arXiv:2603.03211v1 [math.OC] 3 Mar 2026

形状导数信息神经算子及其在风险规避型形状优化中的应用

Xindi Gong[a, *], Dingcheng Luo[b, c], Thomas O'Leary-Roseberry[d], Ruanui Nicholson[e], Omar Ghattas[a, f]

[a]得克萨斯大学奥斯汀分校 *Oden*计算工程与科学研究所，美国得克萨斯州奥斯汀[b]昆士兰科技大学数学科学学院，澳大利亚布里斯班[c]昆士兰科技大学数据科学中心，澳大利亚布里斯班[d]俄亥俄州立大学数学系，美国俄亥俄州哥伦布[e]奥克兰大学工程科学与生物医学工程系，新西兰奥克兰[f]得克萨斯大学奥斯汀分校 *Walker*机械工程系，美国得克萨斯州奥斯汀。

摘要

对于经典的基于偏微分方程的方法，不确定性下的形状优化（OUU）计算成本高昂，因为需要针对大量不确定性实现和变化的几何形状，反复进行基于采样的风险评估；而标准神经代理模型往往无法为优化提供准确、高效的敏感度。我们提出 Shape-DINO，这是一种导数信息神经算子框架，用于学习一系列变化几何形状上的偏微分方程解算子，尤其着眼于加速受偏微分方程约束的形状不确定性下的优化。

Shape-DINO 通过微分同胚映射将几何变异性编码到固定参考域，并采用导数信息算子学习目标，联合学习偏微分方程解及其关于设计变量和不确定参数的弗雷歇导数，从而实现准确的状态预测，并为大规模不确定性下的优化提供可靠的梯度。

我们建立了将代理模型精度与优化误差联系起来的先验误差界，并证明了多输入降阶基神经算子在适当的 C^1 范数下具有通用逼近结果。我们通过三个具有代表性的形状不确定性下优化问题展示了方法的效率和可扩展性，其中包括泊松方程的边界设计，以及由二维和三维稳态纳维–斯托克斯外流所支配的形状设计。在这些示例中，与未使用导数信息训练的算子代理模型相比，Shape-DINO 得出的优化结果更可靠。在我们的示例中，Shape-DINO 在状态和梯度评估方面实现了 3–8 个数量级的加速。计入训练数据生成后，对于单个 OUU问题，与严格基于偏微分方程的方法相比，Shape-DINO 可将所需的偏微分方程求解次数减少 1–2 个数量级。此外，Shape-DINO 的构建成本可分摊到多个目标函数和风险度量上，从而能够针对复杂系统开展大规模形状不确定性下的优化。

关键词：科学机器学习、算子学习、形状优化、不确定性下的优化、通用逼近

目录

1 引言 31.1 贡献 .41.2 相关工作 .51.3 论文结构 .6

*通讯作者电子邮箱：xindi.gong@utexas.edu (Xindi Gong), dingcheng.luo@qut.edu.au (Dingcheng Luo), oleary-roseberry.1@osu.edu (Thomas O'Leary-Roseberry), ruanui.nicholson@auckland.ac.nz (Ruanui Nicholson), omar@oden.utexas.edu (Omar Ghattas)

2 表述与动机 7
2.1 希尔伯特空间上的算子学习符号说明 .72.2 参数化区域上的偏微分方程 .82.3 解算子的导数 .92.4 不确定性下的形状优化 .102.5 通过样本均值近似进行优化 .112.6 使用算子替代模型进行优化 .122.7 示例：期望的导数误差 .12

3 形状导数信息驱动的算子学习 14
3.1 降阶基架构.143.2 降阶基神经算子的通用逼近结果 .153.3 形状导数信息驱动的算子学习 .163.4 实用降维 .173.4.1 状态变量的本征正交分解.173.4.2 不确定参数的活跃子空间 .183.4.3 降阶基计算的基于样本的近似 .20

4 实现细节 21
4.1 基于基展开和弹性延拓的微分同胚变形 .214.2 算子学习的计算工作流程.224.2.1 细节与计算成本 .234.2.2 神经网络训练 .23**5 数值**

结果 24
5.1 潜在空间神经网络架构 .245.2 神经算子的成本精度比较 .24

5.3 不确定渗透率场下的泊松问题中的通量跟踪设计 25
5.3.1 具有变化上边界的泊松问题 .255.3.2 使用傅里叶基的形状参数化 .265.3.3 参考域公式 .265.3.4 不确定性下的通量跟踪形状优化.265.3.5 神经算子构造 .275.3.6 形状优化与成本-精度比较 .29

5.4 多种不确定来流条件下二维流动的耗散降低设计 . . 33
5.4.1 含可变形状物体的纳维–斯托克斯方程 .335.4.2 使用自由形变进行形状参数化 .345.4.3 参考域公式 .355.4.4 不确定来流条件下的二维减耗散形状设计 .365.4.5 神经算子构造 .375.4.6 形状优化与成本-精度比较 .38

5.5 多种不确定来流条件下的三维减阻设计 . . . . . 395.5.1 问题构建 .405.5.2 神经算子构造 .415.5.3 使用神经算子进行形状优化 .425.6 计时比较 .43

6 结论 s **45**

7 致谢 **47**

参考文献 **47**

A 空间域与参考域变换 **52**A.1 常用坐标变换 .52A.2 命题 2.1 的证明 .52**B 优化误差** **53**B.1 导数支配界 .53B.2 期望的导数误差 .55B.2.1 关注量的可微性 .57B.2.2 命题 2.2 的证明 .59**C 通用逼近结果** **60**C.1 有限维神经网络逼近 .60C.1.1 神经网络构造 .61C.1.2 具有受控增长的神经网络近似 . . . . . 63C.1.3 定理 C.1 的证明 .66C.2 Lipschitz光滑算子的降维误差 .67C.3 定理 3.1 的证明 .70C.4 不确定性优化的通用逼近 .71

1. 引言

物理系统的计算模型通常必须考虑可变或不确定的几何形状。前者的一个例子是形状优化：在优化过程中，需要针对不断变化的几何形状进行多次迭代，以最大化给定的性能目标；后者的一个例子是不确定性量化：水深地形或竣工几何等几何特征存在不确定性，因此必须对许多几何形状进行采样，将这种不确定性传播至感兴趣量。在许多情况下，几何变异性通过问题特定参数化来表示；除了空间变化的材料性质或强迫场等其他不确定性来源外，还必须同时考虑这些参数化，这进一步增加了对此类系统进行建模和分析的难度。我们尤其关注由偏微分方程（PDE）控制的系统。下文中，我们将几何（形状）参数记为 z，将不确定参数记为 m，将 PDE状态记为 u，并将隐式定义的解映射写作

$$(m, z) \mapsto u(m, z) \quad \text{subject to} \quad \mathcal{R}(u, m, z) = 0,$$

其中， $\mathcal{R}$ 表示控制偏微分方程的一种抽象残差形式。

在本研究中，我们开发了高效的神经算子学习方法，用于参数化解映射

$$(m, z) \mapsto u(m, z),$$

旨在加速涉及模型参数和区域同时变化的外循环任务。尽管所提出的算子学习框架广泛适用于多种此类任务，但本文将研究范围限定于不确定性下的形状优化（OUU）问题；当必须在模型参数存在不确定性的情况下做出几何设计决策时，自然会遇到这类问题。

我们通过将 m 建模为服从参考概率测度 μ_m的随机变量来处理参数不确定性；例如，该测度可依据先验信念或贝叶斯反问题的解来确定。因此，任何依赖于偏微分方程解的感兴趣量，即$Q = Q(u(m, z), m, z)$，都是随机变量，其分布反映了参数不确定性和几何变异性的综合影响。

与其优化 Q 的单次实现，不如优化一个标量风险度量 ρ，以体现设计者对不确定性 [1]的容忍程度。这便引出了形状不确定性下的优化问题

$$\min_z \mathcal{J}(z) = \rho_{m\sim\mu_m}(Q(u(m, z), m, z)) + \mathcal{P}(z), \tag{1}$$

其中 $\mathcal{P}(z)$ 是作用于几何的正则化项或代价项，用于施加设计约束或体现先验几何偏好。常见的 ρ 包括期望、均值–方差泛函、条件风险价值（CVaR）、最坏情况风险以及指数风险度量 [1–3]。这些度量使从业者能够在风险中性设计与高度保守的设计之间进行权衡；后者强调尾部事件或最坏情况性能 [4–7]。

Co计算挑战。形状不确定性下的优化面临若干相互关联的计算挑战 :

1. 几何变异性。对几何变化建模通常需要重新网格划分或网格变形，这会导致每次设计变量 z 变化都产生大量计算开销。2. 采样复杂度。估计风险度量需要针对不确定参数 m的多次实现反复评估正向模型，从而产生高得令人望而却步的采样成本。3. 快速在线评估。高效的（在线）查询能力对于降低不确定性下的优化所固有的高采样复杂度至关重要。4. 导数精度。非线性优化算法需要精确的梯度；若要快速收敛，还需要海森信息。在风险规避形状优化中，这意味着必须能够计算偏微分方程解相对于设计变量 z 的导数（即形状灵敏度）。

由于这些困难紧密耦合，当单次高保真偏微分方程求解的成本较高时，形状不确定性下的优化很快就会变得计算上难以处理；复杂的三维模拟通常正是如此。在这种情况下，采样、几何变化和导数评估的综合需求会使传统优化流程不堪重负。

这一计算瓶颈推动了先进代理建模技术的发展，例如神经算子，它们显式考虑几何变异性，从而能够实现形状优化等任务。在此背景下，神经算子 [8] 通过由权重 θ参数化的非线性代理模型 $u_\theta(m,z)$来近似参数化解映射$(m,z)\mapsto u(m,z)$，该代理模型在输入和输出中编码离散化不变结构，从而提供高效的近似。

然而，现有的神经算子方法通常尚未同时处理辅助空间场 m 与形状参数 z的参数化，也没有显式控制偏微分方程状态 u关于设计参数和不确定参数的弗雷歇导数之代理模型近似中的误差， $D_z u$ 和 $D_m u$。因此，将其部署于优化场景时，无法保证所得梯度的准确性，最终也无法保证所计算最优设计的质量。这一空白促使了本文所发展方法的提出。

1.1. 贡献

在本研究中，我们开发了形状导数信息 神经算子，以统一的方式应对上述挑战。我们通过微分同胚映射将一系列偏微分方程表述在公共参考域上，从而处理几何变异性。这一表述使整个学习问题能够在固定参考网格上进行，同时训练好的算子仍能够针对广泛的几何变换进行预测。

本研究的核心贡献是将导数信息算子学习扩展到这一几何复杂的场景中。我们构建了一个由形状变量 z 和辅助空间场 m共同参数化的统一代理模型，并明确训练该模型，使其能够准确近似偏微分方程解 u 及其相对于各组参数的弗雷歇导数。

我们的主要贡献如下。

(C1) 先验误差界，适用于代理模型驱动的优化与反演。我们分析了优化问题
- 的问题式（1）中，目标函数 $\mathcal{J}(z)$ 由代理目标函数 $\mathcal{J}_\theta(z)$ 近似: =

$\rho_{m\sim\mu}(Q(u_\theta(m,z)\ m,z))+\mathcal{P}(z)$，使用近似解算子 $u_\theta(m,z)$。具体而言，我们运用优化理论中的标准技术证明，当 $\mathcal{J}$ 为强凸时，代理目标函数 $\mathcal{J}_\theta$ 的一阶驻点的优化误差可以用目标函数导数的误差来界定，$D_z\mathcal{J}-D_z\mathcal{J}_\theta$（定理 2.1）。此外，在 $\rho(\cdot)=\mathbb{E}[\cdot]$的风险中性设定下，我们证明目标函数的导数误差由输出 $\|u(m,z)-u_\theta(m,z)\|_{\mathcal{U}}$的$L^2_\mu$ 误差与 z-导数 $\|D_zu(m,z)-D_zu_\theta(m,z))\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}$的误差共同界定。这些结果为风险中性优化中的导数信息算子学习提供了理论依据，也可以轻松推广到风险规避优化和反问题等其他场景。

(C2) 多输入神经算子的逼近理论。我们建立了神经算子的通用逼近结果，证明多输入神经算子 $u_\theta(m,\ z)$ 能够以任意高的精度逼近解算子 $u(m,\ z)$ 及其 z 阶导数 $D_zu(m,\ z)$，其适用范围是关于 z 的紧域，但在 m (定理 3.1 中，相关区域可能无界)。此外，在风险中性设定下，这意味着代理目标函数可产生优化误差任意小的驻点（定理 3.2）。这些结果证明，在不确定性条件下，可使用神经算子进行形状优化；这类优化问题由相应优化问题的一阶驻点刻画。

(C3) 形状导数信息驱动的算子学习高效算法。我们通过扩展来表示参考网格的任意微分同胚变换 χ_z，从而开发了适用于几何形状族的算子学习实用算法。我们提出了控制偏微分方程的高效表述形式，该方程既可定义在变形域上，也可等价地借助逆微分同胚 χ_z^{-1} 定义在参考域上。我们还阐明了解关于 m 和 z 的导数存在及其高效计算的条件，并结合隐函数定理、伴随算子方法和前向灵敏度方法进行求解。在此基础上，我们将多输入降阶基导数信息神经算子 [9] 扩展至形状参数化设定。

我们通过一系列具有挑战性的算子学习问题验证了所提出框架的有效性。这些问题涉及几何变异性，并最终应用于风险规避的偏微分方程约束形状优化问题。实证结果表明，所提出的方法具有两项关键优势：(i) 优化精度更高，这与针对风险中性设定推导的优化误差界一致；(ii) 单位计算开销下的泛化精度更高。

1.2. 相关工作

几何形状的微分同胚映射。微分同胚已广泛用于在各种场景下对计算域的变换进行参数化。在偏微分方程领域，这类方法在形状微积分和偏微分方程约束形状优化中发挥着核心作用。尤其值得关注的是由基展开构造的微分同胚，例如傅里叶基、B样条和非均匀有理B样条。由于表达能力强，且与计算机辅助设计紧密集成，这些方法常用于形状优化和推断（例如自由形变方法）[10–15]。其他经典进展包括基于速度的微分同胚形状微积分表述 [16–22], 、映射法[23, 24], ，以及通过基于偏微分方程的扩展算子构造微分同胚区域变换的方法，例如弹性形变 [25–28]。基于微分同胚的密切相关表述也见于医学成像领域，尤其是大形变配准和计算解剖学 [29–34]。

风险规避形状优化。关于不确定性下的风险规避形状优化，已有大量研究 [4–6，35–38]。这些方法通常依赖高保真数值离散化，例如有限元法。由于采样和重复进行偏微分方程求解的计算成本相互叠加，这限制了它们在高维随机环境中的适用性。类似方法也已应用于风险厌恶拓扑优化 [39–41]。

用于形状优化的机器学习方法。受传统高保真求解器进行形状优化的计算成本所推动，近年来，人们对基于机器学习的方法日益关注，这些方法旨在学习不同几何形状下的偏微分方程解算子。典型实例包括傅里叶神经算子（FNO）架构的几何感知变体 [42–44], 以及基于 DeepONet 的模型及其多输入扩展（MIONet） [45–47]。

人们提出了多种策略，将几何变异性纳入这些框架。这些策略包括学习变形映射，将不规则区域变换为与 FNO 兼容的规则潜在网格 [43], ；以及采用低维几何参数化方式，如控制点或非均匀有理B样条表示，用于 NACA 翼型优化等应用 [48], ，并在微分同胚域族上进行算子学习 [44, 47, 49]。此外，针对拓扑优化问题，也提出了基于神经算子的相关方法 [50]。

尽管这些方法在处理确定性几何变异性方面表现出色，但并未涉及辅助不确定参数或风险规避优化目标。此外，上述所有工作均采用标准 L^2_μ 训练形式，且未将弗雷歇导数信息纳入学习过程。

导数信息算子学习。本工作将导数信息算子学习 [51] 扩展至降阶基神经算子架构，使其能够同时处理不确定参数和变化的几何形状及其相关导数。本工作还基于 [9], ，该研究为风险规避偏微分方程约束优化开发了一个多输入降阶基导数信息神经算子框架。在此基础上，我们大幅拓展了逼近理论，并聚焦于风险规避形状优化；与以往工作相比，这一问题带来了诸多额外挑战。导数信息神经算子（DINO）在反问题等其他与优化相关的场景中也取得了显著改进 [52, 53], ，并已扩展到 DeepONet[54]和 FNO [55]等多种架构。

现有文献已针对多种架构充分探讨了神经算子的近似能力。与本文最相关的是针对约化基架构的研究成果，例如， [56–59], 在算子满足适当正则性假设的条件下，证明了其具有通用逼近性质，并给出了参数复杂度估计。此类研究通常关注算子的输出误差，而不关注其导数；然而，当使用神经算子近似优化问题的目标函数时，导数对于所得优化器的稳定性至关重要。针对这一点， [60]为约化基架构、 [55]为 FNO 提出了进一步纳入导数精度的通用逼近结果。本文在这些工作的基础上，针对不确定性下的优化（OUU）设置，提出约化基架构的通用逼近结果。

1.3. 论文结构

在第2节中，我们介绍本文所研究的偏微分方程问题类别，并正式阐述目标学习任务和优化任务。本节提供必要的数学背景，包括参考区域与变形域之间微分同胚映射的性质，并为以替代模型取代真实偏微分方程解算子的优化问题建立误差界。第3节提出导数信息算子学习问题，并给出一类降阶基神经算子架构的通用逼近结果。第4 节介绍所提框架的实际实现细节，包括通过基展开和弹性延拓进行网格变形，以及降阶基神经算子的构建。最后，第5 节展示数值实验，验证所提方法在若干具有挑战性的形状优化问题中的有效性，包括具有风险规避设计目标的不确定性下二维和三维流动优化。结果表明，在训练数据量固定的情况下，采用形状导数信息的方法比未使用导数信息训练的神经算子具有显著更高的优化精度。此外，在风险规避形状优化中，Shape-DINO 在生成最优设计方面的精度高于传统的基于偏微分方程的样本均值近似方法，同时所需的状态偏微分方程求解次数（包括生成训练数据的成本）减少了 1–2 个数量级。我们还进一步证明，

Shape-DINO 的构建成本可以分摊到多个设计目标和风险度量上。就在线成本而言，Shape-DINO 可使状态评估和梯度计算提速 3–8 个数量级。

2. 公式表述与动机

2.1. 希尔伯特空间上的算子学习记号

我们首先介绍一些用于希尔伯特空间上算子学习的常用记号。其他记号将在用到时再行介绍。在本文中，我们将考虑属于 $\mathbb{R}$上的可分希尔伯特空间（有限维或无限维）的状态变量、模型参数和设计变量。设 $\mathcal{X}$是这样的希尔伯特空间，我们用 $\langle v_1,\ v_2\rangle_{\mathcal{X}}$表示 $\mathcal{X}$上的内积，它生成范数 $\|v\|_{\mathcal{X}}=\sqrt{\langle v,\ v\rangle_{\mathcal{X}}}$。我们用 $\mathcal{X}'$表示其拓扑对偶，并用 $\langle l,\ \overline{v\rangle_{\mathcal{X}'\times\mathcal{X}}}$表示 $l\in\mathcal{X}'$与 $v\in\mathcal{X}$之间的对偶配对。对于希尔伯特空间，借助里斯映射$\mathrm{R}_{\mathcal{X}}\colon\ \mathcal{X}'\to\mathcal{X}$，可以将 $\mathcal{X}'$与 $\mathcal{X}$等同起来；也就是说，对于 $l\in\mathcal{X}'$, $\langle \mathrm{R}_{\mathcal{X}}l,\ \ v\rangle_{\mathcal{X}}=\langle l,\ \ v\rangle_{\mathcal{X}'\times\mathcal{X}}$.

我们用 $\mathcal{L}(\mathcal{X},\mathcal{Y})$ 表示两个希尔伯特空间 $\mathcal{X}$ 和 $\mathcal{Y}$ 之间的有界线性算子空间。这是一个赋有算子范数 $\|A\|_{\mathcal{L}(\mathcal{X},\mathcal{Y})}=\sup_{\|v\|_{\mathcal{X}}=1}\|Av\|_{\mathcal{Y}}$ 的巴拿赫空间。对于任意 $A\in\mathcal{L}(\mathcal{X},\mathcal{Y})$，我们令 $A^*\in\mathcal{L}(\mathcal{Y},\mathcal{X})$ 表示 A 的伴随算子，并令 $A^T\in\mathcal{L}(\mathcal{Y}',\mathcal{X}')$ 表示 A 的转置算子，使得$\langle w、\ Av\rangle_{\mathcal{Y}}=\langle A^*w、\ v\rangle_{\mathcal{X}}$以及 $\langle l、\ Av\rangle_{\mathcal{Y}'\times\mathcal{Y}}=\langle A^Tl、\ v\rangle_{\mathcal{X}'\times\mathcal{X}}$ 对任意 $v\in\mathcal{X}$、$w\in\mathcal{Y}$ 和 $l\in\mathcal{Y}'$ 均成立。特别地，当 $v\in\mathcal{X}$ 是一个向量时，我们有 $v^T=\mathrm{R}_{\mathcal{X}}^{-1}v$。我们用这一记号定义两个向量 v_1 和 v_2 的外积，记作 $v_1v_2^T$；它是由 $(v_1v_2^T)v_3=v_1\langle v_2、\ v_3\rangle_{\mathcal{X}}$ 给出的线性算子。

我们还将考虑希尔伯特–施密特算子空间 $\mathrm{HS}(\mathcal{X},\ \mathcal{Y})$，其中包含具有有限希尔伯特-施密特范数的连续线性算子，其定义为

$$\|A\|_{\mathrm{HS}(\mathcal{X},\mathcal{Y})}^2=\sum_{i=1}^{\infty}\|Av_i\|_{\mathcal{Y}}^2,$$

其中，$\{v_i\}_{i=1}^{\infty}$ 是 $\mathcal{X}$的任意正交归一基。我们注意到，当 $\mathcal{X}$ 或 $\mathcal{Y}$ 为有限维时，希尔伯特-施密特范数与算子范数等价。

我们将通过希尔伯特空间上的博雷尔概率测度 μ来表示模型参数中的不确定性。对于希尔伯特空间 $\mathcal{X}$和 $\mathcal{Y}$，我们用 $L_\mu^p(\mathcal{X};\ \mathcal{Y})$表示算子 $F\colon\ \mathcal{X}\to\mathcal{Y}$构成的博赫纳空间，这些算子是博赫纳可积的且范数有限，即

$$\|F\|_{L_\mu^p(\mathcal{X};\mathcal{Y})}^p:=\mathbb{E}_{x\sim\mu}[\|F(x)\|_{\mathcal{Y}}^p]=\int\|F(x)\|_{\mathcal{Y}}^p d\mu(x)<\infty$$

当 $p\in[1,\infty)$时，以及

$$\|F\|_{L_\mu^\infty(\mathcal{X};\mathcal{Y})}=\operatorname*{ess\,sup}_x\|F(x)\|_{\mathcal{Y}}.$$

当 $p=\infty$时。请注意，当输入和输出空间明确时，我们偶尔会用简写 L_μ^p 代替 $L_\mu^p(\mathcal{X};\ \mathcal{Y})$。

此外，我们还将考虑弗雷歇可微映射。对于此类映射 $F\colon\mathcal{X}\to\mathcal{Y}$，我们将其在 $x\in\mathcal{X}$处的导数记作 $D_xF(x)\in\mathcal{L}(\mathcal{X},\ \mathcal{Y})$，使得

$$\lim_{\|h\|_{\mathcal{X}}\to0}\frac{\|F(x+h)-F(x)-D_xF(x)h\|_{\mathcal{Y}}}{\|h\|_{\mathcal{X}}}=0.$$

此外，我们将连续弗雷歇可微映射的空间记为 $C^1(\mathcal{X};\ \mathcal{Y})$，或在考虑输入空间 $\mathcal{B}\subset\mathcal{X}$ 的某个子集时记为 $C^1(\mathcal{B};\ \mathcal{Y})$。

2.2. 参数化区域上的偏微分方程

对于形状优化，我们考虑一族由有限维形状变量 $z \in \mathcal{Z} = \mathbb{R}^{d_z}$ 参数化的空间区域 $\Omega_z \subset \mathbb{R}^d$，其中 $d = 2,3$。具体而言，我们假设每个 Ω_z 通过由形状参数定义的微分同胚 $\chi_z : \mathcal{D} \to \mathbb{R}^d$ 与参考域 Ω_0 相耦合，其中 $\mathcal{D} \subset \mathbb{R}^d$ 是包含 Ω_0 闭包的包容域。此外，我们假设参考域 Ω_0 有界且为利普希茨的，从而对于某个闭可行集 $\mathcal{Z}_{ad} \subset \mathcal{Z}$ 中的每个 z，所得空间域 $\Omega_z = \chi_z(\Omega_0)$ 同样有界且为利普希茨的。

我们关注定义在区域族 Ω_z 上的偏微分方程问题。这类偏微分方程问题将形状变量 z 以及任何其他模型参数 $\widetilde{m} \in \mathcal{M}(\Omega_z)$（例如随空间变化的系数场或强迫项）与状态变量 $\widetilde{u} \in \mathcal{U}(\Omega_z)$ 联系起来。这里，$\mathcal{M}(\Omega_z)$ 和 $\mathcal{U}(\Omega_z)$ 分别是模型参数和状态的可分希尔伯特空间，并定义在空间域 Ω_z 上。它们通常取为勒贝格空间 $L^2(\Omega_z)$ 或索伯列夫空间 $H^k(\Omega_z)$。于是，对于 Ω_z 的每个实例，都可以将偏微分方程问题抽象地表述为如下残差形式

$$\text{given } \Omega_z \text{ and } \widetilde{m} \in \mathcal{M}(\Omega_z), \quad \text{find } \widetilde{u} \in \mathcal{U}(\Omega_z) \quad \text{such that} \quad \mathcal{R}_z(\widetilde{u}, \widetilde{m}) = 0, \tag{2}$$

其中，$\mathcal{R}_z : \mathcal{U}(\Omega_z) \times \mathcal{M}(\Omega_z) \to \mathcal{V}(\Omega_z)'$ 是一个微分算子，其输出属于某个检验空间 $\mathcal{V}(\Omega_z)'$ 的拓扑对偶。

或者，我们可以在参考域上表述偏微分方程Ω_0，通过拉回 χ_z来实现。为此，设 $\mathcal{M}(\Omega_0)$ 和 $\mathcal{U}(\Omega_0)$ 为定义在参考域Ω_0上的相应函数空间。那么，式（2）可以等价地在Ω_0 上表述为

$$\text{given } z \in \mathcal{Z} \text{ and } m \in \mathcal{M}(\Omega_0), \quad \text{find } u \in \mathcal{U}(\Omega_0) \quad \text{such that} \quad \mathcal{R}(u, m, z) = 0, \tag{3}$$

其中，式（3）是基于以下事实由式（2）推导得到的：$u = \widetilde{u} \circ \chi_z$、$m = \widetilde{m} \circ \chi_z$和 $\mathcal{R}$：$\mathcal{U}(\Omega_0) \times \mathcal{M}(\Omega_0) \times \mathcal{Z} \to \mathcal{V}(\Omega_0)'$ 可通过对相关算子进行适当的变量替换得到。我们在附录 A.1 节列出了若干常用微分和积分运算的坐标变换。请注意，在式（3）中，形状变量以残差的形式显式出现。假设该偏微分方程对每个 $z \in \mathcal{Z}_{ad}$ 和 $m \in \mathcal{M}(\Omega_0)$均存在唯一解，因此可以抽象地写作 $u = u(m, z)$，其中我们将 u：$\mathcal{M}(\Omega_0) \times \mathcal{Z}_{ad} \to \mathcal{U}(\Omega_0)$称为解算子。

注 2.1（关于记号）。在本文余下部分中，我们采用以下记号约定。我们将在符号上方使用$\widetilde{\cdot}$，表示该量是根据空间域Ω_z定义的；而不带 $\widetilde{\cdot}$ 的量通常表示其参考域对应量，例如上文的 $\widetilde{u}$和 u。对于几何量，我们用 n、dx、ds 分别表示空间域上的单位法向量、体积元和面积元，用 N、dX、dS 表示参考域上的对应量。此外，若书写 $\mathcal{M}$、$\mathcal{U}$和 $\mathcal{V}$ 时未标明相应区域，则默认它们指的是参考域函数空间 $\mathcal{M} := \mathcal{M}(\Omega_0)$、$\mathcal{U} := \mathcal{U}(\Omega_0)$ 和 $\mathcal{V} := \mathcal{V}(\Omega_0)$。

在继续之前，我们先介绍后续章节将用到的推前和拉回算子的概念。对于任意微分同胚 $\chi : \Omega_0 \to \Omega_1$, ，令 T_χ 表示复合算子，$T_\chi u = u \circ \chi$。我们称 T_χ 为拉回算子，并称$T_\chi^{-1} = T_{\chi^{-1}}$为推前算子（即 $T_\chi^{-1} u = u \circ \chi^{-1}$）。根据这些定义，式（2）和（3）中使用的变量替换可表示为

$$u = T_{\chi_z} \widetilde{u}, \quad m = T_{\chi_z} \widetilde{m} \quad \text{or} \quad \widetilde{u} = T_{\chi_z}^{-1} u, \quad \widetilde{m} = T_{\chi_z}^{-1} m,$$

当 m 和 u 是标量场时。

当 $\chi_z \in W^{1,\infty}(\Omega_0)$ 和 $\chi_z^{-1} \in W^{1,\infty}(\Omega_z)$ 时，所得复合算子是参考区域/空间区域上低阶索伯列夫空间之间的有界线性算子。

命题 2.1. 假设 $\chi : \Omega_0 \to \Omega_1$ 是一个微分同胚，使得 $\chi \in W^{1,\infty}(\Omega_0)$ 且 $\chi^{-1} \in W^{1,\infty}(\Omega_1)$。我们有如下结论：

1. 前推算子与拉回算子是介于 $L^2(\Omega_0)$ 与 $L^2(\Omega_1)$ 之间的有界线性算子，即$T_\chi^{-1} \in \mathcal{L}(L^2(\Omega_0)\ L^2(\Omega_1))$ 与 $T_\chi \in \mathcal{L}(L^2(\Omega_1)\ L^2(\Omega_0))$。此外，它们的范数有界于

$$\|T_\chi^{-1}\|^2_{\mathcal{L}(L^2(\Omega_0),L^2(\Omega_1))} \le \|\chi\|^d_{W^{1,\infty}(\Omega_0)}, \tag{4a}$$

$$\|T_\chi\|^2_{\mathcal{L}(L^2(\Omega_1),L^2(\Omega_0))} \le \|\chi^{-1}\|^d_{W^{1,\infty}(\Omega_1)}. \tag{4b}$$

2. 前推算子和拉回算子是 $H^1(\Omega_0)$ 与 $H^1(\Omega_1)$ 之间的有界线性算子，即 $T_\chi^{-1} \in \mathcal{L}(H^1(\Omega_0)\ H^1(\Omega_1))$ 和 $T_\chi \in \mathcal{L}(H^1(\Omega_1)\ H^1(\Omega_0))$。此外，它们的范数有界于

$$\|T_\chi^{-1}\|^2_{\mathcal{L}(H^1(\Omega_0),H^1(\Omega_1))} \le \left(1 + \|\chi^{-1}\|^2_{W^{1,\infty}(\Omega_1)}\right) \|\chi\|^d_{W^{1,\infty}(\Omega_0)}, \tag{5a}$$

$$\|T_\chi\|^2_{\mathcal{L}(H^1(\Omega_1),H^1(\Omega_0))} \le \left(1 + \|\chi\|^2_{W^{1,\infty}(\Omega_0)}\right) \|\chi^{-1}\|^d_{W^{1,\infty}(\Omega_1)}. \tag{5b}$$

证明。 参见第A.2节。 □

将（2）形式的问题等价地表述为（3）形式，这一点对本研究至关重要。正因如此，我们可以通过形状参数 z 的参数化，为所有 Ω_z 上的 Ω_0，表述形状优化问题和算子学习任务。此时，该参数显式出现在残差算子 $\mathcal{R}$ （3）中。

2.3. 解算子的导数

在本文中，我们假设解算子关于 m 和z 连续可微。通常，这使我们能够通过基于导数的优化算法求解涉及偏微分方程的形状优化问题。在后续章节推导正式的理论结果时，我们将提出更具体的假设。

由于我们直接在参考域中进行工作，弗雷歇导数算子 $D_m u(m,z)$和$D_z u(m,z)$可通过偏微分方程(3)的隐式微分来计算，如同标准的正向与伴随灵敏度方法，而无需显式涉及形状微积分。也就是说，在任意 m、z 和$u(m,z)$处，以及对输入空间的任意方向$(h_m, h_z) \in \mathcal{M} \times \mathcal{Z}$，导数作用，即雅可比-向量积(JVP)， $h_u := D_m u(m,z)h_m + D_z u(m,z)h_z$ 由线性化状态方程的解给出

$$D_u\mathcal{R}(u,m,z)h_u = -\left(D_m\mathcal{R}(u,m,z)h_m + D_z\mathcal{R}(u,m,z)h_z\right). \tag{6}$$

另一方面，为了计算导数的伴随算子在某个方向 $v \in \mathcal{U}$上的作用，即向量雅可比积(VJP)，我们求解伴随方程

$$D_u\mathcal{R}(u,m,z)^T p = -v^T, \tag{7}$$

以得到伴随变量 $p \in \mathcal{U}$，使得

$$v^T D_m u(m,z) = D_m\mathcal{R}(u,m,z)^T p, \qquad v^T D_z u(m,z) = D_z\mathcal{R}(u,m,z)^T p. \tag{8}$$

上述形式表明，在已经计算出解 $u = u(m,z)$之后，随后计算单个VJP或JVP的代价相当于用算子 $D_u\mathcal{R}$或其伴随算子（转置算子）求解单个线性偏微分方程。此外，在单个点$(m,z,u(m,z))$处，多个导数作用的计算始终共享同一个线性算子 $D_u\mathcal{R}$，唯一的变化是(6)或(7)的右端项；当线性代数运算可以摊销时，这通常会带来显著的计算节省。

需要指出的是，尽管我们常将关于 z 的导数称为形状导数，但这种求导方式与物质导数 的概念更为接近，因为在形状优化中，这些导数不仅受区域边界 $\partial\Omega_z$的运动影响，也受内部区域运动的影响，而内部区域的运动由 χ_z 的定义决定（例如，参见 [61,第 11]章）。

2.4. 不确定性下的形状优化

令 $Q_z: \mathcal{U}(\Omega_z) \times \mathcal{M}(\Omega_z) \to \mathbb{R}$ 定义一个标量关注量（QoI），它取决于模型参数和状态变量，并以形状变量 z 为索引，这些变量定义了区域 Ω_z。通常，这些量被表述为区域 Ω_z 或其子集上的积分量，例如跟踪目标、力、能量等。在形状优化问题中，我们希望相对于 $Q_z(\widetilde{u}, \widetilde{m})$ 进行最小化，$z \in \mathcal{Z}_{ad}$ 同时满足约束条件：$(\widetilde{u}, \widetilde{m})$ 满足偏微分方程（2）。与偏微分方程残差的情形一样，我们可以通过将关注量表示为参考区域上的量来重新表述它，从而得到

$$Q(u, m, z) = Q_z(\widetilde{u}, \widetilde{m}), \tag{9}$$

其中 $u = T_\chi \widetilde{u}$ 且 $m = T_\chi \widetilde{m}$。接下来，目标是最小化 $Q(u, m, z)$，其中 $u = u(m, z)$ 满足（3）。

通常，模型参数 m 具有不确定性，这意味着对于每个固定的 z，相应的感兴趣量取值也具有不确定性。在本研究中，我们通过其概率分布μ_m来描述 m的不确定性，该分布μ_m定义为 $\mathcal{M} = \mathcal{M}(\Omega_0)$ 上的博雷尔概率测度。随后，对于每个 $z \in \mathcal{Z}_{ad}$，量$Q(u(\cdot, z) \cdot, z)$ 是一个随机变量，其取值取决于 m 的实现。因此，在对 $Q(u(\cdot, z) \cdot)$ 作出适当的矩假设时（例如，它属于 $L^p_{\mu_m}(\mathcal{M}; \mathbb{R})$），可以利用风险度量 ρ 来表述形状优化问题：$L^p_{\mu_m} \to \mathbb{R}$；该泛函提供了分布的统计摘要，其选择旨在表达特定的风险偏好。不确定性下的形状优化问题可表示为

$$\min_{z \in \mathcal{Z}_{ad}} \mathcal{J}(z) = \rho_{m \sim \mu_m}(Q(u(m, z), m, z)) + \mathcal{P}(z) \tag{10}$$

其中 $\mathcal{P}: \mathcal{Z} \to \mathbb{R}$ 是一个惩罚项/正则化项，用于计入与 z 相关的形状的任何额外成本。

在本文中，我们在数值算例中考虑以下风险度量，下面将以一般随机变量 $\xi \sim \mu_\xi$ 为例进行介绍。

均值。 风险度量的典型例子是随机变量的均值（或期望），

$$\rho^{\text{mean}}_{\xi \sim \mu_\xi}(\xi) = \mathbb{E}_{\xi \sim \mu_\xi}[\xi]. \tag{11}$$

均值是风险中性的，并未明确考虑分布的上尾事件，而这些事件往往代表工程系统中的灾难性故障。因此，我们还考虑两个风险规避泛函作为例子。

条件风险价值。 条件风险价值（CVaR）定义为针对分位数 $\beta \in [0, 1]$

$$\rho^{\text{CVaR},\beta}_{\xi \sim \mu_\xi}(\xi) = \min_{t \in \mathbb{R}} t + \frac{1}{1-\beta}\mathbb{E}[(\xi - t)^+], \tag{12}$$

其中 $(x)^+ := \max(x, 0)$。当底层随机变量 ξ 服从连续分布时，CVaR 可解释为在其超过 β 分位数的条件下，ξ 的条件期望。因此，CVaR 可用于明确控制分布的上尾，并已在工程 [3, 62]的风险规避设计中得到研究。

熵风险度量。 熵风险度量定义为

$$\rho^{\text{entropic},\beta}_{\xi \sim \mu_\xi}(\xi) = \frac{1}{\beta}\log\left(\mathbb{E}[\exp(\beta\xi)]\right), \tag{13}$$

其中，参数 $\beta > 0$ 用于指定风险厌恶程度。熵风险度量也已用于物理系统的风险规避优化，例如文献 [63]。

注 2.2（关于模型参数的分布）。 在本文中，我们关注模型参数 μ_m 的分布直接定义在参考域上的情形。在这种情况下，对于任意 z，空间域 Ω_z 中的不确定参数场，都是由定义在参考域 Ω_0，上的随机场通过 χ_z（就空间变量而言）前推得到的，即 $\widetilde{m} = T^{-1}_{\chi_z} m$ 和 $m \sim \mu_m$。这样选择是为了简化记号和所需的变量变换。尽管如此，我们的公式和流程也可以扩展到以下情形：不确定参数场通过对某个底层随机元进行更复杂的变换，定义在空间域上。

2.5. 通过样本均值近似进行优化

求解 OUU问题（10）的一种常规方法，是用从独立同分布样本 $m^{(i)} \sim \mu_m, i = 1, \ldots, n_{\mathrm{SAA}}$. 得到的样本均值近似（SAA）替代期望。由此定义一个新的目标函数，其中（10）中的风险度量部分由其 SAA 替代。

例如，使用 $\rho = \rho^{\mathrm{mean}}$ 时，目标函数的样本平均近似为

$$\widehat{\mathcal{J}}^{\mathrm{mean}}(z) := \frac{1}{n_{\mathrm{SAA}}} \sum_{i=1}^{n_{\mathrm{SAA}}} Q(u(m^{(i)}, z), m^{(i)}, z). \tag{14}$$

类似地，对于熵风险度量，我们取

$$\widehat{\mathcal{J}}^{\mathrm{entropic}}(z) := \frac{1}{\beta} \log \left(\frac{1}{n_{\mathrm{SAA}}} \sum_{i=1}^{n_{\mathrm{SAA}}} \exp(\beta Q(u(m^{(i)}, z), m^{(i)}, z)) \right). \tag{15}$$

对于条件风险价值，除样本平均近似外，我们还将对最大值函数取光滑近似，$(x)_\varepsilon^+ \approx (x)^+$,

$$(x)_\epsilon^+ = \begin{cases} 0, & \text{if } x < 0 \\ \frac{x^3}{\epsilon^2} - \frac{x^4}{2\epsilon^3}, & \text{if } 0 \le x \le \epsilon \\ x - \frac{\epsilon}{2}, & x \ge \epsilon, \end{cases} \tag{16}$$

