arXiv:2602.04720v1 [math.DS] 4 Feb 2026

基于数据的利用 Koopman算子的无偏预测器

Roland Schurig[1], Pieter van Goor[2],Karl Worthmann[3],Rolf Findeisen[1]

摘要——**Koopman算子及其数据驱动的近似方法（如扩展动态模态分解（EDMD））广泛用于分析、建模和控制非线性动力系统。然而，当无法从有限数据中识别真实的 Koopman 特征函数时，多步预测可能会受到结构性误差和系统性偏差的影响。为解决这一问题，我们分析了多步预测残差的一阶和二阶矩。通过将残差分解为单步近似误差的贡献以及累积误差传播的影响，我们推导出一个刻画这些影响的闭式表达式。基于此分析，我们开发了一种新颖且计算高效的算法，可确保无偏性并降低所得预测器的方差。我们通过数值仿真验证了所提出的方法，结果表明，与标准 EDMD 相比，该方法具有更优的不确定性特性。这些结果为具备不确定性感知能力的无偏 Koopman 预测框架奠定了基础，并且该框架可扩展至受控系统和随机系统。**

I. 引言

数据驱动的动力系统分析、建模与控制仍是一项颇具挑战的任务，尤其是在存在非线性的情况下。因此，人们提出了许多基于将动力学 提升 到更高维空间的技术，以借助线性系统理论中的工具对其演化和结构特性进行建模与分析。在这一领域，一种尤为成功的方法是扩展动态模态分解 [1] 的各种算法实现，参见例如 [2]。其分析的理论基础是Koopman算子；自开创性论文 [3], [4], 发表以来，它已成为一项关键工具，另见近期综述 [5]。如今，针对无限字典（投影）和无限数据（估计）极限 [6], ，以及有限样本情形，均已有误差分析；例如， [7] 和 [8] 分别给出了扩展动态模态分解和核扩展动态模态分解的近期结果。

近期研究进展聚焦于 Koopman框架中的可靠预测，旨在解决重复预测过程中出现的不一致问题。在 [9], 一种最大似然重投影方法被提出，用于保持与Koopman 流形的一致性，并将重投影解释为高斯残差下的统计最优操作。该工作 [10]提出了一个在 Grassmann 流形上进行字典学习的几何框架，用于优化可观测子空间，以实现近似的Koopman 不变性。

[1] 控制与信息物理系统实验室，达姆施塔特工业大学，德国， {roland.schurig, rolf.findeisen}@iat.tu-darmstadt.de； [2] 航空航天、机械与机电工程学院，悉尼大学，澳大利亚，pieter.vangoor@sydney.edu.au； [3] 基于优化的控制组，数学研究所，伊尔梅瑙工业大学，德国，karl.worthmann@tu-ilmenau.de

该工作聚焦于提升系统能够以自回归方式进行可靠预测的步数。在有限字典和有限数据的情况下，对提升空间中残差的一阶和二阶矩进行了概率分析。该分析针对提升空间中的通用预测器展开，并指出，由EDMD衍生的预测器在单步预测意义下是最优的，因为它既是无偏的，又是方差最优的。

分析给出了期望残差的解析闭式表达式，使我们能够将这一结果扩展到更长的预测时域。分析展示了一种结构化方法：通过使相关残差相对于由所收集数据点诱导的经验测度的期望为零，来保证多步预测器的无偏性。这导出了多步残差在期望意义下消失的明确条件，并阐明了剩余的自由度。

随后，整体残差的演化可用一个线性系统和局部残差来表示，而局部残差可通过选择合适的预测器加以调节。我们推导出线性预测器空间中的一组基，使无偏性条件一目了然。随后利用剩余的自由度来降低多步方差。我们尤其关注数值适用性：所提出的预测器具有闭式解，可利用数值线性代数工具高效计算。

本文的研究结果为理解基于Koopman的模型中多步预测误差如何累积提供了严谨依据，并揭示了提升预测器在多个时间步内保持无偏性的条件。这带来了一种实用且计算高效的方法，可直接利用有限数据构造无偏且方差降低的多步预测器。因此，所提出的框架夯实了数据驱动Koopman预测的理论基础，并支持更可靠的长时域预测，这对于控制、仿真和不确定性感知决策等应用至关重要。

本文结构如下。在第二节中，介绍了Koopman框架，并将EDMD算法作为Koopman算子的数据驱动近似加以阐述。由此引出第三节中对提升动力学的概率分析，重点关注由扩展动态模态分解所优化的单步行为。在此基础上，我们在第三节中给出预测器在多个时间步上无偏性的充要条件，而这并不能由EDMD预测器保证。此外，一种利用剩余自由度

以实现方差最小化的可能途径得到了概述。所提出的预测器在第VI节中介绍，并在第VII节中给出一个数值示例。我们在第VIII节中作出结论。

记号：记号采用标准约定。自然数记为 $\mathbb{N}$，且 $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$。所有 $(m \times n)$-矩阵构成的 $\mathbb{R}$ 上的欧几里得空间记为 $\mathbf{M}(m \times n, \mathbb{R})$，其具有标准的向量空间结构和内积 $\langle A, B\rangle := \operatorname{tr}(A^\top B)$。由此得到弗罗贝尼乌斯范数 $\|A\|_{\mathrm{F}} := \sqrt{\langle A, A\rangle}$。随机变量 X 的期望和方差分别记为 $\mathbb{E}(X)$ 和 $\operatorname{var}(X)$。

II. Koopman框架中的压缩

我们考虑离散时间非线性系统

$$x(t+1) = f(x(t)), \qquad x(0) = \bar{x}, \tag{1}$$

其具有连续的状态转移映射 $f\colon \mathbb{R}^n \to \mathbb{R}^n$。为便于说明，我们假设 $X \subset \mathbb{R}^n$ 非空、紧致，并且在 f 下正向不变，即 $x \in X$ 蕴含 $f(x) \in X$。从初始状态 $\bar{x} \in X$ 出发的轨迹记为 $x(\cdot, \bar{x})\colon \mathbb{N}_0 \to X$。

