然瑞士, 2022年.[63] Qiang Zheng、Lingzao Zeng 和 George Em Karniadakis. 物理信息语义修复：在地统计建模中的应用. 计算物理学杂志, 419:109676, 2020年。