机器学习研究论文集第 vvv 卷：1–28，2026年

基于 Koopman 算子和保形覆盖保证的可扩展数据驱动可达性分析与控制

DeveshNath[*1] DNATH7@GATECH.EDU **HaoranYin** HYIN95@GATECH.EDU **GlenChou** CHOU@GATECH.EDU

[*1] [2] 佐治亚理工学院。 [1] 电子与计算机工程学院， [2] 网络安全学院、隐私学院与航空航天工程学院。

arXiv:2601.01076v1 [eess.SY] 3 Jan 2026

摘要

我们提出一种可扩展的、基于可达性的框架，用于对未知非线性动力学进行概率化的数据驱动安全性验证。我们利用 Koopman 理论和神经网络（NN）升维函数学习动力学的近似线性表示，并在该空间中设计线性控制器，以实现对参考轨迹分布的闭环跟踪。在升维空间中高效计算闭环可达集，并借助神经网络验证工具将其映射回原始状态空间。为捕捉 Koopman 动力学与真实系统之间的模型失配，我们采用保形预测生成具有统计有效性的误差界，并据此扩张可达集，确保真实轨迹以用户指定的概率包含在其中。这些误差界可推广至不同参考轨迹，因此无需重新计算即可重复使用。在高维 MuJoCo 任务（11 维 Hopper、28 维 Swimmer）和 12 维四旋翼无人机上的实验表明，与现有方法相比，该方法提升了可达集覆盖率和计算效率，同时降低了保守性。

关键词： 安全关键控制、可达性、Koopman 算子、保形预测

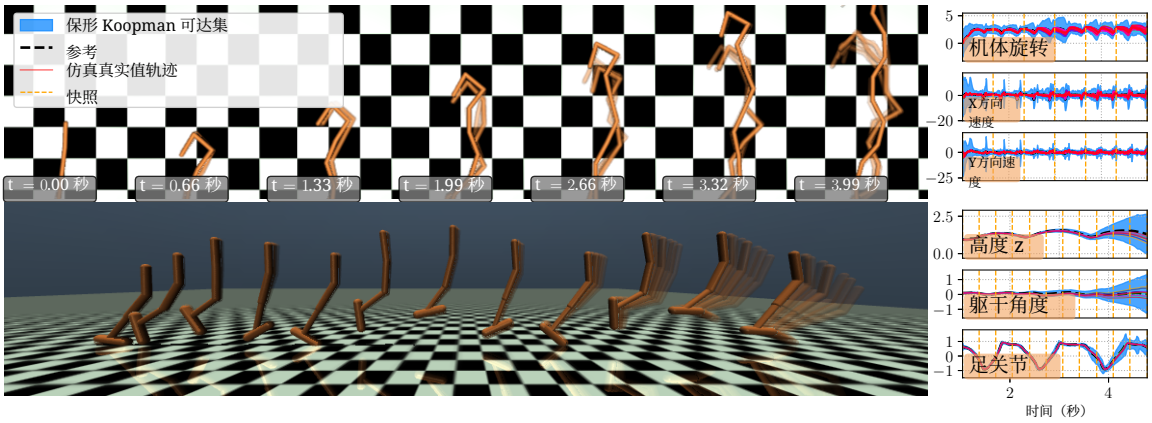


图1： 28维 MuJoCo 游泳者的执行轨迹／可达集（a）；11 维 Hopper 的执行轨迹／可达集（b ）.

1. 引言

计算可达集，即在给定时间范围内可到达的状态集合，是保障机器人安全的关键。然而，机器人的可达性分析颇具难度，因为它们往往（i）缺乏解析模型，（ii）具有高维非线性动力学，且（iii）需要对长时域进行推理。即使具备解析动力学，挑战（ii）和（iii）也会使非线性可达性工具（Bansal等人，2017年；Althoff，2015变得难以承受，或导致过度近似过于保守

© 2026年 D. Nath、H. Yin 和 G. Chou。

在长时域下（Rober and How，2024年）。数据驱动可达性（Park 等人，2024年；Bak 等人，2025）无需解析模型即可使用，但可能数据效率低下，或无法保证估计集合包含所有可能的轨迹。相比之下，线性可达性分析可扩展至长时域和高维情形（Bak 等人，2019）。Koopman 框架在提升状态空间中学习线性动力学（例如通过神经网络（NNs）），以近似非线性系统（Shiet al.，2024年）。虽然针对 Koopman 线性化计算可达集比使用原始非线性动力学更具可扩展性，但现有方法只能为近似的提升动力学提供保证。因此，真实系统的安全性可能无法得到保证（Bak 等人，2021年，2025）。此外，现有的 Koopman 可达性方法只分析单个动力系统。然而，机器人通常需要多个控制器来完成不同任务，这使得现有方法必须针对每个新控制器从头重新计算可达集，效率低下。重用先前计算的信息有助于在不同控制器之间更高效地进行可达性分析。

为弥补可扩展性、速度和可复用性方面的不足，我们提出了 **ScaRe-Kro**（基于 Koopman 算子的可扩展可达性分析），该方法结合 Koopman 理论与神经网络升维函数，进行可达性分析和控制。与以往针对自主系统的 Koopman 可达性方法不同，我们还在升维空间中设计了线性控制器，用于跟踪由参考轨迹构成的分布，其中每条轨迹对应不同的任务。ScaRe-Kro 利用升维的 Koopman 空间中诱导的线性跟踪误差动力学，高效计算闭环可达集，并通过神经网络验证工具将其映射到原始状态空间。为考虑 Koopman 模型误差，我们采用保形预测（CP）推导具有统计有效性的误差界。通过使用这些 CP 误差界扩张映射后的可达集，可保证其以用户指定的概率（例如 97.5%）包含真实系统轨迹。此外，由于 CP 误差界是在轨迹分布上校准的，因此可在不同参考轨迹及相应闭环系统之间复用；这不同于以往仅能认证单个自主系统的 Koopman 方法。总体而言，我们的方法结合线性可达性分析与 CP，实现了可扩展的数据驱动概率验证与控制，在效率和保守性方面均优于以往工作。我们的贡献如下：

1. 一种可扩展的、基于 Koopman 的未知非线性动力学可达性分析框架，利用神经网络提升函数和神经网络验证，在原始状态空间中估计可达集。2. 一种利用保形预测膨胀所得 Koopman 可达集的方法，从而为真实动力学的过度近似提供概率保证。3. 在升维空间中进行控制器设计，以闭环跟踪参考轨迹分布，并使用基于保形预测的膨胀界限，该界限经过校准，可在整个分布中重复使用。4. 通过广泛评估证明，该方法可对高维机器人系统（最高达 28 维）进行可靠验证，并在安全率、计算速度和保守性方面优于基线方法。

2. 相关工作

基于线性松弛的可达性 虽然非线性系统的精确可达集计算通常难以实现，但通常可以高效计算可达集过近似（RSOA），以验证安全性（Althoff 等人，2021年）。当闭环系统包含神经网络组件（例如学习得到的动力学或控制器）时，可以通过基于线性松弛的神经网络验证（NNV）工具（例如 CROWN（Zhang 等人，2018年））（Zhang 等人，2018年；Everett 等人，2021年；Jafarpour 等人，2024年）计算 RSOA。保守性会随着非线性项数量的增加而增大（Nath 等人，2025），对于非线性机器人动力学而言，保守性可能过高。符号方法和单次方法

可以缓解这一问题，但会增加运行时和内存开销（Chen 等人，2023；Rober 和 How，2024；Everett 等人，2021年）。我们改用 Koopman 线性化动力学对机器人建模，减少计算图（CG）中的非线性项数量，从而提升可扩展性（即计算时间和保守性）。

Koopman 可达性 Koopman（Koopman 和 Neumann，1932年）与 Carleman（Amini 等人，2021年；Bayer 和 Leine，2023年）线性化可加速可达性分析，但会引入近似误差。此外，尽管已有误差界，但这些推导仅适用于二次或多项式动力学（Liu 等人，2021年；Forets 和 Pouly，2017年；Amini 等人，2021年），无法应用于许多机器人动力学模型。以往的 Koopman 可达性方法也缺乏可达集覆盖率校准（Thapliyal 和 Hwang，2022年），或仅保证有限数量观测轨迹的包含性（Kochdumper 和 Bak 等人，2022年），因此无法确保真实系统在此前未见轨迹上的安全性。以往方法也不支持神经网络提升，并且只考虑自主系统，这相当于固定单个控制器。更换控制器（机器人领域常见做法）需要重复校准（Forets 和 Schilling，2021年；Bak 等人，2021年、2025年），效率低下。相比之下，我们支持神经网络提升，为真实动力学提供概率覆盖保证，并将控制器设计纳入其中，从而实现跨多个控制器的数据高效可达性分析。

数据驱动可达性数据驱动可达性利用黑盒轨迹数据。一些方法对 Hamilton-Jacobi 分析进行改造（Chilakamarri 等人，2024年），但用于高维系统时成本高昂。另一些方法学习可达性函数（Sun 和 Mitra，2022年），但需要密集的数据才能准确泛化；或者在假设已知系统数据（例如 Lipschitz 常数）的情况下，得到椭球形 RSOA（Park 等人，2024年）。基于采样的方法（Lew 等人，2022年）仅适用于短时域。情景优化可提供概率保证（Dietrich 等人，2024年），但速度慢于基于 Koopman 的方法（Bak 等人，2025）。此外，这些方法假设控制器固定不变；然而，机器人常常会切换控制器以实现不同目标，从而改变闭环动力学，并迫使人们重新计算可达集。轨迹跟踪可达性可针对不同参考轨迹重复使用跟踪误差集（Herbert 等人，2017年；Singh 等人，2018年；Knuth 等人，2021年）。数据驱动的变体通过学习收缩度量（Chou 等人，2021年、2022年；Knuth 等人，2023年），实现未知系统的基于跟踪的可达性，但找到这类度量通常颇具难度。我们在这项工作的基础上，得到 Koopman 可达集，其计算速度快、可扩展至长时域，并且可重复用于跟踪不同轨迹所诱导的各种闭环动力学。

3. 预备知识

我们考虑 未知的黑箱非线性系统，既包括开环（1a），也包括闭环（1b） :

$$x_{t+1} = f(x_t, u_t), \tag{1a}$$

$$x_{t+1} = f(x_t, \pi(x_t, t)) \doteq \tilde{f}(x_t), \tag{1b}$$

时间步为 $t \in \{0, \dots, T\} \doteq \mathcal{T}$，状态为 $x_t \in \mathcal{X} \subseteq \mathbb{R}^n$，控制为 $u_t \in \mathcal{U} \subseteq \mathbb{R}^m$，反馈控制律为 π：$\mathcal{X} \times \mathcal{T} \to \mathcal{U}$。初始状态 x_0 属于集合 $\mathcal{X}_0 \subseteq \mathcal{X}$。将 Int$(a,\ a) = \{a \mid a \le a \le a\}$ 表示为一个区间，其中 $a, a \in \mathbb{R}^A$，且不等式均按元素逐一成立，即 $a_i \le a_i \le a_i,\ 1 \le i \le A$。将 a-球表示为 $\mathcal{B}_a(c) \doteq \{x \mid \|x - c\|_\infty \le a\}$，将序列表示为 $q_{1:Q} \doteq \{q_1, q_2, \dots, q_Q\} = \{q_i\}_{i=1}^{Q}$。我们用 $\oplus$ 和 $\ominus$ 分别表示闵可夫斯基和与闵可夫斯基差。

Koopman 算子 由于我们的工作依赖于 Koopman 算子，因此我们在此讨论其基本概念。考虑自主离散时间非线性动力学 $x_{t+1} = f(x_t)$。Koopman 算子$\mathcal{K}$ 是一个无限维线性算子，它根据状态 $\mathcal{K}\phi^\infty(x)$ =演化可观测函数 ϕ^∞

$\phi^{\infty} \circ f(x)$，其中 $\circ$表示函数复合。由于无限维 Koopman 算子在实践中难以求得，我们使用有限维升维函数 ϕ 近似 $\mathcal{K}$： $\mathcal{X} \to \mathcal{Z} \subseteq \mathbb{R}^l$ 和矩阵 $K_A \in \mathbb{R}^{l \times l}$，它们共同定义提升动力学 $z_{t+1} = K_A z_t$，其中 $z_t = \phi(x_t)$为提升状态，且 $z_t \in \mathcal{Z} \subseteq \mathbb{R}^l$。在这里，$l$ 可通过选择升维函数 ϕ 任意设定。对于具有控制输入的系统（1a），提升动力学也可类似地写为 $z_{t+1} = K_A z_t + K_B u_t$，其中 $K_B \in \mathbb{R}^{l \times m}$。在实践中，$\phi$、$K_A$和 K_B 可从数据中学习，升维函数 g 则通过多项式或神经网络进行参数化（Shi等，2024）。在本研究中，我们采用基于神经网络的自动编码器架构，其中编码器 ϕ： $\mathcal{X} \to \mathcal{Z}$ 定义升维函数，解码器 ψ： $\mathcal{Z} \to \hat{\mathcal{X}}$定义其学习得到的逆函数，即 $\psi(\phi(x)) \approx x$；如果 ψ 是精确的逆函数，则 $\psi(\phi(x)) = x$ 且 $\hat{\mathcal{X}} = \mathcal{X}$。