线性*Koopman*算子 K 将函数 $\varphi \in L^2(X, \mathbb{R})$映射为另一个函数 $K\varphi \in L^2(X, \mathbb{R})$，并由恒等式定义

$$(K\varphi)(x) = (\varphi \circ f)(x) \qquad \forall\, x \in X.$$

函数 $\varphi \in L^2(X, \mathbb{R})$称为可观测量。Koopman算子提供了一种通过线性视角来看待动力系统(1)的方法；然而，它是无限维的。

为了获得 K的有限维近似，考虑 m 个线性无关且连续的可观测量$\psi_1, \ldots, \psi_m \in L^2(X, \mathbb{R})$，并用 $V \subset L^2(X, \mathbb{R})$表示它们的张成空间。该子空间是有限维的，但 $\psi \in V$并不必然意味着经传播的可观测量$K\psi$ 也包含在 V中。例外是所谓的Koopman特征函数以及包含在 V 中的不变子空间，它们已被证明极为有用，参见例如 [11],[12], [13], ，但如果可观测量对应于所考虑被控对象的传感器或输出，则它们可能不可观测。为缓解由于子空间 V 在动力系统预测中缺乏Koopman 不变性而产生的问题，令 P为 $L^2(X, \mathbb{R})$的一个投影，其值域为$\operatorname{ran} P = V$。于是，由下式给出的线性映射 $C\colon V \to V$

$$C\psi := PK\psi$$

定义了 V上 K 的压缩，可表示为一个（$m \times m$）矩阵。若子空间 V 是*Koopman* 不变的，即包含关系 $K\psi \in V$ 对所有 $\psi \in V$均成立，则压缩 C 对应于K 在 V上的限制 $K|_V$。否则会出现投影误差，这是由于忽略了 $K\psi$ 中不属于 V的分量所致。

术语扩展动态模态分解（EDMD）指一种数据驱动方法，用于计算压缩后的 Koopman 算子 C 的矩阵表示。首先，定义与基相关的提升映射 $\Psi\colon X \to \mathbb{R}^m$ $(\psi_1, \ldots, \psi_m)$

$$\Psi(x) := \begin{pmatrix} \psi_1(x) & \ldots & \psi_m(x) \end{pmatrix}^\top,$$

我们假设其为单射（一对一）。接下来，考虑采样状态

$$x_1, \ldots, x_L \in X \tag{2}$$

其中 $L \geq m$，且状态 x_i 是一个随机变量的独立同分布实现。然后，利用该数据集，定义双线性形式 $b\colon L^2(X, \mathbb{R}) \times L^2(X, \mathbb{R}) \to \mathbb{R}$ 如下：

$$b(\psi, \varphi) := \frac{1}{L} \sum_{i=1}^{L} \psi(x_i)\varphi(x_i).$$

此外，对于固定的时间范围 $N \in \mathbb{N}$，我们假设对于每个 $i \in \{1, \ldots, L\}$，都测量了状态轨迹 $x(t; x_i)$，$t \in \{0, 1, \ldots, N\}$。对于每个 $t \in \{0, \ldots, N\}$，相应的数据矩阵定义为

$$G_t := \begin{pmatrix} \Psi(x(t, x_1)) & \ldots & \Psi(x(t, x_L)) \end{pmatrix}.$$

标准形式的扩展动态模态分解算法仅使用矩阵 G_0 和 G_1，即动力系统 (1)[1] 的一步行为信息。我们假设G_0 满行秩，因此 b 在 V上是非退化的，这意味着可以通过以下“正交”性质定义 $L^2(X, \mathbb{R})$的投影 P

$$\forall \varphi \in L^2(X, \mathbb{R})\colon \forall \psi \in V\colon b(\varphi - P\varphi, \psi) = 0.$$

因此，该投影并非相对于 $L^2(X, \mathbb{R})$ 上通常的内积正交，而是相对于数据诱导的双线性形式 b 正交。C相对于基 $(\psi_1, \ldots, \psi_m)$ 的矩阵表示 K为

$$\hat{K} = (G_0 G_0^\top)^{-1} G_0 G_1^\top.$$

这一表示还可以解释为最小二乘问题的唯一极小解

$$\min_{\tilde{K}} \left\{ \|G_1 - \tilde{K}^\top G_0\|_{\mathrm{F}}^2 \right\}.$$

关于扩展动态模态分解及其变体的一般参考资料，请参阅 [1], [2] 。

该压缩可用于在提升空间 M 中传播 (1) 的状态：$= \operatorname{im} \Psi$。对于每个初始状态 $\bar{x} \in X$，可以逐分量应用Koopman算子，以近似后继状态 $\bar{y} = f(\bar{x})$

$$\begin{aligned} \Psi(\bar{y}) = (\Psi \circ f)(\bar{x}) &= ((K\psi_1)(\bar{x}), \ldots, (K\psi_m)(\bar{x})) \\ &\approx ((C\psi_1)(\bar{x}), \ldots, (C\psi_m)(\bar{x})) \\ &= \hat{K}^\top \Psi(\bar{x}), \end{aligned}$$

这意味着，$\hat{K}^\top \Psi(\bar{x})$ 是对 $\Psi(f(\bar{x}))$ 的估计。需要注意的是，虽然由于假设 Ψ 是单射的，可以从 $\Psi(\bar{y})$ 中恢复 $\bar{y}$，但对于 $\hat{K}^\top \Psi(\bar{x})$ 可能无法做到这一点。用其压缩形式 C 替代 Koopman算子 K 会引入上述投影误差，因此

[1]我们评述了第VI节中使用所有矩阵 G_t 的EDMD算法。

可能不存在$\hat{x} \in X$使得 $\Psi(\hat{x}) = \hat{K}^\top \Psi(\bar{x})$，也就是说，我们有$\hat{K}^\top \Psi(\bar{x}) \notin M$。因此，在应用 Ψ^{-1}之前，必须将估计值$\hat{K}^\top \Psi(\bar{x}) \in \mathbb{R}^m$ 重新投影到流形 M上，见 [14], [15]。在 [9], 中，提出了一种基于提升动力学方差估计的几何再投影方法。接下来，我们将在这一概率框架的基础上，分析提升空间中预测残差的一阶和二阶矩，它们构成了第三节多步误差刻画的基础。