其中平滑参数 $\epsilon \ll 1$ [3]较小。这种平滑处理保持了凸性，并确保一致收敛 $(x)_\epsilon^+ \to (x)^+$ 当 $\epsilon \to 0$ 时。由此得到的 CVaR 目标函数近似为

$$\widehat{\mathcal{J}}^{\mathrm{CVaR},\beta,\varepsilon}(z, t) := t + \frac{1}{n_{\mathrm{SAA}}(1-\beta)} \sum_{i=1}^{n_{\mathrm{SAA}}} (Q(u(m^{(i)}, z), m^{(i)}, z) - t)_\varepsilon^+, \tag{17}$$

它保持了关于 (z, t) 的可微性。最后，目标函数的样本平均近似是风险度量与惩罚项的组合，我们对 $z \in \mathcal{Z}_{ad}$ 进行最小化

$$\min_{z \in \mathcal{Z}_{ad}} \widehat{\mathcal{J}}(z) := \widehat{\mathcal{J}}^{\mathrm{risk}}(z) + \mathcal{P}(z) \tag{18}$$

或者对于条件风险价值，还需要对 $t \in \mathbb{R}$ 进行优化。

随后，可通过传统的基于导数的优化算法最小化所得的样本平均近似目标函数，例如拟牛顿法（如果问题足够光滑，也可采用牛顿法），其中导数可通过伴随方法高效计算。也就是说，为了计算映射 $z \mapsto Q(u(m^{(i)}, z), m^{(i)}, z)$ 关于 z 的导数，我们可以利用链式法则，

$$D_z(Q(u(m^{(i)}, z), m^{(i)}, z)) = D_u Q(u(m^{(i)}, z), m^{(i)}, z)) D_z u(m^{(i)}, z) + D_z Q(u(m^{(i)}, z), m^{(i)}, z).$$

第一项只是解算子 $(v^{(i)})^T D_z u(m^{(i)}, z)$ 在方向

$$v^{(i)} = \mathsf{R}_{\mathcal{V}} D_u Q(u(m^{(i)}, z), m^{(i)}, z)$$

上的一个向量-雅可比积，可通过求解伴随偏微分方程(7)来计算，对每个 $m^{(i)}$ 求解一次；而后一项通常有解析形式，无需额外求解偏微分方程。

显然，不确定性下的形状优化的每次迭代计算成本往往主要由以下两部分主导：（1）通过 n_{SAA} 个状态偏微分方程的解对目标函数求值（即对$u(m^{(i)}, z)$求值），以及（2）随后通过 n_{SAA} 个伴随方程的解对其梯度求值。因此，精确求解偏微分方程约束的不确定性下优化问题往往会导致难以承受的计算成本。当风险度量需要大量样本才能准确估计，且底层偏微分方程求解代价高昂时，这一点尤为突出。

2.6. 使用算子替代模型进行优化

为了减轻与偏微分方程约束的 OUU 相关的计算成本，可以用保真度较低但求值速度快的替代模型来替换高保真偏微分方程模型。在本工作中，我们关注算子代理模型，即对解算子的近似 $u_\theta(m,z) \approx u(m,z)$。由此定义了对优化问题（10）的一个近似，即

$$\min_{z\in\mathcal{Z}_{ad}} \mathcal{J}_\theta(z) := \rho_{m\sim\mu_m}(Q(u_\theta(m,z),m,z)) + \mathcal{P}(z) \tag{19}$$

其中使用代理目标函数 $\mathcal{J}_\theta(z)$ 近似真实目标函数 $\mathcal{J}(z)$。由于替代模型的计算速度很快，因此可以使用包含大量样本的样本平均近似（SAA）快速求解 (19)，从而基本消除抽样误差。相反，替代模型的逼近误差会成为求解 OUU问题时的主要误差来源。因此，基于代理的优化需要权衡替代模型的逼近精度与其构建成本，后者以训练数据集的大小衡量。通常，训练数据集由输入—输出对组成，即 $(m^{(i)},z^{(i)}) \mapsto u^{(i)} = u(m^{(i)},z^{(i)})$，这些数据对分布在输入空间中，每个数据元组均通过求解底层偏微分方程生成。

此外，正如 [9], 所指出的，当 $\mathcal{Z}$ 维度较高时，仍需采用基于导数的算法才能有效求解 (19)（或其样本平均近似）。因此，替代模型在近似 $\mathcal{J}(z)$及其导数 $D_z\mathcal{J}(z)$（或等价地，其梯度 $\mathsf{R}_{\mathcal{Z}}D_z\mathcal{J}(z)$）时引入的误差都很重要。特别是，导数不匹配（$D_z\mathcal{J}_\theta(z)$ 相较于 $D_z\mathcal{J}(z)$）可能会在 $\mathcal{J}_\theta(z)$中引入伪局部极小值，即使原始的偏微分方程约束 OUU问题是适定的。

例如，当底层目标函数是强凸且可微时，近似驻点的误差可由其梯度误差明确控制。我们在下述定理中对此进行总结。该定理的证明基于标准变分分析，详见第 B节。

定理 2.1。 假设 $\mathcal{B}\subset\mathcal{Z}$ 是一个开集，且 $\mathcal{J}\in C^1(\mathcal{B};\mathbb{R})$ 是一个 λ-强凸目标函数，即对于给定的 $\lambda>0$,

$$\mathcal{J}(z_2) \geq \mathcal{J}(z_1) + \langle D_z\mathcal{J}(z_1), z_2-z_1\rangle_{\mathcal{Z}'\times\mathcal{Z}} + \frac{\lambda}{2}\|z_2-z_1\|_{\mathcal{Z}}^2 \tag{20}$$

对于任意 z_1、$z_2\in\mathcal{B}$均成立。此外，假设 $\mathcal{Z}_{ad}\subset\mathcal{B}$ 是一个闭凸非空集，并令 $z^\star$ 为 $\mathcal{J}$ 在 $\mathcal{Z}_{ad}$上的唯一极小化点。此外，令 $\mathcal{J}_\theta\in C^1(\mathcal{B};\mathbb{R})$ 为 $\mathcal{J}$ 的一个近似，并令$z^\dagger\in\mathcal{Z}_{ad}$ 为任意满足一阶最优性条件的点，即

$$\langle D_z\mathcal{J}_\theta(z^\dagger), z-z^\dagger\rangle_{\mathcal{Z}'\times\mathcal{Z}} \geq 0 \qquad \forall z\in\mathcal{Z}_{ad}. \tag{21}$$

那么，优化误差有如下界:

$$\|z^\star - z^\dagger\|_{\mathcal{Z}} \leq \frac{1}{\lambda}\|D_z\mathcal{J}(z^\dagger) - D_z\mathcal{J}_\theta(z^\dagger)\|_{\mathcal{Z}'} \tag{22}$$

且最优性间隙有如下界:

$$\mathcal{J}(z^\dagger) - \mathcal{J}(z^\star) \leq \frac{1}{2\lambda}\|D_z\mathcal{J}(z^\dagger) - D_z\mathcal{J}_\theta(z^\dagger)\|_{\mathcal{Z}'}^2. \tag{23}$$

2.7. 示例：期望的导数误差

定理 2.1 表明，控制导数误差 $D_z\mathcal{J} - D_z\mathcal{J}_\theta$ 非常重要。形式上，根据链式法则，该误差取决于代理模型的输出误差 $u(m,\ z) - u_\theta(m,\ z)$ 以及导数误差 $D_zu(m,\ z) - D_zu_\theta(m,\ z)$，如 [9] 所示。作为示例，我们将考察期望的情形 ρ_{mean} 并展示这种依赖关系的具体形式。

为此，我们将关注在紧容许集上满足额外连续性与矩假设的解算子。首先，回顾我们使用简写记号 $\mathcal{M}:=\mathcal{M}(\Omega_0)$ 和 $\mathcal{U}:=\mathcal{U}(\Omega_0)$ 来表示参考域函数空间。随后，我们对（参考域中的）解算子以及 m 的分布作出如下假设。

假设 2.1. 给定 $\mathcal{M}$ 上具有有限 p^{th} 阶矩的博雷尔概率测度 μ_m 、紧容许集 $\mathcal{Z}_{ad} \subset \mathcal{Z} = \mathbb{R}^{d_z}$ 以及开集 $\mathcal{B} \supset \mathcal{Z}_{ad}$，假设以下成立：

1. 解算子 u： $\mathcal{M} \times \mathcal{B} \to \mathcal{U}$ 是连续可微的。

2. 存在常数 $\kappa_u > 0$ 和 $\kappa_{u_z} > 0$，使得对每个 z_1、$z_2 \in \mathcal{B}$ 以及对每个 m_1、$m_2 \in \mathcal{M}$，

$$\|u(m_2, z_2) - u(m_1, z_1)\|_{\mathcal{U}} \leq \kappa_u \left(\|m_2 - m_1\|_{\mathcal{M}} + \|z_2 - z_1\|_{\mathcal{Z}}\right), \tag{24a}$$
$$\|D_z u(m_2, z_2) - D_z u(m_1, z_1)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})} \leq \kappa_{u_z} \left(\|m_2 - m_1\|_{\mathcal{M}} + \|z_2 - z_1\|_{\mathcal{Z}}\right). \tag{24b}$$

我们还将采用关于关注量 Q： $\mathcal{U} \times \mathcal{M} \times \mathcal{Z} \to \mathbb{R}$ 的假设，这些假设能够容纳 u中的二次型，因为它们经常出现在最优设计问题中（例如跟踪目标、减阻）。

假设 2.2. 给定紧容许集 $\mathcal{Z}_{ad} \subset \mathcal{Z} = \mathbb{R}^{d_z}$和开集 $\mathcal{B} \supset \mathcal{Z}_{ad}$，假设以下条件成立：

1. 关注量 Q： $\mathcal{U} \times \mathcal{M} \times \mathcal{B} \to \mathbb{R}$ 是二阶连续可微的。

2. 存在 t常数 $\kappa_{Q_u} > 0$ 和 $\kappa_{Q_{uz}} > 0$，使得对每一个 u_1、$u_2 \in \mathcal{U}$、$z \in \mathcal{B}$以及每一个 $m \in \mathcal{M}$，

$$\|D_u Q(u_2, m, z) - D_u Q(u_1, m, w)\|_{\mathcal{U}'} \leq \kappa_{Q_u} \|u_2 - u_1\|_{\mathcal{U}}, \tag{25a}$$
$$\|D_{uz} Q(u_2, m, z) - D_{uz} Q(u_1, m, z)\|_{\mathcal{L}(\mathcal{U},\mathcal{Z}')} \leq \kappa_{Q_{uz}} \|u_2 - u_1\|_{\mathcal{U}}. \tag{25b}$$

3. 存在常数 M_Q、M_{Q_u}、M_{Q_z}、$M_{Q_{uz}} \geq 0$，使得对每一个 $z \in \mathcal{B}$ 以及每一个 $m \in \mathcal{M}$，

$$|Q(0, m, z)| \leq M_Q, \tag{26a}$$
$$\|D_u Q(0, m, z)\|_{\mathcal{U}'} \leq M_{Q_u}, \tag{26b}$$
$$\|D_z Q(0, m, z)\|_{\mathcal{Z}'} \leq M_{Q_z}, \tag{26c}$$
$$\|D_{uz} Q(0, m, z)\|_{\mathcal{L}(\mathcal{U},\mathcal{Z}')} \leq M_{Q_{uz}}. \tag{26d}$$

在假设 2.1和假设 2.2下，我们可以证明，期望的梯度近似误差同时依赖于以 $L^2_{\mu_m}$ 意义度量的输出误差和导数误差。这一点总结在如下命题中，其证明见第B.2节。

命题 *2.2.* 设 u、u_θ 满足假设 *2.1*，且 Q 满足假设 *2.2*，其中博雷尔概率测度为 μ_m，紧容许集为 $\mathcal{Z}_{ad}$，开集为 $\mathcal{B}$ （对某个 $p \geq 2$）。令 $\mathcal{J}^{\mathrm{mean}}(z) := \mathbb{E}_{m\sim\mu_m}[Q(u(m, z)\ m, z)]$ 且 $\mathcal{J}$均值$_\theta(z) := \mathbb{E}_{m\sim\mu_m}[Q(u_\theta(m, z)\ m, z)]$。则对任意 $z \in \mathcal{Z}_{ad}$，代价函数的导数误差由下式给出

$$\begin{aligned}
&\|D_z \mathcal{J}^{\mathrm{mean}}(z) - D_z \mathcal{J}^{\mathrm{mean}}_\theta(z)\|_{\mathcal{Z}'} \\
&\quad \leq C_1 \left(\|u(\cdot, z) - u_\theta(\cdot, z)\|_{L^2_{\mu_m}(\mathcal{M};\mathcal{U})} + \|D_z u(\cdot, z) - D_z u_\theta(\cdot, z)\|_{L^2_{\mu_m}(\mathcal{M};\mathrm{HS}(\mathcal{Z},\mathcal{U}))}\right) \\
&\quad + C_2 \left(\|u(\cdot, z) - u_\theta(\cdot, z)\|^2_{L^2_{\mu_m}(\mathcal{M};\mathcal{U})} + \|D_z u(\cdot, z) - D_z u_\theta(\cdot, z)\|^2_{L^2_{\mu_m}(\mathcal{M};\mathrm{HS}(\mathcal{Z},\mathcal{U}))}\right),
\end{aligned} \tag{27}$$

其中， C_1 和 C_2 是与 u_θ无关的常数。

注 2.3 (关于假设2.1 和2.2)。 通过考察命题*2.2*的证明，可以得到其他结果：假设*2.1* 和*2.2* 中的利普希茨常数和界常数可以依赖于 m，但需满足适当的矩假设。因此，梯度误差将取决于输出的高阶矩以及导数误差。对于具有以下形式的通用风险度量，也可以考虑梯度误差： $f_2(\mathbb{E}[f_1(\xi)])$，其中 f_1 和 f_2需满足适当的增长条件。类似地，这也可能使梯度误差取决于输出的高阶矩以及导数误差。为简洁起见，我们不探讨这些扩展。

3. 形状导数信息驱动的算子学习

受前述讨论启发，我们旨在为形状变化系统构建算子学习方案，并为控制上一节讨论的导数误差提供理论和实践保证。本节将推导若干通用逼近定理，说明采用降阶基架构时，如何同时控制算子误差以及相对于形状参数的导数。这一通用逼近理论专门针对 OUU问题；该问题并不特别要求控制相对于 m 的导数。随后，我们将讨论形状导数信息神经算子的实用训练方案。我们还将介绍一种有效策略，用于计算参数空间和状态空间的降阶基，并讨论在不确定性下的形状优化背景中所需的调整。

3.1. 降阶基架构

本节将重点放在降阶基架构上，这类架构已广泛应用于算子学习 [56, 64], ，尤其是在导数信息算子学习的背景下 [51, 60]，并应用于反问题和优化 [9, 52, 53]。提醒一下，我们在参考域 Ω_0 上构建算子学习任务，如式（3）所示；解和弗雷歇导数可通过推前／拉回算子映射到物理坐标 T_χ 和 T_χ^{-1}。

由于假设 $\mathcal{M} = \mathcal{M}(\Omega_0)$ 和 $\mathcal{U} = \mathcal{U}(\Omega_0)$ 是可分希尔伯特空间，因此我们可以用标准正交基 $\{\phi_i\}_{i=1}^\infty$ 表示 $\mathcal{U}(\Omega_0)$ 中的函数，以及用标准正交基 $\{\psi_i\}_{i=1}^\infty$ 表示 $\mathcal{M}(\Omega_0)$ 中的函数，从而对于任意 $u \in \mathcal{U}(\Omega_0)$ 和 $m \in \mathcal{M}(\Omega_0)$,

$$u = \sum_{i=1}^{\infty} \langle u, \phi_i \rangle_{\mathcal{U}} \phi_i \qquad \langle \phi_i, \phi_j \rangle_{\mathcal{U}} = \delta_{ij}, \tag{28a}$$

$$m = \sum_{i=1}^{\infty} \langle m, \psi_i \rangle_{\mathcal{M}} \psi_i \qquad \langle \psi_i, \psi_j \rangle_{\mathcal{M}} = \delta_{ij}, \tag{28b}$$

其中 δ_{ij} 表示克罗内克δ。降阶基架构通过截断的基展开来表示每个形式上无限维的场。这对应于线性编码与解码操作，我们在表 1中对此进行了定义。随后，使用神经网络来表示输入和输出的约化基系数之间的映射。

	编码	**解码**
$\mathcal{U}$ 使用基 $\{\phi_i\}_i$	$\mathcal{U} \ni u \mapsto \mathbf{\Phi}_{r_u}^* u = (\langle u, \phi_i \rangle_{\mathcal{U}})_{i=1}^{r_u} \in \mathbb{R}^{r_u}$	$\mathbb{R}^{r_u} \ni c \mapsto \mathbf{\Phi}_{r_u} c = \sum_{i=1}^{r_u} c_i \phi_i \in \mathcal{U}$
$\mathcal{M}$ 使用基 $\{\psi_i\}_i$	$\mathcal{M} \ni m \mapsto \mathbf{\Psi}_{r_m}^* m = (\langle m, \psi_i \rangle_{\mathcal{M}})_{i=1}^{r_m} \in \mathbb{R}^{r_m}$	$\mathbb{R}^{r_m} \ni c \mapsto \mathbf{\Psi}_{r_m} c = \sum_{i=1}^{r_m} c_i \psi_i \in \mathcal{M}$
$\mathcal{Z} = \mathbb{R}^{d_z}$	恒等	恒等

表 1：编码与解码操作

对于输出基 $\{\phi_i\}_{i=1}^\infty$，一种自然的选择是本征正交分解（POD） [56, 64, 65]，也称为主成分分析（PCA）。对于输入基 $\{\psi_i\}_{i=1}^\infty$，算子学习中常用的选择包括 POD（PCA），以及利用底层映射的导数构造的主动子空间（AS）基 [60, 66, 67]。我们将在第3.4节详细介绍这些方法。对于形状，我们假设 z 是微分同胚 χ_z的有限参数化，并且不再进行进一步降维。不过，当假设的可行设计空间为无限维或高维时，也可以考虑类似的降维策略。

在降阶基神经算子（RBNO）中，输入系数空间到输出系数空间的映射通常由全连接神经网络近似， g_θ : $\mathbb{R}^{r_m} \times \mathbb{R}^{d_z} \to \mathbb{R}^{r_u}$。该网络接收输入 $a_0 = (m_r, z) \in \mathbb{R}^{r_m + d_z}$，并通过 d_L映射得到输出 $a_{d_L+1} \in \mathbb{R}^{r_u}$，即 $a_{d_L+1} = g_\theta(a_0)$

宽度为 $d_{W,1}, d_{W,2}, \ldots, d_{W,d_L}$ 的隐藏层，其中每个 a_i 是

$$a_i = \sigma(W_{i-1}a_{i-1} + b_{i-1}), \quad i = 1, 2, \ldots, d_L,$$
$$a_{d_L+1} = W_{d_L}a_{d_L} + b_{d_L}.$$

网络参数集合为 $\theta = \{(W_i,\ b_i)\}_{i=0}^{d_L}$，其中权重的维度为

$$W_0 \in \mathbb{R}^{d_{W,1}\times(r_m+d_z)}, \quad W_i \in \mathbb{R}^{d_{W,i+1}\times d_{W,i}}, i = 1, 2, \ldots, d_{L-1}, \quad W_{d_L} \in \mathbb{R}^{r_u\times d_{W,d_L}},$$

偏置的维度为

$$b_i \in \mathbb{R}^{d_{W,i+1}}, \qquad i = 0, \ldots, d_{L-1}, \quad b_{d_L} \in \mathbb{R}^{r_u}.$$

这里，我们着重 讨论 g_θ 采用高斯误差线性单元（GELU）作为激活函数的情况 ，

$$\sigma_{\mathrm{GELU}}(x) = x\Phi_{\mathrm{Gauss}}(x), \tag{29}$$

其中， $\Phi_{\mathrm{Gauss}}(x)$ 是标准高斯分布的累积分布函数。因此，完整的神经算子架构形式为

$$u_\theta(m, z) = \mathbf{\Phi}_{r_u} g_\theta(\mathbf{\Psi}^*_{r_m}(m - b_m), z) + b_u, \tag{30}$$

其中， $b_u \in \mathcal{U}(\Omega_0)$ 和 $b_m \in \mathcal{M}(\Omega_0)$ 是偏置项，可根据降阶基的具体选择进行设定。

3.2. 降阶基神经算子的通用逼近结果

在假设 2.1下，我们可以证明，降阶基架构（30）可以在以下意义下，以任意精度近似解算子 $u(m,\ z)$ 及其 z 阶导数:

定理 *3.1*（多输入神经算子的通用逼近）。设 u 满足假设 *2.1*，其中博雷尔概率测度为 μ_m，紧容许集为 $\mathcal{Z}_{ad}$，且对某个 $p \geq 2$，开集为 $\mathcal{B} \supset \mathcal{Z}_{ad}$。设 $\{\phi_i\}_{i=1}^\infty$ 和 $\{\psi_i\}_{i=1}^\infty$ 分别为 $\mathcal{U}$ 和 $\mathcal{M}$ 的标准正交基。此外，设 $b_u \in \mathcal{U}$ 和 $b_m \in \mathcal{M}$ 为任意两个向量。那么，对于任意 $\epsilon > 0$，存在秩 r_m 和 r_u，以及一个 *GELU* 神经网络 g_θ：$\mathbb{R}^{r_m} \times \mathbb{R}^{d_z} \to \mathbb{R}^{r_u}$，使得 *RBNO* $u_\theta(m, z) = \mathbf{\Phi}_{r_u} g_\theta(\mathbf{\Psi}^*_{r_m}(m - b_m)\ z) + b_u$ 满足

$$\sup_{z\in\mathcal{Z}_{ad}} \mathbb{E}_{m\sim\mu_m}[\|u(m, z) - u_\theta(m, z)\|^p_{\mathcal{U}}] + \mathbb{E}_{m\sim\mu_m}[\|D_z u(m, z) - D_z u_\theta(m, z)\|^p_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}] \leq \epsilon^p. \tag{31}$$

定理 3.1的证明见第 C节。辅助引理首先在第 C.1节和 C.2节中给出。随后将这些结果结合起来，在第 C.3节完成证明。

请注意，神经算子的逼近误差采用混合范数度量：在 $L^p_{\mu_m}$ 方向上取 m 范数，在紧容许集上沿 z 方向取上确界。这与准确表示 OUU 问题一阶最优性条件所需的逼近误差形式相符（例如，命题 2.2）。特别地，定理 3.1 表明，RBNO 架构能够近似解算子；将其用作期望最小化的代理模型时（假设真实代价泛函为强凸），其驻点可以任意接近真实极小值点。

定理 3.2。 假设 u满足假设 *2.1*，且 Q满足假设 *2.2*，其中博雷尔概率测度为 μ_m，可行集为凸紧集 $\mathcal{Z}_{ad}$，开集为 $\mathcal{B} \supset \mathcal{Z}_{ad}$，且存在某个 $p \geq 2$。现在考虑均值目标函数下的优化问题:

$$\min_{z\in\mathcal{Z}_{ad}} \mathcal{J}(z) := \mathbb{E}_{m\sim\mu_m}[Q(u(m, z), m, z)] + \mathcal{P}(z),$$

其中 Q、 u 和 $\mathcal{P}$ 满足 $\mathcal{J} \in C^1(\mathcal{B}; \mathbb{R})$ 为强凸函数。令 $\{\phi_i\}_{i=1}^\infty$ 和 $\{\psi_i\}_{i=1}^\infty$ 分别为 $\mathcal{U}$ 和 $\mathcal{M}$ 的标准正交基，并令 $b_u \in \mathcal{U}$ 和 $b_m \in \mathcal{M}$ 为任意两个向量。那么，对于任意 $\epsilon > 0$，

存在秩 r_m 和 r_u，以及一个 *GELU* 神经网络 $g_\theta: \mathbb{R}^{r_m} \times \mathbb{R}^{d_z} \to \mathbb{R}^{r_u}$，使得任何满足代理最小化问题一阶最优性条件的点 $z^\dagger$

$$\min_{z \in \mathcal{Z}_{ad}} \mathcal{J}_\theta(z) := \mathbb{E}_{m \sim \mu_m}[Q(u_\theta(m,z),m,z)] + \mathcal{P}(z)$$

*wit*其中，*RBNO* $u_\theta(m,z) = \mathbf{\Phi}_{r_u} g_\theta(\mathbf{\Psi}^*_{r_m}(m - b_m)\ z) + b_u$ 的优化误差有如下界

$$\|z^\star - z^\dagger\|_{\mathcal{Z}} \le \epsilon \tag{32}$$

其最优性间隙的上界为

$$\mathcal{J}(z^\dagger) - \mathcal{J}(z^\star) \le \epsilon^2. \tag{33}$$

定理 3.2结合了定理2.1中的优化误差界、命题2.2中的误差表达式，以及定理3.1中的通用逼近结果。我们在第C.4节给出其证明。尽管强凸性假设较为严格，但值得注意的是，在实践中使用强凸惩罚泛函时通常可以满足该假设 $\mathcal{P}$。或者，可以将此结果重新解释为真实极小值点 $z^\star$ 周围的一个有界邻域，在该邻域内目标函数是强凸的。此时，只要代理模型在该邻域内具有足够的导数精度，就能保证落在该邻域内的任何代理驻点具有有界的优化误差（例如，参见 [55]）。

3.3. 形状导数信息驱动的算子学习

尽管定理3.2 表明，当 $z \in \mathcal{Z}_{ad}$紧致时，$\mathcal{Z}_{ad}$ 上的逼近误差可以得到一致控制，但在实践中，通常无法通过监督学习确定上确界误差界。因此，作为一种具有实际可操作性的算子学习形式，我们采用输出值和算子弗雷歇导数的均方误差（MSE）损失函数，以提高导数精度，遵循 [9, 51]。为此，我们考虑如下联合分布 μ: $=\mu_m \otimes \mu_z$，其中 μ_z 是定义在 $\mathcal{Z}$ 上的一个假设采样分布，其支撑集包含 $\mathcal{Z}_{ad}$。随后，我们通过求解以下问题来构造算子代理模型:

$$\min_\theta \mathbb{E}_{(m,z)\sim\mu}\Big[\|u(m,z) - u_\theta(m,z)\|^2_{\mathcal{U}} + \|D_m u(m,z) - D_m u_\theta(m,z)\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{U})} + \|D_z u(m,z) - D_z u_\theta(m,z)\|^2_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}\Big], \tag{34}$$

我们将这种索博列夫训练形式称为 H^1_μ 形式，因为在有限维情形下，其损失函数类似于H^1 索博列夫范数。这与传统的 L^2_μ 训练形式形成对比，

$$\min_\theta \mathbb{E}_{(m,z)\sim\mu}[\|u(m,z) - u_\theta(m,z)\|^2_{\mathcal{U}}], \tag{35}$$

后者不涉及导数误差。

注意，在（34）中，我们还将相对于不确定参数的导数$D_m u$纳入了训练损失。这不同于 [9]中所介绍的方法。尽管逼近理论（定理2.1 和3.2）并不要求这样做，但它确实提供了额外的训练数据，可用于提高泛化精度。此外，正如我们将在第4.2.1节中讨论的，生成这些额外雅可比数据的计算成本可以忽略不计。

（34）的表述具有通用性，适用于任何神经算子架构。然而，对导数使用完整的 HS($\mathcal{M}$, $\mathcal{U}$) 和 HS($\mathcal{Z}$, $\mathcal{U}$) 范数进行训练，可能会带来难以承受的计算成本，包括生成导数（雅可比矩阵）数据以及在训练过程中评估损失的成本。相反，为了使 H^1_μ学习变得可行，我们利用降阶基架构（30）的特性，将状态和雅可比矩阵直接投影到降阶基张成的空间中。

这便得到训练问题的潜在空间形式：

$$\min_{\theta} \mathbb{E}_{(m,z)\sim\mu}\Big[\|\mathbf{\Phi}^*_{r_u}(u(m,z)-b_u) - g_\theta(\mathbf{\Psi}^*_{r_m}(m-b_m),z)\|_2^2 + \|\mathbf{\Phi}^*_{r_u} D_m u(m,z)\mathbf{\Psi}_{r_m} - D_1 g_\theta(\mathbf{\Psi}^*_{r_m}(m-b_m),z)\|_F^2 + \|\mathbf{\Phi}^*_{r_u} D_z u(m,z) - D_2 g_\theta(\mathbf{\Psi}^*_{r_m}(m-b_m),z)\|_F^2\Big], \tag{36}$$

其中 D_i 表示关于 i^{th} 变量的导数。由于 RBNO 的输出和导数仅位于降阶子空间中，(36) 与 (34) 等价。此外，由于损失函数仅在维数为 r_m、d_z 和 r_u的潜在空间中计算，因此可以高效地进行训练。该方法在 [51] 中提出，并已应用于其他多项研究 [9, 52, 53]。注意，L^2_μ 损失函数 (35) 同样可以投影到潜在空间中，这相当于只取 (36) 中的第一项。

3.4. 实用降维

尽管对于定理 3.1和 3.2的通用逼近结果而言，任意选择标准正交基都足够，但有效的降维策略能够选取较小的输入和输出秩，从而只需较小的潜在空间神经网络。为此，我们采用了 DIPNet 架构中使用的策略 [67]。具体而言，我们对输出使用 POD u，对不确定参数使用主动子空间 m，以兼顾计算上的实用性和降维效果。二者均基于数据生成分布 μ 构造：$= \mu_m \otimes \mu_z$。

在此，我们介绍降维过程，并指出由于（1）需要考虑状态变量中区域的变化，以及（2）仅对两个输入变量之一（即不确定参数）进行降维，因此标准理论结果需要作出的一些小幅修改。

3.4.1. 状态变量的本征正交分解

我们直接在参考域上计算 POD。也就是说，我们使用状态变量的协方差算子，

$$\mathcal{C}_u = \mathbb{E}_{(m,z)\sim\mu}[(u(m,z)-\bar{u})(u(m,z)-\bar{u})^T], \tag{37}$$

其中，$\bar{u} = \mathbb{E}_{(m,z)\sim\mu}[u(m,z)]$ 是训练分布上的均值，而被积函数由外积 $(u(m,z) - \bar{u})(u(m,z) - \bar{u})^T v = (u(m,z) - \bar{u})\langle u(m,z) - \bar{u}, v\rangle_{\mathcal{U}}$构成。随后，POD基由 $\mathcal{C}_u$的前 r_u 个特征向量给出，即 $\{\phi_i\}_{i=1}^{r_u} = \{\phi_i^{\mathrm{POD}}\}_{i=1}^{r_u}$ 满足

$$\mathcal{C}_u \phi_i^{\mathrm{POD}} = \lambda_i^{\mathrm{POD}} \phi_i^{\mathrm{POD}}, \qquad i \in \mathbb{N}, \tag{38}$$

其中 λ_i^{POD}按降序排列。采用均值平移本征正交分解时，我们将 $b_u = \bar{u}$ 作为偏置项。这本质上对应于仿射解码器 $u_r \mapsto \mathbf{\Phi}^*_{r_u} u_r + \bar{u}$。虽然并非必需，但对协方差进行均值平移可确保与上述解码操作对应的仿射子空间 $\bar{u} + \mathrm{span}(\phi_1^{\mathrm{POD}}, \ldots, \phi_{r_u}^{\mathrm{POD}})$满足任意线性、非参数约束（例如，齐次边界条件、无散条件） [60]。

POD的标准结果表明，与POD基相关的截断误差有界于截断特征方向上的尾部特征值之和。下文对此进行总结（证明见 [57, 60]）。

命题 3.1（POD 重构误差界）。设 μ 是 $\mathcal{M}\times\mathcal{Z}$ 上的概率测度，且$u \in L^2_\mu(\mathcal{M}\times\mathcal{Z};\ \mathcal{U})$。令 $\{\phi_i\}_{i=1}^{r_u} = \{\phi_i^{\mathrm{POD}}\}_{i=1}^{r_u}$ 为前 r_u 个*POD*基向量。则投影 $P_{\mathcal{U},\ r_u}: = \sum_{i=1}^{r_u} \phi_i \phi_i^T$的重构误差有界于

$$\mathbb{E}_{(m,z)\sim\mu}[\|(I - P_{\mathcal{U},r_u})(u(m,z)-\bar{u})\|_{\mathcal{U}}^2] \le \sum_{i>r_u} \lambda_i^{\mathrm{POD}}. \tag{39}$$

由于我们在参考域中计算POD，该截断误差界使用参考域函数空间 $\mathcal{U}$。对于 $\mathcal{U}(\Omega_0) = L^2(\Omega_0)$或 $H^1(\Omega_0)$，我们还可以进一步考虑将截断状态变量前推到空间域Ω_z后的误差界。

命题 3.2 (空间域范数下的 POD 重构误差界)。 考虑命题*3.1*中的设定。此外，假设存在常数 γ_0 和 γ_1 使得

$$\|\chi_z\|_{W^{1,\infty}(\Omega_0)} \le \gamma_0 \qquad \forall z \in \mathcal{Z}_{ad}, \tag{40a}$$
$$\|\chi_z^{-1}\|_{W^{1,\infty}(\Omega_z)} \le \gamma_1 \qquad \forall z \in \mathcal{Z}_{ad}. \tag{40b}$$

我们有如下结论：

1. 如果 $\mathcal{U}(\Omega_0) = L^2(\Omega_0)$ 且 $\mathcal{U}(\Omega_z) = L^2(\Omega_z)$，那么

$$\mathbb{E}_{(m,z)\sim\mu}[\|T_{\chi_z}^{-1}((I - P_{\mathcal{U},r_u})(u(m,z) - \bar{u}))\|_{\mathcal{U}(\Omega_z)}^2] \le \gamma_0^d \sum_{i>r_u} \lambda_i^{\mathrm{POD}}. \tag{41}$$

2. 如果 $\mathcal{U}(\Omega_0) = H^1(\Omega_0)$ 且 $\mathcal{U}(\Omega_z) = H^1(\Omega_z)$，那么

$$\mathbb{E}_{(m,z)\sim\mu}[\|T_{\chi_z}^{-1}((I - P_{\mathcal{U},r_u})(u(m,z) - \bar{u}))\|_{\mathcal{U}(\Omega_z)}^2] \le (1+\gamma_1^2)\gamma_0^d \sum_{i>r_u} \lambda_i^{\mathrm{POD}}. \tag{42}$$

证明。 首先考虑 $\mathcal{U}(\Omega_0) = H^1(\Omega_0)$ 和 $\mathcal{U}(\Omega_z) = H^1(\Omega_z)$ 的情形。于是，根据 *Proposition* 2.1，我们有

$$\begin{aligned}\|T_{\chi_z}^{-1}((I - P_{\mathcal{U},r_u})(u(m,z) - \bar{u}))\|_{\mathcal{U}(\Omega_z)}^2 &\le \left(1 + \|\chi_z^{-1}\|_{W^{1,\infty}(\Omega_z)}^2\right) \|\chi_z\|_{W^{1,\infty}(\Omega_0)}^d \|(I - P_{\mathcal{U},r_u})(u(m,z) - \bar{u})\|_{\mathcal{U}(\Omega_0)}^2 \\ &\le (1+\gamma_1^2)\gamma_0^d \|(I - P_{\mathcal{U},r_u})(u(m,z) - \bar{u})\|_{\mathcal{U}(\Omega_0)}^2\end{aligned}$$

对任意 $z \in \mathcal{Z}_{ad}$ 均成立。因此，关于 $(m, z) \sim \mu$ 取期望，并应用标准 POD 界，即可得到所需结果。对于 $\mathcal{U}(\Omega_0) = L^2(\Omega_0)$，证明过程类似。 □

3.4.2. 不确定参数的活跃子空间

然后，我们基于解算子在降阶 POD 坐标中的截断表示，构造 m 的降维。也就是说，我们计算映射$P_{\mathcal{U},r_u}u(m,z) = \mathbf{\Phi}_{r_u}\mathbf{\Phi}_{r_u}^* u(m,z)$的主动子空间，其导数协方差由下式给出

$$\mathcal{H}_m = \mathbb{E}_{(m,z)\sim\mu}[D_m u(m,z)^* \mathbf{\Phi}_{r_u}\mathbf{\Phi}_{r_u}^* D_m u(m,z)]. \tag{43}$$

于是，子空间基由 $\mathcal{H}_m$的前 r_m 个特征向量给出，即

$$\mathcal{H}_m \psi_i^{\mathrm{AS}} = \lambda_i^{\mathrm{AS}} \psi_i^{\mathrm{AS}}, \qquad i \in \mathbb{N}, \tag{44}$$

其中 λ_i^{AS}再次按降序给出。然后取$\{\psi_i\}_{i=1}^{r_m} = \{\psi_i^{\mathrm{AS}}\}_{i=1}^{r_m}$ 与 $b_m = \bar{m}$。主动子空间捕捉参数空间中对输出影响最大的主导方向，其影响由导数 [66]的大小来衡量。由于该降维明确考虑了敏感性信息，它在输入维数约简方面通常比 POD 更有效；POD 仅考虑输入方差，而与映射本身无关。这里我们指出，基于（43）的主动子空间与标准活跃子空间的不同之处在于，仅考虑 m的导数。

当底层分布 μ_m 满足子空间庞加莱不等式时，可以分析主动子空间的理论性质。也就是说，设 $\mathcal{G}(P_{\mathcal{M},r_m})$为由秩为 r的正交投影$P_{\mathcal{M},r_m}$ 生成的子 σ-代数。那么，对任意 $f \in L^2_{\mu_m}$，其条件期望 $\mathbb{E}_{\mu_m}[f|\mathcal{G}(P_{\mathcal{M},r_m})]$ 是 f 在$\mathcal{G}_{P_{\mathcal{M},r_m}}$-可测函数子空间中的最佳 L^2_μ近似。这类函数可由如下事实刻画：对某个可测函数 g，有 $f(m) = g(P_{\mathcal{M},r_m}m)$，即它们是不依赖于$P_{\mathcal{M},r_m}$的零空间的脊函数。我们可以通过考察 f 与 $\mathbb{E}_{\mu_m}[f|\mathcal{G}(P_{\mathcal{M},r_m})]$之间的这一 $L^2_{\mu_m}$ 误差，来量化输入维数约简策略的有效性。为此，令 $\mathcal{W} := \mathbb{R}^{r_u}$ 表示输出的潜在空间。于是我们有如下定义。

定义 3.1（子空间庞加莱不等式）。 我们称博雷尔概率测度 μ_m 满足子空间庞加莱不等式，如果对任意满足$\mathbb{E}_{\mu_m}[\|D_m f\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{W})}] < \infty$的 $f \in C^1(\mathcal{M};\mathcal{W})$，

$$\mathbb{E}_{\mu_m}[\|f - \mathbb{E}_{\mu}[f|\mathcal{G}(P_{\mathcal{M},r_m})]\|^2_{\mathcal{W}}] \le c_{\mu_m}\mathbb{E}_{\mu_m}[\|D_m f(I - P_{\mathcal{M},r_m})\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{W})}], \tag{45}$$

其中， $c_{\mu_m} > 0$ 是一个庞加莱常数，且与 f 和$P_{\mathcal{M},\ r_m}$无关。

满足子空间庞加莱不等式的分布的典型例子是高斯分布；我们将在下文讨论这一情形。虽然还有其他例子和刻画，但这些通常针对有限维输入空间给出。

现在考虑任意可微的 f： $\mathcal{M}\times\mathcal{Z}\to\mathcal{W}$。由于我们只关注$m$的降维，因此我们逐点考察 z中相对于 m 的 u的条件期望，

$$\mathbb{E}_{P_{\mathcal{M},r_m}}[f](m,z) := \mathbb{E}_{\mu_m}[f(\cdot,z)|\mathcal{G}(P_{\mathcal{M},r_m})](m). \tag{46}$$

并研究$\mathbb{E}_{P_{\mathcal{M},\ r}}[f]$(m、 z)的逼近误差。于是，只要 μ_m 满足子空间庞加莱不等式，乘积测度 $\mu = \mu_m \otimes \mu_z$ 就满足部分子空间庞加莱不等式。特别是，由于 z 未被截断，该结果不依赖于 μ_z 满足子空间庞加莱不等式，且所得的庞加莱常数仅取决于 μ_m的子空间庞加莱常数。

命题 3.3（部分降维的子空间庞加莱不等式）。 假设 μ_m 满足子空间庞加莱不等式， μ_z 是定义在 $\mathcal{Z}$上的博雷尔概率测度，且 $f \in C^1(\mathcal{M}\times\mathcal{Z};\mathcal{W})$ 满足关于 $\mu = \mu_m \otimes \mu_z$的矩界，

$$\mathbb{E}_{\mu}[\|f\|^2_{\mathcal{W}}] + \mathbb{E}_{\mu}[\|D_m f\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{W})}] < \infty. \tag{47}$$