通过神经网络验证进行可达性分析（NNV） 我们使用 NNV 计算涉及基于神经网络的提升函数和逆函数的闭环系统的可达集。NNV 工具将神经网络表示为计算图（CG）——一种有向无环图，用于编码应用于输入的运算序列（Rober和How，2024）。对于某些具有输入集 $\mathcal{S} \subseteq \mathbb{R}^{n_i}$ 和输出 $G(\mathcal{S}) \subseteq \mathbb{R}^{n_o}$的计算图 G，我们可以使用基于CROWN的（Zhang等，2018）工具（如 auto LiRPA（Xu等，2020））计算其像 $G(\mathcal{S})$的可保证过度近似。这些工具针对任意区间 $\mathcal{S}$，为输出 $G(\mathcal{S})$提供有保证的仿射下界和上界，G和 G，其形式化表述如下命题：

命题*1*（计算图鲁棒性（*Xuetal.,2020*））对于某个计算图 G 和区间 $\mathcal{S} \doteq \{s \in \mathbb{R}^{n_i} \mid \underline{s} \le s \le \overline{s}\}$，存在仿射函数$\underline{G}$、$\overline{G}$，使得 $\forall s \in \mathcal{S}$、$\underline{G}(s) \le G(s) \le \overline{G}(s)$。这些不等式逐元素成立，且$\underline{G}(s) = \underline{\Psi} s + \underline{\alpha}$、$\overline{G}(s) = \overline{\Psi} s + \overline{\beta}$，其中 $\underline{\Psi}$、$\overline{\Psi} \in \mathbb{R}^{n_o \times n_i}$和 $\underline{\alpha}$、$\overline{\beta} \in \mathbb{R}^{n_o}$。

给定控制器 π、动力学 (1a) 和初始集合 $\mathcal{X}_0$，T 时间步的精确可达集记为 $\mathcal{X}_{0:T} \doteq \{\mathcal{X}_0, f(\mathcal{X}_t, \pi(\mathcal{X}_t, t))\}_{t=0}^{T-1}\}$。虽然非线性系统的精确可达性分析通常难以处理（Althoff 等人，2021），但通常可以计算可达集的过近似（RSOA）$\mathcal{X}_{0:T}$，其满足 $\mathcal{X}_t \subseteq \mathcal{X}_t$ 对所有 $t \in \{0, \ldots, T\}$ 均成立。为获得 $\mathcal{X}_{0:T}$，可将命题 1 应用于闭环动力学 $\tilde{f}$(1b) 的 T 步复合 $\tilde{f}^T(\cdot) \doteq \tilde{f} \circ \cdots \circ \tilde{f}(\cdot)$ 的计算图 F，即 $F(x_0) \doteq \{\tilde{f}(x_t) \doteq f(x_t, \pi(x_t, t))\}_{t=0}^{T}$，并使用 auto LiRPA 计算 RSOA。

保形预测（CP）我们的 RSOA 计算也依赖于保形预测。在本文中，我们采用拆分保形预测，将来自数据集 $\mathcal{D} \doteq \{(v^{(i)}, y^{(i)})\}_{i=1}^{L+K}$ 的独立同分布输入输出对 $(v^{(i)}, y^{(i)}) \in \mathcal{V} \times \mathcal{Y}$ 划分为大小为 L 的训练集 $\mathcal{D}_T$ 和大小为 K 的校准集 $\mathcal{D}_C$。对于给定的预测函数μ：$\mathcal{V} \to \mathcal{Y}$，我们为 $\mathcal{D}_C$ 中的 K 个校准对分别计算非一致性分数 $R^{(i)} \doteq s(y^{(i)}, \mu(v^{(i)}))$，其中 s： $\mathcal{Y} \times \mathcal{Y} \to \mathbb{R}_{\ge 0}$ 是选定的非一致性分数函数。举例而言，我们可以在此框架中表示学习得到的动力学模型 f 的预测误差：定义 $v = (x_t, u_t)$，$y = x_{t+1}$, $\mu(v) = f(v)$，以及 $s(y, \mu(v)) = \|y - \mu(v)\|_2$。给定失覆盖水平 $\delta \in (0, 1)$，阈值 C 定义为校准分数的 $(1 - \delta)$ 分位数：$C \doteq \text{Quantile}_{1-\delta}(R^{(1)}, \ldots, R^{(K)}, \infty)$。该阈值定义了预测集合 $\mathcal{Y}(v) \doteq \{\, y \in \mathcal{Y} \mid s(y, \mu(v)) \le C \,\}$；对于与 $\mathcal{D}$ 可交换的未见数据点 $(y^{(0)}, v^{(0)})$，该集合满足边际覆盖保证 $\mathbb{P}(y^{(0)} \in \mathcal{Y}(v^{(0)})) \ge 1 - \delta$（Lindemann 等人，2024年）。

4. 问题陈述

我们可以黑箱访问 f (1a)，也就是说，我们不知道其解析形式，只能通过数据估计其输出。我们是一个规划器 P: $\mathcal{X} \to \mathcal{X}^{T+1} \times \mathcal{U}^T$，用于生成满足（$(x_{0:T}^{\text{ref}}, u_{0:T-1}^{\text{ref}})$（1a）的开环参考轨迹，初始条件为 $x_0^{\text{ref}} \in \mathcal{X}_0 \ominus \mathcal{B}_\epsilon(0)$

其在 $\mathcal{X}_0 \ominus \mathcal{B}_\epsilon(0)$ 上均匀分布。我们不对 P 作进一步假设；它可以是学习型规划器，也可以是传统规划器。在运行时，真实初始状态 x_0 可能相对于 x_0^{ref} 发生扰动，即 $x_0 \in \mathcal{B}_\epsilon(x_0^{\text{ref}})$，其中 x_0 在 $\mathcal{B}_\epsilon(x_0^{\text{ref}})$ 上均匀分布。为增强对该误差的鲁棒性，我们设计一个控制器 $\pi\colon \mathcal{X} \times \mathcal{T} \to \mathcal{U}$，使其跟踪来自 P 的参考轨迹，并计算相应闭环动力学 (1b) 的高概率 RSOA $\mathcal{X}_{0:T}$，使得 $\mathbb{P}\big(\bigwedge_{t=0}^{T}(x_t \in \mathcal{X}_t)\big) \geq 1-\delta$。这一过程应能扩展至长时域下的高维非线性系统（例如 $n \geq 25$、$T \geq 200$）。我们求解：

问题 1（模型学习与控制器设计） 给定由 P通过对 (1a)进行黑盒查询而生成的 N 长度为 $(T+1)$ 的开环动力学可行轨迹，并形成数据集 $\mathcal{D} = \{(x_0^{(i)}{}_{:T}, u_0^{(i)}{}_{:T-1})\}_{i=1}^{N}$，*(a)* 训练 *Koopman* 模型 (ϕ, ψ, K_A, K_B)，并且 (b) 使用该 Koopman 模型设计时变轨迹跟踪控制器： $\mathcal{X} \times \mathcal{T} \to \mathcal{U}$，以稳定跟踪由 P生成的参考轨迹。

问题 2（() 置信鲁棒安全可达集计算） 给定由 P生成的参考轨迹以及问题 *1* 中的控制器 π 所产生的 M 长度为 $(T+1)$ 的闭环轨迹，计算闭环动力学 (1b)的 $(1-\delta)$-置信鲁棒安全可达集 $\mathcal{X}_{0:T}$。该可达集保证：对于规划器 P利用 $x_0^{ref} \in \mathcal{X}_0$ 生成的任意随机参考轨迹$(x_0^{ref}{}_{:T}, u_0^{ref}{}_{:T-1})$，以及初始状态 (x_0^{ref})，闭环轨迹以至少 的概率始终保持在 $:T$ 内，即 $\mathbb{P}\big(\bigwedge_{t=0}^{T}(x_t \in \mathcal{X}_t)\big) \geq 1-\delta$。

5. 方法

我们的方法（图 2，算法 1）学习一个 Koopman 算子（第 5.1节），将其用于控制器设计（第5.2节），对闭环提升的 Koopman 动力学进行快速线性可达性分析（第5.3节），并通过保形预测扩张这些可达集，以获得概率覆盖保证（第 5.4节）。

5.1. Koopman 算子训练

为求解问题 1，我们在规划器 P 生成的开环轨迹数据集 $\mathcal{D} \doteq \{(x_0^{(i)}{}_{:T}, u_0^{(i)}{}_{:T-1})\}_{i=1}^{N}$上训练 Koopman 算子模型 (ϕ, ψ, K_A, K_B)（详情见第 6 节）。我们采用复合损失 $\mathcal{L} = \sum_{i=1}^{N}(\lambda_1 \mathcal{L}_1^{(i)} + \lambda_2 \mathcal{L}_2^{(i)})$。第一项 $\mathcal{L}_1^{(i)} = \sum_{j=0}^{T}\big[\|x_j^{(i)} - \psi(\phi(x_j^{(i)}))\|^2 + \|\phi(x_j^{(i)}) - \phi(\psi(\phi(x_j^{(i)})))\|^2\big]$通过状态重构误差和潜空间一致性损失来确保自编码器的准确性。第二项 $\mathcal{L}_2^{(i)} = \sum_{j=t+1}^{t+H}\big[\|x_j^{(i)} - \psi(\check{z}_j^{(i)})\|^2 + \|\phi(x_j^{(i)}) - \check{z}_j^{(i)}\|^2\big]$，其中 $\check{z}_j^{(i)} \doteq K_A \check{z}_{j-1}^{(i)} + K_B u_{j-1}^{(i)}$和 $\check{z}_0^{(i)} = \phi(x_0^{(i)})$，是一种多步（时域 H）动力学损失，用于惩罚线性模型 (K_A, K_B) 在潜空间和状态空间中的预测误差。潜空间一致性项对于保证长时域预测精度至关重要。我们将 K_A 初始化为单位矩阵，并使用 Xavier 均匀初始化设置 K_B；训练时采用学习率退火和权重衰减，并将 (ϕ, ψ) 的激活函数设为 ReLU 或 GELU。

5.2. Koopman 跟踪 LQR 设计

为了可靠地模仿规划器生成的状态/控制参考轨迹 P（位于提升状态空间中） $\mathcal{Z}$，需要跟踪控制器 π 来稳定系统，使其不受初始条件扰动 $x_0 - x_0^{\text{ref}}$的影响，并抑制由学习得到的提升动力学中的模型误差引起的扰动 $z_{t+1} = K_A z_t + K_B u_t$。为实现 π，我们利用提升动力学的线性性质，设计一个线性二次调节器（LQR），用于跟踪 $\mathcal{Z}$中的轨迹。给定由 P生成的参考状态空间轨迹 $x_{0:T}^{\text{ref}}$，我们将其映射到提升状态空间，得到 $z_{0:T}^{\text{ref}} \doteq \{\phi(x_t^{\text{ref}})\}_{t=0}^{T}$。随后，计算使提升状态模仿误差最小的相应前馈控制轨迹 $u_{0:T-1}^{\text{ref}}$：

$$u_t^{\text{ref}} = K_B^{\dagger}(z_{t+1}^{\text{ref}} - K_A z_t^{\text{ref}}), \tag{2}$$

其中， $K_B^\dagger$ 是 K_B的伪逆。误差状态 $\delta z_t = z_t - z_t^{\text{ref}}$和控制量 $\delta u_t = u_t - u_t^{\text{ref}}$遵循线性误差动力学 $\delta z_{t+1} = K_A \delta z_t + K_B \delta u_t$，其初始误差为 $\delta z_0 = \phi(x_0) - z_0^{\text{ref}}$，其中 $x_0 \in \mathcal{B}_\epsilon(x_0^{\text{ref}})$。我们求解一个线性二次调节问题，找出使有限时域二次代价 $\delta z_T^\top Q_T \delta z_T + \sum_{t=0}^{T-1} ($ $\delta z_t^\top Q \delta z_t + \delta u_t^\top R \delta u_t)$最小的反馈控制量 $\delta u_{0:T-1}$，并满足线性动力学约束 $\delta z_{t+1} = K_A \delta z_t + K_B \delta u_t$，其中 $Q \succ 0$、 $R \succ 0$和 $Q_T \succ 0$。具体而言，我们通过求解向后的黎卡提递推，得到最优反馈增益 $G_{0:T-1}$。由此得到提升状态空间中的跟踪控制律（3）；通过测量并编码状态 x_t，即可将其应用于非线性系统（1a），如式（4）所示：

$$u_t = u_t^{\text{ref}} - G_t(z_t - z_t^{\text{ref}}), \qquad (3) \qquad\qquad u_t = u_t^{\text{ref}} - G_t(\phi(x_t) - z_t^{\text{ref}}). \qquad (4)$$

