ActaNumerica（2026），第1–100页 英国印刷
doi:10.1017/S09624929XXXXXXXX

arXiv:2602.01397v1 [math.NA] 1 Feb 2026

输运主导问题的非线性模型降阶

Jan S. Hesthaven卡尔
斯鲁厄理工学院
Benjamin Peherstorfer纽约大学库朗数学科学研究所
Benjamin Unger卡尔
斯鲁厄理工学院

本文综述了非线性模型降阶方法，这些方法在某些情况下仍然有效，而线性约化空间逼近在这些情况下本质上效率低下，例如具有类波现象和移动相干结构的输运主导问题；这类问题通常与Kolmogorov屏障相关。本文围绕三个关键要素对非线性模型降阶技术进行梳理：非线性参数化、降阶动力学和在线求解器，并将现有方法归类为基于变换的方法、在线自适应技术，以及结合一般非线性参数化与瞬时残差最小化的方法。

目录

1 引言 1 2 线性模型降阶 5 3 Kolmogorov屏障 17 4 非线性模型降阶的要素 22 5 基于变换的方法 31 6 在线自适应模型降阶 40 7 瞬时残差最小化方法 53 8 结论与展望 79 参考文献 80

1. 引言

我们简要、非技术性地概述模型降阶，并讨论对输运主导问题进行降阶所面临的挑战。

1.1. 模型降阶

在计算科学与工程的许多应用中，数值模型仿真通常嵌套在外循环中，因此需要针对许多不同的参数配置、输入和初始条件进行多次数值仿真。

© 作者，2026。由剑桥大学出版社出版。

例如设计、控制、优化、不确定性量化和反问题。反复进行数值仿真的高昂成本，是解决外循环及其他多查询应用的一大障碍。[1] 模型降阶旨在构建降阶模型，以大幅降低成本进行数值仿真，因此已成为使外循环应用变得可行的关键技术；相关综述参见 Antoulas （2005）、Rozza、Huynh 和 Patera （2008年）、Benner、Gugercin 和 Willcox （2015年）、Antoulas、Beattie 和 Gugercin （2020）、Peherstorfer（2022年）、Hesthaven、Pagliantini 和 Rozza（2022年*b*）、Kramer、Peherstorfer 和 Willcox （2024）。

模型降阶利用了外循环应用的两个基本特性（Rozza 等人 2008年）。首先，在外循环应用所涉及的各种输入、参数和初始条件范围内，数值仿真的解通常会以有规律的方式变化，而非在一般解空间中任意变化。本综述考察的模型降阶方法旨在识别这种有规律的变化，并据此构建降阶模型。与原始高保真数值模型（下文简称完整模型）的解及其对应的完整解相比，这些降阶模型的降阶解仅依赖于少量自由度。其次，外循环应用的性能主要取决于所有数值仿真的总计算成本。由于关注的是总成本，模型降阶可以在离线（或训练）阶段一次性承担较高的预处理成本来构建降阶模型，前提是后续在线（或测试）阶段的降阶模型仿真成本足够低，从而能在整个外循环中摊销预处理成本。利用这两个优势，模型降阶得以在提升外循环应用的可行性方面发挥关键作用。

注 1.1。 模型降阶是一个总称，涵盖了本综述重点介绍的方法之外的许多方法，包括降阶物理模型和纯数据驱动代理模型。本综述聚焦于所谓的基于投影的降阶模型，这类模型由完整模型信息构建，并用随降阶动力学演化的低维状态来表示降阶解；尽管对于非线性参数化，所得构造未必是经典向量空间意义上的投影。

1.2. 线性模型降阶及其局限

（基于投影的）模型降阶的经典方法，是在低维降阶空间中近似全模型解；该空间的维数远小于完整模型所针对的一般完整解空间的维数。一个降阶

[1] 参见 Peherstorfer、Willcox 和 Gunzburger （2018年），了解多保真方法背景下外层循环应用与多查询应用之间区别的讨论。本综述中，我们交替使用这两个术语。

空间是依赖于问题的低维近似空间，通常是完整解希尔伯特空间的子空间。它旨在近似外循环应用中，针对所关注的输入、参数和初始条件所得到的所有全模型解的集合；这一集合通常称为解流形。降阶解是在降阶空间中对完整解进行的线性近似，因此下文将这些方法称为线性模型降阶。需要强调的是，线性模型降阶使用单一的全局降阶空间。线性模型降阶方法包括平衡截断（Mullis 和 Roberts 1976，Moore 1981）、本征正交分解（Sirovich1987，Holmes、Lumley 和 Berkooz 1996）、插值方法（Antoulas 2005，Baur、Beattie、Benner 和 Gugercin 2011）以及降阶基方法（Maday 和 Rønquist 2002，Rozza等2008年）。

有些线性模型降阶方法适用于具有非线性动力学的完整模型，也就是说，数值模型的控制算子——因而时间演化——以非线性方式依赖于解函数。线性模型降阶中的“线性”一词仅指降阶解的试探形式：降阶解是在低维空间中寻求的，因此可以表示为基函数的线性组合。因此，尽管近似试探形式是线性的，所得降阶模型仍可能呈现非线性动力学——这与非线性偏微分方程（PDE）的有限元离散类似，在这种情况下，解场也表示为基函数的线性组合。

线性模型降阶取得了巨大进展，已发展成为一种成熟可靠的技术，涵盖了构造表达能力强的降阶空间的有效方法、后验误差估计、处理解场非线性依赖问题以及结构保持。我们参考综述（Antoulas、Sorensen 和 Gugercin 2001年，Rozza等2008年，Benner等 2015年， Hesthaven等2022年*b*， Kramer等2024）和专著（Antoulas 2005，Quarteroni 和 Rozza 2014， Hesthaven、Rozza 和 Stamm 2016，Quarteroni、Manzoni 和 Negri 2016， Benner、Cohen、Ohlberger 和 Willcox 2017，Antoulas 等 2020， Benner、Grivet-Talocia、Quarteroni、Rozza、Schilders 和 Silveira 2021年*a*，*b*），其中综述了线性模型降阶领域的进展。

然而，线性模型降阶存在一个重大且根本的局限：完整模型的解不仅必须近似位于完整模型解空间中的光滑低维流形上，还必须能够由具有线性向量空间结构的低维空间良好近似。要求降阶解是向量空间中的线性近似，这施加了严格的结构限制。即使解流形光滑且维数较低，它也可能具有显著的曲率，或以其他方式偏离线性，因此无法用低维向量空间良好地捕捉。对于一类与表现良好的扩散（椭圆/抛物）问题相关的问题，已有研究表明，解流形可以用降阶空间良好近似；也就是说，存在一系列维数不断增加的降阶空间 n，其近似误差

呈指数衰减 $\mathcal{O}(e^{-n})$，且衰减率取决于降阶维数 n，在适当的度量下（Maday、Patera 和 Turinici，2002年，Buffa、Maday、Patera、Prud'homme 和 Turinici，2012年）。对于这类问题，线性模型降阶非常适用——长期以来，这类问题一直是模型降阶研究的主要关注对象。相比之下，输运主导问题——其中波前或相变等相干结构在空间域中传播——可能会产生粗糙的解流形；所谓粗糙，是指在降阶空间中对这些解进行线性近似时，误差随降阶维数增加而缓慢衰减（Rowley 和 Marsden，2000年、Ohlberger 和 Rave，2013年、Peherstorfer，2022年）。线性模型降阶对于这类输运主导问题并不奏效。例如，Ohlberger 和 Rave (2016年)、Greif 和 Urban (2019)、Arbes、Greif 和 Urban(2025）推导出了误差下界，其变化规律为 $\mathcal{O}(1/\sqrt{n})$；当用线性近似在维数为 n 的降阶空间中近似线性平流方程和波动方程对应的解流形时，该下界成立。与某些扩散主导问题所达到的指数衰减相比，这种误差衰减要慢得多。这一局限通常称为 Kolmogorov屏障。

除这些逼近理论误差界之外，线性模型降阶对于输运和波动类问题的局限性，已在实践中一系列受关注的问题上得到经验性指出，包括强平流与化学反应流（Huang、Xu、Duraisamy 和 Merkle，2018年，Huang 和 Duraisamy，2023年），野火动力学（Black、Schulze 和 Unger，2021年*a*），图案形成（Geelen 和 Willcox，2022年），太阳天气预测（Issan 和 Kramer，2023年），脉冲爆震燃烧室（Schulze、Reiss 和 Mehrmann，2019年），动力学等离子体问题（Nicolini 和 Teixeira，2021年），以及旋转爆震波（Mendible、Brunton、Aravkin、Lowrie 和 Kutz，2020年，Uy、Wentland、Huang 和 Peherstorfer，2024年）。

1.3. 非线性模型降阶

非线性模型降阶旨在通过寻求全模型解的非线性逼近，克服线性模型降阶表达能力有限的问题。因此，降阶解可以非线性地依赖其自由度，从而比降阶空间中的线性近似更高效地表示全模型解。换言之，非线性逼近所诱导的试探流形能够以少于线性约化空间逼近的自由度，准确表示解流形。从近似理论的角度来看，在逼近中引入非线性，可以克服线性约化空间逼近在表达能力上的根本局限，从而缓解Kolmogorov屏障。然而，非线性逼近只是非线性模型降阶的一个组成部分。与线性模型降阶不同，非线性逼近给出的降阶解不再位于线性向量空间中。具体而言，通常不存在映射到试探流形上的全局线性投影算子，也不存在与之相关的正交性

原理，因此无法将标准伽辽金构造和许多标准稳定性论证原封不动地从完整模型中沿用。因而，非线性模型降阶需要重新审视降阶模型的构造、分析及数值求解方法，以处理逼近的非线性；这正是本文综述的主题。

注 1.2. 非线性逼近在数值分析中有着悠久的历史，远早于近年来非线性模型降阶的发展。例如，DeVore（1998年）研究了非线性逼近相对于线性空间中逼近的近似理论方面，以及构造非线性逼近的方法。神经网络从近似理论角度的研究见于Pinkus （1999年）、DeVore、Hanin 和 Petrova（2021年）、Telgarsky （2016年）。在针对偏微分方程的有限元及相关数值方法中，自适应网格细化会产生以依赖于解的方式选取的逼近空间，因此所得近似是底层自由度的非线性函数（Babuv̌ska 和 Rheinboldt 1978年，Berger 和 Colella 1989年，Berger 和 Oliger 1984年）。同样，有理函数逼近就其自由度而言在本质上是非线性的，并且已被证明，对于具有尖锐特征或近奇异行为的函数，它比多项式逼近和固定线性空间逼近更为有效；参见例如Braess （1986年）、Trefethen （2019）。虽然非线性模型降阶可以借鉴上述诸多进展，但其一个关键的独特之处在于：降阶模型并非用于更好地近似单个解函数，而是用于构建一个非线性降阶模型，该模型可针对一系列参数、输入和初始条件反复使用，以准确近似全模型解。

1.4. 论文结构

在本引言之后，第 2 节将简要讨论线性模型降阶方法，为后文作铺垫。第 3 节讨论线性模型降阶在处理输运类和波动类问题时的局限性。第 4 节综述非线性模型降阶，并介绍非线性参数化、降阶动力学和在线求解器这三个非线性模型降阶的组成要素。随后，我们将讨论三类非线性模型降阶方法。第一类是基于非线性变换的方法，见第 5 节。第二类是在在线阶段调整降阶空间的方法，见第6节。第三类是在在线阶段基于通用非线性参数化并采用瞬时残差最小化的方法，见第 7节。第 8节给出结论并展望未来。

2. 线性模型降阶

本节简要回顾线性模型降阶技术，该技术旨在对全模型解进行线性近似。线性模型降阶通常分为离线阶段和在线

阶段：在离线阶段构建降阶模型，在在线阶段使用降阶模型进行仿真。我们讨论线性模型降阶离线阶段的三个步骤：(i) 收集训练数据，(ii) 训练降阶空间，以及 (iii) 构建降阶模型。接着，我们讨论阻碍在线阶段高效运行的提升瓶颈，并介绍经验插值，说明如何绕过提升瓶颈，从而实现高效的在线降阶模型求解。最后，我们通过一项数值实验展示线性模型降阶在输运主导问题中的有限效果。

2.1. 完整模型

我们考虑如下形式的时变和参数相关偏微分方程

$$\partial_t q_{\mathcal{Q}}(t,\boldsymbol{x};\boldsymbol{\mu}) = f(\boldsymbol{x}, q_{\mathcal{Q}}(t,\cdot;\boldsymbol{\mu});\boldsymbol{\mu}) \tag{2.1}$$

其中 $t \in \mathcal{T} = [0,T]$ 表示时间变量，$\boldsymbol{x} \in \Omega \subseteq \mathbb{R}^d$ 表示空间变量，$\boldsymbol{\mu} \in \mathcal{D} \subseteq \mathbb{R}^{d'}$ 表示一个参数。在本文中，我们通常考虑空间域的维数 d 较低的情况，$d \in \in \{1,2,3\}$，且参数的维数 d' 至多为中等。问题 (2.1) 配备了适当的初始条件 $q_{\mathcal{Q}}(0,\cdot;\boldsymbol{\mu})$：$\Omega \to \mathbb{R}$ 对所有 $\boldsymbol{\mu} \in \mathcal{D}$ 均成立。解函数 $q_{\mathcal{Q}}(t,\cdot;\boldsymbol{\mu})$：$\Omega \to \mathbb{R}$ 属于一个适当的希尔伯特空间 $\mathcal{Q}$，该空间施加边界条件，从而使偏微分方程问题适定。请注意，问题 (2.1) 还可以有多种扩展：问题也可以依赖于控制输入，右端函数 f 可以依赖于时间 t，解也可以是向量值的，例如偏微分方程组。尽管已有适用于这些情形的模型降阶方法，为了便于阐述，我们将讨论范围限定为公式 (2.1) 中的形式。

设解函数 $q_{\mathcal{Q}}(t,\cdot;\boldsymbol{\mu})$：$\Omega \to \mathbb{R}$ 在所有时间 $t \in \mathcal{T}$ 和参数 $\boldsymbol{\mu} \in \mathcal{D}$下于有限维子空间 $\mathcal{Q}_N \subseteq \mathcal{Q}$ 中被近似。空间 $\mathcal{Q}_N$ 的维数为 $N \in \mathbb{N}$，并由基函数$\phi_1, \ldots, \phi_N$：$\Omega \to \mathbb{R}$张成。例如，有限维空间 $\mathcal{Q}_N$ 可通过有限元方法获得（Ern 和 Guermond，2004）。$q_{\mathcal{Q}}(t,\cdot;\boldsymbol{\mu})$ 在 $\mathcal{Q}_N$ 中的近似是一个线性组合

$$q(t,\boldsymbol{x};\boldsymbol{\mu}) = \sum_{i=1}^{N} q_i(t,\boldsymbol{\mu})\phi_i(\boldsymbol{x}), \tag{2.2}$$

具有 N 系数 $q_1(t,\boldsymbol{\mu})$，……，$q_N(t,\boldsymbol{\mu}) \in \mathbb{R}$，这些系数依赖于时间 t 和参数 $\boldsymbol{\mu}$。试探形式（2.2）随后可代入偏微分方程问题（2.1）的适当弱（变分）形式，并可施加（Petrov–）Galerkin条件，使得残差在检验空间上检验时为零。需要指出的是，在模型降阶中，通常假定解（2.2）在 $\mathcal{Q}_N$ 中相对于相应解 $q_{\mathcal{Q}}$ 在空间 $\mathcal{Q}$ 中的误差可以用标准数值分析工具加以控制，并且就模型降阶的目的而言可以忽略。因此，每当我们通过模型降阶构造降阶模型时，它

就足以推导出一个能够准确逼近解q （见（2.2）、位于 $\mathcal{Q}_N$）的降阶模型。相应地，我们将解 q 在 $\mathcal{Q}_N$ 中称为全模型解。

收集全模型解（2.2）的系数，得到

$$\mathbf{q}(t,\boldsymbol{\mu}) = \left[q_1(t,\boldsymbol{\mu}),\ldots,q_N(t,\boldsymbol{\mu})\right]^{\mathsf{T}} \in \mathbb{R}^N, \tag{2.3}$$

一个向量，它描述空间 $\mathcal{Q}_N$ 中的函数 $q(t,\cdot;\boldsymbol{\mu})$。这种基于坐标的表示 $\boldsymbol{q}(t,\boldsymbol{\mu})$ 对全模型解 $q(t,\cdot;\boldsymbol{\mu})$ 的描述，导出了半离散全模型，它是一个常微分方程（ODE）系统，

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{q}(t,\boldsymbol{\mu}) &= \boldsymbol{f}(\mathbf{q}(t,\boldsymbol{\mu});\boldsymbol{\mu}),\\ \mathbf{q}(0;\boldsymbol{\mu}) &= \mathbf{q}_0(\boldsymbol{\mu}),\end{aligned} \tag{2.4}$$

其初始条件为 $\mathbf{q}_0(\boldsymbol{\mu}) \in \mathbb{R}^N$。相应地，向量 $\boldsymbol{q}(t,\boldsymbol{\mu})$ 通常称为半离散全模型的状态。空间离散化可能会产生质量矩阵，我们在形式上将其纳入向量场 $\boldsymbol{f}$ 中。在实际实现中，需要对质量矩阵进行适当处理。

注释 2.1. 本综述中，我们考虑参数化完整模型，但不显式构造输入–输出映射。尽管如此，系统与控制理论中的一类重要模型降阶方法，正是以输入和输出的显式定义为核心。在这种情况下，完整模型被视为输入–输出动力系统，降阶则可表述为对其诱导的输入–输出映射进行近似。代表性方法包括平衡截断、矩匹配与插值，以及相关的基于投影的技术。这些方法构造的降阶模型能够近似完整系统的输入–输出响应，并且通常还能提供稳定性和误差保证，至少对于线性系统动力学而言是如此。关于系统与控制理论模型降阶的综述，请参见 Antoulas(2005）、Benner等(2015年）、Antoulas 等 (2020）。

2.2. 线性模型降阶概述

线性模型降阶旨在低维子空间 $\mathcal{V} \subseteq \mathcal{Q}_N$ 中构造维数为 $n \ll N$的全模型解近似。

降阶空间 $\mathcal{V}$ 必须具有足够的表达能力，才能在时间区间 $\mathcal{T}$和参数域 $\mathcal{D}$ 上良好地近似全模型解。形式上，全模型解诱导出一个所谓的解流形，

$$\mathcal{M} = \{q(t,\cdot;\boldsymbol{\mu}) \,|\, t \in \mathcal{T}, \boldsymbol{\mu} \in \mathcal{D}\}, \tag{2.5}$$

它是与所关注的时间和参数范围相对应的所有全模型解的集合。我们沿用惯例，将 $\mathcal{M}$ 称为解流形，尽管集合 $\mathcal{M}$ 不一定具有流形结构；例如参见 Haasdonk (2017年，例 2.9）。解流形可以定义为

作为空间 $\mathcal{Q}$ 中的解集，而不是 N维全模型空间 $\mathcal{Q}_N$中的解集。

降阶解记为一个函数 $\hat{q}\colon \mathcal{T} \times \Omega \times \mathcal{D} \to \mathbb{R}$。对于给定的$t$ 和 $\boldsymbol{\mu}$，函数 $\hat{q}(t, \cdot; \boldsymbol{\mu})\colon \Omega \to \mathbb{R}$ 是 $\mathcal{V}$的一个元素，并近似全模型解 $q(t, \cdot; \boldsymbol{\mu})$。由于 $\hat{q}(t, \cdot; \boldsymbol{\mu})$ 位于降阶空间 $\mathcal{V}$中，它可以表示为线性组合

$$\hat{q}(t, \boldsymbol{x}; \boldsymbol{\mu}) = \sum_{i=1}^{n} \hat{q}_i(t, \boldsymbol{\mu}) \varphi_i(\boldsymbol{x}) \tag{2.6}$$

降阶空间 $\mathcal{V}$的基函数 $\varphi_1, \ldots, \varphi_n\colon \Omega \to \mathbb{R}$。线性组合 (2.6) 的系数 $\hat{q}_1(t, \boldsymbol{\mu}), \ldots, \hat{q}_n(t, \boldsymbol{\mu}) \in \mathbb{R}$ 依赖于时间 t 和参数 $\boldsymbol{\mu}$。相比之下，张成空间 $\mathcal{V}$ 的基函数在所有时间 t 和参数 $\boldsymbol{\mu}$上都是固定的。因此，在所有时间 t 和参数 $\boldsymbol{\mu}$下，降阶解 (2.6) 是 $\mathcal{V}$ 中的线性近似，其含义是：当将 $\hat{q}$ 解释为这些系数的函数时，系数（自由度）$\hat{q}_1(t, \boldsymbol{\mu}), \ldots \hat{q}_n(t, \boldsymbol{\mu})$ 线性地进入 $\hat{q}$。降阶解 (2.6) 的系数可通过偏微分方程问题 (2.1) 的变分形式以及对 $\mathcal{V}$的 Galerkin 投影或 Petrov–Galerkin 投影得到，这类似于 $\mathcal{Q}_N$中的全模型解 (2.2) 由该偏微分方程问题得到的方式。特别地，在所有时间 t 和参数 $\boldsymbol{\mu}$ 下，降阶解 $\hat{q}(t, \cdot; \boldsymbol{\mu})$ 在 $\mathcal{V}$中计算，从而它仅依赖于 n 个自由度，即该表示中的 n 个系数。这与依赖于 N 个系数（通常 $N \gg n$）的全模型解 (2.2) 形成对比。

2.3. 线性模型降阶离线阶段的三个步骤

我们确定了线性模型降阶离线阶段的三个主要步骤：步骤 1 是收集有关完整模型的信息，用于构造合适的降阶空间 $\mathcal{V}$，以便很好地逼近位于解流形上的完整解 $\mathcal{M}$。在本综述的范围内，这些信息通常以选定时间和参数下的全模型解形式获取，这些解构成训练数据。步骤 2 是利用训练数据训练降阶空间 $\mathcal{V}$，以便很好地逼近 $\mathcal{M}$。第三步是构造一个降阶模型，用于描述降阶空间中降阶解的系数所对应的方程组（2.6） $\mathcal{V}$。完成离线阶段的这三个步骤并构造出降阶模型后，便可以在在线阶段将其用于在外循环应用中计算降阶解。接下来，我们将更详细地讨论离线阶段的三个步骤，然后在第 2.4 节讨论在线阶段。

2.3.1. 离线步骤 *1*：收集训练数据

收集训练数据的一种标准方法是针对一组训练参数 $\mathcal{D}_{\text{train}} = \{\boldsymbol{\mu}_1, \ldots\ldots, \boldsymbol{\mu}_M\} \subseteq \mathcal{D}$ 求解完整模型（ 2.4 ），从而得到一个训练数据集

$$\mathcal{Q}_{\text{train}} = \{\boldsymbol{q}(t, \boldsymbol{\mu}_i) \,|\, i = 1, \ldots, M,\, t \in [0, T]\}, \tag{2.7}$$

where 我们采用式（2.3）中给出的全模型解的半离散表示 .

选择合适的训练参数 $\boldsymbol{\mu}_1$, ..., $\boldsymbol{\mu}_M$ 使训练数据集 $\mathcal{Q}_{\text{train}}$ 能够为构建降阶空间提供有用信息，是一项具有挑战性的工作。大量研究采用贪心过程来逐步构建训练集，以解决这一问题：从一个初始参数样本开始，在候选参数集上反复评估降阶模型，由误差指示器确定降阶模型表现最差的参数，再将该参数加入训练集。重复这一过程，直至误差指示器低于用户指定的容差（Prud' homme、Rovas、Veroy、Machiels、Maday、Patera 和 Turinici 2001，Veroy 和 Patera 2005，Haasdonk 和 Ohlberger 2008，Binev、Cohen、Dahmen、DeVore、Petrova 和 Wojtaszczyk 2011，Hesthaven、Stamm 和 Zhang 2014，Cohen、Dahmen、DeVore 和 Nichols 2020）。这类贪心策略往往在很大程度上依赖后验误差估计方面的进展（Grepl 和 Patera 2005，Veroy 和 Patera 2005，Haasdonk 和 Ohlberger 2011，Urban 和 Patera 2012）。贪心方法还可以通过构建降阶空间，使降阶解能够准确近似感兴趣量，从而明确针对感兴趣量；例如参见 Bui-Thanh、Willcox、Ghattas 和 van Bloemen Waanders (2007)。

2.3.2. 离线步骤 *2*：训练降阶空间

来自离线步骤 1 的训练数据 $\mathcal{Q}_{\text{train}}$ 可用于训练维数为 $n \ll N$ 的低维空间 $\mathcal{V}$，这通常通过构造张成 $\mathcal{V} \subseteq \mathcal{Q}_N$ 的基函数 $\varphi_1, \ldots, \varphi_n$ 来实现。因此，$\mathcal{V}$ 的基函数$\varphi_1, \ldots, \varphi_n$ 可以表示为张成全模型解空间 $\mathcal{Q}_N$ 的基函数$\phi_1, \ldots, \phi_N$的线性组合，

$$\varphi_i = \sum_{j=1}^{N} v_i^{(j)} \phi_j, \qquad i = 1, \ldots, n. \tag{2.8}$$

我们将系数收集到向量 $\boldsymbol{v}_i = [v_i^{(1)}, \ldots, v_i^{(N)}]^\top \in \mathbb{R}^N$中，对应于 $i =$ 1 n n $\mathbb{R}^N$,...,。系数向量 $\boldsymbol{v}_1$,. .., $\boldsymbol{v}_n$ 张成一个 - 维子空间。因此，我们可以用张成 $\mathcal{Q}_N$ 的某个子空间的基函数 $\varphi_1, \ldots, \varphi_n$ 来表示该降阶空间，或者等价地，用张成 $\mathbb{R}^N$ 的某个子空间的基向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$ 来表示。

构造一组基的一种方法 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$，并由此隐式地构造基函数$\varphi_1, \ldots, \varphi_n$，其形式为 (2.8)，并张成 $\mathcal{V}$，可通过奇异值分解 (SVD) 得到。为此，考虑 K 离散时间点 $0 = t_0 < t_1 < \cdots < t_K = T$以及训练数据集 $\mathcal{Q}_{\text{train}}$对应的快照矩阵，

$$\boldsymbol{Q}_{\text{train}} = [\boldsymbol{q}(t_0; \boldsymbol{\mu}_1), \ldots, \boldsymbol{q}(t_K; \boldsymbol{\mu}_1), \boldsymbol{q}(t_0; \boldsymbol{\mu}_2), \ldots, \boldsymbol{q}(t_K; \boldsymbol{\mu}_M)] \in \mathbb{R}^{N \times (K+1)M}. \tag{2.9}$$

计算 $\boldsymbol{Q}_{\text{train}}$ 的奇异值分解，并使用对应于 n 个最大奇异值的前 n 个左奇异向量，可得到基 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$，该基张成 $\mathbb{R}^N$ 的一个子空间。在模型降阶中，这种基于奇异值分解计算基矩阵的过程

快照矩阵的过程被称为本征正交分解（POD）；例如，参见 Sirovich（1987）、Holmes 等人（1996）在流体力学背景下的早期应用，以及 Gubisch 和 Volkwein（2017年）从数值分析角度的论述。除 POD 之外，还有许多其他技术可用于构造降阶空间，包括贪心方法（Prud'homme 等人 2001年，Veroy 和 Patera 2005，Bui-Thanh 等人 2007，Binev 等人 2011，Cohen 等人 2020）、基于插值的方法（Antoulas 2005，Baur 等人 2011，Antoulas 等人 2020）以及平衡（Mullis 和 Roberts 1976，Moore 1981）。

离线步骤 2 的结果是一个 n 维降阶空间 $\mathcal{V}$的基，可用基函数 $\varphi_1, \ldots, \varphi_n \in \mathcal{Q}_N$或系数向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n \in \mathbb{R}^N$表示。

2.3.3. 离线步骤 *3*：构造降阶模型

线性模型降阶的第三个离线步骤是构造降阶模型，即推导一个方程组，用以确定降阶解 $\hat{q}_1(t,\boldsymbol{\mu}), \ldots, \hat{q}_n(t,\boldsymbol{\mu})$ 的系数，该降阶解表示为 $\hat{q}$，如式（2.6）所示。

与推导完整模型类似，可以通过变分形式以及 Galerkin 投影或 Petrov–Galerkin 投影得到降阶模型，但此时投影到降阶空间 $\mathcal{V}$，而非全模型空间 $\mathcal{Q}_N$。在半离散情形下，降阶模型成为一个常微分方程组：表示式（2.8）使我们能够将式（2.6）中给出的函数 $\hat{q}(t,\cdot;\,\boldsymbol{\mu}) \in \mathcal{V}$ 与系数向量

$$\hat{\boldsymbol{q}}(t,\boldsymbol{\mu}) = [\hat{q}_1(t,\boldsymbol{\mu}), \ldots, \hat{q}_n(t,\boldsymbol{\mu})]^\top \in \mathbb{R}^n\,, \tag{2.10}$$

它属于由基矩阵 $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n] \in \mathbb{R}^{N\times n}$ 的列张成的 $\mathbb{R}^n$ 中的 n 维子空间。我们注意到，在许多情况下，构造降阶基向量 $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n$（相应地，降阶基函数 $\varphi_1, \ldots, \varphi_n$）时会使其正交归一。特别是在这些情况下，基矩阵 $\boldsymbol{V}$ 的列是正交归一的。若无特别说明，以下均假设 $\boldsymbol{V}$ 的列是正交归一的。

基于基矩阵 $\boldsymbol{V}$ 和半离散完整模型（2.4），半离散降阶模型为

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{\boldsymbol{q}}(t,\boldsymbol{\mu}) = \hat{\boldsymbol{f}}(\hat{\boldsymbol{q}}(t,\boldsymbol{\mu});\boldsymbol{\mu})\,, \tag{2.11}$$

$\hat{}$ 其中，右端函数 $\boldsymbol{f}$ 通过 Galerkin 投影得到：

$$\hat{\boldsymbol{f}}(\hat{\boldsymbol{q}}(t,\boldsymbol{\mu});\boldsymbol{\mu}) = \boldsymbol{V}^\top \boldsymbol{f}(\boldsymbol{V}\hat{\boldsymbol{q}}(t,\boldsymbol{\mu});\boldsymbol{\mu})\,. \tag{2.12}$$

与半离散完整模型的状态 $\boldsymbol{q}(t,\boldsymbol{\mu})$ 类似，向量 $\hat{\boldsymbol{q}}(t,\boldsymbol{\mu})$ 通常称为降阶状态。初始条件 $\hat{\boldsymbol{q}}(0;\boldsymbol{\mu}) \in \mathbb{R}^n$ 例如可以通过投影 $\hat{\boldsymbol{q}}(0;\boldsymbol{\mu}) = \boldsymbol{V}^\top \boldsymbol{q}(0;\boldsymbol{\mu})$ 得到。

尽管我们在 (2.12) 中通过 Galerkin 投影定义右端函数，但我们注意到，当右端函数 $\boldsymbol{f}$源自非对称

PDE算子时，采用检验空间与试探空间不同的 Petrov–Galerkin 投影对于稳定性至关重要（Quarteroni 和 Rozza 2014年，Hesthaven 等人 2016年，Mehrmann 和 Unger 2023）。基于插值的模型降阶方法通常也会导出彼此不同的检验空间与试探空间，从而通过 Petrov–Galerkin 投影获得降阶模型（Antoulas 2005，Baur 等人 2011，Antoulas 等人 2020）。正如 Otto、Padovan 和 Rowley（2022年）以及 Otto、Macchio 和 Rowley（2023年）的研究所示，在 Petrov–Galerkin框架中使用斜投影也能提升降阶模型的预测能力。提升鲁棒性和稳定性，也是 Carlberg、Bou-Mosleh 和 Farhat（2011）以及 Carlberg、Barone 和 Antil（2017）提出最小二乘 Petrov–Galerkin 方法的动机之一。Dahmen、Plesken 和 Welper（2014年），Brunken、Smetana 和 Urban（2019）以及 Engwer、Ohlberger 和 Renelt（2024）的研究明确针对输运主导问题，开发了在此类问题中仍保持稳定的降阶模型；然而，这些研究侧重于线性模型降阶的鲁棒性和稳定性，而非克服线性近似的 Kolmogorov屏障。另一种方法是将线性模型降阶公式与稳定项相结合（例如，Pacciarini 和 Rozza（2014年）），或者开发在线稳定化技术（例如，Maday、Manzoni 和 Quarteroni（2016年），Balajewicz、Tezaur 和 Dowell（2016年））。

ˆ需要注意的是，$\boldsymbol{f}$ 降阶模型（2.11）的右端函数不一定关于降阶状态 $\hat{\boldsymbol{q}}(t,\ \boldsymbol{\mu})$ 呈线性，即使 $\hat{\boldsymbol{q}}(t,\ \boldsymbol{\mu})$ 是完整模型状态的线性近似 $\boldsymbol{q}(t,\ \boldsymbol{\mu})$，且其位于由 n的列向量张成的维数为 $\boldsymbol{V}$的子空间中。这再次强调，线性模型降阶方法适用于非线性完整模型，而线性模型降阶中的“线性”一词仅指降阶解的线性试探形式。

备注 2.2。 通常会区分侵入式和非侵入式模型降阶。非侵入式方法主要根据快照数据学习降阶表示（例如降阶空间）以及降阶动力学；例如参见 Ghattas 和 Willcox (2021年)、Kramer 等人(2024)。相比之下，本综述侧重于侵入式方法，这类方法通过投影或更一般的残差最小化技术等方式，从控制方程推导降阶动力学。我们将在第8节简要回顾非侵入式模型降阶。

2.4. 在线计算、提升瓶颈与经验插值

现在我们讨论线性模型降阶的在线阶段，即针对给定的新参数和初始条件，利用降阶模型高效地进行数值仿真。具体而言，我们讨论提升瓶颈：直接应用投影来推导降阶右端函数，仍可能需要进行在线计算，而其计算量会随全空间维数 N 不利地增长，从而导致加速效果甚微，甚至毫无加速。随后，我们简要回顾经验插值这一克服提升瓶颈的策略。

2.4.1. 在线效率

在线阶段中，使用降阶模型，通过随时间积分降阶动力系统（2.12），针对新的参数 $\boldsymbol{\mu}$ 和初始条件进行仿真——也称为模型评估。此过程需要 $\hat{}$ 多次评估降阶右端函数 $\boldsymbol{f}$；对于隐式时间积分格式，通常还需要在 $\boldsymbol{f}$ 的求值过程中计算其导数，例如进行牛顿迭代时。因而，计算 $\boldsymbol{f}$ 的成本通常在在线计算成本中占主导地位。

如果在线阶段的计算成本不随完整模型的维度变化，我们就称降阶模型具备在线效率 N（Rozza等2008年）。在一种常见情形下，通常可以实现在线效率：完整模型的右端项关于状态呈线性，并关于参数呈仿射，即

$$\boldsymbol{f}(\boldsymbol{q};\boldsymbol{\mu}) = \sum_{i=1}^{\ell} \rho_i(\boldsymbol{\mu})\boldsymbol{A}_i\boldsymbol{q}, \qquad \rho_i : \mathcal{D} \to \mathbb{R}, \quad \boldsymbol{A}_i \in \mathbb{R}^{N\times N}.$$

在这种情况下，参数无关的降阶算子 $\boldsymbol{V}^\top\boldsymbol{A}_i\boldsymbol{V}$ 可以预先在离线阶段计算，而对于给定参数，组装和计算 $\boldsymbol{f}$ 的成本仅与全模型维数无关 N（Rozza等2008年，Quarteroni 和 Rozza 2014年，Hesthaven 等 2016年）。如果只需在某个时刻或一个较短的时间窗内计算降阶模型，则可使用围道积分方法进一步提高在线效率；例如，参见Guglielmi、L´opez-Fern´andez 和 Manucci（2021年）以及Guglielmi 和 Manucci（2023）。然而，若不具备此类特殊结构，计算降阶右端项 $\boldsymbol{f}$ 通常仍需在高维全模型空间中进行运算，这从根本上限制了在线效率。

2.4.2. 提升瓶颈

要了解为什么通常难以实现在线效率，请回想一下，式 (2.12) 中的降阶右端项 $\boldsymbol{f}$ 是通过对全模型右端项 $\boldsymbol{f}$ 进行（Petrov–）Galerkin 投影得到的。因此，要在时间t 和参数 $\boldsymbol{\mu}$ 下，对降阶状态 $\hat{}\boldsymbol{q}(t,\boldsymbol{\mu}) \in \mathbb{R}^n$ 求值，通常需要按以下步骤进行：首先将降阶状态提升到高维空间，得到 $\boldsymbol{V}\hat{}\boldsymbol{q}(t,\boldsymbol{\mu}) \in \mathbb{R}^N$；然后在这一提升后的状态 $\boldsymbol{V}\hat{}\boldsymbol{q}(t,\boldsymbol{\mu})$处计算全模型右端函数 $\boldsymbol{f}$，这需要计算所有分量函数；最后，将 $\boldsymbol{f}$ 的输出投影到由 $\boldsymbol{V}$的列向量张成的降阶空间中。

在某些特殊情况下，例如上述讨论的仿射线性情形，计算 $\boldsymbol{f}$ 时涉及的项可以在离线阶段预先计算，并在在线阶段重复使用（Rozza 等人 2008年）。然而，在更一般的情况下，提升降阶状态、在 $\boldsymbol{f}$ 中计算 $\mathbb{R}^N$，以及随后进行投影，意味着计算降阶右端项 $\boldsymbol{f}$ 的计算成本会随全模型维数 N增长。在线计算时，需要反复将量提升至

高维空间并在其中进行计算，这通常称为提升瓶颈。

2.4.3. 经验插值

经验插值法由Barrault、Maday、Nguyen 和 Patera (2004) 提出，旨在绕开提升瓶颈。经验插值用插值替代（2.12）中的投影，因此只需计算全模型右端函数中少数几个 N分量函数。为绕开提升瓶颈，人们还提出了其他采用类似插值和回归策略的方法，例如缺失点估计（Astrid、Weiland、Willcox 和 Backx 2004，2008年）、经验算子学习（Haasdonk、Ohlberger 和 Rozza 2008年，Drohmann、Haasdonk 和 Ohlberger 2012年）、高斯–牛顿近似张量（Carlberg 等2011），以及能量守恒采样与加权方法；该方法侧重于保持底层全模型离散化的结构和稳定性（Farhat、Avery、Chapman 和 Cortial 2014，Farhat、Chapman 和 Avery 2015）。

经验插值的步骤 这里我们重点关注 Barrault 等人 (2004)提出的经验插值法，尤其是其离散对应方法——由 Chaturantabut 和 Sorensen(2010)提出的离散经验插值法。在离线阶段，选取两两互异的插值点，

$$p_1, \ldots, p_n \in \{1, \ldots, N\}, \tag{2.13}$$

在离散情形下，这些插值点对应于全模型右端函数 $\boldsymbol{f}$ 的 N 分量函数的索引。由这些插值点可构造插值点矩阵

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{e}_{p_1} & \ldots & \boldsymbol{e}_{p_n} \end{bmatrix}^{\mathsf{T}} \in \mathbb{R}^{N \times n}$$

因此，将向量 $\boldsymbol{z} = [z_1, \ldots, z_N]^{\mathsf{T}} \in \mathbb{R}^N$ 作用于 $\boldsymbol{P}^{\mathsf{T}}$ 时，会选取与索引 $p_1, \ldots, p_n$ 对应的分量：

$$\boldsymbol{P}^{\mathsf{T}} \boldsymbol{z} = \begin{bmatrix} z_{p_1} & \ldots & z_{p_n} \end{bmatrix}^{\mathsf{T}}.$$

在实际实现中，$\boldsymbol{P}$ 不会显式组装插值点矩阵，而是通过切片操作来模拟其对向量的作用（Chaturantabut 和 Sorensen 2010）。随后，通过在插值点 p_1、…、p_n处，在基 $\boldsymbol{v}_1$、…、$\boldsymbol{v}_n$中对全模型右端项 $\boldsymbol{f}$进行插值，即可得到降阶右端函数：

$$\tilde{\boldsymbol{f}}(\hat{\boldsymbol{q}}; \boldsymbol{\mu}) = (\boldsymbol{P}^{\mathsf{T}} \boldsymbol{V})^{-1} \boldsymbol{P}^{\mathsf{T}} \boldsymbol{f}(\boldsymbol{V}\hat{\boldsymbol{q}}; \boldsymbol{\mu}). \tag{2.14}$$

基于经验插值的降阶模型类似于模型˜在（2.11）中给出，不同之处在于使用了降阶右端项 $\boldsymbol{f}$ 在（2.14）中给出，而不是通过投影定义的右端项 $\boldsymbol{f}$ 在（2.12）中定义。我们指出，经验插值也可以应用于右端

项函数 $\boldsymbol{f}$。例如，Barrault 等人（2004）、Chaturantabut 和 Sorensen（2010）仅对 $\boldsymbol{f}$ 的非线性项应用经验插值，同时对线性项进行投影。此外，还可以为非线性项的近似采用单独的降阶空间，即为状态和对 $\boldsymbol{f}$ 的非线性项进行经验插值近似分别使用不同的降阶空间。

˜在线阶段，将通过经验插值（2.14）得到的 $\boldsymbol{f}$ 用作降阶模型中的右端函数，即使使用通过投影（2.12）得到的 $\boldsymbol{f}$ 无法实现运行时加速，也可能实现运行时加速。如果全模型右端函数 $\boldsymbol{f}$ 满足某些结构性质，例如，$\boldsymbol{f}$ 的每个分量函数都能独立于其他所有分量函数进行求值，那么计算（2.14）中的项 $\boldsymbol{P}^\top \boldsymbol{f}(\boldsymbol{V}\hat{\boldsymbol{q}}(t,\boldsymbol{\mu});\boldsymbol{\mu})$ 时，可能只需对 n 求值，而不必对 $\boldsymbol{f}$ 的所有 N 个分量函数求值。因此，在这种情况下，经验插值可以绕过提升瓶颈。更多详情请参阅 Barrault 等人（2004）以及 Chaturantabut 和 Sorensen（2010）的讨论。