III. 提升空间中的基于矩的预测误差分析

现在，我们分析提升空间中的预测残差，重点关注其一阶和二阶矩，并以此作为多步误差分解的基础。为此，我们分析

$$z(t+1) = Az(t), \qquad z(0) = \Psi(\bar{x})$$

对于初始状态 $\bar{x} \in X$。需要设计的量是$A \in \mathrm{M}(m \times m, \mathbb{R})$，即提升映射 Ψ 是固定的。根据前面的讨论，合理的选择是$\hat{A} \coloneqq \hat{K}^\top$，但我们将针对一般的 A进行分析，并偶尔对具体情形 $\hat{A} = A$进行评论。因此，接下来定义的所有内容都取决于 A的选择，但为提高可读性，我们在符号中省略这种依赖关系。

我们用$z(\cdot, \bar{x})$表示提升系统的轨迹：$\mathbb{N}_0 \to \mathbb{R}^m$。特别值得关注的是，$z(\cdot, \bar{x})$对 $\Psi(x(\cdot, \bar{x}))$的近似效果，以及它相对于由训练过程中使用的样本定义的经验测度是否具有零均值性质。为此，我们首先考察单步误差，以获得结构性见解；随后将这些见解用于多步误差分析，并最终用于开发一种新颖的近似方案。

令 $\mathcal{O}$ 表示 $X \subset \mathbb{R}^n$上的子空间拓扑，$\mathcal{A} = \sigma(\mathcal{O})$表示相应的 Borel σ-代数。此外，考虑式 (2) 中的状态 $x_1, \dots, x_L$，以及与之对应的狄拉克测度 δ_i：$\mathcal{A} \to \mathbb{R}$，其定义为

$$\delta_i(B) \coloneqq \begin{cases} 1 & x_i \in B \\ 0 & \text{otherwise} \end{cases}$$

由于 $\delta_i(X) = 1$，每个狄拉克测度 δ_i 都是概率测度，因此，经验测度

$$P_L \coloneqq \frac{1}{L} \sum_{i=1}^{L} \delta_i$$

使 $(X, \mathcal{A}, P_L)$ 成为一个概率空间。如果 L 足够大，那么对于每个事件 $B \in \mathcal{A}$，概率的频率解释表明 $P_L(B) \simeq P_\star(B)$，即，P_L近似于原概率空间 $(X, \mathcal{A}, P_\star)$ 中用于抽取初始状态 x_i的相应测度 $P_\star$。因此，当我们希望计算概率测度 $P_\star$ 的某个定量性质——例如期望或方差——时，可以通过计算经验测度的该性质来近似。我们在全文中都采用这一思路。

对于每个 $\bar{x} \in X$，我们有

$$(\Psi \circ f)(\bar{x}) = A\Psi(\bar{x}) + ((\Psi \circ f)(\bar{x}) - A\Psi(\bar{x})),$$

这促使我们引入残差映射R：将 $X \to \mathbb{R}^m$ 作为

$$R(\bar{x}) \coloneqq (\Psi \circ f)(\bar{x}) - A\Psi(\bar{x}). \tag{3}$$

这描述了我们在提升空间中经过一个时间步所产生的误差。因此，我们希望选择 A，使残差映射在某种意义下“较小”；我们将利用所引入的概念，为提升动力学赋予概率解释，并据此定义一个“较小”的标准。

由于 R 关于子空间拓扑是连续的，因此它必然是 $\mathcal{A}$-可测的。这意味着，残差映射可以被视为一个取值于 $\mathbb{R}^m$ 的随机变量。下面我们将分析其一阶矩和二阶矩。

首先，我们考察残差的期望。对于 $t \in \{0, \dots, N\}$，引入算术平均值将很方便:

$$\nu_t \coloneqq \frac{1}{L} \sum_{i=1}^{L} \Psi(x(t, x_i)). \tag{4}$$

在时间 t 收集到的提升状态的算术平均值。我们可以得出以下直接推论。

引理1。 残差的一阶矩 $\mathbb{E}(R)$为零，当且仅当 $\nu_1 = A\nu_0$。

证明。 我们可以将期望表示为

$$\begin{aligned} \mathbb{E}(R) &= \int R \,\mathrm{d}P_L = \frac{1}{L} \sum_{i=1}^{L} R(x_i) = \frac{1}{L} \sum_{i=1}^{L} \Big[\Psi(y_i) - A\Psi(x_i)\Big] \\ &= \frac{1}{L} \left(\sum_{i=1}^{L} \Psi(y_i) - A \sum_{i=1}^{L} \Psi(x_i) \right) = \nu_1 - A\nu_0, \end{aligned}$$

这直接意味着该断言成立。 □

将 $\mathrm{M}(m \times m, \mathbb{R})$ 中能导出 $\mathbb{E}(R) = 0$ 的矩阵 A 所构成的子集记为 E_0，即有

$$E_0 = \{A \in \mathrm{M}(m \times m, \mathbb{R}) \mid \nu_1 = A\nu_0\}. \tag{5}$$

接下来将证明，在关于字典的一个温和且标准的假设下，通过 EDMD算法得到的预测器 $\hat{A} = \hat{K}^\top$ 属于 E_0，这一点已在 [9] 中提及。

命题1。 假设对于所选基中的某个元素 ψ_j，有 $\psi_j \equiv 1$（$\psi_1, \dots, \psi_m$），其中 V。那么，$\hat{A} \in E_0$。

证明。 令 $\mathbf{1} \in \mathbb{R}^L$ 表示全 1 向量。根据 EDMD算法的定义，$\hat{A} = G_1 G_0^\top (G_0 G_0^\top)^{-1}$，由此得到

$$\hat{A}\nu_0 = \frac{1}{L} \hat{A} G_0 \mathbf{1} = \frac{1}{L} G_1 G_0^\top (G_0 G_0^\top)^{-1} G_0 \mathbf{1}.$$

因此，如果我们能证明 $G_0^\top (G_0 G_0^\top)^{-1} G_0 \mathbf{1} = \mathbf{1}$，则由引理1可得该断言成立。