5.3. 闭环 Koopman 动力学的可达性分析

非线性可达性分析的高计算成本促使我们采用 Koopman 算子。标准的 RSOA 计算会在每个时间步反复对f 和 π 中的非线性项进行过近似（为了稳定 f，它们通常必须是非线性的），导致保守性很高。一个T步的 RSOA 需要进行 $2T$ 次非线性函数求值，随着 T 增加，可达集会变得宽松到失去实际意义。为降低保守性，基于划分的方法会将输入集$\mathcal{X}_t$拆分为多个部分，传播每个部分后再取并集；然而，这会带来计算和内存开销，且开销会随着划分数量的增加而上升。相比之下，基于 Koopman 的可达性分析通过减少非线性项，在无需划分开销的情况下提升可扩展性并降低保守性。映射到升维空间 $\mathcal{Z}$ 后会形成线性动力学，因此传播过程中无需界定非线性项。线性动力学还允许采用线性跟踪控制器。因此，对于任意时间步 t 计算 RSOA $\mathcal{X}_t$ 时，只需界定非线性编码器（ϕ）和解码器（ψ）各一次，将非线性界定操作的数量从 $2T$ 次减少到仅 2 次。为将这一思路正式应用于 RSOA 计算，我们定义跟踪 Koopman 反馈回路（TKFL），记为 Γ，它是一个计算图（CG），用于描述在提升状态空间中使用反馈控制律 (4) 跟踪$z_{0:T}^{\text{ref}} \doteq \{\phi(x_t^{\text{ref}})\}_{t=0}^T$ 的闭环动力学。具体而言，TKFL Γ：$\mathcal{X}_0 \times \mathcal{U}^T \times \mathcal{Z}^{T+1} \to \hat{\mathcal{X}}^{T+1}$将初始状态 $x_0 \in \mathcal{B}_\epsilon(x_0^{\text{ref}})$ 、前馈控制轨迹 $u_{0:T-1}^{\text{ref}}$ (2)和提升参考轨迹 $z_{0:T}^{\text{ref}}$映射为解码轨迹 $\hat{x}_{0:T} \doteq \{\psi(z_t)\}_{t=0}^T$，该轨迹表示原始状态空间中的预测闭环轨迹 (5)，其中提升状态 z_t 通过闭环提升动力学 (6) 更新：

$$\hat{x}_{0:T} = \{x_0, \{\psi(\tilde{f}_z(z_t))\}_{t=0}^{T-1}\}, \qquad (5)$$

$$z_{t+1} = K_A z_t + K_B(u_t^{\text{ref}} - G_t(z_t - z_t^{\text{ref}})) \doteq \tilde{f}_z(z_t). \qquad (6)$$

总之，计算图按如下方式构建（可视化结果见图2）。首先，通过 ϕ 将初始状态映射到潜在空间，计算出 $z_0 = \phi(x_0)$。其次，将单步潜在动力学 $z_{t+1} = K_A z_t + K_B u_t$递归应用 T次并展开，其中 u_t 是使用式（3）计算得到的闭环跟踪控制，从而得到潜在轨迹 $z_{0:T} = \{z_0, z_2, \cdots, z_T\}$。第三，通过解码器 ψ将 $z_{0:T}$映射回原始状态空间，得到 $\hat{x}_{0:T}$。计算图Γ 被传递给 auto LiRPA 库（Xu等，2020），该库可使用 CROWN（Zhang等，2018）对 Γ 进行基于集合的界传播。auto LiRPA 计算向量值仿射界函数 $\bar{\Gamma}$ 和 $\underline{\Gamma}$，使我们能够通过考察 Γ、Γ_t 和 Γ_t的特定分量，计算解码后的闭环提升动力学（5）的 RSOA $\bar{\mathcal{X}}_{0:T}$，这些分量对状态进行了界定

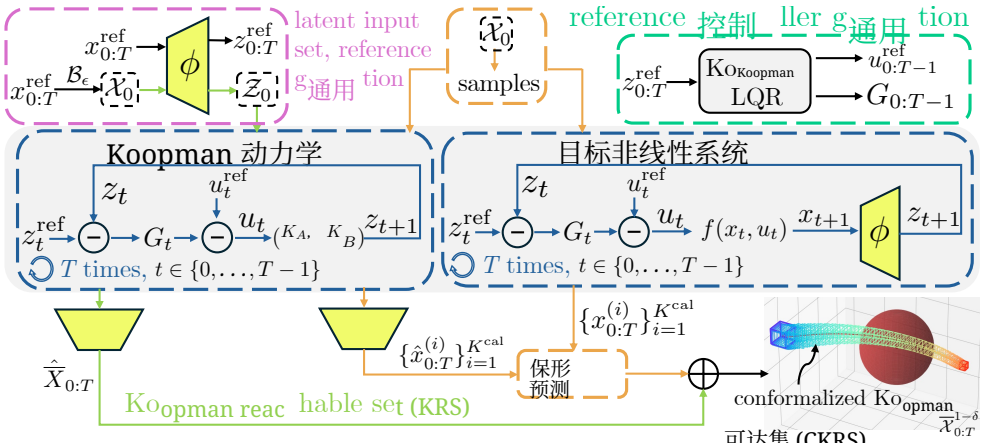


图 2：描述我们方法（ScaRe-Kro）流程的框图。绿色和橙色箭头分别表示 KRS 和 CP 计算。

在（5）下于时间步 t从任意初始条件 $x_0 \in \mathcal{X}_0$出发所到达的。这一超ˆ矩形集合 $\mathcal{X}_{0:T}$，我们称之为 Koopman 可达集（KRS），可通过求解优化问题（7）得到，并可由 `auto LiRPA`高效计算：

$$\hat{\overline{\mathcal{X}}}_{0:T} \doteq \left\{\mathsf{Int}\left(\min_{x_0\in\mathcal{X}_0} \underline{\Gamma}_t(x_0, u^{\text{ref}}_{0:T-1}, z^{\text{ref}}_{0:T}), \max_{x_0\in\mathcal{X}_0} \overline{\Gamma}_t(x_0, u^{\text{ref}}_{0:T-1}, z^{\text{ref}}_{0:T})\right)\right\}_{t=0}^{T} \tag{7}$$

ˆ KRS $\mathcal{X}_{0:T}$对 ˆ$x_{0:T}$进行过近似，使得 ˆ$x_t \in \mathcal{X}_t$对于所有 $t \in \mathcal{T}$均成立，从而确保解码后的闭环ˆ提升动力学（5）始终处于 $\mathcal{X}_{0:T}$中。形式化地，我们给出以下结果（证明见附录 8.1）：

引理 1（KRS 过近似）令 Γ 表示 *TKFL* 的闭环 *CG*，并令 $\mathcal{X}_0$ 为ˆ初始集合。假设 *KRS* $\mathcal{X}_{0:T}$如（7）所定义。那么，对于任意初始状态 $x_0 \in \mathcal{X}_0$，解码后的闭环轨迹（5），即ˆ $x_{0:T} = \Gamma$（x_0、 z_0^{ref} : T、 u_0^{ref} : $T-1$）满足ˆ$x_t \in$ ˆ$\mathcal{X}_t$对于所有 $t \in \mathcal{T}$均成立。

关键在于，KRS 对解码后的 Koopman 线性化动力学（5）的可达集进行过近似，而非真实非线性动力学(1b)，这是由于 Koopman 模型误差所致。这促使我们在第 5.4 节进一步讨论。

5.4. Koopman 可达集的保形化

现在，我们说明如何扩张 KRS，以保证它以用户指定的概率 $1-\delta$ 包含真实系统 (1b) 的轨迹。我们使用保形预测来计算解码后的闭环提升动力学 (5) 的状态空间预测误差的高概率上界。具体而言，我们改进单一非一致性分数方法 (SNSA)（Lindemann 等人，2024年），利用真实闭环动力学 $x_{0:T}$ (1b) 和解码后的闭环提升动力学 ˆ$x_{0:T}$ (5) 的轨迹 $\mathcal{D}_C(x_{0\,T}^{(i)}:, \hat{}x_{0\,T}^{(i)}:)\}_{i=1}^{K^{\text{cal}}}$ 校准数据集 $\mathcal{D}_C = \{$，计算与状态和维度相关的误差界。为了构造 $\mathcal{D}_C$ 中的每个数据元组，我们从规划器P中采样一条参考轨迹 $x^{\text{ref}},_{0\,T}^{(i)}:$，并从 $\mathcal{B}_\epsilon(x_0^{\text{ref}})$ 中均匀采样一个初始状态 $x_0^{(i)}$。然后，我们使用 (2) 计算潜在参考轨迹$z^{\text{ref}},_{0\,T}^{(i)}:$ $= \{\phi(x^{\text{ref}},_{t}^{(i)})\}_{t=0}^{T}$ 和前馈控制序列 $u^{\text{ref}},_{0\,T-1}^{(i)}:$。我们使用 $u^{\text{ref}},_{0\,T-1}^{(i)}:$，通过以下两种方式计算得到的轨迹：1）在提升状态空间 (8) 中跟踪 $z_{0:T}^{\text{ref}}$，并通过 ψ 解码至状态空间；2）在真实非线性动力学 (9) 上执行 Koopman控制器：

$\hat{x}^{(i)}_{1:T} = \{\psi(K_A z_t + K_B(u^{\text{ref}}_t - G_t(z_t - z^{(i)}_t))\}^{T-1}_{t=0}$ (8)　　$x^{(i)}_{1:T} = \{f(x_t, u^{\text{ref}}_t - G_t(\phi(x_t) - z^{(i)}_t)\}^{T-1}_{t=0}$ (9)

( ) 每个状态维度 $j \in \{1,\dots,n\}$ 在时间 t 的预测误差记为 $e^{(i)}_{t,j} = x^{(i)}_{t,j} - \hat{x}^{(i)}_{t,j}$。由此得到一个误差轨迹校准集 $\mathcal{D}_E = \{\{e^{(i)}_t\}^{T}_{t=0}\}^{K^{\text{cal}}}_{i=1}$，其中每个 $e^{(i)}_t \in \mathbb{R}^n$。每条误差轨迹 $e^{(i)}$ 都归一化为一个非一致性分数 $R^{(i)} = \max_{t,j}(\lambda_{t,j}|e^{(i)}_{t,j}|)$，即所有时间步 $t \in \{0,\dots,T\}$ 和状态维度$j \in \{1,\dots,n\}$上的最大加权误差。这里，归一化常数$\lambda_{t,j} \doteq 1/(e^{\max}_{t,j} + \sigma) \in \mathbb{R}_{>0}$使用一个单独的归一化数据集 $\mathcal{D}_N$计算。该数据集 $\mathcal{D}_N$由 M^λ 条额外的误差轨迹 $\{e^{(i)}\}^{K^{\text{cal}}+M^\lambda}_{i=K^{\text{cal}}+1}$ 组成，这些轨迹采用与校准数据相同的流程收集。这里， σ 是一个较小的正数，用于防止除以零。项$e^{\max}_{t,j}$ 是跨越 $\mathcal{D}_N$观测到的各维度和各时间点的最大绝对误差，即 $e^{\max}_{t,j} = \max_i |e^{(i)}_{t,j}|$，其中 $i \in \{K^{\text{cal}}+1,\dots,K^{\text{cal}}+M\}$。

现在，给定用户定义的失覆盖率 $\delta \in (0,1)$，我们通过找出非一致性分数 $C = \text{Quantile}_{1-\delta}(\{R^{(i)}\}^{K^{\text{cal}}}_{i=1} \cup \{\infty\})$ 中第 p 小的值，计算非一致性分数的 $(1-\delta)$ 分位数，其中 $p \equiv \lceil (K^{\text{cal}}+1)(1-\delta)\rceil$。最终，依赖于时间和维度的误差界 $\bar{e}_{t,j} \in \mathbb{R}_{\geq 0}$ 可通过逆转归一化 $\bar{e}_{t,j} = C/\lambda_{t,j}$ 得到。这些误差界确保，对于一个未见过的测试数据点 $(x^{(0)}_{0:T}, \hat{x}^{(0)}_{0:T})$，真实预测误差 $|e^{(0)}_t|$ 将包含在界 $\bar{e}_{t,j}$ 内，对所有 t 和 j 均成立，且概率至少为 $1-\delta$，即

$$\mathbb{P}\big(|e^{(0)}_{t,j}| \leq \bar{e}_{t,j}, \forall t \in \{0,\dots,T\}, \forall j \in \{1,\dots,n\}\big) \geq 1-\delta$$

我们使用这些界对通过（7）计算出的 KRS 进行膨胀，以计算原始闭环动力学（1b）对应的 RSOA，并保证其有效概率至少为 $1-\delta$，使用（11）（证明见附录 8.1）：

定理 1（CKRS 覆盖保证）设 $\mathcal{X}_{0:T}$为 (7)中定义的 *KRS*，并设 $\bar{e}_t = [\bar{e}_{t,1},\dots,\bar{e}_{t,n}]^\top \in \mathbb{R}^n$。定义保形 *Koopman* 可达集（*CKRS*）为

$$\overline{\mathcal{X}}^{1-\delta}_{0:T} \equiv \{\hat{\mathcal{X}}_t \oplus \text{diag}(\bar{e}_t)\mathcal{B}_1(0)\}^{T}_{t=0}, \tag{10}$$

其中 $\text{diag}(\bar{e}_t) \in \mathbb{R}^{n\times n}$ 是一个对角矩阵，其对角线上为 $\bar{e}_t$ 。那么，对于一条来自相同分布的新参考轨迹 $x^{ref}_{0:T}$：由（(1b)）生成的真实闭环系统轨迹以至少 $1-\delta$的概率包含在 *CKRS* 中，即 $\mathbb{P}\big(\bigwedge^{T}_{t=0} x^{(0)}_t \in \mathcal{X}^{1-\delta}_t\big) \geq 1-\delta$。

离线校准 如前所述，第5.4 节中的方法针对特定参考，因为校准轨迹 $\mathcal{D}_C$ 和误差界 $\bar{e}_t$ 仅适用于单个参考 $x^{\text{ref}}_{0:T}$。我们提出一种离线方法，预先计算一组全局误差界 $\bar{e}_{t,j}$ 适用于从规划器 P中采样的任意参考 $x^{\text{ref}}_{0:T}$。实现这一点的方法是构建一个规模更大、更多样化的校准集。首先，我们采样 N^{ref}条参考轨迹 $\{x^{\text{ref},(i)}_{0:T}\}^{N^{\text{ref}}}_{i=1} \sim P$。对于每个参考，我们通过采样 $x^{(i)}_0 \sim \mathcal{B}_\epsilon(x^{\text{ref}}_0)$生成一条校准轨迹，从而得到误差轨迹$e^{(i)}_t = x^{(i)}_t - \hat{x}^{(i)}_t$，其中 $x^{(i)}_t$ 和 $\hat{x}^{(i)}_t$ 分别如式（8）和（9）所定义。随后，针对整个聚合数据集，严格执行一次 SNSA 程序（通过 $\lambda_{t,j}$ 计算 $R^{(i)}$、 C 和 $\bar{e}_{t,j}$ ）。