那么，

$$\mathbb{E}_{\mu}[\|f - \mathbb{E}_{P_{\mathcal{M},r_m}}[f]\|^2_{\mathcal{W}}] \le c_{\mu}\mathbb{E}_{\mu}[\|D_m f(I - P_{\mathcal{M},r_m})\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{W})}]. \tag{48}$$

证明。 这可直接由子空间庞加莱不等式得出。回顾一下，对于任意 z，我们有

$$\mathbb{E}_{m\sim\mu_m}[\|f(m,z) - \mathbb{E}_{P_{\mathcal{M},r_m}}[f](m,z)\|^2_{\mathcal{W}}] \le c_{\mu}\mathbb{E}_{m\sim\mu_m}[\|D_m f(m,z)(I - P_{\mathcal{M},r_m})\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{W})}] \tag{49}$$

现在对 $z\sim\mu_z$ 取期望，即得所需结果。 □

这使我们能够量化式（43）–（44）中给出的部分主动子空间的有效性。

命题 3.4。 假设 μ_m 满足子空间庞加莱不等式，且 μ_z 是定义在 $\mathcal{Z}$上的博雷尔概率测度，并且 $u \in C^1(\mathcal{M}\times\mathcal{Z};\mathcal{U})$ 满足关于 $\mu = \mu_m \otimes \mu_z$ 的矩界

$$\mathbb{E}_{\mu}[\|u\|^2_{\mathcal{U}}] + \mathbb{E}_{\mu}[\|D_m u\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{U})}] < \infty. \tag{50}$$

设 $\{\psi_i\}_{i=1}^{r_m} = \{\psi_i^{\mathrm{AS}}\}_{i=1}^{r_m}$ 为活跃子空间基，$P_{\mathcal{M},\ r_m} := \sum_{i=1}^{r_m}\psi_i\psi_i^T$。那么，给定投影$P_{\mathcal{M},\ r_m}$ 的条件期望满足

$$\mathbb{E}_{(m,z)\sim\mu}[\|\mathbf{\Phi}^*_{r_u}(u(m,z) - \mathbb{E}_{P_{\mathcal{M},r_m}}[u](m,z))\|^2_{\mathcal{W}}] \le c_{\mu}\sum_{i>r_m}\lambda_i^{\mathrm{AS}}. \tag{51}$$

证明。 我们使用上文中的子空间庞加莱不等式，将其应用于映射 f(m， z) =$\mathbf{\Phi}^*_{r_u}u$(m， z)。也就是说，对于任意 $P_{\mathcal{M},\ r_m}$，有

$$\mathbb{E}_{(m,z)\sim\mu}[\|f(m,z) - \mathbb{E}_{P_{\mathcal{M},r_m}}[f](m,z)\|^2_{\mathcal{W}}] \le c_{\mu_m}\mathbb{E}_{(m,z)\sim\mu}[\|D_m f(m,z)(I - P_{\mathcal{M},r_m})\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{W})}].$$

特别地，上界也可以写成

$$\begin{aligned}\mathbb{E}_{\mu}[\|D_m f(I - P_{\mathcal{M},r_m})\|^2_{\mathrm{HS}(\mathcal{M},\mathcal{W})}] &= \mathrm{Tr}((I - P_{\mathcal{M},r_m})\mathbb{E}_{\mu}[D_m f^* D_m f](I - P_{\mathcal{M},r_m})) \\ &= \mathrm{Tr}((I - P_{\mathcal{M},r_m})\mathcal{H}_m(I - P_{\mathcal{M},r_m})).\end{aligned}$$

对于 $P_{\mathcal{M},\,r_m} = \sum_{i=1}^{r_m} \psi_i \psi_i^T$，迹项可简化为

$$\mathrm{Tr}((I - P_{\mathcal{M},r_m})\mathcal{H}_m(I - P_{\mathcal{M},r_m})) = \sum_{i>r_m} \lambda_i^{\mathrm{AS}},$$

从而得到所需结果。 □

该结果表明，存在一个岭函数，即 $\mathbb{E}_{P_{\mathcal{M}}}$，其平均误差由 $\mathcal{H}_m$ 的尾部特征值之和界定。虽然这一结果与上一节所建立的逼近理论并无直接联系（该理论使用的是岭函数 $u_r(m,z) = P_{\mathcal{U},r_u} u(P_{\mathcal{M},r_m} m, z)$，而不是条件期望），但它确实直观地说明了由 $P_{\mathcal{M},r_m}$ 表示的子空间捕捉了输入空间中的重要方向。我们注意到，在对 u 和 μ_m 作出进一步正则性假设的情况下，例如 u 和 Du 具有利普希茨连续性且存在高斯μ_m，我们可以证明 $\bar{u}_r$ 也是连续可微的，并且其导数具有利普希茨连续性，进而基于对 $\bar{u}_r$ 的神经网络近似，建立 RBNO 的逼近误差分析，这与第2节所采用的方法类似。 [9] 采用了类似的方法。

有限维和无限维高斯输入。 在数值算例中，我们主要考虑不确定参数 $\mathcal{M}$ 的高斯随机场离散化，即 $\mathcal{M} = \mathbb{R}^{d_m}$ 是有限维的，而 $\mu_m = \mathcal{N}(0, \mathcal{C}_m)$ 是具有非退化协方差 $\mathcal{C}_m$ 的高斯变量。在这种情况下，通常赋予 $\mathcal{M}$ 精度加权内积。为避免混淆，我们将 $\mathcal{E}$ 写作表示赋予该内积的 $\mathcal{M}$，使得 $\langle m_1, m_2\rangle_{\mathcal{E}} = \langle \mathcal{C}_m^{-1/2} m_1, \mathcal{C}_m^{-1/2} m_2\rangle_{\mathcal{M}}$。对于高斯$\mu_m$，关于 $\mathcal{G}(P_{\mathcal{M},r_m})$ 的条件期望具有显式表达式，

$$\mathbb{E}_{\mu_m}[f|\mathcal{G}(P_{\mathcal{M},r_m})] = \int f(P_{\mathcal{M},r_m} m + (I - P_{\mathcal{M},r_m})\eta) d\mu_m(\eta). \tag{52}$$

此外，μ_m 满足子空间庞加莱不等式

$$\mathbb{E}_{\mu_m}[\|f - \mathbb{E}_{\mu_m}[f|\mathcal{G}(P_{\mathcal{M},r_m})]\|_{\mathcal{U}}^2] \le \mathbb{E}_{\mu_m}[\|D_m f(I - P_{\mathcal{M},r_m})\|_{\mathrm{HS}(\mathcal{E},\mathcal{U})}^2], \tag{53}$$

其中 $c_{\mu_m} = 1$。在这种情况下，式 (43) 中的伴随算子为 $D_m u(m,z)^* = (D_m u(m,z))_{\mathcal{E}}^* = \mathcal{C}_m (D_m u(m,z))_{\mathcal{M}}^*$，其中 $(\cdot)_{\mathcal{E}}^*$ 表示相对于卡梅伦–马丁空间内积的伴随算子，而 $(\cdot)_{\mathcal{M}}^*$ 表示相对于原始内积 $\langle\cdot,\cdot\rangle_{\mathcal{M}}$ 的伴随算子。相应的特征值问题 (44) 可写为广义特征值问题，

$$\widetilde{\mathcal{H}}_m \psi_i^{\mathrm{AS}} = \lambda_i^{\mathrm{AS}} \mathcal{C}_m^{-1} \psi_i^{\mathrm{AS}}, \qquad i \in \mathbb{N}, \tag{54}$$

其中

$$\widetilde{\mathcal{H}}_m = \mathbb{E}_{(m,z)\sim\mu}[(D_m u(m,z))_{\mathcal{M}}^* \mathbf{\Phi}_{r_u} \mathbf{\Phi}_{r_u}^* D_m u(m,z))], \tag{55}$$

重要的是，子空间庞加莱常数 $c_{\mu_m} = 1$以及因此命题 3.4 中的误差界，并不显式依赖于离散化维度 d_m。

尽管 $\mathcal{M}$ 和 $\mathcal{E}$ 在有限维情形下等价，但在无限维情形下则不再如此。具体而言，$\mathcal{C}_m^{-1/2}$可以是作用于 $\mathcal{M}$ 上的无界算子，因此 $\mathcal{E}$ 范数仅在卡梅伦–马丁空间 $\mathrm{Range}(\mathcal{C}_m^{1/2})$ 上有定义，而该空间未必与 $\mathcal{M}$重合。尽管如此，由式（44）构造的降阶基仍可为无限维高斯分布提供一种定义良好的输入维数约简形式（见 [52, 53, 60]）。在（54）和（55）存在有意义的极限时，这可视为 $d_m \to \infty$ 的极限情形。

3.4.3. 用于约化基计算的基于样本的近似

在实践中，POD 和主动子空间都需要根据有限样本对期望进行近似计算。对于 POD，我们采用均值估计量

$$\widehat{u} := \frac{1}{n_{\mathrm{POD}}} \sum_{i=1}^{n_{\mathrm{POD}}} u(m^{(i)}, z^{(i)}) \tag{56}$$

和无偏协方差估计量

$$\widehat{\mathcal{C}}_u := \frac{1}{n_{\mathrm{POD}}-1}\sum_{i=1}^{n_{\mathrm{POD}}}(u(m^{(i)},z^{(i)})-\widehat{u})(u(m^{(i)},z^{(i)})-\widehat{u})^* \tag{57}$$

其中，$(m^{(i)},\ z^{(i)})\sim\mu, i=1,\ \ldots\ldots,\ n_{\mathrm{POD}}$ 是从输入分布中抽取的独立同分布样本。于是，经验$\widehat{}$POD由主特征向量 $\mathcal{C}_u$给出，

$$\widehat{\mathcal{C}}_u\widehat{\phi}_j^{\mathrm{POD}}=\widehat{\lambda}_j^{\mathrm{POD}}\widehat{\phi}_j^{\mathrm{POD}},\qquad j\in\mathbb{N} \tag{58}$$

其中，我们取 $\{\phi_j\}_{i=1}^{r_u}=\{\widehat{\phi}_j^{\mathrm{POD}}\}_{j=1}^{r_u}$，并将 $b_u=\widehat{}u$ 用作RBNO中的偏差。

对于主动子空间，我们取导数协方差估计量

$$\widehat{\mathcal{H}}_m := \frac{1}{n_{\mathrm{AS}}}\sum_{i=1}^{n_{\mathrm{AS}}}D_m u(m^{(i)},z^{(i)})^*\mathbf{\Phi}_{r_u}\mathbf{\Phi}_{r_u}^* D_m u(m^{(i)},z^{(i)}), \tag{59}$$

其中，$(m^{(i)},\ z^{(i)})\sim\mu, i=1,\ \ldots\ldots,\ n_{\mathrm{AS}}$ 同样是从输入分布中抽取的独立同分布样本。经验主动子空间$\widehat{}$由 $\mathcal{H}_m$的主特征向量给出，

$$\widehat{\mathcal{H}}_m\widehat{\psi}_j^{\mathrm{AS}}=\widehat{\lambda}_j^{\mathrm{AS}}\widehat{\psi}_j^{\mathrm{AS}},\qquad j\in\mathbb{N} \tag{60}$$

其中，我们取 $\{\psi_j\}_{j=1}^{r_m}=\{\widehat{\psi}_j^{\mathrm{AS}}\}_{j=1}^{r_m}$。由于我们假设已知 μ_m 的均值 $\bar{m}$，因此只需将$b_m=\bar{m}$ 作为偏差。

4. 实现细节

4.1. 通过基展开和弹性延拓实现微分同胚变形

本节讨论 χ_z 的选取及其关联分布 $z\sim\mu_z$，使得对于 z的所有取值，χ_z 都保持为微分同胚。具体而言，我们将考虑对 $z\mapsto\chi_z$ 的表示，该表示本身采用基展开，

$$\chi_z(X)=X+\sum_{i=1}^{d_z}z_i\mathfrak{d}_i(X), \tag{61}$$

其中 $\{\mathfrak{d}_i\}_{i=1}^{d_z}$ 是一组光滑位移基函数 $\mathfrak{d}_i:\Omega_0\to\mathbb{R}^d$。我们有 tha t if

$$\left\|\sum_{i=1}^{d_Z}z_i\nabla_X\mathfrak{d}_i-I\right\|_{L^\infty(\Omega_0)}<1, \tag{62}$$

则 χ_z 由（61）给出，是一个微分同胚。然而，正如 中所指出的，条件（62）通常过于保守，实际中可以允许更大的可行集 $\mathcal{Z}_{ad}$。在数值算例中，我们通常将 $\mathcal{Z}_{ad}$ 约束在盒约束 $z_l\le z_i\le z_u$ 内，以确保所得变形χ_z 为微分同胚，随后将 μ_z 选为 $\mathcal{Z}_{ad}$上的均匀分布。

公式 (61) 适用于形状优化中常用的多种参数化形状表示。其中包括通过傅里叶基展开给出的区域变形，这种方法可将边界设计为任意函数 [69, 70] 的图像。还包括通过自由形变实现的形状表示 [71], ，其中 $\mathfrak{d}_i$ 通常选自表达能力强的局部支撑基函数类，例如伯恩斯坦多项式或 B样条。在数值实验中，我们在第 5.3 节研究了泊松偏微分方程中使用傅里叶表示的形状优化，并在第 5.4 节和 5.5节研究了稳态纳维–斯托克斯流中采用自由形变方法的情形。两种方法的形状参数化细节分别见第 5.3.2 节和 5.4.2 节。

在许多其他设计问题中，形状边界被规定为基展开的图像，而未显式延拓至区域内部。对于非平凡区域使用特定于区域的基时，情况便是如此，例如用于翼型设计的 Hicks–Henne 基 [72–74]，以及用于船体设计的 Lewis 形式 [75] 。在这类情形下，位移基 $\mathfrak{d}_D^{(i)}$ 仅在设计边界 $\Gamma_D \subset \partial\Omega_0$上给定。尽管如此，我们仍可通过将位移基延拓至整个区域 Ω_0 来利用公式 (61) 表述问题；弹性延拓是形状优化 [76, 77]和流固耦合 [78–80]中常用的网格更新方法。为此，我们将 $\mathfrak{d}_i$ 视为带有狄利克雷边界条件 $\mathfrak{d}_i|_{\Gamma_D} = \mathfrak{d}_D^{(i)}$ 和 $\mathfrak{d}_i|_{\partial\Omega_0\setminus\Gamma_D} = 0$的线弹性偏微分方程的解。

$$-\nabla\cdot\big(a_\mu(\nabla u + \nabla u^T) + a_\lambda(\nabla\cdot u)I\big) = 0 \qquad \text{in } \Omega_0 \tag{63a}$$

$$\mathfrak{d}_i = \mathfrak{d}_D^{(i)} \qquad \text{on } \Gamma_D \tag{63b}$$

$$\mathfrak{d}_i = 0 \qquad \text{on } \partial\Omega_0 \setminus \Gamma_D, \tag{63c}$$

其中 $(\nabla u^T)_{ij} = (\nabla u)_{ji}$ 对应于位移梯度的矩阵转置。拉梅参数 a_λ 和 a_μ 可以采用空间变化的方式定义（例如，根据与边界的距离），以改善变形梯度 $F_{\chi_z}(X) = \nabla_X \chi_z(X)$的条件数。随后，通过求解（63），例如采用有限元法，数值计算位移基函数。

事实上，即使采用自由形变，其中 $\mathfrak{d}_i$ 是在整个 Ω_0, 上指定的，我们仍将使用$\mathfrak{d}_i$ 生成一组新的位移基函数 $\tilde{\mathfrak{d}}_i$，并将 $\mathfrak{d}_i|_{\partial\Omega_0}$ 作为边界条件。这在用于（61）时不会改变形状边界的运动，但可以改善变形梯度的条件数，从而提高所得位移基函数中变形网格的质量。

注 4.1（弹性延拓的正则性）。理论上，变形映射 χ_z 要成为微分同胚，必须属于 C^1类。然而，当通过有限元法求解线弹性方程来生成此类映射时，所得弱解属于索伯列夫空间，例如$W^{1,\infty}$ 而非 C^1。这种光滑度通常足以定义前推/拉回运算，并确保形状优化中的可微性（例如，*[61, 第 11]*节）。此外，只需确保变形梯度在单元级具有正行列式，即可验证数值变形的可逆性。尽管如此，我们指出，自由形变方法并非必须采用弹性延拓；若关注变形映射的光滑性，则可以省略弹性延拓。

4.2. 算子学习的计算工作流程

现在，我们总结构建用于偏微分方程约束的 OUU 的 shape-DINO 代理模型的整体工作流程，该流程大体遵循 [9] 中所述的方法。随后，我们将详细介绍训练数据生成和神经算子训练的具体细节。

1. 状态数据生成：从指定的训练分布中采样参数–控制对 $\{(m^{(i)}, z^{(i)})\}_{i=1}^{n_s}$。对于每个样本，求解高保真偏微分方程，以获得相应的状态解 $\{u^{(i)}\}_{i=1}^{n_s}$。2. 通过 POD 进行状态空间降阶：在一部分 n_{POD} 训练样本上计算样本均值 $\widehat{u}$，并按照 (56)–(58) 对中心化快照 $\{u^{(i)} - \widehat{u}\}_{i=1}^{n_{\mathrm{POD}}}$ 执行 POD。然后，选择秩为 r_u 的降阶基 $\{\phi_j\}_{j=1}^{r_u}$ 来表示状态空间。3. 灵敏度数据生成：对于每个训练样本，使用伴随方法计算降阶输出参数灵敏度$\{\mathbf{\Phi}_{r_u}^* D_m u^{(i)}\}_{i=1}^{n_s}$ 和形状灵敏度 $\{\mathbf{\Phi}_{r_u}^* D_z u^{(i)}\}_{i=1}^{n_s}$。4. 通过主动子空间进行参数空间降阶：根据 (59)–(60)，利用 n_{AS} 降阶输出参数灵敏度 $\{\mathbf{\Phi}_{r_u}^* D_m u^{(i)}\}_{i=1}^{n_{\mathrm{AS}}}$，为随机参数计算主动子空间，并选择秩为 r_m的降阶基 $\{\psi_i\}_{i=1}^{r_m}$。

5. **潜空间编码：** 将参数、状态和雅可比矩阵投影到各自对应的降阶基上，

$$m_r^{(i)} = \mathbf{\Psi}_{r_m}^* \, m^{(i)}, \qquad u_r^{(i)} = \mathbf{\Phi}_{r_u}^* \, (u^{(i)} - \widehat{u}), \qquad J_{m,r}^{(i)} = \mathbf{\Phi}_{r_u}^* D_m u^{(i)} \mathbf{\Psi}_{r_m}, \qquad J_{z,r}^{(i)} = \mathbf{\Phi}_{r_u}^* D_z u^{(i)}.$$

6. **神经算子训练。** 通过经验风险最小化训练神经网络 g_θ，使用缩减后的训练数据集

$$\{(m_r^{(i)}, z^{(i)}, u_r^{(i)}, J_{m,r}^{(i)}, J_{z,r}^{(i)})\}_{i=1}^{n_s},$$

从而学习一个代理模型，将参数和形状映射到潜在空间中的缩减状态和敏感度。

4.2.1. 细节与计算成本

状态数据生成。 通过求解底层偏微分方程 $\mathcal{R}(u^{(i)}, \ m^{(i)}, \ z^{(i)}) = 0$，针对给定的采样输入 $(m^{(i)}, \ z^{(i)})$，生成状态训练数据。这通常是训练数据生成中的主要成本。当 $\mathcal{R}$ 对应线性偏微分方程（即 $\mathcal{R}$ 关于 u是线性的）时，相应的计算成本是求解一个随离散化维度 d_u扩展的线性方程组。另一方面，当 $\mathcal{R}$对应非线性偏微分方程时，求解 $u^{(i)}$ 可能需要依次求解多个线性偏微分方程，例如采用牛顿法或皮卡德迭代。

*POD*计算。 状态数据 $\{u^{(i)}\}_{i=1}^{n_s}$ 生成后，可以使用 n_{POD} 个 n_s 训练样本的协方差样本估计，通过求解特征值问题来计算 POD，如式（56）–（58）所示。特别是，特征值问题（58）可以通过 Lanczos 等迭代方法或随机化方法，以无矩阵方式高效求解 [81] ，而且 [82], 无需额外的偏微分方程求解。

导数数据计算。回顾一下，向量-雅可比积可以通过伴随敏感性方程（7）计算。因此，对于给定的输入样本 $(m^{(i)}, z^{(i)})$，一旦状态 $u^{(i)}$ 计算完成，沿每个$\{\phi_j\}_{j=1}^{n_{\text{POD}}}$POD基方向生成向量-雅可比积的成本，涉及求解 r_u 个线性方程组。值得注意的是，这些 r_u 线性方程组共享同一个算子（左端项），仅右端项不同。这使得解的成本可以摊销：使用直接方法时（例如 LU/Cholesky因子），可以重复利用矩阵分解；使用迭代方法时，则可以重复利用计算成本高昂的预条件器。在我们的数值算例中，我们考虑了适合采用直接方法求解的问题。在这种情况下，生成 r_u 向量-雅可比积（包括 m 和 z 雅可比矩阵）的成本渐近等同于只求解第一个线性方程组的成本，因为将矩阵因子作用于其余 $r_u - 1$右端项的成本，与执行矩阵分解的成本相比可以忽略不计。

主动子空间计算。 在我们的工作流程中，可以直接根据采样得到的向量-雅可比积计算主动子空间$\mathbf{\Phi}_{r_u}^* D_m u$ $(m^{(i)}, \ z^{(i)})$。同样，(44) 中的特征值问题可以通过随机化方法求解$\widehat{\ }$，无需显式构造 $\mathcal{H}_m$，并且在生成灵敏度数据后，无需额外进行偏微分方程求解。这里需要注意，使用传统的 L_μ^2 损失函数 (35) 时，只需为前 n_{AS} 个输入样本生成雅可比数据，即可计算主动子空间，因为训练中的其他环节不会用到这些数据。不过，在这种情况下，实际应用中或许更适合使用输入PCA对$\mathcal{M}$ 进行降维，从而完全避免计算导数。文献 [9], 研究了这种方法，但为了确保 Shape-DINO 与 Shape-NO 之间比较的一致性，我们没有采用这种方法。

4.2.2. 神经网络训练

对于神经网络训练，在给定降阶基的情况下，我们可以使用有限训练数据集，通过蒙特卡洛近似（36）

$$\left\{(m_r^{(i)}, z^{(i)}, u_r^{(i)}, J_{m,r}^{(i)}, J_{z,r}^{(i)}\right\}_{i=1}^{n_s}.$$

具体而言，我们将采用（36 ）

$$\min_{\theta} \sum_{i=1}^{n_s} \frac{\|u_r^{(i)} - g_\theta(m_r^{(i)}, z^{(i)})\|_2^2}{\|u_r^{(i)}\|_2^2} + \alpha_D \sum_{i=1}^{n_s} \left(+ \frac{\|J_{m,r}^{(i)} - D_1 g_\theta(m_r^{(i)}, z^{(i)})\|_2^2}{\|J_{m,r}^{(i)}\|_F^2} + \frac{\|J_{z,r}^{(i)} - D_2 g_\theta(m_r^{(i)}, z^{(i)})\|_2^2}{\|J_{z,r}^{(i)}\|_F^2} \right). \tag{64}$$

在此，我们令 $\alpha_D = 1$ 用于导数信息引导的训练（Shape-DINO），令 $\alpha_D = 0$ 用于常规训练（Shape-NO）。

5. 数值结果

本节针对三个不同的PDE约束形状OUU问题，评估 Shape-DINO 的准确性和样本效率：（i）具有形状参数化上边界和不确定对数渗透率场的泊松方程通量跟踪问题；（ii）不确定来流条件下物体周围二维稳态纳维–斯托克斯流的耗散降低问题；（iii）不确定来流条件下塔周围三维稳态纳维–斯托克斯流的弯矩和力降低问题。在每个示例中，我们旨在通过以下两方面评估 Shape-DINO（和 Shape-NO）的有效性：（1）其对输出和导数预测的泛化准确性；（2）在计算预算相近的情况下，与基于PDE的解相比，其求解OUU问题的准确性。请注意，对于二维问题，我们能够将近似的OUU解与通过大量样本计算得到的高精度基于PDE的参考解进行比较。另一方面，对于三维问题，计算高精度参考解的成本高得难以承受，因此本实验旨在展示我们基于代理模型的方法的可扩展性。

数值结果中，我们使用 FEniCS [83]为偏微分方程构建有限元离散，并使用 PyTorch [84] 构建神经网络近似。此外，数据生成通过 hIPPYlib [85] 和 hIPPYflow [86], 实现，而 OUU问题则使用 SOUPy [87]进行构建和求解。基于中央处理器的计时在 TACC VISTA 系统的单个计算节点上完成，该节点配备 NVIDIA Grace CPU Superchip；神经网络的训练和部署则使用 NVIDIA H200 GPU 完成。

5.1. 潜在空间神经网络架构

在所有实验中，我们均采用全连接前馈架构来表示潜空间神经网络 g_θ。我们使用 Adam优化器 [88] ，在常规L_μ^2 损失（Shape-NO）或导数信息 H_μ^1 损失（Shape-DINO）上训练神经网络。两种情况下，我们都使用学习率调度器，在指定的训练轮次里程碑处将学习率减半。架构和训练超参数，包括 d_L， $d_{W,\,i}$、初始学习率、调度器阶段以及训练轮次数量，均取决于具体问题。这些内容总结于表2 中，并在各数值算例中进一步讨论。为保持一致，Shape-DINO 和 Shape-NO 使用相同的超参数。

示例	$d_{W,i}$	初始学习率	学习率调度器，$\gamma = 0.5$	训练轮次	r_u	r_m	d_z
泊松方程	$[512, 512, 512, 512]$	0.0005	$\{800, 1200, 1800\}$	2000	400	256	11
二维纳维–斯托克斯方程	$[512, 1024, 512]$	0.0001	$\{500, 750\}$	1000	256	50	162
三维纳维–斯托克斯方程	$[512, 512]$	0.0001	$\{1000, 1500\}$	2000	200	128	160

表 2：所有实验的神经网络架构和训练超参数。对于每项实验，Shape-DINO 和 Shape-NO 使用相同的设置，损失函数除外。初始 LR 表示 Adam优化器使用的初始学习率。LR Scheduler 指定学习率阶段，其中整数集合表示学习率按因子 0.5 更新时所对应的训练轮次里程碑。

5.2. 神经算子的成本精度比较

我们针对不同的样本量训练 Shape-DINO 和 Shape-NO。随后，利用训练好的神经算子，通过采用大样本量的目标函数样本平均近似来求解 OUU 问题。我们

将神经算子得到的最优解记为 $z^{\dagger}_{\text{神经网络}}$。作为比较，我们还直接利用控制偏微分方程和不同样本量的样本平均近似求解 OUU 问题，所选样本量旨在反映与训练神经算子所需计算预算相当的计算预算。相应的基于偏微分方程的最优解记为 $z^{\dagger}_{\text{偏微分方程}}$。所有优化问题均采用基于梯度的方法求解。对于带盒约束的线性椭圆泊松问题示例，我们采用 L-BFGS-B 算法。对于二维和三维纳维–斯托克斯形状不确定性下的优化问题，我们采用带盒约束的信赖域约束方法。

通过使用大样本量求解基于偏微分方程的不确定性下优化问题，得到参考解，并以此评估近似的精度。所得参考最优解记为 $z^{\star}_{\text{ref}}$。采用定义如下的相对最优成本误差，将近似最优解与 $z^{\star}_{\text{ref}}$ 进行比较：

$$\frac{|\mathcal{J}(z^{\dagger}_{\text{approx}}) - \mathcal{J}(z^{\star}_{\text{ref}})|}{|\mathcal{J}(z^{\star}_{\text{ref}})|}, \tag{65}$$

其中，$\mathcal{J}(z^{\dagger}_{\text{approx}})$ 和 $\mathcal{J}(z^{\star}_{\text{ref}})$ 使用真实目标函数进行评估，该函数通过大样本量的高精度蒙特卡罗估计量估算。此处，$z^{\dagger}_{\text{approx}}$ 表示神经算子方法所得解 $z^{\dagger}_{\text{神经网络}}$ 或基于偏微分方程的方法所得解 $z^{\dagger}_{\text{偏微分方程}}$，具体取决于所考虑的方法。

计算效率以状态偏微分方程求解的总次数衡量，即为获得最优解所需的 $u(m,\ z)$ 评估次数。该数量代表优化过程中占主导地位的计算成本。对于基于神经算子的优化，该成本对应于使用的训练样本总数。对于基于偏微分方程的优化，则等于目标函数评估总次数与样本平均近似样本量的乘积。

在第5.6节中，我们比较了使用基于偏微分方程的求解器和神经算子求解正向状态及其关于形状参数的灵敏度所需的计算成本。由于这两项操作占总体优化成本的主要部分，我们仅关注这两项。在数值实验中，我们表明，与高保真偏微分方程求解相比，神经算子可带来数个数量级的加速，从而使大规模不确定性下优化问题在计算上切实可行。

5.3. 不确定渗透率场下泊松问题的通量跟踪设计

5.3.1. 顶部边界变化的泊松问题

在第一个数值示例中，我们考虑定义在区域 Ω_z 上的泊松偏微分方程，其中 Ω_z 为矩形，且其上边界可由我们自由设计。我们假设 Ω_z 与参考域 $\Omega_0 = (0,4) \times (0,1) \subset \mathbb{R}^2$ 之间通过一族微分同胚 χ_z相关联，其形式将在下文描述。此示例与 [69] 中所考虑的设置类似。这里，偏微分方程以空间坐标表示为

$$\begin{aligned} -\nabla_x \cdot (\exp(\widetilde{m})\nabla_x \widetilde{u}) &= f \qquad && x \in \Omega_z \\ \nabla_x \widetilde{u} \cdot n &= 0 && x \in \Gamma_{L,z} \cup \Gamma_{R,z} \\ \widetilde{u} &= 0 && x \in \Gamma_{B,z} \\ \widetilde{u} &= 1 && x \in \Gamma_{T,z} \end{aligned} \tag{66}$$

其中，$x = (x_1,\ x_2)^T \in \mathbb{R}^2$，且 $\Gamma_{L,\ z}$，$\Gamma_{R,\ z}$，$\Gamma_{B,\ z}$，$\Gamma_{T,\ z}$ 分别是 $\partial\Omega_z$ 的左、右、下和上边界。这里，$f(x) = 0.1\sin(x_1)\cos(x_2)$ 是一个以空间坐标指定的源项。该模型由对数渗透率场 $\widetilde{m}$ 参数化，其不确定性由参考域上的高斯随机场 $m \sim \mu_m = \mathcal{N}(0,\ \mathcal{C}_m)$ 描述，满足 $\widetilde{m} = T^{-1}_{\chi_z} m$。高斯场的协方差由参考域上的双拉普拉斯算子给出

$$\mathcal{C}_m = (\delta - \gamma\nabla_X \cdot (\Theta\nabla_X))^{-2}, \tag{67}$$

这对应于 Matérn 协方差核 [89, 90]。这里我们取 $\gamma = 1$，$\delta = 25$ 和 Θ 为对称正定矩阵，其沿特征方向 $[\sin(\pi/4),\ \cos(\pi/4)]$ 的特征值为 $\theta_1 = 2$，沿特征方向 $[\cos(\pi/4),\ \sin(\pi/4)]$的特征值为 $\theta_2 = 0.5$。

5.3.2. 使用傅里叶基进行形状参数化

我们使用关于水平坐标的傅里叶展开来表示上边界的形状，

$$x_2(x_1) = a_0 + \sum_{k=1}^{n_z} a_k \cos(k\pi x_1) + b_k \sin(k\pi x_1), \tag{68}$$

最大波数为 $n_z\pi$。这可以通过如下形式的变形来表述（61），其中位移基函数为

$$\begin{aligned} \mathfrak{d}_{2k}(X) &= (0, X_2 \cos(k\pi X_1))^T, \qquad k = 0, \dots, n_z, \\ \mathfrak{d}_{2k+1}(X) &= (0, X_2 \sin(k\pi X_1))^T, \qquad k = 1, \dots, n_z. \end{aligned}$$

以及 $z = [a_0, a_1, b_1, a_2, b_2, \dots, a_{n_z}, b_{n_z}]^T \in \mathbb{R}^{2n_z+1}$。具体而言，我们取 $n_z = 5$，从而得到 11 个位移基函数，并将可行集取为 $\mathcal{Z}_{ad} = [-0.2, 0.2]^{d_z}$。

5.3.3. 参考域公式

通过在弱形式中进行坐标变换，可以在参考域上表述该偏微分方程。在空间域上，我们有试探空间

$$\mathcal{U}(\Omega_z) := \{\widetilde{u} \in H^1(\Omega_z) : \widetilde{u} = 0 \text{ on } \Gamma_{B,z} \text{ and } \widetilde{u} = 1 \text{ on } \Gamma_{T,z}\}$$

以及检验空间

$$\mathcal{V}(\Omega_z) := \{\widetilde{v} \in H^1(\Omega_z) : \widetilde{v} = 0 \text{ on } \Gamma_{B,z} \text{ and } \widetilde{v} = 0 \text{ on } \Gamma_{T,z}\}.$$

对于任意 $\widetilde{v} \in \mathcal{V}(\Omega_z)$，弱形式为

$$\langle R_z(\widetilde{u}, \widetilde{m}), \widetilde{v} \rangle_{\mathcal{V}(\Omega_z)' \times \mathcal{V}(\Omega_z)} = \int_{\Omega_z} \exp(\widetilde{m}) \nabla_x \widetilde{u} \cdot \nabla_x \widetilde{v} \, dx - \int_{\Omega_z} f\widetilde{v} \, dx. \tag{69}$$

然后，我们将A.1 中的坐标变换应用于积分，得到用参考域量表示的偏微分方程残差的弱形式。为此，我们考虑参考域上的状态空间与检验空间，

$$\begin{aligned} \mathcal{U}(\Omega_0) &:= \{u \in H^1(\Omega_0) : u = 0 \text{ on } \Gamma_B \text{ and } u = 1 \text{ on } \Gamma_T\}, \\ \mathcal{V}(\Omega_0) &:= \{v \in H^1(\Omega_0) : v = 0 \text{ on } \Gamma_B \text{ and } v = 0 \text{ on } \Gamma_T\}. \end{aligned}$$

其中， Γ_T 和 Γ_B 分别是 Ω_0，的上边界与下边界。于是有

$$\langle \mathcal{R}(u, m, z), v \rangle_{\mathcal{V}(\Omega_0)' \times \mathcal{V}(\Omega_0)} = \int_{\Omega_0} \exp(m)(F_{\chi_z}^{-T} \nabla_X u) \cdot (F_{\chi_z}^{-T} \nabla_X v) \det F_{\chi_z} dX - \int_{\Omega_0} (T_{\chi_z} f) v \det F_{\chi_z} dX. \tag{70}$$

我们注意到，形状变量 z 通过变形梯度和 $F_{\chi_z} = \nabla_X \chi_z$ 以及拉回 T_{χ_z}，显式地出现在残差（70）中。

如前所述，我们将（70）视为形状优化中偏微分方程问题更为基本的定义。该问题采用伽辽金有限元法进行离散，对状态空间和随机参数空间均使用连续分片线性单元，从而得到 $d_m = d_u = 4,\ 257$。

5.3.4. 不确定性下的通量跟踪形状优化

对于 OUU问题，我们考虑底边界上的通量跟踪型目标函数。也就是说，给定目标通量分布 $\tau_{\text{target}} \in L^2(\Gamma_{B,\ z})$，我们寻求最小化法向通量 $-e^m \nabla_x u \cdot n$ 与目标之间的 $L^2(\Gamma_{B,\ z})$偏差，

$$Q_z(u, m) = \frac{1}{2} \int_{\Gamma_{B,z}} (-e^m \nabla_x u \cdot n - \tau_{\text{target}})^2 ds. \tag{71}$$

这可转换为参考坐标中的以下积分

$$Q(u,m,z)=\frac{1}{2}\int_{\Gamma_B}\left(-\exp(m)F_{\chi_z}^{-T}\nabla_X u\cdot N-\tau_{\text{target}}\right)^2 dS, \tag{72}$$

其中，我们利用 $\chi_z(X)=X$ 在 Γ_B 上成立这一事实进行了简化，因此单位法向量保持不变，即，

$$n(X)=\frac{\det F_{\chi_z}(X)F_{\chi_z}^{-T}(X)N(X)}{|\det F_{\chi_z}(X)F_{\chi_z}^{-T}(X)N(X)|}=N(X).$$

对于惩罚项，我们在 z上考虑一个 ℓ_2 范数，即，

$$\mathcal{P}_{\ell_2}(z;\alpha):=\alpha|\Omega_0|\|z\|_2^2, \tag{73}$$

其中， $\alpha>0$ 为惩罚权重，因子 $|\Omega_0|$表示 Ω_0, 的勒贝格测度，使用该因子是为了便于在 FEniCS 中实现。

我们考虑 用于目标函数的两个示例风险度量。在第一个示例中，我们考虑风险中性情形，使用 将期望作为优化目标，

$$\min_{z\in\mathcal{Z}_{ad}}\rho_{m\sim\mu_m}^{\text{mean}}(Q(u(m,z),m,z)+\mathcal{P}_{\ell_2}(z;\alpha). \tag{74}$$

第二个示例考虑以条件风险价值（CVaR）作为优化目标的风险厌恶情形 e,

$$\min_{z\in\mathcal{Z}_{ad}}\rho_{m\sim\mu_m}^{\text{CVaR},\beta}Q(u(m,z),m,z)+\mathcal{P}_{\ell_2}(z;\alpha), \tag{75}$$

其中 $\beta=0.95$，我们使用光滑CVaR公式（17）求解。在这两种情形下，目标函数都包含对形状变量的 ℓ_2 惩罚项，其中均值目标函数取 $\alpha=0.001$ ，并针对CVaR目标函数考察 $\alpha=0.001$ 和 $\alpha=0.1$ 。

5.3.5. 神经算子构造

神经算子的训练数据通过从 m和 z的输入分布中抽样生成，其中不确定参数分布采用 $\mu_m=\mathcal{N}(0,\ \mathcal{C}_m)$, 形状变量 $\mu_z=$采用均匀分布 Unif$([-0.2,0.2]^{d_z})$。我们使用大小分别为 512、1,024、2,048 和 4,096 的不同训练数据集训练神经网络，并分别考察有无雅可比信息的情况。每种情形下，为生成约化基，我们使用训练集中的 $n_{\text{POD}}=n_{\text{AS}}=512$ 个样本子集计算 POD 和主动子空间（AS），并分别取 $r_u=400$ 个POD基向量和 $r_m=256$ 个AS基向量来定义降维。潜空间映射由多输入全连接神经网络表示，该网络包含 4 个宽度为 512 的隐藏层，并采用 GELU 激活函数。随后使用 Adam 训练 2,000 个训练轮次，初始学习率为 5×10^{-4} ，并在训练 800、1,200 和 1,600 个训练轮次后减半。

在图1中，我们绘制了神经算子输出在 $\mathcal{U}=L^2(\Omega_0)$ 中的平均相对误差，以及其约化雅可比矩阵 $J_{m,\ r}(m,\ z)=\mathbf{\Phi}_{r_u}^*D_m u(m,\ z)\mathbf{\Psi}_{r_m}$ 和 $J_{z,\ r}(m,\ z)=\mathbf{\Phi}_{r_u}^*D_z u(m,\ z)$ 的平均相对误差。误差是在输入分布中抽取的 1,024 个样本组成的测试集上计算的，并对使用不同初始化和训练集/验证集划分进行的 10 次训练过程取平均。我们观察到，使用雅可比信息进行训练时，解近似及其相关敏感度的误差始终较低。

图2 展示了一个具有代表性的测试示例。我们将变形域上的状态可视化，其中（左上）为参考高保真偏微分方程解 u ，（右上）为使用 512 个样本训练得到的 Shape-DINO 预测 u_{true}, （左下）为逐点差异 $u_{\text{DINO}}-u_{\text{true}}$, （右下）为计算中使用的参数 m 。从图中可以看出，尽管训练数据集有限，状态的预测误差仍然较小。

