用于训练数据驱动降阶模型的伴随方法

Donglin Liu[a, b, 1, *], Francisco García Atienza[a, 1], Mengwu Guo[a]

[a]瑞典隆德大学数学科学中心[b]瑞典隆德大学实验医学系

arXiv:2601.07579v1 [cs.CE] 12 Jan 2026

摘要

降阶建模处于数值分析与数据驱动科学计算的交汇处，为压缩科学和工程中的高保真模拟提供了有原则的方法。我们提出了一种训练框架，将算子推断的连续时间形式与伴随状态法相结合，从而获得稳健的数据驱动降阶模型。该方法最小化降阶解与投影快照数据之间基于轨迹的损失，无需从含噪测量中估计时间导数，并通过时间积分实现内在的时间正则化。我们推导了相应的连续伴随方程，以高效计算梯度，并实现了一种基于梯度的优化器来更新降阶模型参数。每次迭代只需进行一次降阶模型的前向求解和一次伴随求解，随后进行计算开销较低的梯度组装，因此该方法适用于大规模模拟。我们在三种偏微分方程上验证了所提方法：黏性伯格斯方程、二维 Fisher-KPP 方程以及对流扩散方程。我们针对两种扰动情形——时间快照密度降低和加性高斯噪声——与标准算子推断进行了系统比较。对于无噪声数据，两种方法的准确度相近；但在采样稀疏且存在噪声的情况下，所提出的基于伴随法的训练准确度更高，滚动预测稳定性也更好。

关键词：降阶建模、算子推断、约束优化、伴随方法、科学机器学习

1. 引言

高保真数值模拟是数字孪生的核心。数字孪生通过双向信息交换，将物理资产与虚拟副本连接起来：测量数据用于校准和更新虚拟模型，而模型预测则为物理系统的监测、控制和决策提供指导 [1]。要投入实际应用，数字孪生必须能够快速、可靠地提供预测；然而，要忠实地解析底层物理过程，往往需要计算成本高得难以承受的高维动力学模型。这一挑战在安全关键场景中尤为突出，在这些场景中，计算效率必须与基于模型的决策的可信度和可解释性相结合 [2]。

降阶模型（ROM）提供了可解释的计算代理模型，将高保真、高维模拟压缩到主要的低维表示中，在保留基本物理行为的同时实现高效计算 [3, 4]。经典的基于投影的降阶模型通过将全阶模型（FOM）投影到低维子空间上，以侵入式方式构建降阶动力学 [3, 5, 6]。相比之下，非侵入式数据驱动降阶模型无需直接与全阶模型实现交互，而是主要从模拟数据中学习低维潜在动力学，从而减少实现

*通讯作者。电子邮箱：donglin.liu@med.lu.se (Donglin Liu)、fragarat@gmail.com (Francisco García Atienza)、mengwu.guo@math.lu.se (Mengwu Guo)[1]前两位作者对本文贡献相同。

负担，并将模型降阶拓展至遗留代码和黑箱模拟器。广泛的数据驱动数值方法均属于这一范畴，包括动态模态分解（DMD） [7, 8, 9, 10],算子推断（OpInf） [4, 11, 12, 13], Loewner 框架 [14], 基于插值的方法 [15, 16],高斯过程替代模型 [17, 18], 以及深度学习方法 [19, 20]。本文聚焦于 OpInf，该方法直接从轨迹数据中识别具有结构的低维算子，同时强制采用预先指定的模型形式。

数据驱动降阶模型的一项关键实践限制在于可用训练数据的数量（例如，稀疏采样的测量值）和质量（例如，含噪观测）。标准 OpInf 通常通过对状态快照的估计时间导数进行回归来识别降阶算子，因此对测量噪声和粗疏的时间采样都很敏感。尤其是，对状态数据进行显式数值微分会放大噪声，并在数据稀缺时使推断出的算子产生偏差，这通常会导致非线性降阶系统的时间积分不稳定。这些问题促使人们探索能够避免此类数值微分、并通过时间积分纳入时间正则化的训练方法，从而在保持复杂系统计算可行性的同时提高鲁棒性。动力学学习领域中可微编程的相关进展 [21]，（例如神经常微分方程）表明，通过伴随方法 [22, 23, 24]对常微分方程求解器进行求导，可以高效优化基于轨迹的目标函数。

本文为 OpInf 开发了一种连续时间、基于梯度的训练框架，并通过伴随方法计算梯度。我们最小化降阶模型解与投影后的降阶状态数据之间的积分轨迹失配，从而避免显式估计时间导数，并通过时间积分实现内在正则化。通过向后求解伴随状态系统，评估该损失函数相对于降阶模型参数的梯度。因此，优化过程中的每次迭代只需进行一次降阶模型正向求解和一次伴随求解，随后再通过额外的代数运算组装参数梯度。由此，我们的方法沿用了神经常微分方程的高效优化机制，同时保留了 OpInf 的可解释性、鲁棒性和物理一致性。我们在三个典型偏微分方程上验证所提方法：黏性伯格斯方程、二维 Fisher-KPP 反应扩散方程，以及对流扩散方程。对于每个数值算例，我们利用全阶快照构造 POD 基，使用标准 OpInf 初始化降阶算子，再通过所提的基于伴随法的训练方法对其进行优化。在三个算例中，我们比较了标准 OpInf 与所提基于伴随法的版本在两种扰动下的表现：为降低降阶快照密度而进行的均匀时间下采样，以及多个噪声等级下的加性高斯噪声。对于无噪声数据，所提方法与标准 OpInf 的准确度相当。然而，在稀疏采样和含噪数据的情况下，基于伴随法的训练能提高准确度，并在时间积分期间表现出更强的稳定性。

本文其余部分的组织结构如下。第 2 节回顾降阶建模和 OpInf。随后，在第 3 节中，我们利用伴随状态方程介绍所提出的 OpInf 轨迹拟合方案，并讨论其计算实现。第 4 节给出并讨论数值算例的结果，最后在第 5 节作出总结性说明。

2. 通过算子推断构建降阶模型

我们考虑一个高维全阶模型

$$\dot{\mathbf{u}}(t) = \mathbf{F}\big(\mathbf{u}(t), \mathbf{s}(t)\big), \qquad t \in [0, T], \tag{1}$$

其状态为 $\mathbf{u}(t) \in \mathbb{R}^n$ ，输入为 $\mathbf{s}(t) \in \mathbb{R}^m$。在本文中，我们关注一种常见的二次结构 [12]

$$\dot{\mathbf{u}}(t) = \boldsymbol{\mathcal{C}} + \boldsymbol{\mathcal{A}}\,\mathbf{u}(t) + \boldsymbol{\mathcal{H}}\big(\mathbf{u}(t) \otimes \mathbf{u}(t)\big) + \boldsymbol{\mathcal{B}}\,\mathbf{s}(t), \tag{2}$$

其中，$\boldsymbol{\mathcal{C}} \in \mathbb{R}^n$、$\boldsymbol{\mathcal{A}} \in \mathbb{R}^{n\times n}$、$\boldsymbol{\mathcal{H}} \in \mathbb{R}^{n\times n^2}$、$\boldsymbol{\mathcal{B}} \in \mathbb{R}^{n\times m}$，且 $\otimes$ 表示克罗内克积 [25]。

降阶状态 $\mathbf{q}(t) \in \mathbb{R}^r$ 在一个 r 维子空间中近似全阶模型轨迹 $\mathbf{u}$ 与 $r \ll n$ ，

$$\mathbf{u}(t) \approx \mathbf{V}_r \mathbf{q}(t),$$

其中，$\mathbf{V}_r \in \mathbb{R}^{n\times r}$ 通过本征正交分解（POD）构造并汇集了一个降阶基，该基由在时间点 t_1、$\cdots$、t_k采集的状态快照 $\mathbf{U} = \begin{bmatrix}\mathbf{u}(t_1) & \cdots \mathbf{u}(t_k)\end{bmatrix} \in \mathbb{R}^{n\times k}$ 构成。具体而言，设快照矩阵 $\mathbf{U}$ 的奇异值分解（SVD）为 $\mathbf{U} = \boldsymbol{\Phi}\boldsymbol{\Sigma}\boldsymbol{\Psi}^\top$，其中 $\boldsymbol{\Sigma}$ 是汇集奇异值 $\sigma_1 \geq \cdots \geq \sigma_{\mathrm{rank}(\mathbf{U})} \geq 0$的对角矩阵。秩为$r$ 的 POD 基为 $\mathbf{V}_r = \boldsymbol{\Phi}_{:,1:r}$，我们根据累积能量准则 $\sum_{i=1}^{r}\sigma_i^2 / \sum_{i=1}^{\mathrm{rank}(\mathbf{U})}\sigma_i^2 \geq 1-\epsilon_r$选择 $r < \mathrm{rank}(\mathbf{U})$，其中 $\epsilon_r \ll 1$ 是预设容差。根据 Eckart–Young–Mirsky 定理 [26]，，POD 投影在最小二乘意义下是最优的，并且

$$\sum_{i=1}^{k}\left\|\mathbf{u}(t_i) - \mathbf{V}_r\mathbf{V}_r^\top\mathbf{u}(t_i)\right\|_2^2 = \left\|\mathbf{U} - \mathbf{V}_r\mathbf{V}_r^\top\mathbf{U}\right\|_F^2 = \sum_{i=r+1}^{\mathrm{rank}(\mathbf{U})}\sigma_i^2.$$

将系统（1）投影到子空间 $\mathrm{Col}(\mathbf{V}_r) \subset \mathbb{R}^n$ 可得到降阶动力学

$$\dot{\mathbf{q}}(t) = \mathbf{V}_r^\top\mathbf{F}(\mathbf{V}_r\mathbf{q}(t), \mathbf{s}(t)).$$

对于二次系统（2），降阶模型保留了多项式形式

$$\dot{\mathbf{q}}(t) = \mathbf{c} + \mathbf{A}\mathbf{q}(t) + \mathbf{H}\big(\mathbf{q}(t)\otimes\mathbf{q}(t)\big) + \mathbf{B}\mathbf{s}(t) =: \mathbf{f}\big(\mathbf{q}(t), \mathbf{s}(t); \boldsymbol{\theta}\big), \tag{3}$$

其中降阶算子为 $\mathbf{c} = \mathbf{V}_r^\top\mathbf{C} \in \mathbb{R}^r$、$\mathbf{A} = \mathbf{V}_r^\top\mathbf{A}\mathbf{V}_r \in \mathbb{R}^{r\times r}$、$\mathbf{H} = \mathbf{V}_r^\top\mathbf{H}(\mathbf{V}_r\otimes\mathbf{V}_r) \in \mathbb{R}^{r\times r^2}$ 和 $\mathbf{B} = \mathbf{V}_r^\top\mathbf{B} \in \mathbb{R}^{r\times m}$。我们将参数记为 $\boldsymbol{\theta} = [\mathbf{c}、\mathbf{A}、\mathbf{H}、\mathbf{B}] \in \mathbb{R}^d$，其中 $d = r + r^2 + r^3 + rm$ 表示参数空间的维数（等价地，$\mathrm{vec}(\boldsymbol{\theta}) \in \mathbb{R}^d$ 通过堆叠 $\mathbf{c}、\mathbf{A}、\mathbf{H}、\mathbf{B}$ 得到）。当高保真算子 $(\mathbf{C}、\mathbf{A}、\mathbf{H}、\mathbf{B})$ 可用时，可以通过投影以侵入式方式组装 $(\mathbf{c}、\mathbf{A}、\mathbf{H}、\mathbf{B})$。然而，在许多应用中，求解器是专有的或属于黑箱，因此需要采用非侵入式的数据驱动方法，从投影状态数据 $(\mathbf{q}\ \dot{\mathbf{q}})$ 和输入数据 $\mathbf{s}$ 中学习降阶算子。

算子推断 (OpInf) 通过线性最小二乘 [12]直接拟合降阶模型 (3)，并在 POD 坐标中进行。给定 $\mathbf{V}_r$，我们构造降阶快照 $\mathbf{q}(t_i) = \mathbf{V}_r^\top\mathbf{u}(t_i)$，并通过有限差分估计 $\dot{\mathbf{q}}(t_i)$ $(1 \leq i \leq k)$。本研究具体采用二阶和六阶差分模板。随后建立回归问题：

$$\min_{\boldsymbol{\theta}}\sum_{i=1}^{k}\left\|\mathbf{c} + \mathbf{A}\mathbf{q}(t_i) + \mathbf{H}\big(\mathbf{q}(t_i)\otimes\mathbf{q}(t_i)\big) + \mathbf{B}\mathbf{s}(t_i) - \dot{\mathbf{q}}(t_i)\right\|_2^2, \tag{4}$$

可选用 Tikhonov（岭回归）正则化 [27, 28]，也可通过对快照的 Gram 矩阵 [29, 30] 进行截断奇异值分解 (TSVD) 求解，以改善条件性。OpInf 是非侵入式的，并保留控制常微分方程的多项式结构，但其准确度取决于可靠的 $\dot{\mathbf{q}}$ 估计。高阶差分模板可降低截断误差，但也可能放大噪声。这种敏感性促使我们在下一节提出连续时间的基于伴随法的训练方法，该方法无需显式进行时间导数估计。

3. 算子推断的伴随方法

3.1. 连续时间损失泛函

为避免像标准 OpInf 方法中那样显式求导，我们引入一个连续时间损失泛函，以 L_2 意义比较通过对学习到的算子进行积分而重构的轨迹与观测数据。令 $\mathcal{T} = [0, T]$ 和 $\mathbf{q}(\cdot) \in \mathcal{C}^1(\mathcal{T}; \mathbb{R}^r)$ 表示随时间变化的 ROM 状态函数。实际应用中，我们将降阶状态的快照 $\{(t_i,\ \mathbf{q}_{\mathrm{true}}(t_i))\}_{i=1}^k$ 作为离散轨迹数据进行评估。随后，降阶空间中的连续时间轨迹损失定义为

$$\ell:\ \mathcal{C}^1(\mathcal{T};\mathbb{R}^r) \to \mathbb{R}, \qquad \mathbf{q}(\cdot) \mapsto \int_0^T \left\|\mathbf{q}(t) - \mathbf{q}_{\mathrm{true}}(t)\right\|_2^2 \mathrm{d}t = \int_0^T g(\mathbf{q}(t), t)\mathrm{d}t, \tag{5}$$