插值点的选择 插值点（2.13）的选择对经验插值近似的精度至关重要。已有多种方法用于选择插值点。Barrault等人(2004）和Chaturantabut 和 Sorensen (2010）提出了贪心方法，在每一步中根据插值残差依次添加点。由Drmaˇc 和 Gugercin (2016年）提出的 QDEIM 算法根据转置基矩阵 $\boldsymbol{V}^\top$ 的 QR 分解中的主元元素选择插值点；与贪心选择插值点所对应的误差界相比，这种方法可能得到更紧的误差界；这种基于 QR 的采样方法与最大体积子矩阵选择有关（Drmaˇc 和 Gugercin 2016）。

过采样和经验回归 标准经验插值会在恰好 n 个采样点处强制执行插值，而一种常见的推广方法是放宽这一要求，使用多于降阶空间维数的采样点。选择 $n_s > n$ 个点 p_1、……、p_n、p_{n+1}、……、p_{n_s} 会得到经验回归，

$$\tilde{\boldsymbol{f}}(\hat{\boldsymbol{q}};\boldsymbol{\mu}) = (\boldsymbol{P}^\top\boldsymbol{V})^+\boldsymbol{P}^\top\boldsymbol{f}(\boldsymbol{V}\hat{\boldsymbol{q}};\boldsymbol{\mu}),$$

其中 $(\boldsymbol{P}^\top\boldsymbol{V})^+$ 表示 $\boldsymbol{P}^\top\boldsymbol{V}$ 的 Moore–Penrose 伪逆。取 $n_s > n$ 个点称为过采样；在模型降阶领域，最早应用过采样的是受缺失数据 POD 方法（Everson 和 Sirovich 1995）启发的方法，例如（Astrid 等 2004、2008、Carlberg 等 2011）。Argaud、Bouriquet、de Caso、Gong、Maday 和 Mula （2018）指出了一种过采样至关重要的情况：他们表明，在使用 $n = n_s$ 个点时，加入噪声或扰动可能会导致经验插值出现不稳定性。Peherstorfer、Drmaˇc 和 Gugercin （2020）提出进行随机过采样，即 $n_s > n$，并证明随机选择点可以缓解这种不稳定性。

需要特别指出的是，Barrault 等人（2004）提出的贪心算法、Chaturantabut 和 Sorensen （2010）以及 Drmaˇc

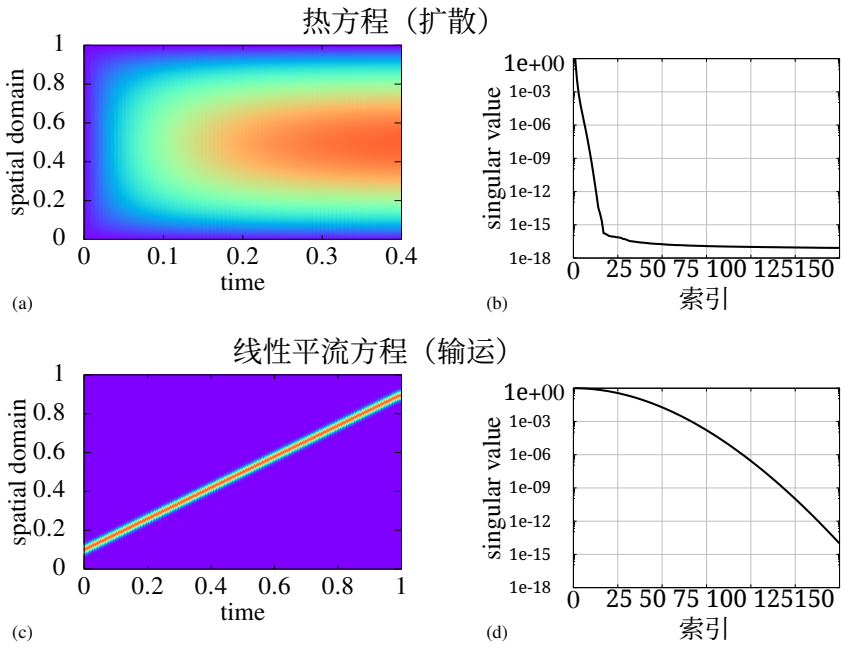


图 2.1：与热方程示例对应的快照矩阵，其奇异值迅速衰减，表明低维子空间能够高效地捕捉快照集。相比之下，平流方程的奇异值衰减速度明显较慢，这凸显了线性约化空间逼近在输运主导动力学中的局限性。（首次发表于美国数学会出版的《美国数学会通告》第 69 卷第 5 期（2022年）。© 2022年 美国数学会。）

和 Gugercin （2016年）均不能直接用于选择 $n_s > n$ 个点。相反，有多种过采样策略可供采用（Astrid等人 2008年、Carlberg等人2011）。一类研究侧重于最小化采样算子的范数$\boldsymbol{P}^\top \boldsymbol{V}$，并将添加一个点解释为对算子进行低秩更新 $\boldsymbol{P}^\top \boldsymbol{V}$，进而推导出选择准则（Zimmermann 和 Willcox 2016，Peherstorfer 等人 2020）。

2.5. 扩散主导与输运主导问题的线性模型降阶

现在，让我们来看一个数值示例，说明线性模型降阶在输运主导问题中何时会变得低效。以下内容参照 Peherstorfer (2022年）。我们将区分扩散主导与输运主导问题。

2.5.1. 扩散问题：热方程

以扩散主导问题为例，考虑热方程

$$\partial_t q_{\mathcal{Q}}(t,x;\mu) - \mu \partial_x^2 q_{\mathcal{Q}}(t,x;\mu) = 1\,, \qquad x \in \Omega\,, \tag{2.15}$$

定义在一维空间域 $\Omega = [0,1]$ 上，并满足齐次狄利克雷边界条件。初始条件为常数函数，在 Ω处取值为零。参数 μ 对应热传导系数，在本例中设为 $\mu = 1$ 。图 2.1a 展示了热方程 (2.15) 随时间和空间变化的数值解。该数值解在 Ω中使用线性有限元，在 $N = 1024$ 个节点（包括边界点）上计算，并采用隐式欧拉法进行时间离散，时间区间为 $\mathcal{T} = [0, 0.4]$，共 $K = 400$个时间步，时间步长为 10^{-3} 。现在，按照第 2.3.1 节概述的线性模型降阶离线步骤 1，收集快照 (2.7)。请记住，$\mu = 1$ 是固定的。我们在 $K + 1$ 个离散时间点处提取快照，构成定义于 (2.9) 的快照矩阵 $\boldsymbol{Q}_{\text{train}}$，并在图 2.1b 中绘制归一化奇异值。归一化奇异值为比值 $\bar{\sigma}_i = \sigma_i/\sigma_1$，其中 $i = 1, \ldots,$ $K + 1$，因此第一个归一化奇异值为 1。

图2.1b 中绘制的归一化奇异值迅速衰减，其含义是：由前 $n = 20$ 个左奇异向量张成的空间，能够在相对 Frobenius 范数下以双精度近似快照数据。由完整模型得到的快照向量维数在本例中为 $N = 1024$ （包括两个边界点），比降阶维数 n高出 50 倍以上。奇异值的快速衰减表明，在该热方程示例中收集的快照可以用低维空间 $\mathcal{V}$很好地近似。虽然奇异值并不能保证快照矩阵之外的完整解的近似误差，但奇异值仍常被用作估计可期望降阶程度的启发式方法。

2.5.2. 输运问题：线性平流方程

现在我们考虑一个典型的输运问题示例，即线性平流方程，

$$\partial_t q_{\mathcal{Q}}(t,x;\mu) + \mu \partial_x q_{\mathcal{Q}}(t,x;\mu) = 0\,, \tag{2.16}$$

在实数轴上，其中 μ 对应于平流速度，在本例中固定为 $\mu = 4/5$。初始条件 $q_{\mathcal{Q}},0$： $\mathbb{R} \to \mathbb{R}$ 是均值为 10^{-1}、标准差为 1.5×10^{-2} 的高斯概率密度函数。问题 (2.16) 的闭式解为 $(t,x) \mapsto q_{\mathcal{Q}},0(x - 4t/5)$，因为我们固定了 $\mu = 4/5$。我们在 $\Omega = [0,1]$ 中的 $N = 1024$ 个等距点处计算解析解，并在 $[0,1]$，中取 $K = 400$ 个等距时间步，从而得到 $K + 1$ 个时间快照。图 2.1c 以时空表示的形式绘出了这些快照。图 2.1d 绘出了相应快照矩阵的归一化奇异值。

归一化奇异值的衰减速度比扩散主导问题中的情况更慢。特别是，结果似乎符合代数衰减，这与扩散情形中观察到的快速衰减形成对比。要以双精度相对 Frobenius 范数精度表示快照数据，需要一个维数为 $n \geq 150$ 的空间

此时压缩比为 N/n，比扩散问题低约一个数量级。缓慢衰减与该输运问题解流形的几何结构相符：动力学仅使一个局部特征（高斯波包）在区域内平移。这意味着快照是同一函数的平移副本，而向量空间必须使用许多基函数，才能表示该特征在不同空间位置上的形态。因此，问题并不在于解流形具有高维性——事实上，在这个例子中，由于参数 μ 固定，解流形是一维的——而在于用低维向量空间近似这个解流形的效率较低，线性模型降阶便是如此。

总之，这一数值示例表明，以达到给定精度所需降阶空间的维数来衡量，线性模型降阶在输运主导问题中的效率可能远低于在扩散主导问题中的效率。

3. Kolmogorov屏障

理解线性模型降阶性能极限的一个核心工具是 Kolmogorov n宽度，它衡量将近似限制在 n维子空间内时所能达到的最优最坏情况误差。本节简要回顾这一概念，并讨论 n宽度的下界；这些下界揭示了线性模型降阶方法在输运主导问题中的根本局限，并促使我们提出 Kolmogorov 屏障这一概念。本节仅旨在概述相关的逼近理论概念。关于模型降阶的逼近理论基础的全面论述，请参阅 Cohen 和 DeVore (2015年、2016年）、Benner等人 (2017年），Cohen、DeVore、Petrova 和 Wojtaszczyk (2022年*b*）。

3.1. 向量空间中的最佳线性逼近：*Kolmogorov* n-宽度

设 $\mathcal{Z}$ 为希尔伯特空间，且 $\mathcal{N} \subseteq \mathcal{Z}$ 为其子集。我们可以将 $\mathcal{Z}$ 视为定义偏微分方程问题（2.1）的空间，或视为与完整模型（2.4）相对应的有限维近似空间。相应地，子集 $\mathcal{N}$ 可以非正式地理解为偏微分方程问题（2.1）或完整模型（2.4）的所有解的集合。对于给定的 n维子空间 $\mathcal{Z}_n \subseteq \mathcal{Z}$，我们将 $\mathcal{N}$到 $\mathcal{Z}_n$ 的偏差定义为

$$d(\mathcal{N}, \mathcal{Z}_n) = \sup_{z \in \mathcal{N}} \inf_{z_n \in \mathcal{Z}_n} \|z - z_n\|, \tag{3.1}$$

这可以解释为最佳逼近的最坏情形。由于 $\mathcal{Z}$ 是希尔伯特空间，下确界可通过将 z 正交投影到 $\mathcal{Z}_n$上得到；由于 $\mathcal{Z}_n$ 的维数 n有限，这种投影总是存在的。此外，投影映射是线性的，这意味着将最佳逼近限制在希尔伯特空间的子空间中，会得到线性（基于子空间的）近似。

现在，我们对 (3.1) 在所有 n 维子空间上取下确界 $\mathcal{Z}$，从而定义 Kolmogorov n 宽度

$$d_n(\mathcal{N},\mathcal{Z}) = \inf_{\substack{\mathcal{Z}_n\le\mathcal{Z}\\ \dim(\mathcal{Z}_n)\le n}} d(\mathcal{N},\mathcal{Z}_n) = \inf_{\substack{\mathcal{Z}_n\le\mathcal{Z}\\ \dim(\mathcal{Z}_n)\le n}} \sup_{z\in\mathcal{N}} \inf_{z_n\in\mathcal{Z}_n} \|z-z_n\|, \tag{3.2}$$

一个最初由Kolmogoroff（1936）提出的想法。按照Pinkus（1985），一个n维子空间 $\mathcal{Z}_n^\star$（属于 $\mathcal{Z}$）被称为 d_n的最优子空间（$\mathcal{N}$， $\mathcal{Z}$），如果

$$d_n(\mathcal{N},\mathcal{Z}) = d(\mathcal{N},\mathcal{Z}_n^\star).$$

3.2. Kolmogorov n-宽度的快速衰减与缓慢衰减及Kolmogorov屏障

现在让我们将Kolmogorov n-宽度的概念应用于模型降阶的设定。如上所述，我们可以将 $\mathcal{Z}$ 视为空间 $\mathcal{Q}$，偏微分方程问题（2.1）在该空间上定义，并将 $\mathcal{N}$ 视为所有解的集合。类似地，我们可以将 $\mathcal{Z}$ 视为有限维近似空间 $\mathcal{Q}_N$，其对应于完整模型（2.4），并将子集 $\mathcal{N}$ 视为完整模型的所有解（2.5）的集合。

在此设定下，Kolmogorov n-宽度（3.2）是在 n维降阶空间中近似这些解时所能达到的最小最坏情况误差。因此，Kolmogorov n-宽度随维数 n （即降阶空间的维数）衰减的快慢，对线性模型降阶的效率具有重大影响。如果它衰减很快，那么一个降阶低维空间通常就足以满足给定的误差容限。因此，在这种情况下，线性模型降阶可以是高效的，至少从近似理论的角度来看是如此。相反，如果Kolmogorov n-宽度衰减缓慢，那么为满足误差容限所需的降阶空间维数可能高到使线性模型降阶变得低效。在这种情况下，我们称之为Kolmogorov屏障，因为它构成了线性模型降阶的硬性限制

我们来看一些模型降阶文献中的例子，以阐明在此背景下快速衰减和缓慢衰减的含义：Maday 等人（2002）研究了一维参数 μ 下的椭圆强制性偏微分方程及其相应的解集。Maday 等人（2002）表明，在这种情况下，Kolmogorov n-宽度存在一个上界，该上界随降阶空间的维数 n 呈指数快速衰减。因此，在线性模型降阶中增加降阶空间的维数，可以实现误差的指数级降低。Cohen、DeVore 和 Schwab（2011），Cohen 和 DeVore（2015，2016），以及 Bachmayr 和 Cohen（2017）针对更一般的椭圆问题，证明了类似的快速衰减结果，尽管这种衰减不一定是指数级的。作为补充说明，除存在性结果外，Binev 等人（2011）、Buffa 等人（2012）和 Cohen 等人（2020）还表明，可以通过贪心方法构造一系列降阶空间，使其达到与 n-宽度相同的衰减率（相差一个常数因子）。

Br宽泛地说，扩散主导问题通常属于这一类，例如所讨论的

第 2.5 节中的问题；对于这类问题，Kolmogorov n-宽度衰减很快，因此非常适合采用线性模型降阶。

3.3. 线性平流问题的 n-宽度缓慢衰减

现在我们来考虑一些偏微分方程问题，这些问题的解流形的 Kolmogorov n-宽度下界已得到证明。

我们来考虑一维空间域 $\Omega = [0, 1]$ 和时间域 $[0, 1]$, 上的线性平流问题

$$\partial_t q_{\mathcal{Q}}(t, x; \mu) + \mu \partial_x q_{\mathcal{Q}}(t, x; \mu) = 0\,, \tag{3.3}$$

满足以下条件

$$q_{\mathcal{Q}}(0, x; \mu) = 0\,, \qquad q_{\mathcal{Q}}(t, 0; \mu) = 1\,,$$

对所有 $\mu \in [0, 1]$ 均成立。我们来考虑仅针对 $\mu = 1$ 的相应解流形,

$$\mathcal{M} = \{q_{\mathcal{Q}}(t, \cdot; \mu) \,|\, t \in [0, 1], \mu = 1\}\,, \tag{3.4}$$

由阶跃函数组成

$$q_{\mathcal{Q}}(t, x; \mu) = \begin{cases} 1\,, & 0 \le x \le t\mu\,, \\ 0\,, & t\mu < x \le 1\,, \end{cases} \tag{3.5}$$

这些阶跃函数在空间域中的某处存在不连续点 $[0, 1]$。注意，此处我们使用解集 $\mathcal{M}$ 而非有限维（完整模型）空间 $\mathcal{Q}_N$，如（2.5）中那样，而是在 $\mathcal{Q} = L^2$（$[0, 1]$）上。Ohlberger 和 Rave（2016年）证明，关于 n范数的 Kolmogorov L^2 宽度的下界为

$$d_n(\mathcal{M}, \mathcal{Q}) \ge c \frac{1}{\sqrt{n}}\,,$$

其中， n 是降阶空间的维数， $c > 0$ 是一个常数。

我们可以将此界理解为：不存在一列 L^2 子空间，能够在 Kolmogorov n宽度（3.2）意义下实现比 $1/\sqrt{n}$更快的近似误差衰减——这比第3.2节讨论的单参数扩散主导问题中可实现的指数误差衰减要慢得多。此结果也与第2.5节所示的数值实验相符：基于线性平流方程的示例中，奇异值所反映的投影误差衰减在经验上较慢；而以热方程为例的扩散主导问题中，奇异值则呈指数衰减。我们再次强调，奇异值的衰减并不直接对应于 Kolmogorov n宽度的衰减；不过，在某些条件下，奇异值可以提供经验性指示；参见第2.5节。

Arbes 等人 (2025) 表明，如果线性平流中的初始条件

示例选择得更平滑，那么这种平滑性也会体现在 Kolmogorov n-宽度的衰减率上。波动方程的类似结果见Greif 和 Urban(2019)。

总的来说，这些近似理论结果，以及第 2.5 节和第 1.2 节所讨论的实证证据表明，具有类波现象和移动的相干局部特征的输运主导问题会导致 n-宽度缓慢衰减，从根本上限制了线性模型降阶技术。

注 3.1。 本综述聚焦于时变问题；然而，即使对于定常问题，线性模型降阶也可能因 Kolmogorov 屏障而变得低效。例如， Welper (2017年), Taddei (2020), Alireza Mirhoseini 和 Zahr (2023) 研究了这样的问题：参数 $\boldsymbol{\mu}$ 会引发空间域中某种相干结构的输运。

3.4. Kolmogorov n-宽度与线性控制系统

本综述始终聚焦于有限维参数空间。为补充这一图景，我们在此给出一个具有无限维参数空间的具体示例，即如下形式的线性、有限维、时不变控制系统

$$\begin{aligned} \dot{\mathbf{q}}(t) &= \boldsymbol{A}\mathbf{q}(t) + \boldsymbol{B}\boldsymbol{u}(t), \\ \boldsymbol{y}(t) &= \boldsymbol{C}\mathbf{q}(t), \end{aligned} \tag{3.6}$$

矩阵 $\boldsymbol{A} \in \mathbb{R}^{N\times N}$、 $\boldsymbol{B} \in \mathbb{R}^{N\times N_{\text{in}}}$和 $\boldsymbol{C} \in \mathbb{R}^{N_{\text{out}}\times N}$。符号 $\boldsymbol{u}$和 $\boldsymbol{y}$ 分别表示系统的输入和输出，而输入$\boldsymbol{u} \in L^2(-\infty,\ 0;\ \mathbb{R}^{N_{\text{in}}})$ 充当参数。我们假设控制系统（3.6）是渐近稳定的，即矩阵 $\boldsymbol{A}$ 的特征值均位于复平面的开左半平面内。在模型降阶方面，我们关注输入到输出算子的近似

$$\mathcal{H}\colon L^2(-\infty, 0; \mathbb{R}^{N_{\text{in}}}) \to L^2(0, \infty; \mathbb{R}^{N_{\text{out}}}), \qquad \boldsymbol{u} \mapsto \boldsymbol{y}, \tag{3.7}$$

该算子将过去的输入映射为未来的输出。在文献中，算子 $\mathcal{H}$ 称为汉克尔算子。引入脉冲响应 $\boldsymbol{h}(t) = \boldsymbol{C}\exp(t\boldsymbol{A})\boldsymbol{B}$，汉克尔算子可通过卷积积分表示为

$$(\mathcal{H}\boldsymbol{u})(t) = \int_{-\infty}^{0} \boldsymbol{h}(t-s)\boldsymbol{u}(s)\,\mathrm{d}s.$$

众所周知， $\mathcal{H}$ 是一个秩至多为 N的有限秩算子；参见Antoulas(2005，第 5.4 节）或Francis (1987，第 5.1 节）。因此， $\mathcal{H}$ 是一个紧算子，其奇异值通常称为汉克尔奇异值，可通过矩阵乘积 $\mathcal{P}_{\text{c}}\mathcal{P}_{\text{o}}$的特征值的平方根计算，其中$\mathcal{P}_{\text{c}}$， $\mathcal{P}_{\text{o}} \in \mathbb{R}^{N\times N}$ 是李雅普诺夫方程的唯一解

$$\boldsymbol{A}\mathcal{P}_{\text{c}} + \mathcal{P}_{\text{c}}\boldsymbol{A}^{\top} + \boldsymbol{B}\boldsymbol{B}^{\top} = 0 \qquad \text{and} \qquad \boldsymbol{A}^{\top}\mathcal{P}_{\text{o}} + \mathcal{P}_{\text{o}}\boldsymbol{A} + \boldsymbol{C}^{\top}\boldsymbol{C} = 0. \tag{3.8}$$

我们按降序排列汉克尔奇异值，并将第 i个汉克尔奇异值记为 $\sigma_i(\mathcal{H})$，其中 $i = 1, \ldots\ldots, N$。对于 $\mathcal{Z} = L^2(-\infty, 0; \mathbb{R}^{N_{\text{in}}})$，Unger 和 Gugercin (2019)证明，Kolmogorov n-宽度与汉克尔奇异值一致，具体而言：

$$d_n(\mathcal{H}(\{\boldsymbol{u} \in \mathcal{Z} \mid \|\boldsymbol{u}\|_{\mathcal{Z}} \leq 1\}), \mathcal{Z}) = \sigma_{n+1}(\mathcal{H}). \tag{3.9}$$

我们强调，与之前的结果不同，(3.9) 并非下界或上界，而是一个可以显式计算的精确恒等式。虽然 n-宽度，也就是汉克尔奇异值的计算，在计算方面已有充分研究，且在大规模情形下可行，参见Simoncini(2016年)的综述，但要计算出一个具有以下形式的显式降阶模型

$$\begin{aligned}\dot{\hat{\mathbf{q}}}(t) &= \hat{\boldsymbol{A}}\hat{\mathbf{q}}(t) + \hat{\boldsymbol{B}}\boldsymbol{u}(t),\\ \hat{\boldsymbol{y}}(t) &= \hat{\boldsymbol{C}}\hat{\mathbf{q}}(t),\end{aligned}$$

能够达到这种近似精度，则是一项并非易事的任务。在文献中，这被称为最优汉克尔范数逼近；GloverGlover(1984年) 首次解决了这一问题，其依据是 Adamyan、Arov 和 Krein (1971) 的研究成果。关于描述符系统的扩展，见 Cao、Saltik 和 Weiland (2015年) 以及Benner 和 Werner (2020)。

虽然式 (3.7) 中解算子的空间便于进行精确的解析处理，但在实际应用中，人们往往关注解算子

$$\begin{aligned}\mathcal{S}_1 &: L^2(0, \infty; \mathbb{R}^{N_{\text{in}}}) \to L^2(0, \infty; \mathbb{R}^{N_{\text{out}}}), && \boldsymbol{u} \mapsto \boldsymbol{y},\\ \mathcal{S}_2 &: L^2(0, \infty; \mathbb{R}^{N_{\text{in}}}) \to L^\infty(0, \infty; \mathbb{R}^{N_{\text{out}}}), && \boldsymbol{u} \mapsto \boldsymbol{y},\end{aligned}$$

它们对应于哈代 $\mathcal{H}_\infty$ 和 $\mathcal{H}_2$ 范数意义下的近似。尽管如此，研究汉克尔算子仍具有重要意义，因为著名的平衡截断先验误差界是以截断的汉克尔奇异值给出的 $\mathcal{H}_\infty$ 范数上的界。

我们以四点评论结束对线性控制系统的简短探讨。首先，Unger 和 Gugercin（2019）的分析表明，控制系统的 Kolmogorov$n$-宽度可被理解为 Constantine（2015）提出的主动子空间方法的对偶概念。其次，如果矩阵 $\boldsymbol{A}$、$\boldsymbol{B}$ 和 $\boldsymbol{C}$ 具有参数依赖，且参数 $\boldsymbol{\mu}$ 属于紧参数集 $\mathcal{D}$，那么 Unger 和 Gugercin（2019，定理 3）证明，Kolmogorov n-宽度的下界为该参数集上汉克尔奇异值的最大值。因此，正如 Wittmuess、Tarin、Keck、Arnold 和 Sawodny（2016）、Gosea、Gugercin 和 Unger（2021）以及 Hund、Mitchell、Mlinari´c 和 Saak（2022）所示，允许使用参数依赖投影矩阵可以改进 Kolmogorov n-宽度。第三，Zuyev、Feng 和 Benner（2024）提出了向双线性控制系统的扩展。第四，Kolmogorov$n$-宽度在很大程度上取决于空间 $\mathcal{Z}$ 及其关联范数。鉴于输运主导问题及其相关方法，标准的 L_p-范数可能

无法完整呈现情况，Wasserstein距离等其他度量可能会带来不同的见解（Ehrlacher、Lombardi、Mula 和 Vialard 2020）；另请参见第5.2.2节。

4. 非线性模型降阶的要素

非线性模型降阶旨在通过寻求全模型解的非线性近似，而非线性近似，来克服 Kolmogorov 屏障。

4.1. 通过非线性参数化克服 *Kolmogorov* 屏障

我们现在介绍参数化的含义，随后讨论非线性参数化如何能够产生克服 Kolmogorov 屏障的非线性近似。

4.1.1. 参数化

参数化是一个具有如下签名的映射

$$\hat{q}\colon \mathbb{R}^n \times \Omega \to \mathbb{R}, \quad (\boldsymbol{\theta}, \boldsymbol{x}) \mapsto \hat{q}(\boldsymbol{\theta}, \boldsymbol{x}), \tag{4.1}$$

因此，对于一个 n 维权重向量 $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]^\top \in \mathbb{R}^n$，函数 $\hat{q}(\boldsymbol{\theta}, \cdot) : \Omega \to \mathbb{R}$ 是感兴趣的函数空间中的一个元素 $\hat{q}(\boldsymbol{\theta}, \cdot) \in \mathcal{Z}$，记为 $\mathcal{Z}$。例如，$\mathcal{Z}$ 可以是我们定义偏微分方程问题 (2.1) 所在的空间 $\mathcal{Q}$，也可以是定义全模型解 (2.2) 所在的空间 $\mathcal{Q}_N$。更准确地说，可以先定义一个映射 $Q : \mathbb{R}^n \to \mathcal{Z}$，以得到 $Q(\boldsymbol{\theta}) \in \mathcal{Z}$，然后在定义 $\mathcal{Z}$ 的空间域 Ω 中的各点处对其求值，得到 $Q(\boldsymbol{\theta})(\boldsymbol{x})$。直接使用 (4.1)，并在意指 Q 时写作 $\boldsymbol{\theta} \mapsto \hat{q}(\boldsymbol{\theta}, \cdot)$，会更方便。目前，我们尚未说明如何获得权重向量 $\boldsymbol{\theta}$，以近似感兴趣函数空间中的一个元素。对于给定函数生成权重向量的映射有时称为编码器映射（相应地，$\boldsymbol{\theta} \mapsto \hat{q}(\boldsymbol{\theta}, \cdot)$ 则称为解码器映射）。然而，在模型降阶中，我们通常无法访问在线阶段希望编码的函数，因为这些函数就是全模型解。因此，在线阶段必须通过降阶动力学计算权重向量；参见后文第 4.4 节。

如果参数化 (4.1) 对权重向量 $\boldsymbol{\theta}$呈线性（或仿射）依赖，即映射 $\boldsymbol{\theta} \mapsto \hat{q}(\boldsymbol{\theta}, \cdot)$ 是线性（或仿射）的，我们就称其为线性（或仿射）参数化。需要注意的是，线性指的是对 $\boldsymbol{\theta}$ 而非 $\boldsymbol{x}$的依赖，也就是说，映射 $\hat{q}(\boldsymbol{\theta}, \cdot) : \Omega \to \mathbb{R}, \boldsymbol{x} \mapsto \hat{q}(\boldsymbol{\theta}, \boldsymbol{x})$ 仍可能对空间坐标 $\boldsymbol{x}$呈非线性，这对所有 $\boldsymbol{\theta} \in \mathbb{R}^n$都成立。此外，权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 可能对时间 t 和参数 $\boldsymbol{\mu}$等量呈非线性依赖。

注 4.1. 向量 $\boldsymbol{\theta}$通常称为参数向量；不过，我们遵循模型降阶领域的惯例，将 $\boldsymbol{\mu}$ 称为参数，因此选择将 $\boldsymbol{\theta}$ 称为权重向量，以便用于神经网络等非线性参数化。

在模型降阶的语境中，权重向量 $\boldsymbol{\theta}$ 也称为降阶状态或潜在状态，因为正如我们将在后续章节中看到的，它可以表示降阶解。

4.1.2. 线性参数化与线性模型降阶

回顾线性模型降阶中的降阶试探形式，它由线性组合（2.6）给出。该试探形式也可以等价地写成系数（或降阶状态（2.10））的函数：

$$\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\boldsymbol{x}) = \sum_{i=1}^{n} \theta_i(t,\boldsymbol{\mu})\varphi_i(\boldsymbol{x}), \tag{4.2}$$

这使系数的线性依赖关系一目了然。注意，在（4.2）中，我们将线性组合（2.6）的系数与 n 权重$\theta_1(t,\boldsymbol{\mu}),\ldots,\theta_n(t,\boldsymbol{\mu})$ 对应起来，并将它们汇集为权重向量 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$，该向量依赖于时间 t 和参数 $\boldsymbol{\mu}$。从这一角度看，我们可以将降阶试探形式（2.6）解释为对降阶空间 $\mathcal{V}$ 的线性参数化（4.2）；该空间由降阶基函数 $\varphi_1,\ldots,\varphi_n$ 张成，权重向量 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$ 的维数为 n。换言之，线性模型降阶通过线性参数化来表示降阶解。

通过线性参数化（4.2）所能表示的所有降阶解构成降阶空间 $\mathcal{V}$，因为对于每个 $\boldsymbol{\theta}\in\mathbb{R}^n$，我们都有 $\hat{q}(\boldsymbol{\theta},\cdot)\in\mathcal{V}$；反过来，$\mathcal{V}$ 中的每个元素都可以写成 $\hat{q}(\boldsymbol{\theta},\cdot)$，其中 $\boldsymbol{\theta}\in\mathbb{R}^n$。降阶解的自由度就是向量 $\boldsymbol{\theta}$中的权重，它们可以依赖于时间 t 和参数 $\boldsymbol{\mu}$，但由基 $\mathcal{V}$ 给出的降阶空间φ_1，……，φ_n 保持固定，且独立于时间和参数。

4.1.3. 非线性参数化

现在来考虑非线性参数化，即关于权重向量 $\boldsymbol{\theta}$ 呈非线性的参数化。给定一种非线性参数化，我们可以通过改变权重 $\boldsymbol{\theta}$，考虑所有可达函数构成的集合：

$$\widehat{\mathcal{M}} = \{\hat{q}(\boldsymbol{\theta},\cdot) \mid \boldsymbol{\theta}\in\mathbb{R}^n\}, \tag{4.3}$$

该集合通常称为试探集或试探流形。线性参数化$\widehat{\ }$得到的集合 $\mathcal{M}$ 具有向量空间结构，而非线性参数化对应于更一般的试探集，它们不一定是向量空间，因此能够以更少的自由度 n更高效地表示目标解。概括而言，对权重向量 $\boldsymbol{\theta}$的非线性依赖意味着非线性参数化可能绕过 Kolmogorov屏障$\widehat{\ }$，因为由此产生的函数集 $\mathcal{M}$ 不再必然具有向量空间结构。

我们也可以将对权重向量 $\boldsymbol{\theta}$ 的非线性依赖理解为允许表示本身发生变化。为说明这一点，请考虑

参数化

$$\hat{q}(\boldsymbol{\theta}, \boldsymbol{x}) = \sum_{i=1}^{r} \beta_i \psi_i(\alpha_i, \boldsymbol{x}) . \tag{4.4}$$

这里，权重向量 $\boldsymbol{\theta} = [\boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top]^\top \in \mathbb{R}^n$ （其中 $n = 2r$ ）由 r 个权重 $\boldsymbol{\alpha} = [\alpha_1, \ldots\ldots \alpha_r]^\top \in \mathbb{R}^r$ 组成，这些权重通过表示函数 $\psi_1 \ldots\ldots \psi_r$ 以非线性方式参与： $\mathbb{R} \times \Omega \to \mathbb{R}$ ，以及 r 个权重 $\boldsymbol{\beta} = [\beta_1, \ldots\ldots \beta_r]^\top \in \mathbb{R}^r$ ，这些权重以线性方式参与。与线性参数化不同，在线性参数化中，表示由基函数 $\varphi_1 \ldots \ldots \varphi_n$ 固定，式 (4.4) 中的权重 $\alpha_1 \ldots\ldots \alpha_r$ 会直接影响函数 $\psi_1(\alpha_1, \cdot) \ldots\ldots \psi_r(\alpha_r, \cdot)$ 。因此，表示会适应所近似的特定函数。对于全模型解 $q(t, \cdot; \boldsymbol{\mu}) \in \mathcal{M}$ ，权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 因而不仅可能改变固定展开的系数，还可能改变底层表示本身的形状。这与线性模型降阶有本质区别：在线性模型降阶中，表示对应于基函数 $\varphi_1 \ldots\ldots \varphi_n$ ，这些基函数张成一个预先固定的向量空间，而线性参数化中的权重只决定相应线性组合的系数。

注 4.2。 机器学习中也有与线性参数化和非线性参数化相对应的概念。线性参数化通常对应于核近似（Sch¨olkopf 和 Smola 2001年）。选定核之后，学习问题就归结为寻找这些核的线性组合，也就是说，表示是固定的，而核空间中近似的系数（权重）以线性方式参与计算。相比之下，非线性参数化例如由至少包含一个隐藏层和非线性激活函数的神经网络给出（LeCun、Bengio 和 Hinton 2015年、Goodfellow、Bengio 和 Courville 2016）。内部层对应于表示，并通过权重进行调整。换句话说，网络会调整基函数本身，这称为表示学习（Rumelhart、Hinton 和 Williams 1986、Bengio、Courville 和 Vincent 2013）。

注 4.3。 除了权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 随时间演化的参数化之外，还可以考虑全局时间参数化，在这种参数化中，时间被视为另一个输入，而权重向量不依赖于时间；这在思想上类似于时空离散。本文不讨论全局时间参数化，而是聚焦于序贯或局部时间参数化，其中时间依赖性通过权重向量体现，这与广泛使用的模型降阶范式相一致，在该范式中，降阶状态由降阶动力系统推进。关于全局时间与局部时间（非线性）参数化的讨论，请参阅 Zhang、Chen、Vanden-Eijnden 和 Peherstorfer (2025)。

4.1.4. 示例：非线性参数化与线性平流让我们重新考察第 3.3 节中的线性平流方程问题。相应的解流形由（3.5）中给出的阶跃函数组成。我们将阶跃函数（3.5）视为平移参考剖面在 $[0,1]$ 上的限制，

$$q_{\mathcal{Q}}(t,x;\mu) = q_0(x - t\mu), \tag{4.5}$$

其中 $q_0\colon\ \mathbb{R} \to \mathbb{R}$ 定义为

$$q_0(x) = \begin{cases} 1, & x \le 0\,, \\ 0, & x > 0\,. \end{cases}$$

将线性平流问题的解表示为（4.5），揭示了一个关键难点：解可以写成平移 $(t,\ x,\ \mu) \mapsto x - t\mu$ 与初始函数 q_0的复合。线性参数化（4.2）能够高效地表示特征叠加，因为它是基函数的线性组合。然而，通过叠加来表示特征的平移，意味着必须在特征原来的空间位置将其停用（在叠加中的权重较低），并在新的空间位置将其激活（在叠加中的权重较高）。这一过程可能需要数量多到难以承受的基函数，才能描述特征的平滑移动。

借助非线性参数化（例如（4.4）中定义的参数化），通过将平移与特征叠加复合，可以更高效地描述平移。具体而言，我们来推导一种形式为（4.4）的非线性参数化，使其能够精确表示阶跃函数（4.5）。一种实例为

$$\hat{q}(\boldsymbol{\theta}(t,\mu),x) = \beta_1(t,\mu)\psi_1(\alpha_1(t,\mu),x), \tag{4.6}$$

其中 $\beta_1(t,\mu) = 1$ 和 $\alpha_1(t,\mu) = t\mu$，由此得到权重向量 $\boldsymbol{\theta}(t,\mu) = [\alpha_1(t,\mu),\ \beta_1(t,\mu)]^\top$。函数 ψ_1 定义为

$$\psi_1(\alpha,x) = q_0(x-\alpha)\,.$$

因此，阶跃函数（4.5）包含第 3.3 节线性平流问题的解流形（3.4）上的所有函数，可以通过非线性参数化（4.4）用仅 $n = 2$ 个权重精确表示。相比之下，使用线性参数化（4.2）时，误差的衰减速度无法快于 $1/\sqrt{n}$ 在 L^2 中的速率，详见第 3.2 节。

尽管这个例子是人为构造的，但它已经表明，通过恰当地选择非线性参数化，可以组合输运和特征叠加等现象，而不必仅靠特征叠加来近似所有现象。这种额外的灵活性有助于减少参数化所需的权重数量，从而提高效率。

4.2. Kolmogorov n-宽度向非线性参数化的扩展

Kolmogorov n-宽度概念可以推广到非线性参数化。有一种广泛使用的概念建立在类似自编码器的结构之上：

编码器将感兴趣函数集中的一个元素映射 $\mathcal{N} \subseteq \mathcal{Z}$ 为一个n维向量，即我们术语中的权重 $\boldsymbol{\theta}$。解码器将该n维向量映射回 $\mathcal{Z}$中的一个函数。随后，人们关注编码器与解码器复合后的最坏情况近似误差，以及该误差随维数 n 的衰减；参见例如 DeVore、Kyriazis、Leviatan 和 Tikhomirov(1993)，以及 DeVore (1998)。我们将在第 7节从数值视角重新审视自编码器。

存在能够实现较低近似误差的编码器-解码器对，并不足以保证数值计算具有实际意义。编码器和解码器映射的正则性至关重要，因为它能排除病态参数化，从而与数值稳定性相关联。这促使人们研究 Cohen 等人（2022年*b*）所考虑的稳定流形宽度；该研究要求编码器和解码器至少满足 Lipschitz 界。

Cohen、Farhat、Maday 和 Somacal （2024）进一步将编码器限制为由以下内容构成： n 对待编码解函数的线性泛函。解码器则可以保持非线性。这类非线性逼近的表达能力可用所谓的感知数来量化。具体而言，Cohen 等人 （2024）讨论了随时间输运的阶跃函数示例；对于这些示例，此类近似的误差衰减速度快于线性逼近所能达到的速度。事实上，Cohen 等人 （2022年*b*、2024）证明，感知数与稳定流形宽度具有相同的衰减率。与感知数相关的数值实验见 Glas 和 Unger（2025）。

4.3. 带有离线与在线权重的非线性参数化

数值分析、机器学习以及计算科学与工程中使用着范围广泛的非线性参数化，这也推动了将非线性参数化用于模型降阶。本节仅作简要综述，并将在本综述的后续章节中对其进行详细讨论。

4.3.1. 离线与在线权重

现在我们考虑可依赖于离线权重向量$\boldsymbol{\theta}_{\text{off}} \in \mathbb{R}^P$的参数化，

$$\hat{q}\colon \mathbb{R}^n \times \Omega \times \mathbb{R}^P \to \mathbb{R}, \qquad (\boldsymbol{\theta}, \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}) \mapsto \hat{q}(\boldsymbol{\theta}, \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}), \tag{4.7}$$

从而对于任意固定的 $\boldsymbol{\theta}_{\text{off}}$，我们得到一个参数化 $\hat{q}(\cdot, \cdot; \boldsymbol{\theta}_{\text{off}})\colon \mathbb{R}^n \times \Omega \to \mathbb{R}$ 与（4.1）中的定义一致。当离线权重向量 $\boldsymbol{\theta}_{\text{off}}$ 的依赖关系可由上下文明确时，我们会在记号中省略它，并如（$q(\boldsymbol{\theta}, \boldsymbol{x})$ 中所示，简记为 $\hat{}$4.1）。

在模型降阶的背景下，离线权重向量 $\boldsymbol{\theta}_{\text{off}}$ 在离线阶段固定，因此独立于时间 t 和参数 $\boldsymbol{\mu}$。相比之下，权重向量 $\boldsymbol{\theta}$ 的作用类似于线性模型降阶中的系数向量或状态向量（2.10），充当降阶（潜在）状态，因此在在线

阶段会随时间 t 和参数 $\boldsymbol{\mu}$变化。为明确区分 $\boldsymbol{\theta}$ 与离线权重向量 $\boldsymbol{\theta}_{\text{off}}$，我们有时将其称为在线权重向量。

我们大致将离线权重分为先验选择和经训练选择两类。在经训练的情况下，向量 $\boldsymbol{\theta}_{\text{off}}$ （或其部分）会在离线阶段根据快照数据进行拟合，这与线性模型降阶中根据快照构造张成降阶空间 $\mathcal{V}$ 的约化基类似。相比之下，在先验情况下，离线权重向量 $\boldsymbol{\theta}_{\text{off}}$ 要么为空，要么无需训练即可指定，例如选择一个带给定平移的解析变换。在线权重向量 $\boldsymbol{\theta}$ 仍需在在线阶段确定。