不失一般性，假设 ψ_m 是常值可观测函数，即 $\psi_m \equiv 1$。那么，数据矩阵的形式为

$$G_0 = \begin{pmatrix} \hat{G} \\ \mathbf{1}^\top \end{pmatrix},$$

ˆ其中 G表示 G_0的前 $m-1$ 行。因此，利用舒尔补

$$S_{\mathrm{c}} = L - \mathbf{1}^\top \hat{G}^\top (\hat{G}\hat{G}^\top)^{-1}\hat{G}\mathbf{1}$$

$G_0G_0^\top$中ˆGˆ$G^\top$ 的逆可写为$(G_0G_0^\top)^{-1} = V^\top HV$，其中

$$V = \begin{pmatrix} I & 0 \\ -\mathbf{1}^\top \hat{G}^\top (\hat{G}\hat{G}^\top)^{-1} & 1 \end{pmatrix}, \quad H = \begin{pmatrix} (\hat{G}\hat{G}^\top)^{-1} & \\ & S_{\mathrm{c}}^{-1} \end{pmatrix},$$

然后，经过简单计算可得

$$VG_0\mathbf{1} = \begin{pmatrix} \hat{G}\mathbf{1} \\ S_{\mathrm{c}} \end{pmatrix}, \qquad HVG_0\mathbf{1} = \begin{pmatrix} (\hat{G}\hat{G}^\top)^{-1}\hat{G}\mathbf{1} \\ 1 \end{pmatrix},$$
$$V^\top HVG_0\mathbf{1} = \begin{pmatrix} \mathbf{0} \\ 1 \end{pmatrix},$$

which implies $G_0^\top (G_0G_0^\top)^{-1}G_0\mathbf{1} = \mathbf{1}$, as required. □

总之，有限字典和有限数据下的 Koopman 压缩，在匹配预测残差的一阶和二阶矩的意义下近似真实的提升动力学。假设残差的期望为零，则对于一般的A，可以推导出残差方差的闭式表达式，该表达式关于 A呈二次关系。

推论1. 令 $A \in E_0$ 为任意值。则， R 的二阶矩 $\Sigma \coloneqq \mathrm{var}(R)$ 满足

$$\Sigma = \frac{1}{L}\left(G_1 - AG_0\right)\left(G_1 - AG_0\right)^\top.$$

此外，方差的迹为

$$\mathrm{tr}(\Sigma) = \frac{1}{L}\|G_1 - AG_0\|_{\mathrm{F}}^2.$$

证明. 利用 $A \in E_0$ 的假设，因此 $\mathbb{E}(R)$ =0，方差公式可直接由以下计算得出

$$\begin{aligned}
\Sigma &= \mathbb{E}\left(\left(R - \mathbb{E}(R)\right)\left(R - \mathbb{E}(R)\right)^\top\right) = \mathbb{E}(RR^\top) \\
&= \frac{1}{L}\sum_{i=0}^{L}\left(\Psi(y_i) - A\Psi(x_i)\right)\left(\Psi(y_i) - A\Psi(x_i)\right)^\top \\
&= \frac{1}{L}\Big(\sum_{i=0}^{L}\Psi(x_i)\Psi(y_i)^\top - \sum_{i=0}^{L}\Psi(y_i)\Psi(x_i)^\top A^\top \\
&\quad - \sum_{i=0}^{L}A\Psi(x_i)\Psi(y_i)^\top + \sum_{i=0}^{L}A\Psi(x_i)\Psi(x_i)^\top A^\top\Big) \\
&= \frac{1}{L}(G_1G_1^\top - G_1G_0^\top A^\top - AG_0G_1^\top + AG_0G_0^\top A^\top),
\end{aligned}$$

其中我们用到了 G_0 和 G_1 的定义。现在，所声称的表达式可直接由迹算子的循环性质推出。

□

IV. 多步预测与预测误差

鉴于这一结果，并假设 V 的基包含一个常值可观测函数，EDMD算法可以解释为：从E_0 中选取使 tr(var(R)) 最小的元素。请注意，这一解释依赖于 A 属于 E_0 这一事实；否则，方差的表达式会发生变化。此外，这一结果不能直接推广到多步预测。接下来我们将讨论这一局限。

对于 $t \in \{1,\dots,N\}$，我们考虑随机变量族 R_t: $X \to \mathbb{R}^m$，其中 R_t 由下式给出

$$R_t(\bar{x}) \coloneqq \Psi(x(t,\bar{x})) - z(t,\bar{x}), \tag{6}$$

描述时刻 t 提升空间中的预测误差。我们的目标是找到 A，使得对于每个 $t \in \{1,\dots,N\}$,

(i) 残差 R_t 的一阶矩 $\mathbb{E}(R_t)$ 在时刻 t 为零。(ii) 残差 R_t 在时刻t 的二阶矩 var(R_t) 最小。

考虑期望

$$\mu_t \coloneqq \mathbb{E}(R_t).$$

为了分析 μ_t 随时间的演化，对于 $t \in \{0,\dots,N-1\}$, 引入随机变量族 D_t: $X \to \mathbb{R}^m$，定义如下

$$D_t(\bar{x}) \coloneqq \Psi(x(t+1,\bar{x})) - A\Psi(x(t,\bar{x})) = R(x(t,\bar{x})), \tag{7}$$

它们描述了时刻 t 的局部误差 。残差 R_t 描述时刻 t 的整体误差，而局部误差描述在时刻 t 新增的误差；如果提升系统从正确状态$\Psi(x(t,\bar{x}))$初始化，就会产生这一误差。特别地， $D_0 = R_1 = R$ 成立，其中残差映射 R 定义于式 (3) 中。