在运行时，给定新的参考轨迹 $x^{\text{ref}}_{0:T}$：从 P 中采样得到后，我们会进行两项高效计算：1）通过（$\hat{}$ $\mathcal{X}_{0:T}$ 7）计算特定于 $x^{\text{ref}}_{0:T}$：的名义 KRS；2）通过闵可夫斯基和 $\bar{e}_t$ 与 $\mathcal{X}^{1-\delta}_t = \hat{}\mathcal{X}_t \oplus \text{diag}(\bar{e}_t)\mathcal{B}_1(0)$，利用预先离线计算的界限对该集合进行扩张。

如（11）所定义。这种方法可在运行时快速计算可达集，但可能会使最终 CKRS 的保守性增加，因为界限 $\bar{e}_t$ 现在必须对从 P 中抽取的所有参考轨迹统一成立，而不是仅对某一固定参考轨迹成立。这种离线方法的有效性依赖于非一致性分数的可交换性；由于每个分数 $R^{(i)}$ 都是数据对 $(x^{\mathrm{ref},(i)}_{0:T}, x^{(i)}_0)$ 的确定性函数，因此满足这一条件。这些数据对均独立同分布地取自由 $x^{\mathrm{ref}}_{0:T} \sim P$ 和 $x_0 \sim \mathcal{B}_\epsilon(x^{\mathrm{ref}}_0)$ 定义的联合分布。因此，所得分数也独立同分布，进而满足可交换性，从而保证标准保形预测保证 $\mathbb{P}\big(\bigwedge_{t=0}^{T} x_t^{(0)} \in \mathcal{X}_t^{1-\delta}\big) \geq 1-\delta$ 成立。

6. 实验

我们首先将本方法与可达性基线进行比较，然后评估其在独轮车（3D）、平面四旋翼飞行器（6D）、三维四旋翼无人机（12D）、MuJoCo（Todorov 等人，2012）Hopper（11 维）和 MuJoCo 游泳者（28D）上的可达性分析效果。所有实验均在一台配备 Intel i9-14900K CPU、64 GB RAM 和 NVIDIA RTX 4090 GPU 的台式机上运行。对于 MuJoCo 示例，我们将规划器设为专家级近端策略优化（Schulman 等人，2017年）策略；对于其他系统，我们使用基于解析动力学的轨迹优化器作为规划器。我们报告四项关键指标：（i）可达性运行时间（秒），越低越好；（ii）可达集的平均对数体积，数值越低表示保守性越低；（iii）滚动轨迹的安全率，越高越好；（iv）基于保形预测的验证的覆盖率变异性，不同校准样本规模下经验覆盖率的贝塔后验方差越低，表示覆盖率越稳定。部分指标仅适用于特定基线（例如，安全过滤器仅报告安全率）。完整结果见表 1。

基线方法 我们选择基线方法，以突出数据驱动可达性中的四个缺口。(i) 学习式可达性估计器（Bansal 和 Tomlin，2021年）缺乏过近似保证，因此有必要采用形式化神经网络验证。(ii) 现有的混合式保形预测与神经网络验证方法（CP-NNV）（Hashemi 等人，2024年）难以扩展到高维非线性动力学和长时域，因此有必要采用Koopman提升。(iii) 纯粹的基于保形预测的可达性方法（Lindemann 等人，2024年）会随着校准数据或样本预算的减少而出现覆盖率下降，因此有必要采用在数据有限时仍能保持可靠性的样本高效可达性分析。(iv) 纯粹的基于神经网络验证的RSOA计算（Nath 等人，2025）可提供确定性保证，但速度可能较慢，因此有必要采用快速的基于Koopman的线性可达性分析。

学习式可达性估计器导致的不安全行为。 在独轮车动力学（12）下，基于价值函数的安全过滤器（Bansal 和 Tomlin，2021年）——一种近似可达性方法——无法始终确保可达-避障安全性；250条滚动轨迹中仅有70%满足约束（图5）。由于学习得到的价值函数是神经偏微分方程的近似解，价值函数及其梯度中的微小误差可能导致安全违规，而且该方法缺乏形式化保证。相比之下，我们的方法可提供校准概率保证，确保闭环系统保持在通过（11）计算得到的CKRS内（我们的实验中为100%），同时保持目标到达能力。

纯保形预测验证。 由于我们学习得到的 Koopman 提升与训练动力学高度吻合，因此得到的 KRS 已经较为紧致；随后进行共形化，仅用于吸收少量残余模型误差。我们将其与纯 SNSA 基线进行比较（Lindemann 等人，2024年），该基线将所有异质的轨迹误差和状态误差汇总为一个非一致性分数。随着时域或状态维度增大，为达到目标 δ，该基线不得不扩大预测集，或需要大得多的校准集，因此难以快速收敛至目标覆盖水平。在实际数据预算下（例如 1000 个样本），覆盖率方差仍然很大，例如真实系统轨迹的覆盖不足最高可达 5%（图3），这可能会损害

表1： SCaRe-Kro 在独轮车、四旋翼无人机和 MuJoCo模型上的数值RSOA结果 s.

实验	系统维度	Time 步数	置信水平 $(1-\delta)$	CP time	KRS time	总计 time (在线)	平均对数成交量 (在线)	平均对数成交量 (离线)	校准数据集 size	经验覆盖率
独轮车	3	100	99。0%	0.038秒	0.277秒	0.315秒	−11.543	−8.690	100	100%
二维四边形	6	100	99.0%	0.055秒	0.349秒	0.404秒	−21.491	−5.333	100	100%
三维四边形	12	200	99.0%	0.175秒	0.500秒	0.675秒	−42.954	−35.275	100	100%
Hopper	11	225	99。0%	10.597秒	0.387秒	10.983秒	−4.969	−1.806	100	100%
Swimmer	28	400	95.0%	20.818秒	0.593秒	21.411秒	32.595	42.617	100	99.5%

系统安全。相比之下，我们的方法在所有测试的校准集规模下均达到 100% 的经验覆盖率，避免了这种方差-数据权衡，并以少得多的样本实现目标覆盖率。

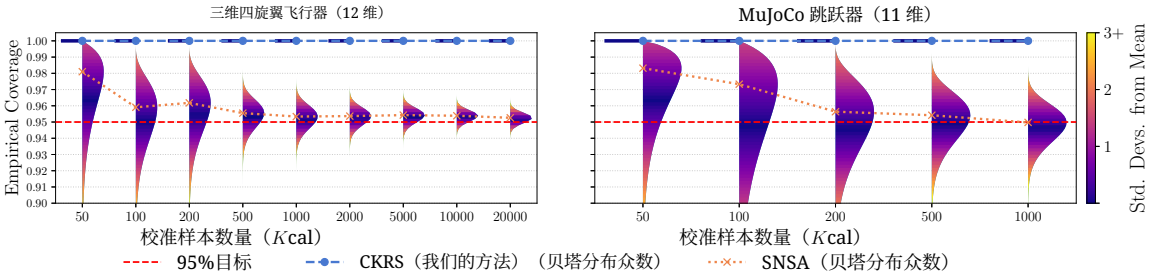


图3：为直观展示经验可达集覆盖率，我们计算了三维四旋翼无人机（左）和 Hopper（右）上 SNSA 基线方法与 ScaRe-Kro 的贝塔后验分布。虚线表示贝塔后验分布的众数。对于 SNSA，随着校准数据集大小增加，覆盖率缓慢收敛至目标值 $1-\delta$；而 ScaRe-Kro 即使在 K^{cal}较小时也能保持 100% 的覆盖率。

+

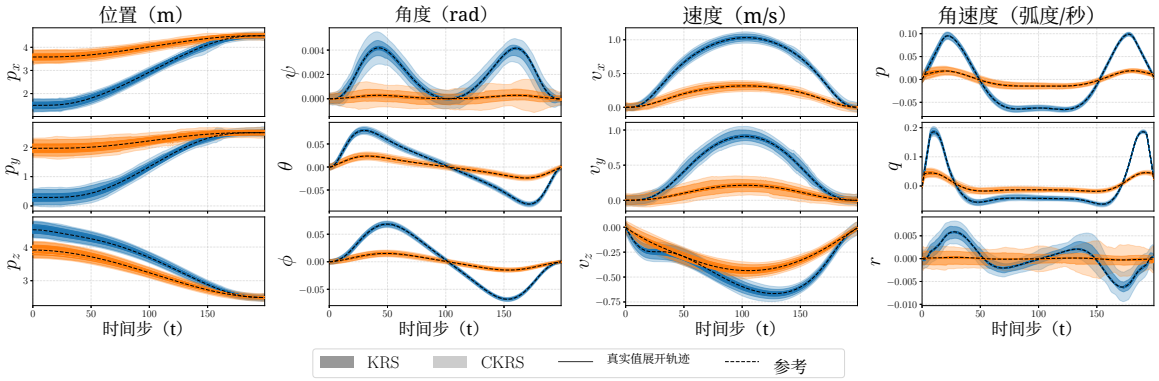
混合可达性方法（保形预测 *NNV*）。对于三维四旋翼无人机（14），我们将其与一种混合数据驱动/NNV 可达性方法进行比较（Hashemi 等人，2024年）（记为 CP-NNV）。该方法通过对学习得到的神经网络轨迹预测器应用 NNV，并使用保形预测来界定由学习误差引起的 RSOA 低估近似误差，从而计算 RSOA。如图6所示，CP-NNV 得到的可达集大得多（平均对数体积：-23.35），尤其是在姿态和速度方面。这反映了 NNV 和保形预测难以扩展至高维状态空间和长时域；我们通过 Koopman 算子来解决这一问题。我们的方法在保持覆盖率的同时，生成更紧致的 RSOA（平均对数体积：-40.66），提升了随时域增长的可扩展性，并降低了保守性。

纯 *NNV* 可达性。对于独轮车模型，我们将其与先前的工作进行比较（Nath 等人，2025），该工作使用 NNV 工具进行一次性 RSOA 计算，需要 51.95s，平均对数体积为 −3.98。相比之下，我们的方法在 0.32s 内计算出 RSOA，对数体积为−11.54（表 1）。这表明我们的方法速度快得多，同时也降低了保守性。

独轮车 我们在独轮车模型上评估 CKRS 计算（12），其中 ϕ 和 ψ 各有 3 个隐藏层，每层 128 个神经元，并且 $z \in \mathbb{R}^{10}$。如图 7所示，经过认证的可达集验证了安全性和目标达成。完整指标见表 1。我们将这一基准测试视为对简单系统的合理性检验，为后续更实质性的可扩展性结果铺路。

平面四旋翼飞行器 我们评估了一架六状态平面四旋翼飞行器（13）绕行圆形障碍物的情况，持续时间为 $T = 100$。其中，ϕ和 ψ 均有 3 个隐藏层，每层包含 128 个神经元，且 $z \in \mathbb{R}^{24}$。我们在图8中展示了针对多条参考轨迹离线校准的 CKRS；相关指标见表1。仅 NNV 基线（Nath 等人，2025）耗时 59。85 秒，平均对数体积为 −18.709。总体而言，这些结果支持了我们关于降低 RSOA 保守性的主张。

三维四旋翼无人机 我们评估了一架解析式 12 状态四旋翼无人机（14）（Sabatino，2015）绕行一个半径为 1 米的障碍物的情况，持续时间为 $T = 200$。其中，ϕ 和 ψ 均有 3 个隐藏层，每层包含 128 个神经元，且 $z \in \mathbb{R}^{24}$。计算得到的 CKRS（图4）验证了安全性和目标到达；相关指标见表1。为展示离线 CP 膨胀能够推广至不同控制器，我们使用相同的离线校准边界，针对两条参考轨迹（图4中一条为橙色，另一条为蓝色）计算 CKRS；更多 CKRS 展示于图9。仅 NNV 基线（Nath 等人，2025）速度慢得多且保守性更强，耗时2153。45 秒，平均对数体积为 -30.32。总体而言，这些结果表明，我们的方法在高维系统中具有高效率、紧致性以及离线 CP 泛化能力。



Figu图 4：针对三维四旋翼无人机计算的 CKRS 和 KRS（保形预测前），按各维度绘制on.