其中 $g(\mathbf{q}(t)\ t) = \left\|\mathbf{q}(t) - \mathbf{q}_{\mathrm{true}}(t)\right\|_2^2$，连续轨迹 $\mathbf{q}_{\mathrm{true}}$ 则通过对降阶快照进行时间插值来近似。

我们按照（3）中的方式对降阶模型进行参数化，并将降阶算子收集在 $\boldsymbol{\theta} = [\,\mathbf{c}、\mathbf{A}、\mathbf{H}、\mathbf{B}\,] \in \mathbb{R}^d$中。给定输入函数 $\mathbf{s}(t)$，预测轨迹 $\tilde{\mathbf{q}}(\cdot;\ \boldsymbol{\theta})$ 是以下方程的解

$$\tilde{\mathbf{q}}(t;\boldsymbol{\theta}) = \tilde{\mathbf{q}}_0 + \int_0^t \Big[\mathbf{c} + \mathbf{A}\,\tilde{\mathbf{q}}(\tau;\boldsymbol{\theta}) + \mathbf{H}\big(\tilde{\mathbf{q}}(\tau;\boldsymbol{\theta}) \otimes \tilde{\mathbf{q}}(\tau;\boldsymbol{\theta})\big) + \mathbf{B}\,\mathbf{s}(\tau)\Big]\,\mathrm{d}\tau.$$

将式（5）与解映射 $\boldsymbol{\theta} \mapsto \tilde{\mathbf{q}}(\cdot;\boldsymbol{\theta})$ 复合后，可得到一个以参数 $\boldsymbol{\theta}$为变量的降阶损失函数:

$$\tilde{\ell}(\boldsymbol{\theta}) := \ell\big(\tilde{\mathbf{q}}(\cdot;\boldsymbol{\theta})\big) = \int_0^T \big\|\tilde{\mathbf{q}}(t;\boldsymbol{\theta}) - \mathbf{q}_{\mathrm{true}}(t)\big\|_2^2\,\mathrm{d}t. \tag{6}$$

学　　习问题旨在确定降阶算子 $\boldsymbol{\theta}$，这便是一个无约束优化

$$\boldsymbol{\theta}^* \ \in\ \arg\min_{\boldsymbol{\theta}\in\mathbb{R}^d}\ \tilde{\ell}(\boldsymbol{\theta}).$$

为了最小化损失函数 $\ell(\boldsymbol{\theta})$，我们计算关于参数的总梯度。根据链式法则，有

$$\frac{\mathrm{d}\tilde{\ell}(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}} = \frac{\mathrm{d}\ell(\tilde{\mathbf{q}}(\cdot,\boldsymbol{\theta}))}{\mathrm{d}\boldsymbol{\theta}} = \underbrace{\frac{\partial \ell}{\partial \boldsymbol{\theta}}}_{\text{direct dependence } = 0} + \frac{\partial \ell}{\partial \mathbf{q}}\bigg|_{\mathbf{q}(\cdot)=\tilde{\mathbf{q}}(\cdot,\boldsymbol{\theta})} \cdot \frac{\mathrm{d}\tilde{\mathbf{q}}}{\mathrm{d}\boldsymbol{\theta}},$$

其中 $\partial\ell/\partial\boldsymbol{\theta} = 0$，因为 ℓ 仅通过降阶状态依赖于 $\boldsymbol{\theta}$。注意，$\partial\ell/\partial\mathbf{q}$实际上是定义在降阶状态函数空间 $\mathcal{C}^1(\mathcal{T};\mathbb{R}^r)$上、而非定义在 $\mathbb{R}^r$上的方向导数（算子）。

然而，直接计算 $\mathrm{d}\tilde{\mathbf{q}}/\mathrm{d}\boldsymbol{\theta}$ 通常计算开销很大，因为此类原始灵敏度分析的成本会随参数数量（即，在我们的设定中为 d）线性增长。接下来，我们将使用一个满足反向常微分方程的连续伴随变量来改写梯度，具体内容见下一小节中的命题 1。尤其是在需要估计的参数数量相对较多时，伴随方法的优势便十分明显：无论参数数量多少，其计算成本大约是一次原始模拟成本的两倍。

3.2. 伴随方法

我们将确定降阶算子 $\boldsymbol{\theta}$ 的训练过程表述为一个微分方程约束优化问题，即

$$\begin{aligned} \min_{\boldsymbol{\theta}} \tilde{\ell}(\boldsymbol{\theta}) \quad &= \quad \min_{\boldsymbol{\theta},\mathbf{q}(\cdot)} \ell(\mathbf{q}(\cdot)) \\ &\qquad \text{s.t.} \quad \dot{\mathbf{q}}(t) = \mathbf{f}\big(\mathbf{q}(t);\boldsymbol{\theta}\big), \qquad \mathbf{q}(0) = \mathbf{q}_0, \qquad t \in [0,T], \end{aligned} \tag{7}$$

其中 ℓ和 ℓ分别定义于（6）和（5）中，而 $\mathbf{f}(\cdot;\ \boldsymbol{\theta})$ 是（3）中降阶模型的右端项。请注意，轨迹损失 ℓ 仅取决于降阶状态函数 $\mathbf{q}(\cdot)$，用于衡量其与参考轨迹 $\mathbf{q}_{\mathrm{true}}(\cdot)$之间的偏差，并不显式依赖于 $\boldsymbol{\theta}$。对于这种情形，经典的连续时间伴随方法（例如参见 [24, 31]）得出以下结果。

命题 1（伴随方法）。假设 $\mathbf{f}(\mathbf{q};\ \boldsymbol{\theta})$ 关于 $\mathbf{q}$和 $\boldsymbol{\theta}$连续可微，且初始条件 $\mathbf{q}_0$ 与 $\boldsymbol{\theta}$无关。令 $\tilde{\mathbf{q}}(\cdot;\ \boldsymbol{\theta})$ 为（(7)中的状态方程的解。降阶损失函数(6)的梯度为

$$\frac{\mathrm{d}\tilde{\ell}(\boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}} = \int_0^T \boldsymbol{\lambda}(t)^\top \frac{\partial \mathbf{f}(\mathbf{q};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\bigg|_{\mathbf{q}(\cdot)=\tilde{\mathbf{q}}(\cdot;\boldsymbol{\theta})} \mathrm{d}t, \tag{8}$$

其中，伴随状态变量 $\boldsymbol{\lambda}$：$[0,T] \to \mathbb{R}^r$ 满足向后求解的常微分方程

$$\dot{\boldsymbol{\lambda}}(t) = -\left[\left(\frac{\partial \mathbf{f}}{\partial \mathbf{q}}\right)^\top \boldsymbol{\lambda}(t) + \left(\frac{\partial g}{\partial \mathbf{q}}\right)^\top\right]\bigg|_{\mathbf{q}(\cdot)=\tilde{\mathbf{q}}(\cdot,\boldsymbol{\theta})}, \quad \boldsymbol{\lambda}(T) = \mathbf{0}. \tag{9}$$

证明见附录 B。为求解伴随方程（9），首先对正向降阶模型进行积分，以获得 $\tilde{\mathbf{q}}(t;\ \boldsymbol{\theta})$ 在 $[0,T]$上的值，然后从 $t=T$ 到 $t=0$向后求解伴随常微分方程（9）。随后，通过沿所得轨迹进行数值积分，计算梯度（8）。

导数记号。 我们采用统一的雅可比对流，并将 $\partial_{\mathbf{q}}$ 和 $\partial/\partial\mathbf{q}$ 互换使用。对于标量 $g:\ \mathbb{R}^r\to\mathbb{R}$, $\partial_{\mathbf{q}}g\equiv\nabla_{\mathbf{q}}g^\top\in\mathbb{R}^{1\times r}$。对于向量场 $\mathbf{f}:\ \mathbb{R}^r\to\mathbb{R}^r$, $\partial_{\mathbf{q}}\mathbf{f}\ \in\mathbb{R}^{r\times r}$ 表示雅可比矩阵，其中 $(\partial_{\mathbf{q}}\mathbf{f})_{ij}=\partial f_i/\partial q_j$。同理，参数导数为 $\partial_{\boldsymbol{\theta}}\mathbf{f}\ \in\mathbb{R}^{r\times d}$。

二次降阶模型：显式导数。 对于式（3）中的二次降阶模型，

$$\mathbf{f}(\mathbf{q};\boldsymbol{\theta})=\mathbf{c}+\mathbf{A}\,\mathbf{q}+\mathbf{H}(\mathbf{q}\otimes\mathbf{q})+\mathbf{B}\,\mathbf{s}(t)\quad\text{with}\quad\boldsymbol{\theta}=[\mathbf{c},\mathbf{A},\mathbf{H},\mathbf{B}],$$

式（8）和（9）中的导数为

$$\begin{aligned}\partial_{\mathbf{q}}\mathbf{f}(\mathbf{q};\boldsymbol{\theta})&=\mathbf{A}+2\mathbf{H}\big(\mathbf{I}_r\otimes\mathbf{q}\big)\ \in\ \mathbb{R}^{r\times r},\\ \partial_{\mathbf{q}}g\big(\mathbf{q}(t),t\big)&=2\big(\mathbf{q}(t)-\mathbf{q}_{\text{true}}(t)\big)^\top\ \in\ \mathbb{R}^{1\times r},\\ \partial_{\boldsymbol{\theta}}\mathbf{f}(\mathbf{q};\boldsymbol{\theta})&=\big[\mathbf{I}_r,\ \mathbf{q}^\top\otimes\mathbf{I}_r,\ (\mathbf{q}\otimes\mathbf{q})^\top\otimes\mathbf{I}_r,\ \mathbf{s}(t)^\top\otimes\mathbf{I}_r\big]\ \in\ \mathbb{R}^{r\times d}.\end{aligned}$$

完整推导见 附录 C。

3.3. 计算实现

参数 $\boldsymbol{\theta}$ 的初始化对非凸优化至关重要。我们使用式（4）中回归得到的 OpInf 估计值进行热启动。为稳定该估计器并增强其对噪声的鲁棒性，我们对 OpInf 采用 Tikhonov（岭）正则化，并可选择应用截断奇异值分解（TSVD），以舍弃受噪声主导的方向；这是因为较小的奇异值主要反映噪声，而较大的奇异值则承载信号。通过网格搜索选择超参数（岭回归权重和 TSVD 秩）：针对每种设置，在训练集上拟合 OpInf，并选择使验证相对状态误差（RSE）最小的模型，

$$\text{RSE}\ =\ \frac{\big\|\mathbf{q}_{\text{true}}(t)-\mathbf{q}_{\text{pred}}(t)\big\|_2}{\big\|\mathbf{q}_{\text{true}}(t)\big\|_2},\tag{10}$$

在整个验证时域内进行聚合。选定的 OpInf 估计器为基于伴随法的训练提供初始猜测 $\boldsymbol{\theta}^0$ 。

在 POD 中，一个模态的“能量”等于其捕获的方差；具体而言，平方奇异值σ_i^2 量化了模态 i [32]的方差。在独立同分布高斯测量噪声下，噪声协方差是各向同性的，因此投影噪声方差在每个正交方向上都近似相同，而信号方差集中在主导模态中；因此，能量较低的模态具有较低的信噪比（SNR），而非更大的绝对噪声 [33]。为了突出信息丰富的方向，同时保持统计原理上的严谨性，我们采用逐模态权重 $\omega_i=\sigma_i^p/(\nu_i^2+\tau)$、 $\tau=10^{-8}$，其中 ν_i^2是投影到模态 i上的噪声方差。令 $p=0$，即可得到广义最小二乘（逆方差）加权 [34]； p =2 则得到信噪比加权 σ_i^2/ν_i^2。本研究采用 $p=1$ 作为温和信噪比方案，以优先考虑高能量模态，同时不过度惩罚其余模态。我们通过自适应窗口 Savitzky–Golay 滤波器对每个 POD 系数进行平滑，并将残差方差（状态减去平滑后的状态）作为 ν_i^2 [35, 36]，以此估计 ν_i^2。令 $\boldsymbol{W}=\text{diag}(\omega_1,\dots,\omega_r)/\sum_{i=1}^r\omega_i$，使得 $\text{tr}(\boldsymbol{W})\ =1$，从而固定整体尺度，并稳定不同数据集上的梯度。加权逐点损失为

$$g(\mathbf{q},t)=\big\|\sqrt{\boldsymbol{W}}\big(\mathbf{q}-\mathbf{q}_{\text{true}}(t)\big)\big\|_2^2,$$

并且由于 $\boldsymbol{W}$ 是对角矩阵且其元素严格为正，其梯度为

$$\partial_{\mathbf{q}}g(\mathbf{q},t)=2\left(\boldsymbol{W}(\mathbf{q}(t)-\mathbf{q}_{\text{true}}(t))\right)^\top.$$

对于大规模问题，可以采用随机梯度下降（SGD）等随机变体 [37]。˜然而，对于本文所考虑的较小数据集，我们使用确定性梯度下降对 $\ell(\boldsymbol{\theta})$ 进行优化，并采用˜Armijo回溯。参数梯度 $\nabla_{\boldsymbol{\theta}}\ell$ 通过伴随公式（8）计算，更新为

$$\boldsymbol{\theta}^{j+1} \;=\; \boldsymbol{\theta}^{j} - \eta_j \, \nabla_{\boldsymbol{\theta}}\tilde{\ell}(\boldsymbol{\theta}^{j}),$$