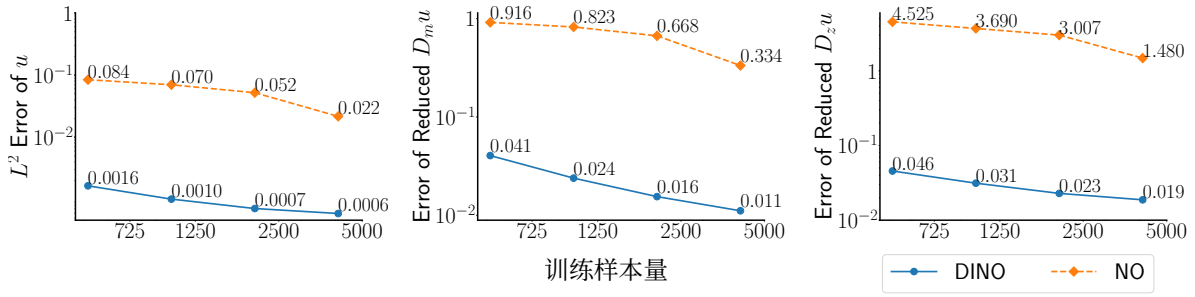


图1：采用导数信息（Shape-DINO）和不采用导数信息（Shape-NO）训练的泊松偏微分方程神经算子的全阶状态（$L^2(\Omega_0)$）（左）以及降阶雅可比矩阵（中、右）的均值相对测试误差。使用雅可比信息进行训练，在解及其敏感度方面均持续获得更低的误差，表明 Shape-DINO 的准确性和稳健性更佳。

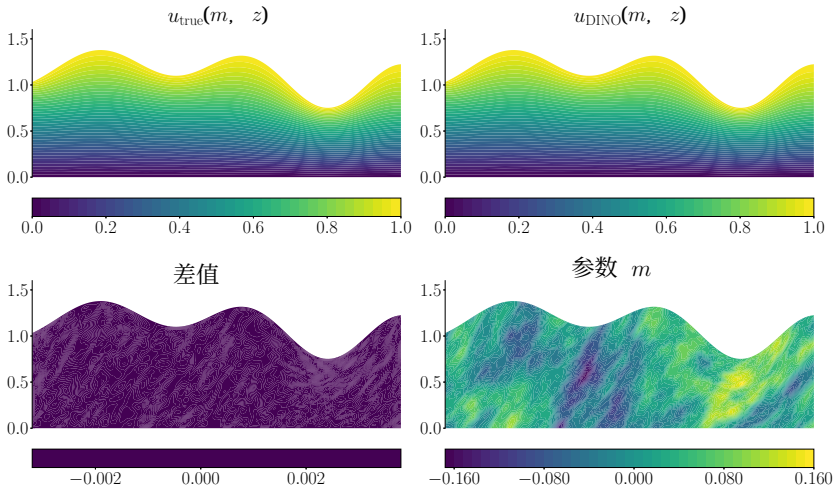


图2：变形域上绘制的状态 u。左上：真实的PDE状态 u_{true}。右上：使用 Shape-DINO 和 512 个训练样本得到的预测状态 u_{DINO}。左下：差值：$u_{\text{DINO}} - u_{\text{true}}$。右下：计算所用的参数 m。从视觉上看，预测误差接近于零，证实了 Shape-DINO 的准确性。

5.3.6. 形状优化与成本-精度比较

为评估 Shape-DINO 和 Shape-NO 的预测精度与优化精度，我们将它们优化后的目标函数值与基于偏微分方程的优化所得结果进行比较。对于基于偏微分方程的方法，目标函数通过样本平均近似（SAA）进行评估，每次迭代使用的样本数依次增加至 16、32、64、128、256 和 512；参考解则通过使用 16,384 个样本的高保真偏微分方程样本平均近似获得。

对于基于代理模型的方法（Shape-DINO 和 Shape-NO），优化时每次迭代使用固定的批量大小，即 2,048 个样本。由于这些样本由训练好的神经代理模型计算和评估，因此与训练数据生成所需的偏微分方程求解相比，其计算成本可以忽略不计。因此，在比较各方法的计算效率时，我们根据训练样本量报告代理模型的结果。训练样本量直接反映了构建各代理模型所需的偏微分方程求解次数，因此能够公平衡量其离线计算成本，并与基于偏微分方程的优化中的函数评估成本进行比较。

比较结果汇总于三组图中，见图3 至5。每组图均展示了优化后的目标函数值与状态偏微分方程求解次数（对于基于偏微分方程的样本平均近似）或等效的神经网络训练规模（对于基于代理模型的方法）之间的关系。除了平均相对误差曲线外，我们还展示了相应的25%˘75% 分位数带。每组图对应一种不同的目标函数配置：(i) 惩罚权重为 $\alpha = 0.001$的均值目标函数；(ii) $\alpha = 0.001$的 CVaR 目标函数；(iii) $\alpha = 0.1$的 CVaR 目标函数。

我们观察到几个关键现象。在所有风险设置下，对于相同甚至少几个数量级的偏微分方程求解次数，Shape-DINO 达到的优化误差始终显著小于基于偏微分方程的样本平均近似。尤其是，Shape-DINO 仅用 512 个训练样本便达到接近参考值的精度，展现出显著提升的样本效率。这一优势在惩罚权重较小的 CVaR 目标函数中最为明显，因为准确的风险估计需要大量样本。另一方面，对于风险中性目标函数或惩罚权重大得多的 CVaR 目标函数，Shape-DINO 与基于偏微分方程的方法之间的性能差距虽明显，但没有那么大。在这些情况下，目标函数
要么相对容易估计（如均值目标函数中的情况），要么受惩罚项主导（如带有 $\alpha = 0.1$ 的条件风险价值目标函数中的情况）。相比之下，在所有情况下，即使使用 4，096 个训练样本，Shape-NO 仍难以生成具有竞争力的解，这表明雅可比信息对于训练出准确且具有实际用途的代理模型可能至关重要；这类代理模型能够摊销离线训练成本，并在广泛的使用场景中提供可靠的性能。

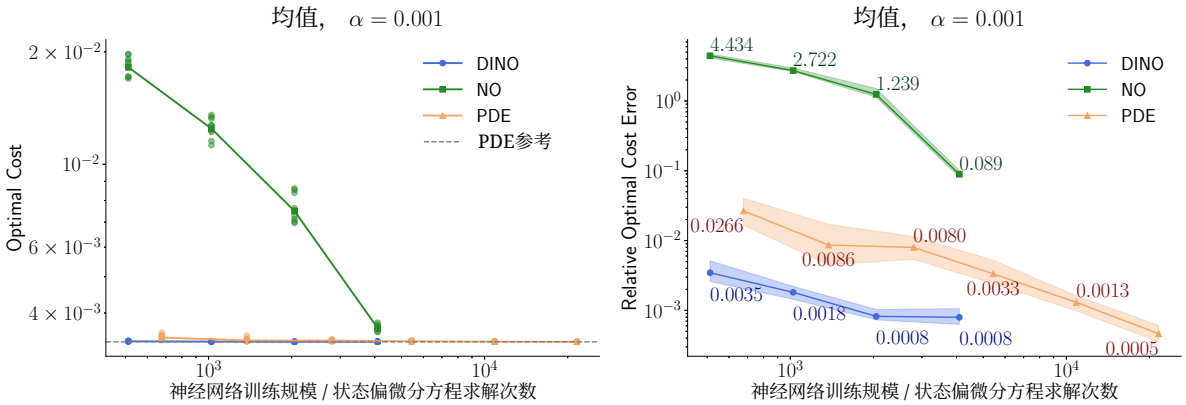


图3：最优解处的最优成本值（左）以及目标函数的相对误差（右）与优化所需状态偏微分方程求解次数的关系；两者均采用均值样本平均近似和惩罚权重 $\alpha = 0.001$。结果展示了 Shape-DINO、Shape-NO 和基于偏微分方程的样本平均近似。实线表示平均相对误差，阴影区域表示多次运行所得的 25%˘75%分位数带。即使目标函数容易估计（如均值目标函数），Shape-DINO 也能以远少于基于偏微分方程的方法所需的偏微分方程求解次数，快速收敛至参考解；而 Shape-NO 的误差更大，收敛也更慢。

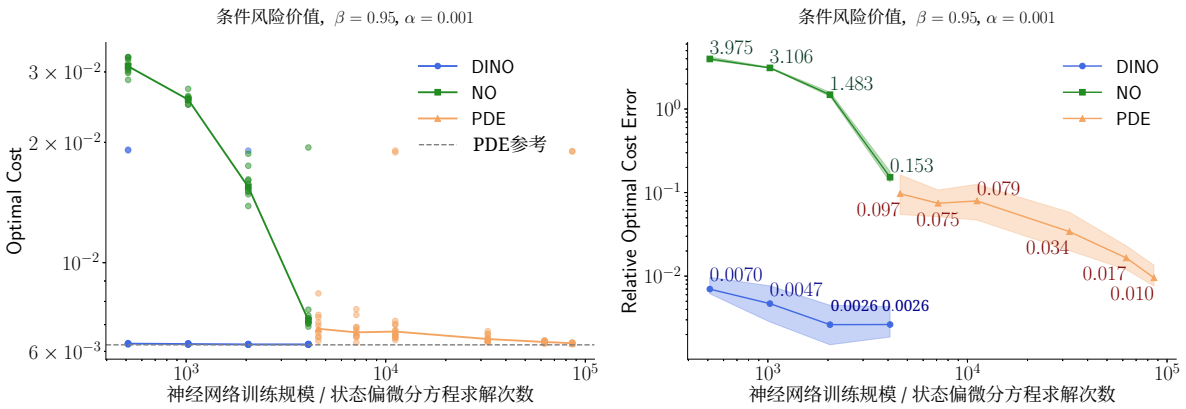


图4：最优解处的最优成本值（左）以及目标函数的相对误差（右）随优化所需状态偏微分方程求解次数的变化，二者均对应分位数 $\beta = 0.95$ 和惩罚权重 $\alpha = 0.001$。实线表示平均相对误差，阴影区域表示 25%˘75% 分位数带。在这一风险敏感度极高的设置中，Shape-DINO 仅需比基于偏微分方程的样本平均近似少几个数量级的偏微分方程求解次数，即可达到接近参考值的精度；而 Shape-NO 在训练规模较小和中等时表现不佳，凸显了导数信息对于准确进行风险估计的重要性。

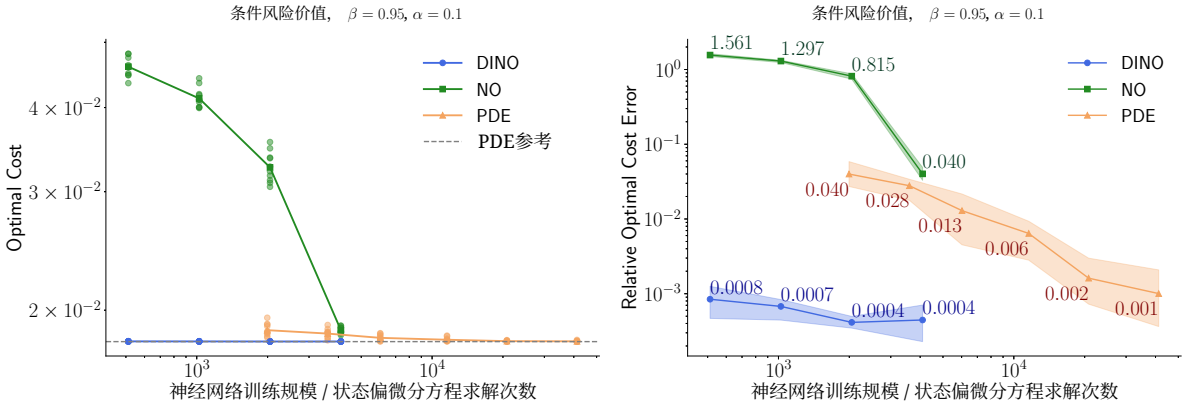
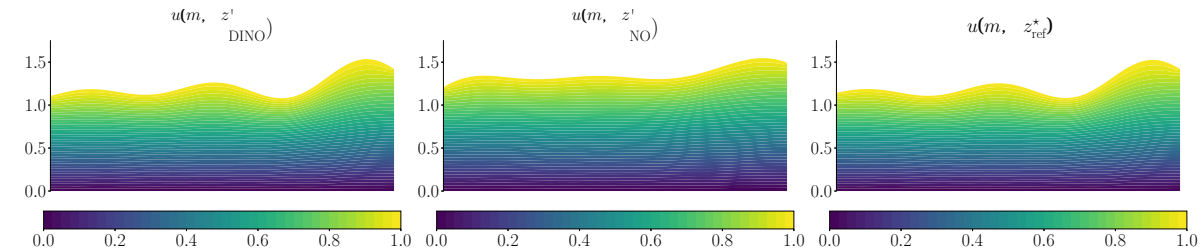


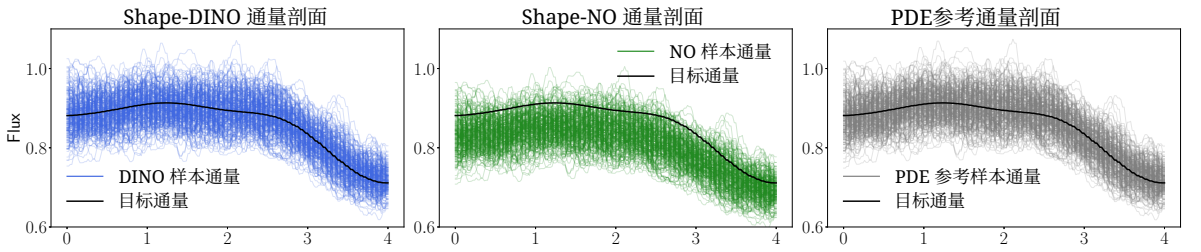
图5：最优解处的最优成本值（左）以及目标函数的相对误差（右）与优化所需的状态偏微分方程求解次数的关系，两者均对应分位数 $\beta = 0.95$ 和惩罚权重 $\alpha = 0.1$。实线表示平均相对误差，阴影区域表示 25%˘75% 分位数带。随着惩罚项占主导地位，Shape-DINO 与基于偏微分方程的方法之间的性能差距缩小，但 Shape-DINO 在准确性和离线计算成本之间仍能提供最佳权衡，而 Shape-NO 的可靠性仍显著较低。

图6 展示了在分位数 $\beta = 0.95$ 和惩罚权重 $\alpha = 0.001$下通过 CVaR–SAA 优化得到的代表性最优设计及其对应的状态解。顶行展示了 Shape-DINO（$z^{\dagger}_{\text{DINO}}$）和 Shape-NO（$z^{\dagger}_{\text{神经算子}}$）得到的最优上边界设计，并与 PDE 参考（$z^{\star}_{\text{ref}}$）进行比较，同时展示了在变形域上绘制的一个 PDE 状态样本。参考解 $z^{\star}_{\text{ref}}$ 通过包含 16,284 个样本的高保真PDE型SAA计算得到。底行给出了在各最优设计处评估的样本通量剖面。黑色曲线表示目标通量 τ_{target}，彩色曲线分别对应使用 $u(m^{(i)}, z^{\dagger}_{\text{DINO}})$、$u(m^{(i)}, z^{\dagger}_{\text{神经算子}})$ 和 $u(m^{(i)}, z^{\star}_{\text{ref}})$ 针对 256 个独立参数样本 $m^{(i)} \sim \mu_m$计算得到的通量剖面。值得注意的是，$z^{\dagger}_{\text{DINO}}$ 得到的上边界形状与参考解 $z^{\star}_{\text{ref}}$ 相近，且其通量剖面与目标通量的吻合程度远高于 Shape-NO，性能非常接近参考解。

为展示算子代理的泛化能力，我们使用训练好的 Shape-DINO 和 Shape-NO 分别求解两个额外的 OUU问题，每个问题对应一个不同的目标通量，并重点考察使用 512 个样本训练的神经算子。随后，我们比较图7–8中与最优形状设计相对应的通量剖面样本。这里，第一个案例对应与训练形状分布一致的光滑目标通量，第二个案例则引入了可行集中的形状无法实现的不连续目标通量。结合图6，在所有案例中，Shape-DINO 产生的通量响应始终更接近指定目标，并且在不同参数实现之间的变异性更小。值得注意的是，即使在分布外设置下，Shape-DINO 得到的通量剖面也与大样本PDE型SAA的结果相当，而离线成本则小了几个数量级。在分布内和分布外设置下观察到的鲁棒性提升，凸显了在训练此类 OUU问题的神经代理模型时纳入导数信息的重要性。



(a) 最优设计



(b) 最优设计处的通量剖面

图 6：上：采用分位数 $\beta = 0.95$ 和惩罚权重 $\alpha = 0.001$进行条件风险价值-SAA 优化得到的最优上边界设计。三个子图分别展示 Shape-DINO、Shape-NO 和 PDE参考的最优设计，以及它们在变形域上的对应 PDE状态解 $u(m, z)$。此处，$z^{\star}_{\text{ref}}$ 由包含 16,284 个样本的大样本基于偏微分方程的样本平均近似获得。下：最优设计处的样本通量剖面。黑线表示目标通量 τ_{target}。彩色线分别对应使用 $u(m^{(i)}, z^{\dagger}_{\text{DINO}})$、$u(m^{(i)}, z^{\dagger}_{\text{NO}})$ 和 $u(m^{(i)}, z^{\star}_{\text{ref}})$ 对 256 个独立参数样本 $m^{(i)} \sim \mu_m$计算得到的通量剖面。与 Shape-NO 相比，Shape-DINO 的最优设计生成的通量剖面与目标通量 σ_{target} 更加吻合，且其性能非常接近 PDE参考。

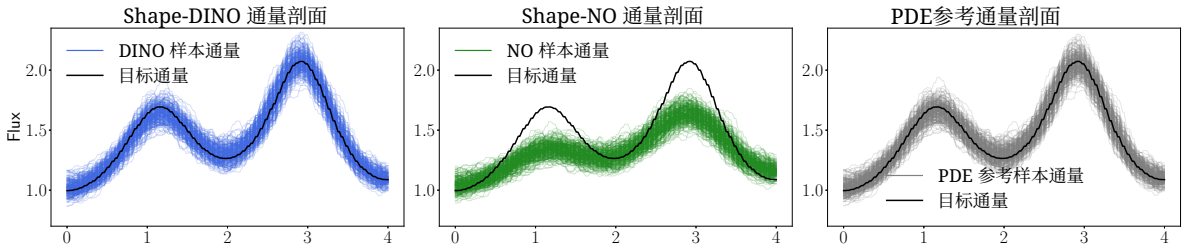


图 7：采用不同于图 $\beta = 0.95$ 和 $\alpha = 0.001$的光滑目标通量分布，通过条件风险价值优化获得的最优设计处的样本通量剖面。6DINO 和神经算子均使用 512 个样本进行训练，而 $z^{\star}_{\text{ref}}$ 则通过含 16,284 个样本的大样本基于偏微分方程的样本平均近似获得。尽管离线阶段使用的偏微分方程求解次数少了好几个数量级，Shape-DINO 生成的通量剖面仍与大样本 PDE 参考结果几乎无法区分，在不同参数实现下紧密跟踪目标通量，并且表现出更低的变异性；相比之下，Shape-NO 的偏差要大得多。

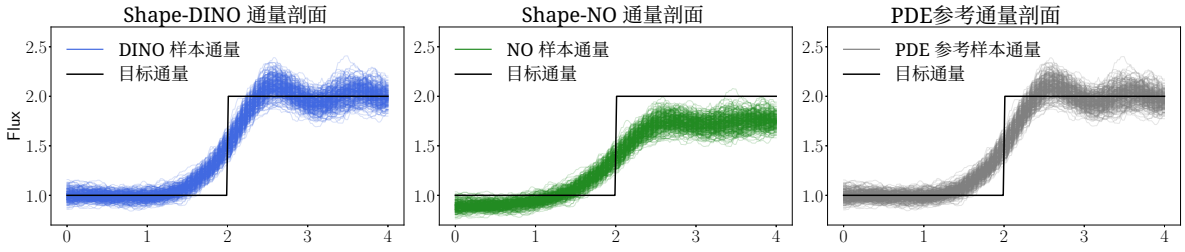


图 8：采用不连续目标通量分布，通过条件风险价值优化获得的最优设计处的样本通量剖面，其中 $\beta = 0.95$ 和 $\alpha = 0.001$；该目标通量分布由 $x_2 = 1$ 定义：对于 $x_1 \in [0, 2]$ 取值为 $x_2 = 2$，对于 $x_1 \in [2, 4]$取值为。DINO 和神经算子均使用 512 个样本进行训练，而 PDE 参考结果则通过大样本基于偏微分方程的样本平均近似获得。尽管目标通量位于训练所用分布之外，Shape-DINO 生成的通量剖面仍能更好地匹配目标，并在不确定性变化下保持稳定；相比之下，Shape-NO 生成的通量剖面与目标存在显著偏差。

5.4. 多种不确定来流条件下二维流动的耗散降低设计 *5.4.1.* 含可变形状物体的纳维–斯托克斯方程

接下来，我们考虑通道区域 $\Omega_{\text{channel}} = (0, 2) \times (0, 1) \subset \mathbb{R}^2$ 中的稳态不可压缩纳维–斯托克斯问题，其中包含一个初始为圆形的物体 $\Omega_{\text{object}} = \{x \in \mathbb{R}^2 : (x_1 - 0.5)^2 + (x_2 - 0.5)^2 < 0.1^2\}$，其半径为 $r = 0.1$，圆心位于 $(0.5, 0.5)^T$。目标是通过自由形变来设计物体的形状，使粘性耗散最小化；物体的形状通过一族微分同胚变形 χ_z 进行变换，从而得到设计 $\Omega_{\text{object},z} = \chi_z(\Omega_{\text{object}})$。这种设置常用作基准配置，用于评估形状诱导的减阻效果以及受限流动中的耗散控制 [91]。

这里，我们通过考虑两个相互独立的入流情景引入了额外的难度：在这两个情景中，物体分别面对水平方向入流（东向流动）或垂直方向入流（北向流动），以此作为不同风荷载方向下形状设计问题的原型。我们通过考虑两组偏微分方程来构造该问题，这两组方程的区域之间的差异在于物体沿直线 $x_1 = x_2$ 发生了反射。对于东向流动情景，我们采用物体的默认方向 $\Omega_{\text{object}}^{\text{East}} = \Omega_{\text{object},\ z}$，此时流动区域变为

$$\Omega_z^{\text{East}} = \Omega_{\text{channel}} \setminus \text{cl}\left(\Omega_{\text{object}}^{\text{East}}\right),$$

其中，$\text{cl}(\cdot)$ 表示闭包。对于北向流动，我们对物体的方向进行反射，$\Omega_{\text{object}}^{\text{North}} = \text{Reflect}(\Omega_{\text{object},\ z})$，使得流动区域为

$$\Omega_z^{\text{North}} = \Omega_{\text{channel}} \setminus \text{cl}\left(\Omega_{\text{object}}^{\text{North}}\right),$$

其中，$\text{Reflect}(\cdot)$ 表示关于直线 $x_1 = x_2$ 的反射。注意，该操作相当于先绕中心点 $(0。5，\ 0。5)^T$ 顺时针旋转 90 度，再沿直线 $(x_1，\ 0。5)$ 进行反射。由于我们随后将考虑关于该反射保持不变的来流分布，因此这次额外的反射不会影响问题定义。

我们现在针对区域 $\Omega_z^s, s \in \{$东,北$\}$上的两种情形建立偏微分方程，其状态在空间坐标中由速度-压力对 $\widetilde{u}^s = (\widetilde{v}^s, \widetilde{p}^s)$ 给出。利用柯西应力张量，

$$\widetilde{\mathbf{T}}(\widetilde{u}^s) := 2\nu\widetilde{\varepsilon}(\widetilde{v}^s) - \widetilde{p}^s I, \tag{76}$$

其中

$$\widetilde{\varepsilon}(\widetilde{v}^s) := \text{sym}(\nabla_x \widetilde{v}^s) = \frac{\nabla_x \widetilde{v}^s + (\nabla_x \widetilde{v}^s)^T}{2} \tag{77}$$

为应变率张量，我们可以将控制方程写为

$$(\widetilde{v}^s \cdot \nabla_x)\widetilde{v}^s - \nabla_x \cdot \widetilde{\mathbf{T}}(\widetilde{u}^s) = 0 \quad \text{in } \Omega_z^s, \tag{78a}$$
$$\nabla_x \cdot \widetilde{v}^s = 0 \quad \text{in } \Omega_z^s, \tag{78b}$$
$$\widetilde{v}^s - (\widetilde{v})_{\text{in}}^s = 0 \quad \text{on } \Gamma_{\text{in}}, \tag{78c}$$
$$\widetilde{\mathbf{T}}(\widetilde{u}^s)n = 0 \quad \text{on } \Gamma_{\text{out}}, \tag{78d}$$
$$\widetilde{v}^s = 0 \quad \text{on } \Gamma_{\text{no-slip}}, \tag{78e}$$
$$\widetilde{\mathbf{T}}(\widetilde{u}^s)n \cdot e_t = 0 \quad \text{on } \Gamma_{\text{free-slip}}, \tag{78f}$$
$$\widetilde{v}^s \cdot n = 0 \quad \text{on } \Gamma_{\text{free-slip}}, \tag{78g}$$

其中 e_t 是边界上的单位切向量，且

$$\Gamma_{\text{in}} = \{x \in \partial\Omega_z^s : x_1 = 0\}, \qquad \Gamma_{\text{no-slip}} = \partial\Omega_{\text{object}}^s,$$
$$\Gamma_{\text{out}} = \{x \in \partial\Omega_z^s : x_1 = 2\}, \qquad \Gamma_{\text{free-slip}} = \{x \in \partial\Omega_z^s : x_2 = 0 \text{ or } x_2 = 1\}.$$

来流剖面由不确定模型参数决定：

$$(\widetilde{v})_{\text{in}}^s(x) = \exp(\widetilde{m}^s(x))e_1, \qquad e_1 = (1, 0)^T.$$

我们定义参考域 $\Omega_0 := \Omega_{\text{channel}} \setminus \text{cl}(\Omega_{\text{object}})$ 上 $\widetilde{m}^s$ 高斯随机场的分布，使得 $m^s \sim \mu_m^s = \mathcal{N}(0, \mathcal{C}_m^s)$，且每个场景$s \in \{$东、北$\}$对应一个不同的双拉普拉斯协方差算子 $\mathcal{C}_m^s$。具体而言， $\mathcal{C}_m^s$ 由式（67）给出，其中我们取 $\Theta = I$ 为恒等矩阵，并取δ^s、 γ^s ，使两个流动场景的相关长度均为 0.25，而$s =$东的点方差为 0.2，$s =$ 北的点方差为 0.1。

5.4.2. 使用自由形变进行形状参数化

可容许形状的变化采用自由形变方法表示，其中 χ_z 由伯恩斯坦多项式给出。由于具有灵活性、光滑性以及相对低维的边界运动参数化方式，这种自由形变策略已广泛应用于偏微分方程约束优化中 [10, 11, 14, 71]。我们首先介绍基于 [92] 的方法，然后说明如何将这种参数化解释为式（61）的形式。

为此，我们考虑包含物体 Ω_{object} 的单位正方形 $\Omega_{\text{control}} := (0, 1)^2$ ，以及控制点格 $\xi_{k,l} = (k/K, l/L)^T$，其中 $k = 0, \ldots, K$，且 $l = 0, \ldots L$，其中 K， L 为整数。接着，我们定义伯恩斯坦多项式；当最高次数为 n_p 时，其形式为

$$b_j^{n_p}(t) := \binom{n_p}{j} t^j (1-t)^{n_p - j}, \quad j = 0, \ldots, n_p. \tag{79}$$

在二维中，我们可以由一维伯恩斯坦多项式构造张量积伯恩斯坦多项式 ls w

$$b_{k,l}^{K,L}(t_1, t_2) = b_k^K(t_1) b_l^L(t_2), \qquad k = 0, \ldots, K, \qquad l = 0, \ldots, L. \tag{80}$$

接下来，令 $\delta_{k,l} = ((\delta_{k,l})_1, (\delta_{k,l})_2)^T \in \mathbb{R}^2$ 表示控制点 $\xi_{k,l}$的位移，使其位移后的位置为 $\xi_{k,l} = \xi_{k,l} + \delta_{k,l}$。我们令控制点的运动定义变形映射，使得

$$\chi(X) := \sum_{k=0}^{K} \sum_{l=0}^{L} b_{k,l}^{K,L}(X_1, X_2) \widetilde{\xi}_{k,l}. \tag{81}$$

显然，控制点位移 $\delta_{k,l}$ 可以对变形进行参数化。为将其与形式为 (61) 的变形联系起来，我们注意到，根据伯恩斯坦多项式的性质 $\sum_{k=0}^{K} k b_k^K(t), \; = Kt$ 可得

$$\chi(X) = \sum_{k=0}^{K} \sum_{l=0}^{L} b_{k,l}^{K,L}(X_1, X_2)(\xi_{k,l} + \delta_{k,l}) = X + \sum_{k=0}^{K} \sum_{l=0}^{L} b_{k,l}^{K,L} \delta_{k,l}. \tag{82}$$

Th因此，我们也可以使用位移基函数 χ_z 将其等价地定义为形式 (61)

$$\mathfrak{d}_{k,l}^1(X) = b_{k,l}^{K,L}(X_1, X_2) e_1, \quad \mathfrak{d}_{k,l}^2(X) = b_{k,l}^{K,L}(X_1, X_2) e_2,$$

其中 $e_1 = (1, 0)^T$和 $e_2 = (0, 1)^T$，以及形状变量 $z_{k,l}^j = (\delta_{k,l})_j$，其中 $k = 0, \ldots K, l = 0, \ldots L$，且 $j = 1, 2$。

在我们的问题中，我们将采用一个 $K = 10$ 乘 $L = 10$ 控制点格，覆盖区域Ω_{control}。不过，为确保外边界保持不动，我们不将第一个和最后一个控制点纳入优化变量；这样也可以将其余所有位移基函数在 $X_1 \geq 1$之外扩展为零。因此，最终得到的可容许变形由位移基函数$\mathfrak{d}_{k,l}^j(X)$ 以及形状变量$z_{k,l}^j$组成，其中 $k = 1, \ldots K-1$， $l = 1, \ldots L-1$和 $j = 1, 2$，共计 $d_z = 162$ 个自由度。这些位移还通过线弹性方程（63）进行预处理，以改善流体域内变形的条件。

为了生成向北配置下的几何变化，我们保持相同的参考域 Ω_0 ，并通过交换位移基函数中的参数坐标来构造镜像位移基，即

$$\mathfrak{d}_{k,l}^1(X) = b_{k,l}^{K,L}(X_2, X_1) e_2, \quad \mathfrak{d}_{k,l}^2(X) = b_{k,l}^{K,L}(X_2, X_1) e_1,$$

这实际上使位移场沿对角线 $X_1 = X_2$发生镜像反射，从而改变物体相对于固定流入边界的方向。图中展示了两种来流情况下的变形示例以及参考通道域 9。

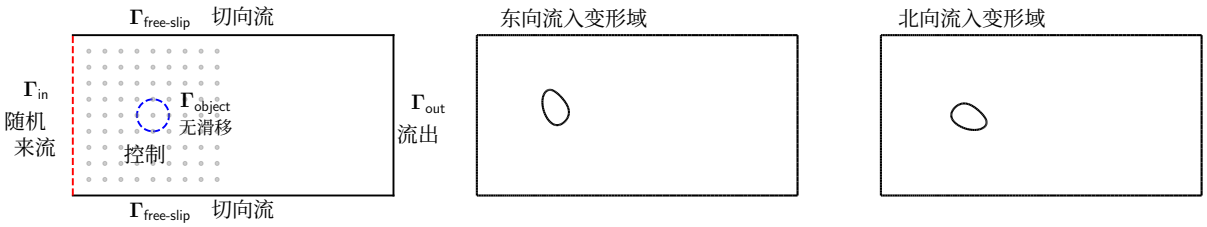


图9：左图显示参考通道域 $\Omega_{\text{channel}} = (0,\ 2) \times (0,\ 1)$及其标注的边界。障碍物是一个半径为 0。1、中心位于 (0。5，0。5) 的圆。灰色点为伯恩斯坦控制点 $\xi_{k,\ l}$。中图和右图分别显示对应的东向和北向位移基作用下的变形域；为清晰起见，仅显示区域边界。

5.4.3. 参考域公式

我们以弱形式表述纳维–斯托克斯方程（78a），并采用 Nitsche方法弱施加涉及模型参数的入口边界条件。在空间域中，我们定义速度和压力的函数空间如下

$$\mathcal{U}_v(\Omega_z^s) := \{v \in H^1(\Omega_z^s) : v = 0 \text{ on } \Gamma_{\text{no-slip}},\ v \cdot n = 0 \text{ on } \Gamma_{\text{free-slip}}\}, \qquad \mathcal{U}_p(\Omega_z^s) := L^2(\Omega_z^s),$$

使得试探空间由 $\mathcal{U}(\Omega_z^s) := \mathcal{U}_v(\Omega_z^s) \times \mathcal{U}_p(\Omega_z^s)$给出，检验空间由 $\mathcal{V}(\Omega_z^s) := \mathcal{U}(\Omega_z^s)$给出。对于任意 $\widetilde{y}^s = (\widetilde{w}^s, \widetilde{q}^s) \in \mathcal{V}(\Omega_z^s)$，我们有

$$r_z^s(\widetilde{u}^s, \widetilde{m}^s, \widetilde{y}^s) := \langle \mathcal{R}_z(\widetilde{u}^s, \widetilde{m}^s), \widetilde{y}^s \rangle_{\mathcal{V}(\Omega_z^s)' \times \mathcal{V}(\Omega_z^s)} = r_{\text{flow}}^s(\widetilde{u}^s, \widetilde{m}^s, \widetilde{y}) + r_{\text{Nitsche}}^s(\widetilde{u}^s, \widetilde{m}^s, \widetilde{y}) \tag{83}$$

其中

$$\begin{aligned} r_{\text{flow}}^s(\widetilde{u}^s, \widetilde{m}^s, \widetilde{y}^s) = &\int_{\Omega_z^s} ((\widetilde{v}^s \cdot \nabla_x)\widetilde{v}^s) \cdot \widetilde{w}^s dx + \int_{\Omega_z^s} 2\nu\, \widetilde{\varepsilon}(\widetilde{v}^s) : \widetilde{\varepsilon}(\widetilde{w}^s) dx \\ &- \int_{\Omega_z^s} \widetilde{p}^s\, (\nabla_x \cdot \widetilde{w}^s)\, dx + \int_{\Omega_z^s} (\nabla_x \cdot \widetilde{v}^s)\, \widetilde{q}^s dx \end{aligned}$$

and

$$\begin{aligned} r_{\text{Nitsche}}^s(\widetilde{u}^s, \widetilde{m}^s, \widetilde{y}^s) = &- \int_{\Gamma_{\text{in}}} \widetilde{w}^s \cdot (\widetilde{\mathbf{T}}(\widetilde{u}^s)n) ds - \int_{\Gamma_{\text{in}}} (\widetilde{v}^s - (\widetilde{v})_{\text{in}}^s) \cdot (2\nu\widetilde{\varepsilon}(\widetilde{w}^s)n + \widetilde{q}^s n) ds \\ &- \int_{\Gamma_{\text{in}}} \widetilde{w}^s \cdot (\widetilde{v}^s - (\widetilde{v})_{\text{in}}^s)(\widetilde{v}^s \cdot n) ds + \int_{\Gamma_{\text{in}}} c_d \nu (\widetilde{v}^s - (\widetilde{v})_{\text{in}}^s) \cdot \widetilde{w}^s ds. \end{aligned}$$

这里，c_d 是　为确保正确收敛速率而选取的常数，针对狄利克雷边界条件。我们　下文讨论其选取。

接下来，我们将上述弱形式转换为参考域形式。由于两个流动场景的参考域相同，因此两者可以使用相同的试探空间和检验空间，即 $\mathcal{U}(\Omega_0) = \mathcal{V}(\Omega_0) = \mathcal{U}_v(\Omega_0) \times \mathcal{U}_p(\Omega_0)$，其中

$$\mathcal{U}_v(\Omega_0) := \{v \in H^1(\Omega_0) : v = 0 \text{ on } \partial\Omega_{\text{object}},\ v \cdot n = 0 \text{ on } \Gamma_{\text{free-slip}}\}, \qquad \mathcal{U}_p(\Omega_0) := L^2(\Omega_0)$$

利用应变率张量和应力张量的变量变换，

$$\varepsilon(v^s) := \text{sym}(\nabla_X v^s F^{-1}), \tag{84}$$

$$\mathbf{T}(u^s) := 2\nu\varepsilon(v^s) - p^s I. \tag{85}$$

我们可得，对于任意 $y \in \mathcal{V}(\Omega_0)$,

$$r(u^s, m^s, z, y) := \langle \mathcal{R}(u^s, m^s, z), y \rangle_{\mathcal{V}' \times \mathcal{V}} = r_{\text{flow}}(u^s, m^s, z, y) + r_{\text{Nitsche}}(u^s, m^s, z, y), \tag{86}$$

其中

$$\begin{aligned} r_{\text{flow}}(u^s, m^s, z, y) = &\int_{\Omega_0} \left(\nabla_X v^s F^{-1} v^s\right) \cdot w^s \det F dX + \int_{\Omega_0} 2\nu\varepsilon(v^s) : \varepsilon(w^s) \det F dX \\ &- \int_{\Omega_0} p^s \text{Tr}(\nabla_X w^s F^{-1}) \det F dX + \int_{\Omega_0} (\text{Tr}(\nabla_X v^s F^{-1}) q^s \det F dX, \end{aligned} \tag{87}$$

and

$$\begin{aligned} r_{\text{Nitsche}}(u^s, m^s, z, y) = &- \int_{\Gamma_{\text{in}}} w^s \cdot \mathbf{T}(u^s) N dS - \int_{\Gamma_{\text{in}}} (v^s - v^s_{\text{in}}) \cdot (2\nu\varepsilon(w^s) + q^s I) N dS \\ &- \int_{\Gamma_{\text{in}}} w^s \cdot (v^s - v^s_{\text{in}})(v^s \cdot N) dS + \int_{\Gamma_{\text{in}}} c_d \nu (v^s - v^s_{\text{in}}) \cdot w^s dS \end{aligned} \tag{88}$$

请注意，我们再次基于以下事实进行了简化：在自由形变下，左边界保持固定，因此单位法向量不会改变。此外，在这种表示下，入流剖面可简单表示为 $v^s_{\text{in}}(X) = \exp(m^s(X)) e_1$。

在参考域上，使用三角网格上的混合伽辽金公式对偏微分方程进行离散化。对于状态 $u = (v, p)$，我们采用 Taylor–Hood 元，其中速度 v 采用连续分片二次单元，压力 p采用连续分片线性单元，从而得到状态维数$d_u = 86,205$。随机参数场也采用连续分片二次单元，因此得到$d_m = 201$。对于 Nitsche 方法中的惩罚项，我们选择 $c_d = 10^5/h_e$，其中 h_e 是参考域中每个三角形单元对应的单元尺寸。控制变量由维数为 $d_{\mathcal{Z}} = 162$ 的有限维欧几里得空间表示。

5.4.4. 不确定流入条件下的二维减耗散形状设计

我们的优化问题旨在同时最小化由两种不同入流配置 $s \in$ {东、北}分别引起的黏性耗散率；这一指标在物理上与作用于物体上的阻力相关。我们用空间坐标表示耗散感兴趣量：

$$Q^s_z(\widetilde{u}^s) := \int_{\Omega^s_z} 2\nu\widetilde{\varepsilon}(\widetilde{v}^s) : \widetilde{\varepsilon}(\widetilde{v}^s)\, dx. \tag{89}$$

通过应变张量的变量替换，我们可以在参考域中表述该问题：

$$Q(u^s, m^s, z) = \int_{\Omega_0} 2\nu\varepsilon(v^s) : \varepsilon(v^s) \det F dX. \tag{90}$$

为了量化不确定性下的性能，我们定义了一个加权熵风险目标函数，以加权和的形式结合两种流动场景：

$$\mathcal{J}^{\text{weighted}}(z) := \mathcal{J}^{\text{East}}(z) + 0.3\mathcal{J}^{\text{North}}(z) + \mathcal{P}_{\ell_2}(z; \alpha), \tag{91}$$

其中，我们取

$$\mathcal{J}^s(z) := \rho^{\text{entropic},\beta}_{m^s \sim \mu^s_m}(Q(u^s(m^s, z), m^s, z)), \tag{92}$$

作为每种入口流动情景 $s \in$ {东、北} 的熵风险度量，并令 $\mathcal{P}_{\ell_2}$ 为 ℓ_2 惩罚泛函（73）。这里，取 $\alpha = 0.001$ 为惩罚权重，取 $\beta = 1$ 为适度的风险厌恶水平。

然后，我们考虑受控制偏微分方程约束以及形状变量盒约束的形状优化问题 z (73)。此外，我们施加体积约束，以防止不切实际的

物体收缩或增长，要求体积保持在其初始值的30%至120%之间 $V_0 := |\Omega_{\text{object}}| = \pi \cdot 0.1^2$。也就是说，我们求解该最小化问题

$$\min_{z \in \mathcal{Z}_{ad}} \mathcal{J}^{\text{weighted}}(z), \qquad \mathcal{Z}_{ad} = \{z \in [-1,1]^{d_z} : \text{Vol}(z) \in [0.3V_0, 1.2V_0]\}, \tag{93}$$

其中，体积泛函实现为 $\text{Vol}(z) = |\Omega^{\text{channel}}| - \int_{\Omega_0} \det F dX$.