下面的命题表明， μ_{t+1} 可以表示为 μ_t 和局部误差的期望的函数,

$$\beta_t \coloneqq \mathbb{E}(D_t).$$

这使我们能够分析残差的期望如何演变。

命题 2。 对于每个 $t \in \{1,\dots,N-1\}$，残差 R_t 的期望μ_t 满足

$$\mu_{t+1} = A\mu_t + \beta_t.$$

证明。 任取 $t \in \{1,\dots,N-1\}$，并令 $\bar{x} \in X$为任意值。注意，根据定义，我们有 $x(t+1,\bar{x}) = f(x(t,\bar{x}))$，且 $z(t+1,\bar{x}) = Az(t,\bar{x})$。因此，根据残差的定义，我们得到

$$\begin{aligned}
R_{t+1}(\bar{x}) &= \Psi(x(t+1),\bar{x}) - z(t+1,\bar{x}) \\
&= (\Psi\circ f)(x(t,\bar{x})) - Az(t,\bar{x}) \\
&\overset{(6)}{=} (\Psi\circ f)(x(t,\bar{x})) - A\left(\Psi(x(t,\bar{x})) - R_t(\bar{x})\right) \\
&= ((\Psi\circ f)(x(t,\bar{x})) - A\Psi(x(t,\bar{x}))) + AR_t(\bar{x}) \\
&\overset{(3)}{=} R(x(t,\bar{x})) + AR_t(\bar{x}) = AR_t(\bar{x}) + D_t(\bar{x}).
\end{aligned}$$

根据期望的线性性即可得出该结果。 □

这一结果有一个有趣的解释：期望残差遵循由所选预测器 A 给出的线性动力学，并在时间 t受到额外的“扰动” β_t 。这使得我们可以应用线性系统理论中的方法来分析期望残差误差。

推论2。 给定任意选择的 $A \in \mathrm{M}(m \times m, \mathbb{R})$，则

$$\mu_1 = \cdots = \mu_N = 0 \quad \Longleftrightarrow \quad \beta_0 = \cdots = \beta_{N-1} = 0$$

证明。证明主要基于以下观察：$R_1 = D_0$，由此得到 $\mu_1 = \beta_0$。于是，线性系统 $\mu_{t+1} = A\mu_t + \beta_t$ 的解对每个 $t \in \{1, \ldots, N\}$ 均成立

$$\mu_t = A^{t-1}\mu_1 + \sum_{k=1}^{t-1} A^{t-1-k}\beta_k = \sum_{k=0}^{t-1} A^{t-1-k}\beta_k,$$

由此我们可以立即看出，$\beta_0 = \cdots = \beta_{N-1} = 0$ 蕴含 $\mu_1 = \cdots = \mu_N = 0$。

对于反向，注意到当 $0 = \mu_1 = \beta_0$ 时，μ_t 的表达式对每个 $t \in \{1, \ldots, N-1\}$ 均意味着

$$0 = \mu_{t+1} = \beta_t,$$

正如所述。 □

推论2表明，当且仅当局部误差的期望为零时，总体残差在所有时刻 $t \in \{1, \ldots, N\}$ 的期望才为零。由于 β_t 对 A的线性依赖关系，我们可以用关于 A的线性条件来表示这种关系。下面的推论将引理1推广到多步情形。

推论3。 对于每个 $t \in \{0, \ldots, N-1\}$，时刻 t 的局部误差的一阶矩$\beta_t = \mathbb{E}(D_t)$ 当且仅当 $\nu_{t+1} = A\nu_t$时为零。

证明。 固定任意一个 $t \in \{0, \ldots, N-1\}$。于是

$$\beta_t = \frac{1}{L}\sum_{i=1}^{L} \Psi(x(t+1, x_i)) - A\Psi(x(t, x_i)) = \nu_{t+1} - A\nu_t,$$

断言得证。 □

因此，对于每个 $t \in \{0, \ldots, N-1\}$，我们引入集合

$$E_t = \{A \in \mathrm{M}(m \times m, \ \mathbb{R}) \mid \forall k \in \{0, \ldots, t\}\colon \ \nu_{k+1} = A\nu_k\},$$

因此，$\beta_0 = \cdots = \beta_t = 0$ 当且仅当 $A \in E_t$。注意，该定义包含我们先前定义的 E_0，并且我们有包含关系 $E_{t+1} \subseteq E_t$。此外，$A \in E_{t-1}$ 等价于 $\mu_1 = \cdots = \mu_t = 0$。

至此，我们完成了对无偏条件的理论分析。

V. 参数化与自由度

在刻画了多步残差无偏的条件之后，我们现在转向该问题的构造性方面。在本节中，我们对集合 E_{N-1} 进行参数化，识别选择 A 时可用的自由度，并利用这些自由度设计一个在降低方差的同时强制无偏性的预测器。

A. 自由度

为此，我们定义矩阵

$$B = \begin{pmatrix} \nu_0 & \ldots & \nu_{N-1} \end{pmatrix}, \qquad C = \begin{pmatrix} \nu_1 & \ldots & \nu_N \end{pmatrix}.$$

假设 $m \geq N$ 并取QR分解[2]

$$B = Q\bar{R} = \begin{pmatrix} Q_1 & Q_2 \end{pmatrix} \begin{pmatrix} \hat{R} \\ 0 \end{pmatrix} = Q_1\hat{R},$$

其中 $\hat{R} \in \mathrm{M}(N \times N, \ \mathbb{R})$。

命题 3. 假设 $m \geq N$。那么，$\mu_1 = \cdots = \mu_N = 0$ 当且仅当 $A = CR^{-1}Q_1^\top + A_2Q_2^\top$ 对于某个$A_2 \in \mathrm{M}(m \times (m-N)\ \mathbb{R})$。

证明。 结合推论2和推论3，并使用新定义的矩阵 B和 C，可知 $\mu_1 = \cdots = \mu_N = 0$ 等价于

$$A \in E_{N-1} \qquad \Longleftrightarrow \qquad AB = C.$$

因此，我们需要对满足$AB = C$的矩阵 A进行参数化。由于 Q 是正交的，我们可以将任意$A \in \mathrm{M}(m \times m, \mathbb{R})$分解为 $A = A_1Q_1^\top + A_2Q_2^\top$，其中令$A_1 = AQ_1$ 和 $A_2 = AQ_2$。该分解给出

$$AB = A_1Q_1^\top Q_1\hat{R} + A_2Q_2^\top Q_1\hat{R} = A_1\hat{R},$$

因此，$AB = C$ 当且仅当 $A = A_1Q_1^\top + A_2Q_2^\top$ 且 $A_1 = C\hat{R}^{-1}$。 □

分解 $A = A_1Q_1^\top + A_2Q_2^\top$ 揭示了在选择 $A \in E_{N-1}$时可用的自由度。在这个基下，必须 选择 $A_1 = C\hat{R}^{-1}$，其余自由度由 A_2给出。