并采用 η_j 通过 Armijo条件选取，以确保充分下降 [38]。线搜索和完整训练循环的总结见算法 1–2，详见 附录 A。

4. 数值实验

我们在三个典型偏微分方程上评估基于伴随法的训练：一维黏性伯格斯方程、二维 Fisher–KPP反应扩散方程，以及二维对流–扩散方程（ADE）。这些示例的数据集生成数值细节见 附录 D。这些基准测试考察了降阶模型的不同方面：低黏度的 Burgers方程会产生尖锐梯度和非线性对流扩散耦合；Fisher–KPP方程展现了二维空间中的前沿传播；ADE 则在允许对流主导（高佩克莱数）状态的同时，单独考察线性输运和扩散。

对于每个偏微分方程，我们在均匀空间网格上生成固定时间跨度T内的全阶模型（FOM）轨迹，收集状态快照，并按时间顺序将其划分为训练集/验证集/测试集。我们根据训练集快照构建 POD 基$\mathbf{V}_r$，并用它将 FOM 数据投影到降维坐标中。除非另有说明，我们选择 r，使累积快照能量满足 $\kappa_r \geq 0.995$（即保留至少 99.5% 的能量）。

扰动方案。 为评估鲁棒性，我们考虑两种受控扰动：

- **快照稀疏化：** 对时间网格进行均匀稀疏化（降低采样密度）
- **加性高斯噪声：** 用独立同分布噪声污染无噪声快照

$$\epsilon(t) \;\sim\; \mathcal{N}\big(0,\; (\delta\,\sigma_{\mathbf{q}})^2\,\mathbf{I}\big), \qquad \delta \in \{0, 40, 80, 120, 160, 200\}\%,$$

其中， $\sigma_{\mathbf{q}}$ 表示训练窗口内无噪声降阶状态的经验标准差。我们将 δ 称为噪声水平（NL）。

基线与初始化（*OpInf*）。 我们在相同的 POD 基和预处理条件下，将伴随训练的降阶模型与标准 OpInf 基线进行比较。OpInf 使用 Python 软件包 `opinf` [39]实现。对于 OpInf 回归，我们测试 2 阶（OpInf-ord2）和 6 阶（OpInf-ord6）有限差分模板，以近似$\dot{\mathbf{q}}$。我们采用岭回归和 TSVD正则化，并通过对 (i) 岭权重 $\{0, 10^{-2}, 10^{-1}, 1\}$ 以及 (ii) 舍弃 1–7 个最小奇异方向进行网格搜索，在验证集上选择超参数。通过最小化验证集相对平方误差（公式（10））来选择差分模板阶数和正则化方法。使用最佳 OpInf 估计初始化伴随训练， $\boldsymbol{\theta}^0$。

训练方案（伴随方法）。当 ODE 在较长时间跨度内端到端积分时，微小的离散化误差和建模误差会不断累积，并可能被动力学中的不稳定模态放大，最终导致较大的预测误差，以及条件不良、甚至发生梯度爆炸的梯度。因此，我们采用多重打靶启发式训练方案，通过短时域滚动预测缓解长时间ODE训练中的梯度爆炸 [40]。轨迹被划分为三个片段。对于每个片段，我们将初始条件重置为片段起始处的观测状态，并仅在该短时间跨度内对降阶模型进行积分。各片段共享参数，但更新依据单个片段的目标函数计算（不对所有片段进行全局求和）。我们依次处理各片段，每个片段最多进行 30 次伴随迭代。在伴随优化过程中，我们对参数向量加入岭惩罚 $\boldsymbol{\theta}$。岭惩罚的权重通过在 $\{0, 10^{-2}, 10^{-1}, 1, 10\}$ 上进行网格搜索来选取，并依据验证集上的相对平方误差进行模型选择。

优化与数值计算。Armijo回溯参数固定为 $\alpha = 10^{-4}$，$\beta = 0.5$，$\gamma = 0.5$，以及$\eta_0 = 10^{-3}$；当 $\|\nabla_{\boldsymbol{\theta}} \tilde{\ell}\|_2 \leq 10^{-8}$时，梯度下降终止。前向和伴随常微分方程以及时间积分均使用`SciPy` 例程 [41]计算；数据失配及其梯度在求解器的内部时间网格上，利用观测轨迹的时间插值函数进行评估。

4.1. 黏性伯格斯方程

我们研究一维黏性Burgers方程

$$u_t + u\,u_x = \nu\,u_{xx}, \qquad x \in [0,1],\ t \in [0,T], \quad \nu = 0.01,$$

满足齐次Dirichlet条件 $u(t,\ 0) = u(t,\ 1) = 0$，初始条件为 $u(x,\ 0) = \sin(2\pi x)$。这一非线性对流–扩散问题会形成陡峭梯度，而黏度会使其平滑，因此它是非线性模型降阶的典型测试问题。

我们在具有 $N = 998$ 个内点的均匀空间网格上求解偏微分方程 $(\Delta x = 1/(N+1))$，并采用包含 $T = 1$ 个 $M = 9999$ 时间步的均匀时间网格 $(\Delta t = T/M \approx 10^{-4})$。时间推进采用二阶半隐式（隐式显式方法）有限差分格式：对流项采用Lax–Wendroff格式，扩散采用Crank–Nicolson方法，并在每个时间步施加Dirichlet边界数据。将所有时刻的离散解叠放起来，得到快照矩阵 $\mathbf{U} \in \mathbb{R}^{1000\times 10,\ 000}$，并将其用于POD和降阶模型训练（见图 1）。

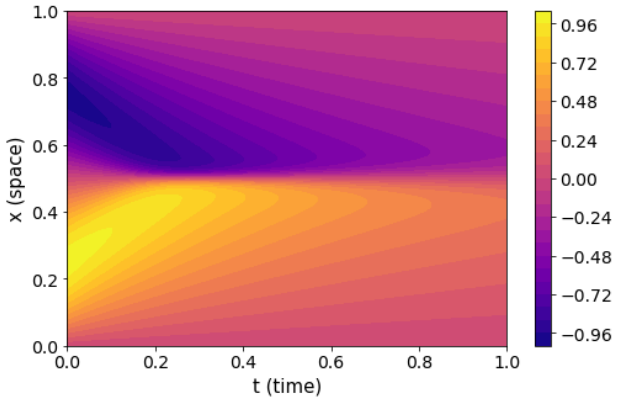


图 1：黏性伯格斯方程的数值解。 $u(x,\ t_i)$ 的时空等值图。

为探究采样效应，我们对时间网格进行均匀子采样，得到 100%（10000 个点）、10%（1000 个点）、1%（100 个点）和 0.2%（20 个点）的快照。随后，我们按时间顺序将轨迹划分为训练集/验证集/测试集： $t \in [0, 0.5]$（50%）、 $t \in [0.5, 0.6]$（10%）和 $t \in [0.6, 1]$（40%）。训练集和验证集快照分别以状态标准差的 $\delta \in \{0, 40, 80, 120, 160, 200\}\%$ 水平叠加独立同分布的高斯噪声；测试数据不含噪声。对于每种情景和 ROM 维数 $r \in \{1, \ldots, 5\}$（其中$\kappa_5 \geq 0.999$），我们（i）拟合 OpInf-ord2 和 OpInf-ord6，并根据验证集相对平方误差选择岭回归/截断奇异值分解超参数；（ii）使用最佳 OpInf 估计初始化伴随方法。随后，所有方法均从测试初始条件出发进行滚动预测，并将测试窗口上的对数10-相对平方误差作为性能指标（图2）。

在不同采样和噪声条件下，伴随训练的降阶模型在大多数子图中取得了最低的测试误差（图 2）。在无噪声数据上（最左列），OpInf-ord6 的表现具有竞争力，但随着噪声增加，其性能迅速下降，这与高阶差分会放大噪声的情况一致。OpInf-ord2 在噪声下通常比 OpInf-ord6 更稳定，但仍不及伴随方法。值得注意的是，在最高噪声水平下（最右列），即使只有 20 个快照，伴随方法仍能保持准确，而两种 OpInf 变体的误差都显著增大。随着 r 增加，误差通常会下降，直到 $r \approx 3 - 4$，随后趋于平稳或在噪声较大的情况下恶化，这表明收益递减且可能出现过拟合。

图3 展示了降维坐标滚动预测（$r = 5$，1000 个快照），在 $\delta = \{0, 80, 160\}\%$：随着噪声增大，伴随降阶模型始终比 OpInf 更贴近无噪声轨迹，凸显了其

在数据损坏情况下具有更优的梯度保真度。图4展示了映射回全阶模型网格的相应时空场；在强噪声下，伴随模型比 OpInf 基线方法更好地保留了演化剖面的结构与相位。

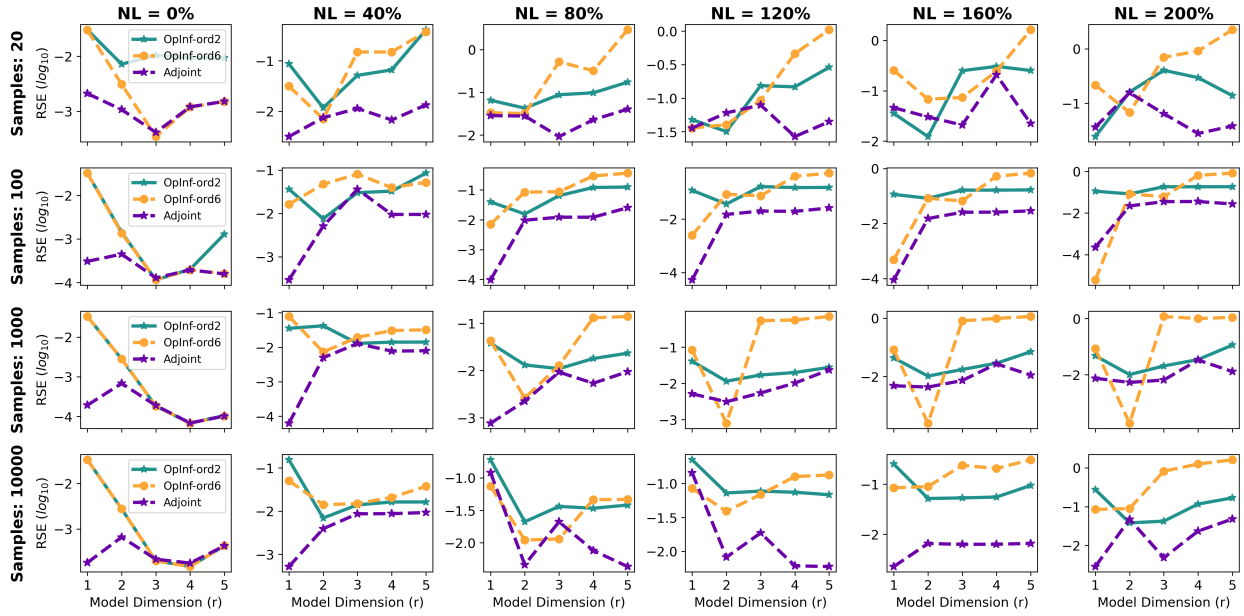


图2：不同噪声水平和采样条件下 Burgers方程的模型性能。列表示不同的噪声水平（NL = 0%, 40%, 80%, 120%, 160%, 200%）；行表示训练集、验证集和测试集中的快照数量（10000 = 26, 100, 1000, 6 ）。每个面板展示测试相对平方误差（$\log_{10}$）随 ROM 维数 r 的变化。方法：伴随方法（本文方法）、OpInf-ord2 和 OpInf-ord 。数值越低越好。

4.2. Fisher-KPP方程

对于第二个合成实验，我们考虑二维 Fisher-KPP方程

$$u_t = \mathfrak{D}\left(u_{xx} + u_{yy}\right) + \rho\, u\,(1-u), \qquad (x,y) \in [0, L_x] \times [0, L_y],\ t \in [0, T],$$

其边界条件为齐次 Neumann（零通量） $\partial u/\partial n|_{\partial\Omega} = 0$ ，并采用以区域中心为中心的高斯分布初始条件，

$$u(x, y, 0) = \exp\big(-10\big[(x - L_x/2)^2 + (y - L_y/2)^2\big]\big).$$

我们使用 $L_x = L_y = 10$、 $\mathfrak{D} = 0.1$和 $\rho = 1.0$。偏微分方程在均匀网格上求解，每个方向包含 $N_x = N_y = 125$个点（$\Delta x = L_x/(N_x - 1)$， $\Delta y = L_y/(N_y - 1)$），并采用包含 $T = 5$、$\Delta t = 0.0025$ ($N_t = T/\Delta t = 2000$)的均匀时间网格。时间推进采用二阶隐式显式方法：扩散项使用 Crank–Nicolson方法，逻辑斯蒂反应项使用显式欧拉法，并在离散拉普拉斯算子中施加 Neumann 数据。将解沿时间方向堆叠可得到快照矩阵 $\mathbf{U} \in \mathbb{R}^{(N_x N_y)\times(N_t+1)}$，用于 POD 和降阶模型训练（见图 5）。

我们采用与 Burgers方程案例相同的评估方案（OpInf初始化、伴随训练和测试窗口相对平方误差），但针对 FKPP 做了两项调整：(i) 通过均匀下采样设置更密集的采样基线，分别为 2000（100%）、1000（50%）、500（25%）和 200（10%）个快照；(ii) 延长时间跨度，并按 $t \in [0, 3.75]$ (训练集，75%)、 $t \in [3.75, 4.25]$ (验证集，10%) 和 $t \in [4.25, 5]$ (测试集，15%) 划分。

图 6 显示的趋势总体上与 Burgers方程一致：对于无噪声数据，所有方法的表现相近，且随着 r 最高达到 5 而有所改善。随着噪声水平（NL）增加，OpInf-ord6 的性能比 ord2 下降得更快——高阶差分会放大噪声，而 FKPP 的反应项会保留这些