5.4.5. 神经算子构造

对于这个问题，我们构造两个独立的神经算子 $u_\theta^s(m^s, z)$，分别近似两个流动场景 $s \in$ {东，北}对应的解算子 $u^s(m^s, z)$。每个神经算子的训练数据通过对输入分布 $m^s \sim \mu_m^s$ 和 $z \sim \mu_z = \text{Unif}([-1,1]^{d_z})$ 采样生成。我们将每个流动场景的训练问题与其他场景完全分开处理。对于每个流动场景，我们使用大小分别为 512、1,024、2,048 和 4,096 的不同训练数据集训练神经网络，并分别考虑使用和不使用雅可比信息的情况。在每种情况下，我们使用 $n_{\text{POD}} = n_{\text{AS}} = 512$ 个样本的子集来计算秩为 $r_u = 256$ 的状态POD基和秩为 $r_m = 50$的参数AS基。我们使用具有 3 个隐藏层（宽度分别为 512、1,024、512）的多输入全连接神经网络表示潜在空间映射，每层均采用 GELU 激活函数。随后，使用 Adam 将神经网络训练 1,000 个训练轮次，初始学习率为 1×10^{-4}，并在第 500 和第 750 个训练轮次后减半。

图中展示了东向流动场景下神经网络所获得的泛化误差10 以及北向流动场景下的泛化误差，见图11，误差随训练样本数量变化。报告的测试误差表示这10 次运行的平均表现。与泊松问题类似，我们再次观察到，使用导数信息训练的神经算子在输出和导数方面都能显著降低泛化误差。这一结论对两种流动场景均成立。

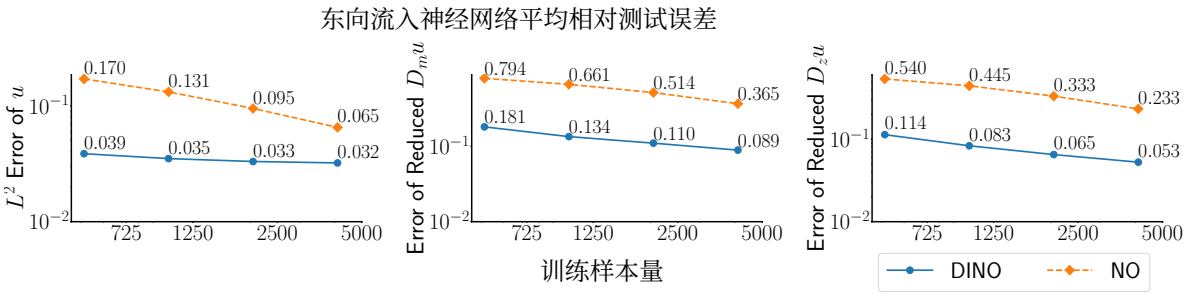


图10：二维纳维-斯托克斯向东流入案例中，采用雅可比训练（Shape-DINO）和未采用雅可比训练（Shape-NO）的神经算子的状态（左）及雅可比矩阵（中、右）测试误差。训练样本规模为512、1,024、2,048、4,096，测试样本规模为1,024。与Shape-NO相比，Shape-DINO在输出和导数方面均显著降低了泛化误差。

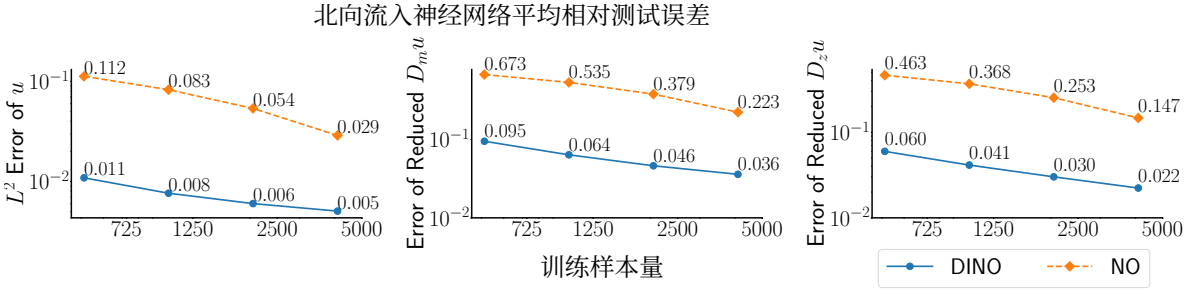


图11：二维纳维-斯托克斯向北来流案例中，采用雅可比训练（Shape-DINO）和未采用雅可比训练（Shape-NO）的神经算子的状态（左）及雅可比矩阵（中、右）测试误差。训练样本规模为512、1024、2048、4096，测试样本规模为1024。与向东流入案例类似，Shape-DINO在预测状态及其导数方面的误差再次显著低于Shape-NO。

5.4.6. 形状优化与成本-精度比较

随后，我们使用训练好的神经算子求解耗散最小化问题。图12 比较了两种入流配置在优化前后的代表性流场。图12a 和12c 分别展示了左侧和顶部入流场景下初始物体几何形状周围的速度场，这些速度场由高保真偏微分方程求解得到。相比之下，图12b 和12d 展示了在形状 DINO 得出的最优形状 $z^{\dagger}_{\text{DINO}}$处评估得到的相应优化后流场；所用形状 DINO 模型以 1024 个样本训练。优化后的几何形状显著减少了物体后方的回流，从而与初始形状相比降低了整体黏性耗散率。此外，物体整体上沿东向呈流线型，但并非完全对称，因为它还需要兼顾权重较小的北向流动。

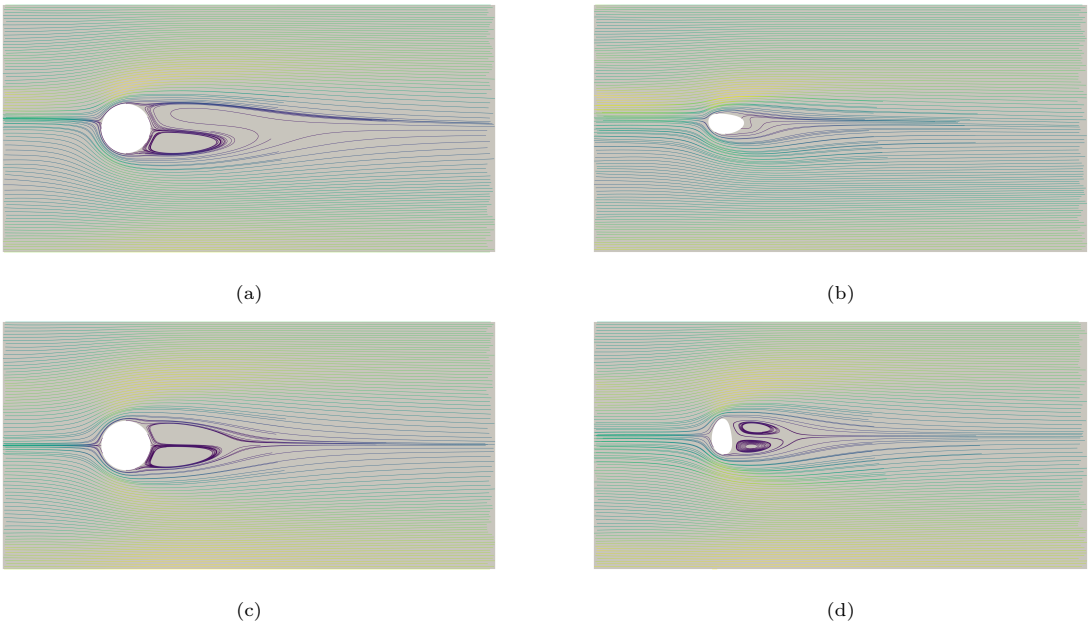


图12：采用经过 1024 个样本训练的 DINO 得到的初始物体几何形状与优化后几何形状所对应的代表性流场对比。每一行对应两种入流情景之一（向东流动和向北流动），每一列展示优化前（左列）和优化后（右列）的速度场。所有流场均通过高保真偏微分方程求解进行评估。尽管优化后的几何形状 $z^{\dagger}_{\text{DINO}}$ 并不对称，但与初始形状相比，它显著减少了物体后方的回流，并降低了整体黏性耗散率。

这些由几何形状引起的流动改善促使我们对风险规避设计性能进行定量评估。为此，我们展示了最优目标函数值随状态偏微分方程求解次数的变化，以及相应的 (25%–75%)分位数带，如图13所示。对于神经算子替代模型，DINO 和神经算子均使用 512、1,024、2,048 和 4,096 个样本进行训练，而所有基于代理模型的不确定性下优化计算均采用每次迭代8,192 个样本。相比之下，基于偏微分方程的不确定性下优化问题采用每次迭代 16、32、64、128、256 和 512 个样本求解，而偏微分方程参考不确定性下优化问题则采用每次迭代 8,192 个样本。随后，所有最优设计 $z^{\star}$ 均使用参考评估中相同的 16,384 个样本进行评估。与泊松偏微分方程示例一样，误差取 10 次运行的平均值；这些运行采用不同的随机入流样本，其随机种子与神经网络训练中使用的相同。

我们从这些结果中得出几项重要观察。首先，即使样本平均近似形式中的样本量相对较小，基于偏微分方程的方法对于黏性耗散目标函数仍能得到相当准确的不确定性下的优化解。尽管如此，DINO 代理模型仅需近一个数量级更少的状态偏微分方程求解次数，就能达到相当或更低的最优耗散值，展现出显著的样本效率提升。相比之下，不采用雅可比学习的神经算子代理模型并未实现同等的成本效益

，因为在相同的计算预算下，其性能并未优于基于偏微分方程的方法。在状态偏微分方程求解次数预算相同的情况下，基于DINO的不确定性下优化得到的最优耗散值低于基于NO的不确定性下优化和直接基于偏微分方程的不确定性下优化，表明DINO在风险规避设计情境下具有更高的样本效率。

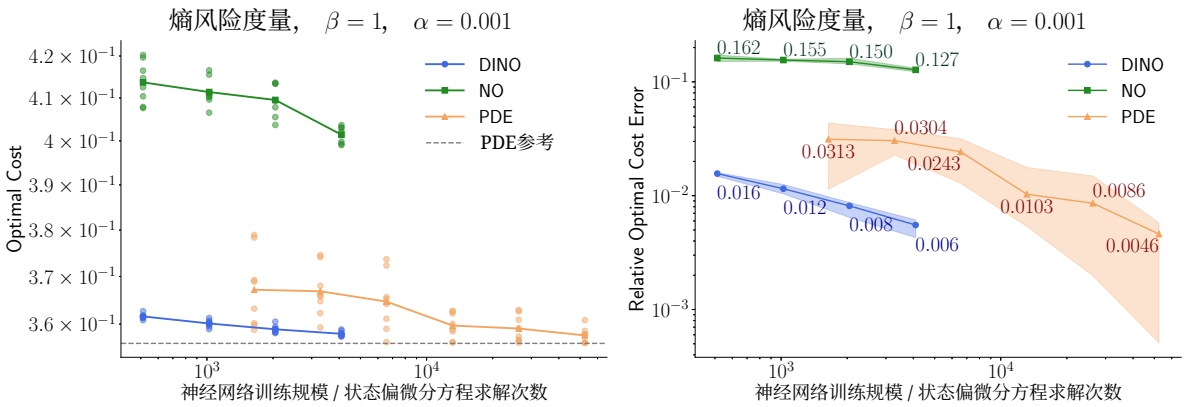


图13：最优解处的最优成本值（左）以及目标函数的相对误差（右）随优化所需的状态偏微分方程求解次数的变化，二者均采用熵风险度量的风险规避参数 $\beta = 1$ 和惩罚权重 $\alpha = 0.001$。神经网络训练规模分别为512、1,024、2,048和4,096，且不确定性下的优化均在每次迭代中使用8,192个样本。偏微分方程不确定性下优化每次迭代分别使用16、32、64、128、512个样本。偏微分方程参考不确定性下优化每次迭代使用8,192个样本。结果均使用相同的16,384个样本进行评估。请注意，上述计数是指每个流动场景的样本数。由于我们对东向流动和北向流动场景均评估感兴趣量，实际偏微分方程求解次数为所报告数量的两倍。

5.5. 多种不确定来流条件下三维减阻设计

现在，我们将第节讨论的二维问题扩展到一个三维示例。也就是说，我们考虑通道 $\Omega_{\text{channel}} = (0,\ 2) \times (0,\ 1) \times (0,\ 1) \subset \mathbb{R}^3$ 中流经塔状物体 Ω_{object}的流动，该物体的几何形状将通过自由形变进行设计。在这种情况下，通道的边界 $\partial\Omega_{\text{channel}}$ 可分解为六个平面，

$$\begin{aligned}
\Gamma_{\text{left}} &= \{(x_1,x_2,x_3) \in \partial\Omega_{\text{channel}} : x_1 = 0\}, \quad \Gamma_{\text{right}} = \{(x_1,x_2,x_3) \in \partial\Omega_{\text{channel}} : x_1 = 2\},\\
\Gamma_{\text{front}} &= \{(x_1,x_2,x_3) \in \partial\Omega_{\text{channel}} : x_3 = 0\}, \quad \Gamma_{\text{back}} = \{(x_1,x_2,x_3) \in \partial\Omega_{\text{channel}} : x_3 = 1\},\\
\Gamma_{\text{base}} &= \{(x_1,x_2,x_3) \in \partial\Omega_{\text{channel}} : x_2 = 0\}, \quad \Gamma_{\text{top}} = \{(x_1,x_2,x_3) \in \partial\Omega_{\text{channel}} : x_2 = 1\}.
\end{aligned}$$

我们再次考虑两个流动场景，分别对应沿东向和北向的来流。我们按照第 5.4 节中的相同方式进行建模，采用物体几何形状的两种不同方向，即东向流动中的 $\Omega^{\text{East}}_{\text{object}} = \chi_z(\Omega_{\text{object}})$ 和北向流动中的 $\Omega^{\text{North}}_{\text{object}} = \text{Reflect}(\Omega^{\text{East}}_{\text{object}})$，其中 Reflect(·) 现在表示关于 $X_1 = X_3$ 平面的反射。空间区域于是定义为 $\Omega^s_z := \Omega_{\text{channel}} \setminus \text{cl}\left(\Omega^s_{\text{object}}\right)$。随后，通过求解带有以下边界条件的式 (78a)，可得到每种流动场景 $u^s = (v^s, p^s)$ $s \in$ {东、北}对应的解

$$\begin{aligned}
\Gamma_{\text{in}} &= \Gamma_{\text{left}}, \qquad \Gamma_{\text{no-slip}} = \Gamma_{\text{base}} \cup \Gamma^s_{\text{object}},\\
\Gamma_{\text{out}} &= \Gamma_{\text{right}}, \quad \Gamma_{\text{free-slip}} = \Gamma_{\text{front}} \cup \Gamma_{\text{top}} \cup \Gamma_{\text{back}}.
\end{aligned}$$

也就是说，左侧表面施加来流条件，右侧表面施加自由流条件，底面（地面） $x_2 = 0$ 和物体表面施加无滑移条件，其余表面施加自由滑移条件。具体而言，入口流动剖面由模型参数定义，该参数对指数等于 $1/5$ 的幂律剖面进行缩放，

$$(v^s_z)_{\text{in}} = \exp(\widetilde{m}^s(x)) x_2^{1/5} e_1, \qquad e_1 = (1,0,0)^T. \tag{94}$$

这种特定的幂律常用于模拟大气环境中的边界层现象 [93]。我们再次将 $\widetilde{m}^s$ 视为定义在参考域上的高斯随机场，即

$\widetilde{m}^s = T_{\chi_z}^{-1} m^s$，$m^s \sim \mathcal{N}\ 0$

(， $\mathcal{C}_m^s$)，其中协方差算子 $\mathcal{C}_m^s$ 仍为双拉普拉斯算子，其相关长度和点方差的选取与二维情形相匹配。

图 14 展示了计算域和初始物体几何形状（一个截锥形塔），以及由偏微分方程解得到的速度场流线。该构型作为后续数值实验的参考几何形状。

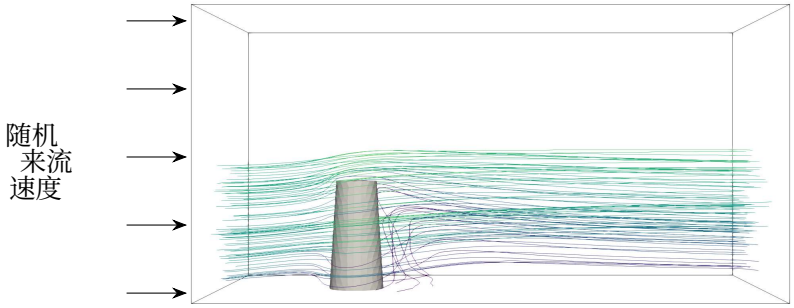


图14：计算域的几何形状和初始截锥形塔，以及由偏微分方程纳维–斯托克斯方程解得到的速度场流线。所示情况对应于左侧入口条件，即来流位于 Γ_{left}，流出位于 Γ_{right}。其余面 $\Gamma_{\text{free-slip}} = \Gamma_{\text{front}} \cup \Gamma_{\text{top}} \cup \Gamma_{\text{back}}$ 作为不透流壁面，而 $\Gamma_{\text{object}} \cup \Gamma_{\text{base}}$上施加无滑移条件。

形状参数化改编自上文所述的二维自由变形框架（79），并沿三个轴采用伯恩斯坦多项式基。具体而言，伯恩斯坦控制点位于 $6\times6\times6$ 单位立方体内部的一致晶格上 $[0,1]^3$,，其中各点沿每个轴位于 $\{0, 0.2, 0.4, 0.6, 0.8, 1\}$。与二维问题类似，我们排除 X_1 和 X_3轴上的第一个和最后一个控制点，以及 X_2 轴上的第一个点，从而使左侧、底面、前侧和后侧边界保持不动。此外，我们仅允许控制点在水平面 $X_1 - X_3$内移动，因此物体高度不会改变。最终，我们得到 $d_z = 160$ 个形状自由度；这些自由度来自总计80个控制点，每个控制点贡献2 个自由度。

为生成北向流动场景的几何变化，我们通过交换参数坐标构造镜像位移基

$$(X_1, X_2, X_3) \mapsto (X_3, X_2, X_1), \qquad (e_1, e_2, e_3) \mapsto (e_3, e_2, e_1)$$

从而使位移场关于平面 $X_1 = X_3$ 反射。

随后，按照 (86) 在参考域中构建偏微分方程。我们使用四面体网格对区域进行离散化，状态采用泰勒–胡德单元，参数采用二次单元，得到 $d_u = 393$，331 和 $d_{\mathcal{M}} = 3$，593。

5.5.1. 问题构建

在这个问题中，我们考虑两个不同的优化感兴趣量，它们都刻画了物体所受阻力的概念。具体而言，分别是：（i）物体所受的总水平力；（ii）物体相对于其底面中心的总弯矩。对于每个由 $s \in \{$东、北$\}$ 索引的入口流动情景，分别定义这些感兴趣量。对于每种情况，令 $u^s = (v^s$、$p^s)$ 表示在空间坐标下稳态不可压缩纳维–斯托克斯方程的速度–压力解。作用于塔体边界 Γ_{object} 且其外法向单位向量为 n 的力和弯矩定义为

$$\mathbf{F}_z^s(\widetilde{u}^s, \widetilde{u}^s) = \int_{\Gamma_{\text{object}}^s} \widetilde{\mathbf{T}}(\widetilde{u}^s)\, n\, ds, \tag{95}$$

$$\mathbf{M}_z^s(\widetilde{u}^s, \widetilde{m}^s) = \int_{\Gamma_{\text{object}}^s} (x - O) \times \left(\widetilde{\mathbf{T}}(\widetilde{u}^s) n\right)\, ds, \tag{96}$$

其中，$x = (x_1, x_2, x_3)^T$为空间坐标，$O = (0.5, 0, 0.5)^T$为结构底面中心，$\times$ 表示叉积。我们令 $\mathbf{F}^s(u^s, m^s, z)$ 和 $\mathbf{M}^s(u^s, m^s, z)$ 分别表示当输入坐标位于参考域中时的对应形式（即，力向量和力矩向量仍以空间坐标表示）。记 $\mathbf{F}^s = (F_1^s, F_2^s, F_3^s)$ 和 $\mathbf{M}^s = (M_1^s, M_2^s, M_3^s)^T$，我们将考虑的两个标量感兴趣量为总水平力

$$Q^s_{\text{force}}(u^s, m^s, z) := (F_1^s)^2 + (F_2^s)^2, \tag{97}$$

其中省略了竖直分量 F_2^s，总弯矩

$$Q^s_{\text{bend}}(u^s, m^s, z) := \|\mathbf{M}^s(u^s, m^s, z)\|_2^2 = (M_1^s)^2 + (M_2^s)^2 + (M_3^s)^2. \tag{98}$$

除了优化感兴趣量外，我们还加入惩罚项，以正则化设计并确保物理一致性。同样，我们考虑两种惩罚：(i) 对形状控制系数的惩罚（ ℓ_2 （73），以及 (ii) 重力力矩惩罚。对于重力力矩，我们假设结构本身具有均匀实体密度 $\rho_{\text{object}} = 2,\ 000$。其底面中心处的力矩为

$$\mathbf{M}^g(z) = \int_{\Omega_{\text{object}}} (x - O) \times \rho_{\text{object}} g\ e_2\ dx, \tag{99}$$

其中 $g = 9.81$ 是重力加速度常数，$e_2 = (0,\ -1,\ 0)^T$ 是竖直方向的单位向量。与上述积分类似，我们可以在参考域中构造该重力力矩积分。然后取

$$\mathcal{P}_{\text{grav}}(z; \alpha) := \frac{\alpha}{2} \|\mathbf{M}^g(z)\|_2^2. \tag{100}$$

不确定性下设计的目标函数由两个流动场景下的熵风险度量加权和适当惩罚项构成。

对于水平力减小，我们采用如下目标函数

$$\mathcal{J}_{\text{force}}(z) = 0.75\mathcal{J}^{\text{East}}_{\text{force}}(z) + 0.25\mathcal{J}^{\text{North}}_{\text{force}}(z) + \mathcal{P}_{\ell_2}(z; \alpha) \tag{101}$$

其中，对于每个流动场景 $s \in \{\text{东，北}\}$，我们有

$$\mathcal{J}^s_{\text{force}}(z) = \rho^{\text{entropic},\beta}_{m^s \sim \mu^s_m}(Q^s(u^s, m^s, z)). \tag{102}$$

此处，惩罚项$\mathcal{P}_{\ell_2}$ 是 ℓ_2 在（73）中使用的惩罚项，其中取 $\alpha = 2 \times 10^{-5}$。

对于弯矩减小，我们采用目标函数

$$\mathcal{J}_{\text{bend}}(z) = 0.75\mathcal{J}^{\text{East}}_{\text{bend}}(z) + 0.25\mathcal{J}^{\text{North}}_{\text{bend}}(z) + \mathcal{P}_{\text{grav}}(z; \alpha) \tag{103}$$

其中 $\alpha = 1$，对于每个流动场景 $s \in \{\text{东，北}\}$，我们有

$$\mathcal{J}^s_{\text{bend}}(z) := \rho^{\text{entropic},\beta}_{m^s \sim \mu^s_m}(Q^s_{\text{bend}}(u^s, m^s, z)). \tag{104}$$

这一公式可以理解为最小化流动诱导力矩与重力诱导力矩之和。

5.5.2. 神经算子构造

神经算子的训练数据通过从 m^s和 z的输入分布中采样生成，其中形状变量采用均匀分布 $\mu_z = \text{Unif}([-0.5, 0.5]^{d_z})$。由于在三维设置下求解偏微分方程的成本高得多，我们使用规模为 384 的小型训练数据集训练神经网络，并分别采用有雅可比信息和无雅可比信息两种设置。对于每种设置，为生成约化基，我们从训练集中选取 $n_{\text{POD}} = n_{\text{AS}} = 256$ 个样本，同时计算POD和主动子空间（AS），并分别取 $r_u = 200$ 个POD基向量和 $r_m = 128$ 个AS基向量。与前两个示例不同，由于数据集规模有限，我们对每种设置仅进行一次训练，而非多次独立运行

。我们使用具有 2 个隐藏层、每层宽度为 512 的多输入全连接神经网络来表示潜在空间映射。随后使用 Adam 训练神经网络 2,000 个训练轮次，初始学习率为 1×10^{-4}，并在训练 1,000 和 1,500 个轮次后减半。在东向流动场景中，在规模为 128 的测试数据集上，全阶 u的测试准确率分别为：Shape-DINO 为 7.22%，Shape-NO 为 18.91%。在北向流动场景中，相应的测试准确率分别为：Shape-DINO 为 4.13%，Shape-NO 为 12.55%。

5.5.3. 使用神经算子的形状优化

随后，我们部署已训练的神经算子来求解力和弯矩最小化问题（无需重新训练）。由于与三维纳维–斯托克斯偏微分方程相关的计算成本在很大程度上使得在缺乏大量高性能计算资源时无法进行基于偏微分方程的风险规避优化，我们转而将基于代理模型的解与由确定性基于偏微分方程的优化 $z^{\dagger}$ 偏微分方程 在 $m^s = \bar{m}^s$ 的标称值下计算得到的最优设计进行比较。这对应于OUU目标函数的单点近似，可以以更为适中的计算成本获得。此外，我们随后通过检视在输入 m^s的200个样本上计算得到的最优设计处所产生的感兴趣量分布，来评估基于偏微分方程的解与基于代理模型的解的质量。

所得分布如图 15 和 16 所示，分别对应力和弯矩目标函数。从所得分布图中，我们观察到，对于两个感兴趣量，由Shape-DINO最优设计诱导的分布比由确定性偏微分方程解得到的分布更为集中，且更接近零，并且显著比Shape-NO产生的分布更接近零。这一观察表明，Shape-DINO能够给出在多项性能指标上同时具有稳健性的设计，从而降低不确定性下响应的幅度和变异性。相比之下，Shape-NO的解对两个感兴趣量都呈现出更宽的分布和较重的尾部，这强调了在代理模型训练过程中纳入导数信息的重要性。

我们进一步指出，基于确定性偏微分方程的优化是在不确定参数 m^s 取其均值时，使用单次实现进行的，因此未考虑由此引起的变异
受不确定性的影响。因此，尽管确定性解在 $\bar{m}^s$ 处可能表现良好，但无法保证其对来流分布具有鲁棒性。另一方面，基于代理的模型能够高效地大批量评估样本平均近似目标；具体而言，Shape-DINO 可提供准确的目标梯度，而使用高保真偏微分方程求解来计算这些梯度的成本将高得难以承受。

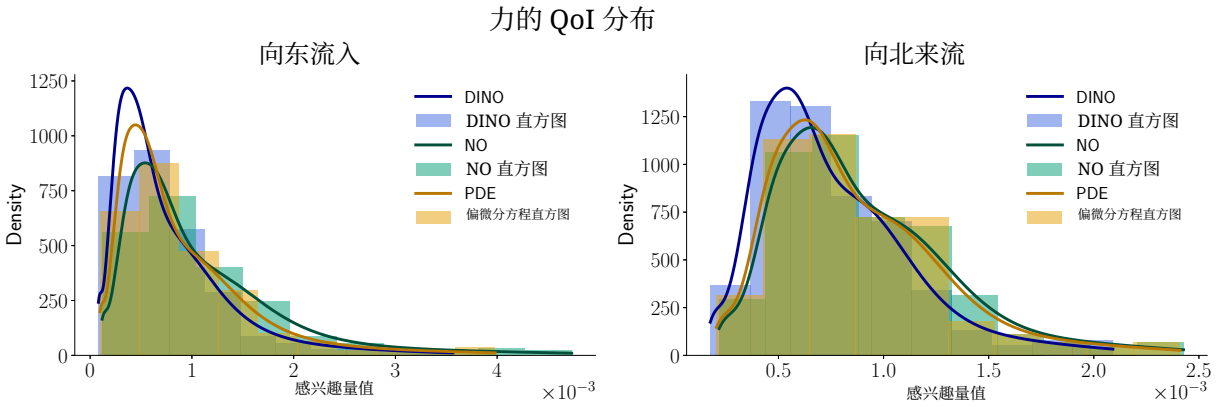


图15：最优设计 $z^{\dagger}$ DINO、 $z^{\dagger}$ 神经算子 和 $z^{\dagger}$ 偏微分方程 对应的水平力 QoI 分布。所有设计均使用 200 个蒙特卡洛样本进行评估。DINO 和神经算子解采用每次迭代 1,024 个样本获得，而偏微分方程解则在 $\bar{m}^s$ 处进行确定性计算。Shape-DINO 设计所得分布更集中，也更接近零；相比之下，Shape-NO 呈现出更宽的分布和较重的尾部，凸显了纳入导数信息对于在不确定性下实现鲁棒设计的益处。

我们展示了对应的最优塔架设计 $z^{\dagger}$，它们针对图 17中的两种感兴趣量分别获得。为展示由此产生的流动行为，我们在平均参数配置下可视化了流线——

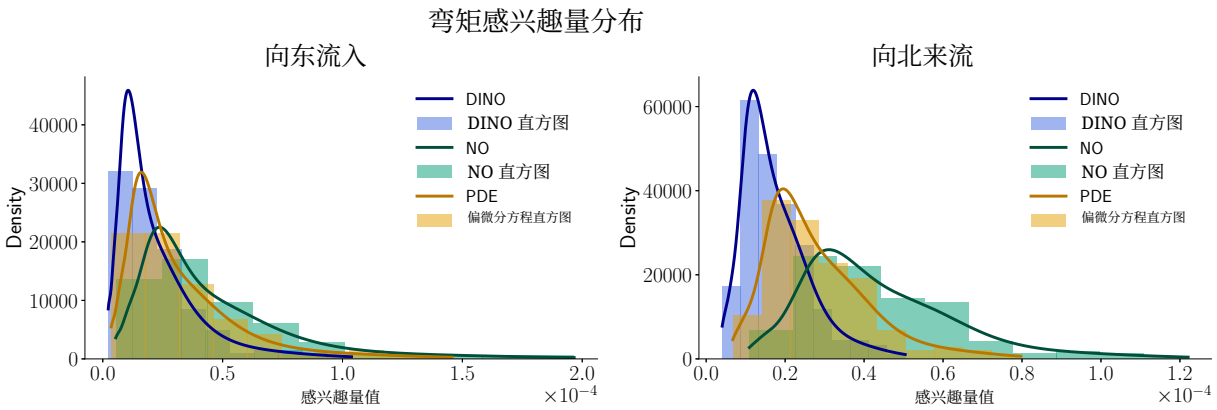


图16：最优设计 $z^{\dagger}_{\text{DINO}}$、$z^{\dagger}_{\text{NO}}$ 和 $z^{\dagger}_{\text{偏微分方程}}$ 的弯矩感兴趣量分布。所有设计均使用 200 个蒙特卡洛样本进行评估。DINO 和神经算子解采用每次迭代 1,024 个样本得到，而偏微分方程解则在 $\bar{m}^s$ 处以确定性方式计算。与确定性偏微分方程和 Shape-NO 设计相比，Shape-DINO 的分布集中得多且更接近零，表明其弯矩感兴趣量的幅值和变异性均有所降低。

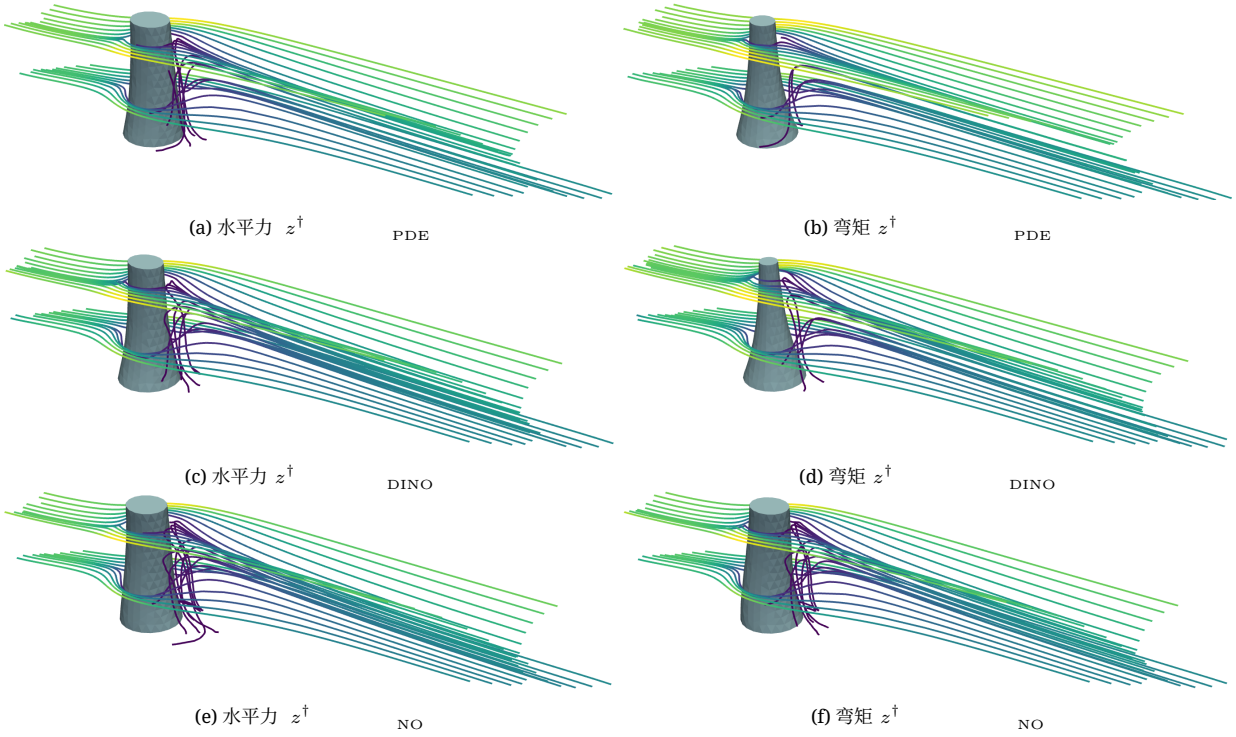
uration $\bar{m}^s$。与上述讨论的感兴趣量分布结果一致，使用 Shape-DINO 得到的最优设计比确定性偏微分方程对应的设计更纤细，因此流动可以顺畅地绕过塔架，受到的阻碍较小。相比之下，Shape-NO 得到的设计更为粗壮，甚至呈现扭曲的几何形状，从而加剧流动偏转，表现为塔架两侧的流线更加密集，下游尾流区也更大。这些差异表明，Shape-DINO 在生成能够有效减小水平力和弯矩的最优设计方面更具优势，也从物理角度解释了此前讨论的 Shape-NO 感兴趣量分布为何范围更广、尾部更重。

5.6. 时间比较

本节报告使用偏微分方程求解器或神经算子近似来评估状态及后续向量-雅可比积的计算成本。执行时间使用适合设备的计时器测量：中央处理器计时以 Python 的 time.perf_counter 测量墙钟经过时间，而图形处理器计时则使用 CUDA事件测量设备执行时间。由于状态和向量-雅可比积评估在训练数据生成、神经算子训练，以及不确定性下的优化中目标函数及其梯度的评估所需的计算成本中占据了大部分，因此我们利用这些计时结果来为前几节中的样本效率比较提供背景，并突出 Shape-DINO方法所实现的在线加速效果。

请注意，对于基于偏微分方程的方法，状态评估对应于求解状态偏微分方程；当状态偏微分方程为非线性时，这可能需要依次求解一系列线性方程组，而向量-雅可比积则需要求解一个对应于伴随偏微分方程的线性方程组。因此，我们报告求解状态偏微分方程和伴随偏微分方程的单核计时，并使用直接求解器 MUMPS [94]作为底层线性求解器。对于基于代理模型的方法，我们只报告神经算子前向（状态）和反向（向量-雅可比积）传播的计时；同时给出单样本评估和批量（1000 个样本）评估的单样本计时。由于 Shape-NO 和 Shape-DINO 采用相同的网络架构，因此这些结果合并报告为一个时间。泊松方程、二维和三维 Navier–Stokes 示例的结果见表 3。

我们首先关注基于偏微分方程的计时。在泊松方程示例中，状态偏微分方程是线性的，因此状态和伴随算子的求解成本确实相当。事实上，使用直接求解器时，可以通过重用状态偏微分方程矩阵分解的转置算子，以几乎可以忽略的成本完成额外的伴随算子求解。另一方面，在纳维–斯托克斯示例中，由于状态偏微分方程求解需要进行牛顿迭代，伴随算子的求解成本平均为状态偏微分方程求解成本的 0。2–0.25× 。此外，由于直接求解器能够以计算单个向量-雅可比积的成本计算多个向量-雅可比积，



(a) 水平力 $z^\dagger$ PDE

(b) 弯矩 $z^\dagger$ PDE

(c) 水平力 $z^\dagger$ DINO

(d) 弯矩 $z^\dagger$ DINO

(e) 水平力 $z^\dagger$ NO

(f) 弯矩 $z^\dagger$ NO

图 17 $z^\dagger$ $m^{\text{East}} = \bar{m}^{\text{East}}$ 17 17 17 17 17 17 东向流动下、在参数均值处评估流线得到的最优塔形状。面板 a、c 和 e 对应水平力感兴趣量，而面板 b、d 和 f 展示了弯矩感兴趣量的结果。与确定性偏微分方程相比，Shape-DINO 设计更纤细，且流动通过更为顺畅；而 Shape-NO 得到的几何形状更粗壮，有时还会扭曲，从而导致更强的流动偏转和更大的下游回流区。这一比较从物理角度解释了 Shape-DINO 实现的力和弯矩为何更小。

使用直接求解器生成 DINO 训练所需额外雅可比数据的成本，确实只是状态数据成本的一小部分（在线性泊松情形下可忽略不计，在纳维–斯托克斯情形下为 0.2–0.25× ）。因此，与前几节中观察到的样本效率提升数个数量级相比，基于导数的训练带来的收益远远超过了每个样本这笔少量的额外成本。

就优化过程中的纯在线成本而言，使用神经算子评估感兴趣量（m， z） $\mapsto Q$（u（m， z）m， z）及其关于形状变量 z 的梯度时，相比偏微分方程求解器可大幅提速。对于状态求解，单样本神经算子评估在泊松问题中快两个数量级以上，在二维纳维–斯托克斯问题中快四个数量级以上。对于批量评估，泊松问题和二维纳维–斯托克斯问题的加速比分别超过五个和六个数量级。值得特别注意的是，批量执行是单 GPU 系统的原生功能。相比之下，在大规模同步环境中，若要让基于 CPU 的偏微分方程求解达到相当的吞吐量，通常需要额外的硬件；而神经算子评估通常不需要。用于评估感兴趣量z 梯度的向量–雅可比积／伴随求解也呈现出类似优势。在单样本情形下，神经算子在泊松问题中的加速比接近四个数量级，在二维纳维–斯托克斯问题中则为三个数量级。在批量情形下，两类问题的加速比均超过四个数量级。

在三维纳维–斯托克斯环境中，这一优势更加显著：相较于二维，离散化维度更高，且三维稀疏分解的复杂性更高，因此高保真偏微分方程求解的计算成本大幅增加。对于该问题，神经算子的状态求解加速比分别在单样本和批量评估中超过六个和八个数量级。向量–雅可比积的加速比同样分别在单样本和批量评估中超过五个和七个数量级。这些结果表明，用神经算子代理模型替代重复的偏微分方程求解能够显著节省计算成本，使大规模优化问题变得可行；若仅使用偏微分方程求解器，其成本可能高得令人难以承受。

每个样本的执行时间（秒）						
	状态求解			向量–雅可比积／伴随		
问题	PDE	神经网络（1）	神经网络 (1000)	PDE	神经网络（1）	神经网络 (1000)
泊松方程	1.06e−1	4.12e−4	5.50e−7	8.84e−2	1.25e−5	1.35e−6
二维纳维-斯托克斯	5.08e+0	3.86e−4	1.70e−6	1.07e+1	1.15e−4	3.16e−6
三维 纳维–斯托克斯	7.18e+2	3.85e−4	5.46e−6	1.62e+2	9.68e−4	1.07e−5