B. 方差最小化

上一小节的分析得出了数值上易于处理的条件，以确保 $A \in E_{N-1}$。具体而言，残差 $\mathbb{E}(R_t)$ 的一阶矩为零。接下来，我们希望利用剩余的自由度来满足第二个设计准则，即最小化二阶矩 $\mathrm{var}(R_t) =: \Sigma_t$。

我们在开篇已经看到，对于 $N = 1$，EDMD预测器可实现这一点；现在我们希望将这一过程推广。具体而言，我们希望推导出一个闭式解，以便随后通过算法加以利用。

为研究时间 t处的方差 Σ_t ，我们首先描述其在 $A \in E_{N-1}$时随时间的演化。由于$R_1 = R$， Σ_1 的表达式由推论1给出。对于未来的时间步，我们依据局部误差 D_t。为此，记

$$\Omega_t \coloneqq \mathrm{var}(D_t), \qquad \Gamma_t \coloneqq \mathbb{E}(R_tD_t^\top),$$

D_t 的方差以及 R_t 和 D_t的协方差。局部误差的方差有一个简单的表达式。

[2]为简化起见，我们在此假设 B 具有满秩，但这一假设可以推广。

推论 4. 若 $A \in E_{N-1}$，则对于每个 $t \in \{0,\dots,N-1\}$，D_t 的方差Ω_t 为

$$\Omega_t = \frac{1}{L}(G_{t+1} - AG_t)(G_{t+1} - AG_t)^\top,$$

带迹

$$\operatorname{tr}(\Omega_t) = \frac{1}{L}\|G_{t+1} - AG_t\|_{\mathrm{F}}^2$$

证明。 该假设意味着 $\mathrm{var}(D_t) = \mathbb{E}(D_t D_t^\top)$，因此可以按照推论1中的相同步骤证明 Ω_t 的表达式和迹。 □

与一阶矩的演化类似， R_t 的二阶矩也可以用其前一时刻的方差和单步方差来表示。不过，还必须加上一个协方差项。

命题 4. 令 $A \in E_{N-1}$。那么，

$$\Sigma_{t+1} = A\Sigma_t A^\top + A\Gamma_t + \Gamma_t^\top A^\top + \Omega_t,$$

对于每个 $t \in \{1,\dots,N-1\}$。

证明。 任取 $t \in \{1,\dots,N-1\}$。根据推论2，该假设 $A \in E_{N-1}$ 尤其意味着 $\mu_{t+1} = 0, \mu_t$ =0 和 $\beta_t = 0$。因此，方差可计算为 Σ_{t+1} =$\mathbb{E}(R_{t+1}R_{t+1}^\top)$。此外，根据命题 2 的证明，我们有 $R_{t+1} = AR_t + D_t$。结合这些可得

$$\begin{aligned}\Sigma_{t+1} &= \mathbb{E}((AR_t + D_t)(AR_t + D_t)^\top) \\ &= A\mathbb{E}(R_t R_t^\top)A^\top + A\mathbb{E}(R_t D_t^\top) + \mathbb{E}(D_t R_t^\top)A^\top \\ &\quad + \mathbb{E}(D_t D_t^\top).\end{aligned}$$

其中我们用到了期望的线性性。由于 μ_t =$\beta_t = 0$，因此我们有 $\mathbb{E}(R_t R_t^\top) = \mathrm{var}(R_t)$ 和 $\mathbb{E}(D_t D_t^\top)$ =$\mathrm{var}(D_t)$，由此得出 Σ_{t+1}的表达式。 □

因此， Σ_{t+1} 的方差由三部分组成：

(i) 最后一步的前向传播方差；由 $A\Sigma_t A^\top$ 给出。 (ii) AR_1 和 D_1的协方差；由 $A\Gamma_1$+$\Gamma^\top A^\top$给出。(iii) D_t 的方差；由 Ω_t 给出。

在这些项中，Ω 是唯一关于A呈二次关系的项，另外两项对 A 的依赖关系更为复杂，这使得对 Σ_{t+1} 进行精确分析十分繁琐。为了解决这一问题，我们提出对真实方差进行近似，从而简化分析。

为此，我们注意到以下理想化结果。

引理 2。 设 $A \in E_{N-1}$和 $t \in \{1,\dots,N\}$。若

$$\forall k \in \{0,\dots,t-2\}\colon \operatorname{tr}(\Omega_k) = 0$$

成立，则 Σ_t =Ω_{t-1}。

证明。 任取 $t \in \{1,\dots,N\}$。由于 $A \in E_{N-1}$，方差 Σ_t 计算为

$$\begin{aligned}\Sigma_t = \mathbb{E}(R_t R_t^\top) &= \frac{1}{L}\sum_{i=1}^{L} R_t(x_i)R_t(x_i)^\top \\ &= \left(G_t - A^t G_0\right)\left(G_t - A^t G_0\right)^\top.\end{aligned}$$

根据推论 4，所作的假设意味着

$$\forall k \in \{0,\dots,t-2\}\colon G_{k+1} = AG_k,$$

由此我们得出

$$A^t G_0 = A^{t-1}AG_0 = A^{t-2}AG_1 = \cdots = AG_{t-1}.$$

将该等式代入 Σ_t 的表达式即可直接得到这一断言。 □

备注 1。 鉴于命题 4，对于 $t > 1$ ，引理 2 中的假设等价于

$$A\Sigma_{t-1}A^\top + A\Gamma_{t-1} + \Gamma_{t-1}^\top A^\top = 0.$$

对此进行详细阐述超出了本文的范围，但对传播方差 $A\Sigma_t A^\top$ 与协方差 $A\Gamma_t$ +$\Gamma_t^\top A^\top$ 之间的关系及其演化进行更深入的分析，是未来研究中一个有趣的问题。