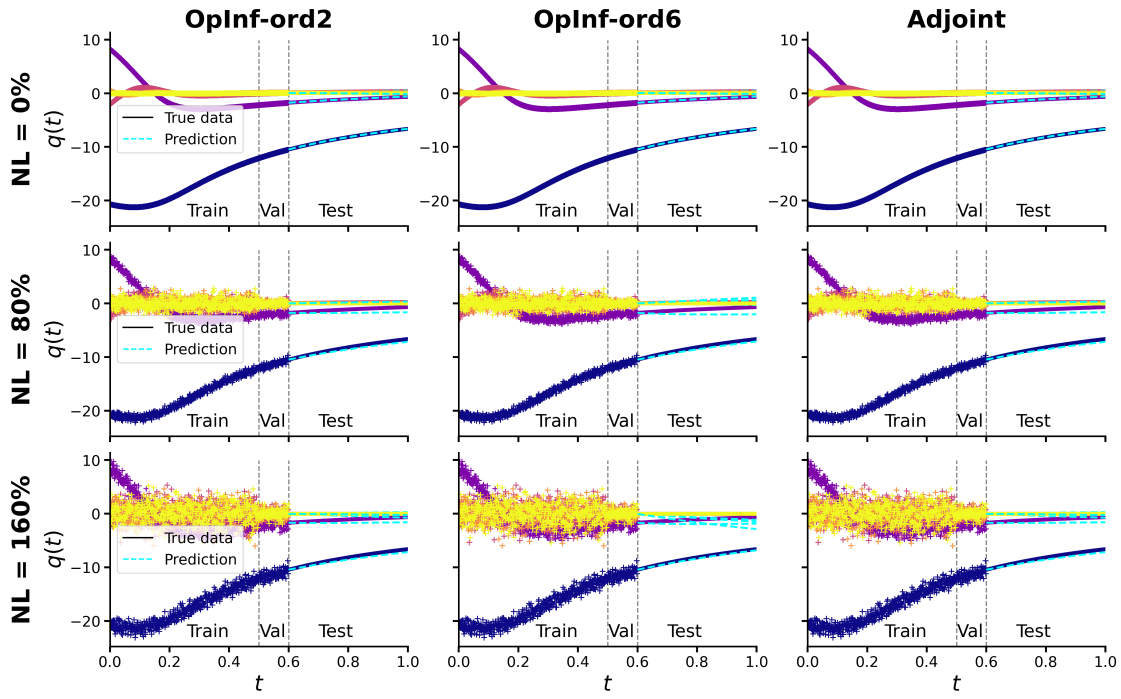


图3：测试窗口内黏性伯格斯方程的降维坐标滚动预测。ROM维数为 $r = 5$；训练集、验证集和测试集共包含 1000 个快照。各行对应不同的噪声水平（NL $= 0\%$、80%、160%）；各列对比 OpInf-ord2、OpInf-ord6 和伴随方法（本文方法）。灰色竖直虚线标示训练/验证/测试划分。彩色标记表示用于训练/验证的含噪观测 $\mathbf{q}_{\text{true}}(t)$；实线表示无噪参考轨迹；青色虚线表示以测试初始条件为起点的模型预测。在强噪声下，伴随法滚动预测更贴近无噪参考轨迹，尤其是在训练窗口之外。

高频伪影。伴随训练的降阶模型在大多数面板中取得最低或接近最低的测试相对平方误差，尤其是在样本数为 200–1000 且 NL 为 $\geq 80\%$ 时；即使在 NL 为 $= 80\%$ 时，性能也能平稳退化。在有噪声的情况下，增加 r 带来的收益会更早达到饱和（通常为 $r = 3$ 或 4）；OpInf 往往趋于平稳或性能变差，而伴随方法则保持稳定。

在降维坐标中（图7），伴随法滚动预测在整个测试窗口内都紧贴无噪参考轨迹；映射回 $t = 5$ 处的全阶模型网格（图8）后，伴随方法保留了平滑的径向对称轮廓，而 OpInf 则出现颗粒状环形伪影；当 NL 为 $\geq 80\%$ 时，ord6 的伪影最为明显。

4.3. 对流–扩散方程（*ADE*）

我们考虑线性对流–扩散方程

$$u_t \;=\; c_x\, u_x \;+\; c_y\, u_y \;+\; \nu\,(u_{xx} + u_{yy}), \qquad (x, y) \in [-1, 1]^2,\; t \in [0, T],$$

其对流速度为常数 $c_x = 1$、 $c_y = 1.5$ ，黏度为 $\nu = 0.005$。我们在两个空间方向上施加周期边界条件，并以高斯分布作为初始条件，

$$u(x, y, 0) \;=\; \exp\!\Big(-\tfrac{(x-x_0)^2+(y-y_0)^2}{2\sigma^2}\Big), \quad (x_0, y_0) = (-0.5, -0.5),\; \sigma = 0.1.$$

偏微分方程在均匀网格上求解，每个方向包含 $N_x = N_y = 201$ 个点（$\Delta x = \Delta y = 2/(N_x - 1)$），并采用均匀时间网格 $T = 0.5$， $\Delta t = 10^{-3}$。空间导数采用带周期延拓的二阶中心差分；时间推进采用显式前向欧拉法。所选 Δt满足标准的对流和扩散 Courant–Friedrichs–Lewy（CFL）约束。将整个时间段内的解堆叠起来可得到快照矩阵 $\mathbf{U} \in \mathbb{R}^{(N_x N_y)\,\times\,(N_t+1)}$ ，我们使用该矩阵进行 POD 和降阶模型训练（见图 9）。

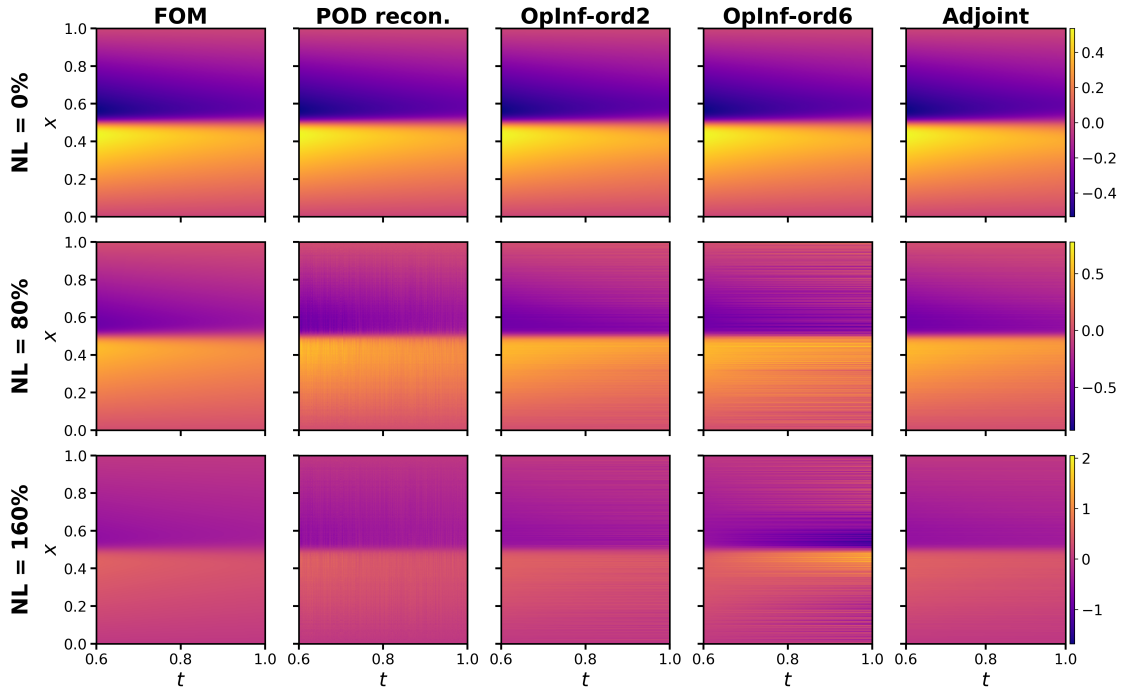


图4：测试窗口内黏性伯格斯方程的时空演化 $u(x,\ t)$。降阶模型维数 $r = 5$；训练集、验证集和测试集中共包含 1000 个快照。各行对应不同的 NL $= 0\%$、80%、160%；各列展示全阶模型，以及分别使用 POD、OpInf-ord2、OpInf-ord6 和伴随降阶模型得到的相应重构。色彩图采用分面尺度。面对噪声，伴随模型最能保留演化剖面的位置和振幅；而 OpInf-ord6 出现明显的噪声印迹，OpInf-ord2 则表现出偏差和平滑效应。

我们采用与 Burgers方程和 FKPP 研究相同的评估方案（OpInf 初始化、伴随训练、测试窗口相对平方误差），但使用更短的时间跨度 $T = 0.5$，数据划分为 $t \in [0, 0.375]$（训练集，75%）、$t \in [0.375, 0.425]$（验证集，10%）和 $t \in [0.425, 0.5]$（测试集，15%），并采用更广的降阶模型范围 $r \in \{1, ..., 15\}$（包含 $\kappa_{15} \geq 99.5$），对应的采样数量为 $\{200, 500, 1000, 2001\}$。

图 10显示，在无噪声数据下，伴随训练与 OpInf-ord6 的准确度相当，并且会随着 r 而提高。随着噪声水平上升（NL ≥ 80），OpInf基线方法变得越来越不稳定，且 r 越大，误差增长越明显，尤其是 ord6 和样本数较少时；相比之下，伴随方法在不同维度下均保持稳定，并且在大多数设置中取得最低或接近最低的测试相对平方误差。这一现象与基于导数的回归易受对流噪声放大影响相符，而伴随形式避免了这种影响。

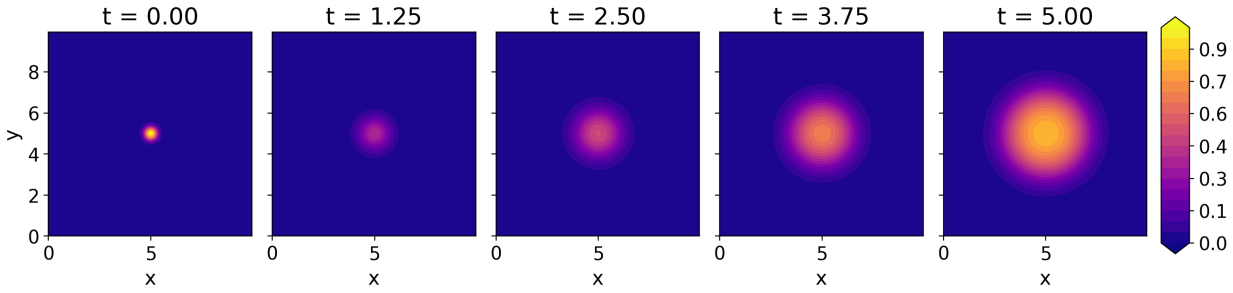


图5：Fisher-KPP方程的数值解。 $u(x,\ t_i)$的时空等值图。

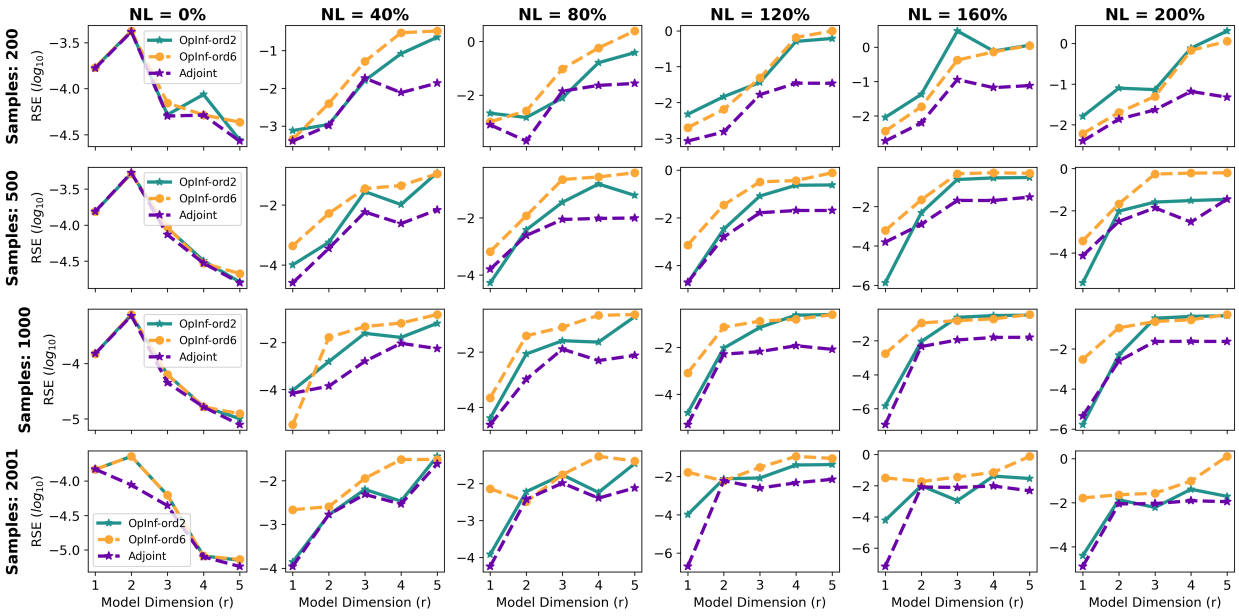


图6：在噪声水平和采样方式变化时，模型在Fisher-KPP方程上的表现。列表示不同的噪声水平（NL = 状态标准差的百分比）；行表示训练集、验证集和测试集中的快照数量（2001, 500, 1000, 200）。每个面板展示测试相对平方误差（$\log_{10}$）随ROM维数 r 的变化。方法包括：伴随方法（本文提出）、OpInf-ord2和OpInf-ord6）。数值越低越好。

在降维坐标（$r = 15$，1000 个快照；图 11）中，随着噪声增加，伴随法滚动预测在整个测试区间内仍保持连贯，而 OpInf轨迹在训练集与验证集的分界处之后逐渐偏移并发散。最终时刻全阶模型网格上的重构结果（图 12）表明，伴随模型保留了平滑、平流输运的高斯剖面，而 OpInf预测则出现斑点状平流伪影，OpInf-ord6 尤为明显。总体而言，ADE结果与 Burgers方程和 FKPP案例一致，在高噪声、平流主导区域中，不同方法之间的差异更加显著。