表3：使用神经算子和基于偏微分方程的求解器求解状态偏微分方程并计算向量–雅可比积（向量-雅可比积）的时间（以秒为单位）。基于偏微分方程的求解器计时是对（m, z）的100个独立随机样本的平均值。对于神经算子，神经网络（1）表示在单样本上执行的推断，而神经网络（1000）对应于对1000个样本同时评估的批量推理，时间按每个样本归一化。神经网络（1）和神经网络（1000）所报告的时间均为100次独立试验的平均值。对于神经算子，我们报告通过反向传播计算向量-雅可比积的时间；而对于基于偏微分方程的方法，我们报告求解伴随偏微分方程所花费的时间，这是向量-雅可比积的主要成本。与基于偏微分方程的求解器相比，神经算子在状态和向量-雅可比积评估方面均实现了数量级的计算时间缩减，在二维和三维纳维–斯托克斯情形中增益尤为显著。这些增益在批量推理（神经网络（1000））下进一步放大，其中原生图形处理器并行评估带来了额外的每样本执行时间缩减。

6. 结论

在本研究中，我们提出了一种新颖的代理模型框架 Shape-DINO，用于加速偏微分方程约束下的形状不确定性下优化问题。该框架通过微分同胚映射将几何变异性表示到固定的参考域上，并采用结合导数信息引导的训练的降阶基神经算子架构，从而联合学习解映射及其相对于随机参数和设计变量的导数。该框架直接应对基于风险的形状优化中面临的几何变异性、采样复杂度、快速在线评估和梯度精度等相互关联的挑战。

除方法上的贡献外，据我们所知，本研究首次对用于不确定性下优化的 RBNO 代理模型进行了端到端的理论分析。我们证明了多输入3.1RBNO 的通用逼近性质（定理 3.2），并在风险中性设定下建立了其诱导的优化误差界，为基于代理的优化提供了严格的理论依据。

通过一系列数值实验，我们证明导数信息神经算子能够高效求解一系列形状不确定性下优化问题。在所有情形中，训练期间纳入雅可比矩阵信息显著提升了状态预测精度、梯度保真度，以及最重要的是，得到了可靠的最优设计。这些观察与第2节推导的先验误差界一致，该误差界表明优化误差直接受解及其形状导数中的误差控制。因此，与非导数神经算子相比，Shape-DINO始终产生更稳健的解和更可靠的优化行为。此外，约化基架构实现了快速推断，而基于代理的样本平均近似允许大批量采样，若仅使用偏微分方程求解则在计算上不可行。

我们的实验进一步表明，与基于PDE的方法相比，Shape-DINO实现了显著更优的成本–精度权衡。在相当的精度水平下，Shape-DINO所需的偏微分方程求解次数比直接基于SAA的PDE优化少几个数量级，同时在训练完成后可实现快速在线评估。这一性能与第3节建立的近似结果一致，该结果从理论上保证约化基多参数神经算子能够在适当的 C^1 范数下任意好地同时近似解及其导数。在实践中，这使得离线训练成本可以在涉及不同目标函数、风险度量或正则化选择的多次优化运行中摊销，从而支持重复的设计查询、交互式设计空间探索以及近实时的不确定性下的形状优化。

此外，我们证明了所提出的框架能够扩展到复杂的三维问题，而在这类问题中，完全基于偏微分方程的优化在计算上难以承受。这些结果凸显了 Shape-DINO 在涉及高维几何、不确定性和复杂物理模型的新型应用中的潜力。

我们也承认该框架存在局限性。当优化目标相对简单，即只需少量偏微分方程求解即可准确近似时，我们的框架可能过于复杂。在这种情况下，经典的基于偏微分方程的优化可能更为合适。当几何变异性、不确定性和采样复杂度使传统方法在计算上难以实现时，Shape-DINO 则更具优势。此外，我们的框架依赖于微分同胚网格变形和弹性延拓，这些方法适用于中等程度的几何变化。然而，所选的形状参数化和训练分布可能无法充分覆盖大幅或极端变形。如果优化探索的形状远超这一范围，就可能发生网格畸变，并影响偏微分方程求解的精度和代理模型训练。

最后，本研究结果表明，未来有若干值得探索的方向，包括纳入高阶导数信息、自适应采样策略（由优化进展引导），以及应用于更复杂的几何参数化。此外，代理模型不必仅作为评估代价泛函时偏微分方程求解的替代方案；它们还可以作为控制变量，与少量基于偏微分方程的评估结合使用，从而在方差缩减框架 [95, 96] 内加速风险度量的估计。这为介于基于神经算子和基于偏微分方程的代价泛函评估之间的方法开辟了道路，也为研究 Shape-DINO 训练期间与基于控制变量的优化期间偏微分方程求解次数之间的最优权衡提供了契机。

更广泛地说，本研究表明，将形状敏感性和参数敏感性纳入神经算子训练，对于实现可靠且可扩展的不确定性下的优化至关重要，这使 Shape-DINO 成为复杂偏微分方程约束设计优化工作流程中的强大工具。除了基于风险的形状优化之外，我们还设想，所提出的形状导数引导的算子学习框架可应用于广泛场景，包括用于几何与形状重建的基于偏微分方程的逆问题、几何不确定性下的贝叶斯反演，以及近实时参数化确定性形状优化。这凸显了 Shape-DINO 作为通用计算框架，在复杂工程系统的设计、推断和决策方面的潜力。

7. 致谢

本研究部分由美国能源部资助（项目编号 DE-SC0023171）、美国空军科学研究局资助（项目编号 FA9550-24-1-0327）、美国国家科学基金会资助（项目编号 DMS-2324643 和 OAC-2313033），以及新西兰皇家学会 Te Ap¯arangi（马斯登基金委员会）资助（项目编号 MFP-24-UOA-279）。作者感谢德克萨斯高级计算中心（TACC）和德克萨斯大学奥斯汀分校提供计算资源，这些资源为本文所报告的研究成果做出了贡献。

参考文献

[1] A. Shapiro、D. Dentcheva、A. Ruszczynski，《随机规划讲义：建模与理论》，工业与应用数学学会，2021年。[2] R. T. Rockafellar、S. Uryasev 等，《条件风险价值的优化》，《风险杂志》2（2000）21–42。[3] D. P. Kouri、T. M. Surowiec，《利用条件风险价值进行风险厌恶的偏微分方程约束优化》，《SIAM优化杂志》26（2016）365–396。[4] S. Conti、H. Held、M. Pach、M. Rumpf、R. Schultz，《风险厌恶形状优化》，《SIAM控制与优化期刊》49（2011）927–947。[5] S. Conti、M. Rumpf、R. Schultz、S. Tölkes，《弹性形状优化中的随机占优约束》，《SIAM控制与优化期刊》56（2018）3021–3034。[6] B. Geihe、M. Lenz、M. Rumpf、R. Schultz，《具有参数化精细尺度几何形状的风险厌恶弹性形状优化》，《数学规划》141（2012）383–403。[7] C. Schillings、S. Schmidt、V. Schulz，《确定性与不确定性气动设计的高效形状优化》，《计算机与流体》46（2011）78–87。[8] N. Kovachki、Z. Li、B. Liu、K. Azizzadenesheli、K. Bhattacharya、A. Stuart、A. Anandkumar，《神经算子：学习函数空间之间的映射及其在偏微分方程中的应用》，《机器学习研究期刊》24（2023）1–97。[9] D. Luo、T. O’Leary-Roseberry、P. Chen、O. Ghattas，《利用导数信息神经算子，在高维不确定性下进行高效偏微分方程约束优化》，《SIAM科学计算杂志》47（2025）C899–C931。[10] J. Samareh，《基于自由形变的气动形状优化》，载于：第10届AIAA/ISSMO多学科分析与优化会议，美国航空航天学会，2004年。[11] G. Rozza、A. Koshakji、A. Quarteroni 等，《应用于三维形状优化问题的自由形变技术》，《应用与工业数学通讯》4（2013）1–26。[12] T. Bui-Thanh、O. Ghattas，《反散射问题的海森矩阵分析。第一部分：声波的逆形状散射》，《反问题》28（2012）055001。[13] V. Akçelik、G. Biros、O. Ghattas、D. Keyes、K. Ko、L.-Q. Lee、E. G. Ng，《国际直线对撞机低损耗腔体电磁形状优化的伴随方法》，《物理学杂志：会议系列》16（2005）435–445。[14] F. Salmoiraghi、A. Scardigli、H. Telib、G. Rozza，《自由形变、网格变形与降阶方法：实现高效气动形状优化的工具》，2018年。

[15] X. He、J. Li、C. A. Mader、A. Yildirim、J. R. Martins，《鲁棒气动形状优化——从圆形到翼型》，《航空航天科学与技术》87（2019）48–61。[16] M. C. Delfour、J.-P. Zolésio，《形状与几何：度量、分析、微分学与优化》，工业与应用数学学会，2011年。[17] D. Henry，《偏微分方程边值问题中的边界扰动》，第318卷，剑桥大学出版社，2005年。[18] J. Sokolowski、J.-P. Zolésio，《形状优化导论》，施普林格，1992年。[19] J. Haslinger、R. A. Mäkinen，《形状优化导论：理论、近似与计算》，工业与应用数学学会，2003年。[20] A. Henrot、M. Pierre，《形状变化与优化》，第28卷，欧洲数学学会出版社，苏黎世，瑞士，2018年。[21] O. Pironneau，《椭圆系统的最优形状设计》，施普林格，1984年。[22] S. Yang、G. Stadler、R. Moser、O. Ghattas，《基于形状海森矩阵的纳维–斯托克斯流边界粗糙度分析》，《SIAM应用数学杂志》71（2011）333–355。[23] R. Hiptmair、A. Paganini，《通过追踪微分同胚进行形状优化》，《应用数学计算方法》15（2015）291–305。[24] A. Paganini、F. Wechsung、P. E. Farrell，《偏微分方程约束形状优化的高阶动网格方法》，《SIAM科学计算杂志》40（2018）A2356–A2382。[25] V. H. Schulz、M. Siebenborn、K. Welker，《基于斯捷克洛夫–庞加莱型度量的高效偏微分方程约束形状优化》，《SIAM优化杂志》26（2016）2800–2819。[26] C. Geiersbach、E. Loayza-Romero、K. Welker，《偏微分方程约束形状优化：迈向乘积形状空间与随机模型》，载于：《计算机视觉与成像中的数学模型与算法手册：数学成像与视觉》，施普林格，2023年，第1585–1630页。[27] S. Onyshkevych、M. Siebenborn，《利用非线性延拓算子保持网格质量的形状优化》，《优化理论与应用杂志》189（2021）291–316。[28] O. Ghattas、X. Li，《非线性气动弹性问题灵敏度分析的区域分解方法》，《国际计算流体动力学杂志》11（1998）113–130。[29] M. F. Beg、M. I. Miller、A. Trouvé、L. Younes，《通过微分同胚的测地流计算大变形度量映射》，《国际计算机视觉杂志》61（2005）139–157。[30] M. I. Miller、A. Trouvé、L. Younes，《论计算解剖学中的度量与欧拉–拉格朗日方程》，《生物医学工程年度评论》4（2002）375–405。[31] A. Trouvé，《图像分析中的微分同胚群与模式匹配》，《国际计算机视觉杂志》28（1998）213–221。[32] A. Mang、G. Biros，《用于约束微分同胚图像配准的非精确牛顿–克雷洛夫算法》，《SIAM成像科学杂志》8（2015）1030–1069。[33] A. Mang、G. Biros，《用于约束微分同胚图像配准的半拉格朗日两层预条件牛顿–克雷洛夫求解器》，《SIAM科学计算杂志》39（2017）B1064–B1101。[34] A. Mang、A. Gholami、C. Davatzikos、G. Biros，《CLAIRE：用于约束大变形微分同胚图像配准的分布式内存求解器》，《SIAM科学计算杂志》41（2019）C548–C584。

[35] A. Kodakkal, B. Keith, U. Khristenko, A. Apostolatos, K.-U. Bletzinger, B. Wohlmuth, R. Wüchner，不确定风况下高层建筑的风险规避设计，《计算机应用力学与工程方法》402（2022年）115371。[36] S. Conti, H. Held, M. Pach, M. Rumpf, R. Schultz，风险厌恶形状优化，《SIAM控制与优化期刊》49（2011年）927–947。[37] A. Kodakkal, B. Keith, U. Khristenko, A. Apostolatos, K.-U. Bletzinger, B. Wohlmuth, R. Wüchner，不确定风况下高层建筑的风险规避设计，《计算机应用力学与工程方法》402（2022年）115371。[38] M. Heinkenschloss, D. P. Kouri，不确定性下由偏微分方程组支配的优化问题，Acta Numerica 34（2025年）491–577。[39] S. Conti, H. Held, M. Pach, M. Rumpf, R. Schultz，风险厌恶形状优化，《SIAM控制与优化期刊》49（2011年）927–947。[40] M. Eigel, J. Neumann, R. Schneider, S. Wolf，风险厌恶随机结构拓扑优化，《计算机应用力学与工程方法》334（2018年）470–482。[41] J. Martínez-Frutos, D. Herrero-Pérez, M. Kessler, F. Periago，基于随机展开方法的随机场风险厌恶结构拓扑优化，《计算机应用力学与工程方法》330（2018年）180–206。[42] Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. Stuart, A. Anandkumar，用于参数化偏微分方程的傅里叶神经算子，arXiv预印本（2021年）。[43] Z. Li, D. Huang, A. Anandkumar, B. Liu, Li，利用学习得到的变形处理一般几何形状上的偏微分方程的傅里叶神经算子，《机器学习研究期刊》24（2023年）1–26。[44] Z. Zhao, C. Liu, Y. Li, Z. Chen, X. Liu，适用于偏微分方程不同区域和参数的微分同胚神经算子，《通讯物理》8（2025年）。[45] L. Lu, P. Jin, G. Pang, G. E. Karniadakis, DeepONet：基于算子通用逼近定理学习非线性算子以识别微分方程，《自然·机器智能》（2021年）。[46] P. Jin, S. Meng, L. Lu，MIONet：通过张量积学习多输入算子，《SIAM科学计算杂志》44（2022年）A3490–A3514。[47] M. Yin, N. Charon, R. Brody, L. Lu, N. Trayanova, M. Maggioni，一种可扩展框架，用于学习偏微分方程的几何相关解算子，《自然·计算科学》4（2024年）928–940。[48] K. Shukla, V. Oommen, A. Peyvan, M. Penwarden, N. Plewacki, L. Bravo, A. Ghoshal, R. M. Kirby, G. E. Karniadakis，深度神经算子作为形状优化的精确代理模型，《人工智能工程应用》129（2024年）107615。[49] Z. Cheng, Z. Hao, X. Wang, J. Huang, Y. Wu, X. Liu, Y. Zhao, S. Liu, H. Su，参考神经算子：学习解对偏微分方程几何变形的平滑依赖关系，PMLR（2024年）8060–8076。[50] C. Kou, Y. Yin, M. Zhu, S. Jia, Y. Luo, X. Yuan, L. Lu，通过主动学习进行神经拓扑优化，以高效设计湍流传质通道，《先进科学》（2025年）。[51] T. O’Leary-Roseberry, P. Chen, U. Villa, O. Ghattas，导数信息神经算子：一种高效的高维参数化导数学习框架，《计算物理学杂志》496（2024年）112555。

[52] L. Cao, J. Chen, M. Brennan, T. O'Leary-Roseberry, Y. Marzouk, O. Ghattas，LazyDINO：利用结构特性和代理模型驱动的测度输运，实现快速、可扩展且高效摊销的贝叶斯反演，arXiv预印本 arXiv:2411.12726（2024）。[53] L. Cao, T. O'Leary-Roseberry, O. Ghattas，导数信息神经算子加速无限维贝叶斯反问题中的几何MCMC，《机器学习研究期刊》26（2025年）1–68。[54] Y. Qiu, N. Bridges, P. Chen，导数增强型深度算子网络，《神经信息处理系统进展》37（2024年）20945–20981。[55] B. Yao, D. Luo, L. Cao, N. Kovachki, T. O'Leary-Roseberry, O. Ghattas，导数信息傅里叶神经算子：通用逼近及其在偏微分方程约束优化中的应用，arXiv预印本 arXiv:2512.14086（2025年）。[56] K. Bhattacharya, B. Hosseini, N. B. Kovachki, A. M. Stuart，参数化偏微分方程的模型降阶与神经网络，《SMAI计算数学杂志》7（2021年）121–157。[57] S. Lanthaler，使用PCA-net进行算子学习：复杂度上下界，《机器学习研究期刊》24（2023年）1–67。[58] L. Herrmann, C. Schwab, J. Zech，神经算子与谱算子代理模型：统一构造与表达速率界，《计算数学进展》50（2024年）。[59] B. Adcock, N. Dexter, S. Moraga，巴拿赫空间之间全纯算子的最优深度学习，载于：第38届神经信息处理系统大会论文集。[60] D. Luo, T. O'Leary-Roseberry, P. Chen, O. Ghattas，导数信息算子学习的降维：逼近误差分析，2025年。[61] A. Manzoni, A. Quarteroni, S. Salsa，《偏微分方程最优控制》，施普林格国际出版，2021年。[62] A. Chaudhuri, B. Kramer, M. Norton, J. O. Royset, K. Willcox，可认证的基于风险的工程设计优化，《美国航空航天学会期刊》60（2022年）551–565。[63] P. A. Guth, V. Kaarnioja, F. Y. Kuo, C. Schillings, I. H. Sloan，基于熵风险度量和准蒙特卡罗积分的不确定性下抛物型偏微分方程约束最优控制，Numerische Mathematik 156（2024年）565–608。[64] J. Hesthaven, S. Ubbiali，使用神经网络对非线性问题进行非侵入式降阶建模，《计算物理学杂志》363（2018年）55–78。[65] A. Quarteroni, A. Manzoni, F. Negri，偏微分方程约化基方法导论，第92卷，施普林格，2015年。[66] O. Zahm, P. G. Constantine, C. Prieur, Y. M. Marzouk，多元向量值函数的基于梯度的降维，《SIAM科学计算杂志》42（2020年）A534–A558。[67] T. O'Leary-Roseberry, U. Villa, P. Chen, O. Ghattas，针对偏微分方程支配的高维参数映射的导数信息投影神经网络，《计算机应用力学与工程方法》388（2022年）114199。[68] F. Ballarin, A. Manzoni, G. Rozza, S. Salsa，通过自由形变进行形状优化：斯托克斯流的存在性结果与数值解，《科学计算杂志》60（2013年）537–563。[69] R. Nicholson, M. Niskanen，泊松问题中罗宾系数与区域边界的联合估计，《反问题》38（2021年）015008。

[70] J. E. Castrillón-Candás, F. Nobile, R. F. Tempone，随机区域上偏微分方程的一种混合配置-扰动方法，《计算数学进展》47（2021年）。[71] T. W. Sederberg, S. R. Parry，实体几何模型的自由形变，计算机图形学与交互技术第13届年度会议论文集（1986）151–160。[72] R. M. Hicks, P. A. Henne，通过数值优化进行机翼设计，《航空器杂志》15（1978）407–412。[73] H.-Y. Wu, S. Yang, F. Liu, H.-M. Tsai，气动优化中翼型三种几何表示的比较，载于：第16届美国航空航天学会计算流体动力学会议，第4095页。[74] C. Sabater, S. Görtz，使用伴随方法和高斯过程进行基于梯度的气动鲁棒优化，载于：《工程与科学中的设计、优化与控制：进化与确定性方法进展》，施普林格，2020年，第211–226页。[75] G. Athanassoulis, T. Loukakis，扩展 Lewis 型船体横剖面族及其在耐波性计算中的应用，《国际造船进展》32（1985）33–43。[76] L. Radtke, G. Bletsos, N. Kühl, T. Suchan, T. Rung, A. Düster, K. Welker，无参数形状优化：面向工程应用的多种形状更新，《航空航天》10（2023年）751。[77] V. Schulz, M. Siebenborn，偏微分方程约束形状优化中曲面度量的计算比较，《应用数学计算方法》16（2016）485–496。[78] A. Shamanskiy, B. Simeon，流固耦合中的网格移动技术：稳健性、累积畸变与计算效率，《计算力学》67（2020）583–600。[79] K. Stein, T. Tezduyar, R. Benney，具有大位移的流固耦合网格移动技术，《应用力学杂志》70（2003）58–63。[80] L. T. Diosady, S. M. Murman，一种用于高阶时空网格变形的线弹性求解器，载于：2018年美国航空航天学会航空航天科学会议，第0919页。[81] Y. Saad，稀疏线性系统的迭代方法，工业与应用数学学会，2003年。[82] N. Halko, P.-G. Martinsson, J. A. Tropp，利用随机性发现结构：用于构造近似矩阵分解的概率算法，《SIAM评论》53（2011）217–288。[83] A. Logg, K.-A. Mardal, G. Wells，通过有限元法自动求解微分方程：FEniCS手册，第84卷，施普林格科学与商业媒体，2012年。[84] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai，PyTorch：一种命令式风格、高性能的深度学习库，2019年。[85] U. Villa, N. Petra, O. Ghattas，hIPPYlib：用于大规模偏微分方程反问题的可扩展软件框架：第一部分：确定性反演与线性化贝叶斯推断，《ACM数学软件汇刊》47（2021年）。[86] T. O’Leary-Roseberry, U. Villa，hippyflow：用于 Python 中参数化偏微分方程映射的降维代理模型构建，https://github.com/hippylib/hippyflow，2021年。Zenodo，https://doi.org/10.5281/zenodo.4608729。[87] D. Luo, P. Chen, T. O’Leary-Roseberry, U. Villa, O. Ghattas，SOUPy：Python 中高维不确定性下的随机偏微分方程约束优化，《开放源代码软件杂志》9（2024）6101–6101。

[88] D. P. Kingma, J. Ba，Adam：一种随机优化方法，2014 .

[89] F. Lindgren、H. Rue、J. Lindström，高斯场与高斯马尔可夫随机场之间的明确联系：随机偏微分方程方法，《皇家统计学会杂志：B辑（统计方法论）》73（2011年）423–498。

[90] U. Villa、T. O'Leary-Roseberry，关于基于偏微分方程的精度算子与Matérn协方差之间关系的注记，arXiv预印本 arXiv:2407.00471（2024）。

[91] A. Quarteroni、R. Sacco、F. Saleri，《数值数学》，施普林格，第3版，201 4.

[92] A. Manzoni、A. Quarteroni、G. Rozza，利用约化基方法和自由形变进行黏性流形状优化，《国际流体数值方法杂志》70（2011年）646–670。

[93] H. Schlichting、K. Gersten，《边界层理论》，施普林格，2016。

[94] P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, J. Koster，MUMPS：一种通用分布式内存稀疏求解器，载于：国际应用并行计算研讨会，施普林格，第 121–130 页。

[95] L. Ng, K. Willcox，不确定性下优化的多保真方法，《国际工程数值方法杂志》100 (2014) 746–772。

[96] P. Chen, U. Villa, O. Ghattas，不确定性下偏微分方程约束最优控制的泰勒近似与方差缩减，《计算物理学杂志》385 (2019年) 163–186。

[97] O. Gonzal ez, A. M. Stuart，《连续介质力学基础课程》，剑桥大学出版社，2001 .

[98] A. Pinkus，神经网络中多层感知器模型的逼近理论，Acta Numerica 8 (1999) 143–195。

A. 空间域与参考域变换

A.1. 常用坐标变换

回顾一下，在微分同胚 χ 下，我们可以通过变量替换在空间域坐标与参考域坐标之间进行转换

$$x = \chi(X), \qquad X = \chi^{-1}(x)$$

对于 $X \in \Omega_0$ 和 $x \in \chi(\Omega_0)$，微分量和积分量会相应地发生变换。在表 A.4 中，我们列出了数值算例中使用的常见量的坐标变换。更多细节可参见标准连续介质力学教材，例如 [97]。

A.2. 命题 *2.1* 的证明

证明。 我们首先推导推前算子的范数界。首先，回顾对于 $F_\chi(X) = \nabla_X \chi(X)$，我们有

$$|\det F_\chi(X)| \le \|F_\chi(X)\|^d_{\mathcal{L}(\mathbb{R}^d,\mathbb{R}^d)} \le \|\chi\|^d_{W^{1,\infty}(\Omega_0)}.$$

现在考虑任意的 $u \in L^2(\Omega_0)$ 及其前推 $T_\chi^{-1}u = u(\chi^{-1}(x))$。通过变量替换， $L^2(\Omega_1)$ 范数中的 $T_{\chi^{-1}}u$可写为

$$\begin{aligned}
\|T_\chi^{-1}u\|^2_{L^2(\Omega_1)} &= \int_{\Omega_1} |(T_\chi^{-1}u)(x)|^2 dx \\
&= \int_{\Omega_1} |u(\chi^{-1}(x))|^2 dx \\
&= \int_{\Omega_0} |u(X)|^2 \det F_\chi(X) dX \\
&\le \|\chi\|^d_{W^{1,\infty}(\Omega_0)} \|u\|^2_{L^2(\Omega_0)},
\end{aligned}$$

表 A.4：常用坐标变换。此处 $F_\chi := \nabla_X \chi$ 为变形梯度张量，$|\cdot|$ 为向量的欧几里得范数，而 $(\cdot)^T$ 表示矩阵转置。

	空间坐标 x	参考坐标 X
坐标	x	$\chi(X)$
函数	$u(x)$	$u(\chi(X)) = (T_\chi^{-1}u)(X)$
梯度（标量函数）	$\nabla_x u(x)$	$F_\chi^{-T}(X)\nabla_X u(\chi(X))$
梯度（向量函数）	$\nabla_x v(x)$	$\nabla_X v(\chi(X))F_\chi^{-1}(X)$
散度	$\nabla_x \cdot v(x)$	$\mathrm{tr}(\nabla_X v(\chi(X))F_\chi^{-1}(X))$
体积元	dx	$\det F_\chi(X)dX$
向量面元	$n(x)ds(x)$	行列式 $F_\chi(X)F_\chi^{-T}(X)N(X)dS(X)$
标量面元	$ds(x)$	$\lvert$行列式 $F_\chi(X)F_\chi^{-T}(X)N(X)\rvert dS(X)$

这就是 L^2 情形下所需的结果。

对于 H^1 情形，我们考虑任意 $u \in H^1(\Omega_0)$。前推的 $H^1(\Omega_1)$ 半范数可写为

$$\begin{aligned}\int_{\Omega_1} |\nabla_x (T_\chi^{-1}u)(x)|^2 dx &= \int_{\Omega_1} |\nabla_x u(\chi^{-1}(x))|^2 dx \\ &= \int_{\Omega_0} |F_\chi(X)^{-T}\nabla_X u(X)|^2 \det F_\chi(X) dX \\ &\le \|\chi\|_{W^{1,\infty}(\Omega_0)}^d \|\chi^{-1}\|_{W^{1,\infty}(\Omega_1)}^2 |u|_{H^1(\Omega_0)}^2,\end{aligned}$$

结合这一 L^2 结果，我们得到

$$\|T_\chi^{-1}u\|_{H^1(\Omega_1)}^2 \le \left(1 + \|\chi^{-1}\|_{W^{1,\infty}(\Omega_1)}^2\right) \|\chi\|_{W^{1,\infty}(\Omega_0)}^d \|u\|_{H^1(\Omega_0)}^2. \tag{A.1}$$

这就得到了 H^1 情形下所需的结果。将论证中的 χ 替换为 χ^{-1} 后重复上述论证，即可得到拉回算子 T_χ的结果。 □

B. 优化误差

B.1. 导数支配界

证明。（定理2.1）证明基于标准的变分分析。在下文中，我们将使用简写记号 $\langle \cdot, \cdot \rangle$ 来表示对偶配对 $\langle \cdot, \cdot \rangle_{\mathcal{Z}' \times \mathcal{Z}}$。首先注意到，根据 λ-强凸性，$\mathcal{J}$对于任意 $z_1, z_2 \in \mathcal{Z}_{ad}$，我们有

$$\mathcal{J}(z_2) \ge \mathcal{J}(z_1) + \langle D_z\mathcal{J}(z_1), z_2 - z_1\rangle + \frac{\lambda}{2}\|z_2 - z_1\|_{\mathcal{Z}}^2,$$

这进一步意味着

$$\langle D_z\mathcal{J}(z_2) - D_z\mathcal{J}(z_1), z_2 - z_1\rangle \ge \lambda\|z_2 - z_1\|_{\mathcal{Z}}^2.$$

最优性间隙。这只是 Polyak-Łojasiewicz 不等式的一种约束优化形式，可以用类似的方法证明。利用 $\mathcal{J}$ 的强凸性，我们得到

$$\mathcal{J}(z^\star) \ge \mathcal{J}(z^\dagger) + \langle D_z\mathcal{J}(z^\dagger), z^\star - z^\dagger\rangle + \frac{\lambda}{2}\|z^\star - z^\dagger\|_{\mathcal{Z}}^2.$$

我们可以重新整理得到

$$\mathcal{J}(z^\dagger) - \mathcal{J}(z^\star) \leq -\langle D_z\mathcal{J}(z^\dagger), z^\star - z^\dagger\rangle - \frac{\lambda}{2}\|z^\star - z^\dagger\|_{\mathcal{Z}}^2. \tag{B.1}$$

由于 $z^\dagger$ 满足 $\mathcal{J}_\theta$的一阶最优性条件，因此我们还可以得到

$$\langle D_z\mathcal{J}_\theta(z^\dagger), z^\star - z^\dagger\rangle \geq 0. \tag{B.2}$$

将（B.2）加到（B.1）上，得到

$$\begin{aligned}\mathcal{J}(z^\dagger) - \mathcal{J}(z^\star) &\leq \langle D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger), z^\star - z^\dagger\rangle - \frac{\lambda}{2}\|z^\star - z^\dagger\|_{\mathcal{Z}}^2 \\ &\leq \|D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger)\|_{\mathcal{Z}'}\|z^\star - z^\dagger\|_{\mathcal{Z}} - \frac{\lambda}{2}\|z^\star - z^\dagger\|_{\mathcal{Z}}^2.\end{aligned} \tag{B.3}$$

（B.3）右侧的形式为二次函数 $at - \lambda t^2/2$，其中 $a = \|D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger)\|_{\mathcal{Z}'} \geq 0$ 且 t 起到 $\|z^\star - z^\dagger\|_{\mathcal{Z}}$的作用。该二次函数在 $t^\star = a/\lambda$处取得最大值。将此最大值代入（B.3）作为上界，可得

$$\mathcal{J}(z^\dagger) - \mathcal{J}(z^\star) \leq \frac{1}{2\lambda}\|D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger)\|_{\mathcal{Z}'}^2, \tag{B.4}$$

这就是所需结果。

优化误差。我们从（B.2）开始，在两边加上并减去导数项 $D_z\mathcal{J}(z^\star)$ 和$D_z\mathcal{J}(z^\dagger)$。得到

$$\langle D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger) + D_z\mathcal{J}(z^\dagger) - D_z\mathcal{J}(z^\star) + D_z\mathcal{J}(z^\star), z^\star - z^\dagger\rangle \geq 0.$$

重新整理可得

$$\langle D_z\mathcal{J}(z^\star) - D_z\mathcal{J}(z^\dagger), z^\star - z^\dagger\rangle \leq \langle D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger), z^\star - z^\dagger\rangle + \langle D_z\mathcal{J}(z^\star), z^\star - z^\dagger\rangle. \tag{B.5}$$

(B.5) 左侧的强凸性意味着

$$\langle D_z\mathcal{J}(z^\star) - D_z\mathcal{J}(z^\dagger), z^\star - z^\dagger\rangle \geq \lambda\|z^\star - z^\dagger\|_{\mathcal{Z}}^2. \tag{B.6}$$

(B.5) 右侧有

$$\langle D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger), z^\star - z^\dagger\rangle \leq \|D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger)\|_{\mathcal{Z}'}\|z^\star - z^\dagger\|_{\mathcal{Z}}, \tag{B.7}$$

and

$$\langle D_z\mathcal{J}(z^\star), z^\star - z^\dagger\rangle \leq 0, \tag{B.8}$$

其中，后一个不等式由 $\mathcal{J}$ 的一阶最优性条件得出。将 (B.6)–(B.8) 代入 (B.5)，可得

$$\lambda\|z^\star - z^\dagger\|_{\mathcal{Z}}^2 \leq \|D_z\mathcal{J}_\theta(z^\dagger) - D_z\mathcal{J}(z^\dagger)\|_{\mathcal{Z}'}\|z^\star - z^\dagger\|_{\mathcal{Z}},$$

由此即可得出所需结果。

□

B.2. 期望的导数误差

本节中，我们给出命题 2.2 的证明。为此，我们首先证明，当 $z \mapsto \mathbb{E}_{\mu_m}[Q(u(\cdot,z)\ \cdot, z)]$ 满足假设 μ_m, u, Q 2.1 和 2.2时，期望会导出一个连续可微泛函。这使我们能够比较将 u 替换为代理模型 u_θ 后导数的差异，如命题 2.2所述。

首先，我们注意到假设 2.1 意味着映射 $z \mapsto u(\cdot, z)$ 和 $z \mapsto D_z u(\cdot, z)$ 是利普希茨连续的。

命题 *B.1.* 设 u 满足假设 *2.1*，其中博雷尔概率测度为 μ_m，紧容许集为 $\mathcal{Z}_{ad}$，开集为 $\mathcal{B}$，且存在某个 $p \geq 2$。那么，映射 $\mathcal{B} \ni z \mapsto u(\cdot, z) \in L^p_{\mu_m}(\mathcal{M}; \mathcal{U})$ 和$\mathcal{B} \ni z \mapsto D_z u(\cdot, z) \in L^p_{\mu_m}(\mathcal{M}; HS(\mathcal{Z}, \mathcal{U}))$ 均为利普希茨连续。也就是说，对于任意 $z_1, z_2 \in \mathcal{B}$

$$\|u(\cdot, z_2) - u(\cdot, z_1)\|_{L^p_{\mu_m}(\mathcal{M};\mathcal{U})} \leq \kappa_u \|z_2 - z_1\|_{\mathcal{Z}}, \tag{B.9}$$

$$\|D_z u(\cdot, z_2) - D_z u(\cdot, z_1)\|_{L^p_{\mu_m}(\mathcal{M};\mathrm{HS}(\mathcal{Z},\mathcal{U}))} \leq \kappa_{u_z} \|z_2 - z_1\|_{\mathcal{Z}}. \tag{B.10}$$

此外，存在常数 M_u 和 M_{u_z} 使得

$$\|u(\cdot, z)\|_{L^p_{\mu_m}(\mathcal{M};\mathcal{U})} \leq M_u, \tag{B.11}$$

$$\|D_z u(\cdot, z)\|_{L^p_{\mu_m}(\mathcal{M};\mathrm{HS}(\mathcal{Z},\mathcal{U}))} \leq M_{u_z}, \tag{B.12}$$

对每个 $z \in \mathcal{Z}_{ad}$均成立。

证明。 首先，对于任意 $z \in \mathcal{B}$, （24a）由假设 2.1 可得

$$\|u(m, z)\|^p_{\mathcal{U}} \leq (\|u(0, z)\|_{\mathcal{U}} + \kappa_u \|m\|_{\mathcal{M}})^p,$$

such 满足 $\mathbb{E}_{m\sim\mu}[\|u(m,\ z)\|^p_{\mathcal{U}} < \infty$，因为 μ_m 具有有限的 p^{th} 阶矩。此外，对于任意 $z_1,\ z_2 \in \mathcal{B}$ ，

$$\|u(\cdot, z_2) - u(\cdot, z_1)\|^p_{L^p_{\mu_m}} = \mathbb{E}_{m\sim\mu_m}[\|u(m, z_2) - u(m, z_1)\|^p_{\mathcal{U}}] \leq \kappa^p_u \|z_2 - z_1\|^p_{\mathcal{Z}},$$

这表明 $z \mapsto u(\cdot,\ z)$ 是利普希茨连续的，其常数为 κ_u。类似地，应用（24b）可得 $z \mapsto D_z u(\cdot,\ z)$ 是利普希茨连续的，其常数为 κ_{u_z}。这种连续性结合可行集 $\mathcal{Z}_{ad} \subset \mathcal{B}$ 的紧致性，进而可推出（B.11）和（B.12）中的上界存在。

□

另一方面，假设 $D_u Q$ 和 $D_{uz} Q$ 中的利普希茨连续性 2.2 意味着以下界成立。

引理 B.1 (二次界)。 设 Q 满足假设 *2.2*，其紧容许集为 $\mathcal{Z}_{ad}$，开集为 $\mathcal{B}$。那么，对于每个 $u_1,\ u_2 \in \mathcal{U}$, $m \in \mathcal{M}$, 以及 $z \in \mathcal{B}$, 我们有

$$|Q(u_2, m, z) - Q(u_1, m, z)| \leq \|D_u Q(u_1, m, z)\|_{\mathcal{U}'} \|u_2 - u_1\|_{\mathcal{U}} + \frac{\kappa_{Q_u}}{2} \|u_2 - u_1\|^2_{\mathcal{U}} \tag{B.13}$$

and

$$\|D_z Q(u_2, m, z) - D_z Q(u_1, m, z)\|_{\mathcal{Z}'} \leq \|D_{uz} Q(u_1, m, z)\|_{\mathcal{L}(\mathcal{U},\mathcal{Z}')} \|u_2 - u_1\|_{\mathcal{U}} + \frac{\kappa_{Q_{uz}}}{2} \|u_2 - u_1\|^2_{\mathcal{U}}. \tag{B.14}$$

证明。 首先考虑 (B.13)。对 Q 作泰勒展开，得到

$$Q(u_2, m, z) - Q(u_1, m, z) = \int_0^1 D_u Q(u_1 + th, m, z) h dt,$$

其中 $h = u_2 - u_1$。取绝对值并利用 $D_u Q$ 的利普希茨连续性，我们有

$$\begin{aligned}
|Q(u_2, m, z) - Q(u_1, m, z)| &\leq \int_0^1 \|D_u Q(u_1 + th, m, z)\|_{\mathcal{U}'} \|h\|_{\mathcal{U}} dt \\
&\leq \int_0^1 \left(\|D_u Q(u_1, m, z)\|_{\mathcal{U}'} + \kappa_{Q_u} t \|h\|_{\mathcal{U}}\right) \|h\|_{\mathcal{U}} dt \\
&\leq \|D_u Q(u_1, m, z)\|_{\mathcal{U}'} \|h\|_{\mathcal{U}} + \frac{\kappa_{Q_u}}{2} \|h\|^2_{\mathcal{U}}.
\end{aligned}$$

对于 D_zQ，我们可以使用类似的泰勒展开

$$D_zQ(u_2,m,z)-D_zQ(u_1,m,z)=\int_0^1 D_uD_zQ(u_1+th,m,z)hdt=\int_0^1 D_{uz}Q(u_1+th,m,z)hdt,$$

这源于 Q 是二阶连续可微的这一事实。取 $\|\cdot\|_{\mathcal{Z}'}$ 并利用 $D_{uz}Q$的利普希茨连续性，可得（B.14）。

□

此外，引理B.1 中的界可写为如下形式。

推论 B.1。 考虑引理 *B.1*的情形。我们有

$$|Q(u_2,m,z)-Q(u_1,m,z)|\le(M_{Q_u}+\kappa_{Q_u}\|u_1\|_{\mathcal{U}})\|u_2-u_1\|_{\mathcal{U}}+\frac{\kappa_{Q_u}}{2}\|u_2-u_1\|_{\mathcal{U}}^2 \tag{B.15}$$

and

$$\|D_zQ(u_2,m,z)-D_zQ(u_1,m,z)\|_{\mathcal{Z}'}\le(M_{Q_{uz}}+\kappa_{Q_{uz}}\|u_1\|_{\mathcal{U}})\|u_2-u_1\|_{\mathcal{U}}+\frac{\kappa_{Q_{uz}}}{2}\|u_2-u_1\|_{\mathcal{U}}^2. \tag{B.16}$$

*Pr*证明。 利用 D_uQ 的利普希茨连续性及其在 $u=0$处的有界性，可得

$$\|D_uQ(u_1,m,z)\|_{\mathcal{U}'}\le\|D_uQ(0,m,z)\|_{\mathcal{U}'}+\kappa_{Q_u}\|u_1\|_{\mathcal{U}}\le M_{Q_u}+\kappa_{Q_u}\|u_1\|_{\mathcal{U}}.$$