在模型降阶中，我们通常尤其关注经训练的非线性参数化，其中离线权重向量 $\boldsymbol{\theta}_{\text{off}}$ 会在离线阶段学习，以提供表达能力强的参数化，从而只需在线计算低维在线权重向量 $\boldsymbol{\theta}$。

*4.3.2. 带有离线-在线权重的非线性参数化示例*一类常见的非线性参数化是关于在线权重的多项式参数化。一个二次示例如下：

$$\hat{q}(\boldsymbol{\theta}, \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}) = \sum_{i=1}^{n} \theta_i \varphi_i(\boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}) + \sum_{1 \le i \le j \le n} \theta_i \theta_j \psi_{ij}(\boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}) \tag{4.8}$$

具有函数 $\varphi_i(\cdot;\ \boldsymbol{\theta}_{\text{off}})\colon\ \Omega \to \mathbb{R}$ 以及 $\psi_{ij}(\cdot;\ \boldsymbol{\theta}_{\text{off}})\colon\ \Omega \to \mathbb{R}$ 对于 $i,\ j = 1,\ \ldots,\ n$。这些函数（或它们在离散化中的系数）通过拟合离线权重向量 $\boldsymbol{\theta}_{\text{off}}$来拟合快照数据；我们将在第7节重新讨论此类多项式近似。

将线性展开与非线性变换复合，会得到另一类重要形式。设 $\mathrm{T}(\cdot;\ \boldsymbol{\theta}_{\text{off}})\colon\ \mathbb{R}^r \times \Omega \to \Omega$ 是一个依赖于离线权重向量的变换。此外，设 $\varphi_1(\cdot;\ \boldsymbol{\theta}_{\text{off}}),\ \ldots,\ \varphi_r(\cdot;\ \boldsymbol{\theta}_{\text{off}})$ 是一组基函数。将在线权重向量分解为 $\boldsymbol{\theta} = [\boldsymbol{\alpha}; \boldsymbol{\beta}]$，其中$\boldsymbol{\alpha} \in \mathbb{R}^r$ 且 $\boldsymbol{\beta} \in \mathbb{R}^r$，可得

$$\hat{q}(\boldsymbol{\theta}, \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}) = \sum_{i=1}^{r} \beta_i \varphi_i(\mathrm{T}(\boldsymbol{\alpha}, \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}); \boldsymbol{\theta}_{\text{off}}), \tag{4.9}$$

为简化表述，我们假设 $n = 2r$。例如，$\mathrm{T}(\alpha,\ x) = x - \alpha$ 可以表示由标量平移 α 给出的平移，如第4.1.4节所述。在这里，可以训练离线权重向量 $\boldsymbol{\theta}_{\text{off}}$ 以确定基 $\varphi_1,\ \ldots,\ \varphi_r$ 并确定变换 T 的参数，而在线权重向量 $\boldsymbol{\theta}$ 则选择平移量和系数。我们将在第5节重新讨论此类非线性参数化。

我们在（4.4）中简要提及的一类非线性参数化，是自适应表示的参数化。令在线权重向量$\boldsymbol{\theta} = [\boldsymbol{\alpha};\ \boldsymbol{\beta}] \in \mathbb{R}^n$ 分解为特征向量 $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_{n_1}]^\top \in \mathbb{R}^{n_1}$和系数向量 $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_{n_2}]^\top \in \mathbb{R}^{n_2}$，其中 $n = n_1 + n_2$。此外，令 $\psi_1(\cdot, \cdot; \boldsymbol{\theta}_{\text{off}}), \ldots, \psi_{n_2}(\cdot, \cdot; \boldsymbol{\theta}_{\text{off}})\colon\ \mathbb{R}^{n_1} \times \Omega \to \mathbb{R}$ 为表示函数，这些函数

依赖于离线权重向量 $\boldsymbol{\theta}_{\text{off}}$，并在离线阶段基于快照数据进行训练。然后，该非线性参数化

$$\hat{q}(\boldsymbol{\theta}, \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}) = \sum_{i=1}^{n_2} \beta_i \psi_i(\boldsymbol{\alpha}, \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}), \tag{4.10}$$

对于固定的 $\boldsymbol{\alpha}$，它关于系数 $\boldsymbol{\beta}$ 呈线性，但通过特征向量 $\boldsymbol{\alpha}$关于在线权重向量 $\boldsymbol{\theta}$呈非线性。特别是，如果 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 随时间（以及参数）变化，那么该表示

$$\psi_1(\boldsymbol{\alpha}(t, \boldsymbol{\mu}), \cdot; \boldsymbol{\theta}_{\text{off}}), \ldots, \psi_{n_2}(\boldsymbol{\alpha}(t, \boldsymbol{\mu}), \cdot; \boldsymbol{\theta}_{\text{off}}) \colon \Omega \to \mathbb{R}$$

也会随时间变化。我们将在第 6 节讨论这类非线性参数化。

另一类常见的模型是神经网络。例如，考虑一个带线性输出层的全连接前馈神经网络，

$$\hat{q}(\boldsymbol{\theta}, \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}) = \mathcal{C}_L(\sigma(\mathcal{C}_{L-1}(\ldots \sigma(\mathcal{C}_1(\boldsymbol{x})) \ldots))) \tag{4.11}$$

其中 σ 表示逐分量应用的激活函数，而 $\mathcal{C}_\ell(\boldsymbol{y}) = \boldsymbol{W}_\ell \boldsymbol{y} + \boldsymbol{b}_\ell$ 表示第 l 层，其矩阵为 $\boldsymbol{W}_\ell$，偏差为 $\boldsymbol{b}_\ell$，且维数适当。随后，可以将部分权重矩阵 $\boldsymbol{W}_1$、……、$\boldsymbol{W}_L$ 和偏差 $\boldsymbol{b}_1$、……、$\boldsymbol{b}_L$收集到离线权重向量 $\boldsymbol{\theta}_{\text{off}}$ 中，将其他部分收集到在线权重向量 $\boldsymbol{\theta}$中。一种方法是将前l 层的所有权重矩阵和偏差纳入离线权重向量，并将最后 $L-l$ 层的权重矩阵和偏差纳入在线权重向量 $\boldsymbol{\theta}$。不过还有其他多种变体，我们将在第 7 节中详细讨论。

4.4. 非线性模型降阶的要素

我们确定了非线性模型降阶不可或缺的三个要素——参数化的选择、针对在线权重构建降阶动力学，以及用于高效在线评估的数值求解器——并将在综述的其余部分反复讨论这些要素。

4.4.1. 非线性模型降阶要素综述

截至目前，我们的讨论主要集中于非线性模型降阶的近似理论方面，尤其是非线性参数化如何克服线性模型降阶固有的局限性及其相关的Kolmogorov屏障。尽管富有表现力的非线性参数化是关键要素，但它们只是成功的非线性模型降阶方法的组成部分之一。

除了选择非线性参数化之外，还必须构造与该参数化相兼容的降阶动力学，也就是说，为在线权重构建演化方程，使其解能够在不同时间和参数下生成准确的降阶解，并且能够稳健求解；这类似于为完整模型选择合适的公式和离散化。此外，模型降阶的最终目的

是在达到指定精度的前提下，相较于全模型模拟降低计算成本，而不仅仅是将自由度数量从N减少到 n。因此，单独实现 $n \ll N$ 还不够，除非降阶模型能够在在线阶段高效求值。高效的在线计算通常既需要合适的降阶动力学，也需要用于时间积分和非线性求解器的高效数值方法。

基于这些考虑，我们识别并区分非线性模型降阶的三个要素：

(i) 能提供高表达能力近似并绕过 Kolmogorov屏障的非线性参数化；(ii) 与所选参数化相兼容，并以准确、稳健的方式控制在线权重演化的降阶动力学；(iii) 高效的在线评估策略和数值求解器，使降阶模型仿真相较于全模型模拟具有计算优势。

这些要素与数值求解偏微分方程的标准概念相呼应：首先选择一个近似类（例如有限元空间），再选择一种用于定义问题的公式（例如 Galerkin 投影、最小二乘、混合方法），然后选择一种高效求解问题的数值方法。

注释4.4。 误差估计和认证通常构成模型降阶中的另一个独立要素。尽管线性模型降阶的严格且高效的后验误差估计已相当成熟（Grepl 和 Patera 2005，Veroy 和 Patera 2005，Rozza 等 2008年，Haasdonk 和 Ohlberger 2008年，2011），非线性模型降阶的相关研究尚处于起步阶段。我们举例提及Klein 和 Ohlberger （2025）以及Black、Schulze 和 Unger （2020）。不过，我们注意到，非线性降阶模型可以与完整模型结合，应用于多保真方法，从而为外层循环结果的近似提供精度保证；关于多保真方法的综述，参见Peherstorfer等 （2018年）。

4.4.2. 要素*1*：非线性参数化

第一个要素是非线性参数化，它定义了用于近似全模型解的降阶解的试探形式。通常需要指定合适的参数化架构 $\hat{q}$，例如多项式试探形式（4.8）或神经网络（4.11），然后通过在快照数据上训练，在离线阶段确定离线权重 $\boldsymbol{\theta}_{\text{off}}$。

利用快照数据进行训练，以拟合离线权重 $\boldsymbol{\theta}_{\text{off}}$，旨在使参数化能够捕捉全模型解的主要特征，从而使 n 在线权重向量的维数 $\boldsymbol{\theta} \in \mathbb{R}^n$ 保持较低。此外，参数化还应满足适当的正则性性质，例如，在线权重向量 $\boldsymbol{\theta}$ 的小幅变化会引起受控的变化 $\hat{q}(\boldsymbol{\theta}, \cdot; \boldsymbol{\theta}_{\text{off}})$，这对于稳健的在线计算很重要；另请参见第 4.2节。

更一般地，选择参数化时必须考虑预期的降阶动力学和在线评估，因为其结构会显著影响评估 $\hat{q}$ 以及组装在线演化降阶动力学所需量的成本。

4.4.3. 要素 *2*：降阶动力学

非线性模型降阶的第二个要素是对非线性参数化的权重施加降阶动力学，即推导出相应方程，以确定权重向量 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$ 如何随时间演化并依赖于参数，从而使 $\hat{q}(\boldsymbol{\theta}(t,\ \boldsymbol{\mu}),\ \cdot)$ 在某种合适的意义下逼近全模型解。

在本综述始终考虑的设定中，通过以降阶方式将控制方程施加于非线性参数化上，$\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$ 会针对参数 $\boldsymbol{\mu} \in \mathcal{D}$ 在时间区间 $t \in [0, T]$ 上推进。线性模型降阶中的对应做法是推导关于降阶坐标的降阶模型；然而，对于非线性参数化，Galerkin 投影等标准技术不能直接照搬，因为经典投影框架严重依赖于线性参数化所具有的向量空间结构。

因此，非线性模型降阶的一个关键方面，是开发适用的降阶动力学公式，使其与非线性参数化兼容，并支持可靠的在线计算。特别是，降阶动力学高度依赖于映射 $\boldsymbol{\theta} \mapsto \hat{q}(\boldsymbol{\theta},\ \cdot; \boldsymbol{\theta}_{\text{off}})$，因此其鲁棒性取决于该映射的正则性和条件性。在实践中，理想情况下，所得方法应能够处理映射条件变差的区域。

4.4.4. 要素*3*：高效的在线评估策略和求解器

第三个要素涉及对降阶动力学进行高效的在线求解，即采用数值方法推进在线权重向量 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$ 随时间演化（或者在稳态情况下求解相应的降阶代数问题）。

根据时间离散的不同，这涉及时间积分器；对于隐式格式，还需要在每个时间步求解非线性系统，通常采用迭代方法，而这些方法可能需要反复求解由参数化 $\hat{q}$ 所诱导的雅可比系统（或类似雅可比的系统）。对于许多与非线性参数化兼容的降阶动力学（例如瞬时残差最小化），每个时间步都可以解释为一个优化问题，用于求得权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 的更新量。因此，非线性模型降阶中的在线计算，其本质通常更接近优化，而不是经典的降阶线性系统求解。此外，在许多变分公式中，降阶动力学是通过要求残差与测试方向正交而得到的，这会产生涉及非线性参数化 $\hat{q}$ 的内积和范数，并且通常还涉及其关于 $\boldsymbol{\theta}$ 的导数。在在线阶段，必须通过数值求积或其他数值手段计算这些量。在离线阶段，适用于线性（及仿射）情形的标准技术通常是预先计算局部矩阵，然后通过叠加快速地在线组装矩阵；但这些技术

通常已不再直接适用，因为这些量通过 $\hat{}q$ 非线性地依赖于 $\boldsymbol{\theta}$，因此一般无法仅由预计算项组装得到。这在本质上与线性模型降阶中的提升瓶颈密切相关。因此，与线性模型降阶相比，将自由度数量从 N 减少到 n 更不能单独意味着在线计算成本的降低。

5. 基于变换的方法

本节综述的非线性模型降阶方法基于变换。其核心思想是找到合适的坐标系，使解流形适于进行线性降阶。方法并非直接在原始变量中近似解，而是引入一个或多个变换，以捕捉输运、旋转或变形等主要非线性特征。基于变换的方法有一个共同特点：变换要么由对底层问题的物理洞察所启发，要么是在从数据中学习得到时，能够在事后获得有意义的物理解释。

5.1. 使用基于变换的方法进行模型降阶

第 3.3 节讨论的一维对流方程这一启发性示例，以及第4.1.4 节中将解表示为模板平移的方式，表明可以引入一个或多个显式变换映射 T，这些映射通常采用坐标映射或输运映射的形式。这类变换可以先验给定，例如已知空间平移的情况；也可以从数据中推断得到。确定合适的变换后，可通过两种主要方式将其纳入模型降阶方法：

(i) 变换可以直接作用于解，从而对变换后的全模型解 $q \circ \mathrm{T}$ 应用线性模型降阶技术，而不是作用于 q 本身。(ii) 或者，变换也可以作用于降阶基函数 φ_i，从而得到一种基于变换模态的近似，如（4.9）所述。

根据变换的性质，这两种方法在某些情况下可能等价。然而，一般而言，它们会产生不同的降阶模型。当不存在能够将控制方程变换为适于高效线性降阶表示形式的微分同胚时，这种区别尤为重要。脉冲爆震燃烧室建模中就存在这样的情况（Schulze 等 2019）。

此外，变换不必局限于空间域或时间域的映射。还可以考虑直接作用于降阶基函数的更一般的变换算子 $\mathcal{T}$，从而得到形式为 $\mathcal{T}_i\varphi_i$ 的变换模态。在所有这些变体中，将线性模型降阶与

变换相结合，可得到具有清晰物理解释的降阶模型，因为解被分解为不同的特征，而这些特征的演化则通过针对具体应用定制的显式变换来描述。

5.2. 基于变换的方法文献综述

我们综述基于变换的非线性模型降阶文献。

5.2.1. 通过群作用与移位实现的变换

一个经典且被广泛探索的想法，是利用底层系统的已知对称性，例如平移不变性或旋转不变性。在这一设定中，与群作用相关联的动力学被从解场的其余演化中分离出来，随后使用线性模型降阶技术进行近似（Kirby 和 Armbruster 1992，Rowley 和 Marsden，2000年，Rowley、Kevrekidis、Marsden 和 Lust 2003，Mowlavi 和 Sapsis 2018年）。与之密切相关的冻结法由Beyn 和 Th¨ummler （2004）提出，它构成了若干降阶技术的概念基础。例如，Ohlberger 和 Rave （2013）在模型降阶的背景下将这一想法应用于Burgers方程，并表明一旦将主导输运分离出去，剩余的解演化便呈现出低维结构，从而可以使用线性技术高效地进行降阶。

许多基于对称性的方法的一个局限是，它们通常只考虑单一的全局群作用，例如均匀平移。因此，若不加以扩展，要捕捉多个同时输运的结构（例如以不同速度传播的多个波）会面临很大挑战。其中一种扩展是 Reiss、Schulze、Sesterhenn 和 Mehrmann Reiss, Schulze, Sesterhennand Mehrmann (2020年Black et al. (2020年)、Reiss (2021年)、Papapicco、Demo、Girgoflio、Stabile 和 Rozza (2022)、Burela、Krah 和 Reiss (2025年)、Krah、Marmin、Zorawski、Reiss 和 Schneider (2025年)等人都介绍了移位 POD 的扩展、变体和推广。类似地，Rim、Moe 和 LeVeque (2018年)的研究旨在通过将解的快照移回参考模板来逆转输运效应。该方法采用迭代过程依次提取多个输运结构，然后对这些对齐后的数据应用线性模型降阶。Cagniart、Maday 和 Stamm (2019年)提出了一种数据驱动标定方法，通过训练变换，使变换后的解快照可以进行线性降阶。在线阶段，对变换后的解应用线性模型降阶，并在残差的驱动下推进降阶状态和变换。Greif、Junk 和 Urban (2022年)从线性降阶近似出发，然后添加通过数据学习得到的脊项Reiss, Schulze, Sesterhennand Mehrmann (2018年Black 等(2020年）、Reiss (2021年）、Papapicco、Demo、Girgoflio、Stabile 和 Rozza (2022年）、Burela、Krah 和 Reiss (2025年）、Krah、Marmin、Zorawski、Reiss 和 Schneider (2025年）。类似地，Rim、Moe 和 LeVeque(2018年）的研究旨在通过将解的快照移回参考模板来逆转输运效应。该方法采用迭代过程依次提取多个输运结构，然后对这些对齐后的数据应用线性模型降阶。 Cagniart、Maday 和 Stamm (2019年）提出了一种数据驱动标定方法，通过训练变换，使变换后的解快照可以进行线性降阶。在线阶段，对变换后的解应用线性模型降阶，并在残差的驱动下推进降阶状态和变换。 Greif、Junk 和 Urban(2022年）从线性降阶近似出发，然后添加通过数据学习得到的脊项

，这些脊项能够表征输运特征。Issan 和 Kramer (2023年)提出了一种同样使用移位的非侵入式方法。Rim、Peherstorfer 和 Mandli (2023)提出将线性近似与沿守恒律特征曲线演化的全局输运动力学相结合。尤其是，基于特征线的输运动力学可在在线阶段高效计算。Rim 和 Welper (2026年）提出了这项工作的一个最新推广。Issan and Kramer (2023年）。将线性近似与沿守恒律特征曲线演化的全局输运动力学相结合的方案由 Rim、Peherstorfer 和 Mandli (2023提出。尤其是，基于特征线的输运动力学可在在线阶段高效计算。这项工作的最新推广由Rim 和 Welper (2026年）提出。

5.2.2. 最优传输与度量空间中的表述

另一条关于基于变换的模型降阶的研究路线受到最优输运理论的启发。Iollo 与 Lombardi （ 2014 年）计算快照之间的输运映射，并利用这些映射随时间平流降阶基函数。另一种视角由Ehrlacher 等人（2020）以及Battisti、Blickhan、Enchery、Ehrlacher、Lombardi 与 Mula （2023）所采用，他们在度量空间中、尤其是在 Wasserstein空间中表述模型降阶。这意味着将解表示为测度，并经由例如重心在 Wasserstein几何中进行降阶；重心可粗略地解释为由变换诱导的非线性表示，非常适合具有移动特征的解。将这些思想经由稀疏 Wasserstein 重心推广至高维情形的工作已在Do、Feydy 与 Mula （2025）中提出。其他方法将最优传输的概念引入深度学习方法（Khamlich、Pichi 与 Rozza 2025），并搭建通向基于配准的方法的桥梁（Blickhan 2024）。与之密切相关的是依赖 Radon 累积分布变换的工作（Kolouri、Park 与 Rohde 2016年），并在模型降阶的背景下加以研究，例如由Ren、Wolf 与 Mao（ 2021年）以及Long、Barnett、Jefferson-Loveday、Stabile 与 Icardi （ 2025）；参见Rim（2018年）以了解有关 Radon变换的计算方面。

5.2.3. 基于配准的方法

基于配准的方法直接从数据中计算变换，通常采用与方程无关且纯粹数据驱动的方式，灵感来自图像配准，旨在对齐不同参数值或时间点上的主要特征，使其能够进行线性降阶。这类配准概念已以多种形式应用于模型降阶。在参数化情形中，参数$\boldsymbol{\mu}$ 会导致解流形出现非光滑行为，Welper(2017年，2020）提出，先学习适当的变换以对齐移动特征，再在变换空间中进行插值；在该空间中，对$\boldsymbol{\mu}$ 的依赖更为平滑。值得注意的是，Amsallem 和 Farhat (2008年）早已在模型降阶背景下，使用非线性映射在局部降阶基之间进行Grassmann插值。

该工作 Taddei （2020）提出了一种与方程无关的变分配准流程，该流程已与高效的在线计算相结合，见Ferrero、Taddei 和 Zhang （2022）。Sarna 和 Benner（2021）采用纯粹的数据驱动方法学习变换，并将其与一种

基于回归的方法相结合，以在线计算降阶状态。若干研究改进了变换的构造方式及其在线应用： Grundel 和 Sarna （2024）将变换与自适应约化网格相结合，以高效计算降阶近似。 Sarna、Giesselmann 和 Benner（2024）通过对间断进行单调特征匹配来计算变换，而Torlo（2025）则采用移动网格。配准也可以在时空框架下进行，即像Taddei 和 Zhang（2021）以及Kleikamp、Ohlberger 和 Rave （2022）那样，在组合的时空域上学习映射。其他关于基于配准的方法的工作还包括Nonino、Ballarin、Rozza 和 Maday（2023），其明确针对流固耦合问题；以及Iollo、Labatut、Mounoud 和 Taddei（2026），其研究了有界域上配准的基础。

5.2.4. 拉格朗日表述与其他变换

基于变换的模型降阶也可以在拉格朗日框架下表述，其中动力学由随流动移动的坐标来描述。由Mojgani 和 Balajewicz (2017年) 提出的拉格朗日基方法是这一视角的早期示例。由Lu 和 Tartakovsky(2020、2021年) 提出了一种驱动方法，将动态模态分解扩展到拉格朗日情形。最后，人们还提出了更具问题针对性的变换，以明确识别激波曲线或移动界面：Taddei、Perotto 和 Quarteroni (2015年) 明确追踪激波曲线，以利用分段光滑分量；而Razavi 和 Yano (2025) 和Taddei、Tifouti 和 Barral(2025) 则计算配准映射以对齐激波。Alireza Mirhoseini 和 Zahr (2023) 也对齐主导特征，但这是通过在线优化问题隐式实现的。

5.3. 移位本征正交分解（移位 *POD*）

现在，我们将更详细地讨论基于变换的方法——移位 POD。该方法由Reiss 等人(2018年提出，并在例如Schulze 等人(2019)、Black 等人(2020)、Reiss (2021年)、Papapicco 等人(2022)、Schulze (2023)、Burela 等人(2025)、Krah 等人(2025)等工作中得到了进一步发展。尽管名称中有“移位”，该方法并不局限于移位变换，因此我们将结合一般变换来介绍它。我们将重点讨论移位 POD 的非线性模型降阶要素，并以简短讨论变换的设计作结。

5.3.1. 移位 *POD*：示例引言

为说明移位 POD 的动机，我们在此再次讨论典型的输运主导问题，即一维对流方程；参见第 3.3 节。具体而言，我们考虑如下偏微分方程问题实例：将时间域设为 $\mathcal{T} = (0,\ 1)$，将空间域设为 $\Omega = (0,\ 1)$，

并考虑

$$\begin{aligned} \partial_t q_{\mathcal{Q}}(t,\boldsymbol{x}) &= \partial_{\boldsymbol{x}} q_{\mathcal{Q}}(t,\boldsymbol{x}), \quad (t,\boldsymbol{x}) \in \mathcal{T}\times\Omega, \\ q_{\mathcal{Q}}(t,0) &= q_{\mathcal{Q}}(t,1), \qquad\qquad t \in \mathcal{T}, \\ q_{\mathcal{Q}}(0,\boldsymbol{x}) &= q_{\mathcal{Q},0}(\boldsymbol{x}), \qquad\qquad \boldsymbol{x} \in \Omega, \end{aligned} \tag{5.1}$$

具有周期边界条件，但不依赖于 $\boldsymbol{\mu}$。解可以表示为 $q_{\mathcal{Q}}(t,\boldsymbol{x}) = q_{\mathcal{Q},0}(\boldsymbol{x}-t)$，其中 $(t,\boldsymbol{x}) \in \mathcal{T}\times\Omega$。尽管我们已经看到，对于这一类问题，Kolmogorov n-宽度至多仅按多项式速度衰减，但我们仍明确计算该解的最佳线性逼近。最佳线性基 $\{\varphi_1,\ldots,\varphi_n\}$ 可通过求解以下优化问题来计算

$$\begin{aligned} &\min \tfrac{1}{2}\int_{\mathcal{T}} \|q_{\mathcal{Q}}(t,\cdot) - \sum_{j=1}^{n} \langle q_{\mathcal{Q}}(t,\cdot),\varphi_j\rangle_{\mathcal{P}}\varphi_j\|_{\mathcal{P}}^2 \,\mathrm{d}t, \\ &\text{s.t. } \varphi_i \in \mathcal{Q} \text{ and } \langle\varphi_i,\varphi_j\rangle_{\mathcal{P}} = \delta_{ij} \text{ for } i,j = 1,\ldots,n, \end{aligned} \tag{5.2}$$

其中 δ_{ij} 表示克罗内克δ，而 $\mathcal{P}$ 是一个希尔伯特空间，例如枢轴空间。对于平流方程这一具体例子（5.1），我们选择希尔伯特空间 $\mathcal{Q} = H^1_{\text{per}}(\Omega)$ 和 $\mathcal{P} = L^2(\Omega)$。优化问题（5.2）称为POD最小化问题，参见 Sirovich (1987)、Gubisch 和 Volkwein (2017年)。优化问题（5.2）具有闭式解，因为基函数是该非负、自伴、紧算子的 n 主特征函数

$$\mathcal{R}\colon \mathcal{P} \to \mathcal{P}, \qquad \mathcal{R}\varphi = \int_{\mathcal{T}} \langle q_{\mathcal{Q}}(t,\cdot),\varphi\rangle_{\mathcal{P}} q_{\mathcal{Q}}(t,\cdot)\,\mathrm{d}t,$$

例如参见Gubisch 和 Volkwein (2017年)。对于（5.1）中的平流方程这一特定选择，可以证明（参见Holmes 等（1996）、Black等（2020））， 最优基函数是傅里叶模态，因此并非针对具体的解而设计；具体的解仅取决于初始函数 $q_{\mathcal{Q},0}$。事实上，这一观察解释了输运主导问题中基于POD的线性降阶模型为何会出现振荡行为。

5.3.2. 移位 *POD*：变换模态

POD 优化问题（5.2）给出了给定数据的最佳线性逼近。ReissReiss等人（2018年）提出引入多个移位算子$\mathcal{T}_i$，以表示解中固有的不同波形，从而缓解动机示例中输运主导问题带来的困难。移位算子由额外的参数 γ_i进行参数化，该参数称为路径变量，表示每个模态的平移量；在多维情形下，还表示平移方向。为简化表述，这里仅考虑一维路径变量，即 $\gamma_i \in \mathbb{R}$ 对于 $i = 1, \ldots\ldots, n$ 并参见Black 等人（2020）关于

th向量值设定。接下来，我们寻求如下形式的降阶近似

$$\hat{q}(t,\boldsymbol{x},\boldsymbol{\mu}) = \sum_{i=1}^{n} \hat{q}_i(t,\boldsymbol{\mu})\,[\mathcal{T}_i(\gamma_i(t,\boldsymbol{\mu}))\varphi_i]\,(\boldsymbol{x}), \tag{5.3}$$

其中， $\mathcal{T}_i$ 是作用于基函数 φ_i 的移位算子，由此得到的函数 $[\mathcal{T}_i(\gamma_i(t,\boldsymbol{\mu}))\varphi_i]$ 是变换模态，随后在空间坐标处对其进行求值。如果路径变量 $\gamma_1(t,\boldsymbol{\mu})$、……、 $\gamma_n(t,\boldsymbol{\mu})$并非先验固定的，那么（5.3）就是一种非线性参数化，构成了第4.4节所述非线性模型降阶的第一个要素。用第4.3节的术语来说，在线权重因此由系数 $\hat{q}_1(t,\boldsymbol{\mu})$、……、 $\hat{q}_n(t,\boldsymbol{\mu})$以及路径变量 $\gamma_1(t,\boldsymbol{\mu})$、……、 $\gamma_n(t,\boldsymbol{\mu})$给出。我们强调，该方法并不局限于将 $\mathcal{T}_i$ 作为移位算子。事实上，可以使用算子的任意参数族，这也是我们将 $\mathcal{T}_i$ 称为变换算子的原因。典型要求包括： $\gamma_i \mapsto \mathcal{T}_i(\gamma_i)\varphi_i$ 是连续的，变换算子具有 $\mathcal{Q}$不变性，并且 $\mathcal{T}_i$ 是一致有界的。

为了将 POD 优化问题（5.2）推广到具有变换模态的分解（5.3），需要注意的是，通常无法保证$\{\mathcal{T}_i(\gamma_i)\varphi_i\}_{i=1}^n$ 构成一组正交归一基函数。因此，受 Reiss 等人（2018年）、Schulze等人（2019）以及 Black等人（2020）的启发，提出如下变换 POD 优化问题：

$$\begin{aligned} &\min \tfrac{1}{2}\int_{\mathcal{T}} \|q_{\mathcal{Q}}(t,\cdot) - \sum_{i=1}^{n} \hat{q}_i(t,\boldsymbol{\mu})\,[\mathcal{T}_i(\gamma_i(t,\boldsymbol{\mu}))\varphi_i]\,(\boldsymbol{x})\|_{\mathcal{P}}^2 \,\mathrm{d}t, \\ &\text{s.t. } \varphi_i \in \mathcal{Q} \text{ and } \|\varphi_i\|_{\mathcal{P}} = 1 \text{ for } i,j = 1,\ldots,n, \end{aligned} \tag{5.4}$$

该问题需要针对系数函数 $\hat{q}_i$、路径变量 γ_i和模态 φ_i进行最小化。通过采用合适的泛函分析框架，包括对变换算子 $\mathcal{T}_i$ 的假设，以及对系数函数 $\hat{q}_i$ 和路径变量 γ_i所处的适当空间及其界的假设，可以证明平移 POD 优化问题（5.4）是可解的；参见 Black等人（2020）以及Black、Schulze 和 Unger （2022年）。与 POD 优化问题（5.2）相比，后者可通过将（5.4）中的 $\mathcal{T}_i$ 设为恒等算子而得到，对（5.4）性质的分析更加复杂；参见 Black等人（2022）。一种值得注意的例外情况是：如果强制所有变换算子及其路径变量均相同，即近似（5.3）采取如下形式：

$$\hat{q}(t,\boldsymbol{x},\boldsymbol{\mu}) = \sum_{i=1}^{n} \hat{q}_i(t,\boldsymbol{\mu})\,[\mathcal{T}(\gamma(t,\boldsymbol{\mu}))\varphi_i]\,(\boldsymbol{x}). \tag{5.5}$$

如果我们进一步假设 $\mathcal{T}$ 是等距变换，那么相应的最小化问题可以通过先对快照数据进行相应变换，再对变换后的快照应用标准 POD 来求解。更多详情请参阅Black 等 (2020)；另见（Reiss等 2018， Schulze 等 2019）， 以及（Cagniart等 2019）中关于相关变换及其计算的讨论。具体情形

变换算子由岭函数给出的情形，见 Greif等(2022年)。

5.3.3. 移位 *POD*：优化问题的计算方面

一般来说，（5.4）没有可用的闭式解，必须数值计算近似极小化解。除空间离散化和时间离散外，这还需要数值求解优化问题，例如采用基于梯度的方法，如Black 等人（2022年）。如Schulze （2023）中所讨论的，即使对于中等数量 n ，计算代价也会变得很高，因为自由度随时间自由度与空间自由度的数量而增长。与其求解完整的优化问题，也可以采用变量投影方法，详见Schulze （2023）。在实践中，良好的初始化往往很重要。例如，可通过初始化路径变量，使其适当地对齐快照来获得；另见Cagniart等人（2019）、Mendible等人（2020）、Reiss （2021年）中讨论的技术。给定路径变量的初始化后，可通过应用带有路径变量的变换算子来初始化模态，并将模态推断为变换后快照的前导奇异向量，例如在野地火灾模拟中 Black 等人（2021年*a*）所做的那样。

5.3.4. 移位 *POD*：降阶模型的构造

非线性模型降阶的第二个要素构成降阶动力学。我们首先假设路径变量 γ_i 是固定的来展开讨论；在合适的参数化假设下，这意味着我们可以将它们视为离线权重；参见第4.3节。在这种情况下，变换模态假设（5.3）可以在经典 Galerkin 投影框架中使用，类似于（2.12）。更详细地说，回顾（2.1）中所述的偏微分方程问题以及降阶状态 $\hat{\boldsymbol{q}}(t,\boldsymbol{\mu}) = [\hat{q}_1(t,\boldsymbol{\mu}),\ldots,\hat{q}_n(t,\boldsymbol{\mu})]^\top \in \mathbb{R}^n$。我们现在引入降阶路径向量为

$$\hat{\boldsymbol{\gamma}}(t,\boldsymbol{\mu}) = [\hat{\gamma}_1(t,\boldsymbol{\mu}),\ldots,\hat{\gamma}_n(t,\boldsymbol{\mu})]^\top \in \mathbb{R}^n,$$

我们目前假设该量已知。相应的 Galerkin降阶模型为

$$\boldsymbol{M}_1(\hat{\boldsymbol{\gamma}}(t,\boldsymbol{\mu}))\dot{\hat{\boldsymbol{q}}}(t,\boldsymbol{\mu}) + \boldsymbol{M}_s(\hat{\boldsymbol{\gamma}}(t,\boldsymbol{\mu}))\boldsymbol{D}(\hat{\boldsymbol{q}}(t,\boldsymbol{\mu}))\dot{\hat{\boldsymbol{\gamma}}}(t,\boldsymbol{\mu}) = \hat{\boldsymbol{f}}_1(\hat{\boldsymbol{q}}(t,\boldsymbol{\mu}),\hat{\boldsymbol{\gamma}}(t,\boldsymbol{\mu});\boldsymbol{\mu}) \quad (5.6)$$

with

$$\begin{aligned}
\boldsymbol{M}_1(\hat{\boldsymbol{\gamma}}) &= \left[\langle \mathcal{T}_i(\hat{\gamma}_i)\varphi_i, \mathcal{T}_j(\hat{\gamma}_j)\varphi_j\rangle_{\mathcal{P}}\right]_{i,j=1}^n \in \mathbb{R}^{n\times n},\\
\boldsymbol{M}_s(\hat{\boldsymbol{\gamma}}) &= \left[\langle \mathcal{T}_i(\hat{\gamma}_i)\varphi_i, \mathcal{T}_j'(\hat{\gamma}_j)\varphi_j\rangle_{\mathcal{P}}\right]_{i,j=1}^n \in \mathbb{R}^{n\times n},\\
\boldsymbol{D}(\hat{\boldsymbol{q}}) &= \operatorname{diag}(\hat{q}_1,\ldots,\hat{q}_n) \in \mathbb{R}^{n\times n},\\
\hat{\boldsymbol{f}}_1(\hat{\boldsymbol{q}},\hat{\boldsymbol{\gamma}};\boldsymbol{\mu}) &= \left[\left\langle \mathcal{T}_i(\hat{\gamma}_i)\varphi_i, f\left(\cdot, \sum_{k=1}^n \hat{q}_k \mathcal{T}_k(\hat{\gamma}_k)\varphi_k;\boldsymbol{\mu}\right)\right\rangle_{\mathcal{P}}\right]_{i=1}^n \in \mathbb{R}^n,
\end{aligned}$$

其中 $\mathscr{T}_i'$表示 $\mathscr{T}_i$ 的导数。

如果降阶路径变量未知，即如果它们构成在线权重，则（5.6）构成一个欠定的微分-代数方程；见Mehrmann 与 Unger （2023）的近期综述。为使该欠定方程（5.6）适定，我们必须添加附加方程，例如如下形式

$$\boldsymbol{\psi}(\hat{\boldsymbol{\gamma}}, \dot{\hat{\boldsymbol{\gamma}}}, \hat{\boldsymbol{q}}, \dot{\hat{\boldsymbol{q}}}; \boldsymbol{\mu}) = 0, \tag{5.7}$$

它们被称为相位条件或重构方程；文献中例如参见Beyn 与 Th¨ummler （2004），Ohlberger 与 Rave （2013），Rowley 等人（2003），Rowley 与 Marsden （2000）。先验而言，相位条件（5.7）的最优选择尚不明确，文献中报告了若干实例，例如通过最小化降阶系数的时间变化来选取。请参阅Schulze （2023）的综述。Black 等人（2020）提出通过 Petrov–Galerkin 投影来获得相位条件，其中检验空间选取为

$$\operatorname{span}\{\hat{q}_1 \mathscr{T}_1'(\gamma_1)\varphi_1, \ldots, \hat{q}_n \mathscr{T}_n'(\gamma_n)\varphi_n\}.$$

采用这一特定选择时，相位条件为

$$\boldsymbol{\psi}(\hat{\boldsymbol{\gamma}}, \dot{\hat{\boldsymbol{\gamma}}}, \hat{\boldsymbol{q}}, \dot{\hat{\boldsymbol{q}}}; \boldsymbol{\mu}) = \boldsymbol{D}(\hat{\boldsymbol{q}})^\top \left(\boldsymbol{M}_s(\hat{\boldsymbol{\gamma}})^\top \dot{\hat{\boldsymbol{q}}} + \boldsymbol{M}_2(\hat{\boldsymbol{\gamma}}) \boldsymbol{D}(\hat{\boldsymbol{q}}) \dot{\hat{\boldsymbol{\gamma}}} - \hat{\boldsymbol{f}}_2(\hat{\boldsymbol{q}}, \hat{\boldsymbol{\gamma}}; \boldsymbol{\mu}) \right) \tag{5.8}$$

with

$$\boldsymbol{M}_2(\hat{\boldsymbol{\gamma}}) = \left[\langle \mathscr{T}_i'(\hat{\gamma}_i)\varphi_i, \mathscr{T}_j'(\hat{\gamma}_j)\varphi_j \rangle_{\mathscr{P}} \right]_{i,j=1}^{n} \in \mathbb{R}^{n \times n},$$

$$\hat{\boldsymbol{f}}_2(\hat{\boldsymbol{q}}, \hat{\boldsymbol{\gamma}}; \boldsymbol{\mu}) = \left[\left\langle \mathscr{T}_i'(\hat{\gamma}_i)\varphi_i, f\left(\cdot, \sum_{k=1}^{n} \hat{q}_k \mathscr{T}_k(\hat{\gamma}_k)\varphi_k; \boldsymbol{\mu} \right) \right\rangle_{\mathscr{P}} \right]_{i=1}^{n} \in \mathbb{R}^n.$$

因此，降阶状态 $\hat{\boldsymbol{q}}$ 和降阶路径变量 $\hat{\boldsymbol{\gamma}}$ 的耦合降阶模型为（忽略参数）

$$\begin{bmatrix} \boldsymbol{I}_n & 0 \\ 0 & \boldsymbol{D}(\hat{\boldsymbol{q}})^\top \end{bmatrix} \begin{bmatrix} \boldsymbol{M}_1(\hat{\boldsymbol{\gamma}}) & \boldsymbol{M}_s(\hat{\boldsymbol{\gamma}}) \\ \boldsymbol{M}_s(\hat{\boldsymbol{\gamma}})^\top & \boldsymbol{M}_2(\hat{\boldsymbol{\gamma}}) \end{bmatrix} \begin{bmatrix} \boldsymbol{I}_n & 0 \\ 0 & \boldsymbol{D}(\hat{\boldsymbol{q}}) \end{bmatrix} \begin{bmatrix} \dot{\hat{\boldsymbol{q}}} \\ \dot{\hat{\boldsymbol{\gamma}}} \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{f}}_1(\hat{\boldsymbol{q}}, \hat{\boldsymbol{\gamma}}; \boldsymbol{\mu}) \\ \boldsymbol{D}(\hat{\boldsymbol{q}})^\top \hat{\boldsymbol{f}}_2(\hat{\boldsymbol{q}}, \hat{\boldsymbol{\gamma}}; \boldsymbol{\mu}) \end{bmatrix}. \tag{5.9}$$

可以证明，该降阶模型在某种意义下使残差最小，参见Black 等人(2020年连续最优性这一术语由Carlberg 等人(2017年)、Lee 和 Carlberg(2020年Miller与Miller(1981年在移动有限元方法的背景下提出并使用。我们强调，这是 Dirac–Frenkel 变分原理的一个特例，更一般的情形将在第7节讨论。

通常无法保证（5.9）左端的矩阵是非奇异的，在这种情况下可能需要进行正则化；另请参见即将介绍的第7.5节。事实上，只要降阶状态的某个系数 $\hat{q}_i$ 取值为零，就无法求解相应的路径变量 $\hat{\gamma}_i$。为缓解这一问题，最好对多个降阶系数强制使用相同的路径变量和相同的变换算子。在只有一个变换算子和一个路径变量的情况下，这相当于

表示式（5.5）。一般情形详见Schulze （2023）。当偏微分方程的右端项为线性时，可以利用半群理论（例如Pazy （1983）来构造后验误差估计器（Black等，2020）。我们强调，由此得到的误差估计器具有相当的通用性，也可用于验证物理信息神经网络的预测误差（Raissi、Perdikaris 和 Karniadakis 2019）；参见Hillebrecht 和 Unger （2025*b*，*a*）。