5. 总结

我们提出了一种用于二次降阶模型的基于伴随法的训练框架，通过最小化连续时间轨迹损失，并利用反向伴随求解计算精确梯度，从而避免了标准算子推断所需的易出错时间导数估计；在时间采样稀疏且存在测量噪声时，这类估计尤其不可靠。在所提出的基于伴随法的训练中，每个优化步骤的计算成本为一次正向和一次反向时间积分，与模型参数数量无关。

在三个具有代表性的偏微分方程基准测试中，伴随训练的降阶模型在无噪声数据上始终达到或优于标准OpInf；在两种扰动方案下也表现出明显更强的鲁棒性：通过均匀下采样降低快照密度，以及逐步增加加性高斯噪声水平。具体而言，在高阶有限差分OpInf（ord6）迅速退化的情况下，伴随形式仍保持稳定；随着噪声增加或数据变得稀疏，其表现通常优于对噪声更稳健的低阶基线（ord2）。这些结果支持了核心观点：与匹配估计导数相比，在连续时间内拟合轨迹，能在训练数据不完美时得到更可靠的降阶算子。其主要局限在于依赖多项式模型类、对时间离散化的准确度较为敏感，以及运行时间长于最小二乘回归。

未来工作的自然方向包括若干方面。例如，将保结构约束纳入训练方案，或可进一步提升长时域泛化能力和稳定性。此外，

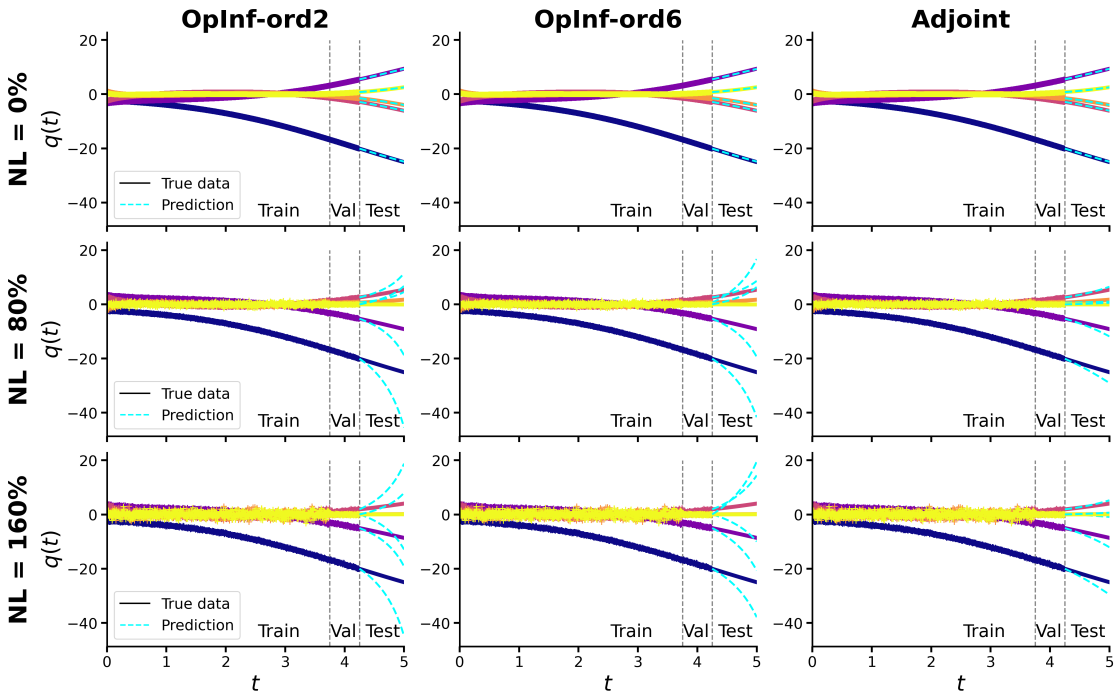


图7：Fisher-KPP方程在测试窗口上的降维坐标滚动预测。ROM维数为 $r = 5$；训练集、验证集和测试集共包含1000个快照。各行对应不同的噪声水平（NL $= 0\%$、80%、160%）；各列对比 OpInf-ord2、OpInf-ord6 和伴随方法（本文方法）。灰色竖向虚线标示训练/验证/测试划分。彩色标记表示用于训练/验证的含噪观测 $\mathbf{q}_{\text{true}}(t)$；实线表示无噪参考轨迹；青色虚线表示以测试初始条件为起点的模型预测。在强噪声下，伴随方法的滚动预测更贴近无噪声轨迹，尤其是在训练窗口之外。

将所提出的方法扩展到处理物理性质的参数变化和由不同初始条件产生的多条轨迹，并评估其在非高斯分布噪声和相关噪声下的鲁棒性，将有助于提升其在实验环境和传感器数据集成中的适用性。总之，所提出的基于伴随的策略为学习多项式ROM提供了一种实用且灵活的方法，可显著降低其对稀疏和含噪数据的敏感性。

利益冲突声明

作者声明不存在利益冲突。

数据可用性

用于复现实验的代码可在 https://github.com/lindliu/adjoint_opinf。

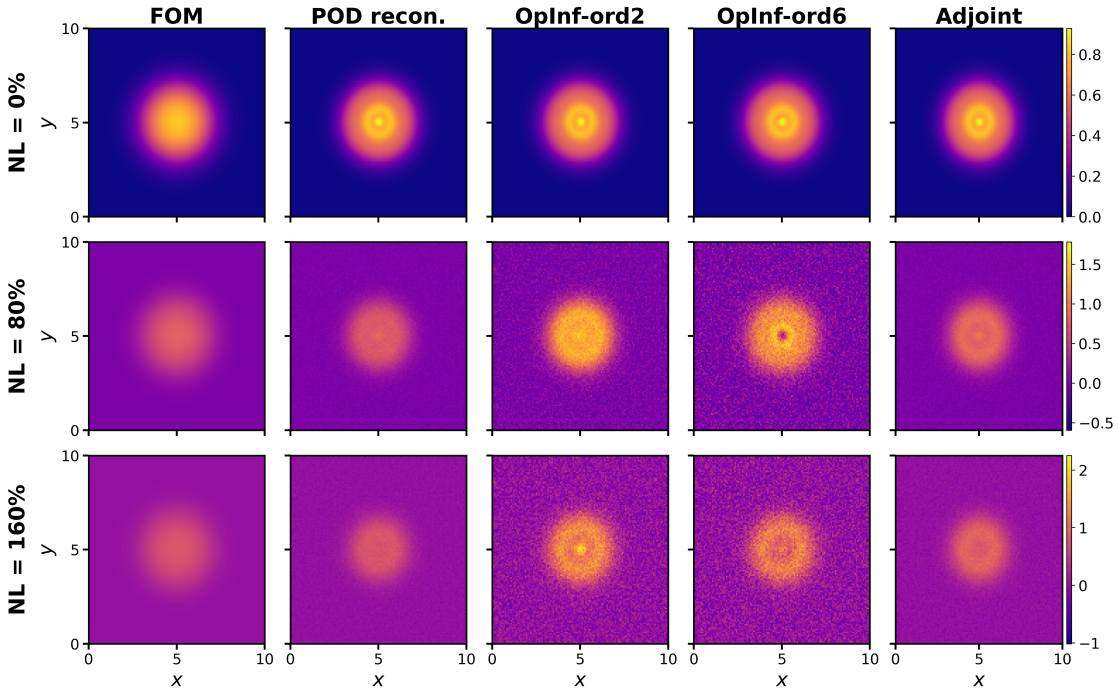


图8：Fisher-KPP方程在 $t = 0.5$ 时的空间场。ROM维数为 $r = 5$；训练集、验证集和测试集共包含 1000 个快照。各行对应不同的 NL $= 0\%$、80%、160%；各列依次展示全阶模型及使用 POD、OpInf-ord2、OpInf-ord6 和伴随降阶模型进行的相应重构。色彩图采用分面尺度。伴随模型在噪声影响下最能保留演化剖面的位置和幅值，而 OpInf-ord6 呈现出明显的噪声印迹，OpInf-ord2 则表现出偏差和平滑效应。

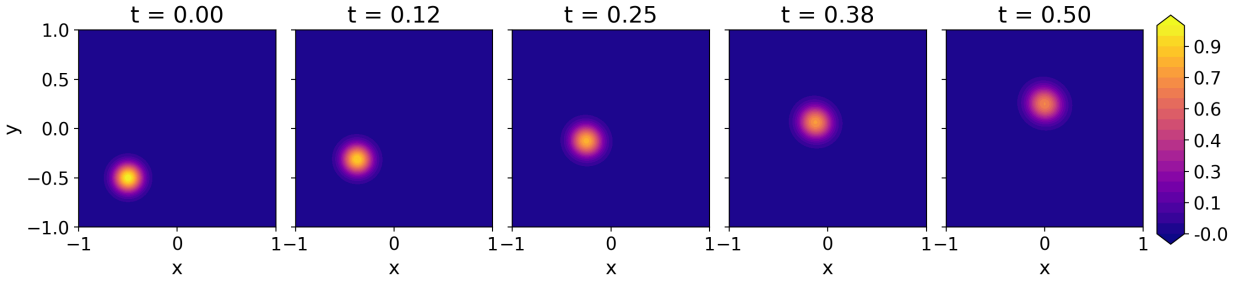


图9：对流-扩散方程的数值解。 $u(x,\ t_i)$ 的时空等值图。

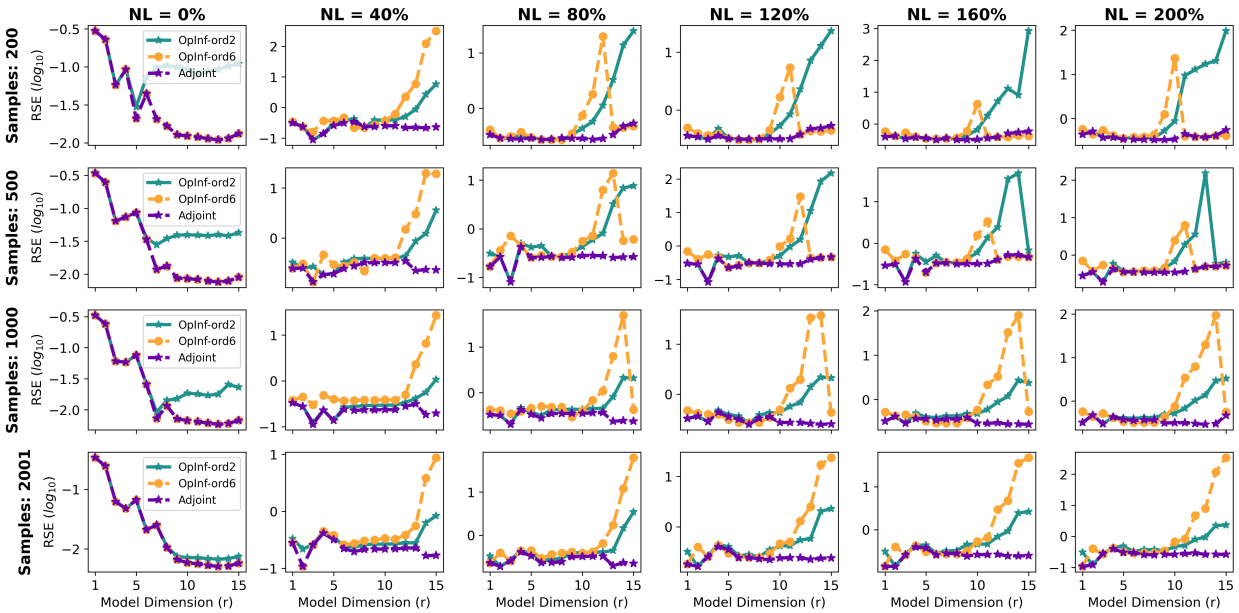


图10：不同噪声水平和采样设置下，模型在对流-扩散方程上的性能。列表示不同的噪声水平（NL $=0$–状态标准差的200%）；行表示训练集、验证集和测试集中的快照数量（200、500、1000、2001）。每个面板显示测试相对平方误差（$\log_{10}$）随ROM维数 r 的变化。方法：伴随方法（本文方法）、OpInf-ord2 和 OpInf-ord6。数值越低越好。

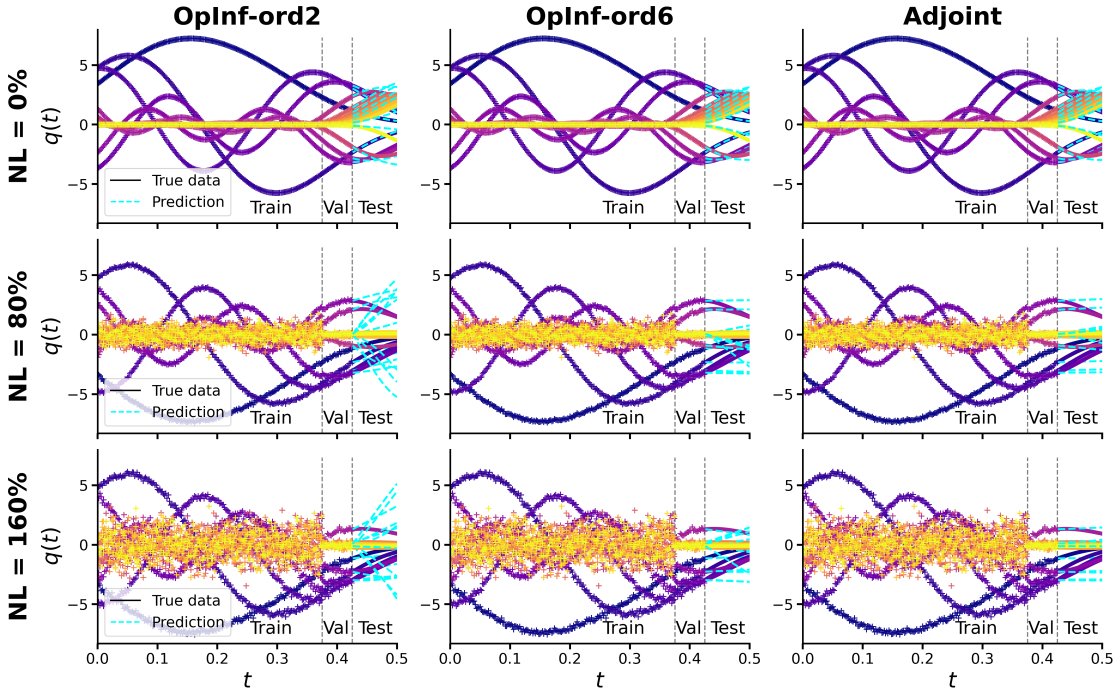


图11：对流-扩散方程在测试窗口上的降维坐标滚动预测。ROM维数 $r = 15$；训练集、验证集和测试集中共包含1000个快照。行表示不同的噪声水平（NL $=0$%、80%、160%）；列比较 OpInf-ord2、OpInf-ord6 和伴随方法（本文方法）。灰色竖直虚线标记训练集/验证集/测试集划分。彩色标记表示用于训练/验证的含噪观测$\mathbf{q}_{\text{true}}(t)$；实线表示无噪参考轨迹；青色虚线表示以测试初始条件初始化的模型预测。在强噪声下，伴随方法的滚动预测更贴近无噪参考轨迹，尤其是在训练窗口之后。