MuJoCo：Hopper 我们在 MuJoCo 的 Hopper-v5（Todorov 等人，2012）上进行评估。这是一个具有黑箱动力学的非线性欠驱动系统。我们使用一个经过 2×10^6 步训练的 PPO 智能体（Stable-Baselines3，默认超参数）作为规划器 P，以生成训练数据集所需的轨迹。这里，ϕ 和 ψ 均包含 3 个隐藏层，每层有 128 个神经元，且 $z \in \mathbb{R}^{24}$。数值结果见表 1，若干状态维度对应的 RSOA 切片见图 1（全部 11 个维度见图 11）。本实验表明，我们的方法能够扩展至黑箱、高维且富接触的动力学系统，同时保持较低的运行时开销并提供概率覆盖保证。

MuJoCo：定制 Swimmer 我们进一步在定制的 MuJoCo Swimmer-v5（Todorov 等人，2012）上评估可扩展性。该系统采用延长连杆，状态维度为 28。规划器 P 是一个 PPO 智能体（Stable-Baselines3 多层感知机），经过 1×10^7 步训练，并采用经过修改的超参数（学习率 1×10^{-3}，$\gamma = 0.9999$，批量大小 512；其他参数为默认值）。这里，ϕ 和 ψ 均包含 3 个隐藏层，每层有 128 个神经元，且 $z \in \mathbb{R}^{38}$。数值结果见表 1，若干状态维度对应的 RSOA 切片见图 1（全部 28 个维度见图 12）。本实验表明，对于 MuJoCo 中的黑箱动力系统，我们的方法在长时域、高维系统上能够以可行的计算开销求得 RSOA，同时维持经认证的概率安全性。

7. 结论

我们提出了 ScaRe-Kro，这是一种针对复杂非线性黑箱动力学的可扩展可达性分析方法，并提供形式化覆盖保证。我们的方法利用数据学习非线性动力学的提升线性表示，从而便于进行跟踪控制设计，并高效传播闭环可达集。随后，我们通过扩张可达集对其进行“保形化”，确保目标非线性系统的滚动轨迹以

指定概率（即至少 $1-\delta$）落在扩张后的集合内。在包括独轮车、平面/三维四旋翼飞行器、MuJoCo 跳跃器和 13连杆 MuJoCo 游泳器在内的多个基准测试中，我们的方法能够高效地计算较长时间范围内的可达集。未来的工作旨在：1）消除对参考轨迹的依赖；2）以鲁棒控制策略替代线性二次调节器，使其在存在模型误差时仍满足约束，从而延长系统可验证的时间范围。

参考

Matthias Althoff。《CORA 2015 简介》。载于 ARCH@CPSWeek，EPiC 计算系列第 34 卷，第 120–151 页，2015。

Matthias Althoff、Goran Frehse 和 Antoine Girard。《用于可达性分析的集合传播技术》。年度控制、机器人与自主系统评论，2021年。

Arash Amini、Qiyu Sun 和 Nader Motee。《一般非线性系统 Carleman 线性化的误差界》，第 1–8 页。2021年。

Stanley Bak、Hoang-Dung Tran 和 Taylor T. Johnson。《对维度高达十亿的仿射系统进行数值验证》。载于第 22 届 ACM 混合系统国际会议：计算与控制论文集，HSCC '19，第 23–32 页，2019。

Stanley Bak、Sergiy Bogomolov、Parasara Sridhar Duggirala、Adam R. Gerlach 和 Kostiantyn Potomkin。《koopman 算子线性化后黑箱非线性系统的可达性》。载于 IFAC，第 54 卷，第 253–258 页，2021年。

Stanley Bak、Sergiy Bogomolov、Brandon Hencey、Niklas Kochdumper、Ethan Lew 和 Kostiantyn Potomkin。《使用显式核近似和多项式 zonotope 精化分析 Koopman 线性化系统的可达性》。Formal Methods Syst. Des.，66(2)：307–333，2025。

Somil Bansal 和 Claire J. Tomlin。《DeepReach：一种用于高维可达性的深度学习方法》。载于 ICRA，第 1817–1824 页，2021年。

Somil Bansal、Mo Chen、Sylvia Herbert 和 Claire J. Tomlin。《Hamilton-Jacobi 可达性：简要概述与最新进展》，2017年。

Fabia Bayer 和 Remco I. Leine。通过 Koopman 分析对周期解进行无需排序的基于 Hill 的稳定性分析。《非线性动力学》，111(9)：8439–8466，2023年5月。Shaoru Chen、Victor M. Preciado 和 Mahyar Fazlyab。神经网络动力系统的一次性可达性分析。载于 ICRA，2023年。Vamsi Krishna Chilakamarri、Zeyuan Feng 和 Somil Bansal。黑箱动力系统的可达性分析，2024年。Glen Chou、Necmiye Ozay 和 Dmitry Berenson。通过学习得到的收缩度量传播模型误差，以实现未知系统的安全反馈运动规划。载于 2021年第60届 IEEE 决策与控制会议（CDC），美国德克萨斯州奥斯汀，2021年12月14日至17日，第3576–3583页，2021年。

Glen Chou、Necmiye Ozay 和 Dmitry Berenson。通过学习得到的感知模块和压缩理论，从图像进行安全输出反馈运动规划。载于 WAFR，《Springer 高级机器人学会议论文集》第25卷，第349–367页，2022年。Elizabeth Dietrich、Alex Devonport 和 Murat Arcak。面向数据驱动可达性的非凸情景优化。载于《第六届动力学与控制学习年会论文集》，第242卷，第514–527页，2024年。Michael Everett、Golnaz Habibi、Chuangchuang Sun 和 Jonathan P. How。神经反馈回路的可达性分析。《IEEE Access》，9，2021年。Marcelo Forets 和 Amaury Pouly。Carleman 线性化的显式误差界。2017年。Marcelo Forets 和 Christian Schilling。利用 Carleman 线性化实现弱非线性系统的可达性。载于 RP，第13035卷，第85–99页，2021年。Navid Hashemi、Lars Lindemann 和 Jyotirmoy V. Deshmukh。分布偏移下随机信息物理系统的统计可达性分析。《IEEE 计算机辅助设计的集成电路与系统学报》，43(11)：4250–4261，2024年。Sylvia L. Herbert、Mo Chen、SooJean Han、Somil Bansal、Jaime F. Fisac 和 Claire J. Tomlin。Fastrack：用于快速且有保证安全运动规划的模块化框架。载于 CDC，第1517–1522页，2017年。Saber Jafarpour、Akash Harapanahalli 和 Samuel Coogan。神经网络反馈回路的高效交互感知区间分析。《TAC》，2024年。Craig Knuth、Glen Chou、Necmiye Ozay 和 Dmitry Berenson。基于学习动力学的规划：通过 Lipschitz 常数对安全性和可达性提供概率保证。《IEEE 机器人学与自动化快报》，2021年。Craig Knuth、Glen Chou、Jamie Reese 和 Joseph Moore。未知欠驱动随机系统反馈运动规划的统计安全性与鲁棒性保证。载于 ICRA，2023年。Niklas Kochdumper 和 Stanley Bak。面向 Koopman 算子线性化控制系统的符合性综合。载于 CDC，第 7327–7332 页，2022 年。B. O. Koopman 和 J. V. Neumann。连续谱动力系统。载于《美国国家科学院院刊》，1932年。Thomas Lew、Lucas Janson、Riccardo Bonalli 和 Marco Pavone。一种简单高效的一般可达性分析采样算法。载于 Roya Firoozi、Negar Mehr、Esen Yel、Rika Antonova、Jeannette Bohg、Mac Schwager 和 Mykel Kochenderfer（编），L4DC，第168卷，第1086–1099页，2022年6月23日至24日。Lars Lindemann、Yiqi Zhao、Xinyi Yu、George J. Pappas 和 Jyotirmoy V. Deshmukh。利用保形预测进行形式化验证与控制。2024年。

Jin-Peng Liu、Herman Øie Kolden、Hari K. Krovi、Nuno F. Loureiro、Konstantina Trivisa 和 Andrew M. Childs。《耗散非线性微分方程的高效量子算法》。美国国家科学院院刊，118(35):e2026805118，2021年。Devesh Nath、Haoran Yin 和 Glen Chou。《通过经认证的局部稳定化，对生成式运动规划器进行形式化安全性验证与精化》。2025年。Shaowu Pan、Eurika Kaiser、Brian M. de Silva、J. Nathan Kutz 和 Steven L. Brunton。《PyKoopman：用于 Koopman 算子数据驱动近似的 Python 软件包》。开放源代码软件期刊，9(94):5881，2024年。doi：10.21105/joss.05881。Hyunsang Park、Vishnu Vijay 和 Inseok Hwang。《非线性系统的数据驱动可达性分析》。IEEE 控制系统快报，8：2661–2666，2024年。Nicholas Rober 和 Jonathan P. How。《用于神经反馈回路安全性验证的约束感知精化》。IEEE 控制系统快报，2024年。F. Sabatino。《四旋翼控制：建模、非线性控制设计与仿真》。理学硕士论文，KTH，瑞典斯德哥尔摩，2015年。John Schulman、Filip Wolski、Prafulla Dhariwal、Alec Radford 和 Oleg Klimov。《近端策略优化算法》，2017年。Lu Shi、Masih Haseli、Giorgos Mamakoukas、Daniel Bruder、Ian Abraham、Todd D. Murphey、Jorge Cort´es 和 Konstantinos Karydis。《机器人学习中的 Koopman 算子》。2024年。Sumeet Singh、Mo Chen、Sylvia L. Herbert、Claire J. Tomlin 和 Marco Pavone。《面向快速安全规划的模型失配鲁棒跟踪：一种 SOS 优化方法》。载于 WAFR，第 14 卷 Springer 高级机器人学会议论文集，第 545–564 页，2018年。Dawei Sun 和 Sayan Mitra。《Neureach：从仿真中学习可达性函数》。载于 Dana Fisman 和 Grigore Rosu 编，《系统构造与分析工具和算法》，第 322–337 页，尚，2022年。施普林格国际出版社。Omanshu Thapliyal 和 Inseok Hwang。《利用混合单调性对 Koopman 系统进行近似可达性分析》。IEEE Access，10：84754–84760，2022年。Emanuel Todorov、Tom Erez 和 Yuval Tassa。《MuJoCo：用于基于模型控制的物理引擎》。载于 IROS，第 5026–5033 页，2012年。Kaidi Xu、Zhouxing Shi、Huan Zhang、Yihan Wang、Kai-Wei Chang、Minlie Huang、Bhavya Kailkhura、Xue Lin 和 Cho-Jui Hsieh。《面向可扩展经认证鲁棒性及更多应用的自动扰动分析》。载于 NeurIPS，2020年。Huan Zhang、Tsui-Wei Weng、Pin-Yu Chen、Cho-Jui Hsieh 和 Luca Daniel。《利用一般激活函数高效认证神经网络鲁棒性》。载于 NeurIPS，第 4944–4953 页，2018年。

8. 附录

我们为理论结果提供证明（附录 8.1），并提供详细说明方法的算法框（附录 8.2），以及补充实验和实验详情（附录 8.3）。

8.1. 证明

引理 1（KRS 过近似）令 Γ 表示 *TKFL* 的闭环计算图，令 $\mathcal{X}_0$ 为初始集合。假设 *KRS* $\mathcal{X}_{0:T}$ 如(7)中所定义。那么，对于任意初始状态 $x_0 \in \mathcal{X}_0$，解码后的闭环轨迹(5)，即 $x_{0:T} = \Gamma(x_0, z_{0:T}^{ref}, u_{0:T-1}^{ref})$ 满足 $x_t \in \mathcal{X}_t$，对所有 $t \in \mathcal{T}$ 均成立。

证明闭环图 Γ 是一个计算图 G，其定义见命题1。使用`auto LiRPA`库Xu 等人(2020)，我们计算出可靠的向量值仿射界函数$\underline{\Gamma}_t$和 $\overline{\Gamma}_t$（分别对应于 G 和 G，见命题1）针对第 t个状态输出。

根据命题1，这些界保证是可靠的。也就是说，对于任意特定输入$x_0 \in \mathcal{X}_0$，预测状态 $\hat{x}_t = \Gamma_t(x_0,\ z_{0:T}^{\text{ref}},\ u_{0:T-1}^{\text{ref}})$ 满足

$$\underline{\Gamma}_t(x_0, z_{0:T}^{\text{ref}}, u_{0:T-1}^{\text{ref}}) \le \hat{x}_t \le \overline{\Gamma}_t(x_0, z_{0:T}^{\text{ref}}, u_{0:T-1}^{\text{ref}}).$$

KRS $\mathcal{X}_t$在式（7）中定义为区间 $\mathsf{Int}(\hat{x}_t^L,\ \hat{x}_t^U)$，其中 $\hat{x}_t^L = \min_{x_0' \in \mathcal{X}_0} \Gamma_t(x_0',\ \ldots\ldots)$ 和 $\hat{x}_t^U = \max_{x_0' \in \mathcal{X}_0} \overline{\Gamma}_t(x_0',\ \ldots\ldots)$。根据最小值和最大值的定义，我们有

$$\hat{x}_t \ge \underline{\Gamma}_t(x_0, \dots) \ge \min_{x_0' \in \mathcal{X}_0} \underline{\Gamma}_t(x_0', \dots) = \hat{x}_t^L,$$
$$\hat{x}_t \le \overline{\Gamma}_t(x_0, \dots) \le \max_{x_0' \in \mathcal{X}_0} \overline{\Gamma}_t(x_0', \dots) = \hat{x}_t^U.$$

因此，$\hat{x}_t \in \mathsf{Int}(\hat{x}_t^L, \hat{x}_t^U) = \mathcal{X}_t$。由于这对所有 $t \in \mathcal{T}$ 均成立，因此轨迹 $\hat{x}_{0:T}$包含于 $\hat{\mathcal{X}}_{0:T}$ 中。 ■

定理 1（CKRS 覆盖保证）设 $\mathcal{X}_{0:T}$为式(7）中定义的 *KRS*，并设 $\bar{e}_t = [\bar{e}_{t,1}, \dots, \bar{e}_{t,n}]^\top \in \mathbb{R}^n$。将保形 *Koopman* 可达集（*CKRS*）定义为

$$\overline{\mathcal{X}}_{0:T}^{1-\delta} \equiv \left\{\hat{\mathcal{X}}_t \oplus \operatorname{diag}(\bar{e}_t)\mathcal{B}_1(0)\right\}_{t=0}^{T}, \tag{11}$$