同理，

$$\|D_{uz}Q(u_1,m,z)\|_{\mathcal{L}(\mathcal{U},\mathcal{Z}')}\le\|D_{uz}Q(0,m,z)\|_{\mathcal{L}(\mathcal{U},\mathcal{Z}')}+\kappa_{Q_{uz}}\|u_1\|_{\mathcal{U}}\le M_{Q_{uz}}+\kappa_{Q_{uz}}\|u_1\|_{\mathcal{U}}.$$

将上述表达式代入（B.13）和（B.14），即可得到本推论给出的另一种界。

□

或者，我们可以在点 $u_1=0$ 处应用泰勒展开，直接推导出 Q 和 D_zQ 的上界。

推论 B.2. 考虑引理 *B.1*中的情形。对于 $u\in\mathcal{U}$， $m\in\mathcal{M}$ 和 $z\in\mathcal{B}$，有

$$|Q(u,m,z)|\le M_Q+M_{Q_u}\|u\|_{\mathcal{U}}+\frac{\kappa_{Q_u}}{2}\|u\|_{\mathcal{U}}^2 \tag{B.17}$$

and

$$\|D_zQ(u,m,z)\|_{\mathcal{Z}'}\le M_{Q_z}+M_{Q_{uz}}\|u\|_{\mathcal{U}}+\frac{\kappa_{Q_{uz}}}{2}\|u\|_{\mathcal{U}}^2. \tag{B.18}$$

证明。 证明与引理 B.1的证明几乎完全相同，只需改为取 $u_2=u$ 和 $u_1=0$，使得 $h=u_2-u_1=u$。因此有

$$|Q(u,m,z)|\le|Q(0,m,z)|+\|D_uQ(0,m,z)\|_{\mathcal{U}'}\|u\|_{\mathcal{U}}+\frac{\kappa_{Q_u}}{2}\|u\|_{\mathcal{U}}^2 \tag{B.19}$$

and

$$\|D_zQ(u,m,z)\|_{\mathcal{Z}'}\le\|D_zQ(0,m,z)\|_{\mathcal{Z}'}+\|D_{uz}Q(0,m,z)\|_{\mathcal{L}(\mathcal{U},\mathcal{Z}')}\|u\|_{\mathcal{U}}+\frac{\kappa_{Q_{uz}}}{2}\|u\|_{\mathcal{U}}^2. \tag{B.20}$$

代入假设2.2中的上界，即可得到所需结果。 □

B.2.1. 感兴趣量的可微性

如前所述，对于任意固定的 z，量 $Q(u(m,z)\ m,z)$ 是一个随机变量，其取值取决于 $m\in\mathcal{M}$ 的实现。因此，我们可以将映射 $z\mapsto Q(u(\cdot,z)\ \cdot,z)$ 视为从变量 z到随机变量空间的映射。在假设 2.1 和 2.2 下，我们可以证明这是一个从 $\mathcal{B}$ 到 $L^1_{\mu_m}$的连续可微映射。

命题 *B.2.* 假设 u 满足假设 *2.1*，且 Q 满足假设 *2.2*，其中 *Borel* 概率测度为 μ_m、可行集为 $\mathcal{Z}_{ad}$，并且对某个 $p\geq 2$，开集为 $\mathcal{B}$。那么，映射 $f_Q\colon z\mapsto Q(u(\cdot,z)\ \cdot,z)$ 是一个从 $\mathcal{B}$ 到 $L^1_{\mu_m}(\mathcal{M};\mathbb{R})$ 的连续可微函数。

证明。定义良好性：我们首先验证 $f_Q(z)$ 属于 $L^1_{\mu_m}$ 对任意 $z\in\mathcal{B}$。由推论 B.17中的 (B.2)，可得

$$\|f_Q(z)\|_{L^1_{\mu_m}}=\mathbb{E}_{m\sim\mu}[|Q(u(m,z),m,z)|]\leq\mathbb{E}_{m\sim\mu_m}\left[M_Q+M_{Q_u}\|u(m,z)\|_{\mathcal{U}}+\frac{\kappa_{Q_u}}{2}\|u(m,z)\|^2_{\mathcal{U}}\right].$$

由于 $u(\cdot,z)\in L^p_{\mu_m}(\mathcal{M},\mathcal{U})$ 且 $p\geq 2$，因此所得期望有限，即 $f_Q(z)\in L^1_{\mu_m}$

可微性：由于 Q和 u 均为连续可微，导数的形式为 具体为

$$D_zf_Q(z)=D_uQ(u(\cdot,z),\cdot,z)D_zu(\cdot,z)+D_zQ(u(\cdot,z),\cdot,z).$$

我们验证这确实是导数。特别地，我们考虑残差

$$r(z,h):=\|Q(u(\cdot,z+h),\cdot,z+h)-Q(u(\cdot,z),\cdot,z)-D_zf_Q(z)h\|_{L^1_{\mu_m}}$$

是 $o(\|h\|_{\mathcal{Z}})$。我们首先令 $A(m,\ z):=D_uQ(u(m,\ z)\ m,\ z)D_zu(m,\ z)+D_zQ(u(m,\ z)\ m,\ z)$，使得 hat

$$\frac{r(z,h)}{\|h\|_{\mathcal{Z}}}=\mathbb{E}_{m\sim\mu_m}\left[\frac{|Q(u(m,z+h),m,z+h)-Q(u(m,z),m,z)-A(m,z)h|}{\|h\|_{\mathcal{Z}}}\right].$$

特别地，由于 Q 和 u的可微性，被积函数在 m 中逐点收敛于零，当 $h\to 0$时。现在我们要为被积函数寻找一个控制函数。为此，定义

$$\mathsf{A}_1:=D_uQ(u(m,z),m,z)D_zu(m,z),\qquad \mathsf{A}_2:=D_zQ(u(m,z),m,z)$$

使得 $A=\mathsf{A}_1+\mathsf{A}_2$, 且

$$\mathsf{D}_1:=Q(u(m,z+h),m,z+h)-Q(u(m,z),m,z+h),\quad \mathsf{D}_2:=Q(u(m,z),m,z+h)-Q(u(m,z),m,z),$$

使得 $Q(u(m,\ z+h)\ m,\ z+h)-Q(u(m,\ z)\ m,\ z)=\mathsf{D}_1+\mathsf{D}_2$。我们可以分别对每一项进行界定。

1. 对于 A_1, , 我们利用 D_uQ的利普希茨连续性，使得

$$\frac{|\mathsf{A}_1(m,z)h|}{\|h\|_{\mathcal{Z}}}\leq\|\mathsf{A}_1(m,z)\|_{\mathcal{Z}'}\leq M_{Q_u}\|D_zu(m,z)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})}+\kappa_{Q_u}\|u(m,z)\|_{\mathcal{U}}\|D_zu(m,z)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})}.\quad\text{(B.21)}$$

2. 对于 A_2, , 我们使用推论 B.18中的（B.2）, 使得

$$\frac{|\mathsf{A}_2(m,z)h|}{\|h\|_{\mathcal{Z}}}\leq\|\mathsf{A}_2(m,z)\|_{\mathcal{Z}'}\leq M_{Q_z}+M_{Q_{uz}}\|u(m,z)\|_{\mathcal{U}}+\frac{\kappa_{Q_{uz}}}{2}\|u(m,z)\|^2_{\mathcal{U}}.\quad\text{(B.22)}$$

3. 对于 D_1, , 我们使用推论 B.15 中的式子（B.1）, 从而得到

$$\begin{aligned}|\mathsf{D}_1(m,z)|&\leq(M_{Q_u}+\|u(m,z)\|_{\mathcal{U}})\|u(m,z+h)-u(m,z)\|_{\mathcal{U}}+\frac{\kappa_{Q_u}}{2}\|u(m,z+h)-u(m,z)\|^2_{\mathcal{Z}}\\&\leq(M_{Q_u}+\|u(m,z)\|_{\mathcal{U}})\kappa_u\|h\|_{\mathcal{Z}}+\frac{\kappa_{Q_u}\kappa_u^2}{2}\|h\|^2_{\mathcal{Z}},\end{aligned}$$

其中，后一个不等式由 u 的利普希茨连续性以及假设2.1可得。因此，

$$\frac{|\mathsf{D}_1(m,z)|}{\|h\|_{\mathcal{Z}}}\leq(M_{Q_u}+\|u(m,z)\|_{\mathcal{U}})\kappa_u+\frac{\kappa_{Q_u}\kappa_u^2}{2}\|h\|_{\mathcal{Z}}.\quad\text{(B.23)}$$

4. 对于 D_2, ，我们首先应用均值定理，从而得到

$$|\mathsf{D}_2(m,z)| \le \|D_z Q(u(m,z),m,\zeta)\|_{\mathcal{Z}'}\|h\|_{\mathcal{Z}},$$

其中 $\zeta = z + th$ 对于某个 $t \in [0,1]$。当 h足够小时，我们有 $\zeta \in \mathcal{B}$。因此，可以应用推论 B.18中的（B.2）得到

$$\frac{|\mathsf{D}_2(m,z)|}{\|h\|_{\mathcal{Z}}} \le \|D_z Q(u(m,z),m,\zeta)\|_{\mathcal{Z}'} \le M_{Q_z} + M_{Q_{uz}}\|u(m,z)\|_{\mathcal{U}} + \frac{\kappa_{Q_{uz}}}{2}\|u(m,z)\|_{\mathcal{U}}^2. \tag{B.24}$$

在每种情况下，上界都是 $\|u(m,z)\|_{\mathcal{U}}$ 和 $\|D_z u(m,z)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})}$ 的二次型（或等价地，$\|D_z u(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}$），二者都是 $L^2_{\mu_m}$ 函数。因此，被积函数受一个 $L^1_{\mu_m}$函数控制。随后利用控制收敛定理，可以取极限，从而得到

$$\frac{r(z,h)}{\|h\|_{\mathcal{Z}}} \to 0 \qquad \text{as } h \to 0.$$

导数的连续性： 为说明连续性，任取 $z_1, z_2 \in \mathcal{B}$。我们有

$$\|D_z f_Q(z_2) - D_z f_Q(z_1)\|_{\mathcal{Z}'} \le \mathbb{E}_{m\sim\mu_m}[\|\mathsf{A}_1(m,z_2) - \mathsf{A}_1(m,z_1)\|_{\mathcal{Z}'} + \|\mathsf{A}_2(m,z_2) - \mathsf{A}_2(m,z_1)\|_{\mathcal{Z}'}].$$

, $D_z Q$, u以及 $D_z u$ 的连续性意味着被积函数对每个 m都收敛到零。再次需要找到一个控制函数，使得可以将极限 $z_2 \to z_1$ 移到期望之外。具体而言，（B.21）给出

$$\|\mathsf{A}_1(m,z_1)\|_{\mathcal{Z}'} \le M_{Q_u}\|D_z u(m,z_1)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})} + \kappa_{Q_u}\|u(m,z_1)\|_{\mathcal{U}}\|D_z u(m,z_1)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})}.$$

利用 u 和 $D_z u$的利普希茨连续性，对于 $h = z_2 - z_1$，我们还有

$$\begin{aligned}\|\mathsf{A}_1(m,z_2)\|_{\mathcal{Z}'} &\le M_{Q_u}\|D_z u(m,z_2)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})} + \kappa_{Q_u}\|u(m,z_2)\|_{\mathcal{U}}\|D_z u(m,z_2)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})}\\ &\le M_{Q_u}\left(\|D_z u(m,z_1)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})} + \kappa_{u_z}\|h\|_{\mathcal{Z}}\right)\\ &\quad + \kappa_{Q_u}(\|u(m,z_1)\|_{\mathcal{U}} + \kappa_u\|h\|_{\mathcal{Z}})(\|D_z u(m,z_1)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})} + \kappa_{u_z}\|h\|_{\mathcal{Z}}),\end{aligned}$$

当 h足够小时， $\|\mathsf{A}_1(m,z_1)\|_{\mathcal{Z}'}$和 $\|\mathsf{A}_1(m,z_2)\|_{\mathcal{Z}'}$均受某个共同的 $L^1_{\mu_m}$ 函数支配。同理，对于A_2, (B.22)可得

$$\|\mathsf{A}_2(m,z_1)\|_{\mathcal{Z}'} \le M_{Q_z} + M_{Q_{uz}}\|u(m,z_1)\|_{\mathcal{U}} + \frac{\kappa_{Q_{uz}}}{2}\|u(m,z_1)\|_{\mathcal{U}}^2,$$

而 u 和 $D_z u$ 的利普希茨连续性意味着

$$\|\mathsf{A}_2(m,z_2)\|_{\mathcal{Z}'} \le M_{Q_z} + M_{Q_{uz}}(\|u(m,z_1)\|_{\mathcal{U}} + \kappa_u\|h\|_{\mathcal{Z}}) + \frac{\kappa_{Q_{uz}}}{2}(\|u(m,z_1)\|_{\mathcal{U}} + \kappa_u\|h\|_{\mathcal{Z}})^2.$$

使得 $\|\mathsf{A}_2($ $m,z_1)\|_{\mathcal{Z}'}$ 和 $\|\mathsf{A}_2(m,z_2)\|_{\mathcal{Z}'}$ 也受某个 $L^1_{\mu_m}$ 函数支配。支配收敛定理 从而允许取极限，使得

$$\|D_z f_Q(z_2) - D_z f_Q(z_1)\|_{\mathcal{Z}'} \to 0 \qquad \text{as } z_2 \to z_1,$$

这表明了 $D_z f_Q$的连续性。 □

该结果的一个推论是，由 Q 的期望定义的代价函数是连续可微的。

推论 B.3. 考虑命题 *B.2* 的设定。则函数 $\mathcal{J}$： $\mathcal{B} \to \mathbb{R}$ 由下式给出

$$\mathcal{J}(z) := \mathbb{E}_{m\sim\mu_m}[Q(u(m,z),m,z)] \tag{B.25}$$

是连续可微的，且其导数由下式给出

$$D_z\mathcal{J}(z) = \mathbb{E}_{m\sim\mu_m}[D_u Q(u(m,z),m,z)D_z u(m,z) + D_z Q(u(m,z),m,z)]. \tag{B.26}$$

证明。 这由链式法则以及如下事实得出：期望 $\mathbb{E}_{\mu_m}[\cdot]$是 $L^1_{\mu_m}$上的连续线性泛函，因此是连续可微的。

□

B.2.2. 命题*2.2*的证明

我们现在给出命题2.2的证明 .

证明。 回顾推论 B.3，我们有以下表达式，分别表示 $\mathcal{J}^{\text{mean}}(z) = \mathbb{E}_{m\sim\mu_m}[Q(u(m,z)\ m,z)]$ 的导数，以及 $\mathcal{J}_{均值\theta}(z) = \mathbb{E}_{m\sim\mu_m}[Q(u_\theta(m,z)\ m,z)]$，

$$D_z\mathcal{J}^{\text{mean}}(z) = \mathbb{E}_{m\sim\mu_m}[D_uQ(u(m,z),m,z)D_zu(m,z) + D_zQ(u(m,z),m,z)],$$
$$D_z\mathcal{J}_\theta^{\text{mean}}(z) = \mathbb{E}_{m\sim\mu_m}[D_uQ(u_\theta(m,z),m,z)D_zu_\theta(m,z) + D_zQ(u_\theta(m,z),m,z)].$$

我们可以用以下三个项来表示它们的差

$$\begin{aligned}
\mathsf{D}_1(m,z) &:= (D_uQ(u(m,z),m,z) - D_uQ(u_\theta(m,z),m,z))\, D_zu(m,z),\\
\mathsf{D}_2(m,z) &:= D_uQ(u_\theta(m,z),m,z)\,(D_zu(m,z) - D_zu_\theta(m,z))\,,\\
\mathsf{D}_3(m,z) &:= D_zQ(u(m,z),m,z) - D_zQ(u_\theta(m,z),m,z),
\end{aligned}$$

使得

$$D_z\mathcal{J}^{\text{mean}}(z) - D_z\mathcal{J}_\theta^{\text{mean}}(z) = \mathbb{E}_{m\sim\mu_m}[\mathsf{D}_1(m,z) + \mathsf{D}_2(m,z) + \mathsf{D}_3(m,z)].$$

现在考虑任意一个 $z \in \mathcal{Z}_{ad}$。我们可以分别对每个 D_i 的范数进行界定。为简化记号，我们用 $\|u(\cdot,z)\|_{L^2_{\mu_m}}$ 表示状态博赫纳范数 $\|u(\cdot,z)\|_{L^2_{\mu_m}(\mathcal{M};\mathcal{U})}$，并用 $\|D_zu(\cdot,z)\|_{L^2_{\mu_m}}$ 表示导数博赫纳范数 $\|D_zu(\cdot,z)\|_{L^2_{\mu_m}(\mathcal{M};\mathrm{HS}(\mathcal{Z},\mathcal{U}))}$（误差也同理）。此外，我们将使用以下事实： $\|\cdot\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})} \le \|\cdot\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}$，并且在下文的推导中不再特别说明。于是有：

1. 对于 D_1，我们有

$$\|\mathsf{D}_1(m,z)\|_{\mathcal{Z}'} \le \|D_uQ(u(m,z),m,z) - D_uQ(u_\theta(m,z),m,z)\|_{\mathcal{U}'}\|D_zu(m,z)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})}.$$

于是， D_uQ的利普希茨连续性给出

$$\|\mathsf{D}_1(m,z)\|_{\mathcal{Z}'} \le \kappa_{Q_u}\|u(m,z) - u_\theta(m,z)\|_{\mathcal{U}}\|D_zu(m,z)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})}$$

然后，取期望并利用柯西–施瓦茨不等式，我们有

$$\mathbb{E}_{m\sim\mu_m}[\|\mathsf{D}_1(m,z)\|_{\mathcal{Z}'}] \le \kappa_{Q_u}\|u(\cdot,z) - u_\theta(\cdot,z)\|_{L^2_{\mu_m}}\|D_zu(\cdot,z)\|_{L^2_{\mu_m}} \le \kappa_{Q_u}M_{u_z}\|u(\cdot,z) - u_\theta(\cdot,z)\|_{L^2_{\mu_m}},$$

其中我们利用了命题 B.1。

2. 对于 D_2，我们有

$$\|\mathsf{D}_2(m,z)\|_{\mathcal{Z}'} \le \|D_uQ(u_\theta(m,z),m,z)\|_{\mathcal{U}'}\|D_zu(m,z) - D_zu_\theta(m,z)\|_{\mathcal{L}(\mathcal{Z},\mathcal{U})}.$$

我们可以进一步利用 D_uQ的利普希茨连续性，从而写出

$$\begin{aligned}
\|D_uQ(u_\theta(m,z),m,z)\|_{\mathcal{U}'} &\le \|D_uQ(0,m,z)\|_{\mathcal{U}'} + \kappa_{Q_u}\|u(m,z) - u_\theta(m,z)\|_{\mathcal{U}} + \kappa_{Q_u}\|u(m,z)\|_{\mathcal{U}}\\
&\le M_{Q_u} + \kappa_{Q_u}\|u(m,z) - u_\theta(m,z)\|_{\mathcal{U}} + \kappa_{Q_u}\|u(m,z)\|_{\mathcal{U}}.
\end{aligned}$$

取期望，并利用柯西–施瓦茨不等式以及上界

$\|u\cdot z\|_{L^2_{\mu_m}} \le (,)$由命题 B.我们同样有 M_u 1,

$$\begin{aligned}
\mathbb{E}_{m\sim\mu_m}[\|\mathsf{D}_2(m,z)\|_{\mathcal{Z}'}] \le{}& (M_{Q_u} + \kappa_{Q_u}M_u)\|D_zu(\cdot,z) - D_zu_\theta(\cdot,z)\|_{L^2_{\mu_m}}\\
&+ \kappa_{Q_u}\|u(\cdot,z) - u_\theta(\cdot,z)\|_{L^2_{\mu_m}}\|D_zu(\cdot,z) - D_zu_\theta(\cdot,z)\|_{L^2_{\mu_m}}.
\end{aligned}$$

注意，根据杨氏不等式， $ab \le \frac{a^2}{2} + \frac{b^2}{2}$，我们可以将其写成不含交叉项的形式：

$$\begin{aligned}
\mathbb{E}_{m\sim\mu_m}[\|\mathsf{D}_2(m,z)\|_{\mathcal{Z}'}] \le{}& (M_{Q_u} + \kappa_{Q_u}M_u)\|D_zu(\cdot,z) - D_zu_\theta(\cdot,z)\|_{L^2_{\mu_m}}\\
&+ \frac{\kappa_{Q_u}}{2}\|u(\cdot,z) - u_\theta(\cdot,z)\|^2_{L^2_{\mu_m}} + \frac{\kappa_{Q_u}}{2}\|D_zu(\cdot,z) - D_zu_\theta(\cdot,z)\|^2_{L^2_{\mu_m}}.
\end{aligned}$$

3. 最后，对于 $\mathsf{D}_3,$，我们直接使用推论 B.1中的式（B.1），

$$\|\mathsf{D}_3(m,z)\|_{\mathcal{Z}'} \le (M_{Q_{uz}} + \kappa_{Q_{uz}}\|u(m,z)\|_{\mathcal{U}})\|u(m,z)-u_\theta(m,z)\|_{\mathcal{U}} + \frac{\kappa_{Q_{uz}}}{2}\|u(m,z)-u_\theta(m,z)\|_{\mathcal{U}}^2.$$

再次取期望，并利用柯西–施瓦茨不等式以及命题B.1的上界，我们有

$$\mathbb{E}_{m\sim\mu_m}[\|\mathsf{D}_3(m,z)\|_{\mathcal{Z}'}] \le (M_{Q_{uz}} + \kappa_{Q_{uz}}M_u)\,\|u(\cdot,z)-u_\theta(\cdot,z)\|_{L^2_{\mu_m}} + \frac{\kappa_{Q_{uz}}}{2}\|u(\cdot,z)-u_\theta(\cdot,z)\|^2_{L^2_{\mu_m}}.$$

综合上述结果，我们有

$$\begin{aligned}\|D_z\mathcal{J}^{\mathrm{mean}}(z) - D_z\mathcal{J}_\theta^{\mathrm{mean}}(z)\|_{\mathcal{Z}'} \le{}& C_1\left(\|u(\cdot,z)-u_\theta(\cdot,z)\|_{L^2_{\mu_m}} + \|D_zu(\cdot,z)-D_zu_\theta(\cdot,z)\|_{L^2_{\mu_m}}\right)\\ &+ C_2\left(\|u(\cdot,z)-u_\theta(\cdot,z)\|^2_{L^2_{\mu_m}} + \|D_zu(\cdot,z)-D_zu_\theta(\cdot,z)\|^2_{L^2_{\mu_m}}\right),\end{aligned}$$

其中 C_1、 $C_2>0$ 是依赖于 M_{Q_u}、 $M_{Q_{uz}}$、 κ_{Q_u}、 $\kappa_{Q_{uz}}$、 M_u 但不依赖于 u_θ的常数。 □

C. 通用逼近结果

在本节中，我们给出RBNO在输出和导数两方面的通用逼近性质的证明（定理3.1），以及其随后在风险中性优化问题中的应用（定理3.2）。为证明定理3.1，我们将RBNO的逼近误差分解为与输入和输出投影相关的降维误差，以及所得潜在空间中的神经网络逼近误差。我们分别对每一类逼近误差来源给出界，首先是第C.1 节中的潜在空间神经网络，随后是第C.2节中的降维，再将它们合并为定理3.1 在第C.3节中的证明。定理3.2 的证明则由定理3.1 连同定理2.1 以及命题2.2得出。

在本节中，我们将偏离注释 2.1 中关于空间区域与参考域之间坐标变换的符号约定。具体而言， x 将用于表示 $\mathbb{R}^{d_x}$中的一般向量，而 $\tilde{\cdot}$ 不再表示相关量属于空间区域还是参考域。

C.1. 有限维神经网络逼近

我们首先证明有限维情形下神经网络通用逼近定理的一个变体，表述如下。

定理 C.1 （半有界集上的通用逼近）。设 $g\in C^1(\mathbb{R}^{d_x}\times\mathcal{B};\ \mathbb{R}^{d_g})$，其中 $\mathcal{B}\subset\mathbb{R}^{d_y}$是开集，且 μ_x 是 $\mathbb{R}^{d_x}$上的博雷尔概率测度。假设对于某个 $p\ge 1$,

1. 映射 $\mathcal{B}\ni y\mapsto g(\cdot,y)\in L^p_{\mu_x}(\mathbb{R}^{d_x};\ \mathbb{R}^{d_g})$ 连续；*2.* 映射 $\mathcal{B}\ni y\mapsto D_yg(\cdot,y)\in L^p_{\mu_x}(\mathbb{R}^{d_x};\ \mathbb{R}^{d_g\times d_y})$ 连续，其中 $\mathbb{R}^{d_g\times d_y}$配备弗罗贝尼乌斯范数 $\|\cdot\|_F$。

Then 对于任意 $\epsilon>0$ 以及任意紧集 $K_y\subset\mathcal{B}$，存在一个 *GELU* 神经网络 g_θ，使得 *at*

$$\sup_{y\in K_y}\mathbb{E}_{x\sim\mu_x}\left[\|g(x,y)-g_\theta(x,y)\|_2^p\right] + \mathbb{E}_{x\sim\mu_x}\left[\|D_yg(x,y)-D_yg_\theta(x,y)\|_F^p\right] < \epsilon^p.$$

该结果与传统的通用逼近结果有两点不同。首先，近似精度同时考虑输出误差和导数误差。其次，我们不再将近似误差视为在紧集上一致，或视为关于分布的积分量（其支撑集可能无界），而是将输入分解为两组变量，分别记为 x 和 y，对前者进行积分，对后者取上确界。这体现了

两个输入在解算子 $(m,\ z) \mapsto u(m,\ z)$ 中所起的不同作用：积分是关于 m进行的，而优化是关于 z进行的。

在本文中，我们沿用C.1 中采用的截断论证，证明定理 [55]。也就是说，我们利用前馈神经网络在紧集上的传统通用逼近结果，构造一个神经网络 $\tilde{g}$，使其在紧子集 $K_x \times K_y \subset \mathbb{R}^{d_x} \times \mathcal{B}$上具有较高精度。接着，我们构造一个由第二个神经网络表示的截断函数，使其作用于 $\tilde{g}$时，能够控制紧集 $K_x \times K_y$之外的逼近误差。最后，我们证明，当 K_x 趋近于整个 $\mathbb{R}^{d_x}$时，这一构造可以取极限，从而使 $L^p_{\mu_x}$ 在 $\mathbb{R}^{d_x}$ 上的误差在 $y \in K_y$上一致地任意小。

我们给出定理C.1 的证明如下。首先，我们说明如何将适当的截断函数构造为神经网络（引理C.2至C.4）。接着，我们说明如何将截断函数与标准通用逼近结果结合起来，从而构造 g 的近似，使其在紧集内精确，同时在紧集之外具有受控的增长率（引理C.5和C.6）。最后，在C.1.3节中，我们通过极限论证，将C.1 趋近于 K_x，结合上述结果，完成定理 $\mathbb{R}^{d_x}$的证明。

C.1.1. 神经网络构造

为了构造截断函数，我们首先需要将范数函数 $\|\cdot\|$近似为神经网络。如 [55, 引理 D.4], 所示，我们可以使用 GELU 激活函数构造绝对值函数 $|\cdot|$的近似。我们将这一结果总结如下。

引理 C.1（绝对值函数）。给定 $\theta > 0$，定义 $x \in \mathbb{R}$

$$f_{\text{abs}}(x;\theta) := \frac{\sigma_{\text{GELU}}(\theta x) + \sigma_{\text{GELU}}(-\theta x)}{\theta}, \tag{C.1}$$

那么，对于任意 $\epsilon > 0$，存在 $\theta > 0$ 使得

$$|f_{\text{abs}}(x;\theta) - |x|| \le \epsilon \tag{C.2}$$

并且 $f_{\text{abs}}(x;\ \theta)\ \ge 0$ 对所有 $x \in \mathbb{R}$均成立。

然后，我们可以利用这一结果构造 E范数的近似，如下所示。

引理 C.2 (ℓ_1 范数)。给定 $\theta > 0$，定义 $x \in \mathbb{R}^{d_x}$

$$f_{\ell_1}(x;\theta) := \mathbf{1}^T \frac{\sigma_{\text{GELU}}(\theta x) + \sigma_{\text{GELU}}(-\theta x)}{\theta}, \tag{C.3}$$

其中 $\mathbf{1} := (1, \ldots, 1) \in \mathbb{R}^{d_x}$。于是，对于任意 $\epsilon > 0$，存在 $\theta > 0$ 使得

$$|f_{\ell_1}(x;\theta) - \|x\|_1| \le \epsilon \tag{C.4}$$

并且 $f_{\ell_1}(x;\theta)\ \ge 0$ 对所有 $x \in \mathbb{R}^{d_x}$均成立。

证明。如引理 C.1所示，GELU激活函数可用于在 $\mathbb{R}$上构造 ϵ的绝对值函数近似，存在 $\theta > 0$ 使得

$$f_{\text{abs}}(t;\theta) := \frac{\sigma_{\text{GELU}}(\theta t) + \sigma_{\text{GELU}}(-\theta t)}{\theta}, \tag{C.5}$$

满足 $f_{\text{abs}}(t;\theta)\ \ge 0$和 $|f_{\text{abs}}(t;\ \theta)\ - |t|| \le \epsilon/d_x$对所有 $t \in \mathbb{R}$均成立。这意味着对于 $x \in \mathbb{R}^{d_x}$，我们有$f_{\ell_1}(x;\theta)$ ≥ 0 和

$$|f_{\ell_1}(x;\theta) - \|x\|_1| \le \sum_{j=1}^{d_x} |f_{\text{abs}}(x_j;\theta) - |x_j|| \le \epsilon.$$

□

˜ ˜GELU激活函数也可用于构造一个截断函数 ˜ f_{cut} $\mathbb{R}$ f_{cut} t ≈ 1 在上，即
()对于 $t \le R$ 而 $f_{\text{cut}}(t) \approx 0$ 对于 $t \gg R$。以下是 [55, 引理 D.5]。

引理 C.3 (截断函数)。 给定 $\theta, b > 0$，定义 $x \in \mathbb{R}$

$$\tilde{f}_{\text{cut}}(x;\theta,b) := \frac{\sigma_{\text{GELU}}(\theta(x-b-1)) - \sigma_{\text{GELU}}(\theta(x-b))}{\theta} + 1. \tag{C.6}$$

那么，对于任意 $\epsilon > 0$ 和 $R > 1$，存在 $\theta, b > 0$ 使得以下条件成立：

$$\begin{cases} |\tilde{f}_{\text{cut}}(x;\theta,b) - 1| \le \epsilon & 0 \le x \le R, \\ |\tilde{f}_{\text{cut}}(x;\theta,b)| \le \epsilon & x \ge 4R, \end{cases} \tag{C.7}$$

˜此外，还存在一个与 θ, b 无关的 $M > 0$，使得对于所有 $x \in$， $|f_{\text{cut}}(x;\theta,b)| \le M$ $\mathbb{R}$.

引理 C.2 和 C.3 可用于构造一个基于输入的 ℓ_2 范数的截断函数。

引理 C.4 (使用 ℓ_2 范数的截断函数)。 存在 $M > 0$，使得对于任意 $\epsilon > 0$ 和 $R > 1$，存在一个两层 *GELU* 神经网络 f_{cut}： $\mathbb{R}^{d_x} \to \mathbb{R}$ 满足

$$\begin{cases} |f_{\text{cut}}(x) - 1| \le \epsilon & 0 \le \|x\|_2 \le R, \\ |f_{\text{cut}}(x)| \le \epsilon & \|x\|_2 \ge 5(\sqrt{d_x}+1)R, \\ |f_{\text{cut}}(x)| \le M & x \in \mathbb{R}^{d_x}. \end{cases} \tag{C.8}$$

˜证明。 我们将 $f_{\text{cut}}(x) = f_{\text{cut}}(x;\theta_{\ell_1}, \theta_{\text{cut}}, b_{\text{cut}})$ 定义为 f_{ℓ_1} 和 f_{cut}的复合，即，

$$f_{\text{cut}}(x;\theta_{\ell_1},\theta_{\text{cut}},b_{\text{cut}}) = \tilde{f}_{\text{cut}}(f_{\ell_1}(x;\theta_{\ell_1});\theta_{\text{cut}},b_{\text{cut}}). \tag{C.9}$$

__˜我们将为 f_{cut} 选择截断半径，以考虑从 ℓ_1 到 ℓ_2的转换，即 $\|x\|_2 \le \|x\|_1 \le \sqrt{d_x}\|x\|_2$。

为此，设 ϵ, R 已知。我们构造 f_{ℓ_1} 使其满足

$$|f_{\ell_1}(x) - \|x\|_1| \le R, \qquad \forall x \in \mathbb{R}^{d_x},$$

˜并令 f_{cut} 满足

$$\begin{cases} |\tilde{f}_{\text{cut}}(x;\theta,b) - 1| \le \epsilon & 0 \le x \le \tilde{R}, \\ |\tilde{f}_{\text{cut}}(x;\theta,b)| \le \epsilon & x \ge 4\tilde{R}, \\ |\tilde{f}_{\text{cut}}(x;\theta,b)| \le M & x \in \mathbb{R}^{d_x}, \end{cases}$$

其中 $\tilde{R} := (\sqrt{d_x}+1)R$。

接下来考虑以下情形。

1. $\|x\|_2 \le R$。我们有

$$f_{\ell_1}(x) \le \|x\|_1 + R \le (\sqrt{d_x}+1)R = \tilde{R},$$

implying that $|\tilde{f}_{\text{cut}}(f_{\ell_1}(x)) - 1| \le \epsilon$.

2. $\|x\|_2 \ge 5(\sqrt{d_x}+1)R$。我们有

$$f_{\ell_1}(x) \ge \|x\|_1 - R \ge 5(\sqrt{d_x}+1)R - R \ge 4\tilde{R},$$

˜这意味着 $|f_{\text{cut}}(f_{\ell_1}(x))| \le \epsilon$。

最后， $|f_{\text{cut}}(x)| \le M$ 对任意 $x \in \mathbb{R}^{d_x}$ 成立，且具有相同的 M，参见引理C.3。由于 $\tilde{f}_{\text{cut}}$ 与f_{ℓ_1} 均可解释为单层神经网络，其所得复合 $f_{\text{cut}}(x)$ 是一个双层神经网络。

□

C.1.2. 具有受控增长的神经网络近似

通用逼近定理，例如 [98, 定理 4。1] 表明，在紧输入域上，单层神经网络能够在输出值及其导数方面以任意精度逼近连续可微函数。作为推论，我们可以证明，对于 GELU 激活函数，尽管 GELU 本身无界，仍可将所得逼近函数构造为使其输出和导数全局有界。

引理 C.5 (紧集上的有界逼近)。 设 $g \in C^1(\mathbb{R}^{d_x} \times \mathcal{B};\ \mathbb{R}^{d_g})$，其中 $\mathcal{B} \subset \mathbb{R}^{d_y}$ 是开集。那么，对于任意 $\epsilon > 0$ 以及紧集 $K_x \subset \mathbb{R}^{d_x}$ 和 $K_y \subset \mathcal{B}$，存在一个单层*GELU*神经网络 g_θ，使得以下条件同时成立

$$\sup_{(x,y)\in K_x \times K_y} \|g(x,y) - g_\theta(x,y)\|_2 \leq \epsilon, \tag{C.10}$$

and

$$\sup_{(x,y)\in K_x \times K_y} \|D_y g(x,y) - D_y g_\theta(x,y)\|_F \leq \epsilon. \tag{C.11}$$

此外， g $_\theta$ 及其导数有界，即存在 $M = M(\epsilon) > 0$，使得以下条件同时成立

$$\sup_{(x,y)\in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}} \|g_\theta(x,y)\|_2 \leq M \tag{C.12}$$

and

$$\sup_{(x,y)\in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}} \|D_y g_\theta(x,y)\|_F \leq M. \tag{C.13}$$

证明。 首先，考虑函数 $\sigma(x) = \sigma_{\text{GELU}}(x+1) - \sigma_{\text{GELU}}(x-1)$。可以证明， $\sigma \in C^\infty(\mathbb{R})$，且 $\sup_{x\in\mathbb{R}} |\sigma^{(k)}(x)| < \infty$ 对所有 $k \in \{0\} \cup \mathbb{N}$均成立。也就是说， σ 是一个光滑的非恒定函数，且其输出和导数有界。

现在，令 $\xi := (x,\ y)$ 且 $K := K_x \times K_y$。根据标准的通用逼近定理（例如 [98,定理 4.1], ），存在一个如下形式的单层神经网络

$$g_\theta(\xi) = W_2\sigma(W_1\xi + b_1) + b_2 \tag{C.14}$$

使得以下条件同时成立

$$\sup_{\xi\in K} \|g(\xi) - g_\theta(\xi)\|_2 \leq \epsilon \tag{C.15}$$

and

$$\sup_{\xi\in K} \|D_\xi g(\xi) - D_\xi g_\theta(\xi)\|_F \leq \epsilon. \tag{C.16}$$

我们注意到， $g_\theta(\xi)$ 也可以写作

$$g_\theta = W_2(\sigma_{\text{GELU}}(W_1\xi + b_1 + 1) - \sigma_{\text{GELU}}(W_1\xi + b_1 - 1)) + b_2, \tag{C.17}$$

这可以看作一个单层 GELU 神经网络。此外，由于 σ 的值及导数有界，因此对所有 $\xi \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}.$， $D_\xi g_\theta(\xi)$ 都有界。最后，只需注意到

$$\|D_y g(x,y) - D_y g_\theta(x,y)\|_F \leq \|D_\xi g(\xi) - D_\xi g_\theta(\xi)\|_F.$$

□

注 C.1. 在上述证明中，我们引用了通用逼近定理，这些定理考虑的是在整个输入空间上定义的函数的近似，例如 $g \in C^1(\mathbb{R}^{d_x} \times \mathbb{R}^{d_y};\ \mathbb{R}^{d_g})$，而我们仅假设 g 定义在开子集 $\mathbb{R}^{d_x} \times \mathcal{B}$上。这在很大程度上无关紧要，原因如下。由于 $K \Subset \mathbb{R}^{d_x} \times \mathcal{B}$，其中 $\Subset$ 表示紧嵌入，存在开集 $\mathcal{B}_1$、$\mathcal{B}_2$，使得$K \Subset \mathcal{B}_1 \Subset \mathcal{B}_2 \Subset \mathcal{B}$。此外，对于闭包 $\text{cl}(\mathcal{B}_1)$，我们考虑示性函数的一个光滑外部逼近，

$\mathbb{1}_{\mathcal{B}_1} \in C^\infty(\mathbb{R}^{d_x} \times \mathbb{R}^{d_y}; \mathbb{R})$，使得 $\mathbb{1}_{\mathcal{B}_1}|_{\mathcal{B}_1} = 1$ 且 $\mathbb{1}_{\mathcal{B}_1}|_{\mathcal{B}_2^c} = 0$。这可用于定义 $\tilde{g}(x,y) = \mathbb{1}_{\mathcal{B}_1}(x,y)g(x,y)$，并将其延拓至 $y \notin \mathcal{B}$，延拓部分取值为 0。因此，延拓后的 $\tilde{g}$ 属于 $C^1(\mathbb{R}^{d_x} \times \mathbb{R}^{d_y}; \mathbb{R}^{d_g})$，且其输出和导数在集合 K 上与 g 一致。于是，在使用传统通用逼近定理时，只需考虑对 $\tilde{g}$ 的神经网络近似 g_θ，而非 g。

引理 C.6（半有界集合上的有界近似）。设 $g \in C^1(\mathbb{R}^{d_x} \times \mathcal{B};\ \mathbb{R}^{d_g})$，其中 $\mathcal{B} \subset \mathbb{R}^{d_y}$ 是开集。那么，对于任意紧子集 $K_y \subset \mathcal{B}$、$\epsilon > 0$、$R > 1$ 和 $p \geq 1$，存在一个三层 *GELU* 神经网络 g_θ，使得以下条件同时成立

$$\sup_{\|x\|_2 \leq R, y \in K_y} \|g(x,y) - g_\theta(x,y)\|_2 \leq \epsilon, \tag{C.18}$$

and

$$\sup_{\|x\|_2 \leq R, y \in K_y} \|D_y g(x,y) - D_y g_\theta(x,y)\|_F \leq \epsilon. \tag{C.19}$$

此外，g_θ 可以选取得满足

$$\|g(x,y) - g_\theta(x,y)\|_2^p \leq C_p(\|g(x,y)\|_2^p + 1) \tag{C.20}$$

and

$$\|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p \leq C_p(\|D_y g(x,y)\|_F^p + 1) \tag{C.21}$$

对于所有 $x \in \mathbb{R}^{d_x}$ 和 $y \in K_y$，其中 $C_p > 0$ 是一个固定常数，不依赖于 ϵ 或 R .