5.3.5. 移位 *POD*：离线–在线分解

现在我们转向非线性模型降阶的第三个要素（参见第 4.4 节），即如何高效计算降阶模型（5.9）。由于采用了非线性试探形式，所有矩阵和右端项在形式上都依赖于全模型维数，因此可能难以预先计算。一个值得注意的例外是，在（5.5）中使用单个变换算子和单个路径变量，且变换算子为酉算子时；对于具有周期边界条件的一维移位算子，这一条件成立。在这种情况下，大多数内积都与全模型维数无关，可以预先计算并高效求值；参见 Schulze (2023，定理 4.3.1）。在非常特殊的情况下，这对一般试探形式（5.3）也成立，正如 Black、Schulze 和 Unger (2021*b*) 针对一维波动方程所展示的那样。然而，一般而言情况并非如此，数值计算（5.9）的成本会随完整模型的维数增长。

为缓解这一问题，Black 等人（2021a）提出了不同策略，将（5.9）左端矩阵的内积与对应于可能非线性的全模型右端项的内积求值加以区分。由于（5.9）左端的矩阵仅依赖于 $\hat{\boldsymbol{\gamma}}$，Black 等人（2021a）建议针对 $\hat{\boldsymbol{\gamma}}$ 的典型取值对这些矩阵进行采样，再通过对采样矩阵进行插值来组装这些矩阵。若全模型右端项 f 为线性，也可采用类似策略；详情参见 Cagniart 等人（2019）和 Schulze（2023）。利用变换算子的性质，并采用 Constantine（2015年）提出的主动子空间方法，可用于降低待采样空间的维数。若 f 为非线性，则可应用第 2.4.3 节所讨论的经验插值的一种扩展，以计入变换算子。相关思想例如见于 Black 等人（2021a）、Rim 等人（2023）、Grundel 与 Sarna（2024）。通过采用具有自适应步长的时间积分器，还可获得进一步加速的机会，因为特殊试探形式（5.3）与线性降阶模型相比通常允许较大的时间步。

另一种对降阶模型（5.9）进行数值积分的方法，是通过机器学习来学习降阶系数和降阶路径。对于变换模态，Papapicco 等人(2022)以及Burela 等人(2025)采用了这种方法。

5.3.6. 移位 *POD*：变换算子

尚待回答的问题是如何设计变换算子 $\mathcal{T}_i$。对于具有周期性边界条件的一维情形，简单的移位算子或许就足够了。然而，对于非周期性边界条件，或在二维及更高维计算域中更复杂的波传播问题，情况则有所不同。非周期性边界条件的情形见Schulze (2023，第3.3节）；例如，可以采用常数外推来虚拟地扩展计算域，或者使用零填充。通常，理想的做法是依据具体应用领域的专家知识来构造变换算子，例如研究相应问题所基于的半群的性质。特别是，半群方法建议构造相对于完整模型右端项 f等变的变换算子。我们强调，融入此类专家知识，可能是确保在远离训练数据的区域仍能获得高保真近似的关键。如果没有此类专家知识，则可以转而学习变换算子，例如Krah,B¨uchholz,H¨aringer 和 Reiss (2023)所展示的方法。

6. 在线自适应模型降阶

本节综述在在线阶段随时间调整降阶空间的模型降阶方法。其背后的动机是，即使在单一全局降阶空间中的近似效率不高，输运主导问题的解通常也能在局部时间范围内得到很好的近似。

6.1. 采用在线基更新的模型降阶

我们考虑在线自适应模型降阶方法，这些方法将降阶解表示为随时间 t（也可能随参数 $\boldsymbol{\mu}$）演化的基函数的线性组合。相应地，由这些基函数张成的降阶空间也随时间演化。

实现时变基函数的一种常见方式是采用非线性参数化，使基依赖于在线权重中包含的时变特征 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$，如 (4.10) 所示。此时，在线阶段还包括对权重向量 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$ 的调整步骤，以随时间演化这些特征（从而演化基）。在半离散情形下，时变降阶空间$\mathcal{V}(t)$ 由时变基矩阵 $\boldsymbol{V}(t) \in \mathbb{R}^{N\times n}$表示，该基矩阵也可能依赖于参数 $\boldsymbol{\mu}$，并在在线阶段进行更新。值得注意的是，某些基于变换的方法也可以解释为时变基方法，例如，其中的移位随时间演化基。

能够克服Kolmogorov屏障的时变基方法所依据的原理是局部时间近似：在较短的时间窗内，全模型解可以在低维向量空间中得到良好近似，

但随着解的演化，该空间需要随时间进行调整。与线性模型降阶中使用一个随时间固定的单一降阶空间不同，在线自适应降阶空间可以近似具有移动相干结构和输运的解，而无需高维的在线权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$。

在线自适应的一个后果是，线性模型降阶中的离线–在线分解被放弃，或至少有所弱化。具体而言，在在线模拟降阶模型时，基矩阵 $\boldsymbol{V}(t)$（或等价地说，相应的基函数）会进行自适应更新；这些基更新可能产生与全模型解空间维数 N 成比例的成本，从而限制加速效果。因此，一个核心挑战是设计能够限制对 N的依赖的基更新机制，例如采用稀疏基更新。与此同时，离线–在线分解的弱化使降阶空间能够在线自适应，并捕捉离线阶段所用训练数据中未曾出现的动力学和解的特征。当参数 $\boldsymbol{\mu}$ 具有高维度，且相应的参数域 $\mathcal{D}$ 无法进行穷尽采样时，或者当完整模型的模拟成本极高、无法生成大量训练数据轨迹时，这一点尤为重要。我们关注通过在线自适应模型降阶克服 Kolmogorov 屏障，而非其他优势，例如避免全面的离线训练阶段；关于后者，请参阅 Peherstorfer 和 Willcox (2015年*b*) 以及注释 6.1。

注释6.1。 关于在外循环迭代（例如优化）过程中调整降阶空间，已有一系列研究，例如 Arian、Fahl 和 Sachs (2000), Kunisch、Karl 和 Volkwein, Stefan (2008年), Zahr 和 Farhat (2015 年), Amsallem 、 Zahr 和 Washabaugh (2015 年), Qian、Grepl、Veroy 和 Willcox(2017年)。逐步构建降阶空间，通常是因为无法对参数域进行穷尽采样，因此会在外循环迭代过程中构建降阶模型。与此不同的是，本综述聚焦于随时间进行的自适应，以缓解 Kolmogorov 屏障。

6.2. 在线自适应模型降阶与时变基方法综述

下面将在线自适应模型降阶置于更广泛的时变基方法背景下加以讨论。时变基方法是指近似空间随时间演化，而非像线性模型降阶那样保持不变的方法。

6.2.1. 局部化模型降阶

局部化模型降阶方法依赖多个低维空间，并通过选择或加权机制在线使用这些空间。例如，一种常见方法是离线预先计算多个降阶空间，然后在线根据指示函数选择其中一个；该指示函数可基于当前降阶解、参数或时间等信息。根据在线信息选择降阶空间，会产生非线性

参数化。关键在于，局部降阶空间的划分方式应使每个局部降阶空间都对应一个局部状态区域，且该区域可由低维降阶空间良好近似；即使维数相同的单个全局降阶空间会导致近似不准确，也仍然如此。

寻找局部状态区域分解的一种方法，是针对参数域 $\mathcal{D}$ 中的不同区域定制局部降阶空间（Eftang、Patera 和 Rønquist 2010，Eftang、Knezevic 和 Patera 2011，Haasdonk、Dihlmann 和 Ohlberger 2011，Eftang 和 Stamm 2012，Wieland 2015，Bonito、Cohen、DeVore、Guignard、Jantsch 和 Petrova 2021）。另一种方法是分解时间区间，并为每个时间窗使用不同的局部降阶空间（Dihlmann、Drohmann 和 Haasdonk 2011，Drohmann、Haasdonk 和 Ohlberger 2011，Shimizu 和 Parish 2021，Parish 和 Carlberg 2021，Copeland、Cheung、Huynh 和 Choi 2022）；这可以解释为采用切换系统或混合系统进行近似。也可以直接基于状态空间（在半离散情形下）对降阶空间进行局部化（Washabaugh、Amsallem、Zahr 和 Farhat 2012，Amsallem、Zahr 和 Farhat 2012，Peherstorfer、Butnaru、Willcox 和 Bungartz 2014，Amsallem 等 2015，Grimberg、Farhat、Tezaur 和 Bou-Mosleh 2021，Hess、Alla、Quaini、Rozza 和 Gunzburger 2019，Amsallem 和 Haasdonk 2016，Geelen 和 Willcox，2022年）。一种实现方式是聚类：在离线阶段对快照应用聚类方法，将相似的快照归入同一簇，并为每个快照簇构造一个局部降阶空间。局部化模型降阶也可以理解为产生非线性参数化，这种参数化由多个降阶空间的并集诱导而成，正如 Cohen、Dahmen、Mula 和 Nichols（2022a）所述。受区域分解和多尺度方法的启发，Ohlberger 和 Schindler（2015，2017）、Buhr 和 Smetana（2018）、Smetana 和 Taddei（2023）在空间域上进行局部化，并为每个子域计算局部降阶空间；这种方法专为具有多尺度结构的问题而设计。这些工作的一个核心进展是严格的误差控制，并以此实现自适应在线增广（Ohlberger 和 Schindler 2017）。如果能够以这样的方式划分区域，使得每个子域中都可能出现几乎相同的局部特征，那么 Van Gastelen、Edeling 和 Sanderse（2026）建议使用单个局部基。局部化还可用于支持在线阶段的局部修改，从而使更新量和重新计算保持局部化且轻量（Phuong Huynh、Knezevic 和 Patera 2013，Smetana 和 Patera 2016，Buhr、Engwer、Ohlberger 和 Rave 2017）。

字典方法将通过离线预先计算大量候选基函数，并在线从中选出少数候选基函数来张成降阶空间的思路进一步拓展（Kaulmann 和 Haasdonk 2013，Abgrall、Amsallem 和 Crisovan 2016，Daniel、Casenave、Akkari 和 Ryckelynck 2020，Balabanov 和 Nouy 2021，Nouy 和 Pasco 2024，Herkert、Buchfink 和 Haasdonk 2024）。

6.2.2. 动态低秩近似及相关方法

Koch 和 Lubich （2007）提出了动态低秩近似，通过低秩矩阵对解进行参数化，并推导出低秩矩阵表示的各因子随时间演化的降阶动力学。动态低秩近似降阶动力学的核心原理是 Dirac–Frenkel变分原理（Dirac 1930年, Frenkel 1934年），该原理将完整动力学投影到固定秩矩阵流形的切空间上；我们将在第7节重新讨论适用于一般非线性参数化的 Dirac–Frenkel变分原理。

动态低秩近似面临的一项挑战是，由于低秩矩阵表示中的奇异值在数值上很小，低秩因子的降阶动力学可能会变得病态。因此，使用标准积分器可能会受到严格的步长限制。相应地，人们开发了专用于动态低秩近似的时间积分格式。Lubich 和 Oseledets (2014年) 提出了投影分裂积分器，Kieri、Lubich 和 Walach (2016年)，这类积分器对低秩矩阵表示中的小奇异值具有鲁棒性。另一种选择是名为基更新与Galerkin（BUG）的积分器，由 Ceruti 和 Lubich (2022年)、Ceruti、Einkemmer、Kusch 和 Lubich (2024*a*) 提出，其计算效率可能更高。总的来说，如何对动态低秩近似对应的降阶动力学进行高效、稳定且鲁棒的时间积分，仍是一个活跃的研究方向，例如（Ceruti、Kusch 和 Lubich 2022、Hauck 和 Schnake 2023、Hochbruck、Neher 和 Schrammer 2023、Ceruti、Kusch 和 Lubich 2024*b*、Lam、Ceruti 和 Kressner 2025、Nobile 和 Trindade 2025*b*、Bachmayr、Eisenmann 和 Uschmajew 2025*b*、Carrel（2026））。

另一个活跃的研究方向是结构保持。例如，针对哈密顿波动方程，一种辛动态低秩逼近方法由Musharbash、Nobile和Vidli˘ckov´a （2020）提出。对于动理学问题，尤其是Vlasov型方程，已有一系列在动态低秩近似中强制守恒律的工作（Einkemmer、Hu和Wang 2021年，Coughlin、Hu和Shumlak 2024，Einkemmer、Ostermann和Scalone 2023，Einkemmer和Joseph 2021年，Einkemmer和Lubich 2019，Hesthaven、Pagliantini和Ripamonti 2024）。

动态低秩近似向张量分解的推广亦由Koch和Lubich（2010）、Lubich、Rohwedder、Schneider和Vandereycken（2013）、Arnold和Jahnke（2014年）、Lubich、Oseledets和Vandereycken（2015年）完成。此外，还有一些用于演化基的方法，这些基专门仅表示全模型动力学的部分内容，例如不稳定方向（Babaee和Sapsis 2016年）。

动态低秩近似与随机（偏）微分方程的动态正交分解有关，该分解由 Sapsis 和 Lermusiaux（2009）提出。关于动态正交近似的误差分析（Musharbash、Nobile 和 Zhou 2015年、Musharbash 和 Nobile 2018年、Kazashi、Nobile 和 Zoccolan 2025年）以及时间

积分（Feppon 和Lermusiaux 2018年、Charous 和 Lermusiaux 2024年、2023）。此外，针对随机偏微分方程情形，也有对Petrov–Galerkin格式的推广（Nobile 和 Trindade 2025年*a*、*b*）。

这里只能简要综述动态低秩逼近方法及其相关方法。关于更多详情，请参阅 Einkemmer、Kormann、Kusch、McClarren 和 Qiu (2025年) 撰写的专题综述。

6.2.3. 模型降阶中的时间自适应基

现在我们将重点介绍专为模型降阶开发的时间自适应基方法。关于时变问题模型降阶的综述，请参阅 Hesthaven等人(2022年*b*)。

一种在线自适应降阶空间的研究思路，是通过模拟输运动力学来推导基演化。例如，Iollo 和 Lombardi (2014年) 借鉴最优传输的思想，使张成降阶空间的基函数随时间演化。类似地，Gerbeau 和 Lombardi (2014年) 基于近似的 Lax 对演化约化基。沿着这一思路，Rim等人(2023) 将降阶空间沿特征线输运。另一种方法见于 Black 等人 (2020)，该方法通过瞬时残差最小化使约化基的模态随时间演化；详细讨论请见第 5 节。

Rap´un 和 Vega (2010) 通过偶尔切换回完整模型以生成新的快照，从而更新降阶模型。Carlberg(2015年) 和 Etter 和 Carlberg(2020) 提出的方法旨在通过拆分降阶基向量来丰富降阶空间，从而模拟有限元中的自适应网格细化。其重点是实现故障安全细化，即在极限情况下恢复全模型空间。

另一类研究通过纳入新获得的在线信息（例如全阶模型残差和不完整状态场），推导出对降阶空间的低秩更新量，从而实现数据驱动的在线基更新。例如，对于线性问题，Peherstorfer 和 Willcox (2015年*a*) 提出的动态数据驱动降阶模型采用增量奇异值分解方法调整降阶空间，并通过相应的低秩更新量调整降阶模型算子，无需从头重建降阶模型。

对于非线性问题，Peherstorfer 和 Willcox (2015年*b*提出了在线自适应离散经验插值法（ADEIM）。该方法以经验插值法为基础（Barrault 等人 2004，Chaturantabut 和 Sorensen 2010），并根据稀疏残差样本对降阶空间进行低秩更新。基更新被表述为通过采样算子从一个小型（可能是随机的）草图中推断出的低秩修正。利用稀疏残差样本有助于降低在线计算成本。更新量的稀疏性还有助于缓解降阶非线性系统时的提升瓶颈（见第2.4.2节）。ADEIM 的一个核心环节是如何随时间选择并调整采样点。Peherstorfer (2020)将在线基更新与自适应采样相结合，研究表明这对输运主导问题至关重要。稀疏采样引出了

不仅要考虑在哪里采样，还要考虑何时采样：对当前状态对应的残差信息进行采样可能会滞后。前瞻策略利用短期预测规则来评估残差和能够预判即将发生的动力学的预测状态，从而提高在线更新的鲁棒性并减少不稳定性；参见Singh、Uy 和 Peherstorfer (2023)，Zucatti 和 Zahr(2024)。

有多种方法可根据稀疏采样计算低秩更新。原始 ADEIM 工作（Peherstorfer 和 Willcox 2015年*b*）中提出的基更新并不能保证自适应基是正交的。Zimmermann、Peherstorfer 和 Willcox （2018年）借鉴信号处理中的子空间跟踪方法（Balzano、Nowak 和 Recht 2010年），推导出一种高效的 ADEIM 秩一更新规则，可随时间保持自适应基的正交归一性。Huang 和 Duraisamy （2023）、Mohaghegh 和 Huang （2026）提出了一种相关的在线自适应策略，专门针对最小二乘公式，而这些公式对于混沌问题与多尺度问题的降阶至关重要。

在模型降阶中，也存在并不借助在线稀疏残差信息来更新基的时变基方法。例如，Ramezanian、Nouri 和 Babaee（2021年）、Naderi 和 Babaee （2023）、Aitzhan、Nouri、Givi 和 Babaee （2025）直接从控制方程推导并积分低秩表示的演化方程。从这个意义上说，这些方法与动态低秩近似具有相似之处。

此外，还有其他受动态低秩近似启发的在线自适应降阶基方法，例如 Hesthaven 和 Pagliantini(2021年)、Pagliantini(2021年)、Hesthaven、Pagliantini 和 Ripamonti (2022年*a*)，这些方法具有结构保持特性。特别是，相关研究（Pagliantini 和 Vismara 2025，2026）展示了如何在通过经验插值等稀疏采样型降阶策略调整基时仍保持结构。

在线自适应降阶空间的另一种方法，是在离线阶段根据训练轨迹预先计算随时间变化的降阶空间，然后在线使用该空间，而不再对降阶空间进行进一步调整或增广（Billaud-Friess 和 Nouy 2017年）。

6.3. 自适应离散经验插值法 *(ADEIM)*

现在，我们以 ADEIM 为例，讨论如何通过在线更新实现在线自适应模型降阶。基于 ADEIM 的降阶模型利用稀疏残差信息进行低秩更新，从而在线调整降阶空间。本节的介绍大致遵循 Peherstorfer 和 Willcox (2015年*b*) 和Peherstorfer(2020)。为便于阐述，本节中省略状态对 $\boldsymbol{\mu}$ 的依赖。

6.3.1. 时间离散全模型
我们考虑如下形式的全离散全模型

$$\boldsymbol{q}_{k-1} = \boldsymbol{f}^{(\Delta)}(\boldsymbol{q}_k; \boldsymbol{\mu}), \qquad k = 1, \ldots, K, \tag{6.1}$$

它可由半离散全模型（2.4）通过例如隐式时间积分格式得到。特别地，$\boldsymbol{q}_k$ 是对 $\boldsymbol{q}(t_k)$ 在某个 $t_k \in [0, T]$下的近似。形式（6.1）通常需要一次非线性求解才能得到 $\boldsymbol{q}_k$，其来自 $\boldsymbol{q}_{k-1}$。注意，右端函数 $\boldsymbol{f}$（半离散全模型（2.4）的）与右端函数 $\boldsymbol{f}^{(\Delta)}$（全离散全模型（6.1）的）并不相同。

按照第2.3节中的步骤，可由离散完整模型得到如下离散降阶模型

$$\hat{\boldsymbol{q}}_{k-1} = \hat{\boldsymbol{f}}^{(\Delta)}(\hat{\boldsymbol{q}}_k; \boldsymbol{\mu}), \qquad k = 1, \ldots, K, \tag{6.2}$$

其中，右端函数$\hat{\boldsymbol{f}}^{(\Delta)}$ 由全模型右端函数 $\boldsymbol{f}^{(\Delta)}$ 经经验插值得到；见第2.4节。特别地，映射$\hat{\boldsymbol{f}}^{(\Delta)}$依赖于基矩阵 $\boldsymbol{V} = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_n] \in \mathbb{R}^{N \times n}$ 以及插值点矩阵 $\boldsymbol{P}$，后者对应于两两互异的插值点 $p_1, \ldots, p_n \in \{1, \ldots, N\}$。

在本节讨论 ADEIM 以及在线自适应降阶建模时，我们将降阶模型（6.2）称为静态降阶模型，因为基矩阵 $\boldsymbol{V}$ 和插值点 $\boldsymbol{P}$ 在所有时间步 $k = 1, \ldots, K$上均保持固定。

6.3.2. ADEIM：基与点自适应
现在我们讨论 ADEIM 对基和插值点的更新机制。

基更新和非线性参数化 现在令基矩阵 $\boldsymbol{V}_k$ 依赖于时间步 $k = 1, \ldots, K$。ADEIM 将基矩阵 $\boldsymbol{V}_{k-1}$ 自适应调整为 $\boldsymbol{V}_k$，即

$$\boldsymbol{V}_k = \boldsymbol{V}_{k-1} + \boldsymbol{\alpha}_{k-1}\boldsymbol{\beta}_{k-1}^{\top}, \tag{6.3}$$

其中 $\boldsymbol{\alpha}_{k-1} \in \mathbb{R}^{N \times r}$ 和 $\boldsymbol{\beta}_{k-1} \in \mathbb{R}^{n \times r}$ 提供秩为 $r \in \mathbb{N}$ 的更新。通常，秩较小，$r \ll n$，甚至只是秩一更新，$r = 1$。除了基矩阵，ADEIM 还会自适应调整插值点，使矩阵 $\boldsymbol{P}_{k-1}$ 适应于 $\boldsymbol{P}_k$。

随时间调整基矩阵 $\boldsymbol{V}_k$（以及插值点矩阵 $\boldsymbol{P}_k$），可以理解为降阶状态表示的演化。按照非线性参数化的术语，这意味着时间步 k 的在线权重包含降阶状态（例如 $\hat{\boldsymbol{q}}_k$）以及低秩更新变量（例如 $\boldsymbol{\alpha}_{k-1}$ 和 $\boldsymbol{\beta}_{k-1}$）。更一般地，在线权重还可以包含与插值点更新相关的变量，以及更新所需的其他辅助变量，例如采样点和窗口数据。离线权重由离线阶段固定的量组成，例如初始降阶基矩阵 $\boldsymbol{V}_0$ 和初始插值点矩阵 $\boldsymbol{P}_0$。从这

从这一角度来看，ADEIM 可以理解为实现了前文（4.10）所讨论的时间自适应非线性参数化的一个全离散实例：降阶解仍以低维形式表示，但其表示会通过随时间变化的基更新在线调整。

基更新 我们首先考虑基自适应。为便于说明，我们在每个时间步 $k = 1, \ldots, K$ 都对基矩阵进行自适应；然而在实际中，为降低在线计算成本，通常只在选定的自适应步（例如每隔 z 个时间步）更新基。在时间步 k，给定前一时间步 $k-1$ 的基矩阵 $\boldsymbol{V}_{k-1}$，ADEIM 将其自适应为基矩阵 $\boldsymbol{V}_k$。

更新量的构造方式使得由更新后的基矩阵 $\boldsymbol{V}_k$张成的空间能够很好地近似滑动窗口中的向量

$$\boldsymbol{F}_{k-1} = [\mathsf{f}_{k-w}, \ldots, \mathsf{f}_{k-1}] \tag{6.4}$$

长度为 $w \in \mathbb{N}$。我们或许想使用全模型解作为滑动窗口 $\boldsymbol{F}_{k-1}$中的向量，但这些解无法获得。因此，ADEIM 使用替代量$\mathsf{f}_{k-w}, \ldots, \mathsf{f}_{k-1}$，这些替代量能够捕捉全模型解中的部分信息，同时无需承担对完整模型进行时间步进的高昂计算成本即可计算得到。我们将在第6.3.4节讨论如何获取滑动窗口 $\boldsymbol{F}_{k-1}$ 中的向量。现在，先假设滑动窗口 $\boldsymbol{F}_{k-1}$ 已知。

我们正式将目标函数 J_{nl} 定义为：在自适应基矩阵 $\boldsymbol{V}_{k-1} + \boldsymbol{\alpha}\boldsymbol{\beta}^\top$ 张成的更新空间中，使用插值点矩阵 $\boldsymbol{P}$近似 $\boldsymbol{F}_{k-1}$ 的各列时残差的范数，即

$$J_{\mathrm{nl}}(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{P}) = \|(\boldsymbol{V}_{k-1} + \boldsymbol{\alpha}\boldsymbol{\beta}^\top)(\boldsymbol{P}^\top(\boldsymbol{V}_{k-1} + \boldsymbol{\alpha}\boldsymbol{\beta}^\top))^{-1}\boldsymbol{P}^\top\boldsymbol{F}_{k-1} - \boldsymbol{F}_{k-1}\|_F\,. \tag{6.5}$$

使 J_{nl} 滑动窗口中的向量在自适应空间中近似误差最小的低秩更新和插值点矩阵（6.3）。请注意，满足条件的插值点矩阵还需符合一些额外条件，而这些条件并未包含在目标函数（6.5）中。目标函数 J_{nl} 关于更新量 $\boldsymbol{\alpha}\boldsymbol{\beta}^\top$ 和插值点矩阵 $\boldsymbol{P}$是非线性的。在线阶段求解以目标函数 J_{nl} 为目标的非线性优化问题，计算成本可能很高。

ADEIM 将目标函数 J_{nl} 简化为

$$J(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \| \left(\boldsymbol{V}_{k-1} + \boldsymbol{\alpha}\boldsymbol{\beta}^\top\right) \boldsymbol{C}_{k-1} - \boldsymbol{F}_{k-1}\|_F\,, \tag{6.6}$$

其中，来自时间步 $k-1$的系数矩阵为，

$$\boldsymbol{C}_{k-1} = (\boldsymbol{P}_{k-1}^\top\boldsymbol{V}_{k-1})^{-1}\boldsymbol{P}_{k-1}^\top\boldsymbol{F}_{k-1}\,, \tag{6.7}$$

被用来替代（6.5）中依赖于调整后量的系数。相应地，目标函数 J 与调整后的插值点矩阵无关。为进一步降低在线更新的成本，残差仅

在稀疏采样点处最小化，

$$\min_{\substack{\boldsymbol{\alpha}\in\mathbb{R}^{N\times r}\\ \boldsymbol{\beta}\in\mathbb{R}^{n\times r}}} \|\boldsymbol{S}_{k-1}^{\top}\left(\left(\boldsymbol{V}_{k-1}+\boldsymbol{\alpha}\boldsymbol{\beta}^{\top}\right)\boldsymbol{C}_{k-1}-\boldsymbol{F}_{k-1}\right)\|_F \tag{6.8}$$

其中 $\boldsymbol{S}_{k-1}=[\boldsymbol{e}_{s_1^{(k-1)}},\ldots,\boldsymbol{e}_{s_{n_s}^{(k-1)}}]$ 是与两两不同的采样点 $s_1^{(k-1)},\ldots,s_{n_s}^{(k-1)}$相对应的采样矩阵。通常，$n\le n_s\ll N$。注意，问题（6.8）仅在以下位置指定 $\boldsymbol{\alpha}_{k-1}$：$n_s$ 采样点 $s_1^{(k-1)},\ldots,s_{n_s}^{(k-1)}$。

Peherstorfer 和 Willcox （2015年*b*）基于特征值分解计算求解（6.8）的更新量；然而，同样为了降低在线更新的成本，文献（Peherstorfer 2020，算法 3 ）表明，可以通过对 $n_s\times w$ 矩阵 $\boldsymbol{S}_{k-1}^{\top}\boldsymbol{R}_{k-1}$ 进行薄奇异值分解，高效地计算出近似；该矩阵是残差矩阵 $\boldsymbol{R}_{k-1}=\boldsymbol{V}_{k-1}\boldsymbol{C}_{k-1}-\boldsymbol{F}_{k-1}$ 在与采样点 $s_1^{(k-1)},\ldots,s_{n_s}^{(k-1)}$相对应的分量上的限制：令$\boldsymbol{U}_{1:r}\in\mathbb{R}^{n_s\times r}$ 和 $\boldsymbol{W}_{1:r}\in\mathbb{R}^{w\times r}$ 分别包含 r 个对应于奇异值$\boldsymbol{S}_{k-1}^{\top}\boldsymbol{R}_{k-1}$的前几个左奇异向量和右奇异向量 $\sigma_1,\ldots,\sigma_r$。将前 r 个奇异值 $\sigma_1,\ldots,\sigma_r$ 收集到 $\boldsymbol{\Sigma}_{1:r}$的对角线上，则 ADEIM 更新量为

$$\boldsymbol{\alpha}_{k-1}=-\boldsymbol{S}_{k-1}\boldsymbol{U}_{1:r}\boldsymbol{\Sigma}_{1:r}\in\mathbb{R}^{N\times r},\qquad \boldsymbol{\beta}_{k-1}=(\boldsymbol{C}_{k-1}^{+})^{\top}\boldsymbol{W}_{1:r}\in\mathbb{R}^{n\times r}, \tag{6.9}$$

这会改变基矩阵 $\boldsymbol{V}_{k-1}$，但仅限于采样点 $s_1^{(k-1)}$，……，$s_{n_s}^{(k-1)}$按（6.3）应用时。这与定义更新量 $\boldsymbol{\alpha}_{k-1}$ 仅在采样点 $s_1^{(k-1)}$，……，$s_{n_s}^{(k-1)}$对应于 $\boldsymbol{S}_{k-1}$的目标函数（6.8）一致。不过请注意，随后对 $\boldsymbol{V}_k$ 进行标准正交化通常会改变所有行。求解该问题时计算奇异值分解的成本随 $\mathcal{O}(n_s w^2$变化，常见情况下 $w\le n_s$。

更新插值点将基矩阵从$\boldsymbol{V}_{k-1}$ 更新为 $\boldsymbol{V}_k$后，插值点矩阵 $\boldsymbol{P}_{k-1}$ 更新为 $\boldsymbol{P}_k$。Peherstorfer 和 Willcox(2015年*b*提出了一种算法，可对$\boldsymbol{P}_{k-1}$进行秩一更新，这意味着每次基更新最多更新一个点。相比之下，Peherstorfer (2020) 从头开始使用 QDEIM 算法（Drmač 和 Gugercin 2016年）重新计算插值点。从头开始计算这些点的成本随 $\mathcal{O}(Nn^2$变化，因此与全模型维数呈线性关系 N。

6.3.3. ADEIM：降阶模型的前向步

在时间步 k，基矩阵更新为 $\boldsymbol{V}_k$ 且插值点矩阵更新为 $\boldsymbol{P}_k$后，相应的降阶模型右端函数 $\hat{\boldsymbol{f}}_k^{(\Delta)}$ 通过经验插值从全模型右端函数 $\boldsymbol{f}^{(\Delta)}$ 推导得到。请注意，降阶右端函数 $\hat{\boldsymbol{f}}_k^{(\Delta)}$ 取决于 k，因为基矩阵和插值点均取决于 k。随后，将降阶模型从 $k-1$ 向前积分一个时间步至 k，以计算时间步 k的降阶解 $\hat{\boldsymbol{q}}_k$，这意味着需要求解一般为非线性的方程组 $\hat{\hat{\boldsymbol{q}}}_{k-1}=\hat{\boldsymbol{f}}_k^{(\Delta)}(\hat{\boldsymbol{q}}_k;\boldsymbol{\mu})$，其中 $\hat{\hat{\boldsymbol{q}}}_{k-1}$ 是 $\boldsymbol{V}_{k-1}\hat{\boldsymbol{q}}_{k-1}$ 在自适应基 $\boldsymbol{V}_k$ 上的投影。

6.3.4. ADEIM：滑动窗口

请记住，滑动窗口的各列（见（6.4））充当不可用的全模型解的替代量。ADEIM利用了这样一个事实：若将时间步 $\boldsymbol{q}_k$ 的全模型解代入全离散全模型（6.1），则可得到 $\boldsymbol{q}_{k-1}$。

在 ADEIM 中，在时间步 k，滑动窗口从 $\boldsymbol{F}_{k-1}$ 更新为 $\boldsymbol{F}_k$。计算出降阶解 $\hat{}\boldsymbol{q}_k$ 及其自适应基 $\boldsymbol{V}_k$（参见第 6.3.3 节）后，利用它推导

$$\mathfrak{f}_k = \boldsymbol{f}^{(\Delta)}(\boldsymbol{V}_k\hat{\boldsymbol{q}}_k;\boldsymbol{\mu}), \tag{6.10}$$

它是前一时间步的全模型解 $\boldsymbol{q}_{k-1}$ 的近似$k-1$。随后，通过移除最旧的列并添加 $\mathfrak{f}_k$，将窗口从 $\boldsymbol{F}_{k-1}$ 更新为 $\boldsymbol{F}_k$。接着，在下一个时间步 $k+1$，使用窗口 $\boldsymbol{F}_k$ 驱动基从 $\boldsymbol{V}_k$ 更新为 $\boldsymbol{V}_{k+1}$。

代理量 $\mathfrak{f}_k$ 通常包含由 $\boldsymbol{V}_k$ 的列张成的降阶空间之外的分量，而提升后的降阶解 $\boldsymbol{V}_k\hat{}\boldsymbol{q}_k$ 按构造属于 $\boldsymbol{V}_k$ 的值域。因此，在滑动窗口中使用形式为（6.10）的替代量，可以揭示当前基中缺失的方向，并促成有意义的基更新（6.9）。从这个意义上说，$\mathfrak{f}_k$ 可视为残差信息：它在当前降阶解处探测全模型右端项，并捕捉当前基无法表示的分量。

现在回想一下，ADEIM 仅在 6.8 中的残差在 $n_s < N$ 个对应于 $\boldsymbol{S}_{k-1}$ 的采样点处最小化，而并非在所有 N 个分量处最小化。这意味着，只需在对应于采样点的分量处计算全模型右端函数，我们将这些分量记为

$$\boldsymbol{S}_{k-1}^{\top}\mathfrak{f}_k = \boldsymbol{S}_{k-1}^{\top}\boldsymbol{f}^{(\Delta)}(\boldsymbol{V}_k\hat{\boldsymbol{q}}_k;\boldsymbol{\mu}). \tag{6.11}$$

未由 $\boldsymbol{S}_{k-1}$ 编入索引的分量可以通过经验回归近似为

$$\check{\boldsymbol{S}}_{k-1}^{\top}\mathfrak{f}_k = \check{\boldsymbol{S}}_{k-1}^{\top}\boldsymbol{V}_k(\boldsymbol{S}_{k-1}^{\top}\boldsymbol{V}_k)^{+}\boldsymbol{S}_{k-1}^{\top}\boldsymbol{f}^{(\Delta)}(\boldsymbol{V}_k\hat{\boldsymbol{q}}_k;\boldsymbol{\mu}), \tag{6.12}$$

˘其中 $\boldsymbol{S}_{k-1}$ 表示与补集 $\{1,\dots,N\}\setminus\{s_1^{(k-1)},\dots,s_{n_s}^{(k-1)}\}$相关联的采样点矩阵。注意，我们使用伪逆 $(\boldsymbol{S}_{k-1}^{\top}\boldsymbol{V}_k)^{+}$ 进行最小二乘回归，以近似未观测到的分量，而非通过插值；关于经验插值中的过采样，请参见第 2.4.3 节。通过这一过程，ADEIM 仅在完整模型的 $\boldsymbol{f}^{(\Delta)}$ 个 n_s 采样点处计算右端函数。

计算 $\mathfrak{f}_k$ 只需在提升后的降阶解 $\boldsymbol{f}^{(\Delta)}$ 处计算时间离散全模型右端项函数 $\boldsymbol{V}_k\hat{}\boldsymbol{q}_k$，而不必求解积分完整模型（6.1）一个时间步所对应的非线性系统。因此，获取滑动窗口所需向量的成本与全模型残差计算相当。不过请注意，如果完整模型采用显式格式进行离散

那么，计算 $\boldsymbol{f}^{(\Delta)}$ 并将完整模型推进一个时间步的成本通常相当。

此处所述的滑动窗口填充过程采用向后看的方式：时刻$\boldsymbol{q}_k$ 的降阶解 k 用于获得在时间步 $k-1$近似全模型解的代理。人们已开发出前瞻策略，利用低成本预测生成向前看的代理，从而避免一步滞后，进而提高稳定性和鲁棒性（Singh 等 2023，Zucatti 和 Zahr，2024）。

6.3.5. ADEIM：采样

ADEIM 基更新由滑动窗口驱动，该窗口通过对全模型右端函数进行稀疏评估获得，如式（6.11）–（6.12）所示。评估全模型右端函数时所用分量的索引由采样点给出，记为 $s_1^{(k-1)}$，……，$s_{n_s}^{(k-1)}$，这些采样点汇集为采样点矩阵 $\boldsymbol{S}_{k-1}$。Peherstorfer 和 Willcox(2015*b*)主要关注均匀分布的采样点。然而，对于可能呈现波前等局部特征的输运主导问题，均匀采样效率低下，如（Peherstorfer 2020）所示。因此，Peherstorfer (2020)提出了一种自适应采样策略。

我们通过以下方式衡量由基矩阵 $\boldsymbol{V}$和 $\boldsymbol{V}_k$分别张成的两个空间之间的距离：

$$d(\bar{\boldsymbol{V}}, \boldsymbol{V}_k) = \|\bar{\boldsymbol{V}} - \boldsymbol{V}_k \boldsymbol{V}_k^\top \bar{\boldsymbol{V}}\|_F^2 , \tag{6.13}$$

其中，为便于阐述，假设 $\boldsymbol{V}_k$ 和 $\boldsymbol{V}$ 均为正交归一。考虑这样一种情形：我们希望将基矩阵 $\boldsymbol{V}_k$ 自适应调整为 $\boldsymbol{V}_{k+1}$，使得 $\boldsymbol{V}_{k+1}$张成的空间在式（6.13）所述的意义下接近张成的空间。请注意，我们现在考虑的是从时间步 k 到 $k+1$ 的自适应调整，而不是从 $k-1$ 到 k。原因在于，算法会在时间步 k更新基之后执行采样点自适应，因此这是为在后续时间步 $k+1$调整空间做准备；参见第6.3.6节给出的算法。

从 $\boldsymbol{V}_k$ 自适应调整到 $\boldsymbol{V}_{k+1}$时，滑动窗口为 $\boldsymbol{F}_k$。目前我们假设 它可表示为 $\boldsymbol{F}_k = \boldsymbol{V}\boldsymbol{F}_k$，其中 $\boldsymbol{F}_k$ 具有满秩，且 $w \geq n$。这意味着，滑动窗口 $\boldsymbol{F}_k$ 的列空间与张成的空间相同。考虑系数矩阵 $\boldsymbol{C}_k$，其定义方式与式（6.7）中的 $\boldsymbol{C}_{k-1}$ 类似；再考虑残差矩阵 $\boldsymbol{R}_k = \boldsymbol{V}_k\boldsymbol{C}_k - \boldsymbol{F}_k$。现在设 $\boldsymbol{S}$ 为一个有效的采样点矩阵。Peherstorfer（2020）证明，采用 $\boldsymbol{F}_k$ 和采样点矩阵 $\boldsymbol{S}$，通过式（6.8）进行 ADEIM 秩-r 更新 $\boldsymbol{V}_{k+1} = \boldsymbol{V}_k + \boldsymbol{\alpha}_k\boldsymbol{\beta}_k^\top$，可得以下界：

$$d(\bar{\boldsymbol{V}}, \boldsymbol{V}_{k+1}) \leq \frac{1}{\sigma_{\min}^2(\boldsymbol{F}_k)} \left(\|\check{\boldsymbol{S}}^\top \boldsymbol{R}_k\|_F^2 + \sum_{i=r+1}^{\bar{r}} \sigma_i^2 \right) , \tag{6.14}$$

其中，$\bar{r}$是 $\boldsymbol{S}^\top\boldsymbol{R}_k$的秩，$\sigma_1 \geq \cdots \geq \sigma_{\bar{r}} > 0$ 是 $\boldsymbol{S}^\top\boldsymbol{R}_k$的奇异值，$\check{}$，而 $\sigma_{\min}(\boldsymbol{F}_k)$是 $\boldsymbol{F}_k$的最小非零奇异值。回想一下，$\boldsymbol{S}$ 是

互补采样点矩阵对应于 $\boldsymbol{S}$，其对应的采样点为$\{1,\ldots,N\}\setminus\{s_1,\ldots,s_{n_s}\}$。界限（6.14）促使我们引入目标函数

$$b(\boldsymbol{S}) = \|\check{\boldsymbol{S}}^\top \boldsymbol{R}_k\|_F^2 + \sum_{i=r+1}^{\bar{r}} \sigma_i^2 \,. \tag{6.15}$$

选择采样点 $\{s_1,\ldots,s_{n_s}\}\subseteq\{1,\ldots,N\}$，使得相应矩阵 $\boldsymbol{S}$ 最小化目标函数 b，但这会在全模型维数下引发组合子集选择问题 N，使得这种直接方法难以处理。相反，Peherstorfer (2020) 建议选择采样点，使其仅最小化范数 $\|\check{\boldsymbol{S}}^\top \boldsymbol{R}_k\|_F^2$，该范数对应于稀疏残差 $\check{\boldsymbol{S}}^\top \boldsymbol{R}_k$，而稀疏残差对应于互补采样点 $\{1,\ldots,N\}\setminus\{s_1,\ldots,s_{n_s}\}$，这是目标函数 b中的一项。由于等式

$$\|\boldsymbol{R}_k\|_F^2 = \|\boldsymbol{S}^\top \boldsymbol{R}_k\|_F^2 + \|\check{\boldsymbol{S}}^\top \boldsymbol{R}_k\|_F^2$$

对于任意采样点矩阵 $\boldsymbol{S}$ 及其互补采样点矩阵 $\check{\boldsymbol{S}}$均成立，因此，寻找一组采样点，使其补集最小化 $\|\check{\boldsymbol{S}}^\top \boldsymbol{R}_k\|_F^2$，等价于寻找一组采样点，使 $\|\boldsymbol{S}^\top \boldsymbol{R}_k\|_F^2$ 最大化。具体步骤如下：首先，计算 $\boldsymbol{R}_k$ 所有 N 行的欧几里得范数，然后按降序排列，再选取对应于最大 n_s 个行范数的 n_s 个索引。所选索引即为新的采样点 $s_1^{(k)},\ldots,s_{n_s}^{(k)}$，其采样点矩阵为 $\boldsymbol{S}_k$。