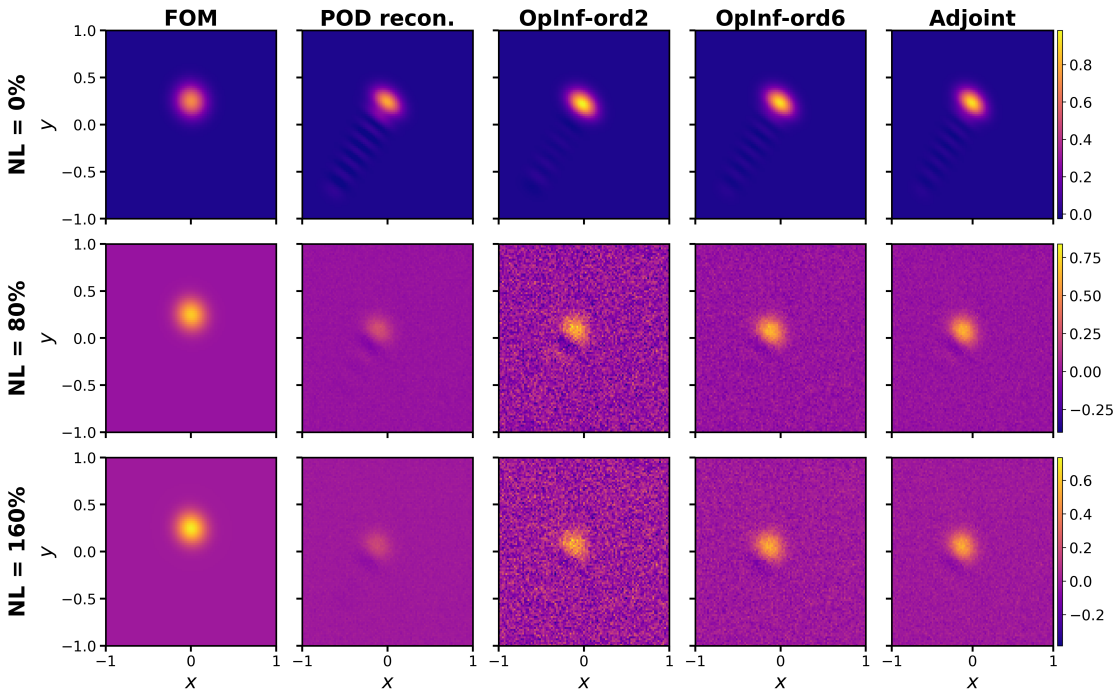


图12： $t = 0.5$处对流-扩散方程的空间场。ROM维数为 $r = 15$；训练集、验证集和测试集共包含 1000 个快照。各行对应不同的 NL $= 0\%$、80%、160%；各列依次展示全阶模型以及使用 POD、OpInf-ord2、OpInf-ord6 和伴随降阶模型进行的相应重构。色彩图采用分面尺度。伴随模型在噪声下最能保留演化剖面的位置和幅值，而 OpInf-ord6 呈现明显的噪声印迹，OpInf-ord2 则表现出偏差和平滑效应。

附录 A. 算法

算法 1： Armijo回溯线搜索（最速下降法）

输入： 当前迭代点 $\boldsymbol{\theta}^j$；目标函数 $\tilde{\ell}$；参数 $\alpha \in (0,\ 1)$、$\beta \in (0,\ 1)$、$\gamma \in (0,\ 1)$；初始步长 $\eta_0 > 0$；最大回溯次数 $N_{\max}$

输出： 下一个迭代点 $\boldsymbol{\theta}^{j+1}$、更新后的初始值 η_0、接受的步长 η_j

集合 $\mathscr{G}^j \leftarrow \nabla\tilde{\ell}(\boldsymbol{\theta}^j)$；$\eta \leftarrow \eta_0$；接受 $\leftarrow$false；

对于 $i = 1, \ldots, N_{\max}$ **do**

  if $\tilde{\ell}(\boldsymbol{\theta}^j - \eta\mathscr{G}^j) \leq \tilde{\ell}\ (\boldsymbol{\theta}^j) - \alpha\,\eta\,\|\mathscr{G}^j\|_2^2$ **然后**

    已接受 $\leftarrow$true;

    $\boldsymbol{\theta}^{j+1} \leftarrow \boldsymbol{\theta}^j - \eta\mathscr{G}^j$;

    $\eta_j \leftarrow \eta$;

    返回

  end

  $\eta \leftarrow \beta\,\eta$; // 回溯

end

如果未接受**则**

  $\boldsymbol{\theta}^{j+1} \leftarrow \boldsymbol{\theta}^j$; $\eta_0 \leftarrow \gamma\,\eta_0$; $\eta_j \leftarrow 0$;

  返回

end

算法 2： 基于伴随法的降阶模型参数训练

输入： 全阶模型快照 $\mathbf{U}$；时间窗口 $[t_1, t_k]$；Armijo 参数（α，β，γ，η_0）；容差 ϵ；最大迭代次数 $J_{\max}$；POD 秩 r；权重指数 p

输出： 优化参数 $\boldsymbol{\theta}^*$

POD 投影： 计算 POD 基$\mathbf{V}_r$（能量阈值或固定 r）和降阶快照 $[\mathbf{q}_{\text{true}}(t_1) \ldots, \mathbf{q}_{\text{true}}(t_k)] \leftarrow \mathbf{V}_r^\top\mathbf{U}$。令 $\{\lambda_i\}_{i=}^{r}{}_{1}$ 为奇异值平方；

逐状态权重： 通过对每个状态进行 Savitzky–Golay 平滑来估计噪声方差 ν_i^2，并残差方差。设置 $\omega_i \leftarrow \sigma_i^p/\nu_i^2$（默认值 $p = 1$ 为温和信噪比）和 $\boldsymbol{W} \leftarrow \text{diag}(\omega_1, \ldots, \omega_r)$；

初始化： 通过 $\boldsymbol{\theta}^0$ 从 OpInf（公式（4））获取，并在一个验证集上选择岭回归/截断奇异值分解超参数验证集；

对于 $j = 0,\ 1,\ \ldots,\ J_{\max} - 1$ **执行**

  前向求解： 在 $[t_1, t_k]$ 上对降阶常微分方程（3）进行积分，使用参数 $\boldsymbol{\theta}^j$ 以获得 $\tilde{\mathbf{q}}(t;\ \boldsymbol{\theta}^j)$；

  加权残差： 对于每个 t，设置 $\partial_{\mathbf{q}}g(t) \leftarrow 2\,W\big(\tilde{\mathbf{q}}(t;\ \boldsymbol{\theta}^j) - \mathbf{q}_{\text{true}}(t)\big)$；

  伴随求解： 从 9 反向积分 t_k 至 t_1，终端条件为 $\boldsymbol{\lambda}(t_k) = \mathbf{0}$，使用 $\partial_{\mathbf{q}}\,g(t)$ 和 $\partial_{\mathbf{q}}\,\mathbf{f}$。插值 $\partial_{\mathbf{q}}\,g(t)$ 在求解器网格上；

  参数梯度： 计算 $\nabla_{\boldsymbol{\theta}}\ \tilde{\ell}\ (\boldsymbol{\theta}^j)$ 通过（8）；

  线搜索与更新： 应用 Armijo回溯（算法 1）以获得 $\boldsymbol{\theta}^{j+1}$ 以及更新后的种子 η_0；

  停止：如果 $\|\nabla_{\boldsymbol{\theta}}\ \tilde{\ell}\ (\boldsymbol{\theta}^{j+1})\ \|_2 \leq \epsilon$ **则**

    中断

  end

end

返回 $\boldsymbol{\theta}^* \leftarrow \boldsymbol{\theta}^{j+1}$。

附录 B. 命题 1

命题1 遵循标准的伴随论证（例如，参见 [23, 24]）。我们将其调整为适用于我们的情形。

附录 *B.1.* 拉格朗日公式证明

证明。 约束优化式（7）对应的拉格朗日函数为

$$\mathcal{L}\big(\mathbf{q}(\cdot),\boldsymbol{\theta},\boldsymbol{\lambda}(\cdot)\big) := \ell\big(\mathbf{q}(\cdot)\big) + \int_0^T \boldsymbol{\lambda}(t)^\top[\mathbf{f}(\mathbf{q}(t);\boldsymbol{\theta}) - \dot{\mathbf{q}}(t)]\ dt - \boldsymbol{\lambda}(0)^\top\big(\mathbf{q}(0) - \mathbf{q}_0\big).$$

Be由于初始条件 $\mathbf{q}(0) = \mathbf{q}_0$ 与 $\boldsymbol{\theta}$ 无关，因此可将拉格朗日函数简化为 s

$$\mathcal{L}(\mathbf{q}(\cdot;\boldsymbol{\theta}),\boldsymbol{\theta},\boldsymbol{\lambda}(\cdot)) := \int_0^T \Big(g(\mathbf{q},t)\ +\ \boldsymbol{\lambda}(t)^\top\big(\mathbf{f}(\mathbf{q};\boldsymbol{\theta}) - \dot{\mathbf{q}}\big)\Big)\,\mathrm{d}t.$$

由于常微分方程约束对 $\tilde{\mathbf{q}}(\cdot;\boldsymbol{\theta})$ 成立，因此对于任意 $\boldsymbol{\lambda}(\cdot) \in \mathcal{C}^1(\mathcal{T};\mathbb{R}^r)$,

$$\mathcal{L}\big(\tilde{\mathbf{q}}(\cdot;\boldsymbol{\theta}),\boldsymbol{\theta},\boldsymbol{\lambda}(\cdot)\big) = \tilde{\ell}(\boldsymbol{\theta}).$$

在所述光滑性条件下，积分号下求导和分部积分均成立。对 $\boldsymbol{\theta}$ 求导，同时将 $\boldsymbol{\lambda}$视为辅助函数，可得

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\tilde{\ell}(\boldsymbol{\theta}) = \frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\mathcal{L}(\tilde{\mathbf{q}}(\cdot;\boldsymbol{\theta}),\boldsymbol{\theta},\boldsymbol{\lambda}(\cdot)) = \int_0^T \left[\underbrace{\frac{\partial g}{\partial \boldsymbol{\theta}}}_{=0} + \frac{\partial g}{\partial \tilde{\mathbf{q}}}\frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}} + \boldsymbol{\lambda}(t)^\top\left(\frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}} + \frac{\partial \mathbf{f}}{\partial \tilde{\mathbf{q}}}\frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}}\right)\right]\mathrm{d}t.$$

对最后一项进行分部积分，得

$$-\int_0^T \boldsymbol{\lambda}^\top \frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}}\Big)\,\mathrm{d}t = -\boldsymbol{\lambda}^\top \frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}}\Big|_0^T + \int_0^T \dot{\boldsymbol{\lambda}}^\top \frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}}\,\mathrm{d}t.$$

由于 $\tilde{\mathbf{q}}(0)$ 不依赖于 $\boldsymbol{\theta}$，因此 $\frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}}(0) = \mathbf{0}$，从而

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\tilde{\ell}(\boldsymbol{\theta}) = \int_0^T \Big(\boldsymbol{\lambda}^\top \frac{\partial \mathbf{f}}{\partial \boldsymbol{\theta}} + \underbrace{\big[\tfrac{\partial g}{\partial \tilde{\mathbf{q}}} + \boldsymbol{\lambda}^\top \tfrac{\partial \mathbf{f}}{\partial \tilde{\mathbf{q}}} + \dot{\boldsymbol{\lambda}}^\top\big]\frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}}}_{\text{sensitivity term}}\Big)\,\mathrm{d}t\ -\ \boldsymbol{\lambda}(T)^\top \frac{\partial \tilde{\mathbf{q}}}{\partial \boldsymbol{\theta}}(T). \tag{B.1}$$

此处，可以选择伴随变量 $\boldsymbol{\lambda}$，通过施加以下条件消除灵敏度项

$$\dot{\boldsymbol{\lambda}}(t) = -\Big(\frac{\partial \mathbf{f}}{\partial \tilde{\mathbf{q}}}\Big)^\top \boldsymbol{\lambda}(t) - \Big(\frac{\partial g}{\partial \tilde{\mathbf{q}}}\Big)^\top, \qquad \boldsymbol{\lambda}(T) = \mathbf{0},$$