证明。我们的策略如下。首先，我们使用引理 C.5 在一个紧集上构造一个有界近似 $\tilde{g}_\theta$ 来近似 g。随后，我们通过（近似地）将近似 $\tilde{g}_\theta$ 在 x 方向上的增长控制在 ˜范围内，方法是将其乘以一个截断函数 f_{cut}。

为此，给定 g，ϵ，R。对于任意 $\epsilon_g > 0$，引理 C.5 允许我们构造 $\tilde{g}_\theta$，它是 g 在紧集 K_x 上的有界近似: $= \{\|x\|_2 \leq 4(\sqrt{d_x}+1)R\}$ 和 K_y，使得

$$\sup_{(x,y) \in K_x \times K_y} \|g(x,y) - \tilde{g}_\theta(x,y)\|_2 \leq \epsilon_g, \qquad \sup_{(x,y) \in K_x \times K_y} \|D_y g(x,y) - D_y \tilde{g}_\theta(x,y)\|_F \leq \epsilon_g,$$

and

$$\sup_{(x,y) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}} \|\tilde{g}_\theta(x,y)\|_2 \leq M_g, \qquad \sup_{(x,y) \in \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}} \|D_y \tilde{g}_\theta(x,y)\|_F \leq M_g,$$

注意到 $M_g = M_g(\epsilon_g)$。对于任意的 $\epsilon_{\text{cut}} > 0$，我们还考虑引理C.4 中的截断函数 $f_{\text{cut}}(x)$，满足

$$\begin{cases} |f_{\text{cut}}(x) - 1| \leq \epsilon_{\text{cut}} & \|x\|_2 \leq R \\ |f_{\text{cut}}(x)| \leq \epsilon_{\text{cut}} & \|x\|_2 \geq 5(\sqrt{d_x}+1)R \\ |f_{\text{cut}}(x)| \leq M_{\text{cut}} & x \in \mathbb{R}^{d_x}, \end{cases}$$

其中 M_{cut} 不依赖于 ϵ_{cut} 或 R。最后，我们考虑乘法函数 $\times(t_1,\ t_2) = t_1 t_2$ 的单层神经网络近似 $f_\times$，使得

$$\|f_\times - \times(\cdot,\cdot)\|_{C^1(B)} \leq \epsilon_\times,$$

对于任意的 $\epsilon_\times > 0$，其中 $B = [-M_{\text{cut}}, M_{\text{cut}}] \times [-M_g, M_g]$。这同样可以通过标准的普适近似结论实现 [98]。然后我们取

$$g_\theta(x,y) := f_\times(f_{\text{cut}}(x), \tilde{g}_\theta(x,y)) \approx f_{\text{cut}}(x)\tilde{g}_\theta(x,y),$$

即 ˜是一个三层神经网络，因为 f_{cut} 是一个两层神经网络。这具有导数 e

$$D_y g_\theta(x,y) = D_2 f_\times(f_{\text{cut}}(x), \tilde{g}_\theta(x,y)) D_y \tilde{g}_\theta(x,y),$$

其中 D_2 表示对第二个变量进行偏微分。注意，这种构造中， $f_\times$ 是根据 $f_{\text{cut}}(x)$ 的输出范围选取的，而 $\tilde{g}$ 确保了

$$\|f_{\text{cut}}(x)\tilde{g}(x,y) - f_\times(f_{\text{cut}}(x), \tilde{g}(x,y))\|_2 \le \epsilon_\times$$

and

$$\|f_{\text{cut}}(x)I - D_2 f_\times(f_{\text{cut}}(x), \tilde{g}(x,y))\|_F \le \epsilon_\times$$

对于所有 $x \in \mathbb{R}^{d_x}$ 和 $y \in \mathbb{R}^{d_y}$。

现在，给定 $x \in \mathbb{R}^{d_x}$，我们考虑以下情况： $y \in K_y$

1. $\|x\|_2 \le R_\circ$ 根据三角不等式，我们有

$$\begin{aligned}&\|g(x,y) - g_\theta(x,y)\|_2\\ &\quad\le \|g(x,y) - \tilde{g}_\theta(x,y)\|_2 + \|(1 - f_{\text{cut}}(x))\tilde{g}_\theta(x,y)\|_2 + \|f_{\text{cut}}(x)\tilde{g}_\theta(x,y) - f_\times(f_{\text{cut}}(x), \tilde{g}_\theta(x,y))\|_2.\end{aligned}$$

由于 $\|x\|_2 \le R$，我们有 $\|g(x,\ y) - g_\theta(x,\ y)\|_2 \le \epsilon_g$ 和 $|1 - f_{\text{cut}}(x)| \le \epsilon_{\text{cut}\circ}$ 因此，

$$\|g(x,y) - g_\theta(x,y)\|_2 \le \epsilon_g + \epsilon_{\text{cut}} M_g + \epsilon_\times.$$

通过类似的推理，我们有

$$\begin{aligned}\|D_y g(x,y) - D_y g_\theta(x,y)\|_F &\le \|D_y g(x,y) - D_y \tilde{g}_\theta(x,y)\|_F\\ &\quad + \|(1 - f_{\text{cut}}(x))D_y\tilde{g}_\theta(x,y)\|_F\\ &\quad + \|f_{\text{cut}}(x)D_y\tilde{g}_\theta(x,y) - D_2 f_\times(f_{\text{cut}}(x), \tilde{g}_\theta(x,y))D_y\tilde{g}_\theta(x,y)\|_F\\ &\le \epsilon_g + \epsilon_{\text{cut}} M_g + \epsilon_\times M_g.\end{aligned}$$

2. $\|x\|_2 \ge 5(\sqrt{d_x} + 1)R_\circ$ 我们有

$$\begin{aligned}\|g(x,y) - g_\theta(x,y)\|_2 &\le \|g(x,y)\|_2 + \|g_\theta(x,y)\|_2\\ &\le \|g(x,y)\|_2 + |f_{\text{cut}}(x)|\|\tilde{g}_\theta(x,y)\|_2 + \|f_\times(f_{\text{cut}}(x), \tilde{g}_\theta(x,y)) - f_{\text{cut}}(x)\tilde{g}_\theta(x,y)\|_2\\ &\le \|g(x,y)\|_2 + \epsilon_{\text{cut}} M_g + \epsilon_\times\end{aligned}$$

and

$$\begin{aligned}&\|D_y g(x,y)\ - D_y g_\theta(x,y)\|_F \le \|D_y g(x,y)\|_F + \|D_y g_\theta(x,y)\|_F \le \|D_y g(x,y)\|_F + \|f_{\text{cut}}(x)D_y\tilde{g}_\theta(x,y)\\ &\|_F + \|f_{\text{cut}}(x)D_y\tilde{g}_\theta(x,y)\ - D_2 f_\times(f_{\text{cut}}(x)\ \tilde{g}_\theta(x,y))D_y\tilde{g}_\theta(x,y)\|_F \le \|D_y g(x,y)\|_F + \epsilon_{\text{cut}} M_g + \epsilon_\times M_g.\end{aligned}$$

3. $\|x\|_2 < 5(\sqrt{d_x} + 1)R_\circ$ 我们有

$$\begin{aligned}\|g(x,y) - g_\theta(x,y)\|_2 &\le \|g(x,y)\|_2 + \|g_\theta(x,y)\|_2\\ &\le \|g(x,y)\|_2 + |f_{\text{cut}}(x)|\|\tilde{g}_\theta(x,y)\|_2 + \|f_\times(f_{\text{cut}}(x), \tilde{g}_\theta(x,y)) - f_{\text{cut}}(x)\tilde{g}_\theta(x,y)\|_2\\ &\le \|g(x,y)\|_2 + M_{\text{cut}}(\|g(x,y)\|_2 + \epsilon_g) + \epsilon_\times,\end{aligned}$$

and

$$\begin{aligned}&\|D_y g(x,y)\ - D_y g_\theta(x,y)\|_F \le \|D_y g(x,y)\|_F + \|D_y g_\theta(x,y)\|_F \le \|D_y g(x,y)\|_F + \|f_{\text{cut}}(x)D_y\tilde{g}_\theta(x,y)\\ &\|_F + \|f_{\text{cut}}(x)D_y\tilde{g}_\theta(x,y)\ - D_2 f_\times(f_{\text{cut}}(x)\ \tilde{g}_\theta(x,y))D_y\tilde{g}_\theta(x,y)\|_F \le \|D_y g(x,y)\|_F + M_{\text{cut}}(\|D_y g(x,y)\|_F + \epsilon_g)\\ &+ \epsilon_\times M_g.\end{aligned}$$

因此，只需选择具有相应误差容限的近似

$$\epsilon_g \le \min\left\{\frac{\epsilon}{3}, \frac{1}{2M_{\text{cut}}}\right\}, \qquad \epsilon_{\text{cut}} \le \min\left\{\frac{\epsilon}{3M_g}, \frac{1}{2M_g}\right\}, \qquad \epsilon_\times \le \min\left\{\frac{\epsilon}{3}, \frac{\epsilon}{3M_g}, \frac{1}{2}, \frac{1}{2M_g}\right\}.$$

在这些选择下，误差满足 (C.18)–(C.21)，其中 $p = 1$ 使用常数 $C_1 := 1 + M_{\text{cut}}$。现在，对于任意 $p \ge 1$，有

$$\|g(x,y) - g_\theta(x,y)\|_2^p \le C_1^p \left(\|g(x,y)\|_2 + 1\right)^p \le 2^{p-1} C_1^p \left(\|g(x,y)\|_2^p + 1\right),$$

其中我们用到了不等式 $(a+b)^p \le 2^{p-1}(a^p + b^p)$。因此，对于任意 $p \ge 1$，(C.20) 和 (C.21) 均成立，其中 $C_p = 2^{p-1} C_1^p$。 □

C.1.3. 定理C.1的证明

现在我们可以完成定理 C.1的证明。

证明。 我们考虑神经网络近似 g_θ，它们属于满足引理 C.6的近似族。给定这样的 g_θ，考虑特定 $y \in K_y$对应的误差，即

$$\mathbb{E}_{x\sim\mu_x}[\|g(x,y) - g_\theta(x,y)\|_2^p] = \int \|g(x,y) - g_\theta(x,y)\|_2^p d\mu_x(x),$$
$$\mathbb{E}_{x\sim\mu_x}[\|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p] = \int \|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p d\mu_x(x).$$

对于任意 $R > 1$，令 $B_R := \{\|x\|_2 \le R\}$ 和 $B_R^c := \{\|x\|_2 > R\}$。我们可以将积分分解为

$$\int \|g(x,y) - g_\theta(x,y)\|_2^p d\mu_x(x) = \int_{B_R} \|g(x,y) - g_\theta(x,y)\|_2^p d\mu_x(x) + \int_{B_R^c} \|g(x,y) - g_\theta(x,y)\|_2^p d\mu_x(x)$$

and

$$\int \|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p d\mu_x(x), = \int_{B_R} \|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p d\mu_x(x) + \int_{B_R^c} \|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p d\mu_x(x).$$

特别地，对于 g_θ 由引理C.6得出，我们可以利用误差的全局界将其写为

$$\int \|g(x,y) - g_\theta(x,y)\|_2^p d\mu_x(x) \le \int_{B_R} \|g(x,y) - g_\theta(x,y)\|_2^p d\mu_x(x) + C_p \int_{B_R^c} \left(\|g(x,y)\|_2^p + 1\right) d\mu_x(x) \quad \text{(C.22)}$$

and

$$\int \|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p d\mu_x(x) \le \int_{B_R} \|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p d\mu_x(x) + C_p \int_{B_R^c} \left(\|D_y g(x,y)\|_F^p + 1\right) d\mu_x(x), \quad \text{(C.23)}$$

其中 C_p 与 g_θ的具体选择无关。

近似策略如下。我们首先选择 $R > 1$ 足够大，使得误差

$$e_R^0(y) := C_p \int_{B_R^c} \left(\|g(x,y)\|_2^p + 1\right) d\mu_x(x)$$
$$e_R^1(y) := C_p \int_{B_R^c} \left(\|D_y g(x,y)\|_F^p + 1\right) d\mu_x(x)$$

在 $y \in K_y$ 上一致较小。随后，对于选定的 R，我们选择 g_θ，使其能够在紧集 $B_R \times K_y$ 上准确近似 g，从而将总误差控制在 ϵ^p 以内。

我们首先证明，当 $R \to \infty$ 时，$e_R^0(y)$ 和 $e_R^1(y)$ 在 K_y 上一致收敛到零。先考虑输出误差 $e_R^0(y)$。根据控制收敛定理，对于所有 $y \in \mathcal{B}$，有 $e_R^0(y) \downarrow 0$，当 $R \to \infty$ 时，其中 $\downarrow$ 表示收敛是单调的，即 $e_R^0 \to 0$，并且只要 $R_1 \leq R_2$，就有 $e_{R_1}^0(y) \geq e_{R_2}^{(0)}(y)$。此外，$e_R^0(y)$ 是连续的。为此，首先注意到 $e_R^0(y)$ 可以分解为两个积分，

$$e_R^0(y) = C_p \int_{B_R^c} \|g(x,y)\|_2^p d\mu_x(x) + C_p \int_{B_R^c} 1 d\mu_x(x).$$

项 $C_p \int_{B_R^c} 1 d\mu_x(x) = C_p \mu_x(B_R^c)$ 关于 y 显然是连续的。因此，我们将重点放在其连续性上：

$$y \mapsto C_p \int_{B_R^c} \|g(x,y)\|_2^p d\mu_x(x) = C_p \|\mathbb{1}_{B_R^c} g(\cdot, y)\|_{L^p_{\mu_x}}^p, \tag{C.24}$$

其中 $\mathbb{1}_{B_R^c}$ 表示集合 B_R^c 的示性函数。为此，我们考虑 $y_1, y_2 \in \mathcal{B}$。由反三角不等式可得

$$\left| \|\mathbb{1}_{B_R^c}(\cdot) g(\cdot, y_2)\|_{L^p_{\mu_x}} - \|\mathbb{1}_{B_R^c}(\cdot) g(\cdot, y_1)\|_{L^p_{\mu_x}} \right| \leq \|\mathbb{1}_{B_R^c}(\cdot) \left(g(\cdot, y_2) - g(\cdot, y_1)\right)\|_{L^p_{\mu_x}}. \leq \|g(\cdot, y_2) - g(\cdot, y_1)\|_{L^p_{\mu_x}}.$$

由映射 $\mathcal{B} \ni y \mapsto g(\cdot, y_2) \in L^p_{\mu_x}$ 的连续性假设可知

$$\|g(\cdot, y_2) - g(\cdot, y_1)\|_{L^p_{\mu_x}} \to 0 \qquad \text{as } y_2 \to y_1.$$

因此，映射 $\mathcal{B} \ni y \mapsto \mathbb{1}_{B_R^c} g(\cdot, y) \in L^p_{\mu_x}$ 是连续的，进而映射 (C.24) 也是连续的。根据狄尼定理，$e_R^0(y) \downarrow 0$ 在 $y \in K_y \Subset \mathcal{B}$上一致收敛。采用类似的推理，我们还可以证明 $e_R^1(y)$也具有同样的性质。

现在，令 $\epsilon > 0$ 为任意值。根据上述结果，存在一个 $R > 1$，使得

$$\sup_{y \in K_y} e_R^0(y) + e_R^1(y) \leq \frac{\epsilon^p}{2}. \tag{C.25}$$

对于这个 $R > 1$，引理 C.6 保证存在一个 g_θ，使得

$$\sup_{x \in B_R, y \in K_y} \|g(x,y) - g_\theta(x,y)\|_2^p + \|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p \leq \frac{\epsilon^p}{2}. \tag{C.26}$$

将 (C.26) 和 (C.25) 代入 (C.22) 和 (C.23)，得到

$$\int \|g(x,y) - g_\theta(x,y)\|_2^p + \|D_y g(x,y) - D_y g_\theta(x,y)\|_F^p d\mu_x(x) \leq \epsilon^p$$

对每个 $y \in K_y$ 均成立，这就是所需的结果。

□

*C.2. Lipschitz*光滑算子的降维误差

回顾一下，我们关注的是降阶基架构的近似能力

$$u_\theta(m, z) = \mathbf{\Phi}_{r_u} g_\theta(\mathbf{\Psi}_{r_m}^*(m - b_m), z) + b_u,$$

其中 $\mathbf{\Phi}_{r_u}$ 和 $\mathbf{\Psi}_{r_m}$ 分别表示到 $\mathcal{U}$ 和$\mathcal{M}$的降阶正交归一基$\{\phi_i\}_{i=1}^{r_u}$ 和$\{\psi_i\}_{i=1}^{r_m}$ 上的解码。我们首先证明，对于任意正交归一基 $\{\phi_i\}_{i=1}^\infty$ 和 $\{\psi_i\}_{i=1}^\infty$，通过选取足够大的 r_u 和 r_m，可使降维误差任意小。我们从输出情形开始。

引理 C.7（输出维数约简误差）。假设 u 满足假设2.1，其中对某个 $p \geq 2$采用博雷尔概率测度 μ_m、紧容许集 $\mathcal{Z}_{ad}$和开集 $\mathcal{B} \supset \mathcal{Z}_{ad}$。设$\{\phi_i\}_{i=1}^{\infty}$ 为 $\mathcal{U}$ 的正交归一基，且$P_{\mathcal{U},\,r} = \sum_{i=1}^{r} \phi_i \phi_i^T$为到约化基 $\{\phi_i\}_{i=1}^{r}$上的正交投影。则对任意 $\epsilon > 0$，存在 $r_\epsilon \in \mathbb{N}$ 使得同时

$$\sup_{z \in \mathcal{Z}_{ad}} \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{U},r})u(m,z)\|_{\mathcal{U}}^p] \leq \epsilon^p$$

and

$$\sup_{z \in \mathcal{Z}_{ad}} \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{U},r})D_z u(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \leq \epsilon^p,$$

只要 $r \geq r_\epsilon$。

证明。首先考虑输出投影误差。我们定义

$$e_r^0(z) := \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{U},r})u(m,z)\|_{\mathcal{U}}^p].$$

我们声称，对于每个 $r \in \mathbb{N}$， $e_r^0(z)$ 在紧集 $\mathcal{Z}_{ad}$ 上连续，并且对于每个 $z \in \mathcal{Z}_{ad}$，单调收敛至零。根据狄尼定理，这便意味着在$\mathcal{Z}_{ad}$上一致收敛至零。

为说明连续性，我们考虑映射 $\mathcal{B} \ni z \mapsto (I - P_{\mathcal{U},r})u(\cdot,z) \in L_{\mu_m}^p(\mathcal{M};\mathcal{U})$。由于 $(I - P_{\mathcal{U},r})$ 是一个投影，对于每个 $z_1, z_2 \in \mathcal{B}$，我们有

$$\|(I - P_{\mathcal{U},r})u(\cdot,z_2) - (I - P_{\mathcal{U},r})u(\cdot,z_1)\|_{L_{\mu_m}^p}^p \leq \|u(\cdot,z_2) - u(\cdot,z_1)\|_{L_{\mu_m}^p}^p. \tag{C.27}$$

根据命题 B.1， $\mathcal{B} \ni z \mapsto u(\cdot,z) \in L_{\mu_m}^p(\mathcal{M};\mathcal{U})$ 连续，因此结合 (C.27) 可知， $\mathcal{B} \ni z \mapsto (I - P_{\mathcal{U},r})u(\cdot,z) \in L_{\mu_m}^p(\mathcal{M};\mathcal{U})$ 也连续。

为说明逐点单调收敛，我们注意到，对任意 $m \in \mathcal{M}$ 和 $z \in \mathcal{Z}_{ad}$，都有 $\|(I - P_{\mathcal{U},r})u(m,z)\|_{\mathcal{U}}^p \downarrow 0$，当 $r \to \infty$ 时，其中 $\downarrow$ 再次表示单调收敛。这意味着误差序列单调递减，因为

$$e_{r_1}^0(z) = \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{U},r_1})u(m,z)\|_{\mathcal{U}}^p] \leq \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{U},r_2})u(m,z)\|_{\mathcal{U}}^p] = e_{r_2}^0(z)$$

当 $r_1 \geq r_2$ 时。此外，由于被积函数有如下界

$$\|(I - P_{\mathcal{U},r})u(m,z)\|_{\mathcal{U}}^p \leq \|u(m,z)\|_{\mathcal{U}}^p,$$

我们有控制函数 $\|u(\cdot,z)\|_{\mathcal{U}}^p$。由于 $u(\cdot,z) \in L_{\mu_m}^p(\mathcal{M};\mathcal{U})$，该控制函数可积。因此，由控制收敛定理可知，对于每个 $z \in \mathcal{Z}_{ad}$，都有 $e_r^0(z) \downarrow 0$，

对导数误差也可以采用类似的步骤。也就是说，我们可以定义

$$e_r^1(z) := \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{U},r})D_z u(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p],$$

由命题 B.1可知，该函数同样连续。此外，我们可以将希尔伯特-施密特范数改写为

$$\|(I - P_{\mathcal{U},r})D_z u(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p = \|D_z u(m,z)^*(I - P_{\mathcal{U},r})\|_{\mathrm{HS}(\mathcal{U},\mathcal{Z})}^p = \left(\sum_{i>r}^{\infty} \|D_z u(m,z)^* \phi_i\|_{\mathcal{Z}}^2\right)^{p/2},$$

对每个 $z \in \mathcal{Z}_{ad}$，当 $r \to \infty$ 时，该式单调收敛至零，且受可积函数 $\|D_z u(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p$ 控制。再次结合控制收敛定理和狄尼定理可知，对于 $z \in \mathcal{B}$， $e_r^1(z) \downarrow 0$ 一致。

由于 e_r^0 和 e_r^1 的收敛均为单调且一致的，因此只需选取 r_ϵ，使得两个误差都足够小，即：

$$\sup_{z \in \mathcal{Z}_{ad}} e_{r_\epsilon}^0(z) \leq \epsilon^p, \quad \sup_{z \in \mathcal{Z}_{ad}} e_{r_\epsilon}^1 \leq \epsilon^p.$$

那么，对于任意 $r \geq r_\epsilon$，误差界均成立。

□

引理 C.8 (部分输入降维)。 假设 u 满足假设 *2.1*，其博雷尔概率测度为 μ_m，紧容许集为 $\mathcal{Z}_{ad}$，并且对于某个 $p \geq 2$，开集为 $\mathcal{B} \supset \mathcal{Z}_{ad}$。设 $\{\psi_i\}_{i=1}^{\infty}$ 是 $\mathcal{M}$ 的一组正交归一基，令 $P_{\mathcal{M},\ r} = \sum_{i=1}^{r} \psi_i \psi_i^T$ 为投影到约化基$\{\psi_i\}_{i=1}^{r}$上的正交投影。那么，对于任意 $\epsilon > 0$，存在 $r_\epsilon \in \mathbb{N}$，使得以下结论同时成立

$$\sup_{z \in \mathcal{Z}_{ad}} \mathbb{E}_{m \sim \mu_m}[\|u(m,z) - u(P_{\mathcal{M},r}m, z)\|_{\mathcal{U}}^p] \leq \epsilon^p$$

and

$$\sup_{z \in \mathcal{Z}_{ad}} \mathbb{E}_{m \sim \mu_m}[\|D_z u(m,z) - D_z u(P_{\mathcal{M},r}m, z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \leq \epsilon^p$$

只要 $r \geq r_\epsilon$

证明。 根据假设 2.1中的利普希茨连续性，我们有

$$\mathbb{E}_{m \sim \mu_m}[\|u(m,z) - u(P_{\mathcal{M},r}m, z)\|_{\mathcal{U}}^p\| \leq \kappa_u^p \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{M},r})m\|_{\mathcal{M}}^p].$$

and

$$\mathbb{E}_{m \sim \mu_m}[\|D_z u(m,z) - D_z u(P_{\mathcal{M},r}m, z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \leq \kappa_{u_z}^p \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{M},r})m\|_{\mathcal{M}}^p].$$

对于上界，被积函数 $\|(I - P_{\mathcal{M},r})m\|_{\mathcal{M}}^p \downarrow 0$ 当 $r \to \infty$ 时趋于 $m \in \mathcal{M}$，对每个 $m \in \mathcal{M}$均成立。此外，$\|(I - P_{\mathcal{M},r})m\|_{\mathcal{M}}^p \leq \|m\|_{\mathcal{M}}^p$，其中根据我们对 μ_m 的矩假设，假定 $\|m\|_{\mathcal{M}}^p$ 可积。因此，上界 $\mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{M},r})m\|_{\mathcal{M}}^p] \downarrow 0$ 当 $r \to \infty$时趋于 $r \to \infty$。只需选择足够大的 r_ϵ，使得

$$\mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{M},r})m\|_{\mathcal{M}}^p] \leq \epsilon^p \min\left\{\frac{1}{\kappa_u^p}, \frac{1}{\kappa_{u_z}^p}\right\},$$

对于该值，所需的误差界成立。 □

引理 C.9 (联合截断误差)。 假设 u 满足假设 *2.1*，其博雷尔概率测度为μ_m，紧容许集为 $\mathcal{Z}_{ad}$，开集为 $\mathcal{B} \supset \mathcal{Z}_{ad}$，其中 $p \geq 2$。设 $\{\phi_i\}_{i=1}^{\infty}$ 和 $\{\psi_i\}_{i=1}^{\infty}$ 分别为 $\mathcal{M}$ 和 $\mathcal{U}$的标准正交基。则对于任意 $\epsilon > 0$，存在秩 r_m 和 r_u，使截断算子 $u_r(m,z) := P_{\mathcal{U},r_u} u(P_{\mathcal{M},r_m} m, z)$ 满足

$$\sup_{z \in \mathcal{Z}_{ad}} \mathbb{E}_{m \sim \mu_m}[\|u(m,z) - u_r(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m \sim \mu_m}[\|D_z u(m,z) - D_z u_r(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \leq \epsilon^p. \tag{C.28}$$

证明。 这是引理 C.7 和 C.8的推论。根据三角不等式，有

$$\|u(m,z) - u_r(m,z)\|_{\mathcal{U}} \leq \|u(m,z) - P_{\mathcal{U},r_u} u(m,z)\|_{\mathcal{U}} + \|P_{\mathcal{U},r_u} u(m,z) - P_{\mathcal{U},r_u} u(P_{\mathcal{M},r_m} m, z)\|_{\mathcal{U}}.$$

由于 $P_{\mathcal{U},\ r_u}$ 是正交投影，该界可简化为

$$\|u(m,z) - u_r(m,z)\|_{\mathcal{U}} \leq \|u(m,z) - P_{\mathcal{U},r_u} u(m,z)\|_{\mathcal{U}} + \|u(m,z) - u(P_{\mathcal{M},r_m} m, z)\|_{\mathcal{U}}.$$

类似地，对于导数，我们有

$$\begin{aligned}
&\|D_z u(m,z) - D_z u_r(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})} \\
&\quad \leq \|(I - P_{\mathcal{U},r_u}) D_z u(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})} + \|D_z u(m,z) - D_z u(P_{\mathcal{M},r_m} m, z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}.
\end{aligned}$$

利用不等式 $(a+b)^{\ p} \leq 2^{p-1}(a^p + b^p)$，我们可得，对于和 $z \in \mathcal{Z}_{ad}$,

$$\begin{aligned}
&\mathbb{E}_{m \sim \mu_m}[\|u(m,z) - u_r(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m \sim \mu_m}[\|D_z u(m,z) - D_z u_r(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \\
&\quad \leq 2^{p-1}\left(\mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{U},r_u})u(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m \sim \mu_m}[\|(I - P_{\mathcal{U},r_u}) D_z u(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p]\right) \\
&\quad + 2^{p-1}\left(\mathbb{E}_{m \sim \mu_m}[\|u(m,z) - u(P_{\mathcal{M},r_m} m, z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m \sim \mu_m}[\|D_z u(m,z) - D_z u(P_{\mathcal{M},r_m} m, z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p]\right).
\end{aligned} \tag{C.29}$$

根据引理 C.7，我们可以选取 r_u 使得

$$\sup_{z\in\mathcal{Z}_{ad}} \mathbb{E}_{m\sim\mu_m}[\|(I-P_{\mathcal{U},r_u})u(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m\sim\mu_m}[\|(I-P_{\mathcal{U},r_u})D_z u(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \le \frac{\epsilon^p}{2^p}, \tag{C.30}$$

而由引理C.8，我们可以选取 r_m 使得

$$\sup_{z\in\mathcal{Z}_{ad}} \mathbb{E}_{m\sim\mu_m}[\|u(m,z)-u(P_{\mathcal{M},r_m}m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m\sim\mu_m}[\|D_z u(m,z)-D_z u(P_{\mathcal{M},r_m}m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \le \frac{\epsilon^p}{2^p}. \tag{C.31}$$

对（C.29）取上确界，并代入（C.30）和（C.31），可得

$$\sup_{z\in\mathcal{Z}_{ad}} \mathbb{E}_{m\sim\mu_m}[\|u(m,z)-u_r(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m\sim\mu_m}[\|D_z u(m,z)-D_z u_r(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \le \epsilon^p.$$

□

C.3. 定理 *3.1* 的证明

我们现在可以将降维误差的界与引理 C.9 以及神经网络近似结果定理C.1 结合起来，以证明定理 3.1。

证明。（定理3.1）我们首先假设 $b_m = b_u = 0$。$b_m \neq 0$ 与 $b_u \neq 0$的情形作为推论得出。在此假设下，令 $u_r(m,z) := P_{\mathcal{U},r_u}u(P_{\mathcal{M},r_m}m,z) = \mathbf{\Phi}_{r_u}\mathbf{\Phi}_{r_u}^* u(\mathbf{\Psi}_{r_m}\mathbf{\Psi}_{r_m}^* m,z)$，我们将逼近误差分解为截断误差 $u(m,z) - u_r(m,z)$ 与潜在近似误差 $u_r(m,z) - u_\theta(m,z)$。即，

$$\begin{aligned}
&\mathbb{E}_{m\sim\mu_m}[\|u(m,z)-u_\theta(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m\sim\mu_m}[\|D_z u(m,z)-D_z u_\theta(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \\
&\quad\le 2^{p-1}\left(\mathbb{E}_{m\sim\mu_m}[\|u(m,z)-u_r(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m\sim\mu_m}[\|D_z u(m,z)-D_z u_r(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p]\right) \\
&\quad+ 2^{p-1}\left(\mathbb{E}_{m\sim\mu_m}[\|u_r(m,z)-u_\theta(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m\sim\mu_m}[\|D_z u_r(m,z)-D_z u_\theta(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p]\right).
\end{aligned} \tag{C.32}$$

由引理 C.9，存在 r_u 和 r_m 使得

$$\sup_{z\in\mathcal{Z}_{ad}} \mathbb{E}_{m\sim\mu_m}[\|u(m,z)-u_r(m,z)\|_{\mathcal{U}}^p] + \mathbb{E}_{m\sim\mu_m}[\|D_z u(m,z)-D_z u_r(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p] \le \frac{\epsilon^p}{2^p}. \tag{C.33}$$

我们现在针对所选的 r_u 和 r_m，寻求 u_r 的一个近似。为此，我们定义映射g： $\mathbb{R}^{r_m}\times\mathcal{B}\to\mathbb{R}^{r_u}$，

$$g(x,z) = \mathbf{\Phi}_{r_u}^* u(\mathbf{\Psi}_{r_m}x, z), \tag{C.34}$$

使得

$$u_r(m,z) = \mathbf{\Phi}_{r_u} g(\mathbf{\Psi}_{r_m}^* m, z). \tag{C.35}$$

我们可以验证 g 满足定理 C.1中的假设。首先，由于 u 是连续可微的，故 g 也是连续可微的。此外，映射 $\mathcal{B}\ni z\mapsto g(\cdot,z)\in L^2_{\mu_r}(\mathbb{R}^{r_m};\mathbb{R}^{r_u})$和$\mathcal{B}\ni z\mapsto D_z g(\cdot,z)\in L^2_{\mu_r}(\mathbb{R}^{r_m};\mathbb{R}^{r_u\times d_z})$关于 z是利普希茨连续的，其中 $\mu_r = \left(\mathbf{\Psi}_{r_m}^*\right)^\sharp \mu_m$表示 μ_m 在投影 $\mathbf{\Psi}_{r_m}^*$下的推前测度。为看出这一点，考虑 z_1、$z_2\in\mathcal{B}$以及该差

$$\|g(\cdot,z_2)-g(\cdot,z_1)\|_{L^p_{\mu_m}}^p = \int \|g(x,z_2)-g(x,z_1)\|_2^p d\mu_r(x).$$

使用 defin　　对 g 的ition以及变量替换 $x = \mathbf{\Psi}_{r_m}^* m$（推前测度为 μ_r），我们有

$$\int \|g(x,z_2)-g(x,z_1)\|_2^p d\mu_r(x) = \int \|\mathbf{\Phi}_{r_u}^* u(\mathbf{\Psi}_{r_m}\mathbf{\Psi}_{r_m}^* m,z_2) - \mathbf{\Phi}_{r_u}^* u(\mathbf{\Psi}_{r_m}\mathbf{\Psi}_{r_m}^* m,z_2)\|_2^p d\mu_m(m).$$

由于 $\mathbf{\Phi}_{r_u}$ 是正交归一基，我们有$\|\mathbf{\Phi}^*_{r_u}\|_{\mathcal{L}(\mathcal{U},\mathbb{R}^{r_u})} = 1$，使得

$$\int \|g(x,z_2) - g(x,z_1)\|_2^p d\mu_r(x) \le \int \|u(\mathbf{\Psi}_{r_m}\mathbf{\Psi}^*_{r_m} m, z_2) - u(\mathbf{\Psi}_{r_m}\mathbf{\Psi}^*_{r_m} m, z_2)\|_{\mathcal{U}}^p d\mu_m(m) \le \kappa_u^p \|z_2 - z_1\|_{\mathcal{Z}}^p.$$

这意味着 $z \mapsto g(\cdot,\ z)$ 关于 z 是利普希茨连续的。类似地，可以用同样的过程证明 $z \mapsto D_z g(\cdot,\ z)$ 关于 z 也是利普希茨连续的。

这使我们能够应用定理 C.1 的结果。也就是说，给定 $\epsilon > 0$，存在一个神经网络近似 g_θ 对 g 满足

$$\sup_{z\in\mathcal{Z}_{ad}} \mathbb{E}_{x\sim\mu_r}[\|g(x,z) - g_\theta(x,z)\|^p] + \mathbb{E}_{x\sim\mu_r}[\|D_z g(x,z) - D_z g_\theta(x,z)\|_F^p] \le \frac{\epsilon^p}{2^p}. \tag{C.36}$$

那么，$u_r(m,\ z)$ 与 RBNO $u_\theta(m,\ z) = \mathbf{\Phi}_{r_u} g_\theta(\mathbf{\Psi}^*_{r_m},\ z)$ 之间的误差可直接用 g 和 g_θ 表示。也就是说，

$$\begin{aligned}
&\mathbb{E}_{m\sim\mu_m}\left[\|u_r(m,z) - u_\theta(m,z)\|_{\mathcal{U}}^p\right] \\
&\quad= \mathbb{E}_{m\sim\mu_m}\left[\|\mathbf{\Phi}_{r_u}\left(g(\Psi^*_{r_m} m, z) - g_\theta(\Psi^*_{r_m} m, z)\right)\|_{\mathcal{U}}^p\right] \\
&\quad= \mathbb{E}_{x\sim\mu_r}\left[\|g(x,z) - g_\theta(x,z)\|_2^p\right],
\end{aligned} \tag{C.37}$$

and

$$\begin{aligned}
&\mathbb{E}_{m\sim\mu_m}\left[\|D_z u_r(m,z) - D_z u_\theta(m,z)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p\right] \\
&\quad= \mathbb{E}_{m\sim\mu_m}\left[\|\mathbf{\Phi}_{r_u}\left(D_z g(\mathbf{\Psi}^*_{r_m} m, z) - D_z g_\theta(\mathbf{\Psi}^*_{r_m} m, z)\right)\|_{\mathrm{HS}(\mathcal{Z},\mathcal{U})}^p\right] \\
&\quad= \mathbb{E}_{x\sim\mu_r}\left[\|D_z g(x,z) - D_z g(x,z)\|_F^p\right].
\end{aligned} \tag{C.38}$$

将界 (C.33) 和 (C.36) 与 (C.37) 和 (C.38) 结合，并代入 (C.32)，即可得到所需的逼近误差。

当 $b_m \neq 0$ 时，我们有

$$\mathbf{\Psi}^*_{r_m} m = \mathbf{\Psi}^*_{r_m}(m - b_m) + \mathbf{\Psi}^*_{r_m} b_m.$$

因此，对于使用上述过程构造的神经网络 g_θ，我们有

$$g_\theta(\mathbf{\Psi}^*_{r_m} m, z) = g_\theta(\mathbf{\Psi}^*_{r_m}(m - b_m) + \mathbf{\Psi}^*_{r_m} b_m, z) = \tilde{g}_\theta(\mathbf{\Psi}^*_{r_m}(m - b_m), z),$$

其中，$\tilde{g}_\theta$ 也是一个 GELU 神经网络，其输入层偏置通过$\mathbf{\Psi}^*_{r_m} b_m \in \mathbb{R}^{r_m}$进行平移。因此，RBNO 具有所需的表示形式

$$u_\theta(m,z) = \mathbf{\Phi}_{r_m}\tilde{g}_\theta(\mathbf{\Psi}^*_{r_m}(m - b_m), z).$$

最后，当 $b_u \neq 0$时，我们可以构造近似 $\tilde{u}_\theta(m,\ z)$，用于近似平移后的算子 $u(m,\ z) - b_u$，方法与上述步骤相同。注意，$u(m,\ z) - b_u$ 同样满足假设 2.1。于是，$u(m,\ z)$ 的近似可通过仿射平移得到：

$$u_\theta(m,z) = \tilde{u}_\theta(m,z) + b_u = \mathbf{\Phi}_{r_u} g_\theta(\mathbf{\Psi}^*_{r_m}(m - b_m), z) + \bar{u}.$$

□

C.4. 不确定性优化的通用逼近

现在，我们可以结合定理 3.1 和定理 2.1 来证明定理 3.2，该定理保证存在一个 RBNO，其输出和导数足够准确；当将其用作期望最小化中的代理模型（假设目标函数是凸的）时，其驻点也足够准确。

证明。（定理 3.2）首先，我们注意到，任何 GELU 神经网络 g_θ 都是光滑的，并且由于 GELU 激活函数的特性，其各阶导数均有全局界。因此，给定任意降阶基$\mathbf{\Phi}_{r_u}$ 和 $\mathbf{\Psi}_{r_m}$，由其构造的 RBNO u_θ 由 g_θ 构造而成，也同样是光滑的，且导数有界。这意味着 u_θ 满足假设 2.1 ，且开集 $\mathcal{B}$保持不变。

现在考虑任意一个 $0 < \epsilon_u < 1$。根据定理 3.1的通用逼近结果，存在一个 RBNO，使得

$$\sup_{z\in\mathcal{Z}_{ad}} \|u(\cdot,z) - u_\theta(\cdot,z)\|_{L^2_{\mu_m}} \le \epsilon_u, \tag{C.39}$$

and

$$\sup_{z\in\mathcal{Z}_{ad}} \|D_z u(\cdot,z) - D_z u_\theta(\cdot,z)\|_{L^2_{\mu_m}} \le \epsilon_u. \tag{C.40}$$

由于 u 和 u_θ 均满足假设 2.1 ，而 Q满足假设 2.2，命题 2.2 意味着存在一个常数 $C > 0$ ，且与 u_θ 无关，使得

$$\sup_{z\in\mathcal{Z}_{ad}} \|D_z\mathcal{J}(z) - D_z\mathcal{J}_\theta(z)\|_{\mathcal{Z}'} \le C\epsilon_u, \tag{C.41}$$

其中，当 $\epsilon_u < 1$时，线性误差项占主导。根据定理 2.1，我们得到优化误差和最优性间隙的界，分别为

$$\|z^\star - z^\dagger\|_{\mathcal{Z}} \le \frac{C}{\lambda}\epsilon_u, \tag{C.42}$$

$$\mathcal{J}(z^\dagger) - \mathcal{J}(z^\star) \le \frac{C^2}{2\lambda}\epsilon_u^2. \tag{C.43}$$

由此可见，只需使用满足以下条件的 ϵ_u 来构造 u_θ

$$\epsilon_u \le \min\left\{\frac{\lambda\epsilon}{C}, \frac{\sqrt{2\lambda\epsilon}}{C}, 1\right\}.$$

□