Cortinovis、Kressner、Massei 和 Peherstorfer (2020) 表明，基于最大化残差范数的方法所得到的目标函数值（6.15）至多比通过最小化 b 而获得的最优采样点选择对应的目标函数值差两倍。通过残差范数选择采样点的成本，主要来自于在所有 N 个分量上计算全模型右端函数，以及随后对 N 行的范数进行排序。ADEIM 及相关在线自适应模型降阶技术还有一系列其他采样策略（Wentland、Duraisamy 和 Huang 2023, Zucatti 和 Zahr，2024，Mohaghegh 和 Huang 2026）。例如，Wentland 等人 (2023) 聚焦于计算流体动力学应用，并提出了物理驱动策略。

6.3.6. ADEIM 降阶模型的在线算法

文献中有 ADEIM 的若干变体（Peherstorfer 和 Willcox 2015*b*, Peherstorfer 2020，Singh等人2023），但它们的核心步骤相同：

（1）低秩基自适应（第6.3.2节），（2）插值点更新（第6.3.2节），（3）使用经验插值的降阶时间积分（第6.3.3节），（4）稀疏全模型查询以填充滑动窗口（第6.3.4节），以及（5）采样点自适应（第6.3.5节）。

下面总结一个通用的在线 ADEIM 流程。为便于说明，我们假设每个时间步都会进行基自适应；在实际应用中，可以仅间歇性地执行步骤（5）和/或（1）–（2）（例如，每隔 z 步执行一次），以降低在线计算成本。

我们遵循Peherstorfer (2020，算法 1) 中给出的在线 ADEIM 过程。对于前 $w_{\text{init}} \in \mathbb{N}$ 个时间步，对完整模型进行时间积分以生成快照 $\boldsymbol{Q}_{1:w_{\text{init}}} \in \mathbb{R}^{N \times w_{\text{init}}}$，并据此通过经验插值构造初始基矩阵$\boldsymbol{V}_{w_{\text{init}}}$、插值点矩阵 $\boldsymbol{P}_{w_{\text{init}}}$ 以及降阶右端函数 $\hat{\boldsymbol{f}}^{(\Delta)}_{w_{\text{init}}}$。此外，用这些快照初始化滑动窗口 $\boldsymbol{F}_{w_{\text{init}}} = \boldsymbol{Q}_{w_{\text{init}}-w+1:w_{\text{init}}}$。随后，对 $k = w_{\text{init}} + 1, \ldots\ldots, K$ 迭代执行以下步骤：

- 步骤 *1*（基更新）：将 ADEIM 基从 $\boldsymbol{V}_{k-1}$ 调整为 $\boldsymbol{V}_k$，所用更新量（6.9）基于滑动窗口 $\boldsymbol{F}_{k-1}$ 和采样点矩阵 $\boldsymbol{S}_{k-1}$（第6.3.2节）。

- 步骤 *2*（插值点更新）：通过例如对自适应基矩阵 $\boldsymbol{P}_k$ 应用 QDEIM，计算插值点矩 $\boldsymbol{V}_k$ 阵（第6.3.2节）。

- 步骤 *3*（前向步）：利用自适应基和插值点，求解由右端函数 $\hat{\boldsymbol{f}}^{(\Delta)}_k$ 给出的降阶模型，并存储提升后的降阶解 $\boldsymbol{q}_k \hat{\boldsymbol{V}}_k$，方法是将 $\boldsymbol{q}_k$ 扩展$\boldsymbol{Q}_{1:k-1}$到 $\boldsymbol{Q}_{1:k}$。注意，$\hat{\boldsymbol{f}}^{(\Delta)}_k$ 现在依赖于步 k，因为底层基 $\boldsymbol{V}_k$ 和插值点矩阵 $\boldsymbol{P}_k$ 依赖于 k（第6.3.3节）。

- 步骤 4（滑动窗口更新）和可选的步骤 5（采样点更新）：如果应在此时间步 k调整采样点，则在所有 N 个分量处计算完整模型右端函数 $\boldsymbol{f}^{(\Delta)}(\boldsymbol{V}_k \hat{\boldsymbol{q}}_k)$，并将其追加到滑动窗口中，得到 $\boldsymbol{F}_k$。从 $\boldsymbol{F}_k$中删除最旧的一列。计算系数矩阵 $\boldsymbol{C}_k$ 和残差矩阵 $\boldsymbol{R}_k = \boldsymbol{V}_k\boldsymbol{C}_k - \boldsymbol{F}_k$，并根据逐行残差范数选择新的采样点 $\boldsymbol{S}_k$（第 6.3.5 节）。否则，如果此时间步 k不应更新采样点，则仅在与 $\boldsymbol{S}_{k-1}$对应的采样点处计算 $\boldsymbol{f}^{(\Delta)}(\boldsymbol{V}_k \hat{\boldsymbol{q}}_k)$，并用 (6.11)–(6.12) 更新滑动窗口 $\boldsymbol{F}_k$。然后设定 $\boldsymbol{S}_k = \boldsymbol{S}_{k-1}$。

采用 ADEIM 降阶模型进行在线时间步计算时，计算成本通常主要来自评估用于更新滑动窗口的全模型右端函数。在调整采样点的时间步，全模型右端函数会在所有 N 个分量上进行评估，因此在线计算成本至少随 N 线性增长；但通常仍比在该时间步求解与完整模型对应的非线性系统（6.1）成本更低。在其他所有时间步，全模型右端函数仅在$n_s \ll N$ 个采样点上进行评估，因此这些稀疏全模型查询的成本随n_s 而非 N增长。

7. 瞬时残差最小化方法

现在考虑一般参数化（4.1），其中权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 依赖于时间和参数，并以非线性方式参与参数化；也可能像（4.7）那样与离线权重向量结合。这里的一般是指，我们不限制权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 参与参数化时必须遵循某种特定的结构形式。这与第 5 节讨论的基于变换的方法以及第6节讨论的在线自适应基方法形成对比；在这些方法中，权重向量以特定的预设方式参与计算，从而可利用这种结构实现高效计算。为推导此类一般非线性参数化的降阶动力学，我们回顾瞬时残差最小化方法。其核心思想是在每个时间 t，通过最小化残差范数来确定权重向量的时间导数 $\dot{\boldsymbol{\theta}}(t, \boldsymbol{\mu})$。

7.1. 本节综述

我们将在第 7.2 节介绍 Dirac–Frenkel 变分原理，阐述其作为一般非线性参数化下瞬时残差最小化核心构件的作用，并回顾神经伽辽金格式等基于该原理构建的数值方法。基于 Dirac–Frenkel 变分原理的方法在推导和实际实现时涉及若干数值方面的考量。首先，残差范数需要进行数值近似。我们将在第 7.3 节回顾神经伽辽金格式中的静态和自适应配点方法。其次，需要对残差最小化所导出的权重向量演化方程进行时间离散。我们将在第 7.4 节简要探讨时间步进选择及其影响。第三，瞬时残差最小化可能不足以完全确定权重向量，从而导致系统奇异或条件性不佳。因此，我们将在第 7.5 节讨论此类情况下的正则化策略。第四，残差最小化会导出由一阶最优性条件给出的代数方程组，每个时间步都必须对其进行数值求解。我们将介绍基于随机草图化、旨在降低计算成本的求解器，并在第 7.5 节讨论随机化如何改善条件性。

我们进一步将 Dirac–Frenkel变分原理与先优化后离散格式联系起来，并在第 7.6 节讨论作为 Dirac–Frenkel变分原理替代方案的瞬时残差最小化的先离散后优化格式。在第 7.7 节中，我们将在基于快照数据训练的非线性参数化背景下考察瞬时残差最小化，其中包括基于自编码器的方法和神经表示。

*7.2. Dirac–Frenkel*变分原理

表述瞬时残差最小化的一种常见方法是借助Dirac–Frenkel变分原理。其起源可以追溯到Dirac(1930年)和Frenkel (1934年)。更现代的表述可见于麦克拉克伦 (1964年),

Kramer和Saraceno (1981),Broeckhove、Lathouwers、Kesteloot和Van Leuven(1988年), Lubich (2008年)。Dirac–Frenkel变分原理的简史载于Lasser和Lubich (2020，第3.8节)。

*7.2.1. Dirac–Frenkel*变分原理

带有权重向量 $\boldsymbol{\theta} \in \mathbb{R}^n$ （其可能依赖于时间 t 和参数 $\boldsymbol{\mu}$）的非线性参数化 $\hat{q}(\boldsymbol{\theta}, \cdot)\colon\ \Omega \to \mathbb{R}$ 会导出式 (4.3) 中定义的集合 $\mathcal{M}$ ，该集合包含通过改变 $\mathbb{R}^n$ 中的权重向量可以表示的所有函数。注意，目前我们并未明确表示 $\hat{q}$ 可能对离线权重向量的依赖。回想一下， $\mathcal{M}$ 表示式 (2.5) 中定义的解流形 $\widehat{\ }$，而我们将集合 (4.3) 记为 $\mathcal{M}$。尽管我们不限制参数化必须具有特定的结构形式，但确实需要作出适当的正则性假设，例如，$\hat{q}$ 关于权重向量 $\boldsymbol{\theta}$ 可微，并且 $\widehat{\ }$ 所有函数 $\hat{q}(\boldsymbol{\theta}, \cdot) \in \mathcal{M}$ 也属于构建偏微分方程问题所用的空间 $\mathcal{Q}$ ；参见第 2.1 节。

$\widehat{\ }$现在，我们将 $\mathcal{M}$ 视为求解偏微分方程问题（2.1）的试探流形，并将带有依赖于时间和参数的权重向量的非线性参数化 $\hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}),\cdot)$ $\boldsymbol{\theta}(t, \boldsymbol{\mu})$代入偏微分方程（2.1），得到残差函数$r\colon\ \mathbb{R}^n \times \mathbb{R}^n \times \Omega \to \mathbb{R}$,

$$r(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \dot{\boldsymbol{\theta}}(t, \boldsymbol{\mu}), \cdot) = \nabla_{\boldsymbol{\theta}} \hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \cdot)^\top \dot{\boldsymbol{\theta}}(t, \boldsymbol{\mu}) - f(\cdot, \hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \cdot); \boldsymbol{\mu})\,. \tag{7.1}$$

我们使用链式法则

$$\partial_t \hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \cdot) = \sum_{i=1}^{n} \partial_{\theta_i} \hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \cdot)\dot{\theta}_i(t, \boldsymbol{\mu}) = \nabla_{\boldsymbol{\theta}} \hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \cdot)^\top \dot{\boldsymbol{\theta}}(t, \boldsymbol{\mu})$$

以便更明确地表示时间导数 $\partial_t \hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}),\cdot)$ 对权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 的时间导数$\dot{}$的依赖关系。

Dirac–Frenkel变分原理旨在寻找权重向量的时间导数 $\dot{\boldsymbol{\theta}}(t, \boldsymbol{\mu}) \in \mathbb{R}^n$ ，使得残差与时间 t 处 $\hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}),\cdot)$ 的所有容许变分$\widehat{\ }$正交。如果 $\mathcal{M}$ 具有由参数化的正则性所诱导的适当流形结构，那么这等价于

$$\left\langle v, r(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \dot{\boldsymbol{\theta}}(t, \boldsymbol{\mu}), \cdot)\right\rangle = 0 \qquad \forall v \in \mathcal{T}_{\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)}\widehat{\mathcal{M}}\,, \tag{7.2}$$

$\widehat{\ }$ $\widehat{\ }$其中 $\mathcal{T}_{\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)}\mathcal{M}$ 表示 $\mathcal{M}$ 在 $\hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \cdot)$ 处的切空间。注意，条件 (7.2) 未必能唯一确定时间导数，若 $\widehat{\ }$空间 $\mathcal{T}_{\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)}\mathcal{M}$ 的维数小于权重的数量 n ；这一点我们将在第 7.5 节再次讨论，并且已在第 5.3.4节遇到过。

正交条件（$\langle\cdot,\cdot\rangle$ ）7.2所采用的内积 L^2 内积；然而， L^2 内积对于某些问题可能并不合适，因为残差上的正交条件无法控制感兴趣范数下的误差。同时，数值计算正交条件

采用其他内积表述的条件可能难以处理。深入讨论请参见Bachmayr、Dahmen 和 Oster (2025*a*)。

正交条件（7.2）具有瞬时性，因为它们在时间上局部作用：在每个时刻 t，选择一个导数 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$，使残差 $\widehat{}$ 与切空间 $\mathcal{T}_{\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)}\mathcal{M}$中的所有元素正交，从而为权重向量 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$定义一个演化方程。对残差施加瞬时条件，这与全局时间方法形成对比，例如基于张量的时空方法（Nouy 2010）和物理信息神经网络（Raissi等，2019），后者通过在时空域上最小化基于残差的目标函数来确定全局近似。关于非线性参数化背景下瞬时（或按时间顺序）方法与全局时间方法的详细比较，见 Zhang 等 (2025)；关于线性模型降阶背景下的讨论，参见（Glas、Mayerhofer 和 Urban 2017）。

注记7.1. 尽管Dirac–Frenkel变分原理为推导降阶动力学提供了一种自然且被广泛使用的机制，但它仅代表更广泛几何设定中的一种特定选择。Buchfink、Glas、Haasdonk和Unger (2024*b*)提出了一种面向非线性模型降阶的微分几何框架，并采用了更为一般的观点。在该框架中，降阶模型不必由Dirac–Frenkel变分原理导出；相反，允许采用定义在解流形切丛上的更一般的降阶映射，这可为使用非线性参数化的保结构模型降阶提供蓝图。

*7.2.2. Dirac–Frenkel*与瞬时残差最小化

给定残差函数（7.1），瞬时残差最小化是指在当前时间 t使残差范数最小，

$$\dot{\boldsymbol{\theta}}(t,\boldsymbol{\mu}) \in \operatorname*{arg\,min}_{\boldsymbol{\eta}\in\mathbb{R}^n} \|r(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\boldsymbol{\eta},\cdot)\|^2, \tag{7.3}$$

其中 $\|\cdot\|$ 是由内积$\langle\cdot,\cdot\rangle$诱导的范数。将（7.3）的一阶最优性条件置为零，即

$$\nabla_{\boldsymbol{\eta}}\|r(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\boldsymbol{\eta},\cdot)\|^2 = 0,$$

得到方程组

$$\langle \partial_{\theta_i}\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot), r(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\boldsymbol{\eta},\cdot)\rangle = 0, \qquad i = 1,\dots,n. \tag{7.4}$$

$\widehat{}$ 现在注意到，如果 $\mathcal{M}$ 配备了适当的流形结构，则$\hat{}q(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)$处的切空间由$\hat{}q$ 关于权重向量各分量所有方向的偏导数张成，

$$\mathcal{T}_{\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)}\widehat{\mathcal{M}} = \operatorname{span}\left\{\partial_{\theta_1}\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot),\dots,\partial_{\theta_n}\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)\right\}. \tag{7.5}$$

因此，由7.4）所给出的条件，即由瞬时残差范数最小化（7.3）导出的条件，等价于Dirac–Frenkel变分原理（7.2）所施加的条件。

If t该参数化是线性的，即形式为（4.2），那么（7.4）等价于 ，

$$\langle \varphi_i, \partial_t \hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot) - f(\cdot, \hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot);\boldsymbol{\mu})\rangle = 0, \qquad i = 1,\ldots,n,$$

如果基函数 φ_i 是正交归一的，则这在坐标表示下恰好对应于 Galerkin降阶模型（2.12）。因此，Dirac–Frenkel变分原理可以被解释为 Galerkin 正交条件的推广。

注释 7.2。 Dirac–Frenkel变分原理，以及更一般的瞬时残差最小化，为在非线性试探流形上推导降阶动力学提供了一个统一框架。特别是，在基于变换的降阶（第5节）和在线自适应基（第6节）部分讨论的若干方法，都可以被解释为瞬时残差最小化的实例。然而，当非线性参数化具有可进一步利用的结构时，降阶动力学通常可以写成更具体的形式，并得到更高效的求解（例如，动态低秩近似中的矩阵分解，或基于变换的模型降阶中的移位参数）。因此，在本综述中，我们将这类结构更为特殊的非线性参数化单独讨论。

7.2.3. 采用非线性参数化的瞬时残差最小化文献综述

关于 Dirac–Frenkel变分原理及其在科学与工程学中的应用，已有大量文献。本节重点讨论基于 Dirac-Frenkel变分原理、针对不一定在离线阶段利用快照数据训练的参数化方法所发展出的方案，即在思想上更接近 PDE求解器的方法。然而，从非线性模型降阶的角度来看，将 Dirac–Frenkel原理应用于经训练的非线性参数化的方法尤其值得关注；我们将在第 7.7 节中单独综述这些方法。

计算化学界已经基于 Dirac-Frenkel变分原理发展出一些方法，用于数值求解例如 Schr¨odinger 方程（Heller 1976、Meyer、Manthe 和 Cederbaum 1990、Beck、J¨ackle、Worth 和 Meyer 2000、Lubich2005、2008）。Dirac–Frenkel变分原理也被用于以矩阵分解作为参数化形式的动态低秩近似和动态正交分解；我们将在第 6.2 节关于时变基方法的背景下对其进行综述。

移动有限元方法中，Miller与Miller（1981）通过有限元系数以及随时间变化的节点位置对解进行参数化，其演化方程由Dirac–Frenkel变分原理得到。但Dirac–Frenkel可应用于更一般的非线性参数化。例如，Anderson 和 Farazmand （2022年）使用Dirac–Frenkel动力学

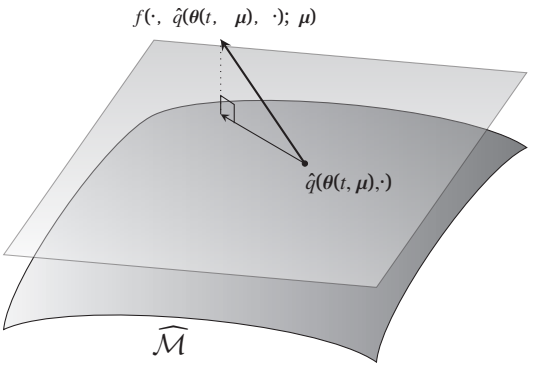


图 7.1：Dirac–Frenkel变分原理通过瞬时残差最小化来演化权重向量 $\theta(t,\ \mu)$：它将时间导数 $\dot{\theta}(t,\ \mu)$ 选为右端函数 $f(\cdot,\ \hat{q}(\theta(t,\ \mu),\ \cdot);\ \mu)$ 在试探流形 $\widehat{\text{形}}$ $\mathcal{M}$ 上、当前解 $\hat{q}(\theta(t,\ \mu),\ \cdot)$ 处的切空间上的正交投影的坐标。

结合形状变形参数化。神经网络参数化因其表达能力而尤为受到关注，Lee 和 Carlberg（2020）以及 Du 和 Zaki（2021年）在Dirac–Frenkel格式的背景下对此进行了研究。类似地，Bruna、Peherstorfer 和 Vanden-Eijnden（2024）提出的神经伽辽金格式为神经网络参数化的权重建立了Dirac–Frenkel演化方程，并由此催生了关于高维问题（Wen、Vanden-Eijnden 和 Peherstorfer，2024）、非线性模型降阶（Berman 和 Peherstorfer，2024；Weder、Schwerdtner 和 Peherstorfer，2025）以及快速保结构在线求解器（Berman 和 Peherstorfer，2023；Schwerdtner、Schulze、Berman 和 Peherstorfer，2024）的后续工作。相关思想也被用于滤波（Aghili、Atokple、Billaud-Friess、Garnier、Mula 和 Tognon，2025）。将Dirac–Frenkel变分原理与神经网络结合使用的诸多方面与挑战已得到讨论：Kast 和 Hesthaven（2024）建议使用拉普拉斯–贝尔特拉米位置嵌入来处理几何复杂区域；Finzi、Potapczynski、Choptuik 和 Wilson（2023）引入重新训练步骤以改善数值稳定性；Zhao、Sun、Cohen、Stokes 和 Veerapaneni（2024）发展了一种变分蒙特卡洛型采样估计量，以实现高维偏微分方程问题的Dirac–Frenkel参数动力学。将Dirac–Frenkel变分原理应用于神经网络等一般非线性参数化时，一个重要变体是其正则化版本；Feischl、Lasser、Lubich 和 Nick（2024）以及 Lubich 和 Nick（2025）对此进行了系统研究，我们将在第 7.5 节中再次讨论。

7.3. 神经伽辽金格式

现在我们讨论神经伽辽金格式，它们使用神经网络参数化来实现 Dirac–Frenkel变分原理（Bruna 等 2024，Berman、Schwerdtner 和 Peherstorfer 2024）。我们重点关注一个特定的数值方面，它对于神经伽辽金格式以及其他基于 Dirac–Frenkel变分原理的格式都至关重要：式（7.2）给出的正交条件涉及必须通过数值方法计算的内积。神经伽辽金

格式通过在空间域上进行求积或配置（采样），来近似这些内积，或者等价地，近似残差最小化表述中的范数。

7.3.1. 神经伽辽金：*Dirac–Frenkel*变分原理的一种实例化受神经网络参数化的启发，Bruna等人（2024）提出了神经伽辽金格式，将Dirac-Frenkel动力学施加于神经网络权重 $\boldsymbol{\theta}$（t、 $\boldsymbol{\mu}$）。这意味着神经网络权重依赖于时间，并依据Dirac–Frenkel变分原理进行演化。我们注意到，较早的工作Du 和 Zaki （2021年）同样在神经网络参数化的背景下使用了Dirac–Frenkel变分原理。

从数值视角来看，神经伽辽金格式直接考虑瞬时残差范数最小化（7.3），因此时间导数 $\boldsymbol{\theta}$（t、 $\boldsymbol{\mu}$）是时变最小二乘问题的解

$$\dot{\boldsymbol{\theta}}(t,\boldsymbol{\mu}) \in \operatorname*{arg\,min}_{\boldsymbol{\eta}\in\mathbb{R}^n} \|\nabla_{\boldsymbol{\theta}}\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)^{\top}\boldsymbol{\eta} - f(\cdot,\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot);\boldsymbol{\mu})\|^2 , \tag{7.6}$$

这与（7.3）中的优化问题相同，只是残差函数（7.1）被明确写出。表述（7.6）强调了这是一个关于未知量 $\boldsymbol{\eta}$的线性最小二乘问题，尽管使用了非线性参数化 $\hat{q}$ 来表示偏微分方程解。此外，表述（7.6）明确表明，时间导数 $\boldsymbol{\theta}$（t、 $\boldsymbol{\mu}$）是右端函数在由梯度的分量函数张成的空间（7.5）上的最佳逼近（投影）的系数向量，如图7.1所示。

最小二乘问题（7.6）的正规方程导出一个动力系统，其形式为 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$,

$$\boldsymbol{E}(\boldsymbol{\theta}(t,\boldsymbol{\mu}))\dot{\boldsymbol{\theta}}(t,\boldsymbol{\mu}) = \boldsymbol{G}(\boldsymbol{\theta}(t,\boldsymbol{\mu});\boldsymbol{\mu}) , \tag{7.7}$$

其矩阵为

$$\boldsymbol{E}_{ij}(\boldsymbol{\theta}(t,\boldsymbol{\mu})) = \langle \partial_{\theta_i}\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot), \partial_{\theta_j}\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)\rangle , \quad i,j = 1,\ldots,n , \tag{7.8}$$

$$\boldsymbol{G}_i(\boldsymbol{\theta}(t,\boldsymbol{\mu});\boldsymbol{\mu}) = \langle \partial_{\theta_i}\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot), f(\cdot,\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot);\boldsymbol{\mu})\rangle , \quad i = 1,\ldots,n . \tag{7.9}$$

动力系统（7.7）可通过时间积分，得到权重轨迹 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$，使得每个时刻都满足正交条件（7.4）。或者，也可以对最小二乘问题（7.6）进行时间积分；与求解正规方程（Berman 等人，2024）相比，这通常能使每个时间步上的中间数值问题条件更好。再次说明，（7.7）可能是欠定的；参见后续第7.5节。

7.3.2. 使用采样点估计内积

本节讨论在神经伽辽金格式的背景下，用于数值估计最小二乘问题（7.6）中范数的配置策略。

数值估计内积的必要性 考虑与神经伽辽金格式相对应的最小二乘公式（7.6）。最小二乘问题的目标函数通过范数 $\|\cdot\|$来表示，该范数由内积 $\langle\cdot,\cdot\rangle$ 诱导，而该内积是 Dirac–Frenkel变分原理给出的残差正交条件（7.2）的基础。因此，为求解关于 ¨的最小二乘问题，必须对目标函数进行数值计算 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$。

正如 Berman 等人 (2024) 所指出，数值分析中也有类似情况：在有限元方法中，组装伽辽金系统需要计算基函数的内积，以求出质量矩阵和刚度矩阵的各项。对于线性参数化，例如由以固定网格点为中心的局部基函数线性组合构成的参数化，这些内积通常可以预先计算并高效复用（Ern 和 Guermond，2004）。相比之下，对于非线性参数化，由于检验空间会随参数化的权重变化而随时间改变，因此内积中的量也会随时间变化，无法直接复用，预先计算便不再可行。此外，由于参数化具有非线性，叠加原理不再成立，因此无法根据各个独立分量求得残差范数。

在应用 Dirac–Frenkel变分原理时，有些情况下残差具有闭式表达式，因此无需数值计算即可精确求值。例如，在某些情况下，可以选择基于高斯波包的非线性参数化，使其对 Schr¨odinger 方程的某些变体给出闭式残差范数；例如参见 Lubich(2008年),Lasser 和 Lubich (2020)。在其他情况下，可以像Anderson 和 Farazmand (2024) 那样使用符号计算。然而，在许多场景中，特别是像模型降阶那样，在离线阶段通过快照数据训练非线性参数化时，并不存在闭式解，因此必须通过数值方法计算残差范数。

通过静态配置点估计内积为了对（7.6）中的范数进行数值计算，我们考虑配置点 $\boldsymbol{x}_1, \ldots\ldots, \boldsymbol{x}_m \in \Omega$，这些点是从测度 ν 中采样得到的，该测度的支撑集为 Ω。例如，若 ν 是 Ω上的均匀测度，则点 $\boldsymbol{x}_1, \ldots\ldots, \boldsymbol{x}_m$ 是在 Ω上均匀采样得到的。我们称这些配置点为静态配置点，因为它们在时间上保持固定，并且与参数 $\boldsymbol{\mu}$无关。

这些配置点用于估计（7.6）中的范数，由此得到具有经验目标函数的优化问题

$$\min_{\boldsymbol{\eta}\in\mathbb{R}^n} \frac{1}{m}\sum_{i=1}^{m} \left|\nabla_{\boldsymbol{\theta}}\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\boldsymbol{x}_i)^{\top}\boldsymbol{\eta} - f(\boldsymbol{x},\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot);\boldsymbol{\mu})|_{\boldsymbol{x}=\boldsymbol{x}_i}\right|^2 . \tag{7.10}$$

如果这些点 $\boldsymbol{x}_1$、……、$\boldsymbol{x}_m$ 是从 ν中抽样得到的，那么式（7.10）中的经验目标函数就是一个蒙特卡罗估计量。使用蒙特卡罗估计量（7.6）估计目标函数（7.10），这与机器学习中的情形类似：总体损失函数通过经验损失函数来估计（Vapnik 1991）。

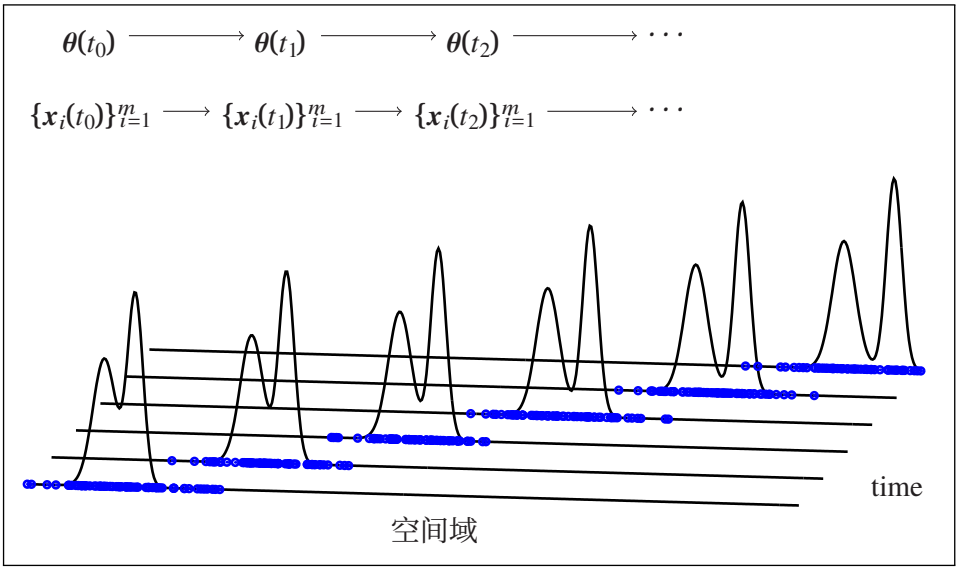


图 7.2：神经伽辽金格式按照 Dirac–Frenkel变分原理的规定演化权重向量，同时可以根据由残差确定的势函数自适应采样点。这种自适应机制会将采样点集中在主导残差的区域，从而比静态采样更高效地进行残差范数估计。

数值求积是蒙特卡罗估计的另一种选择：点 $\boldsymbol{x}_1$、……、$\boldsymbol{x}_m$ 根据某种求积规则选取，该规则考虑了测度 ν。使用求积规则可得到与式（7.10）类似的估计量，但权重可能不同于 $1/m$。例如，Du 和 Zaki(2021年) 和 Schwerdtner 等人(2024) 研究了 Dirac–Frenkel变分原理背景下用于计算内积的数值求积。

7.3.3. 自适应采样点

神经伽辽金格式支持让采样点随权重向量一起演化 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$；参见图7.2。

输运主导问题中静态配置点的局限性回到第2.5节图2.1中的动机示例，输运主导问题通常会表现出移动的波前或其他局部的相干特征——尖锐的解的梯度、凸起、波前——它们会穿过空间域。

使用一组静态配置点时，也就是说这些点在所有时间内都固定不变 t，必须对这些点进行选择，使其能够在所有时间内充分分辨移动局部特征。然而，要确保所有时间内的精度均匀并不容易，因为局部特征可能穿过

空间域的任意区域，因此整个空间域都需要较高的分辨率，这通常需要大量采样点（Bruna等，2024，Wen等，2024）。在图 2.1所示的线性输运示例中，均匀静态采样必须在所有时间内覆盖域的各个部分，以追踪移动的窄高斯波包。尽管在大多数时间步中，这些点中的大多数都位于解接近于零的区域，因此提供的有效信息很少，但仍需要计算梯度和右端项，以便像式（7.6）中那样，通过经验方式估计式（7.10）的目标函数。因此，依赖静态采样点可能需要大量采样点，才能分辨移动局部特征。

注释 7.3. 物理信息机器学习方法中也存在类似的采样难题，这些方法旨在空间（及时间）区域上最小化残差范数。如果解的特征具有局部性，就需要大量配点，从而需要进行大量残差范数评估，这可能成为计算瓶颈（Nabian、Gladstone和Meidani 2021，Rotskoff、Mitchell和Vanden-Eijnden 2022，Wu、Zhu、Tan、Kartha和Lu 2023）。

利用自适应配点估计内积 利用时间自适应配点估计内积，是一种即使解的特征具有局部性，也能将配点数量保持在较低水平的方法。Wen等人（2024）通过时变测度 $\nu_{\theta(t,\mu),\dot{\theta}(t,\mu)}$ that are allowed to change with the weights $\theta(t,\mu)$ and the derivative $\dot{\theta}(t,\mu)$来构建神经伽辽金格式。非线性模型降阶中的自适应采样也在Bon、Caris和Mula（2025）的研究中有所探讨，他们提出了一种直接针对降阶动力学条件性的采样准则；Black等人（2021a）则研究了随时间更新插值点，以高效估计非线性项。此外，Gruhlke、Nouy和Trunschke（2024）的工作也与此相关，该研究引入了用于自然梯度下降的采样，这对应于神经伽辽金格式所描述的类比动力学（Zhang等人 2025）。我们强调，在数值分析中，偏微分方程逼近领域的采样点自适应选择也应用于移动有限元方法等场景（Miller与Miller 1981）。

根据 Wen 等人 (2024) 的研究，我们通过一个依赖时间的测度来表述正交条件，这会导出与（7.6）类似的最小二乘问题，唯一不同的是 L^2 范数 $\|\cdot\|_{\nu_t}$ 现在依赖于 ν_t，即

$$\min_{\dot{\theta}(t,\mu)\in\mathbb{R}^n} \|\nabla_{\theta}\hat{q}(\theta(t,\mu),\cdot)^{\top}\dot{\theta}(t,\mu) - f(\cdot,\hat{q}(\theta(t,\mu),\cdot);\mu)\|^2_{\nu_t},$$

其中我们将 $\nu_{\theta(t,\mu),\dot{\theta}(t,\mu)}$ 简记为 ν_t。 Wen 等人 (2024) 指出，如果残差相对于静态测度 ν可以设为零，那么在测度在空间域 Ω上均有完全支撑等标准条件下，最优解不受测度变化的影响。Wen Wen 等人(2024) 还指出，假设残差接近于零是合理的，因为

通常人们关注的是参数化足够丰富、从而能够使残差保持较小的情况。我们注意到，除了用时间依赖范数来表述目标函数外，还可以在保持基准测度不变的同时，利用时间依赖偏置测度推导目标函数的重要性采样估计量；相关的重要性采样方法见 (Bruna 等人，2024) 的讨论。

一个主要问题是如何将测度 ν_t 随时间向前传播，从而使来自 ν_t 的样本能够给出残差范数的准确估计量。Wen等人（2024）提出让残差本身驱动 ν_t 的演化，方法是将 ν_t 定义为吉布斯测度

$$\nu_t \propto \exp\left(-U_{\theta(t,\mu),\dot{\theta}(t,\mu)}\right)$$

其中势函数$U_{\theta,\,\dot{\theta}}$取为残差平方，即

$$U_{\theta(t,\mu),\dot{\theta}(t,\mu)}(\boldsymbol{x}) = |r(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\dot{\boldsymbol{\theta}}(t,\boldsymbol{\mu}),\boldsymbol{x})|^2\,. \tag{7.11}$$

测度 ν_t 的高概率区域对应于绝对残差较大的空间位置。也可以使用其他势函数，例如，若解函数是概率密度函数，则可直接从中采样（Bruna等人 2024），或基于其他采样准则，例如在Bon等人（2025）中引入的基于稳定性的准则。

演化该时间相关测度的一种形式化方法由势函数（7.11）驱动的福克–普朗克方程给出，

$$\partial_t \nu_t = \gamma\nabla\cdot\left(\nabla\nu_t + \nu_t\nabla U_{\theta(t,\mu),\dot{\theta}(t,\mu)}\right)\,, \tag{7.12}$$

它可以与权重动力学耦合，

$$\begin{aligned}\boldsymbol{E}_t(\boldsymbol{\theta}(t,\boldsymbol{\mu}))\dot{\boldsymbol{\theta}}(t,\boldsymbol{\mu}) &= \boldsymbol{G}_t(\boldsymbol{\theta}(t,\boldsymbol{\mu});\boldsymbol{\mu})\,,\\ \partial_t\nu_t &= \gamma\nabla\cdot\left(\nabla\nu_t + \nu_t\nabla U_{\theta(t,\mu),\dot{\theta}(t,\mu)}\right)\,,\end{aligned} \tag{7.13}$$

其中，矩阵 $\boldsymbol{E}_t$和 $\boldsymbol{G}_t$分别在式（7.8）和（7.9）中给出；由于它们是相对于与 ν_t相关的内积构造的，现在也依赖于时间。参数 γ 控制测度 ν_t 相对于权重 $\boldsymbol{\theta}$（t，$\boldsymbol{\mu}$）进行调整的时间尺度。

关键在于，无须显式求解福克-普朗克方程（7.12），因为只需要从对应于 ν_t的密度中抽取样本。因此，测度 ν_t 可以通过经验测度近似

$$\hat{\nu}_t = \frac{1}{m}\sum_{i=1}^{m}\delta_{\boldsymbol{x}_i(t)}$$

由时间依赖样本给出 $\boldsymbol{x}_1$（t），……，$\boldsymbol{x}_m$（t）。由于势函数（7.11）由残差给出，因此可以计算残差梯度以得到得分 $\nabla\log\,\nu_t$的 ν_t从而推进样本 $\boldsymbol{x}_1$（t），……，$\boldsymbol{x}_m$（t）随

时间用粒子方法推进，而不是求解福克-普朗克方程（7.12）。Wen 等人（2024）提出使用朗之万采样器和Stein变分梯度下降，二者均通过得分 $\nabla \log \nu_t$传播样本。这样一来，权重轨迹 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$与样本集合 $\{\boldsymbol{x}_i(t)\}_{i=1}^m$ 随时间共同推进，从而逼近耦合动力学（7.13），如图 7.2。

7.4. 半离散与全离散瞬时残差最小化

现在，我们讨论由 Dirac–Frenkel变分原理 得到的最小二乘问题的半离散和全离散变体，所采用的表述和术语遵循神经伽辽金格式。

半离散最小二乘问题 给定 m 采样点 $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m \in \Omega$，这些采样点可以像第 7.3.2 节那样静态获取，也可以像第 7.3.3 节那样自适应获取。考虑所谓的批量梯度

$$\boldsymbol{J}(\boldsymbol{\theta}) = \begin{bmatrix} \nabla_{\boldsymbol{\theta}} \hat{q}(\boldsymbol{\theta}, \boldsymbol{x}_1)^\top \\ \vdots \\ \nabla_{\boldsymbol{\theta}} \hat{q}(\boldsymbol{\theta}, \boldsymbol{x}_m)^\top \end{bmatrix} \in \mathbb{R}^{m \times n}, \tag{7.14}$$

其各行是非线性参数化 $\nabla_{\boldsymbol{\theta}} \hat{q}(\boldsymbol{\theta}, \cdot)$ 的梯度转置，其中 q 是在 m 采样点处求值得到的。批量梯度这一名称源于如下考虑：像式（7.10）那样近似目标函数时，$\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$是当前批次的样本；原则上，这些样本可以随时间步（以及时间离散情形下的优化迭代；见下文）而变化。类似地，我们将批次右端函数定义为

$$\boldsymbol{F}(\boldsymbol{\theta}; \boldsymbol{\mu}) = \begin{bmatrix} f(\boldsymbol{x}, \hat{q}(\boldsymbol{\theta}, \cdot); \boldsymbol{\mu})|_{\boldsymbol{x}=\boldsymbol{x}_1} \\ \vdots \\ f(\boldsymbol{x}, \hat{q}(\boldsymbol{\theta}, \cdot); \boldsymbol{\mu})|_{\boldsymbol{x}=\boldsymbol{x}_m} \end{bmatrix} \in \mathbb{R}^m. \tag{7.15}$$

利用批量梯度和批量右端项函数，我们可以将半离散瞬时残差最小化问题写为

$$\dot{\boldsymbol{\theta}}(t, \boldsymbol{\mu}) \in \operatorname*{arg\,min}_{\boldsymbol{\eta} \in \mathbb{R}^n} \|\boldsymbol{J}(\boldsymbol{\theta}(t, \boldsymbol{\mu}))\boldsymbol{\eta} - \boldsymbol{F}(\boldsymbol{\theta}(t, \boldsymbol{\mu}); \boldsymbol{\mu})\|_2^2, \tag{7.16}$$

它是样本点 $\boldsymbol{x}_1$、……、$\boldsymbol{x}_m$ 上经验目标函数（7.10）的矩阵表示，仅差一个常数缩放因子。请注意，（7.16）中的范数 $\|\cdot\|_2$ 表示欧几里得向量范数，可以直接计算。值得注意的是，最小二乘问题（7.16）关于未知量是线性的；在进行时间离散时，人们通常希望保留这一性质。

时间离散 考虑离散时间步 $0 = t_0 < t_1 < \cdots < t_K = T$，时间步长为常数 $\tau > 0$。采用龙格-库塔方法对最小二乘问题（7.6进行显式时间离散，会得到一系列最小二乘

问题，这些问题关于未知量也都是线性的。例如，使用显式欧拉法会在所有时间步 $k = 1, \ldots, K$得到线性最小二乘问题，

$$\min_{\delta\boldsymbol{\theta}_k(\boldsymbol{\mu})\in\mathbb{R}^n} \|\boldsymbol{J}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}))\delta\boldsymbol{\theta}_k(\boldsymbol{\mu}) - \boldsymbol{F}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu});\boldsymbol{\mu})\|_2^2\,, \tag{7.17}$$

连同更新量

$$\boldsymbol{\theta}_k(\boldsymbol{\mu}) = \boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}) + \tau\delta\boldsymbol{\theta}_k(\boldsymbol{\mu})\,. \tag{7.18}$$

因此，若采用显式格式进行离散化，则每个时间步只需求解一个线性最小二乘问题。

相比之下，隐式时间离散会在每个时间步导致非线性优化问题。为演示起见，采用隐式欧拉法，我们得到

$$\min_{\delta\boldsymbol{\theta}_k(\boldsymbol{\mu})\in\mathbb{R}^n} \|\boldsymbol{J}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}) + \tau\delta\boldsymbol{\theta}_k(\boldsymbol{\mu}))\delta\boldsymbol{\theta}_k(\boldsymbol{\mu}) - \boldsymbol{F}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}) + \tau\delta\boldsymbol{\theta}_k(\boldsymbol{\mu});\boldsymbol{\mu})\|_2^2, \tag{7.19}$$