其中，$\operatorname{diag}(\bar{e}_t) \in \mathbb{R}^{n\times n}$ 是一个对角线上为 $\bar{e}_t$ 的对角矩阵。随后，对于一个新的参考轨迹 $x_{0:T}^{ref}$，$x_0^{(0)}$：从相同分布中抽取，*CKRS* 以至少 $1-\delta$：的概率包含由 (1b) 生成的真实闭环系统轨迹

$$\mathbb{P}\Big(\bigwedge_{t=0}^{T} x_t^{(0)} \in \overline{\mathcal{X}}_t^{1-\delta}\Big) \ge 1-\delta.$$

证明 该证明将确定性包含保证与概率覆盖保证相结合。

步骤 *1*：*KRS* 包含性。 根据引理 1，解码后的闭环提升轨迹 $\hat{x}_t^{(0)}$，j 确定性地包含于超矩形 KRS 中：

$$\hat{x}_{t,j}^{(0)} \in \hat{\mathcal{X}}_{t,j} = \mathsf{Int}(\hat{x}_{t,j}^L, \hat{x}_{t,j}^U), \quad \forall t \in \mathcal{T}, \quad \forall j \in \{1, \dots, n\}.$$

步骤 *2*：概率覆盖。保形预测保证要求非一致性分数具有可交换性 $\{R^{(i)}\}_{i=0}^{K^{\text{cal}}}$。在我们的框架中，所有校准轨迹和测试轨迹共享相同的参考$(z_{0:T}^{\text{ref}},\ u_{0:T-1}^{\text{ref}})$和控制器 G_t，且它们的初始状态 $x_0^{(i)}$ 从 $\mathcal{B}_\epsilon(x_0^{\text{ref}})$中独立同分布采样。由于解码后的闭环提升轨迹（8）、真实闭环轨迹（9）和非一致性分数都是 $x_0^{(i)}$ 的确定性函数，因此得到的分数 $R^{(i)}$ 也独立同分布，进而具有可交换性。根据保形预测保证，事件

$$\mathcal{E} \equiv \{|e_{t,j}^{(0)}| \le \bar{e}_{t,j}, \forall t, j\}$$

发生的概率至少为 $1-\delta$。
步骤 *3*：组合保证。真实状态可分解为 $x_{t,j}^{(0)} = \hat{x}_{t,j}^{(0)} + e_{t,j}^{(0)}$。当 $\mathcal{E}$发生时，我们有

$$\hat{x}_{t,j}^L - \bar{e}_{t,j} \le \hat{x}_{t,j}^{(0)} + e_{t,j}^{(0)} \le \hat{x}_{t,j}^U + \bar{e}_{t,j},$$

根据定义，这意味着 $x_{t,j}^{(0)} \in \mathcal{X}_{t,j}^{1-\delta}$。由于这对所有 t 和 j均成立，因此轨迹 $x_{0:T}^{(0)}$ 以至少 $1-\delta$的概率包含在 CKRS $\mathcal{X}_{0:T}^{1-\delta}$ 中。 ■

8.2. 算法

算法 1 详细说明了计算 CKRS 的过程。首先，将规划器提供的参考轨迹 $x_{0:T}^{\text{ref}}$提升到潜在空间，通过最小二乘法计算前馈控制u^{ref}，并设计时变 LQR 控制器以跟踪该参考轨迹。为考虑学习型Koopman动力学与真实系统之间的差异，算法从系统滚动轨迹中生成校准数据，并利用基于 SNSA 的共形预测推导出具有统计有效性的误差界 $\bar{e}_t$。与此同时，通过在提升后的 Koopman 状态空间中构建闭环跟踪动力学的计算图，并使用 auto LiRPA 计算线性 Koopman 动力学的可达集，得到名义 KRS $\mathcal{X}_{0:T}$。最后，算法通过闵可夫斯基和将推导出的共形误差界叠加到名义 KRS 上，从而返回保形 Koopman 可达集（ϵKRS）$\mathcal{X}_{0:T}^{1-\delta}$。关键的是，CKRS 能以 $1-\delta$的概率保证覆盖由真实系统 生成的轨迹。

8.3. 实验详情与补充实验

以下部分介绍了本文所用的解析动力学（附录 8.3.1）以及更多实验结果。具体而言，我们将本文方法与以下基线方法进行基准测试：利用多项式提升的 Koopman 动力学学习方法（附录 8.3.2）、学习式可达集预测方法（附录 8.3.3），以及不利用基于 Koopman 的传播的基于共形预测的可达集计算方法（附录 8.3.4）。我们还展示了更多实验图，包括独轮车系统（附录 8.3.5）、平面四旋翼系统（附录 8.3.6）、MuJoCo 跳跃器系统（附录 8.3.8）和 MuJoCo 游泳者系统（附录 8.3.9）中按状态逐项展示的完整可达集可视化结果。

算法 1 在线保形 Koopman 可达集（CKRS）

输入：Koopman 模型 (ϕ, ψ, K_A, K_B)；真实动力学 f（仿真器）；参考轨迹 $x^{\text{ref}}_{0:T}$；初始集合 $\mathcal{X}_0 = \mathcal{B}_\epsilon(x^{\text{ref}}_0)$；RSOA 置信度 $1-\delta$；校准样本规模 $M^\lambda, K^{\text{cal}}$；代价矩阵$Q, R$；验证工具 auto LiRPA 输出：保形 Koopman 可达集 $\mathcal{X}^{1-\delta}_{0\,T}$：（CKRS）// 1. 初始化 1：$z^{\text{ref}}_{0:T} \leftarrow \phi(x^{\text{ref}}_{0:T})$ // 编码参考状态轨迹 2：$u^{\text{ref}}_{0:T-1} \leftarrow$公式（2）// 计算前馈控制序列 3：$G_{0:T-1} \leftarrow$线性二次调节器$(K_A, K_B)$ // 计算 LQR 增益 4：$\pi(\cdot) \leftarrow$公式（3）// 使用 LQR 和前馈输入计算提升控制器 // 2. 通过保形预测（CP）计算误差界 $\bar{e}_t$ 5：$\mathcal{D}_E, \mathcal{D}_N \leftarrow$ GenerateCalibrationData$(M^\lambda, K^{\text{cal}})$ 6：$\bar{e}_t \leftarrow$SNSA$(\mathcal{D}_N, \mathcal{D}_E)$ ˆ // 3. 名义 KRS $\mathcal{X}_{0:T}$7：$\Gamma(\cdot) \leftarrow \psi$(PropagateLatent$(\phi(\cdot)\ z^{\text{ref}}, u^{\text{ref}}, G)$) // 定义计算图 8：ˆ $\mathcal{X}_{0:T} \leftarrow$AutoLiRPA$(\Gamma, \mathcal{X}_0)$ // 通过计算图计算 KRS 9：$\mathcal{X}^{1-\delta}_{0}{}_{:T} \leftarrow$ˆ$\mathcal{X}_{0:T} \oplus$ diag$(\bar{e}_t)\mathcal{B}_1(0)$ // 4. 合并以获得 CKRS 并返回 10：返回 $\mathcal{X}^{1-\delta}_{0\,T}$：

—

—　　—

—

8.3.1. 动力学。

对于独轮车实验（$x \in \mathbb{R}^3$、 $u \in \mathbb{R}^2$），我们使用了通过前向欧拉法进行时间离散化的动力学模型（$dt = 0.1$秒）

$$\dot{\mathbf{x}} \;=\; f(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} u_1 \cos(x_2) \\ u_1 \sin(x_2) \\ u_2 \end{bmatrix}. \tag{12}$$

对于平面四旋翼实验（$x \in \mathbb{R}^6, u \in \mathbb{R}^2$），动力学模型采用参数 $m = 0.5$kg、 $g = -9.81$ m/s^2、 $I_y = 0.01$kg $\cdot$ m^2，并以 $dt = 0.05$s 为时间离散步长：

$$\dot{\mathbf{x}} \;=\; f(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} x_4 \\ x_5 \\ x_6 \\ -\frac{u_1}{m}\sin(x_3) \\ g + \frac{u_1}{m}\cos(x_3) \\ \frac{u_2}{I_y} \end{bmatrix} \tag{13}$$

对于三维四旋翼实验，动力学采用参数 $m = 1\text{kg}$、$g = -9.81\text{m/s}^2$ $I_{x,y,z} = [0.5, 0.1, 0.3]$ kg·m^2，时间离散步长为 $dt = 0.025\,\text{s}$：

$$\dot{\mathbf{x}} = f(\mathbf{x}, \mathbf{u}) = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \\ q \cdot \sin(\phi)/\cos\theta + r \cdot \cos\phi/\cos\theta \\ q \cdot \cos\phi - r \cdot \sin\phi \\ p + q \cdot \sin\phi \cdot \tan\theta + r \cdot \cos\phi \cdot \tan\theta \\ \frac{u_1}{m} \cdot (\sin\phi \cdot \sin\psi + \cos\phi \cdot \cos\psi \cdot \sin\theta) \\ \frac{u_1}{m} \cdot (\cos\phi \cdot \sin\phi - \cos\phi \cdot \sin\psi \cdot \sin\theta) \\ g + u_1 \cdot (\cos\phi \cdot \cos\theta)/m \\ ((I_y - I_z)/I_x) \cdot q \cdot r + \frac{u_2}{I_x} \\ ((I_z - I_x)/I_y) \cdot p \cdot r + \frac{u_3}{I_y} \\ ((I_x - I_y)/I_z) \cdot p \cdot q + \frac{u_4}{I_z} \end{bmatrix} \tag{14}$$

8.3.2. 神经网络 Koopman 提升与多项式提升的比较

我们比较了多项式提升扩展动态模态分解基线与我们提出的神经网络提升方法在 Hopper 系统上的均方根误差（RMSE）。对于多项式提升，我们使用 PyKoopman（Pan 等人，2024年），采用三次单项式，并使用相同的训练数据和预处理方法。在开环评估期间，多项式模型会变得数值不稳定，也就是说，其预测值会无限增大，并在第八个时间步发散，因此我们仅报告表2中前 5 个预测步的均方根误差。相比之下，我们的神经网络提升方法产生了稳定的潜在动力学模型，能够在连续三个完整跳跃过程中准确跟踪，从而支持更长时域的评估。这些结果表明，对于某些系统，例如具有强非线性、富接触动力学的系统（如 MuJoCo 跳跃器），固定的多项式基并不足够，而神经网络提升函数能够捕捉必要的结构，并在较长的滚动轨迹中保持稳定。总体而言，这些结果推动我们开发将神经网络纳入计算流程的 Koopman 可达集计算框架；我们通过 `auto LiRPA`实现了这一点。

状态	多项式提升	Ours
z 坐标（高度）	2728.481591	0.009848
躯干角度	117.374908	0.004201
大腿关节	2832.025934	0.013581
腿部关节	10791.162550	0.015690
足关节	8063.167273	0.017901
x 坐标速度	321013.541866	0.047526
z 坐标速度	1691803.060196	0.059763
躯干速度	4512148.680118	0.082879
大腿速度	596074.452102	0.073777
腿部速度	6164730.041985	0.164258
脚部速度	135554800.971296	0.259034

表 2：Hopper（11 维）系统中，学习得到的模型和多项式拟合模型所执行滚动轨迹的均方根误差对比。

8.3.3. 与 Bansal 和 Tomlin（2021年）基线方法的比较。

图 5 对比了基线 DeepReach 方法（Bansal 和 Tomlin，2021年）与 ScaRe-Kro 在安全违规方面的定性表现。虽然正文中的统计数据表明基线方法的安全率较低，但这一可视化明确呈现了“可达—避障”失败模式：由 DeepReach 值函数生成的橙色轨迹明显偏入红色障碍区域。这凸显了依赖近似神经偏微分方程解来处理安全关键约束的风险。相比之下，我们的方法计算出的可达集（蓝色所示）形成了一条狭窄且经过验证的通道，以较高的校准概率包含真实值轨迹；这在视觉上证实，控制器能够在遵守概率界的同时，成功引导动力学避开危险区域。

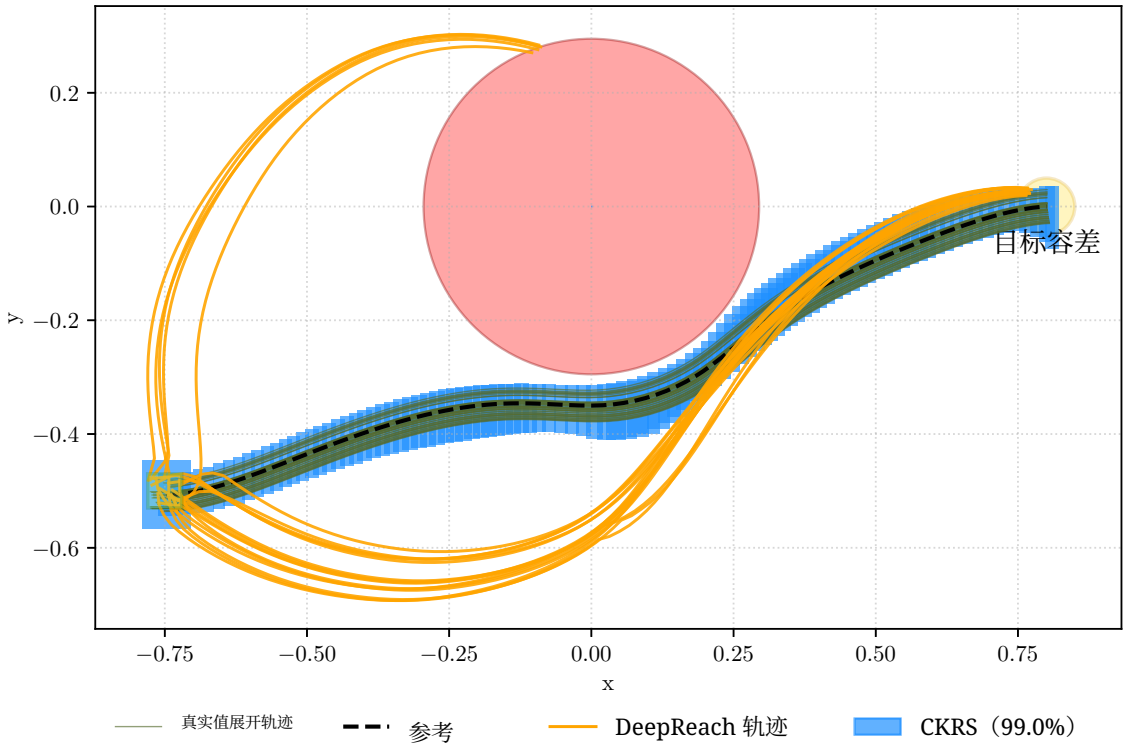


图5：基于 DeepReach 的控制器得到的闭环轨迹（橙色）可能无法满足可达-避障约束（到达黄色目标区域，同时避开红色不安全集）；而我们的控制器（绿色滚动轨迹）以 0.99 的概率保证安全性，并确保轨迹包含于蓝色 CKRS 中。