这使得括号中的项和终端边界项都消失，从而得到伴随常微分方程组。最终，(B.1) 简化为

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{\theta}}\tilde{\ell}(\boldsymbol{\theta}) = \int_0^T \boldsymbol{\lambda}(t)^\top \frac{\partial \mathbf{f}(\tilde{\mathbf{q}}(t);\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\,\mathrm{d}t,$$

证明至此完成。

□

附录 *B.2.* 原始灵敏度分析的证明

我们在此 给出另一种证明，因为它通过原始灵敏度分析对该问题提供了更多见解 l

敏感性分析 分析。

证明。我们先写出状态解 $\tilde{\mathbf{q}}(t,\ \boldsymbol{\theta})$ 关于参数 $\theta_i, i=1,\ \ldots\ldots,\ d$ 的原始灵敏度方程：

$$\dot{\tilde{\mathbf{q}}}(t)=\mathbf{f}\big(\tilde{\mathbf{q}}(t,\boldsymbol{\theta}),\boldsymbol{\theta}\big),\qquad \partial_{\theta_i}\dot{\tilde{\mathbf{q}}}(t)=\frac{\partial\mathbf{f}}{\partial\tilde{\mathbf{q}}}\big(\tilde{\mathbf{q}}(t,\boldsymbol{\theta}),\boldsymbol{\theta}\big)\,\partial_{\theta_i}\tilde{\mathbf{q}}+\partial_{\theta_i}\mathbf{f}\big(\tilde{\mathbf{q}},\boldsymbol{\theta}\big).$$

将第二个方程乘以任意测试函数 $\mathbf{v}(\cdot)\in\mathcal{C}^1(\mathcal{T};\ \mathbb{R}^r)$，在 0 到 T 上积分并重新整理，可将原始问题改写为其弱形式

$$\int_0^T\left[\partial_{\theta_i}\dot{\tilde{\mathbf{q}}}(t,\boldsymbol{\theta})-\left(\frac{\partial\mathbf{f}}{\partial\tilde{\mathbf{q}}}\right)\partial_{\theta_i}\tilde{\mathbf{q}}(t,\boldsymbol{\theta})\right]^\top\mathbf{v}(t)\,\mathrm{d}t=\int_0^T\big(\partial_{\theta_i}\mathbf{f}\big)^\top\mathbf{v}(t)\,\mathrm{d}t,$$

或者，用算子记号表示，

$$\langle\mathscr{L}_{\boldsymbol{\theta}}\ \partial_{\theta_i}\tilde{\mathbf{q}},\ \mathbf{v}\rangle=\langle\partial_{\theta_i}\mathbf{f},\ \mathbf{v}\rangle,$$

其中，算子 $\mathscr{L}_{\boldsymbol{\theta}}\colon\ \mathcal{C}^1(\mathcal{T};\mathbb{R}^r)\ \to\mathcal{C}(\mathcal{T};\ \mathbb{R}^r)$ 定义为

$$(\mathscr{L}_{\boldsymbol{\theta}}\mathbf{w})(t)=\dot{\mathbf{w}}(t)-\frac{\partial\mathbf{f}}{\partial\tilde{\mathbf{q}}}(t,\boldsymbol{\theta})\mathbf{w}(t),\quad \mathbf{w}(\cdot)\in\mathcal{C}^1(\mathcal{T};\mathbb{R}^r),$$

且 $\langle\cdot,\cdot\rangle$ 表示通常的 L_2-内积。

另一方面，损失函数关于 θ_i 的梯度为 ℓ

$$\frac{\mathrm{d}}{\mathrm{d}\theta_i}\,\tilde{\ell}(\boldsymbol{\theta})=\int_0^T\frac{\partial g}{\partial\tilde{\mathbf{q}}}\big(\tilde{\mathbf{q}}(t,\boldsymbol{\theta}),t\big)\,\partial_{\theta_i}\tilde{\mathbf{q}}(t,\boldsymbol{\theta})\,\mathrm{d}t=\langle\nabla_{\tilde{\mathbf{q}}}\,g,\,\partial_{\theta_i}\tilde{\mathbf{q}}\rangle.$$

现在，我们通过引入伴随变量 $\boldsymbol{\lambda}(\cdot)\in\mathcal{C}^1(\mathcal{T};\mathbb{R}^r)$ 来定义伴随问题，使其满足

$$\langle\mathbf{w},\,\mathscr{L}_{\boldsymbol{\theta}}^*\,\boldsymbol{\lambda}\rangle=\langle\mathbf{w},\nabla_{\tilde{\mathbf{q}}}\,g\,\rangle,\quad\forall\,\mathbf{w}(\cdot)\in\mathcal{C}^1(\mathcal{T},\mathbb{R}^r)\ \text{with}\ \mathbf{w}(0)=\mathbf{0},$$

其中 $\mathscr{L}_{\boldsymbol{\theta}}^*$ 是 $\mathscr{L}_{\boldsymbol{\theta}}$ 的伴随算子。根据伴随算子的定义，这等价于

$$\langle\mathscr{L}_{\boldsymbol{\theta}}\,\mathbf{w},\,\boldsymbol{\lambda}\rangle=\langle\nabla_{\tilde{\mathbf{q}}}\,g,\,\mathbf{w}\rangle.$$

在此恒等式中取 $\mathbf{w}\ =\partial_{\theta_i}\tilde{\mathbf{q}}$，立即得到

$$\frac{\mathrm{d}}{\mathrm{d}\theta_i}\,\tilde{\ell}(\boldsymbol{\theta})=\big\langle\mathscr{L}_{\boldsymbol{\theta}}\,\partial_{\theta_i}\tilde{\mathbf{q}},\,\boldsymbol{\lambda}\big\rangle\underset{\substack{\text{primal problem}\\ \text{with }\mathbf{v}=\boldsymbol{\lambda}}}{=}\big\langle\partial_{\theta_i}\mathbf{f},\,\boldsymbol{\lambda}\big\rangle=\int_0^T\boldsymbol{\lambda}^\top\,\partial_{\theta_i}\mathbf{f}(t,\boldsymbol{\theta})\,\mathrm{d}t.$$

为推导伴随常微分方程本身，我们将伴随问题的弱形式展开为

$$\int_0^T\left[\dot{\mathbf{w}}(t)-\frac{\partial\mathbf{f}}{\partial\tilde{\mathbf{q}}}\mathbf{w}(t)\right]^\top\boldsymbol{\lambda}(t)\,\mathrm{d}t=\int_0^T\frac{\partial g}{\partial\tilde{\mathbf{q}}}\mathbf{w}(t)\,\mathrm{d}t.$$

分部积分可得

$$\int_0^T\dot{\mathbf{w}}(t)^\top\boldsymbol{\lambda}(t)\,\mathrm{d}t=\big[\mathbf{w}(t)^\top\boldsymbol{\lambda}(t)\big]_0^T-\int_0^T\dot{\boldsymbol{\lambda}}(t)^\top\mathbf{w}(t)\ \mathrm{d}t.$$

利用 $\mathbf{w}(0)\ =\mathbf{0}$ 并施加伴随终端条件 $\boldsymbol{\lambda}(T)\ =\mathbf{0}$ 以消去边界项，我们得到

$$\int_0^T\big[-\dot{\boldsymbol{\lambda}}(t)\big]^\top\mathbf{w}(t)\,\mathrm{d}t=\int_0^T\left[\left(\frac{\partial\mathbf{f}}{\partial\tilde{\mathbf{q}}}\right)^\top\boldsymbol{\lambda}(t)+\left(\frac{\partial g}{\partial\tilde{\mathbf{q}}}\right)^\top\right]^\top\mathbf{w}(t)\,\mathrm{d}t.$$

由于 $\mathbf{w}$ 是任意的，两侧的被积函数应当相等，由此得到伴随方程的最终表达式

$$\dot{\boldsymbol{\lambda}}(t) = -\left[\left(\frac{\partial \mathbf{f}}{\partial \tilde{\mathbf{q}}}\right)^{\top} \boldsymbol{\lambda}(t) + \left(\frac{\partial g}{\partial \tilde{\mathbf{q}}}\right)^{\top}\right], \quad \boldsymbol{\lambda}(T) = \mathbf{0}.$$

□

附录 C. 微分恒等式

我们使用方向导数 $\mathcal{D}h(\mathbf{q})[\mathbf{v}] := \lim_{\varepsilon \to 0} {}^{h(\mathbf{q}+\varepsilon \mathbf{v}) - h(\mathbf{q})}/_{\varepsilon}$。对于向量场 $h: \mathbb{R}^r \to \mathbb{R}^r$，雅可比矩阵 $\partial_{\mathbf{q}} h(\mathbf{q})$ 是满足以下条件的唯一线性映射：对所有 $\mathbf{v} \in \mathbb{R}^r$，均有 $\mathcal{D}h(\mathbf{q})[\mathbf{v}] = \big(\partial_{\mathbf{q}} h(\mathbf{q})\big)\mathbf{v}$。

设定。对于二次降阶模型

$$\mathbf{f}(\mathbf{q}; \boldsymbol{\theta}) = \mathbf{c} + \mathbf{A}\,\mathbf{q} + \mathbf{H}(\mathbf{q} \otimes \mathbf{q}) + \mathbf{B}\,\mathbf{s}(t), \quad \boldsymbol{\theta} = [\mathbf{c}, \mathbf{A}, \mathbf{H}, \mathbf{B}],$$

我们计算式（9）–（8）中出现的导数。

状态雅可比矩阵 $\partial_{\mathbf{q}} \mathbf{f}$。对于任意方向 $\mathbf{v} \in \mathbb{R}^r$，

$$\mathbf{f}(\mathbf{q} + \varepsilon \mathbf{v}; \boldsymbol{\theta}) = \mathbf{c} + \mathbf{A}\mathbf{q} + \varepsilon\,\mathbf{A}\mathbf{v} + \mathbf{H}(\mathbf{q} \otimes \mathbf{q}) + \varepsilon\,\mathbf{H}(\mathbf{q} \otimes \mathbf{v} + \mathbf{v} \otimes \mathbf{q}) + \mathbf{B}\mathbf{s} + \mathcal{O}(\varepsilon^2),$$

so

$$\mathcal{D}\mathbf{f}(\mathbf{q})[\mathbf{v}] = \mathbf{A}\,\mathbf{v} + \mathbf{H}(\mathbf{v} \otimes \mathbf{q}) + \mathbf{H}(\mathbf{q} \otimes \mathbf{v}) = \big(\mathbf{A} + \mathbf{H}(\mathbf{I}_r \otimes \mathbf{q}) + \mathbf{H}(\mathbf{q} \otimes \mathbf{I}_r)\big)\mathbf{v}.$$

因此，

$$\partial_{\mathbf{q}} \mathbf{f}(\mathbf{q}; \boldsymbol{\theta}) = \mathbf{A} + \mathbf{H}(\mathbf{I}_r \otimes \mathbf{q}) + \mathbf{H}(\mathbf{q} \otimes \mathbf{I}_r) \in \mathbb{R}^{r \times r}.$$

由于二次张量在其最后两个模上对称，$\mathbf{H}_{i,jk} = \mathbf{H}_{i,kj}$（即对所有 $\mathbf{q}, \mathbf{v}$，$\mathbf{H}(\mathbf{q} \otimes \mathbf{v}) = \mathbf{H}(\mathbf{v} \otimes \mathbf{q})$），这可简化为 $\partial_{\mathbf{q}} \mathbf{f} = \mathbf{A} + 2\,\mathbf{H}(\mathbf{I}_r \otimes \mathbf{q})$。

损失梯度 $\partial_{\mathbf{q}} g$。对于 $g(\mathbf{q}\ t) = \|\mathbf{q} - \mathbf{q}_{\text{true}}(t)\|_2^2$,

$$\mathcal{D}g(\mathbf{q})[\mathbf{v}] = 2\big(\mathbf{q} - \mathbf{q}_{\text{true}}(t)\big)^{\top}\mathbf{v} \quad \Rightarrow \quad \partial_{\mathbf{q}} g(\mathbf{q}, t) = 2\big(\mathbf{q} - \mathbf{q}_{\text{true}}(t)\big)^{\top} \in \mathbb{R}^{1 \times r}.$$

采用对角权重 $\boldsymbol{W}$, $g = \|\boldsymbol{W}^{1/2}(\mathbf{q} - \mathbf{q}_{\text{true}})\|_2^2$ 得到 $\partial_{\mathbf{q}} g = 2\left(\boldsymbol{W}(\mathbf{q} - \mathbf{q}_{\text{true}})\right)^{\top}$。

参数导数 $\partial_{\boldsymbol{\theta}} \mathbf{f}$。令 $\mathrm{vec}(\cdot)$ 将矩阵列堆叠起来。利用 $\mathrm{vec}(\mathbf{A}\mathbf{q}) = (\mathbf{q}^{\top} \otimes \mathbf{I}_r)\,\mathrm{vec}(\mathbf{A})$ 以及 $\mathrm{vec}\big(\mathbf{H}(\mathbf{q} \otimes \mathbf{q})\big) = ((\mathbf{q} \otimes \mathbf{q})^{\top} \otimes \mathbf{I}_r)\,\mathrm{vec}(\mathbf{H})$，可得

$$\partial_{\boldsymbol{\theta}} \mathbf{f}(\mathbf{q}; \boldsymbol{\theta}) = \big[\ \underbrace{\mathbf{I}_r}_{\partial \mathbf{f}/\partial \mathbf{c}}\ ,\ \underbrace{\mathbf{q}^{\top} \otimes \mathbf{I}_r}_{\partial \mathbf{f}/\partial\,\mathrm{vec}(\mathbf{A})}\ ,\ \underbrace{(\mathbf{q} \otimes \mathbf{q})^{\top} \otimes \mathbf{I}_r}_{\partial \mathbf{f}/\partial\,\mathrm{vec}(\mathbf{H})}\ ,\ \underbrace{\mathbf{s}(t)^{\top} \otimes \mathbf{I}_r}_{\partial \mathbf{f}/\partial\,\mathrm{vec}(\mathbf{B})}\ \big] \in \mathbb{R}^{r \times d},$$