其中更新量 $\delta\boldsymbol{\theta}_k(\boldsymbol{\mu})$ 按（7.18）中的方式施加。在隐式时间离散的情形下，时间离散最小二乘问题关于更新量 $\delta\boldsymbol{\theta}_k(\boldsymbol{\mu})$ 变为非线性，因为批量梯度和批次右端函数都在下一迭代 $\boldsymbol{\theta}_k(\boldsymbol{\mu})$ 处求值。因此，数值求解（7.19）相当于求解关于更新量 $\delta\boldsymbol{\theta}_k(\boldsymbol{\mu})$ 的非线性最小二乘问题，从而继承了拟合非线性参数化（如神经网络）的典型困难。特别是，拟合必须在每个时间步进行。因此，人们对显式时间积分有强烈兴趣，因为它对许多基于狄拉克–弗兰克尔的方法在离散时间下保持了最小二乘问题的线性性。

*7.5. Dirac–Frenkel*动力学的正则化与随机化时间积分

如前所述，Dirac–Frenkel变分原理中的正交条件所产生的最小二乘问题可能条件性较差，甚至秩亏。第5.3.4节在讨论基于变换的方法时已给出一个具体例子。一般而言，病态性促使我们采用正则化。我们考虑通过蒂霍诺夫方法和截断奇异值分解进行显式正则化。我们还简要讨论通过右侧草图实现的随机化时间积分；这种方法既能起到正则化的作用，又能通过将更新量限制在低维随机子空间中，减少每个时间步要求解的未知数数量。

7.5.1. 最小二乘问题的条件性较差与切空间塌缩

考虑式（7.17）给出的时间离散最小二乘问题，该问题是对式（7.16）给出的时间连续问题应用显式欧拉法得到的。

基于 Dirac–Frenkel 的方案面临的一项重大挑战是数值求解最小二乘问题。除了计算成本之外，由于条件性较差，最小二乘问题在数值求解时也可能颇具挑战。

最小二乘问题的条件性（7.17）主要由权重向量 $\boldsymbol{J}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu})$ 处批量梯度的条件数决定 $\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu})$。对于非线性参数化，批量梯度可能条件性很差，甚至（数值上）秩亏；在这种情况下，最小二乘问题是欠定的，权更新量也不唯一。此外，秩亏与 Dirac–Frenkel 方法的精度密切相关：误差界表明，误差受将右端项投影到批量梯度列空间上所产生的投影误差控制（等价地，在连续时间下，是梯度的分量函数或切空间）。当批量梯度秩亏，且用于投影右端项的相应空间无法很好地表示右端项时，残差和近似误差就可能很大，从而导致精度较低（Lubich2005）。

但秩亏在 Dirac–Frenkel变分原理的背景下有着更广泛的影响。在满足适当正则性假设时，梯度的分量函数张成切空间；离散化后，这些分量函数对应于批量梯度的各列 $\boldsymbol{J}$，也张成了使残差满足正交条件的检验空间（7.4）。因此，秩亏还意味着 Dirac–Frenkel 条件仅要求残差相对于一个低维检验空间正交，检验空间之外的残差分量不受控制，因而可能增大，导致精度下降，即使某一时刻的中间近似仍然准确。在 Dirac–Frenkel变分原理的背景下，这种退化现象称为矩阵奇异性问题（Kay 1989，Sawada、Heather、Jackson 和 Metiu 1985，Rowan、Schatzki、Zaklama、Suzuki、Watanabe 和 Varga 2020）。文献（Zhang 等 2025）将其称为切空间塌缩，因为梯度的分量函数所张成的空间会变为低维空间，并且在适当条件下，这些分量函数张成切空间。

正则化是一种自然的补救方法，在实践中也很重要，但一个主要难点在于，这里求解的并非单个孤立的最小二乘问题，而是一系列最小二乘问题，其条件性和解都取决于先前的时间步。因此，正则化引入的任何偏差都可能随时间累积；要控制并解释其对整个时间轨迹的影响，就需要采用审慎的技术，并进行系统而严谨的分析（Feischl 等人 2024）。此外，使用正则化时需要谨慎，因为即便是常用的正则化项，也可能导致所选更新量（例如最小范数更新）保留参数冗余（例如，相同或冗余的分量仍保持相同），从而阻碍系统从退化状态中恢复（Zhang等人 2025）。

7.5.2. 正则化的狄拉克–弗兰克尔变分原理

Feischl 等人对 *Dirac–Frenkel* 最小二乘问题的 *Tikhonov* 正则化进行了系统且严谨的研究（2024）。为缓解残差最小化问题中的病态性，Feischl 等人（2024）在时间连续形式中对时间导数应用 Tikhonov

正则化作用于时间导数 $\dot{\boldsymbol{\theta}}(t,\ \boldsymbol{\mu})$，并在时间离散形式中作用于更新量（“速度”） $\delta\boldsymbol{\theta}_k(\boldsymbol{\mu})$，例如，

$$\delta\boldsymbol{\theta}_k(\boldsymbol{\mu}) = \arg\min_{\boldsymbol{\eta}\in\mathbb{R}^n} \|\boldsymbol{J}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}))\boldsymbol{\eta} - \boldsymbol{F}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu});\boldsymbol{\mu})\|_2^2 + \alpha\|\boldsymbol{\eta}\|_2^2,$$

通过更新量 $\boldsymbol{\theta}_k(\boldsymbol{\mu}) = \boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}) + \tau\delta\boldsymbol{\theta}_k(\boldsymbol{\mu})$。这里，参数 $\alpha \geq 0$ 控制 Tikhonov 正则化，它起到平滑滤波器的作用：最小二乘问题的系统矩阵中与小奇异值相关的方向会受到阻尼，从而抑制较大且可能不稳定的权重更新。

Feischl 等人 (2024年) 的分析强调，仅对解本身进行准确近似，并不足以保证可靠的时间演化。还需要沿解轨迹的切空间能够良好近似右端项（等价地说，即解的时间导数），因为在切空间中近似右端项所产生的误差会计入总体近似误差；另见 Lubich(2005)、Zhang 等人(2025)。这一现象并非非线性模型降阶所独有。在线性模型降阶中也会出现类似现象：仅准确近似状态，并不能保证时间演化也得到准确近似。这一认识是线性模型降阶中广泛采用 Petrov–Galerkin 投影的基础；在这种方法中，选择不同的试验空间和检验空间，以改善对系统动力学的近似（Otto 等人2022年）。

此外，正则化强度必须谨慎选择：较弱的正则化会减小偏差，但可能导致更新量沿着对应于小奇异值的方向变化；较强的正则化有助于提高稳定性，但最终可能使残差无法足够小。 Feischl 等人 (2024) 提出了一种基于缺陷的自适应方法，其中缺陷是正则化最小二乘目标函数的最小值。 Feischl 等人 (2024) 还将 Tikhonov 正则化与截断奇异值分解正则化联系起来：截断奇异值分解会完全舍弃对应于小奇异值的方向，而 Tikhonov 正则化则只是对这些方向进行平滑。

Lubich 和 Nick (2025) 针对刚性问题，因此考虑在 Dirac–Frenkel 框架下对隐式而非显式时间积分格式进行正则化。在每个时间步，都要求解一个带有权更新正则化项的正则化非线性最小二乘问题。非线性问题通过少数几次高斯–牛顿法迭代求解。

7.5.3. 通过随机子空间近似进行正则化

Berman与Peherstorfer （2023），并在Dong、Schwerdtner与Peherstorfer （2025）中得到进一步发展，提出通过在随机子空间中求解权更新，对与Dirac–Frenkel变分原理相对应的动力学进行正则化，这可解释为从右侧进行草图化。

Bu基于时间离散问题（7.17）–（7.18），权更新变为

$$\boldsymbol{\theta}_k(\boldsymbol{\mu}) = \boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}) + \frac{\tau}{\ell} \sum_{i=1}^{\ell} \delta\boldsymbol{\theta}_k(\boldsymbol{\mu}; \boldsymbol{\Gamma}_{k,i}), \tag{7.20}$$

其中 $\delta\boldsymbol{\theta}_k(\boldsymbol{\mu};\ \boldsymbol{\Gamma}_{k,\ 1})$，…，$\delta\boldsymbol{\theta}_k(\boldsymbol{\mu};\ \boldsymbol{\Gamma}_{k,\ \ell})$ 是 l 个相互独立的随机更新，它们依赖于相互独立的随机嵌入 $\boldsymbol{\Gamma}_{k,\ 1}$，…，$\boldsymbol{\Gamma}_{k,\ \ell}$。注意，这些$l$ 个更新的平均值用于（7.20）中，有助于控制随机更新的方差。这些更新由以下最小二乘问题得到

$$\delta\boldsymbol{\theta}_k(\boldsymbol{\mu}; \boldsymbol{\Gamma}) = \underset{\boldsymbol{\eta} \in \mathrm{Range}(\boldsymbol{\Gamma})}{\arg\min} \ \|\boldsymbol{J}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}))\boldsymbol{\eta} - \boldsymbol{F}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}); \boldsymbol{\mu})\|_2^2, \tag{7.21}$$

该问题将解限制在随机矩阵 $\boldsymbol{\Gamma}$ 的列空间内。矩阵 $\boldsymbol{\Gamma}$ 的大小为 $n \times s$，其中 $s \le n$。

对于 $\boldsymbol{\Gamma}$，典型的选择包括高斯嵌入和随机酉嵌入。对于这类嵌入，Dong 等人 (2025) 推导出了草图批量梯度 $\boldsymbol{J}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}))\boldsymbol{\Gamma}$ 的条件数界，该界与 $\boldsymbol{J}(\boldsymbol{\theta}_{k-1}(\boldsymbol{\mu}))$ 的谱及草图维数 s有关，并直接控制最小二乘问题（7.21）的条件数。早期工作 Berman 和 Peherstorfer (2023) 通过选择 $\delta\boldsymbol{\theta}_k(\boldsymbol{\mu};\ \boldsymbol{\Gamma})$ 来考虑稀疏更新，使 $\boldsymbol{\Gamma}$ 对应于随机稀疏坐标嵌入。

随机化表述（7.21）从右侧对最小二乘问题进行草图化，即将权更新限制在一个随机子空间中。因此，式（7.21）中的未知量是一个 s 维向量，因此只需求解 s 个未知量，而不是求解对应于权更新所有分量的全部 n 个未知量。相比之下，另有一系列研究针对动态低秩近似以及其他更一般的非线性参数化，探讨从左侧进行草图化（Lam 等，2025，Carrel（2026）；Lindsey（2025））。从左侧进行草图化有助于改善条件性，但更新量中的未知数仍为 n，而非 s。

7.6. 采用先离散后优化（*DtO*）与先优化后离散（*OtD*）方法的瞬时残差最小化

Dirac–Frenkel变分原理可理解为先优化后离散（OtD）类型的方法：首先，通过令残差与试探流形的切空间正交，推导出权重 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$ 的连续时间演化方程，然后再对该演化进行时间离散。

正如 Zhang 等人（2025）详细讨论的那样，另一类概念上截然不同的瞬时（或按时间顺序）残差最小化方案源于先离散后优化（DtO）策略：在 DtO 方案中，首先在偏微分方程层面对时间进行离散，从而得到一系列关于空间变量的边值问题。只有在完成时间离散之后，才引入非线性参数化，并针对每个边值

问题（对应于离散时间步）构造残差最小化问题。针对非线性参数化的 DtO 残差最小化方案，例如 Kvaal、Lasser、Pedersen 和 Adamowicz （2023）的研究、Chen、Wu、Grinspun、Zheng 和 Chen（2023*a*）、Chen、Mccarran、Vizcaino、Soljacic 和 Luo（2024）。值得注意的是，对于线性参数化，DtO 对应于 Rothe 方法（Rothe 1930年、Deuflhard 和 Weiser 2012年）。由于针对每个边值问题进行残差最小化，通常会在每个时间步产生一个非凸优化问题，因此 DtO 方案的计算成本往往高于 OtD 方案；后者在每个时间步都会得到一个线性最小二乘问题——至少对于显式时间积分而言。与此同时，DtO 方案可以避免 OtD 方案中通常出现的病态问题，例如切空间塌缩或矩阵奇异问题。关于 OtD 与 DtO 方案的深入讨论，请参阅 Zhang 等人（2025）。

我们指出，线性模型降阶中也存在与 OtD 和 DtO 公式类似的对应关系。Carlberg等人（2011、2017年）提出并分析了基于线性参数化的最小二乘Petrov–Galerkin降阶模型这一概念。按照 Carlberg 等人 （2017年）使用的术语，DtO 对应最小二乘Petrov–Galerkin，而 OtD 对应 Galerkin公式。

7.7. 使用经训练的非线性参数化进行非线性模型降阶的瞬时残差最小化

现在，我们综述在线阶段采用瞬时残差最小化、针对离线阶段基于快照数据训练的参数化的模型降阶方法。参数化通常需要依赖离线权重（以向量形式表示为 $\boldsymbol{\theta}_{\text{off}}$；见第 4.3.1 节），才能进行训练。参数化训练完成后，即离线权重固定后，在线权重 $\boldsymbol{\theta}$（t、$\boldsymbol{\mu}$）即可在在线阶段通过瞬时残差最小化求得。

本节重点介绍通过神经网络自编码器、具有先验固定特征映射的自编码器以及神经表示得到的经训练的非线性参数化。我们强调，此外还有基于变换的方法和在线自适应模型降阶，它们也能生成经训练的非线性参数化，可在线用于残差最小化；相关内容将在各自章节中介绍。

7.7.1. 神经网络自编码器给出的非线性参数化

Lee 和 Carlberg （2020）提出了一种方法：使用由深度卷积神经网络参数化的自编码器，在离线阶段用快照数据对其进行训练，然后在在线阶段使用自编码器的解码器函数，构建非线性参数化来表示降阶解。此设定中的向量 （t，$\boldsymbol{\mu}$）对应于降阶解的潜在状态，并通过瞬时残差最小化计算得出。尽管此前已有使用神经网络自编码器进行模型降阶的研究（Kashima 2016，

Hartman 和 Mestha 2017），Lee 和 Carlberg （2020）的工作首次提出了一个全面的框架，并由此推动了一系列关于使用自编码器进行非线性模型降阶的论文。Buchfink 等人 （2024*b*）介绍了在非线性模型降阶中使用自编码器的一般微分几何框架。我们的介绍大体上遵循 Lee 和 Carlberg （2020）。

自编码器 自编码器是两个映射的复合 $\Phi^{\uparrow} \circ \Phi^{\downarrow}$：$\mathbb{R}^N \to \mathbb{R}^N$ 编码器 $\Phi^{\downarrow}$：$\mathbb{R}^N \to \mathbb{R}^n$ 和解码器 $\Phi^{\uparrow}$：$\mathbb{R}^n \to \mathbb{R}^N$。编码器$\Phi^{\downarrow}$ 将维度为 N 的高维向量映射$\Phi^{\downarrow}$ 为维度为 $n \ll N$ 的低维向量。解码器则执行相反方向的操作，将低维向量提升为高维向量。

在模型降阶框架中，通常关注这样的编码器-解码器对：其复合映射 $\Phi^{\uparrow} \circ \Phi^{\downarrow}$ 接近于数据上的恒等映射，这些数据对应于由完整模型诱导的解流形 (2.5)：将自编码器应用于半离散全模型 (2.4) 的状态 $\boldsymbol{q}(t, \boldsymbol{\mu}) \in \mathbb{R}^N$ 时，编码器会将全阶模型状态 $\boldsymbol{q}(t, \boldsymbol{\mu})$ 映射为低维近似$\tilde{\boldsymbol{q}}(t, \boldsymbol{\mu}) = \Phi^{\downarrow}(\boldsymbol{q}(t, \boldsymbol{\mu})) \in \mathbb{R}^n$。解码器将 n 维的近似$\tilde{\boldsymbol{q}}(t, \boldsymbol{\mu})$ 提升为 N 维向量 $\Phi^{\uparrow}(\tilde{\boldsymbol{q}}(t, \boldsymbol{\mu})) \in \mathbb{R}^N$，该向量近似于全阶模型状态 $\boldsymbol{q}(t, \boldsymbol{\mu})$。对于 Galerkin框架中的线性模型降阶 (2.12)，编码器由 $\Phi^{\downarrow}(\boldsymbol{q}) = \boldsymbol{V}^{\top}\boldsymbol{q}$ 给出，解码器则由 $\Phi^{\uparrow}(\boldsymbol{q}) = \boldsymbol{V}\boldsymbol{q}$ 给出。在 Petrov–Galerkin框架中，可以用另一个矩阵 $\boldsymbol{W}^{\top}$ 替换编码器中的矩阵 $\boldsymbol{V}^{\top}$。

在快照数据上训练自编码器 Lee 和 Carlberg（2020）使用神经网络对编码器和解码器映射进行参数化。按照第4.3节的思路，我们将参数化的编码器记为 $\Phi^{\downarrow}(\cdot;\, \boldsymbol{\theta}^{\downarrow}\text{off}) : \mathbb{R}^N \to \mathbb{R}^n$，将参数化的解码器记为 $\Phi^{\uparrow}(\cdot;\, \boldsymbol{\theta}^{\uparrow}\text{off}): \mathbb{R}^n \to \mathbb{R}^N$，其离线权重向量 $\boldsymbol{\theta}^{\downarrow}\text{off}$ 和 $\boldsymbol{\theta}^{\uparrow}\text{off}$ 的维数适当。参数化的架构取决于具体问题。例如，Lee 和 Carlberg（2020）、Buchfink、Glas和Haasdonk（2023）使用深度卷积神经网络。相比之下，Kim、Choi、Widemann和Zohdi（2022）使用在输出层应用稀疏掩码的浅层网络，以降低在线阶段的成本。Gruber、Gunzburger、Ju和Wang（2022）对不同架构进行了比较。

离线权重 $\boldsymbol{\theta}^{\downarrow}{}_{\text{off}}$ 和 $\boldsymbol{\theta}^{\uparrow}{}_{\text{off}}$ 通过完整模型（2.9）的快照数据（2.4）进行训练。Lee 和 Carlberg (2020)中使用的一种标准损失函数是快照重构的均方误差，

$$\min_{\boldsymbol{\theta}^{\downarrow}_{\text{off}}, \boldsymbol{\theta}^{\uparrow}_{\text{off}}} \left\| \boldsymbol{Q}_{\text{train}} - \Phi^{\uparrow}(\Phi^{\downarrow}(\boldsymbol{Q}_{\text{train}}; \boldsymbol{\theta}^{\downarrow}_{\text{off}}); \boldsymbol{\theta}^{\uparrow}_{\text{off}}) \right\|_F^2, \tag{7.22}$$

其中，$\boldsymbol{Q}_{\text{train}}$ 是快照矩阵，编码器和解码器逐列应用于其矩阵输入。

还可以在（7.22）中添加其他项，以惩罚结构违反。例如，Buchfink 等人 （2023）添加了一个惩罚结构违反的惩罚项

解码器的辛性，从而得到弱辛自编码器。在 Lee 和 Carlberg （2021年）中加入了一个物理信息惩罚项。 Otto 等人 （2023）还训练了一个专为所学习的自编码器量身定制的投影算子。

基于自编码器的非线性参数化的在线残差最小化Lee 和 Carlberg (2020) 提出使用解码器 $\Phi^{\uparrow}$ 来构造非线性参数化。为便于说明，我们现在在解码器的记号中省略对离线权重的依赖。所构造的参数化$\Phi^{\uparrow}\colon \mathbb{R}^n \to \mathbb{R}^N$ 是向量值的，且不将空间坐标作为输入。相反，我们可以将 $\Phi^{\uparrow}$ 视为

$$\Phi^{\uparrow}(\boldsymbol{\theta}(t,\boldsymbol{\mu})) = \begin{bmatrix} \hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\boldsymbol{x}_1) & \dots & \hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\boldsymbol{x}_N) \end{bmatrix}^{\top}, \tag{7.23}$$

其中 $\hat{q}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\cdot)\colon \Omega \to \mathbb{R}$ 是一个隐式给定的标量值非线性参数化，其形式为（4.1），并在点 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_N \in \Omega$ 处求值，这些点对应于定义半离散全模型（2.4）的空间坐标。需要强调的是，$\hat{q}$允许将不同于 $\boldsymbol{x}_1,\dots,\boldsymbol{x}_N$的其他空间坐标作为输入，但通常无法直接从解码器函数 $\Phi^{\uparrow}$ 中获得。尤其是，当自编码器通常用于模型降阶时，它们直接作用于半离散全模型的 N 维状态向量，因此从根本上受到底层全模型离散化的约束。关于独立于底层全模型离散化的神经表示，请参阅第 7.7.3 节。

与解码器 $\Phi^{\uparrow}$ 所诱导的向量值参数化相对应，可以定义向量值残差函数 $\boldsymbol{r}\colon \mathbb{R}^n \times \mathbb{R}^n \times \mathcal{D} \to \mathbb{R}^N$,

$$\boldsymbol{r}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\dot{\boldsymbol{\theta}}(t,\boldsymbol{\mu});\boldsymbol{\mu}) = \boldsymbol{J}(\boldsymbol{\theta}(t,\boldsymbol{\mu}))\dot{\boldsymbol{\theta}}(t,\boldsymbol{\mu}) - \boldsymbol{f}(\Phi^{\uparrow}(\boldsymbol{\theta}(t,\boldsymbol{\mu}));\boldsymbol{\mu}),$$

其中， $\boldsymbol{f}$ 是半离散全模型 (2.4) 的右端函数，而 $\boldsymbol{J}(\boldsymbol{\theta}(t,\boldsymbol{\mu}))$ 是在 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$ 处计算的 $\Phi^{\uparrow}$ 的雅可比矩阵。回顾一下，对于标量值非线性参数化，式 (7.1) 中定义的残差函数取决于 $\hat{q}$ 关于 $\boldsymbol{\theta}(t,\boldsymbol{\mu})$ 的梯度；在半离散情形下，该梯度变为与采样点对应的批量梯度 (7.14)。这里，由于参数化是在 N 个坐标 $\boldsymbol{x}_1$、……、$\boldsymbol{x}_N$ 上取向量值，因此 $\boldsymbol{J}(\boldsymbol{\theta}(t,\boldsymbol{\mu}))$ 是 $\Phi^{\uparrow}$ 的雅可比矩阵。

随后，通过最小化向量值残差函数的范数，得到时间导数 $\dot{\boldsymbol{\theta}}(t,\ \boldsymbol{\mu})$,

$$\dot{\boldsymbol{\theta}}(t,\boldsymbol{\mu}) \in \operatorname*{arg\,min}_{\boldsymbol{\eta}\in\mathbb{R}^n} \|\boldsymbol{r}(\boldsymbol{\theta}(t,\boldsymbol{\mu}),\boldsymbol{\eta};\boldsymbol{\mu})\|_2^2 . \tag{7.24}$$

Lee 和 Carlberg （2020）重点研究隐式时间离散，这会产生需要在每个时间步通过数值方法求解的非线性最小二乘问题。可以在残差最小化问题（7.24）（或其离散对应形式）中加入约束，以保持质量、动量和能量等量不变（Lee 和 Carlberg 2021年）。

Th瞬时残差最小化方法由Lee 和 Carlberg提出

(2020)与 Dirac–Frenkel变分原理密切相关，但Lee 和 Carlberg (2020)直接在 N 半离散全模型对应的坐标中进行运算。

基于自编码器的非线性参数化的在线计算成本与经验插值 采用Lee 和 Carlberg (2020中使用的基于自编码器的非线性参数化时，要在在线阶段实现相较于完整模型的运行加速颇具挑战性。尽管用于表示 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$ 的优化变量是低维的，其维数为 $n \ll N$，但残差本身定义在 $\mathbb{R}^N$ 上。因此，最小化问题（7.24）是在全模型状态空间 $\mathbb{R}^N$中求解的。当采用隐式时间离散时，这个问题尤其棘手，因为每个时间步都必须求解一个非线性优化问题。例如，如果使用高斯–牛顿法（或类似方法）求解相应的非线性优化问题，那么在每次优化迭代中，权重 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$ 都会通过解码器 $\Phi^{\uparrow}$ 映射为维数为 N 的全模型状态近似，随后需要使用全模型右端函数 $\boldsymbol{f}$ 计算所有 N 个分量的残差。此外，还要针对完整模型所定义的所有 N 个空间坐标计算雅可比矩阵。

为了计算残差而提升回高维全模型状态空间的问题，类似于线性模型降阶中的提升瓶颈；当全模型动力学关于状态呈非线性时，情况尤为如此；参见第2.4节。相应地，已有研究提出了与经验插值类技术兼容的自编码器参数化。由 Kim 等人 (2022年)提出使用浅层掩码解码器函数，并将其与类似于经验插值的稀疏采样方法结合。研究 Romor、Stabile 和 Rozza(2023)更进一步，提出了两种互补机制以降低在线计算成本。首先，仅在选定的少数配点上执行时间离散残差最小化；这种方法以 Chen、Gottlieb、Ji 和 Maday (2021年)提出的超配置方法为基础，以稀疏地评估解码器函数。其次，采用压缩解码器，使其仅计算配点集上的值，从而避免在在线阶段对与全模型状态对应的所有 N 空间坐标进行求值。

注 7.4。 有些方法会训练神经网络自编码器来表示解场，并进一步从数据中学习降阶动力学，而不是通过瞬时残差最小化计算在线权重（Xu 和 Duraisamy 2020、Maulik、Lusch 和 Balaprakash 2021年、Fresca 和 Manzoni 2022年）。我们将在第 8 节简要展望这类非侵入式模型降阶方法。

7.7.2. 具有先验特征映射与二次近似的自编码器我们现在综述基于自编码器的非线性模型降阶技术，其解码器通过先验选定的固定非线性特征映射来表述，而编码器函数是线性的。我们尤其关注多项式特征

映射，尤其是二次特征映射，它们会导出所谓的二次（试探）流形。

采用先验选定的固定特征映射的自编码器 回顾第 7.7.1 节，自编码器由编码器 $\Phi^{\downarrow}(\cdot;\boldsymbol{\theta}^{\downarrow}{}_{\text{off}}) : \mathbb{R}^N \to \mathbb{R}^n$ 和解码器$\Phi^{\uparrow}(\cdot;\boldsymbol{\theta}^{\uparrow}{}_{\text{off}}) : \mathbb{R}^n \to \mathbb{R}^N$组成，它们分别依赖于可训练的离线权重 $\boldsymbol{\theta}^{\downarrow}{}_{\text{off}}$和 $\boldsymbol{\theta}^{\uparrow}{}_{\text{off}}$。此外，回顾一下，我们可以将解码器 $\Phi^{\uparrow} : \mathbb{R}^n \to \mathbb{R}^N$ 解释为诱导出式（7.23）中的向量值非线性参数化。

现在我们考虑具有以下形式的解码器 $\Phi^{\uparrow}$，等价地，也就是向量值非线性参数化

$$\Phi^{\uparrow}(\boldsymbol{\theta};\boldsymbol{\theta}^{\uparrow}_{\text{off}}) = \boldsymbol{V}_1\boldsymbol{\theta} + \boldsymbol{V}_2 g(\boldsymbol{\theta}), \tag{7.25}$$

其中 $\boldsymbol{V}_1 \in \mathbb{R}^{N\times n}$和 $\boldsymbol{V}_2 \in \mathbb{R}^{N\times S}$是可离线训练的矩阵，即 $\boldsymbol{\theta}^{\uparrow}{}_{\text{off}} = [\text{vec}(\boldsymbol{V}_1); \text{vec}(\boldsymbol{V}_2)]$。函数 $g : \mathbb{R}^n \to \mathbb{R}^S$是一个先验给定的非线性特征映射函数。采用试探形式（7.25）有两个目的：(i) 通过非线性项（如截断泰勒展开）丰富由 $\boldsymbol{V}_1$ 的列张成的线性子空间；(ii) 对于某些应用，Cohen 等人 (2024) 证明，高阶 POD 模态的系数可以表示为低阶模态的函数。

一种常用于非线性模型降阶中的特征映射 g，是收集所有不重复二次单项式的二次函数

$$g(\boldsymbol{\theta}) = \boldsymbol{\theta}\hat{\otimes}\boldsymbol{\theta}, \tag{7.26}$$

其中 $\hat{\otimes}$ 表示对称（无重复）Kronecker 积，即包含以下乘积的向量：$\{\theta_i\theta_j\}_{1\le i\le j\le n}$所有 n分量的 $\boldsymbol{\theta}$，并按固定顺序排列。对于这种二次特征映射，特征数量为 $S = n(n+1)/2$。我们可以这样理解（7.25）中的两项：第一项 $\boldsymbol{V}_1\boldsymbol{\theta}$ 提供线性近似，第二项 $\boldsymbol{V}_2 g(\boldsymbol{\theta})$ 是非线性修正。形式如（7.25）的非线性参数化会形成一个试探流形；当特征映射为二次映射时，这种流形有时称为二次流形。

Jain、Tiso、Rutzmoser 和 Rixen（2017年），Rutzmoser、Rixen、Tiso 和 Jain（2017年），Cenedese、Ax˚as、B¨auer-lein、Avila 和 Haller（2022年），Barnett 和 Farhat（2022年），以及 Geelen、Wright 和 Willcox（2023b）开展了二次近似和多项式逼近的研究，并促成了一系列关于用于非线性模型降阶的二次近似的出版物，其中包括在快照数据上训练二次近似的方法（Geelen、Balzano 和 Willcox 2023a，Geelen、Balzano、Wright 和 Willcox 2024，Schwerdtner 和 Peherstorfer 2024，Schwerdtner、Mohan、Pachalieva、Bessac、O'Malley 和 Peherstorfer 2025b），以及在在线阶段高效使用这些近似的方法（Benner、Goyal、Heiland 和 Pontes Duff 2023，Sharma、Mu、Buchfink、Geelen、Glas 和 Kramer 2023，Goyal 和 Benner 2024，Yıldız、Goyal、Bendokat 和 Benner 2024，Weder 等 2025，Glas 和 Mu 2025）。较早的 Gu（2012）也讨论了模型降阶背景下基于多项式特征映射的非线性参数化。总的来说，我们指出，二次

系统和二次动力学，以及对相应张量结构的利用，在模型降阶中无处不在；例如参见 Gu(2011)、Benner 和 Breiten (2015)、Schlegel 和 Noack (2015)、Benner、Goyal 和 Gugercin (2018)、Kramer 和 Willcox (2019)、Qian、Kramer、Peherstorfer 和 Willcox(2020）。有理流形的扩展由 Klein、Sanderse、Costa、Pecnik 和 Henkes (2025）。

形如（7.25）的非线性参数化的表达能力由先验选定的特征映射 g主导。Buchfink、Glas和Haasdonk（2024*a*）给出了Kolmogorov n-宽度的下界，这些下界针对具有多项式特征映射 g的此类非线性参数化。特别地，粗略而言，Buchfink 等人（2024*a*）表明，对于第3.2节中研究的线性平流示例， n-宽度的下界表现为 $\mathcal{O}\left(n^{-\alpha/2}\right)$，其中 α 是多项式的次数。

依据有关稳定流形宽度和感知数的近似理论结果，见第4.2节，并由Barnett和Farhat（2022年）、Geelen 等人（2023*b*）、Cohen等人（2024）加以推广，通常选择使用线性编码器 $\Phi^{\downarrow}(\cdot;\, \boldsymbol{\theta}^{\downarrow}\,\text{off}) \colon \mathbb{R}^N \to \mathbb{R}^n$，即，

$$\Phi^{\downarrow}(\boldsymbol{q}) = \boldsymbol{W}^{\top}\boldsymbol{q}\,, \tag{7.27}$$

对于某个矩阵 $\boldsymbol{W} \in \mathbb{R}^{N\times n}$。尽管可以根据 Buchfink 等人（2024b）提出的通用框架推断矩阵 $\boldsymbol{W}$ 的一般条件，但一种常见的选择是使用 $\boldsymbol{W}^{\top} = \boldsymbol{V}_1^{+}$，它对应于矩阵 $\boldsymbol{V}_1$ 的 Moore–Penrose 伪逆。注意，如果 $\boldsymbol{V}_1$ 具有正交归一列，那么 $\boldsymbol{V}_1^{+} = \boldsymbol{V}_1^{\top}$。Cohen 等人（2024）、Barnett、Farhat 和 Maday（2023b,a）以及 Ballout、Maday 和 Prud'homme（2024）还将线性编码器与比二次解码器更通用的非线性解码器结合使用。具体而言，Cohen 等人（2024）从感知数量的角度研究了这类线性编码器和非线性解码器对，并证明，对于表示输运特征的函数，其近似误差相对于维数 n 的 Kolmogorov n-宽度可以显著更快地衰减。Bensalah、Nouy 和 Soffo（2025）也采用了通过线性编码器和非线性解码器实现的类似策略。此外，经验观察还表明，线性编码器能够达到与更通用的非线性编码器相似的重构误差，同时还具有训练鲁棒性更高的额外优势（Glas 和 Unger 2025，Schwerdtner、Gugercin 和 Peherstorfer 2025a）。

利用结构形式高效训练快照数据回顾线性编码器（7.27）及其特定选择 $\boldsymbol{W}^{\top} = \boldsymbol{V}_1^{+}$，以及非线性解码器对（7.25）及其可训练矩阵 $\boldsymbol{V}_1 \in \mathbb{R}^{N\times n}$和 $\boldsymbol{V}_2 \in \mathbb{R}^{N\times S}$，它们构成了离线权重 $\boldsymbol{\theta}^{\downarrow}\,\text{off} = \text{vec}(\boldsymbol{V}_1)$ 和 $\boldsymbol{\theta}^{\uparrow}\,\text{off} = [\text{vec}(\boldsymbol{V}_1); \text{vec}(\boldsymbol{V}_2)]$。还要回顾，特征映射 g 是先验给定的。

A 训练此类非线性参数化的典型策略是首先

选择矩阵 $\boldsymbol{V}_1$，然后通过线性最小二乘问题拟合 $\boldsymbol{V}_2$ ，

$$\boldsymbol{V}_2 = \underset{\boldsymbol{Z}\in\mathbb{R}^{N\times S}}{\arg\min} \|\boldsymbol{Q}_{\text{train}} - \boldsymbol{V}_1 \underbrace{\Phi^{\downarrow}(\boldsymbol{Q}_{\text{train}}; \boldsymbol{\theta}^{\downarrow}_{\text{off}})}_{\boldsymbol{V}_1^{+}\boldsymbol{Q}_{\text{train}}} - \boldsymbol{Z} g(\underbrace{\Phi^{\downarrow}(\boldsymbol{Q}_{\text{train}}; \boldsymbol{\theta}^{\downarrow}_{\text{off}})}_{\boldsymbol{V}_1^{+}\boldsymbol{Q}_{\text{train}}})\|_F^2 + \lambda\|\boldsymbol{Z}\|_F^2 , \tag{7.28}$$

这对应于快照的重构误差 $\boldsymbol{Q}_{\text{train}} - \Phi^{\uparrow}(\Phi^{\downarrow}(\boldsymbol{Q}_{\text{train}}; \boldsymbol{\theta}^{\downarrow}\text{关闭}); \boldsymbol{\theta}_{\text{off}})$，这些快照 ↑(2.9）作为列存储在快照矩阵 $\boldsymbol{Q}_{\text{train}}$中。编码器和特征映射按列应用于快照矩阵。最小二乘问题中加入了正则化项，其正则化参数为 $\lambda > 0$。

现在考虑拟合矩阵 $\boldsymbol{V}_1$，它完全决定编码器，同时也出现在解码器的线性项中。Barnett 和 Farhat（2022）、Geelen 等（2023b）的研究采用这样的解释：根据（7.25）的定义， $\Phi^{\uparrow}(\boldsymbol{\theta}; \boldsymbol{\theta}^{\uparrow}\text{ off}) = \boldsymbol{V}_1\boldsymbol{\theta} + \boldsymbol{V}_2 g(\boldsymbol{\theta})$ 是一种线性近似 $\boldsymbol{V}_1\boldsymbol{\theta}$ ，并带有一个非线性校正项$\boldsymbol{V}_2 g(\boldsymbol{\theta})$。因此，Barnett 和 Farhat（2022）、Geelen 等（2023b）通过将矩阵 $\boldsymbol{V}_1 \in \mathbb{R}^{N\times n}$ 的列设为快照矩阵 $\boldsymbol{Q}_{\text{train}}$的前$n$ 个主导左奇异向量，来拟合 $\boldsymbol{V}_1$ ，使相应的线性项 $\boldsymbol{V}_1\boldsymbol{\theta}$ 在快照数据上达到最低的线性近似误差。然而，Cohen 等人（2024）以及 Schwerdtner 和 Peherstorfer（2024）强调，非线性校正项会在编码点$\boldsymbol{\theta} = \Phi^{\downarrow}(\boldsymbol{q}; \boldsymbol{\theta}^{\downarrow}\text{ off}) = \boldsymbol{V}_1^{+}\boldsymbol{q}$ 处计算特征映射 g 。因此，单纯拟合 $\boldsymbol{V}_1$ 以最小化线性近似的近似误差，可能导致编码 $\Phi^{\downarrow}(\boldsymbol{q}; \boldsymbol{\theta}^{\downarrow}\text{ off})$丢弃对于特征映射 g 提供有效信息并实现有效校正至关重要的信息。为此，Schwerdtner 和 Peherstorfer（2024）提出了一种贪心策略，逐列构建 $\boldsymbol{V}_1$ 。该策略不再将 $\boldsymbol{V}_1$ 固定为由快照矩阵 $\boldsymbol{Q}_{\text{train}}$的前 n 个左奇异向量（主成分）构成，而是采用 Schwerdtner 和 Peherstorfer（2024）提出的贪心方法，从更大的候选集中选择 n 基向量；该集合包含靠前及靠后的主成分，其索引可能大于 n。在贪心过程中，每一步都会选取一个向量：在通过线性最小二乘问题（7.28）拟合 $\boldsymbol{V}_2$ 后，该向量能够最大程度地改善重构。因此，最终得到的 $\boldsymbol{V}_1$ 可能包含非主导左奇异向量（即索引大于 n的向量），前提是这些向量对于提升特征映射的有效性、降低目标函数（7.28）中的残差范数十分重要。

在另一项研究中，Geelen 等人(2023*a*，2024)提出优化问题，以构造 $\boldsymbol{V}_1$ 和 $\boldsymbol{V}_2$，从而也可能得到近似误差更低的矩阵，而不是通过快照矩阵的 $\boldsymbol{V}_1$ 的 n 主导奇异向量来定义该矩阵。

利用结构形式提升在线阶段的效率 形式为（7.25）的二次（或多项式）近似有一个关键优势：其结构高度规整，可在在线阶段的残差最小化中加以利用。

Re将与瞬时7.16残差最小化相对应的半离散最小二乘问题称为（ -

基于 Dirac–Frenkel变分原理的残差最小化。此外，回想一下，由于 $\Phi^{\uparrow}: \mathbb{R}^n \to \mathbb{R}^N$是向量值非线性参数化，批量梯度会变成雅可比矩阵。对于具有二次特征映射的参数化（7.26），雅可比矩阵 $\boldsymbol{J}(\boldsymbol{\theta}(t,\ \boldsymbol{\mu}) \in \mathbb{R}^{N\times n}$ 关于 $\Phi^{\uparrow}$ 存在闭式表达式，且关于 $\boldsymbol{\theta}(t,\ \boldsymbol{\mu})$ 为仿射函数。更准确地说，由于 $\partial(\boldsymbol{\theta}\hat{}\otimes\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ 关于 $\boldsymbol{\theta}$是线性的，因此可以写成

$$\boldsymbol{J}(\boldsymbol{\theta}(t,\boldsymbol{\mu})) = \boldsymbol{V}_1 + \sum_{\ell=1}^{n} \theta_\ell(t,\boldsymbol{\mu})\,\boldsymbol{J}_\ell, \tag{7.29}$$

用于矩阵 $\boldsymbol{J}_\ell \in \mathbb{R}^{N\times n}$，这些矩阵可在离线阶段预先计算，且与在线权重向量 $\boldsymbol{\theta}$ 无关（$t,\ \boldsymbol{\mu}$）。

当完整模型关于状态是线性的，且具有仿射参数依赖时，可在在线阶段组装瞬时残差最小化最小二乘问题的目标函数，而无需进行任何计算量随环境维度 N增长的计算；Sharma等人（2023）和Weder等人(2025)展示并利用了这种离线–在线分离以及与 N无关的在线组装。当完整模型具有非线性状态依赖时，预先计算雅可比矩阵分量仍然可能有用。具体而言，Weder等人(2025)提出了一种二次近似公式，它不依赖于 N 全模型空间坐标，而是可以在一组选定的配点上实现，从而避免更高昂的在线计算成本。

Barnett 和 Farhat（2022年）提出了另一种利用二次结构的方法：将二次参数化代入隐式时间最小二乘 Petrov–Galerkin 离散化中。由于二次参数化具有闭式仿射雅可比矩阵，因此可以针对它设计稀疏采样（类似于经验插值；见第2.4节），并高效应用。因此，在在线阶段，残差和梯度评估仅限于少数选定点；在所展示的数值实验中，与完整模型相比，这带来了在线运行时间的加速。

Geelen 等人 （2023*b*）进一步利用结构，将雅可比矩阵（7.29）近似为其一阶近似，也就是取值为 $\boldsymbol{V}_1$的常数函数。在这种情况下，雅可比矩阵与在线权重 $\boldsymbol{\theta}$（$t,\ \boldsymbol{\mu}$）无关，因此在线阶段的降阶模型评估速度更快；在线运行时间方面的比较见Weder 等人（2025）。此外，Geelen 等人（2023*b*）表明，将雅可比矩阵近似为其一阶近似 $\boldsymbol{V}_1$ 是推导一种基于二次流形的非侵入式降阶建模方法的关键步骤；关于侵入式与非侵入式模型降阶的更详细讨论见第8 节。

二次（以及其他先验固定特征映射近似）在雅可比矩阵（7.29）中所呈现的结构，与诸如神经网络这类通用自编码器参数化所对应的梯度形成鲜明对比；后者并不具备

可比的低阶结构，因此必须通过自动微分在线反复求得，甚至可能需要在每个时间步的每次高斯-牛顿迭代中求得。不过，我们强调，先验固定的特征映射往往会导向表达能力低于深度神经网络结合表示学习所形成的通用架构的非线性参数化。关于这一背景下较低的逼近界，另请参阅Buchfink等人（2024*a*）。