原则上，因此可以在满足 $\mathrm{tr}(\Omega_t) = 0$for $t < N-1$ 的约束下，通过最小化 $\mathrm{tr}(\Omega_{N-1})$ 来最小化 $\mathrm{tr}(\Sigma_N)$。然而，对于方差而言，通常即使是对单个时间步，也无法实现这一约束，因为 A没有提供足够的自由度。因此，我们不能指望对任何 $t > 0$都能精确地 满足引理 2 中的假设。

然而，根据引理 2，我们仍可以认为，只要对于 $k \in \{0,\dots,t-2\}$，$\mathrm{tr}(\Omega_k)$“足够”小，那么近似 $\Sigma_t \approx \Omega_{t-1}$ 就是合理的。因此我们得出结论，获得 A的一种可能方法是同时最小化

$$\frac{1}{L}\|G_1 - AG_0\|^2 \qquad \dots \qquad \frac{1}{L}\|G_N - AG_{N-1}\|^2$$

并满足约束 $A \in E_{N-1}$。这可以看作是（近似地）最小化 $\mathrm{tr}(\Sigma_1)\dots,\mathrm{tr}(\Sigma_N)$。

VI. 预测器设计

我们终于可以介绍一种用于 A 的设计流程。首先，我们要求满足 $A \in E_{N-1}$；根据命题 3，这对于 $m > N$ 很容易实现：令 $A = A_1 Q_1^\top + A_2 Q_2^\top$，其中 $A_1 = C\hat{}R^{-1}$。

第二个要求是最小化 $\mathrm{tr}(\Sigma_N)$，这基于前面的讨论。具体而言，我们寻找一个 A_2，使得对所有 $t \in \{1,\dots,N-1\}$

$$\frac{1}{L}\|G_{t+1} - AG_t\|_{\mathrm{F}}^2 \approx 0.$$

这确保近似 $\Sigma_t \approx \Omega_{t-1}$ 是合理的，因此也能直接最小化 $\mathrm{tr}(\Sigma_t)$ $\approx$$\mathrm{tr}(\Omega_{t-1})$。

然而，对 A_2 的这一要求并不容易实现，因为各个单个目标函数 $\|G_{t+1} - AG_t\|_{\mathrm{F}}^2$彼此竞争。在本工作中，我们建议选择$A$ ，以使简单的加性目标函数最小化

$$\sum_{t=0}^{N-1}\|G_{t+1} - AG_t\|_F^2 \tag{8}$$

在 E_{N-1}上。凭借 A的显式参数化，这可以转化为一个无约束最小二乘问题。引入聚合的数据矩阵

$$G = \begin{pmatrix} G_0 & \dots & G_{N-1} \end{pmatrix}, \quad S = \begin{pmatrix} G_1 & \dots & G_N \end{pmatrix}, \quad (9)$$

将目标函数 (8) 转化为

$$\|S - AG\|_{\mathrm{F}}^2 = \|(S - A_1 Q_1^\top G) - A_2 Q_2 G\|_{\mathrm{F}}^2 =: \|\tilde{S} - A_2 \tilde{G}\|_{\mathrm{F}}^2.$$

假设 G 满行秩，则唯一的最小二乘解为 $A_2 = \tilde{S}\tilde{G}^\top (\tilde{G}\tilde{G}^\top)^{-1}$，由此可以得到我们的所提出的预测器

$$A = C\hat{R}^{-1}Q_1^\top + \tilde{S}\tilde{G}^\top(\tilde{G}\tilde{G}^\top)^{-1}Q_2^\top. \quad (10)$$

尽管这一过程在概念上很简单，但它有几个优点。首先，它保证了 $\mu_1 = \cdots = \mu_N = 0$，从而确保预测器在整个预测时域内无偏。其次，由此得到的无偏性使得计算方差时无需考虑期望。第三，该预测器通过求解凸优化问题得到，从而保证全局最优性。最后，最小二乘问题的最优值为每个 tr(Ω$_t$) 提供了一个（粗略的）上界；如果这个上界过大，则可以相应地缩短预测时域 N。

需要注意的是，所提出的方法并不等同于对聚合数据矩阵 G 和 S 应用 EDMD算法。在 EDMD 中，预测器将由 $\bar{A} = SG^\top(GG^\top)$ 给出，但这不会导致 $\mu_1 = \cdots = \mu_N = 0$。

VII. 数值示例

我们考虑无外力、无阻尼的杜芬振子$\ddot{x}(t) = x(t) - x(t)^3$。我们使用状态空间表示$x_1 = x$ 和 $x_2 = \dot{x}$，并以采样时间 $\Delta t = 0.1$对连续动力学进行采样。我们从 $X = [-1, 1]^2$ 中均匀抽取 $L = 10,000$个状态，并据此向前模拟$N = 5$ 步。可观测量（$\psi_1, \dots, \psi_m$）对应于最高十次的单项式，由此得到$m = 66$。此设置定义了数据矩阵 $G_0, \dots, G_5$。

我们考虑三个预测器：

(i) 基于 G_1 和 G_2的原始EDMD预测器，即$\hat{A} = G_1 G_0^\top (G_0 G_0^\top)^{-1}$。(ii) 使用式（9）中聚合数据矩阵的EDMD预测器，即 $\bar{A} = SG^\top(GG^\top)^{-1}$。(iii) 我们在式（10）中提出的预测器 A。

首先，来看图1。它展示了期望残差（由经验测度衡量）偏离零的程度。可以看到，我们的预测器确实如预期般无偏。单步 EDMD预测器 $\bar{A}$ 产生的期望残差误差最高。这是意料之中的，因为训练 $\bar{A}$时没有使用任何多步信息。聚合预测器 $\tilde{A}$所产生残差的期望较小——它确实能够利用多步信息——但该预测器仍然有偏。这凸显了本研究所采用的分析与合成的重要性。