其中包含 $d = r + r^2 + r^3 + rm$。

附录 D. 数值离散化细节

附录 *D.1.* 黏性伯格斯方程的离散化

我们在空间和时间上采用均匀网格进行离散。令 $x_j = j\,\Delta x$ 对 $j = 0, \ldots, N+1$，其中 $N = 998$ 且 $\Delta x = 1/(N+1)$。记 $^{m} = (u_0^m, \ldots, u_{N+1}^m)^{\top} \approx u(\cdot, t^m)$ 为时刻 $t^m = m\,\Delta t$, $m = 0, \ldots, M$ 处的 $\mathbf{u}$，其中 $T = 1$、$M = 9999$ 和 $\Delta t = T/M \approx 10^{-4}$。施加齐次狄利克雷值 $u_0^m = u_{N+1}^m = 0$。

二阶半隐式时间推进。 对流通过 Lax–Wendroff 更新显式推进至中间场$\mathbf{w}^m$ （其中 $w_0^m = w_{N+1}^m = 0$）。 对于 $j = 1, \dots, N$,

$$w_j^m = u_j^m - \frac{\Delta t}{2\,\Delta x}\, u_j^m\,(u_{j+1}^m - u_{j-1}^m) + \frac{\Delta t^2}{2\,\Delta x^2}\, u_j^m \Big[\tfrac{1}{2}(u_{j+1}^m - u_{j-1}^m)^2 + u_j^m\,(u_{j+1}^m - 2u_j^m + u_{j-1}^m)\Big].$$

扩散采用梯形（Crank–Nicolson方法）规则进行隐式处理，使用离散拉普拉斯算子 $\mathbf{T}_{\Delta x}$ =Δx^{-2} 三对角矩阵$\{1, -2, 1\}$ （修改首行和末行以施加 Dirichlet 边界）。半隐式更新式为

$$\Big(\mathbf{I} - \tfrac{\nu\Delta t}{2}\,\mathbf{T}_{\Delta x}\Big)\mathbf{u}^{m+1} = \mathbf{w}^m + \tfrac{\nu\Delta t}{2}\,\mathbf{T}_{\Delta x}\,\mathbf{u}^m,$$

对此，我们求解 每一步都使用三对角／直接求解器求解内部节点，然后恢复边界值 [42] .

快照。将各时间状态堆叠起来即可得到快照矩阵 $\mathbf{U} = [\mathbf{u}^0 \ \cdots \mathbf{u}^M] \in \mathbb{R}^{(N+2)\times(M+1)}$，我们用它来构建 POD 基并训练降阶模型。

附录 *D.2. Fisher-KPP*方程的离散化

设空间域为 $\Omega = [0, L_x] \times [0, L_y]$，其中 $L_x = L_y = 10$。我们采用均匀网格 $x_i = i\,\Delta x$,$i = 0, \dots, N_x - 1$ 和 $y_j = j\,\Delta y$, $j = 0, \dots, N_y - 1$, 其中 $N_x = N_y = 125$, $\Delta x = L_x/(N_x - 1)$,$\Delta y = L_y/(N_y - 1)$。令 $\mathbf{u}^m \in \mathbb{R}^{N_x N_y}$在 $t^m = m\Delta t$处堆叠为 $u(x_i, y_j, t^m)$, 其中 $T = 5$, $\Delta t = 0.005$, $m = 0, \dots, N_t$,$N_t = 1000$。

带 *Neumann* 边界条件的离散拉普拉斯算子。构造一维二阶差分矩阵 $\mathbf{L}_x^{1\mathrm{D}} \in \mathbb{R}^{N_x \times N_x}$ 和 $\mathbf{L}_y^{1\mathrm{D}} \in \mathbb{R}^{N_y \times N_y}$，其内部模板为 $\frac{1}{\Delta x^2}\{1, -2, 1\}$ 和 $\frac{1}{\Delta y^2}\{1, ~~~

采用显式欧拉法进行时间推进。一个时间步为

$$u_{i,j}^{m+1} = u_{i,j}^{m} + \Delta t\Big[- \, c_x (D_x u)_{i,j}^{m} - \, c_y (D_y u)_{i,j}^{m} + \nu \, (\Delta_h u)_{i,j}^{m} \Big],$$

适用于所有 i, j 并采用周期索引回绕。

稳定性的 *CFL* 条件。 使显式格式保持稳定的一个充分条件是

$$\frac{|c_x|\Delta t}{\Delta x} + \frac{|c_y|\Delta t}{\Delta y} \le 1, \qquad \Delta t \, \nu \left(\frac{1}{\Delta x^2} + \frac{1}{\Delta y^2} \right) \le \frac{1}{2},$$

上述参数设置满足该条件。

快照。 $\mathbf{U} = [\mathbf{u}^0 \ \cdots \mathbf{u}^{N_t}] \in \mathbb{R}^{(N_x N_y)\times(N_t+1)}$ 用于 POD 和降阶模型训练。

参考文献

[1] 美国国家工程院和美国国家科学院、工程院和医学院，《数字孪生的基础研究空白与未来方向》，美国国家科学院出版社，华盛顿特区（2024年）。doi:https://doi.org/10.17226/26894.[2] D. Gunning、D. W. Aha，美国国防高级研究计划局的可解释人工智能项目，《AI Magazine》40（2）（2019）44–58。doi:https://doi.org/10.1609/aimag.v40i2.2850.[3] P. Benner、S. Gugercin、K. Willcox，面向参数化动力系统的基于投影的模型降阶方法综述，SIAM评论 57（4）（2015）483–531。doi:https://doi.org/10.1137/ 130932715.[4]O. Ghattas、K. E. Willcox，从数据中学习基于物理的模型：反问题与模型降阶的视角，《Acta Numerica》30（2021）445–554。doi:https://doi.org/10.1017/S0962492921000064.[5] G. Berkooz、P. Holmes、J. L. Lumley，本征正交分解在湍流分析中的应用，Annual Review of Fluid Mechanics 25（1）（1993）539–575。doi:https://doi.org/10.1146/annurev.fl.25.010193.002543.[6] A. Quarteroni、A. Manzoni、F. Negri，《偏微分方程的降阶基方法：导论》，施普林格，查姆，瑞士，2016年。[7] S. L. Brunton、J. L. Proctor、J. H. Tu、J. N. Kutz，压缩感知与动态模态分解，《Journal of Computational Dynamics》2（2）（2015）165–191。doi:https://doi.org/10.3934/jcd. 2015002.[8] P. J. Schmid，动态模态分解及其变体，Annual Review of Fluid Mechanics 54（第54卷，2022年）（2022）225–254。doi:https://doi.org/10.1146/annurev-fluid-030121-015835.[9] F. Andreuzzi、N. Demo、G. Rozza，一种用于参数化动力系统预测的动态模态分解扩展，《SIAM 应用动力系统杂志》22（3）（2023）2432–2458。doi:https://doi.org/10.1137/22M1481658.[10] T. Sayadi、P. J. Schmid、F. Richecoeur、D. Durox，用于分岔分析的参数化数据驱动分解及其在热声不稳定系统中的应用，《Physics of Fluids》27（3）（2015）037102。doi:https://doi.org/10.1063/1.4913868.[11] B. Kramer、B. Peherstorfer、K. E. Willcox，利用算子推断从数据中学习非线性降阶模型，Annual Review of Fluid Mechanics 56（1）（2024）521–548。doi:https://doi.org/10.1146/ annurev-fluid-121021-025220.

[12] B. Peherstorfer、K. Willcox，用于非侵入式投影模型降阶的数据驱动算子推断，《计算力学与工程方法》306（2016年）196–215。doi:https: //doi.org/10.1016/j.cma.2016.03.025.[13] S. A. McQuarrie、A. Chaudhuri、K. E. Willcox、M. Guo，利用高斯过程对时变非线性系统的低维表示进行贝叶斯学习，《物理学D：非线性现象》475（2025年）134572。[14] A. C. Ionita、A. C. Antoulas，基于 Loewner 框架的数据驱动参数化模型降阶，《工业与应用数学学会科学计算杂志》36（3）（2014年）A984–A1007。doi:https://doi.org/10.1137/ 130914619.[15] T. Franz、R. Zimmermann、S. Görtz、N. Karcher，利用流形学习对定常跨声速流进行基于插值的降阶建模，《国际计算流体动力学杂志》28（3-4）（2014年）106–121。doi:https://doi.org/10.1080/10618562.2014.918695.[16] B. G. Zastrow、A. Chaudhuri、K. E. Willcox、A. Ashley、M. C. Henson，用于耦合多物理场模型降阶的块结构算子推断，《美国航空航天学会杂志》0（0）（0）1–16。doi:https://doi. org/10.2514/1.J065798.[17] M. Guo、J. S. Hesthaven，利用高斯过程回归进行非线性结构分析的降阶建模，《计算力学与工程方法》341（2018年）807–826。doi:https://doi.org/10.1016/j.cma.2018.07.017.[18] L. Cicci、S. Fresca、M. Guo、A. Manzoni、P. Zunino，利用结合高斯过程回归的降阶模型对非线性固体力学进行不确定性量化，《Computers & Mathematics with Applications》149（2023年）1–23。doi:https://doi.org/10.1016/j.camwa.2023.08.016.[19] S. Fresca、A. Manzoni，POD-DL-ROM：通过本征正交分解增强用于非线性参数化偏微分方程的基于深度学习的降阶模型，《计算力学与工程方法》388（2022年）114181。[20] N. R. Franco、A. Manzoni、P. Zunino，一种针对参数相关偏微分方程的深度学习降阶建模方法，《计算数学》92（2023年）483–524。doi:https: //doi.org/10.1090/mcom/3781.[21] F. Sapienza、J. Bolibar、F. Schäfer、B. Groenke、A. Pal、V. Boussange、P. Heimbach、G. Hooker、F. Pérez、P.-O. Persson等，可微编程在微分方程中的应用：综述，arXiv预印本 arXiv:2406.09699（2024年）。[22] R. T. Q. Chen、Y. Rubanova、J. Bettencourt、D. Duvenaud，神经常微分方程，《神经信息处理系统进展》31（2018年）。doi:https://doi.org/10.48550/arXiv.1806.07366.[23]H. Antil、D. Leykekhman，偏微分方程约束优化简介，载于：《偏微分方程约束优化前沿》，施普林格，2018年，第3–40页。doi:https://doi.org/10.1007/ 978-1-4939-8636-1_1.[24] A. M. Bradley，偏微分方程约束优化与伴随方法，技术报告，斯坦福大学（2024年）。[25] C. F. Loan，无处不在的克罗内克积，《计算与应用数学杂志》123（1）（2000年）85–100，数值分析2000年。第三卷：线性代数。doi:https://doi.org/10. 1016/S0377-0427(00)00393-9。URL https://www.sciencedirect.com/science/article/pii/S0377042700003939

[26] C. Eckart, G. Young，以较低秩矩阵逼近另一个矩阵，Psychometrika 1 (3)（1936年）211–218。doi:https://doi.org/10.1007/BF02288367。[27] R. A. Willoughby，《不适定问题的解》（A. N. Tikhonov 和 V. Y. Arsenin），SIAM评论 21 (2)（1979年）266–267。doi:https://doi.org/10.1137/1021044。[28] A. E. Hoerl, R. W. Kennard，岭回归：非正交问题中的有偏估计，Technometrics 42 (1)（2000年）80–86。[29] P. C. Hansen，将截断SVD作为正则化方法，BIT数值数学 27 (4)（1987年）534–553。doi:https://doi.org/10.1007/BF01937276。[30] N. Halko, P. G. Martinsson, J. A. Tropp，利用随机性发现结构：用于构造近似矩阵分解的概率算法，SIAM评论 53 (2)（2011年）217–288。doi:https: //doi.org/10.1137/090771806。[31] P. Luchini, A. Bottaro，《伴随问题导论》，arXiv预印本 arXiv:2404.17304（2024年）。doi:https://doi.org/10.48550/arXiv.2404.17304。[32]C. W. Rowley, S. T. Dawson，流动分析与控制中的模型降阶，Annual Review of Fluid Mechanics 49（第49卷，2017年）（2017年）387–417。doi:https://doi.org/10.1146/ annurev-fluid-010816-060042。[33] M. E. Tipping, C. M. Bishop，概率主成分分析，英国皇家统计学会会刊B辑：统计方法学 61 (3)（2002年）611–622。doi:https://doi.org/10.1111/ 1467-9868.00196。[34] T. Strutz，《数据拟合与不确定性：加权最小二乘及其扩展的实用入门》，Vieweg and Teubner，威斯巴登，DEU，2010年。[35] A. Savitzky, M. J. E. Golay，利用简化最小二乘程序进行数据平滑与求导，Analytical Chemistry 36 (8)（1964年）1627–1639。doi:https://doi.org/10.1021/ac60214a047。[36] F. Lejarza, M. Baldea，通过移动时域优化对动力系统控制方程进行数据驱动发现，Scientific Reports 12 (1)（2022年）1–15。[37] L. Bottou, F. E. Curtis, J. Nocedal，大规模机器学习的优化方法，SIAM评论 60 (2)（2018年）223–311。doi:https://doi.org/10.1137/16M1080173。[38] J. Nocedal, S. J. Wright，《数值优化》，施普林格，纽约，NY，1999年。[39] S. A. McQuarrie, Willcox研究组，opinf：Python中的算子推断，https://pypi.org/ project/opinf/，Python软件包版本0.5.16（2025年）。[40] B. Houska, F. Logist, M. Diehl, J. Van Impe，非线性动态系统状态与参数估计数值方法教程，Springer London，伦敦，2012年，第5章，第67–88页。doi: https://doi.org/10.1007/978-1-4471-2221-0_5。[41] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Har- ris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt，SciPy 1.0贡献者，SciPy 1.0：科学计算的基础算法，Nature Methods 17（2020年）261–272。doi:https://doi.org/10.1038/s41592-019-0686-2。[42] R. J. LeVeque，《守恒律的数值方法》，第2版，《数学讲义》，Birkhäuser

’ 1992
年。