7.7.3. 采用离线–在线权重的神经表示

现在我们来考虑神经表示，它使用神经网络将非线性参数化直接表示为 $\hat{q}$。

模型降阶中解函数的神经表示 解场的神经网络参数化有时也称为隐式神经表示，这一概念起源于计算机图形学界（Sitzmann、Martel、Bergman、Lindell 和 Wetzstein2020），但也已用于对物理场解进行降阶（压缩），例如参见 Pan、Brunton 和 Kutz (2023）。

神经表示的一项关键特性是，可以利用来自不同空间离散的快照数据对其进行训练，因为它们将场表示为空间坐标的函数；这与上文讨论的典型神经网络自编码器形成对比，后者依赖于固定的空间离散。当全模型解是在自适应网格或点云上计算时，这一特性尤其有利；Chen、Xiang、Cho、Chang、Pershing、Maia、Chiaramonte、Carlberg 和 Grinspun（2023b）在非线性模型降阶中就利用了这一点。

在模型降阶中使用神经表示时，一个重要的设计选择是如何将神经网络的权重分解为离线权重向量$\boldsymbol{\theta}_{\text{off}}$ 和在线权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$：前者可以针对所有快照数据仅训练一次，后者的分量较少，可通过例如瞬时残差最小化等方法在线获得。基于权重的离线–在线分解，可以将非线性参数化表示为 $\hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \cdot; \boldsymbol{\theta}_{\text{off}}) : \Omega \to \mathbb{R}$，这就是 (4.7) 中给出的形式，其中 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 对应网络的在线权重，可能随时间 t 和参数 $\boldsymbol{\mu}$变化。神经网络的输入是 Ω 中的空间坐标。而离线权重 $\boldsymbol{\theta}_{\text{off}}$ 与时间 t和参数 $\boldsymbol{\mu}$ 无关，并且在在线阶段保持固定。

采用离线–在线权重的神经表示 实现神经网络权重离线–在线分解的一种方法，是让在线权重通过调制离线权重的方式参与计算。一种常见的权重调制方法是采用低秩矩阵，见（Hu, yelong shen, Wallis, Allen- Zhu, Li, Wang, Wang 和 Chen 2022年），由此得到低秩适配层。针对模型降阶，Berman 和 Peherstorfer （2024）提出了连续低秩适配（CoLoRA）层，其中调制可连续地依赖于时间和参数 $\boldsymbol{\mu}$。

CoLoRA 层的形式为

$$\mathcal{C}_\ell(\boldsymbol{y}) = \boldsymbol{W}_\ell \boldsymbol{y} + \alpha_\ell(t, \boldsymbol{\mu}) \boldsymbol{A}_\ell \boldsymbol{B}_\ell \boldsymbol{y} + \boldsymbol{b}_\ell \,, \tag{7.30}$$

其中 $\boldsymbol{W}_\ell$ 是一个权重矩阵，$\boldsymbol{b}_\ell$ 是一个偏置项。此外，还有一个秩为 r 、秩-r 因子为 $\boldsymbol{A}_\ell$ 和 $\boldsymbol{B}_\ell$ 的低秩矩阵 $\boldsymbol{A}_\ell \boldsymbol{B}_\ell$ ，以及调制$\alpha_\ell(t, \boldsymbol{\mu}) \in \mathbb{R}$。这些层（7.30）可以组合成 CoLoRA 网络，形式如下：

$$\hat{q}(\boldsymbol{\theta}(t, \boldsymbol{\mu}), \boldsymbol{x}; \boldsymbol{\theta}_{\text{off}}) = \mathcal{C}_L(\sigma(\mathcal{C}_{L-1}(\ldots \sigma(\mathcal{C}_1(\boldsymbol{x}))))) \,, \tag{7.31}$$

这是一种具有（4.7）形式的非线性参数化。离线权重向量 $\boldsymbol{\theta}_{\text{off}}$ 由网络（7.31）中所有与时间 t和参数 $\boldsymbol{\mu}$无关的权重组成，包括所有权重矩阵 $\boldsymbol{W}_1$、……、 $\boldsymbol{W}_L$，低秩矩阵$\boldsymbol{A}_1$、……、 $\boldsymbol{A}_L$， $\boldsymbol{B}_1$、……、 $\boldsymbol{B}_L$，以及偏置项 $\boldsymbol{b}_1$、……、 $\boldsymbol{b}_L$。在线权重向量由以下调制组成

$$\boldsymbol{\theta}(t, \boldsymbol{\mu}) = \big[\alpha_1(t, \boldsymbol{\mu}), \ldots, \alpha_L(t, \boldsymbol{\mu})\big]^\top \in \mathbb{R}^L \,.$$

在（7.31）所示的 CoLoRA 网络特定架构中，层数 L 决定了在线权重的数量 n ，即 $n = L$。CoLoRA 网络还有其他变体，例如，将部分 L CoLoRA 层替换为没有调制的常规线性层，从而可以实现 $n < L$个在线权重。此外，还可以通过引入对角矩阵，在每层引入多个调制，以获得

$$\mathcal{C}_\ell(\boldsymbol{y}) = \boldsymbol{W}_\ell \boldsymbol{y} + \boldsymbol{A}_\ell \operatorname{diag}(\alpha_\ell^{(1)}(t, \boldsymbol{\mu}), \ldots, \alpha_\ell^{(r)}(t, \boldsymbol{\mu})) \boldsymbol{B}_\ell \boldsymbol{y} + \boldsymbol{b}_\ell \,,$$

其中低秩因子 $\boldsymbol{A}_\ell$ 和 $\boldsymbol{B}_\ell$ 的秩为 r。这类变体见Berman 和 Peherstorfer（2024）的讨论。一般而言，权重调制的方式有很多种（Perez、Strub、de Vries、Dumoulin 和 Courville，2018年）。

除了权重调制之外，还有其他方式可以构建具有离线与在线权重的神经网络。例如，Chen等人 （2023*b*）使用多层感知机来表示 $\hat{q}$ ，并将潜在状态作为输入传入网络。潜在状态类似于我们所说的在线权重向量$\boldsymbol{\theta}$（t， $\boldsymbol{\mu}$），但它是作为输入传入的。按照我们的术语，多层感知机的权重就是离线权重。

基于快照数据训练神经表示的离线权重 直接训练离线权重向量 $\boldsymbol{\theta}_{\text{off}}$ 和在线权重向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$ 在所有时间和参数上，会使在线权重随时间任意变化，从而可能导致在线权重对时间的依赖不平滑。为促进在线权重在时间上的正则性 t，Berman 和 Peherstorfer (2024) 在训练过程中使用一种所谓的调制（或超）网络：设 $h_{\boldsymbol{\psi}}\colon [0, T] \times \mathcal{D} \to \mathbb{R}^n$是一个将时间 t 和参数 $\boldsymbol{\mu}$ 映射为一个 n维向量 $\boldsymbol{\theta}(t, \boldsymbol{\mu})$。网络 $h_{\boldsymbol{\psi}}$ 的可训练权重向量为 $\boldsymbol{\psi} \in \mathbb{R}^{n'}$，其维度通常较低。

使用网络 h_ψ 会得到以下训练问题

$$\min_{\theta_{\text{off}},\psi} \frac{1}{Mm(K+1)} \sum_{i=1}^{M} \sum_{j=1}^{m} \sum_{k=0}^{K} \frac{|q(t_k, \boldsymbol{x}_j; \boldsymbol{\mu}_i) - \hat{q}(h_\psi(t_k, \boldsymbol{\mu}_i); \boldsymbol{x}_j; \boldsymbol{\theta}_{\text{off}})|^2}{|q(t_k, \boldsymbol{x}_j; \boldsymbol{\mu}_i)|^2} \tag{7.32}$$

关于离线权重 $\boldsymbol{\theta}_{\text{off}}$ 以及网络 $\boldsymbol{\psi}$ 的权重h_ψ。问题（7.32）取决于 M 训练参数 $\boldsymbol{\mu}_1$，……，$\boldsymbol{\mu}_M$ 以及 $K+1$ 时间步 $0 = t_0 < t_1 < \cdots < t_K = T$，与训练数据（2.9）中的情况相同。此外，我们使用 m 空间点$\boldsymbol{x}_1$，……，$\boldsymbol{x}_m \in \Omega$。回顾一下，$q(t_k, \boldsymbol{x}_j; \boldsymbol{\mu}_i)$ 是时间t_k和参数 $\boldsymbol{\mu}_i$在空间坐标 $\boldsymbol{x}_j$ 处的全模型解（2.2）q。具体而言，在训练神经表示时，可以直接使用全模型解函数，而非半离散解向量；这突显了神经表示独立于完整模型所使用的网格和离散化。

调制网络h_ψ 的作用可以看作类似于基于自编码器的参数化中的编码器函数。然而，主要区别在于，调制网络接收时间 t 和参数 $\boldsymbol{\mu}$ （以及可能的其他输入）作为输入，但不接收全模型状态向量。

使用训练后的神经表示进行在线残差最小化 训练完成后，可在在线阶段将神经表示与瞬时残差最小化相结合。

与先前考虑的、针对半离散完整模型构建的非线性参数化截然不同，后者的计算结果是维数为 N 的向量（参见第7.7.1 节和第7.7.2节），神经表示是坐标到数值映射，其形式与我们在（4.1）中定义的非线性参数化完全一致（采用离线权重时则对应（4.7））。因此，在应用基于 Dirac–Frenkel变分原理的格式进行残差最小化时，可以使用一组 m 采样点对内积和残差范数进行数值估算，其中 m 的选取可以独立于全模型维数 N 及其网格。具体而言，Berman 和 Peherstorfer (2024) 在在线阶段采用神经伽辽金格式。此外，通过在瞬时残差最小化中加入约束和非线性嵌入步骤，可以守恒质量、动量和能量（Schwerdtner等2024）。

在在线阶段的每个时间步，如果采用显式时间积分格式，则必须求解一个线性最小二乘问题（Berman 和 Peherstorfer 2024）。未知量的数量为 n，其目标是在基于快照数据训练参数化时，使 n 保持较小。因此，在线计算成本主要取决于用于近似最小二乘问题中残差范数的采样点数量 m，以及批量梯度的计算和采样点处完整模型右端函数的评估。特别是，完整模型右端函数可能包含空间导数，这些导数可以通过在采样点处对神经表示进行自动微分来获得。

由于在线阶段是在 m 采样点上运行，而不是在对应于N 半离散全模型的 N维状态向量的点上运行，因此避免了基于半离散全模型构建的降阶模型中常见的经典提升瓶颈。同时，精度取决于是否选择了数量充足且信息充分的采样点。特别是，尽管采样点数量 m 与完整模型网格点数量 N 在形式上相互独立，但二者在本质上是相关的：如果完整模型需要细网格来解析细尺度解特征，那么采样点数量 m 通常也必须对此有所反映，尽管自适应性有助于使 m 保持较小。这促使人们采用自适应采样策略，正如第 7.3.3节在神经伽辽金格式的背景下所讨论的那样。

注 7.5。 在线残差最小化的一种替代方法是，在在线阶段重新使用经过训练的调制网络 h_ψ 来计算在线权重：$\theta(t,\ \mu) = h_\psi(t,\ \mu)$，从而只需评估一个神经网络即可获得解的近似。由于调制网络 h_ψ 仅依赖于诸如 t 和 μ 等已知输入（而不是像自编码器中的编码器函数那样依赖于完整状态），因此得到的在线阶段是完全非侵入式且数据驱动的，但它失去了与狄拉克–弗伦克尔残差最小化相关的伽辽金最优性。关于通过调制网络或超网络在线计算权重的非侵入式模型降阶相关方法，可参见例如 de Avila Belbute-Peres、fan Chen 和 Sha (2021年)、Cho、Lee、Rim 和 Park (2023)、Berman 和 Peherstorfer (2024)；第 8 节将进一步讨论。

8. 结论与展望

输运主导问题的模型降阶已取得显著进展，这从大量且迅速增长的文献中可见一斑。许多非线性模型降阶方法都可以统一地理解为：将通用的非线性参数化与瞬时残差最小化相结合。与此同时，我们有意将基于变换的方法和在线自适应方法与这一通用视角区分开来，因为这种分类凸显了可加以利用、以提升效率和鲁棒性的其他结构特征。

尽管本综述所引文献涵盖广泛，但一些重要研究方向仍未得到充分关注，包括保结构模型降阶、随机系统的降阶，以及面向控制应用的模型降阶。另一条截然不同且仍处于早期阶段的研究路线，是将双曲型偏微分方程重新表述为时滞方程，再对这种替代表述应用模型降阶。关于这一方向的初步进展，可参见 Schulze、Unger、Beattie 和 Gugercin (2018年)。

一个特别富有成果且发展迅速的领域，是针对输运主导问题的非侵入式模型降阶。与本综述着重讨论的侵入式方法不同，非侵入式模型降阶方法旨在学习低维表示及相应的降阶动力学，主要依据快照数据，而

不是直接从控制方程推导降阶动力学。我们强调，非侵入式模型降阶不同于方程发现；方程发现利用数据推断控制物理定律本身。关于非侵入式模型降阶已有大量文献；综述可参见 Ghattas 和 Willcox (2024), Kramer等人(2024) 的综述。

有几种非侵入式方法专门针对输运主导区域，在这些区域中，线性模型降阶效率不高。一类方法先通过非线性变换分离出输运效应，再应用线性非侵入式模型降阶方法（Sesterhenn 和 Shahirpour，2019，Sarna 和 Benner，2021年，Lu 和 Tartakovsky，2020，Issan 和 Kramer，2023年，Gowrachari、Demo、Stabile 和 Rozza，2025），这与我们上文回顾的基于变换的侵入式方法类似。另一类非侵入式方法学习非线性低维表示（例如，通过自编码器），随后在潜在表示中使用单独的多项式模型或神经网络模型学习降阶动力学（Fresca、Dede' 和 Manzoni，2021年，Maulik 等，2021年，Fresca 和 Manzoni，2022年，Franco、Manzoni 和 Zunino，2023，Cenedese 等，2022年，Geelen 等，2023*b*，Regazzoni、Pagani、Salvador、Dede' 和 Quarteroni，2024，Tomasetto、Williams、Braghin、Manzoni 和 Kutz，2025）。需要强调的是，本展望中引用的文献仅展现了非侵入式模型降阶这一快速发展的领域中的部分研究。引用这些文献旨在说明非侵入式方法在非线性模型降阶中的应用前景，而非提供全面综述。

致谢

B.P. 部分获得了美国国家科学基金会项目（编号 2046521）和美国空军科学研究办公室（AFOSR，美国）奖项（编号 FA9550- 24-1-0327）的资助。B.U. 获得德国研究联合会（DFG）项目（项目编号 258734477，SFB 1173）以及德国联邦教育与研究部（BMBF）项目（编号 05M22VSA）的资助。

参考文献

R. Abgrall, D. Amsallem and R. Crisovan (2016), 通过 l^1-范数最小化实现稳健模型降阶并利用字典进行近似：在非线性双曲型问题中的应用，*Adv. Model.and Simul.***3**(1), 1. V. M. Adamyan, D. Z. Arov and M. G. Krein (1971), 汉克尔算子的施密特对的解析性质及广义舒尔–高木问题，*Math. USSR Sbornik***15**(1), 31–73. J. Aghili, J. Z. Atokple, M. Billaud-Friess, G. Garnier, O. Mula and N. Tognon (2025), 用于滤波问题的动力学神经伽辽金格式，*ESAIM: ProcS* **81**, 2–15. A. Aitzhan, A. G. Nouri, P. Givi and H. Babaee (2025), 低秩矩阵流形上的湍流反应流降阶建模，*J.Comput.Phys.***521**, 113549. M. Alireza Mirhoseini and M. J. Zahr (2023), 通过基于优化的隐式特征跟踪对对流主导偏微分方程进行模型降阶，*J. Comput.Phys.***473**, 111739.

D. Amsallem and C. Farhat (2008), 用于调整降阶模型的插值方法及其在气动弹性中的应用，*AIAA Journal* **46**(7), 1803–1813. D. Amsallem and B. Haasdonk (2016), PEBL-ROM：基于投影误差的局部降阶模型，*Adv. Model. andSimul.***3**(1), 1–25. D. Amsallem, M. J. Zahr and C. Farhat (2012), 基于局部降阶基的非线性模型降阶，*Internat. J. Numer. Methods Engrg.***92**(10), 891–916. D. Amsallem, M. J. Zahr and K. Washabaugh (2015), 通过快速局部降阶基更新，高效降低带有超降阶的非线性系统的阶数，*Adv.Comput.Math.***41**(5), 1187–1230. W. Anderson and M. Farazmand (2022), 守恒量偏微分方程非线性降阶解的演化，*SIAM J. Sci. Comput.***44**(1), A176–A197. W. Anderson and M. Farazmand (2024), 形状变化非线性解的快速可扩展计算及其在演化神经网络中的应用，*J. Comput. Phys.***498**, 112649. A. C. Antoulas (2005), 大规模动力系统的近似, Advances in Design and Control, 工业与应用数学学会, 费城, 宾夕法尼亚州, 美国. A. C. Antoulas, C. Beattie and S. Gugercin (2020), 模型降阶的插值方法, Computational Science & Engineering, 工业与应用数学学会, 费城, 宾夕法尼亚州, 美国. A. C. Antoulas, D. C. Sorensen and S. Gugercin (2001), 大规模系统模型降阶方法综述, *Contemp. Math.***280**, 193–220. F. Arbes, C. Greif and K. Urban (2025), 线性输运的 Kolmogorov n-宽度：精确表示及数据的影响，*Adv.Comput.Math.* **51**(2), 13. J.-P. Argaud, B. Bouriquet, F. de Caso, H. Gong, Y. Maday and O. Mula (2018), 基于广义经验插值法的核反应堆传感器布置，*J.Comput. Phys.* **363**, 354 – 370. E. Arian, M. Fahl and E. W. Sachs (2000), 用于流动控制的信赖域本征正交分解，Technical Report 2000-25, ICASE. A. Arnold and T. Jahnke (2014), 关于层次 Tucker 格式中高维微分方程的近似，*BIT Numer.Math.***54**(2), 305–341. P. Astrid, S. Weiland, K. Willcox and T. Backx (2004), 本征正交分解模型中的缺失点估计，载于第*43*届 *IEEE* 决策与控制会议 *(CDC)*, Vol. 2, pp. 1767–1772. P. Astrid, S. Weiland, K. Willcox and T. Backx (2008), 本征正交分解模型中的缺失点估计，*IEEE Trans. Automat. Control* **53**(10), 2237–2251. H. Babaee and T. P. Sapsis (2016), 用于描述与有限时间不稳定性相关模态的最小化原理，*Proc. R. Soc.A* **472**(2186), 20150779. I. Babuv̌ska and W. C. Rheinboldt (1978), 自适应有限元计算的误差估计，*SIAMJ. Numer. Anal.***15**(4), 736–754. M. Bachmayr and A. Cohen (2017), 参数椭圆偏微分方程的 Kolmogorov 宽度与低秩近似, *Math.Comp.***86**(304), 701–724. M. Bachmayr, W. Dahmen and M. Oster (2025*a*), 参数偏微分方程的变分一致神经残差回归：论可控精度的可行性，*I MAJ. Numer.Anal*

. pp. 1–50.

M. Bachmayr、H. Eisenmann 和 A. Uschmajew（2025*b*），高维抛物问题的动态低秩张量逼近：空间离散的存在性与收敛性，*Numer. Math.* **157**(3)，781–822。O. Balabanov 和 A. Nouy（2021），用于模型降阶的随机线性代数——第二部分：最小残差方法与基于字典的逼近，*Adv. Comput. Math.***47**(2)，26。M. Balajewicz、I. Tezaur 和 E. Dowell（2016），针对可压缩Navier–Stokes方程的基于投影降阶模型，在 Stiefel流形上进行最小子空间旋转以实现稳定与增强，*J.Comput.Phys.***321**，224–241。H. Ballout、Y. Maday 和 C. Prud'homme（2024），多参数椭圆问题的非线性压缩降阶基逼近，载于*Multiscale, Nonlinear andAdaptive Approximation II*（R. DeVore 和 A. Kunoth 编），Springer Nature Switzerland，沙姆，第55–73页。L. Balzano、R. Nowak 和 B. Recht（2010），根据高度不完整的信息在线辨识和跟踪子空间，载于*Proceedings ofthe 48th AnnualAllerton Con-ference on Communication, Control,and Computing (Allerton)*，IEEE，蒙蒂塞洛，伊利诺伊州，美国，第704–711页。J. Barnett 和 C. Farhat（2022），用于缓解非线性基于投影的模型降阶中 Kolmogorov屏障的二次逼近流形，*J. Comput.Phys.***464**，111348。J. Barnett、C. Farhat 和 Y. Maday（2023*a*），通过神经网络增强的基于投影的模型降阶缓解可约化性方面的Kolmogorov屏障，*J. Comput.Phys.***492**，112420。J. L. Barnett、C. Farhat 和 Y. Maday（2023*b*），利用神经网络增强的降阶模型缓解气动模型降阶中的Kolmogorov屏障，载于*AIAA SCITECH2023Forum*。M. Barrault、Y. Maday、N. C. Nguyen 和 A. T. Patera（2004），一种"经验插值"方法：应用于偏微分方程的高效约化基离散，*C. R. Math. Acad. Sci. Paris* **339**(9)，667–672。B. Battisti、T. Blickhan、G. Enchery、V. Ehrlacher、D. Lombardi 和 O. Mula（2023），用于多孔介质参数化流动问题的 Wasserstein模型降阶方法，*ESAIM: ProcS***73**，28–47。U. Baur、C. Beattie、P. Benner 和 S. Gugercin（2011），参数化模型降阶的插值投影方法，*SIAM J.Sci.Comput.***33**(5)，2489–2518。M. Beck、A. J¨ackle、G. Worth 和 H.-D. Meyer（2000），多构型含时 Hartree（MCTDH）方法：一种高效传播波包的算法，*PhysicsReports* **324**(1)，1–105。Y. Bengio、A. Courville 和 P. Vincent（2013），表示学习：综述与新视角，*IEEE Trans. Pattern Anal. Mach.Intell.***35**(8)，1798–1828。P. Benner 和 T. Breiten（2015），非线性模型降阶的双边投影方法，*SIAMJ.Sci. Comput.***37**(2)，B239–B260。P. Benner 和 S. W. R. Werner（2020），大规模描述符系统的 Hankel范数逼近，*Adv. Comput. Math.* **46**，40。P. Benner、A. Cohen、M. Ohlberger 和 K. Willcox（2017），模型降阶与近似，Advances in Design and Control，工业与应用数学学会，费城

, PA, USA.

P. Benner、P. Goyal 和 S. Gugercin（2018），用于二次双线性控制系统的 $\mathcal{H}_2$-准最优模型降阶，*SIAM J. MatrixAnal.Appl.***39**(2)，983–1032。P. Benner、P. Goyal、J. Heiland 和 I. Pontes Duff（2023），一种用于非线性动力系统非侵入式降阶建模的二次解码器方法，载于*Proceedings inApplied Mathematics and Mechanics 2020*，第23卷。P. Benner、S. Grivet-Talocia、A. Quarteroni、G. Rozza、W. Schilders 和 L. M. Silveira（2021*a*），模型降阶：第*1*卷——系统驱动与数据驱动方法及算法，De Gruyter，柏林、波士顿。P. Benner、S. Grivet-Talocia、A. Quarteroni、G. Rozza、W. Schilders 和 L. M. Silveira（2021*b*），模型降阶：第*2*卷——基于快照的方法与算法，De Gruyter，柏林、波士顿。P. Benner、S. Gugercin 和 K. Willcox（2015），参数化动力系统基于投影的模型降阶方法综述，*SIAM Rev.***57**(4)，483–531。A. Bensalah、A. Nouy 和 J. Soffo（2025），利用复合多项式网络进行非线性流形近似，*arXiv* **2502**(05088)，1–28。M. Berger 和 P. Colella（1989），用于激波流体动力学的局部自适应网格细化，*J.Comput. Phys.* **82**(1)，64–84。M. J. Berger 和 J. Oliger（1984），双曲型偏微分方程的自适应网格细化，*J. Comput. Phys.* **53**(3)，484–512。J. Berman 和 B. Peherstorfer（2023），用于借助深度网络求解演化方程的随机稀疏神经伽辽金格式，载于*Advances in Neural InformationProcessing Systems*（A. Oh、T. Naumann、A. Globerson、K. Saenko、M. Hardt 和 S. Levine 编），第36卷，Curran Associates, Inc.，第4097–4114页。J. Berman 和 B. Peherstorfer（2024），CoLoRA：用于参数化偏微分方程降阶隐式神经建模的连续低秩适配，载于*Proceedingsof the 41st International Conference on Machine Learning*（R. Salakhutdinov、Z. Kolter、K. Heller、A. Weller、N. Oliver、J. Scarlett 和 F. Berkenkamp 编），机器学习研究论文集第235卷，PMLR，第3565–3583页。J. Berman、P. Schwerdtner 和 B. Peherstorfer（2024），第8章——利用深度网络按时间顺序求解偏微分方程的神经伽辽金格式，载于*Numerical Analysis Meets MachineLearning*（S. Mishra 和 A. Townsend 编），《数值分析手册》第25卷，爱思唯尔，第389–418页。W. J. Beyn 和 V. Th¨ummler（2004），等变演化方程的解冻结，*SIAM J.Appl. Dyn. Syst.* **3**(2)，85–116。M. Billaud-Friess 和 A. Nouy（2017），用于求解参数依赖动力系统的动力学模型降阶方法，*SIAM J.Sci.Comput.***39**(4)，A1766–A1792。P. Binev、A. Cohen、W. Dahmen、R. DeVore、G. Petrova 和 P. Wojtaszczyk（2011），约化基方法中贪心算法的收敛速率，*SIAM J. Math. Anal.***43**(3)，1457–1472。F. Black、P. Schulze 和 B. Unger（2020），利用动态变换模态进行基于投影的模型降阶，*ESAIM: Math. Model. Numer.Anal.***54**(6)，2011–2043。F. Black、P. Schulze 和 B. Unger（2021*a*），通过非线性模型降阶高效模拟野地火灾，*Fluids* **6**(8)，280。F. Black、P. Schulze 和 B. Unger（2021*b*），波动方程的动态变换模态模型降阶，载于*Proceedingsin AppliedMathematicsandMechanics 2020*

F. Black、P. Schulze 和 B. Unger（2022），通过基于梯度的输运优化对流动数据进行模态分解，载于*Active Flow and Combustion Control2021*（R. King 和 D. Peitsch 编），Springer International Publishing，沙姆，页码 203–224。T. Blickhan（2024），一种利用线性最优输运处理降阶基问题的配准方法，*SIAMJ. Sci. Comput.***46**（5），A3177–A3204。D. Bon、B. Caris 和 O. Mula（2025），利用动力学采样实现稳定的非线性动力学近似，*arXiv* **2505**（11938），1–28。A. Bonito、A. Cohen、R. DeVore、D. Guignard、P. Jantsch 和 G. Petrova（2021），模型降阶的非线性方法，*ESAIM*：数学建模与数值分析 **55**（2），507–531。D. Braess（1986），非线性逼近理论，Springer。J. Broeckhove、L. Lathouwers、E. Kesteloot 和 P. Van Leuven（1988），关于含时变分原理的等价性，*Chem.Phys.Lett.***149**（5），547–550。J. Bruna、B. Peherstorfer 和 E. Vanden-Eijnden（2024），用于高维演化方程的主动学习神经伽辽金格式，计算物理学杂志 **496**，112588。J. Brunken、K. Smetana 和 K. Urban （ 2019 ），（参数化）一阶输运方程：最优稳定 Petrov–Galerkin 方法的实现，*SIAMJ. Sci.Comput.* **41**（1），A592–A621。P. Buchfink、S. Glas 和 B. Haasdonk（2023），非线性流形上哈密顿系统的辛模型降阶及利用弱辛自编码器进行近似，*SIAMJ.Sci. Comput.* **45**（2），A289–A311。P. Buchfink、S. Glas 和 B. Haasdonk（2024*a*），多项式映射流形上模型降阶的逼近界，巴黎科学院数学报告 **362**，1211–1226。P. Buchfink、S. Glas、B. Haasdonk 和 B. Unger（2024*b*），流形上的模型降阶：微分几何框架，物理学*D*辑 **468**，134299。A. Buffa、Y. Maday、A. T. Patera、C. Prud'homme 和 G. Turinici（2012），参数化降阶基方法中贪心算法的先验收敛性，*ESAIM*：数学建模与数值分析**46**（3），595–603。A. Buhr 和 K. Smetana（2018），随机局部模型阶约简，*SIAMJ. Sci.Comput.* **40**（4），A2120–A2151。A. Buhr、C. Engwer、M. Ohlberger 和 S. Rave（2017），ArbiLoMod：一种专为任意局部修改设计的仿真技术，*SIAM J.Sci.Comput.***39**（4），A1435–A1465。T. Bui-Thanh、K. Willcox、O. Ghattas 和 B. van Bloemen Waanders（2007），面向目标且受模型约束的大规模系统降阶优化，计算物理学杂志**224**（2），880–896。S. Burela、P. Krah 和 J. Reiss（2025），利用基于移位 POD 的深度学习方法对野地火灾模型进行参数化模型阶数约简，计算数学进展 **51**（1），9。N. Cagniart、Y. Maday 和 B. Stamm（2019），强对流效应问题的模型阶数约简，载于偏微分方程与应用的贡献（B. N. Chetverushkin、W. Fitzgibbon、Y. A. Kuznetsov、P. Neittaanm¨aki、J. Periaux 和 O. Pironneau 编），应用科学计算方法，Springer 出版社，瑞士沙姆，页码 131–150。X. Cao、M. B. Saltik 和 S. Weiland（2015），描述符系统的 Hankel 模型降阶，载于第*54*届 *IEEE* 决策与控制会议（*CDC*），页码 4668–4673。K. Carlberg（2015），降阶模型的自适应 h 加密，*Internat.J. Numer.MethodsEngrg.* **102**（5），1192–1210。K. Carlberg、M. Barone 和 H. Antil（2017），非线性模型降阶中的 Galerkin 投影与最小二乘 Petrov–Galerkin 投影，计算物理学杂志**330**，693–734。

K. Carlberg、C. Bou-Mosleh 与 C. Farhat（2011年），基于最小二乘 Petrov–Galerkin 投影与压缩张量近似的高效非线性模型降阶，*Internat. J. Numer. Methods Engrg.***86**(2)，155–181。B. Carrel（2026），动力学低秩近似的随机化方法，计算物理学杂志 **544**，114421。M. Cenedese、J. Ax˚as、B. B¨auerlein、K. Avila 与 G. Haller（2022年），基于谱子流形对不可线性化动力学进行数据驱动建模与预测，自然通讯**13**(1)，872。G. Ceruti 与 C. Lubich（2022年），一种用于动力学低秩近似的非常规稳健积分器，*BIT*数值数学**62**(1)，23–44。G. Ceruti、L. Einkemmer、J. Kusch 与 C. Lubich（2024*a*），基于中点法则的稳健二阶低秩 BUG 积分器，*BIT*数值数学**64**(3)，30。G. Ceruti、J. Kusch 与 C. Lubich（2022年），用于动力学低秩近似的秩自适应稳健积分器，*BIT*数值数学**62**(4)，1149–1174。G. Ceruti、J. Kusch 与 C. Lubich（2024*b*），用于动力学低秩近似的并行秩自适应积分器，*SIAM J. Sci. Comput.* **46**(3)，B205–B228。A. Charous 与 P. F. J. Lermusiaux（2023），用于动力学低秩近似的带扰动回缩的动态正交 Runge–Kutta 格式，*SIAMJ. Sci.Comput.* **45**(2)，A872–A897。A. Charous 与 P. F. J. Lermusiaux（2024），稳定的秩自适应动态正交 Runge–Kutta 格式，*SIAM J. Sci. Comput.* **46**(1)，A529–A560。S. Chaturantabut 与 D. C. Sorensen（2010年），基于离散经验插值的非线性模型降阶，*SIAM J. Sci. Comput.* **32**(5)，2737–2764。H. Chen、R. Wu、E. Grinspun、C. Zheng 与 P. Y. Chen（2023*a*），含时偏微分方程的隐式神经空间表示，载于第*40*届国际机器学习会议论文集 （A. Krause、E. Brunskill、K. Cho、B. Engelhardt、S. Sabato 与 J. Scarlett 编），机器学习研究论文集第202卷，PMLR，第5162–5177页。P. Y. Chen、J. Xiang、D. H. Cho、Y. Chang、G. A. Pershing、H. T. Maia、M. M. Chiaramonte、K. T. Carlberg 与 E. Grinspun（2023*b*），CROM：使用隐式神经表示的偏微分方程连续降阶建模，载于第十一届国际学习表征会议。Y. Chen、S. Gottlieb、L. Ji 与 Y. Maday（2021年），基于过度配置与残差超降阶误差估计的无 EIM 退化的约化基方法，计算物理学杂志**444**，110545。Z. Chen、J. Mccarran、E. Vizcaino、M. Soljacic 与 D. Luo（2024），TENG：面向机器精度、用深度神经网络求解偏微分方程的时间演化自然梯度，载于第*41*届国际机器学习会议论文集 （R. Salakhutdinov、Z. Kolter、K. Heller、A. Weller、N. Oliver、J. Scarlett 与 F. Berkenkamp 编），机器学习研究论文集第235卷，PMLR，第7143–7162页。W. Cho、K. Lee、D. Rim 与 N. Park（2023），面向低秩物理信息神经网络的基于超网络的元学习，载于第三十七届神经信息处理系统会议。A. Cohen 与 R. DeVore（2015年），高维参数偏微分方程的近似，数值数学学报 **24**，1–159。A. Cohen 与 R. DeVore（2016年），全纯映射下的 Kolmogorov 宽度，*IMA J.Numer.Anal.* **36**(1)，1–12。

A. Cohen、W. Dahmen、R. DeVore 与 J. Nichols（2020），使用随机训练集的贪心选择降阶基，ESAIM：数学建模与数值分析 54(5)，1509–1524。A. Cohen、W. Dahmen、O. Mula 与 J. Nichols（2022年a），用于状态与参数估计的非线性降阶模型，SIAM/ASA 不确定性量化期刊 10(1)，227–267。A. Cohen、R. DeVore 与 C. Schwab（2011年），参数与随机椭圆型偏微分方程的解析正则性与多项式近似，分析与应用 09(01)，11–47。A. Cohen、R. DeVore、G. Petrova 与 P. Wojtaszczyk（2022年b），最优稳定非线性逼近，计算数学基础 22(3)，607–648。A. Cohen、C. Farhat、Y. Maday 与 A. Somacal（2024），偏微分方程的非线性压缩降阶基逼近，C. R. M´ec. 352，267–279。P. G. Constantine（2015年），主动子空间：参数研究中降维的新兴思想，SIAM 聚焦丛书，工业与应用数学学会。D. M. Copeland、S. W. Cheung、K. Huynh 与 Y. Choi（2022年），拉格朗日流体动力学的降阶模型，计算方法与应用力学工程 388，114259。A. Cortinovis、D. Kressner、S. Massei 与 B. Peherstorfer（2020），结合自适应经验插值、用于在线自适应模型降阶中学习基更新的准最优采样，载于美国控制会议（ACC）2020，IEEE。J. Coughlin、J. Hu 与 U. Shumlak（2024），基于新型宏观-微观分解的 Vlasov 方程稳健守恒动力学低秩方法，计算物理学杂志 509，113055。W. Dahmen、C. Plesken 与 G. Welper（2014年），双重贪心算法：面向输运主导问题的约化基方法，ESAIM：数学建模与数值分析 48(3)，623–663。T. Daniel、F. Casenave、N. Akkari 与 D. Ryckelynck（2020），深度神经网络辅助的模型降阶（ROM-net），Adv. Model. and Simul. 7(1)，16。F. de Avila Belbute-Peres、Y. fan Chen 与 F. Sha（2021年），HyperPINN：利用物理信息超网络学习参数化微分方程，载于深度学习与微分方程的共生。P. Deuflhard 与 M. Weiser（2012年），偏微分方程的自适应数值求解，De Gruyter。R. A. DeVore（1998年），非线性逼近，数值数学学报 7，51–150。R. A. DeVore、G. Kyriazis、D. Leviatan 与 V. M. Tikhomirov（1993年），小波压缩与非线性 n-宽度，计算数学进展 1(2)，197–214。R. DeVore、B. Hanin 与 G. Petrova（2021年），神经网络逼近，数值数学学报 30，327–444。M. Dihlmann、M. Drohmann 与 B. Haasdonk（2011年），采用带自适应时间划分的降阶基方法对参数化演化问题进行模型降阶，载于第五届自适应建模与仿真国际会议论文集（ADMOS 2011）（D. Aubry、P. Dı́ez、B. Tie 与 N. Par´es 编）。P. A. Dirac（1930年），关于托马斯原子中交换现象的注记，剑桥哲学学会数学会刊 26(3)，376–385。M.-H. Do、J. Feydy 与 O. Mula（2025），稀疏 Wasserstein 重心及其在降阶建模中的应用，科学计算杂志 102(3)，64。Y. Dong、P. Schwerdtner 与 B. Peherstorfer（2025），演化问题非线性参数化解的随机化时间步进，arXiv 2512(19009)，1–43。Z. Drmaˇc 与 S. Gugercin（2016年），离散经验插值法的一种新选择算子——改进的先验误差界及其推广，SIAM J. Sci. Comput. **38**(2)，A631–A648。

M. Drohmann、B. Haasdonk 和 M. Ohlberger（2011），非线性对流–扩散方程的自适应降阶基方法，载于复杂应用中的有限体积法 *VI*：问题与展望（J. Fo˘rt、J. F¨urst、J. Halama、R. Herbin 和 F. Hubert 编），Springer Berlin Heidelberg，柏林，海德堡，第 369–377 页。M. Drohmann、B. Haasdonk 和 M. Ohlberger（2012），基于经验算子插值的非线性参数化演化方程约化基近似，*SIAM J.Sci. Comput.* **34**（2），A937–A969。Y. Du 和 T. A. Zaki（2021），演化深度神经网络，物理评论 *E* **104**，045303。J. L. Eftang 和 B. Stamm（2012），参数多域'hp'经验插值，*Internat. J. Numer. Methods Engrg.***90**（4），412–428。J. L. Eftang、D. J. Knezevic 和 A. T. Patera（2011），参数化抛物型偏微分方程的 hp 认证降阶基方法，数学计算与动力系统建模 **17**（4），395–422。J. L. Eftang、A. T. Patera 和 E. M. Rønquist（2010），参数化椭圆型偏微分方程的 hp 认证降阶基方法，*SIAMJ. Sci. Comput.* **32**（6），3170–3200。V. Ehrlacher、D. Lombardi、O. Mula 和 F.-X. Vialard（2020），度量空间上的非线性模型降阶：在 Wasserstein 空间中的一维守恒偏微分方程上的应用，*ESAIM*：数学建模与数值分析 **54**（6），2159–2197。L. Einkemmer 和 I.Joseph（2021），用于 Vlasov 方程的质量、动量和能量守恒动态低秩方案，计算物理学杂志**443**，110495。L. Einkemmer 和 C. Lubich（2019），用于 Vlasov 方程的拟守恒动态低秩算法，*SIAM J. Sci. Comput.* **41**（5），B1061–B1081。L. Einkemmer、J. Hu 和 Y. Wang（2021），用于多尺度多维线性输运方程的渐近保持动态低秩方法，计算物理学杂志**439**，110353。L. Einkemmer、K. Kormann、J. Kusch、R. G. McClarren 和 J.-M. Qiu（2025），含时动力学模拟低秩方法综述，计算物理学杂志 **538**，114191。L. Einkemmer、A. Ostermann 和 C. Scalone（2023），稳健守恒动态低秩算法，计算物理学杂志**484**，112060。C. Engwer、M. Ohlberger 和 L. Renelt（2024），参数化反应输运问题中超弱且最优稳定变分形式的模型降阶，*SIAM J.Sci. Comput.* **46**（5），A3205–A3229。A. Ern 和 J.-L. Guermond（2004），有限元理论与实践，Springer。P. A. Etter 和 K. T. Carlberg（2020），通过向量空间筛选实现降阶模型的在线自适应基细化与压缩，计算方法与应用力学工程 **364**，112931。R. Everson 和 L. Sirovich（1995），用于不完整数据的 Karhunen-Loeve 过程，美国光学学会杂志 **12**，1657–1664。C. Farhat、P. Avery、T. Chapman 和 J. Cortial（2014），通过有限转动以及基于能量的网格采样与加权，实现非线性有限元动力学模型的维数降阶并提高计算效率，*Internat.J.Numer.Methods Engrg.* **98**（9），625–662。C. Farhat、T. Chapman 和 P. Avery（2015），用于非线性有限元动力学模型超降阶的能量守恒采样与加权方法的结构保持、稳定性和精度性质，*Internat. J. Numer. MethodsEngrg.***102**（5），1077–1110。