8.3.4. 与 Hashemi 等人 （2024年）基线方法的比较。

图 6 逐状态展示了基线 CP-NNV（Hashemi 等人，2024年）方法（黄色）与我们提出的方法（蓝色）计算出的可达集之间的保守性差距。第 6 节中的表格结果量化了体积差异，而图6则直观展示了这种差异的来源。总体而言，我们的方法与

Hashemi 等人（2024年）的一个关键区别在于，我们使用学习型多步预测器而非学习型Koopman动力学来计算可达集。我们注意到，CP-NNV方法得到的可达集保守性更强，尤其是在角速度（p，q，r）和角度（ϕ，θ，ψ）维度上。我们认为，这是因为在长时域内使用学习型多步预测器难以准确计算可达集。相比之下，ScaRe-Kro 在全部 12 个维度上始终保持紧致的界限，表明 Koopman 线性化有效缓解了导致标准神经网络验证中出现过度近似的“包裹效应”。

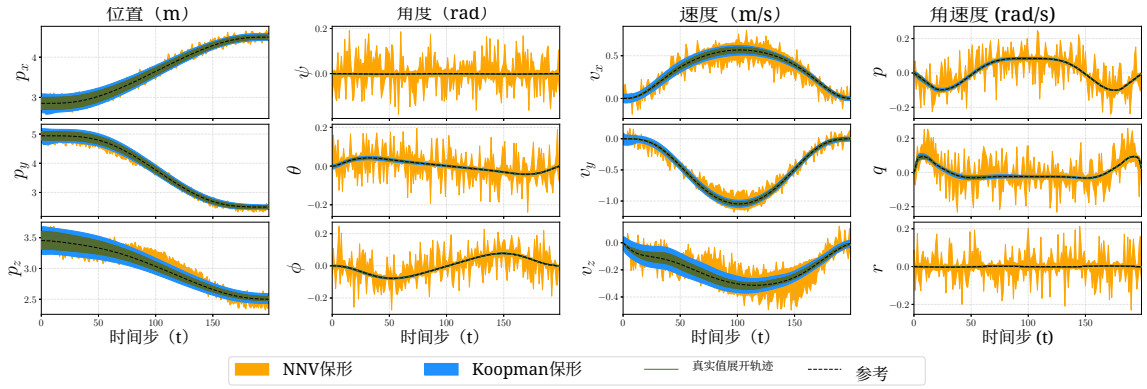


图6：CP NNV方法与我们方法的基线比较

8.3.5. 独轮车系统的其他结果。

图7 展示了独轮车的 CKRS 在空间中的演变。除了第 6 节中的数值覆盖率外，该图还可作为几何合理性检查，展示可达集在通过狭窄约束区域时包含参考轨迹（黑色虚线）和闭环轨迹（绿色）的能力。蓝色管状区域是 CKRS，它表示系统至少以 $1-\delta$ 的概率保证留在其中的区域。这表明，即使共形化步骤导致可达集膨胀，所得集合仍足够紧凑，可以验证安全性，尽管闭环运行轨迹非常接近红色障碍物。

8.3.6. 平面四旋翼系统的其他结果。

图 8 展示了六维平面四旋翼飞行器的完整维度轨迹演化。这些图单独呈现了 Koopman 跟踪控制器在欠驱动系统上的性能。值得注意的是，可达集（浅蓝色）及其保形化对应集（深蓝色）能够准确捕捉所有状态维度上的行为，并界定闭环轨迹。这证实，提升的Koopman空间中学得的线性结构保留了原始非线性动力学中的关键耦合效应，同时保形界有效吸收了预测 KRS 中的残余误差。我们还注意到，在此示例中，KRS 的共形化对可达集估计总体体积的影响相对较小，不过这并非总是如此，例如图 11 所示的跳跃器系统。此外，本实验展示了离线 CP 膨胀界（见第 5.4 节）能够泛化到多条参考轨迹：我们共针对四条参考轨迹计算 CKRS（图 8a-d），并表明

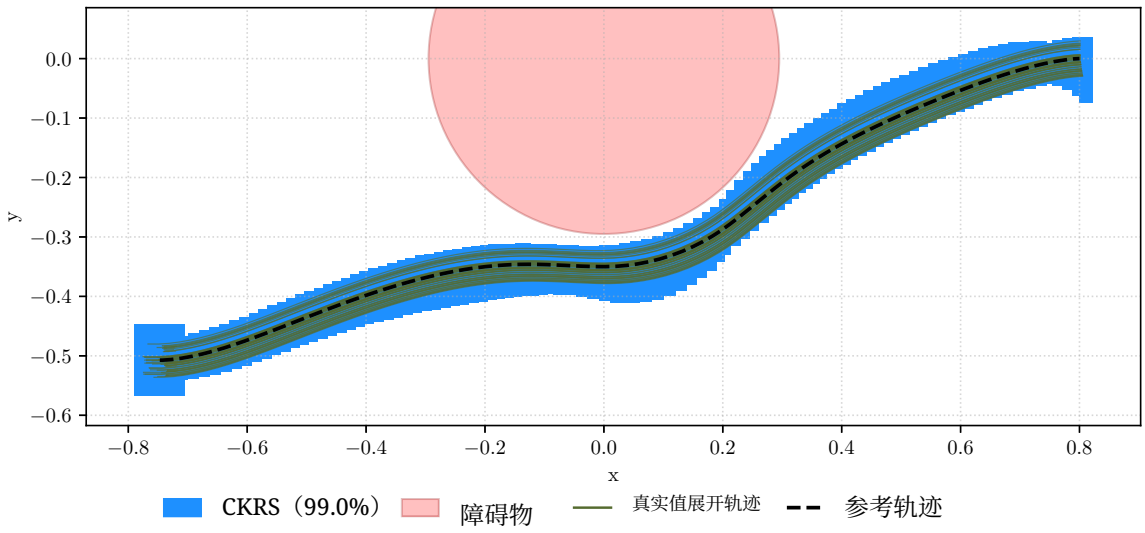


图7：针对三维独轮车模型计算的 CKRS，即使闭环轨迹与障碍物（红色）的距离很小，也能够进行安全性验证。

由此得到的闭环轨迹始终处于计算出的集合内。这表明，单次离线 CP 校准可重复使用，从而高效膨胀在线计算的可达集。

8.3.7. 三维四旋翼系统的补充结果。

为验证离线 CP 膨胀界（见第 5.4 节）能否推广到多条参考轨迹，我们针对三条不同的参考轨迹计算 CKRS，并展示由此得到的闭环轨迹均处于计算出的集合内（图 9）。这表明，只需进行一次离线 CP 校准，即可重复使用该校准，高效地膨胀在线计算的可达集。我们还展示了其中两个 CKRS，以及四旋翼飞行器绕行的障碍物（图 10）；由于 RSOA 与障碍物不相交，因此可以保证闭环系统在跟踪任一参考轨迹时都不会发生碰撞。

8.3.8. MuJoCo Hopper 系统的补充结果。

图11展示了 11 维 MuJoCo 跳跃器（Todorov 等人，2012）系统在向前跳跃动作期间的完整状态空间演化。该系统具有混合动力学的特征，在地面冲击期间会发生非平滑过渡，因此带来了独特的挑战。在“高度 z”和“躯干角度”轨迹中，可以清晰地观察到步态的周期稳定性；ScaRe-Kro 框架成功捕捉了与飞行与支撑阶段相关的复杂动力学。值得注意的是，尽管位置状态始终紧密限制在参考轨迹（黑色虚线）附近，速度维度（例如“足部速度”）却表现出保形界必要的扩张。这种扩张正确地反映了接触力和冲击时的建模不确定性所导致的状态瞬时跳变；与此同时，真实值滚动轨迹始终严格包含在预测管内，从而验证了 Koopman 跟踪控制器在不连续接触动力学下的鲁棒性。

8.3.9. MuJoCo 游泳者系统的其他结果。

图12 展示了对 28 维 MuJoCo Swimmer Todorov 等人 (2012) 的分量可达性分析，这是对我们框架最高维度的验证。图中展示了运动所需的高度耦合波动步态，这从连杆关节和速度沿链条呈现的相位偏移振荡中可见一斑。尽管状态维数显著增加，CKRS 仍在全部 28 个维度上持续覆盖闭环轨迹。游泳者连杆段上的界限均匀且紧密，这证实了所学习的全局线性动力学结构能够有效表征多体动力学，从而在保持计算可处理性的同时，为复杂的关节机器人实现安全的长时域验证。