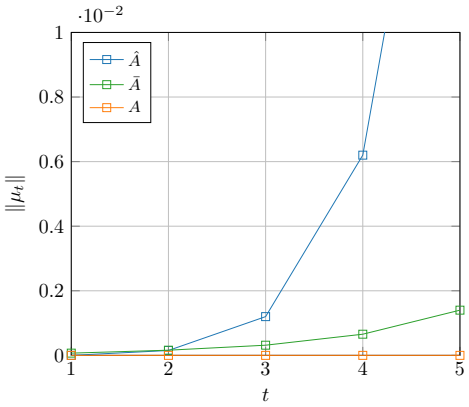

图1：三个预测器的残差期望随时间的演化（$\hat{A}$：单步 EDMD，$\bar{A}$：聚合 EDMD，A：所提出的无偏预测器）。

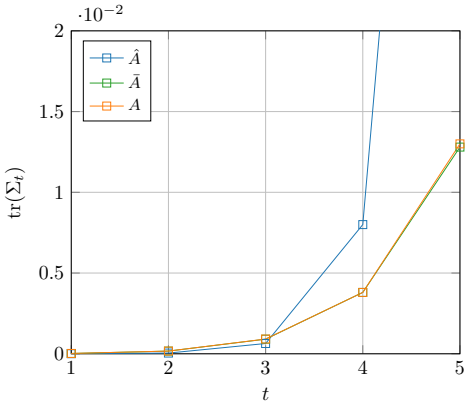

图2：三个预测器的残差方差随时间的演化（$\hat{A}$：单步 EDMD，$\bar{A}$：聚合 EDMD，A：所提出的无偏预测器）。

再来看图2，可以看到三个预测器的方差如何随时间变化。与之前一样，预测器 $\hat{A}$的表现最差，这在意料之中。需要注意的关键点是，所提出的预测器 A与 $\bar{A}$的方差相匹配。因此，相应的残差围绕各自期望值的波动幅度相同。然而，基于 $\bar{A}$的预测更容易解释，因为残差围绕零波动，而基于 $\hat{A}$ 的预测则存在偏差。因此，只要字典——以及 $\hat{A}$ 中的自由度——足够大，我们就不必以更高的方差为代价来获得无偏性 A。

第八节：结论

准确的长时域预测对于数据驱动建模与控制至关重要，但基于 Koopman 算子的预测器往往存在结构性偏差，并且这种偏差会随时间累积。这凸显了采用有原则的方法来确保无偏性并控制方差的必要性。

构建了一种多步预测器，可保证无偏性，并在 Koopman 提升空间中实现低方差。对残差进行概率刻画，可得出实现无偏

多步预测的结构条件，并揭示预测器设计的可用自由度。随后利用这些自由度，通过凸优化过程确保无偏性并最小化多步方差。

数值研究证实了所得预测器的有效性。当提升空间具有足够的表达能力时，该方法能够达到与现有基于扩展动态模态分解的方法相当的方差水平，同时还确保在整个预测时域内保持无偏。

所提出的框架为利用有限数据构建可靠的多步Koopman预测器提供了一种有理论依据的方法。它为不确定性感知预报以及兼具学习能力与基于模型的控制奠定了基础；在这些领域，无偏预测和一致的方差传播至关重要。

参考文献

[1] M. O. Williams、I. G. Kevrekidis 和 C. W. Rowley，“Koopman算子的一个数据驱动近似：扩展动态模态分解”，《非线性科学杂志》，第 25 卷，页 1307–1346，2015。[2] M. J. Colbrook，“动态模态分解算法的多重宇宙”，载于《数值分析手册》，第 25 卷，页 127–230，爱思唯尔，2024。[3] I. Mezi´c，“动力系统的谱性质、模型降阶与分解”，《非线性动力学》，第 41 卷，325，2005 1，309 期，页 –。[4] C. W. Rowley、I. Mezi´c、S. Bagheri、P. Schlatter 和 D. S. Henningson，“非线性流动的谱分析”，《流体力学杂志》，第 641 卷，页 115–127，2009。[5] R. Str¨asser、K. Worthmann、I. Mezi´c、J. Berberich、M. Schaller 和 F. Allg¨ower，“基于Koopman的控制综述：从误差界到闭环保证”，《控制年评》，第 61 卷，页 101035，2026。[6] M. Korda 和 I. Mezi´c，“扩展动态模态分解向Koopman算子的收敛性”，《非线性科学杂志》，第 28 卷，第 2 期，页 687–710，2018。[7] F. M. Philipp、M. Schaller、S. Boshoff、S. Peitz、F. N¨uske 和 K. Worthmann，“Koopman算子数据驱动近似的方差表示与收敛速率”，arXiv预印本 arXiv:2402.02494，2024。[8] F. K¨ohne、F. M. Philipp、M. Schaller、A. Schiela 和 K. Worthmann，“L^∞-核扩展动态模态分解对Koopman算子近似的误差界”，《SIAM应用动力系统杂志》，第 24 卷，第 1 期，页 501–529，2025。[9] P. van Goor、R. Mahony、M. Schaller 和 K. Worthmann，“用于可靠Koopman预测及参数动力系统分岔分析的最大似然再投影”，arXiv预印本 arXiv:2506.17817，2025。[10] R. Schurig、P. van Goor、K. Worthmann 和 R. Findeisen，“通过在格拉斯曼流形上学习来塑造Koopman字典”，arXiv预印本 arXiv:2511.07234，2025。[11] I. Mezi´c，“通过Koopman算子的谱性质分析流体流动”，《流体力学年评》，第 45 卷，第 1 期，页 357–378，2013。[12] A. Mauroy 和 I. Mezi´c，“利用Koopman算子的特征函数进行全局稳定性分析”，《IEEE自动控制汇刊》，第 61 卷，第 11 期，页 3356–3369，2016。[13] M. Haseli 和 J. Cort´es，“从数据中学习Koopman特征函数与不变子空间：对称子空间分解”，《IEEE自动控制汇刊》，第 67 卷，第 7 期，页 3442–3457，2021。[14] A. Mauroy 和 J. Goncalves，“用于非线性系统辨识的基于Koopman的提升技术”，《IEEE自动控制汇刊》，第 65 卷，第 6 期，页 2550–2565，2019。[15] P. van Goor、R. Mahony、M. Schaller 和 K. Worthmann，“用于基于Koopman的建模与预测的再投影方法”，载于第 62 届 IEEE决策与控制会议（CDC），页 315–321，2023。