M. Feischl、C. Lasser、C. Lubich 和 J. Nick（2024），正则化动力学参数近似，arXiv 2403(19234)，1–38。F. Feppon 和 P. F. J. Lermusiaux（2018），用于高效随机平流和拉格朗日输运的动态正交数值格式，SIAM Rev. 60(3)，595–625。A. Ferrero、T. Taddei 和 L. Zhang（2022），参数化二维守恒律的基于配准的模型降阶，J. Comput. Phys. 457，111068。M. A. Finzi、A. Potapczynski、M. Choptuik 和 A. G. Wilson（2023），一种利用神经网络求解初值偏微分方程的稳定且可扩展方法，载于第十一届国际学习表征会议。B. A. Francis（1987），《H_∞ 控制理论教程》，Springer-Verlag，柏林、海德堡。N. R. Franco、A. Manzoni 和 P. Zunino（2023），一种针对参数相关偏微分方程降阶建模的深度学习方法，Math. Comp. 92(340)，483–524。J. Frenkel（1934年），《波动力学：高级通论》，第436卷，牛津。S. Fresca 和 A. Manzoni（2022），POD-DL-ROM：通过本征正交分解增强基于深度学习的非线性参数化偏微分方程降阶模型，Comput. Meth. Appl. Mech. Eng. 388，114181。S. Fresca、L. Dede’ 和 A. Manzoni（2021），一种针对非线性时变参数化偏微分方程降阶建模的综合性深度学习方法，J. Sci. Comput. 87(2)，61。R. Geelen 和 K. Willcox（2022），算子推断框架中的局部非侵入式降阶建模，Philos. Trans. R. Soc. A 380(2229)，20210206。R. Geelen、L. Balzano 和 K. Willcox（2023a），利用多项式流形构造学习高维状态空间中的潜在表示，载于第62届IEEE决策与控制会议（CDC），第4960–4965页。R. Geelen、L. Balzano、S. Wright 和 K. Willcox（2024），利用非线性流形从数据中学习基于物理的降阶模型，Chaos 34(3)，033122。R. Geelen、S. Wright 和 K. Willcox（2023b），基于二次流形的非侵入式模型降阶算子推断，Comput. Methods Appl. Mech. Engrg. 403，115717。J.-F. Gerbeau 和 D. Lombardi（2014），用于非线性演化方程降阶积分的近似Lax对，J. Comput. Phys. 265，246–269。O. Ghattas 和 K. Willcox（2021），从数据中学习基于物理的模型：反问题与模型降阶视角，数值数学学报 30，445–554。S. Glas 和 H. Mu（2025），二次嵌入流形上的分段辛模型降阶，载于《数值数学与高级应用 ENUMATH 2023》第1卷（A. Sequeira、A. Silvestre、S. S. Valtchev 和 J. Janela 编），Springer Nature Switzerland，沙姆，第355–364页。S. Glas 和 B. Unger（2025），利用时间和参数进行非线性模型降阶，IFAC在线论文集 59(19)，550–555。S. Glas、A. Mayerhofer 和 K. Urban（2017），在降阶基方法中处理时间的两种方式，载于《参数化系统的模型降阶》（P. Benner、M. Ohlberger、A. Patera、G. Rozza 和 K. Urban 编），Springer International Publishing，沙姆，第1–16页。K. Glover（1984），线性多变量系统的所有最优Hankel范数逼近及其 l^∞ 误差界，Internat. J. Control 39(6)，1115–1193。I. Goodfellow、Y. Bengio 和 A. Courville（2016），《深度学习》，MIT Press。http: //www.deeplearningbook.org。

I. V. Gosea、S. Gugercin 和 B. Unger（2021），沿参数方向进行有理插值的参数化模型降阶，载于第*60*届*IEEE*决策与控制会议（*CDC*），第6895–6900页。H. Gowrachari、N. Demo、G. Stabile 和 G. Rozza（2025），利用神经网络移位增强流形变换，对流主导双曲问题的非侵入式模型降阶，*Comput.Fluids***300**，106758。P. Goyal 和 P. Benner（2024），利用深度学习实现非线性动力学的广义二次嵌入，物理学*D*辑**463**，134158。C. Greif 和 K. Urban（2019），波问题 Kolmogorov *n*-宽度的衰减，应用数学快报 **96**，216–222。C. Greif、P. Junk 和 K. Urban（2022），线性/岭展开：利用岭函数增强线性近似，计算数学进展 **48**(3)，15。M. A. Grepl 和 A. T. Patera（2005），参数化抛物型偏微分方程降阶基近似的后验误差界，*ESAIM*：数学建模与数值分析 **39**(1)，157–181。S. Grimberg、C. Farhat、R. Tezaur 和 C. Bou-Mosleh（2021），采用局部降阶基的非线性Petrov–Galerkin降阶模型超降阶的网格采样与加权，*Internat. J. Numer. Methods Engrg.***122**(7)，1846–1874。A. Gruber、M. Gunzburger、L. Ju 和 Z. Wang（2022），数据驱动降阶建模中神经网络架构的比较，计算方法与应用力学工程**393**，114764。R. Gruhlke、A. Nouy 和 P. Trunschke（2024），随机梯度下降与自然梯度下降的最优采样，*arXiv* **2402**(03113)，1–37。S. Grundel 和 N. Sarna（2024），通过自适应约化网格对参数化输运主导问题进行超降阶，偏微分方程应用 **5**(1)，3。C. Gu（2011），QLMOR：一种利用非线性系统二次-线性表示的基于投影的非线性模型阶数约化方法，*IEEE*计算机辅助设计与集成电路系统汇刊 **30**(9)，1307–1320。C. Gu（2012），《非线性动力系统的模型阶数约化》，博士学位论文，电气工程与计算机科学系，加州大学伯克利分校。M. Gubisch 和 S. Volkwein（2017），第1章：线性-二次最优控制的本征正交分解，载于模型降阶与近似（P. Benner、A. Cohen、M. Ohlberger 和 K. Willcox 编），工业与应用数学学会，费城，宾夕法尼亚州，美国，第3–63页。N. Guglielmi 和 M. Manucci（2023），参数偏微分方程围道积分方法中的模型降阶，*SIAM J. Sci. Comput.* **45**(4)，A1711–A1740。N. Guglielmi、M. L´opez-Fern´andez 和 M. Manucci（2021），用于对流-扩散方程的伪谱漫游围道积分方法，科学计算杂志 **89**(1)，22。B. Haasdonk（2017），参数偏微分方程的约化基方法——定常与非定常问题入门教程，载于模型降阶与近似（P. Benner、M. Ohlberger、A. Cohen 和 K. Willcox 编），工业与应用数学学会，第65–136页。B. Haasdonk 和 M. Ohlberger（2008），参数化线性演化方程有限体积近似的降阶基方法，*ESAIM*：数学建模与数值分析**42**(2)，277–302。B. Haasdonk 和 M. Ohlberger（2011），通过离线/在线分解实现参数化动力系统的高效降阶模型与后验误差估计，*Math.Comput*

· 模型 · *Dyn. Sys.* **17**(2)，145–161 ·

B. Haasdonk、M. Dihlmann 和 M. Ohlberger（2011），一种基于参数空间自适应网格的参数化模型降阶训练集与多基生成方法，数学计算与动力系统建模 **17**(4)，423–442。B. Haasdonk、M. Ohlberger 和 G. Rozza（2008年），一种用于具有参数依赖显式算子的演化格式的降阶基方法，电子数值分析汇刊 **32**，145–161。D. Hartman 和 L. K. Mestha（2017年），一种用于动力系统模型降阶的深度学习框架，载于*2017*年 *IEEE*控制技术与应用会议 *(CCTA)*，第1917–1922页。C. D. Hauck 和 S. Schnake（2023），动力低秩近似自适应的一种预测校正策略，*SIAM J.Matrix Anal.Appl.***44**(3)，971–1005。E. J. Heller（1976），半经典动力学的含时变分方法，化学物理杂志 **64**(1)，63–73。R. Herkert、P. Buchfink 和 B. Haasdonk（2024），用于参数化哈密顿系统的基于字典的在线自适应保结构模型阶数约简，计算数学进展**50**(1)，12。M. Hess、A. Alla、A. Quaini、G. Rozza 和 M. Gunzburger（2019），一种用于具有分岔解的偏微分方程的局部降阶建模方法，计算方法与应用力学工程**351**，379–403。J. S. Hesthaven 和 C. Pagliantini（2021年），泊松系统的保结构降阶基方法，*Math. Comp.* **90**(330)，1701–1740。J. S. Hesthaven、C. Pagliantini 和 N.Ripamonti(2022*a*)，哈密顿系统的秩自适应保结构模型阶数约简，*ESAIM*：数学建模与数值分析**56**(2)，617–650。J. S. Hesthaven、C. Pagliantini 和 N. Ripamonti（2024），参数化粒子型韦拉索夫–泊松方程的自适应辛模型阶数约简，*Math. Comp.* **93**(347)，1153–1202。J. S. Hesthaven、C. Pagliantini 和 G. Rozza（2022*b*），时变问题的降阶基方法，数值数学学报**31**，265–345。J. S. Hesthaven、G. Rozza 和 B. Stamm（2016年），参数化偏微分方程的认证降阶基方法，Springer数学简明丛书，Springer出版社，沙姆，瑞士。J. S. Hesthaven、B. Stamm 和 S. Zhang（2014年），用于高维参数空间的高效贪心算法及其在经验插值和降阶基方法中的应用，*ESAIM*：数学建模与数值分析**48**(1)，259–283。B. Hillebrecht 和 B. Unger（2025*a*），物理信息神经网络的预测误差认证：理论、计算及其在斯托克斯流中的应用，*arXiv* **2508**(07994)，1–22。B. Hillebrecht 和 B. Unger（2025*b*），由偏微分方程定义的物理信息神经网络的严格后验误差界，*IEEE*神经网络与学习系统汇刊**36**(1)，1583–1593。M. Hochbruck、M. Neher 和 S. Schrammer（2023），用于一阶和二阶矩阵微分方程的秩自适应动力低秩积分器，*BIT*数值数学**63**(1)，9。P. Holmes、J. Lumley 和 G. Berkooz（1996），湍流、相干结构、动力系统与对称性，剑桥大学出版社，剑桥，英国。E. J. Hu、yelong shen、P. Wallis、Z. Allen-Zhu、Y. Li、S. Wang、L. Wang 和 W. Chen（2022年），LoRA：大语言模型的低秩适配，载于国际学习表征会议

.

C. Huang and K. Duraisamy (2023)，利用自适应采样投影预测混沌多尺度问题的降阶建模，计算物理学杂志 **491**，112356。C. Huang、J. Xu、K. Duraisamy 和 C. Merkle (2018)，火箭燃烧应用中的降阶模型探索，载于*2018 AIAA*航空航天科学会议。M. Hund、T. Mitchell、P. Mlinari´c 和 J. Saak (2022)，通过 $\mathcal{H}_2 \otimes \mathcal{L}_2$ 一阶必要条件进行基于优化的参数化模型阶数约简，*SIAM J. Sci. Comput.* **44**(3)，A1554–A1578。A. Iollo 和 D. Lombardi (2014)，基于最优质量输运的平流模态，物理评论 *E***89**，022923。A. Iollo、J. Labatut、P. Mounoud 和 T. Taddei (2026)，有界域中配准方法的数学方面，*arXiv* **2601**(03010)，20。O. Issan 和 B. Kramer (2023)，通过移位算子推断预测从内日球层到地球的太阳风流，计算物理学杂志**473**，111689。S. Jain、P. Tiso、J. B. Rutzmoser 和 D. J. Rixen (2017)，用于非线性结构动力学模型阶数约简的二次流形，*Comput.Struct.* **188**，80–94。K. Kashima (2016)，利用噪声响应数据的深度自编码器进行非线性模型降阶，载于第*55*届决策与控制会议（*CDC*），第5750–5755页。M. Kast 和 J. S. Hesthaven (2024)，利用演化深度神经网络求解偏微分方程的位置嵌入，计算物理学杂志**508**，112986。S. Kaulmann 和 B. Haasdonk (2013)，利用字典在线贪心构建约化基，载于自适应建模与仿真*2013*：第六届自适应建模与仿真国际会议（*ADMOS 2013*）论文集 （J. P. Moitinho de Almeida、P. D´ıez、C. Tiago 和 N. Par´es 编），国际工程数值方法中心（CIMNE），西班牙巴塞罗那，第365–376页。会议于2013年6月3日至5日在葡萄牙里斯本举行。K. G. Kay (1989)，含时变分法中的矩阵奇异性问题，*Chem.Phys.* **137**(1)，165–175。Y. Kazashi、F. Nobile 和 F. Zoccolan (2025)，随机微分方程的动力学低秩近似，*Math. Comp.* **94**(353)，1335–1375。M. Khamlich、F. Pichi 和 G. Rozza (2025)，受最优输运启发的深度学习框架，用于缓慢衰减的 Kolmogorov n-宽度问题：利用 Sinkhorn损失和 Wasserstein核，*SIAM J. Sci. Comput.* **47**(2)，C235–C264。E. Kieri、C. Lubich 和 H. Walach (2016)，存在小奇异值时的离散动力学低秩近似，*SIAM J. Numer.Anal.***54**(2)，1020–1038。Y. Kim、Y. Choi、D. Widemann 和 T. Zohdi (2022)，采用浅层掩码自编码器的快速准确物理信息神经网络降阶模型，计算物理学杂志**451**，110841。M. Kirby 和 D. Armbruster (1992)，从偏微分方程仿真中重构相空间，*Z.Angew.Math. Phys.***43**(6)，999–1022。H. Kleikamp、M. Ohlberger 和 S. Rave (2022)，利用时空域的微分同胚变换进行非线性模型阶数约简，载于*MATHMOD 2022* 讨论论文集，MATHMOD 2022，ARGESIM 出版社，维也纳。B. Klein 和 M. Ohlberger (2025)，用于抛物型偏微分方程约束优化的降阶模型多保真学习，*arXiv* **2503**(21252)，1–36。R. Klein、B. Sanderse、P. Costa、R. Pecnik 和 R. Henkes (2025)，利用有理二次流形对一维双曲系统进行熵稳定模型降阶，计算物理学杂志**528**，113817。

O. Koch 和 C. Lubich (2007)，动力学低秩近似，SIAM J. Matrix Anal. Appl. 29(2)，434–454。O. Koch 和 C. Lubich (2010)，动力学张量近似，SIAM J. Matrix Anal. Appl. 31(5)，2360–2375。¨A. Kolmogoroff (1936)，关于给定函数类中函数的最佳近似，Ann. of Math. (2) 37(1)，107–110。S. Kolouri、S. R. Park 和 G. K. Rohde (2016)，Radon 累积分布变换及其在图像分类中的应用，IEEE Trans. Image Process. 25(2)，920–934。P. Krah、S. B¨uchholz、M. H¨aringer 和 J. Reiss (2023)，复杂移动前沿的前沿输运降阶，J. Sci. Comput. 96，28。P. Krah、A. Marmin、B. Zorawski、J. Reiss 和 K. Schneider (2025)，稳健的移位本征正交分解：利用近端方法分解具有多重输运的流动，SIAM J. Sci. Comput. 47(2)，A633–A656。B. Kramer 和 K. E. Willcox (2019)，通过提升变换和本征正交分解进行非线性模型阶数约简，AIAA杂志 57(6)，2297–2307。B. Kramer、B. Peherstorfer 和 K. E. Willcox (2024)，通过算子推断从数据中学习非线性降阶模型，Annu. Rev. Fluid Mech. 56，521–548。P. Kramer 和 M. Saraceno (1981)，量子力学中含时变分原理的几何，载于《物理学讲义》第140卷，Springer。Kunisch, Karl and Volkwein, Stefan (2008)，最优性系统的本征正交分解，ESAIM：数学建模与数值分析 42(1)，1–23。S. Kvaal、C. Lasser、T. B. Pedersen 和 L. Adamowicz (2023)，无需网格：用于含时薛定谔方程的自适应全柔性高斯函数，arXiv 2207(00271)，1–8。H. Y. Lam、G. Ceruti 和 D. Kressner (2025)，随机低秩 Runge–Kutta 方法，SIAM J. Matrix Anal. Appl. 46(2)，1587–1615。C. Lasser 和 C. Lubich (2020)，半经典情形下的量子动力学计算，数值数学学报 29，229–401。Y. LeCun、Y. Bengio 和 G. Hinton (2015)，深度学习，《自然》521(7553)，436–444。K. Lee 和 K. T. Carlberg (2020)，利用深度卷积自编码器对非线性流形上的动力系统进行模型降阶，J. Comput. Phys. 404，108973。K. Lee 和 K. T. Carlberg (2021)，深度守恒：实现物理守恒定律精确满足的潜在动力学模型，AAAI 人工智能大会论文集 35(1)，277–285。M. Lindsey (2025)，MNE：具有扩散过程和采样应用的过参数化神经演化，arXiv 2502(03645)，1–22。T. Long、R. Barnett、R. Jefferson-Loveday、G. Stabile 和 M. Icardi (2025)，基于 Radon 累积分布变换的对流主导问题降阶模型，计算数学进展 51(1)，5。H. Lu 和 D. M. Tartakovsky (2020)，用于构建对流主导现象降阶模型的拉格朗日动态模态分解，J. Comput. Phys. 407，109229。H. Lu 和 D. M. Tartakovsky (2021)，用于构建含激波双曲问题降阶模型的动态模态分解，J. Mach. Learn. Model. Comput. 2(1)，1–29。C. Lubich (2005)，关于量子分子动力学中的变分近似，Math. Comp. 74(250)，765–779。

C. Lubich（2008），从量子到经典分子动力学：降阶模型与数值分析，第12卷，欧洲数学会。C. Lubich 和 J. Nick（2025），刚性演化问题的正则化动力学参数近似，*arXiv***2501**(12118), 1–33。C. Lubich 和 I. V. Oseledets（2014），用于动力学低秩近似的投影算子分裂积分器，*BIT*数值数学**54**(1), 171–188。C. Lubich、I. V. Oseledets 和 B. Vandereycken（2015），张量列车的时间积分，*SIAM J.Numer. Anal.***53**(2), 917–941。C. Lubich、T. Rohwedder、R. Schneider 和 B. Vandereycken（2013），通过层级 Tucker 张量和张量列车张量进行动力学近似，*SIAM J.Matrix Anal. Appl.* **34**(2), 470–494。Y. Maday 和 E. M. Rønquist（2002），一种降阶基元方法，科学计算杂志**17**(1), 447–459。Y. Maday、A. Manzoni 和 A. Quarteroni（2016），一种用于参数化对流主导问题约化基近似的在线内禀稳定化策略，巴黎科学院数学报告 **354**(12), 1188–1194。Y. Maday、A. T. Patera 和 G. Turinici（2002），单参数对称强制椭圆偏微分方程约化基近似的全局先验收敛理论，巴黎科学院数学报告 **335**(3), 289–294。R. Maulik、B. Lusch 和 P. Balaprakash（2021），采用循环神经网络和卷积自编码器对对流主导系统进行降阶建模，流体物理**33**(3), 037106。A. McLachlan（1964年），含时薛定谔方程的变分解，分子物理学 **8**(1), 39–44。V. Mehrmann 和 B. Unger（2023），端口哈密顿微分代数系统的控制及其应用，数值数学学报**32**, 395–515。A. Mendible、S. L. Brunton、A. Y. Aravkin、W. Lowrie 和 J. N. Kutz（2020），行波物理中的降维与降阶建模，理论与计算流体动力学**34**(4), 385–400。H.-D. Meyer、U. Manthe 和 L. Cederbaum（1990），多构型含时 Hartree 方法，化学物理快报 **165**(1), 73–78。K. Miller 和 R. N. Miller（1981），移动有限元 I, *SIAM J. Numer. Anal.* **18**(6), 1019–1032。A. Mohaghegh 和 C. Huang（2026），用于对流主导问题自适应模型阶数降阶的特征引导采样策略，计算物理学杂志 **545**, 114468。R. Mojgani 和 M. Balajewicz（2017），用于对流主导非线性流动降维的拉格朗日基方法，*arXiv* **1701**(04343), 1–10。B. Moore（1981），线性系统中的主成分分析：可控性、可观测性与模型降阶，*IEEE Trans. Automat. Control* **26**(1), 17–32。S. Mowlavi 和 T. P. Sapsis（2018），具有连续对称性的随机动力系统模型阶数降阶，*SIAM J. Sci. Comput.***40**(3), A1669–A1695。C. Mullis 和 R. Roberts（1976），最小舍入噪声定点数字滤波器的综合，*IEEE Trans. Circuits andSystems* **23**(9), 551–562。E. Musharbash 和 F. Nobile（2018），具有随机边界条件的不可压缩纳维–斯托克斯方程的对偶动力学正交近似，计算物理学杂志

. **354**, 135–162 .

E. Musharbash、F. Nobile 和 E. Vidlič ckov´a (2020)，随机参数波动方程的辛动力低秩逼近，*BIT*数值数学**60**(4), 1153–1201。E. Musharbash、F. Nobile 和 T. Zhou (2015)，时变随机偏微分方程动态正交逼近的误差分析，*SIAM J. Sci. Comput.* **37**(2), A776–A810。M. A. Nabian、R. J. Gladstone 和 H. Meidani (2021)，通过重要性采样高效训练物理信息神经网络，*Comput.-Aided Civ.Eng.* **36**(8), 962–977。M. H. Naderi 和 H. Babaee (2023)，用于加速采用时变基的非线性随机降阶建模的自适应稀疏插值，计算方法与应用力学工程**405**, 115813。J. L. Nicolini 和 F. L. Teixeira (2021)，采用移位本征正交分解对对流占优动理学等离子体问题进行降阶建模，*IEEE*等离子体科学汇刊**49**(11), 3689–3699。F. Nobile 和 T. T. Trindade (2025*a*)，SUPG稳定化动力低秩逼近的误差估计，载于数值数学与高级应用 *ENUMATH 2023，* 第*2*卷(A. Sequeira、A. Silvestre、S. S. Valtchev 和 J. Janela 编)，Springer Nature Switzerland，沙姆，第438–447页。F. Nobile 和 T. T. Trindade (2025*b*)，Petrov–Galerkin动力低秩逼近：对流主导问题的SUPG稳定化，计算方法与应用力学工程**433**, 117495。M. Nonino、F. Ballarin、G. Rozza 和 Y. Maday (2023)，借助输运映射，对 Kolmogorov *n*-宽度缓慢衰减的流固耦合问题采用约化基方法，*Adv. Comput. Sci.Eng.***1**(1), 36–58。A. Nouy (2010)，通过正规广义分解进行先验模型降阶，以求解时变偏微分方程，计算方法与应用力学工程**199**(23), 1603–1626。A. Nouy 和 A. Pasco (2024)，用于状态估计的基于字典的模型降阶，计算数学进展 **50**(3), 32。M. Ohlberger 和 S. Rave (2013)，通过冻结法对参数化演化方程进行非线性约化基近似，巴黎科学院数学报告**351**(23–24), 901–906。M. Ohlberger 和 S. Rave (2016)，约化基方法：成效、局限与未来挑战，*Algoritmy*会议论文集第1–12页。M. Ohlberger 和 F. Schindler (2015)，采用自适应在线增广的局部化约化基多尺度方法的误差控制，*SIAM J.Sci.Comput.***37**(6), A2865–A2895。M. Ohlberger 和 F. Schindler (2017)，在线增广的非协调局部化模型降阶：迈向PDE约束优化中的最优复杂度，载于复杂应用的有限体积法 *VIII*——双曲型、椭圆型与抛物型问题 (C. Canc`es 和 P. Omnes 编)，Springer International Publishing，沙姆，第357–365页。S. E. Otto、G. R. Macchio 和 C. W. Rowley (2023)，利用约束自编码器学习动力系统降阶建模的非线性投影，*Chaos***33**(11), 113130。S. E. Otto、A. Padovan 和 C. W. Rowley (2022)，利用轨迹优化非线性系统的斜投影，*SIAM J.Sci.Comput.***44**(3), A1681–A1702。P. Pacciarini 和 G. Rozza (2014)，参数化对流-扩散偏微分方程的稳定化约化基方法，计算方法与应用力学工程**274**, 1–18。

C. Pagliantini（2021），哈密顿系统的动力学降阶基方法，*Numer.Math.* **148**(2), 409–448。C. Pagliantini 和 F. Vismara（2025），参数化哈密顿系统的全自适应保结构超降阶，*SIAM J. Sci. Comput.***47**(1), A124–A152。C. Pagliantini 和 F. Vismara（2026），非稀疏算子的自适应超降阶：在基于粒子的参数化动力学等离子体模型中的应用，*J. Comput. Phys.* **549**, 114609。S. Pan、S. L. Brunton 和 J. N. Kutz（2023），神经隐式流：一种网格无关的时空数据降维范式，*J. Mach. Learn. Res.* **24**(41), 1–60。D. Papapicco、N. Demo、M. Girfoglio、G. Stabile 和 G. Rozza （ 2022 ），神经网络移位本征正交分解：一种用于双曲方程非线性降阶的机器学习方法，*Comput. Meth.Appl.Mech.Eng.* **392**, 114687。E. J. Parish 和 K. T. Carlberg（2021），动力系统的窗口化最小二乘模型降阶，*J. Comput. Phys.***426**, 109939。A. Pazy（1983），线性算子半群及其在偏微分方程中的应用，应用数学科学，Springer，纽约，纽约州，美国。B. Peherstorfer（2020），通过在线自适应基和自适应采样对输运主导问题进行模型降阶，*SIAM J. Sci.Comput.***42**, A2803–A2836。B. Peherstorfer（2022），以非线性模型降阶突破 Kolmogorov 屏障，*Not. Am. Math. Soc.***69**, 725–733。B. Peherstorfer 和 K. Willcox（2015*a*），动态数据驱动降阶模型，*Comput. Meth. Appl. Mech. Eng.***291**, 21–41。B. Peherstorfer 和 K. Willcox（2015*b*），通过低秩更新对非线性系统进行在线自适应模型降阶，*SIAM J. Sci. Comput.***37**(4), A2123–A2150。B. Peherstorfer、D. Butnaru、K. Willcox 和 H.-J. Bungartz（2014），局部离散经验插值方法，*SIAM J. Sci. Comput.***36**(1), A168–A192。B. Peherstorfer、Z. Drmaˇc 和 S. Gugercin（2020），随机与确定性采样点下离散经验插值和缺失数据适当正交分解的稳定性，*SIAM J. Sci. Comput.* **42**(5), A2837–A2864。B. Peherstorfer、K. Willcox 和 M. Gunzburger（2018），不确定性传播、推断和优化中的多保真度方法综述，*SIAM Rev.***60**(3), 550–591。E. Perez、F. Strub、H. de Vries、V. Dumoulin 和 A. Courville（2018），FiLM：通过通用条件层进行视觉推理，载于第三十二届*AAAI*人工智能会议、第三十届人工智能创新应用会议和第八届*AAAI*人工智能教育进展研讨会论文集，AAAI’18/IAAI’18/EAAI’18, AAAI出版社。D. B. Phuong Huynh、D. J. Knezevic 和 A. T. Patera（2013），静力凝聚降阶基元方法：近似与后验误差估计，*ESAIM*：数学建模与数值分析 **47**(1), 213–251。A. Pinkus（1985），逼近理论中的*N*宽度，数学及其交叉领域成果，Springer 出版社，海德堡，德国。A. Pinkus（1999），神经网络中多层感知机模型的逼近理论，数值数学学报**8**, 143–195。C. Prud’homme、D. V. Rovas、K. Veroy、L. Machiels、Y. Maday、A. T. Patera 和 G. Turinici（ 2001），参数化偏微分方程的可靠实时求解：降阶基输出界方法，*J. Fluids Eng.* **124**(1), 70–80。

E. Qian、M. Grepl、K. Veroy 和 K. Willcox（2017），用于PDE约束优化的认证信赖域降阶基方法，*SIAM J. Sci. Comput.* **39**(5), S434–S460。E. Qian、B. Kramer、B. Peherstorfer 和 K. Willcox（2020），提升与学习：面向大规模非线性动力系统的物理信息机器学习，物理学*D*辑 **406**, 132401。A. Quarteroni 和 G. Rozza 编（2014），用于建模和计算约化的降阶方法，‘MS&A - 建模、仿真与应用’第9卷，Springer，沙姆，瑞士。A. Quarteroni、A. Manzoni 和 F. Negri（2016），偏微分方程降阶基方法：导论，UNITEXT，Springer 出版社，沙姆，瑞士。M. Raissi、P. Perdikaris 和 G. Karniadakis（2019），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与反问题的深度学习框架，*J. Comput.Phys.***378**, 686–707。D. Ramezanian、A. G. Nouri 和 H. Babaee（2021），利用时变流形对被动与反应性组分进行在线降阶建模，*Comput. Meth. Appl. Mech.Eng.***382**, 113882。M.-L. Rap´un 和 J. M. Vega（2010），基于局部本征正交分解与 Galerkin 投影的降阶模型，*J. Comput. Phys.* **229**(8), 3046–3063。A. H. Razavi 和 M. Yano（2025），参数化空气动力学问题的基于配准的非线性模型降阶及其在跨声速欧拉和RANS流动中的应用，*J.Comput. Phys.* **521**, 113576。F. Regazzoni、S. Pagani、M. Salvador、L. Dede’ 和 A. Quarteroni（2024），通过潜动力学网络学习时空过程的内禀动力学，*Nat.Commun.* **15**(1), 1834。J. Reiss（2021），针对多重输运系统的基于优化的模态分解，*SIAMJ. Sci. Comput.***43**(3), A2079–A2101。J. Reiss、P. Schulze、J. Sesterhenn 和 V. Mehrmann（2018），移位本征正交分解：一种针对多重输运现象的模态分解方法，*SIAM J. Sci.Comput.* **40**(3), A1322–A1344。J. Ren、W. R. Wolf 和 X. Mao（2021），通过 Radon 累积分布变换对行波问题进行模型降阶，*Phys.Rev.Fluids* **6**, L082501。D. Rim（2018），利用 Radon 变换对双曲型偏微分方程进行维数分裂，*SIAM J. Sci. Comput.* **40**(6), A4184–A4207。D. Rim 和 G. Welper（2026），熵解的低秩神经表示，计算数学进展**52**(1), 2。D. Rim、S. Moe 和 R. J. Le Veque（2018），双曲型偏微分方程模型降阶中的输运反转，*SIAM/ASA* 不确定性量化期刊**6**(1), 118–150。D. Rim、B. Peherstorfer 和 K. T. Mandli（2023），通过输运子空间进行流形逼近：输运主导问题的模型降阶，*SIAMJ. Sci. Comput.***45**(1), A170–A199。F. Romor、G. Stabile 和 G. Rozza（2023），基于卷积自编码器和约化超配点法的非线性流形降阶模型，科学计算杂志 **94**(3), 74。E. Rothe（1930年），作为一维边值问题极限情形的二维抛物型边值问题，*Math. Ann.***102**, 650–670。G. M. Rotskoff、A. R. Mitchell 和 E. Vanden-Eijnden（2022），针对由稀有事件主导的变分目标进行主动重要性采样：对优化和泛化的影响，载于第二届数学与科学机器学习会议论文集 （J. Bruna、J. Hesthaven 和 L. Zdeborova 编），机器学习研究论文集第145卷，PMLR，第757–780页。

K. Rowan, L. Schatzki, T. Zaklama, Y. Suzuki, K. Watanabe 和 K. Varga（2020），利用含时变分原理模拟激光场中的氢原子，物理评论 ***E*101**，023313。C. W. Rowley 和 J. E. Marsden（2000），具有对称性的系统的重构方程与 Karhunen–Lo`eve 展开，物理学*D*辑 **142**(1–2)，1–19。C. W. Rowley、I. G. Kevrekidis、J. E. Marsden 和 K. Lust（2003），自相似动力系统的降阶与重构，*Nonlinearity* **16**(4)，1257。G. Rozza、D. Huynh 和 A. Patera（2008），仿射参数化椭圆强制偏微分方程的约化基近似与后验误差估计，*Arch.Comput. Methods Eng.* **15**(3)，229–275。D. E. Rumelhart、G. E. Hinton 和 R. J. Williams（1986），通过误差反向传播学习表征，《自然》**323**(6088)，533–536。J. Rutzmoser、D. Rixen、P. Tiso 和 S. Jain（2017），用于非线性结构动力学降阶建模的二次流形推广，*Comput. Struct.* **192**，196–209。T. P. Sapsis 和 P. F. Lermusiaux（2009），连续随机动力系统的动力正交场方程，物理学*D*辑 **238**(23)，2347–2360。N. Sarna 和 P. Benner（2021），针对参数依赖跳跃不连续问题的数据驱动模型阶数约简，计算方法与应用力学工程 **387**，114168。N. Sarna、J. Giesselmann 和 P. Benner（2024），通过单调特征匹配进行数据驱动的快照校准，*Finite Elem. Anal. Des* **230**，104065。S. Sawada、R. Heather、B. Jackson 和 H. Metiu（1985），使用波函数的高斯波包表示进行含时量子力学计算的策略，化学物理杂志 **83**(6)，3009–3027。M. Schlegel 和 B. R. Noack（2015），论 Galerkin 模型的长期有界性，*J.FluidMech.***765**，325–352。B. Sch¨olkopf 和 A. J. Smola（2001），核方法学习：支持向量机、正则化、优化及其他，MIT Press，马萨诸塞州剑桥。P. Schulze（2023），输运主导现象的基于能量的模型降阶，博士论文，柏林工业大学数学研究所。P. Schulze、J. Reiss 和 V. Mehrmann（2019），通过移位本征正交分解对脉冲爆震燃烧器进行模型降阶，载于《主动流动与燃烧控制 *2018*》（R. King 编），Springer International Publishing，沙姆，第271–286页。P. Schulze、B. Unger、C. Beattie 和 S. Gugercin（2018），数据驱动的结构化实现，*Linear AlgebraAppl.* **537**，250–286。P. Schwerdtner 和 B. Peherstorfer（2024），用于非线性降维和非线性模型降阶的二次流形贪心构造，*arXiv* **2403**(06732)，1–35。P. Schwerdtner、S. Gugercin 和 B. Peherstorfer（2025*a*），二次流形上的经验稀疏回归，*SIAM J. Sci. Comput.* **47**(6)，A3085–A3107。P. Schwerdtner、P. Mohan、A. Pachalieva、J. Bessac、D. O'Malley 和 B. Peherstorfer（2025*b*），从流式数据在线学习二次流形，以实现非线性降维和非线性模型降阶，*Proc.R.Soc.A* **481**(2314)，20240670。P. Schwerdtner、P. Schulze、J. Berman 和 B. Peherstorfer（2024），利用神经伽辽金格式构造守恒哈密顿量及其他量的非线性嵌入，*SIAM J.Sci.Comput.***46**(5)，C583–C607。

J. Sesterhenn 和 A. Shahirpour（2019），特征动态模态分解，*Theor.Comput. FluidDyn.***33**(3)，281–305。H. Sharma、H. Mu、P. Buchfink、R. Geelen、S. Glas 和 B. Kramer（2023），利用数据驱动二次流形对哈密顿系统进行辛模型降阶，计算方法与应用力学工程 **417**，116402。Y. S. Shimizu 和 E. J. Parish（2021），非线性动力系统的窗化时空最小二乘 Petrov–Galerkin 模型阶数降阶，计算方法与应用力学工程**386**，114050。V. Simoncini（2016），线性矩阵方程的计算方法，*SIAM Rev.***58**(3)，377–441。R. Singh、W. Uy 和 B. Peherstorfer（2023），针对输运主导问题在线自适应模型降阶的前瞻式数据采集策略，*Chaos* **33**(11)，113112。L. Sirovich（1987），湍流与相干结构的动力学。第1部分：相干结构，*Q.Appl. Math.***45**(3)，561–571。V. Sitzmann、J. Martel、A. Bergman、D. Lindell 和 G. Wetzstein（2020），具有周期激活函数的隐式神经表示，载于神经信息处理系统进展（H. Larochelle、M. Ranzato、R. Hadsell、M. Balcan 和 H. Lin 编），第33卷，Curran Associates, Inc.，第7462–7473页。K. Smetana 和 A. T. Patera（2016），用于组件式静态凝聚过程的最优局部近似空间，*SIAM J.Sci.Comput.***38**(5)，A3318–A3356。K. Smetana 和 T. Taddei（2023），非线性椭圆偏微分方程的局部化模型降阶：局部化训练、单位分解与自适应增广，*SIAMJ.Sci. Comput.* **45**(3)，A1300–A1331。T. Taddei（2020），一种用于模型阶数约简的配准方法：数据压缩与几何降阶，*SIAM J. Sci. Comput.***42**(2)，A997–A1027。T. Taddei 和 L. Zhang（2021），参数化一维双曲型偏微分方程的基于时空配准的模型降阶，*ESAIM*：数学建模与数值分析 **55**(1)，99–130。T. Taddei、S. Perotto 和 A. Quarteroni（2015），非线性守恒律的降阶基技术，*ESAIM*：数学建模与数值分析**49**(3)，787–814。T. Taddei、I. Tifouti 和 N. Barral（2025），具有时空与参数自适应性的参数化偏微分方程局部配准模型降阶，载于*AIAA SCITECH 2025* 论坛，第1–23页。M. Telgarsky（2016），神经网络深度的优势，载于第*29*届年度学习理论会议（V. Feldman、A. Rakhlin 和 O. Shamir 编），《机器学习研究论文集》第49卷，PMLR，哥伦比亚大学，纽约，美国纽约，第1517–1539页。M. Tomasetto、J. P. Williams、F. Braghin、A. Manzoni 和 J. N. Kutz（2025），使用浅层循环解码器网络进行降阶建模，*Nat.Commun.* **16**(1)，10260。D. Torlo（2025），任意拉格朗日-欧拉框架下对流主导双曲问题的模型降阶：离线与在线阶段及误差估计器，载于《数值数学与高级应用 *ENUMATH 2023，第2卷*》（A. Sequeira、A. Silvestre、S. S. Valtchev 和 J. Janela 编），第409–419页。L. N. Trefethen（2019），逼近理论与逼近实践，工业与应用数学学会。B. Unger 和 S. Gugercin（2019），线性动力系统的 Kolmogorov *n*-宽度，计算数学进展 **45**(5–6)，2273–2286。

K. Urban 和 A. T. Patera（2012），抛物型偏微分方程约化基近似的一种新误差界，巴黎科学院数学报告**350**(3-4)，203–207。W. I. T. Uy、C. R. Wentland、C. Huang 和 B. Peherstorfer（2024），用于预混火焰问题的具有潜在动力学非线性逼近的降阶模型，载于降阶、近似、机器学习、代理模型、仿真替代模型与模拟器：*RAMSES*（G. Rozza、G. Stabile、M. Gunzburger 和 M. D'Elia 编），Springer Nature Switzerland，沙姆，第241–259页。T. Van Gastelen、W. Edeling 和 B. Sanderse（2026），利用空间局部降阶模型对对流主导流动进行建模，*Comput. Fluids* **305**，106911。V. Vapnik（1991），学习理论中的风险最小化原理，载于神经信息处理系统进展（J. Moody、S. Hanson 和 R. Lippmann 编），第4卷，Morgan-Kaufmann。K. Veroy 和 A. Patera（2005），参数化稳态不可压缩纳维-斯托克斯方程的经认证实时求解：严格的约化基后验误差界，*Int.J. Numer. MethodsFluid.* **47**(8–9)，773–788。K. Washabaugh、D. Amsallem、M. Zahr 和 C. Farhat（2012），使用局部降阶基的计算流体力学问题非线性模型降阶，载于第*42*届*AIAA*流体动力学会议暨展览。P. Weder、P. Schwerdtner 和 B. Peherstorfer（2025），二次流形上利用神经伽辽金格式进行非线性模型降阶，计算物理学杂志**539**，114249。G. Welper（2017），通过变换后的快照对具有参数依赖跳跃的函数进行插值，*SIAM J. Sci. Comput.***39**(4)，A1225–A1250。G. Welper（2020），利用高分辨率变换进行变换快照插值，*SIAM J.Sci. Comput.* **42**(4)，A2037–A2061。Y. Wen、E. Vanden-Eijnden 和 B. Peherstorfer（2024），神经伽辽金格式中参数与粒子动力学的耦合及其在自适应采样中的应用，物理学*D*辑 **462**，134129。C. R. Wentland、K. Duraisamy 和 C. Huang（2023），面向大型多尺度流体系统的可扩展基于投影的降阶模型，*AIAA*杂志**61**(10)，4499–4523。B. Wieland（2015），未知参数集的隐式分区方法，计算数学进展 **41**(5)，1159–1186。P. Wittmuess、C. Tarin、A. Keck、E. Arnold 和 O. Sawodny（2016），利用泰勒级数表示通过平衡截断进行参数化模型阶数约简，*IEEE Trans.Automat. Control* **61**(11)，3438–3451。C. Wu、M. Zhu、Q. Tan、Y. Kartha 和 L. Lu（2023），物理信息神经网络中非自适应采样与基于残差的自适应采样的综合研究，计算方法与应用力学工程 **403**，115671。J. Xu 和 K. Duraisamy（2020），用于时空动力学参数化预测的多层卷积自编码器网络，计算方法与应用力学工程**372**，113379。S. Yıldız、P. Goyal、T. Bendokat 和 P. Benner（2024），利用弱辛提升对非线性哈密顿系统二次表示的数据驱动辨识，机器学习、建模与计算杂志 **5**(2)，45–71。M. J. Zahr 和 C. Farhat（2015），用于PDE约束优化的参数化降阶模型的渐进构建，*Internat.J. Numer.Methods Engrg.* **102**(5)，1111–1135。H. Zhang、Y. Chen、E. Vanden-Eijnden 和 B. Peherstorfer（2025），通过按时间顺序训练非线性参数化求解时变偏微分方程，*SIAM*评论。（已接收）。

T. Zhao、C. Sun、A. Cohen、J. Stokes 和 S. Veerapaneni（2024），量子启发变分算法用于偏微分方程：在金融衍生品定价中的应用，*Quant.Finance* **24**(1)，1–11。R. Zimmermann 和 K. Willcox（2016），一种加速的贪心缺失点估计过程，*SIAMJ. Sci. Comput.***38**(5)，A2827–A2850。R. Zimmermann、B. Peherstorfer 和 K. Willcox（2018），几何子空间更新及其在在线自适应非线性模型降阶中的应用，*SIAM J.MatrixAnal.Appl.***39**(1)，234–261。V. Zucatti 和 M. J. Zahr（2024），一种基于混合快照、面向对流主导问题的自适应免训练降阶模型，*Int. J. Numer.Methods Fluid.***96**(2)，189–208。A. Zuyev、L. Feng 和 P. Benner（2024），非线性变换的Kolmogorov n-宽度估计及其在分布参数控制系统中的应用，*IEEE ControlSystemsLett.* **8**，1877–1882。