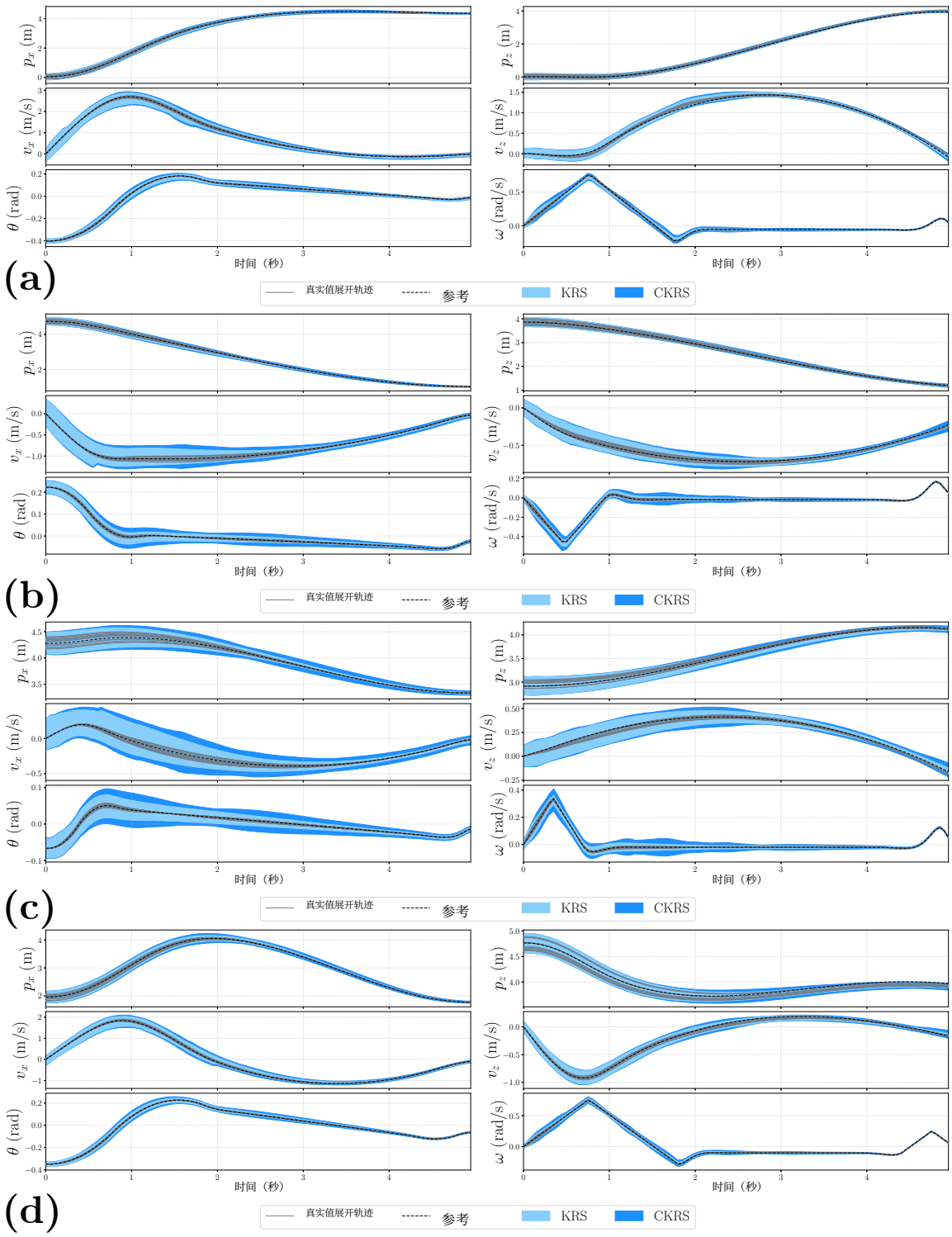


图 8：在经过离线校准的保形预测界限下，针对二维四旋翼模型的四条参考轨迹计算得到的 CKRS。

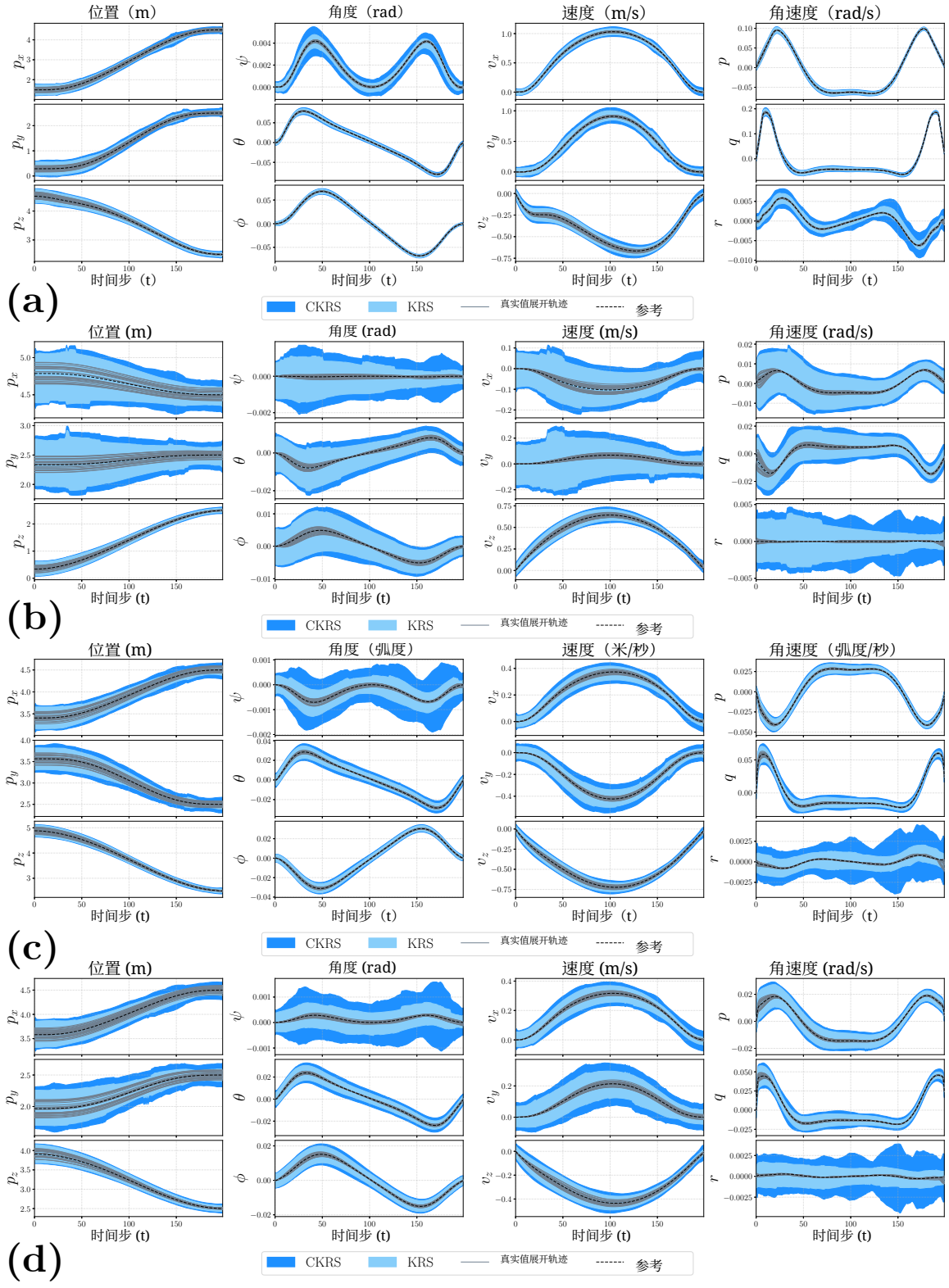


图9：高度 针对不同参考轨迹计算的三维四旋翼无人机模型的替代 CKRS。

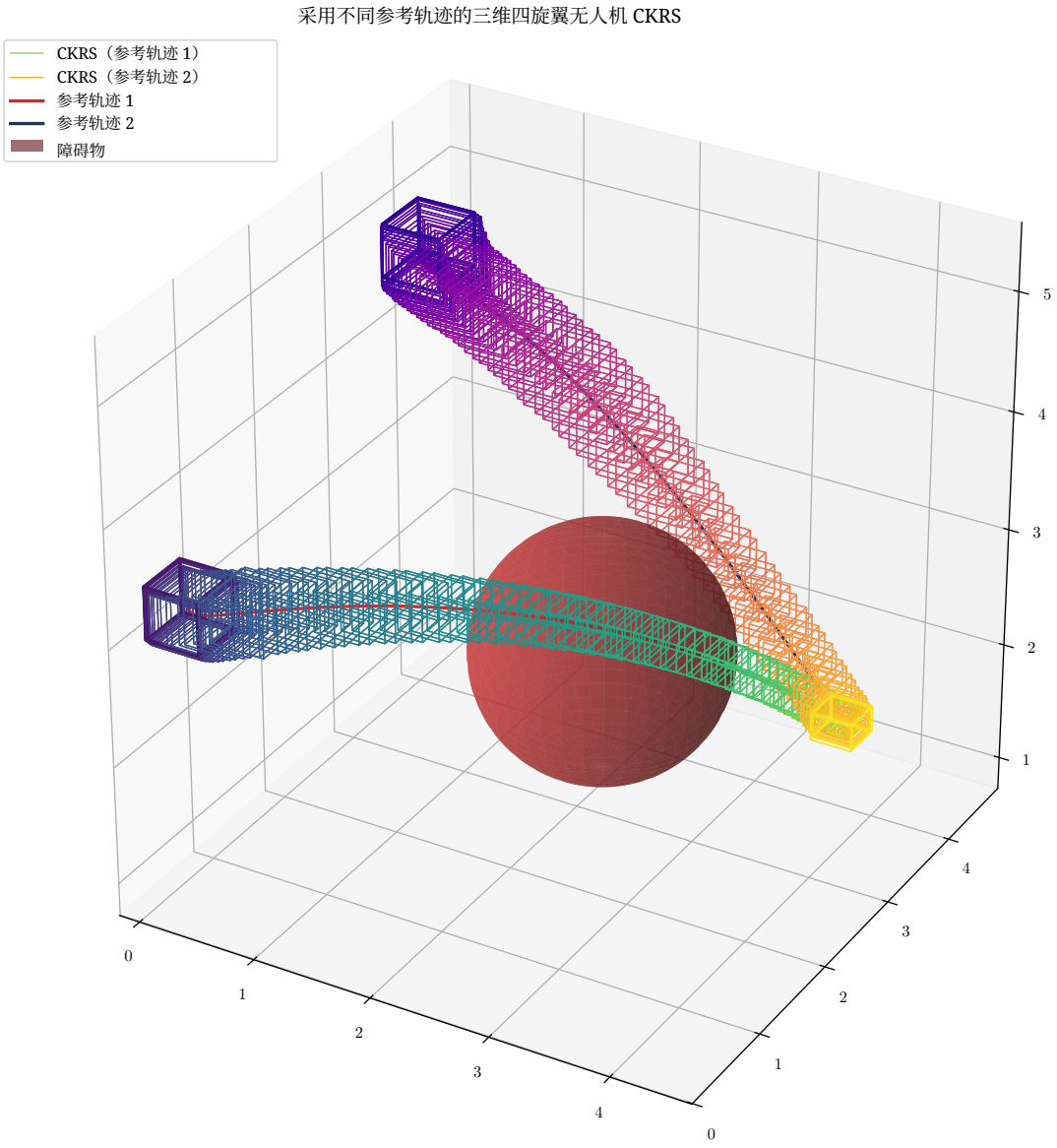


图10：针对三维四旋翼无人机模型（位置分量）计算的 CKRS，并叠加显示障碍物。

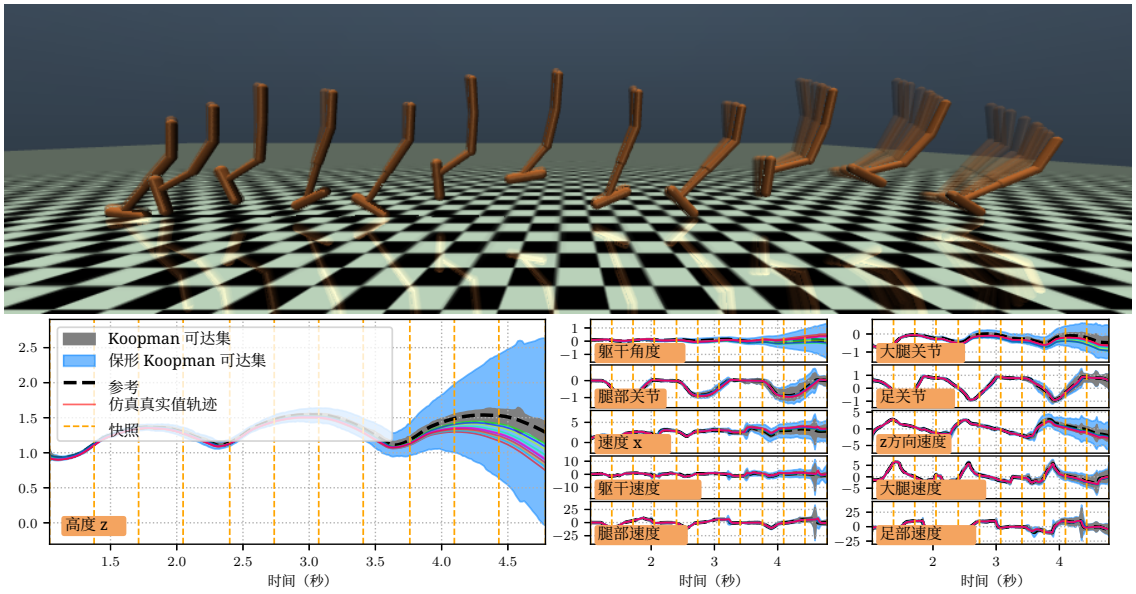


图11：针对 11维 MuJoCo Hopper 计算的 CKRS 和 KRS。细的多色线条表示不同的闭环轨迹滚动轨迹。

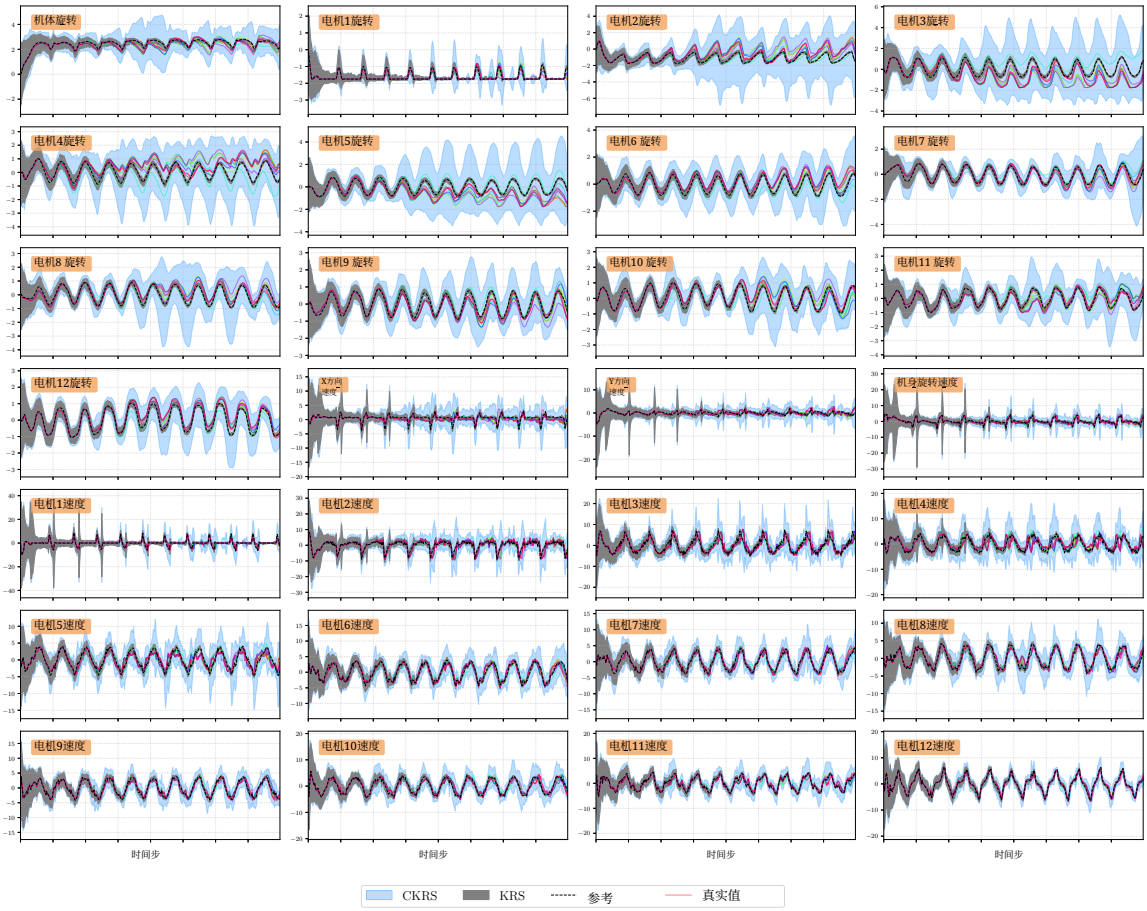


图12：为28维 MuJoCo 游泳者计算的 CKRS 和 KRS。细小的多色线条表示不同的闭环轨迹滚动。